

**UNIVERSIDADE DE BRASÍLIA  
FACULDADE DE TECNOLOGIA  
DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

**UM MODELO DE VISUALIZAÇÃO DE DADOS UTILIZANDO  
BANCO DE DADOS ORIENTADO A GRAFO SUPORTADO  
POR BIG DATA**

**GUSTAVO HENRIQUE MOREIRA ALVARES DA SILVA**

**ORIENTADOR: FLÁVIO ELIAS GOMES DE DEUS**

**DISSERTAÇÃO DE MESTRADO EM ENGENHARIA ELÉTRICA  
ÁREA DE CONCENTRAÇÃO INFORMÁTICA FORENSE E  
SEGURANÇA DA INFORMAÇÃO**

**PUBLICAÇÃO: PPGENE.DM – 635/2016**

**BRASÍLIA / DF: JULHO/2016**



**UNIVERSIDADE DE BRASÍLIA  
FACULDADE DE TECNOLOGIA  
DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

**UM MODELO DE VISUALIZAÇÃO DE DADOS  
UTILIZANDO BANCO DE DADOS ORIENTADO A GRAFO  
SUPPORTADO POR BIG DATA**

**GUSTAVO HENRIQUE MOREIRA ALVARES DA SILVA**

DISSERTAÇÃO DE MESTRADO SUBMETIDA AO DEPARTAMENTO DE ENGENHARIA ELÉTRICA DA FACULDADE DE TECNOLOGIA DA UNIVERSIDADE DE BRASÍLIA, COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE PROFISSIONAL EM INFORMÁTICA FORENSE E SEGURANÇA DA INFORMAÇÃO.

APROVADA POR:

---

**FLÁVIO ELIAS GOMES DE DEUS, Dr., ENE/UnB  
(ORIENTADOR)**

---

**ROBSON DE OLIVEIRA ALBUQUERQUE, Dr., ENE/UnB  
(EXAMINADOR INTERNO)**

---

**MARCELO STEHLING DE CASTRO, Dr., EEEEC/UFG  
(EXAMINADOR EXTERNO)**

**DATA: BRASÍLIA/DF, 07 DE JULHO DE 2016.**



## FICHA CATALOGRÁFICA

SILVA, GUSTAVO H. M. A.

Um modelo de visualização de dados utilizando banco de dados orientado a grafo suportado por *big data* [Distrito Federal] 2016.

xxiii, 150p., 297 mm (ENE/FT/UnB, Mestre, Engenharia Elétrica, 2016).

Dissertação de Mestrado – Universidade de Brasília, Faculdade de Tecnologia. Departamento de Engenharia Elétrica.

1. Investigação policial 2. *Big data*  
3. Visualização de dados 4. *NoSQL*

I. ENE/FT/UnB. II. Título (Série)

## REFERÊNCIA BIBLIOGRÁFICA

SILVA, G. H. M. A. (2016). Um modelo de visualização de dados utilizando banco de dados orientado a grafo suportado por big data. Dissertação de Mestrado, Publicação PPGENE.DM – 635/2016, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 150p.

## CESSÃO DE DIREITOS

NOME DO AUTOR: Gustavo Henrique Moreira Alvares da Silva

TÍTULO DA DISSERTAÇÃO: Um modelo de visualização de dados utilizando banco de dados orientado a grafo suportado por *big data*.

GRAU/ANO: Mestre/2016.

É concedida à Universidade de Brasília permissão para reproduzir cópias desta Dissertação de Mestrado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. Do mesmo modo, a Universidade de Brasília tem permissão para divulgar este documento em biblioteca virtual, em formato que permita o acesso via redes de comunicação e a reprodução de cópias, desde que protegida a integridade do conteúdo dessas cópias e proibido o acesso a partes isoladas desse conteúdo. O autor reserva outros direitos de publicação e nenhuma parte deste documento pode ser reproduzida sem a autorização por escrito do autor.

---

Gustavo Henrique Moreira Alvares da Silva  
Universidade de Brasília  
Campus Universitário Darcy Ribeiro – Asa Norte  
CEP 70.910-900 – Brasília – DF - Brasil



Este trabalho é dedicado à minha  
querida esposa, Raquel, pelo  
companheirismo e incentivo.



## **AGRADECIMENTOS**

Ao meu orientador Prof. Dr. Flávio Elias Gomes de Deus, pelo constante apoio e dedicação para o desenvolvimento desta pesquisa.

Ao Prof. Luciano Caixeta Moreira pelo seu domínio prático sobre o tema e contribuição com as ideias e diretrizes iniciais deste trabalho.

Ao Perito Criminal Federal Hέλvio Pereira Peixoto, e demais envolvidos, pelo empenho para que este curso de Mestrado se viabilizasse.

Aos amigos Vladimir de Paula Brito, Rodrigo Benjamim Gesteira e Rodrigo Fileto Cuerci Maciel pelo incentivo à busca pelo aperfeiçoamento profissional e acadêmico.

O presente trabalho foi realizado com o apoio do Departamento Polícia Federal – DPF com recursos da Secretaria Nacional de Segurança Pública - SENASP, do Ministério da Justiça.



## **RESUMO**

### **UM MODELO DE VISUALIZAÇÃO DE DADOS UTILIZANDO BANCO DE DADOS ORIENTADO A GRAFO SUPORTADO POR BIG DATA**

**Autor: Gustavo Henrique Moreira Alvares da Silva**

**Orientador: Flavio Elias Gomes de Deus**

**Programa de Pós-graduação em Engenharia Elétrica**

**Brasília, julho de 2016**

Este trabalho propõe um modelo de plataforma de dados baseado em ferramentas big data voltado para o processo de investigação criminal das polícias judiciárias brasileiras.

Os organismos de investigação criminal necessitam buscar soluções frente aos desafios tecnológicos que lhes são regularmente impostos, particularmente em relação ao novo contexto informacional, em que variáveis como volume, velocidade, variedade, validade, veracidade e volatilidade dos dados que aportam nessas organizações impactam diretamente suas atividades. Sabe-se que os bancos de dados relacionais não foram arquitetados para lidar com esse novo paradigma informacional, sendo necessário, portanto, a busca por outras alternativas.

A solução apresentada utiliza um banco de dados orientado a grafos, suportado por big data, que é capaz de melhorar o trabalho investigativo e pericial, à medida em que suporta visualização e análise de vínculo de milhares de entidades e seus relacionamentos. A solução também contempla um cluster de máquinas baseado em sistema de arquivos distribuído, no qual é possível aplicar variados tipos de operações sobre um conjunto de dados semiestruturadas.

Neste estudo é apresentado um cenário hipotético de investigação criminal que foi utilizado para testar e demonstrar os benefícios da plataforma proposta.

## **ABSTRACT**

### **A MODEL OF DATA PLATFORM BASED ON BIG DATA AS SUPPORT FOR POLICE INVESTIGATION**

**Author: Gustavo Henrique Moreira Alvares da Silva**

**Supervisor: Flavio Elias Gomes de Deus**

**Programa de Pós-graduação em Engenharia Elétrica**

**Brasília, July of 2016**

This work proposes a data platform based on big data tools, geared towards the criminal investigation processes of the Brazilian law enforcement agencies.

Este trabalho propõe um modelo de plataforma de dados baseado em ferramentas big data voltado para o processo de investigação criminal das polícias judiciárias brasileiras.

Criminal investigation agencies need to find solutions to the technological challenges often imposed on them, particularly due to new information context, in which variables as volume, velocity, variety, validity, veracity and volatility directly impact their activities. It is well known that relational databases management systems were not designed to deal with this new informational paradigm. Therefore, the search for viable alternatives is needed.

The solution proposed in this thesis uses a graph-oriented database, supported by big data, which is capable of improving the investigative and forensic tasks, as in that supports visualization and link analysis of thousands of entities and their relationships. The solution also includes a cluster of machines based on distributed file system in which it is possible to apply various types of operations on the mass of semi-structured data.

As a proof of concept, it is presented a hypothetical investigation scenario to test and demonstrate the benefits of the proposed data platform.





# SUMÁRIO

<b>1. INTRODUÇÃO .....</b>	<b>25</b>
1.1. JUSTIFICATIVA / HIPÓTESE .....	25
1.2. OBJETIVOS .....	26
1.3. METODOLOGIA DE TRABALHO .....	27
1.4. RESULTADOS ESPERADOS .....	28
1.5. ORGANIZAÇÃO DO TEXTO .....	28
<b>2. ESTUDO BIBLIOGRÁFICO – ESTADO DA ARTE .....</b>	<b>31</b>
2.1. O TERMO BIG DATA .....	31
2.2. A EXPLOSÃO INFORMACIONAL .....	32
2.3. BANCO DE DADOS RELACIONAIS x NOSQL .....	36
2.4. O SISTEMA DE ARQUIVOS HDFS (HADOOP DISTRIBUTED FILE SYSTEM) .....	38
2.5. O MODELO INFORMACIONAL BASEADO EM BIG DATA .....	39
2.6. BIG DATA NA SEGURANÇA PÚBLICA .....	42
<b>3. A POLÍCIA INVESTIGATIVA NO BRASIL E SUAS FONTES DE INFORMAÇÃO .....</b>	<b>47</b>
3.1. ESTRUTURA E ORGANIZAÇÃO DAS POLÍCIAS JUDICIÁRIAS .....	47
3.2. FONTES DE INFORMAÇÃO .....	50
3.2.1. Processo tradicional de investigação .....	50
3.2.2. Sistemas de informações transacionais .....	51
3.2.3. Laudos periciais e relatórios de identificação .....	51
3.2.4. Tabelas ou planilhas oriundas de medidas cautelares .....	51
3.2.5. Interceptação telefônica .....	52
3.2.6. Interceptação telemática .....	52
3.2.7. Vigilância a alvos .....	52
3.2.8. Fontes abertas .....	53
3.3. APLICAÇÃO DAS FONTES DE DADOS NO MODELO INFORMACIONAL .....	53
3.4. APLICAÇÃO DE CARACTERÍSTICAS DO MODELO INFORMACIONAL NA ESTRUTURA DE INVESTIGAÇÃO .....	55
3.5. O FORMATO DE DADOS DA INVESTIGAÇÃO NA POLÍCIA FEDERAL .....	57
3.5.1. Informações cartorárias .....	57
3.5.2. Base de dados de passaporte .....	58
3.5.3. Laudos periciais .....	59
3.5.4. Impressões digitais .....	60
3.5.5. Informações bancárias .....	61

3.5.6. Interceptação telefônica .....	62
3.5.7. Interceptação telemática.....	62
3.5.8. Dados das redes sociais.....	63
3.5.9. Mineração na Web .....	64
3.5.10. Dados cadastrais diversos .....	65
3.5.11. Atividades operacionais .....	65
<b>4. A PLATAFORMA BIG DATA X INVESTIGAÇÃO.....</b>	<b>68</b>
4.1. TIPOS DE BANCO DE DADOS NoSQL.....	68
4.1.1. Banco de Dados Chave-Valor.....	69
4.1.2. Banco de Dados de Documento .....	70
4.1.3. Banco de Dados de Família de Colunas.....	72
4.1.4. Banco de Dados de Grafo .....	73
4.2. A PLATAFORMA DE DADOS INVESTIGATIVA BASEADA EM NoSQL.....	75
4.2.1. Suporte ao relacionamento entre entidades.....	77
4.2.2. Desempenho na consulta das entidades e seus relacionamentos .....	78
4.2.3. Escalabilidade dos dados .....	79
4.3. UTILIZAÇÃO DO BANCO DE DADOS DE GRAFO NA SOLUÇÃO .....	80
4.4. UTILIZAÇÃO DO HDFS NA SOLUÇÃO .....	81
4.5. A PLATAFORMA PROPOSTA .....	82
4.5.1. Definição dos Insumos Informativos.....	84
4.5.2. Inserção dos Dados Catalogáveis no Banco de Dados de Grafos.....	85
4.5.3. Inserção de Dados Semiestruturados no Cluster Hadoop.....	87
4.5.4. Realização de Operações MapReduce .....	87
<b>5. APLICAÇÃO DA PLATAFORMA DE DADOS NUM EXEMPLO DE INVESTIGAÇÃO.....</b>	<b>90</b>
5.1. DEFINIÇÃO DO INSUMO INFORMATIVO .....	90
5.1.1. Atividades operacionais .....	92
5.1.2. Interceptação telefônica .....	93
5.1.3. Interceptação telemática.....	94
5.1.4. Dados bancários .....	95
5.2. INSERÇÃO DOS DADOS CATALOGÁVEIS NO BANCO DE DADOS ORIENTADO A GRAFOS .....	96
5.2.1. Inserção dos investigados .....	98
5.2.2. Inserção dos aparelhos telefônicos e sua vinculação com os investigados.....	98
5.2.3. Inserção das contas de e-mail e sua vinculação com os investigados.....	99

5.2.4. Inserção das contas bancárias .....	100
5.2.5. Inserção de encontros entre investigados filmados pela equipe operacional....	101
5.2.6. Inserção das chamadas telefônicas oriunda da quebra do sigilo telefônico .....	102
5.2.7. Inserção da movimentação financeira do grupo oriunda da quebra do sigilo bancário: .....	103
5.2.8. Inserção de mensagem eletrônica oriunda da interceptação telemática:.....	104
5.2.9. Exploração da Base de Dados .....	106
5.3. INSERÇÃO DOS DADOS SEMIESTRUTURADOS NO CLUSTER HADOOP ....	113
5.4. REALIZAÇÃO DE OPERAÇÕES MAPREDUCE.....	114
<b>6. CONCLUSÕES.....</b>	<b>120</b>
6.1. TRABALHOS FUTUROS .....	121



## LISTA DE TABELAS

Tabela 2.1: Quantidade de Usuários nas Redes Sociais.....	34
Tabela 3.1: Legislação aplicada à organização das polícias judiciárias.....	48
Tabela 3.2: Exemplo de rotinas realizadas com implantação do big data.....	54
Tabela 3.3: Aplicação das características do modelo big data na investigação .....	55
Tabela 3.4: Formato dos dados dos insumos informacionais do DPF .....	66
Tabela 4.1: Vantagens e desvantagens dos tipos de bases de dados NoSQL.....	77
Tabela 5.1: Registro de Atividade Operacional .....	93
Tabela 5.2: Registro de Interceptação Telefônica.....	94
Tabela 5.3: Registro de Interceptação Telefônica.....	95
Tabela 5.4: Registro de transação financeira .....	96
Tabela 5.5: Informações de Investigados .....	98
Tabela 5.6: Informações de linhas telefônicas .....	98
Tabela 5.7: Informações de contas de e-mail.....	99
Tabela 5.8: Informações de contas bancária .....	100
Tabela 5.9: Informações sobre encontros.....	101
Tabela 5.10: Informações sobre chamadas telefônicas .....	102
Tabela 5.11: Informações sobre movimentação financeira.....	103
Tabela 5.12: Informações de mensagens eletrônicas (e-mails).....	104
Tabela 5.13: Operações MapReduce passíveis de serem aplicadas em fluxos de dados não estruturados. ....	118

## LISTA DE FIGURAS

Figura 2.1: Quantidade de usuários do Facebook (2005 - 2014) .....	34
Figura 2.2: Modelo Informacional do big data .....	40
Figura 3.1: Organograma genérico das instituições policiais estudadas .....	49
Figura 4.1: Exemplo de Registro no Banco de Dados de Chave-Valor .....	69
Figura 4.2: Exemplo de Registro no Banco de Dados de Documentos .....	71
Figura 4.3: Exemplo de Registros no Banco de Dados de Família de Colunas .....	72
Figura 4.4: Exemplo de Registros no Banco de Dados Orientado a Grafos .....	74
Figura 4.5: Escalabilidade x Complexidade e Flexibilidade dos bancos de dados NoSQL (adaptada de Schmutz (2013)) .....	79
Figura 4.6: Plataforma de big data para investigação policial .....	83
Figura 4.7: Exemplo de processamento MapReduce .....	89
Figura 5.1: Processo de incorporação de fontes de informação numa investigação .....	92
Figura 5.2: Inserção de investigados .....	98
Figura 5.3: Inserção de linhas telefônicas .....	99
Figura 5.4: Inserção de contas de e-mail .....	100
Figura 5.5: Inserção de contas de contas bancárias .....	101
Figura 5.6: Inserção de um encontro de investigados .....	102
Figura 5.7: Inserção de chamadas telefônicas .....	103
Figura 5.8: Inserção da movimentação bancária .....	104
Figura 5.9: Inserção da troca de e-mails entre investigados .....	104
Figura 5.10: Diagrama de relacionamento entre investigados, contas de e-mail, contas bancárias e telefones da operação policial .....	106
Figura 5.11: Consulta referente à rede de encontros entre investigados .....	107
Figura 5.12: Rede de encontro entre investigados .....	107
Figura 5.13: Consulta referente a chamadas telefônicas, e-mails e encontros de investigados .....	108

Figura 5.14: Chamadas telefônicas, e-mails e encontros de investigados .....	109
Figura 5.15: Consulta referente a informações bancárias de investigados .....	110
Figura 5.16: Informações bancárias de investigados .....	110
Figura 5.17: Consulta referente ao fluxo financeiro ente contas bancárias .....	111
Figura 5.18: Fluxo financeiro ente contas bancárias .....	112
Figura 5.19: Transferência de arquivo para o HDFS .....	114
Figura 5.20: Fluxo de processamento de arquivos não estruturados.....	116

## **LISTA DE SÍMBOLOS, NOMENCLATURA E ABREVIACÕES**

ACID	Atomicity, Consistency, Isolation, Durability
AFIS	Automated Fingerprint Identification System
ANP	Academia Nacional de Polícia
ANSI	American National Standards Institute
API	Application Programming Interface
ASCII	American Standard Code for Information Interchange
BLOB	Binary Large Object
CPF	Cadastro de Pessoa Física
DPF	Departamento de Polícia Federal
ePOOLICE	Early Pursuit Against Organized Crime Using Environmental Scanning, the Law and Intelligence Systems
ERB	Estação Rádio Base
FBI	Federal Bureau of Investigation
GFS	Google File System
HDFS	Hadoop Distributed File System
IDC	International Data Corporation
IP	Internet Protocol
JSON	JavaScript Object Notation
M2M	Machine to Machine
NIST	National Institute of Standards and Technology
NoSQL	Not Only SQL
ODM	Oracle Data Mining
PCAP	Packet Capture
PGR	Procuradoria-Geral da República
RAM	Random Access Memory
SGBDR	Sistema Gerenciador de Banco de Dados Relacional
SIMBA	Sistema de Investigação de Movimentação Bancária
SIS	Sistema de Interceptação de Sinais
SISCART	Sistema de Controle de Atividades Cartorárias
SMS	Short Message Service
SOCMINT	Social Media Intelligence
SQL	Structured Query Language

TI	Tecnologia da Informação
UNICART	União dos SISCART's do Departamento de Polícia Federal
URL	Uniform Resource Locator
VoIP	Voice over Internet Protocol
XML	eXtensible Markup Language



# 1. INTRODUÇÃO

Hoje, um dos principais desafios de uma organização é definir como tratar a quantidade de informações decorrentes de inovações técnicas que permitem grande capacidade de coleta, mas que não possibilitaram avanços imediatos em termos de armazenamento, recuperação, processamento e análise daqueles dados que se apresentam em diferentes formatos e em múltiplas bases de dados, tanto em relação a dados estruturados quanto em relação àqueles que não possuem estrutura rígida.

A universalização da utilização da Tecnologia da Informação – TI é notada por meio do grau em que as pessoas se comunicam e realizam registros em seu cotidiano. A presença de dispositivos portáteis na rotina das pessoas, aliada ao acesso massificado à internet e à produção de informações através de múltiplos canais (redes sociais, registros de filmagem e fotográfico, bases cartográficas, georreferenciamento, *blogs*, mensagens de diversos tipos, dados provenientes de sites de compras, dentre outros) são fontes geradoras de informações que estão diretamente ligadas ao dia a dia do cidadão (Manyka et al., 2011).

Antes uma preocupação restrita ao universo científico ou às grandes companhias de tecnologia ou mídias sociais, atualmente as empresas dos mais diversos portes e negócios estão interessadas em obter vantagens competitivas ao explorar novas soluções na esperança de descobrir qualquer informação valiosa que possa estar camuflada num universo tão amplo de dados. As organizações de segurança pública também devem estar atentas a esse novo cenário tecnológico no intuito de fornecer melhores serviços para a sociedade.

## 1.1. JUSTIFICATIVA / HIPÓTESE

Ao contrário da polícia eminentemente ostensiva, que possui como função básica o policiamento preventivo de rua, as polícias judiciárias no Brasil trabalham principalmente com investigação policial, recebendo informações das mais variadas fontes de dados, com o objetivo de processá-las e apresentar o resultado das análises, as quais buscam indicar o fato criminoso ocorrido e seus responsáveis. (ANP/DPF, 2009).

Os organismos de investigação criminal necessitam buscar novas soluções frente aos desafios tecnológicos que lhe são regularmente impostos. Um deles é o crescente volume de dados gerados por cidadãos e empresas e a variedade com que essas informações ficam acessíveis aos órgãos de segurança pública. Estudar alternativas que permitam manipular, de

maneira célere, grandes volumes e diferentes formatos de dados possibilitaria oferecer resposta em tempo hábil à sociedade e enfrentar com mais eficácia às diversas ações criminosas.

A aplicação de conceitos de *big data* poderia melhorar os processos dos núcleos de investigação e das perícias forenses dos órgãos de segurança. Enquanto a inteligência analítica tradicional oferece subsídios para analisar dados e transformá-los em informações para gerar vantagem competitiva a fim de elucidar casos criminais, a utilização de *big data* possibilitaria agregar mecanismos essenciais a esse novo contexto, inclusive a revelação de relações e vínculos mais sutis e que são difíceis de serem encontrados pelos métodos tradicionais de processamento.

Poder-se-ia chegar aos inocentes, suspeitos ou culpados mediante análise exaustiva e relacionamento dos dados que são disponibilizados por meio da busca e apreensão de informações em campo (ou em meios digitais, no caso de investigações relacionadas a essa temática). Perfis de determinada rede social, imagens, vídeos, dados geográficos de aparelhos móveis, informações biométricas, dentre outros, formam o universo analítico que possibilitaria solucionar crimes ou até mesmo prever a ocorrência deles.

As ferramentas do tipo *big data* seriam, portanto, protagonistas no contexto tecnológico de inundação de dados e possibilitariam o atendimento das demandas da segurança pública no que tange à produção de informações, ao mesmo tempo em que podem ser capazes de realizar até mesmo previsões voltadas à ocorrência de determinados crimes.

## **1.2. OBJETIVOS**

O objetivo deste trabalho é propor um modelo de armazenamento e recuperação de informações baseado em *big data* para que as organizações policiais estejam aptas a atuar no presente cenário de manipulação de grandes massas de dados, as quais também se apresentam em formato cada vez mais diversificado.

Para alcançar o objetivo principal, os seguintes objetivos específicos foram definidos:

1. Descrever atribuições e responsabilidades dos principais órgãos policiais que lidam diretamente com investigações policiais no Brasil;

2. Mapear e descrever os insumos informacionais dos órgãos policiais, bem como detalhar o formato de dados de cada um deles;
3. Verificar a possibilidade de aplicação de um modelo informacional de *big data* nos tipos de informação utilizados por esses órgãos;
4. Comparar os tipos mais comuns de bancos de dados existentes para utilização em *big data* e propor soluções que melhor se adaptem à realidade da polícia investigativa;
5. Propor, por meio de um exemplo de investigação policial, uma plataforma de dados orientado a grafo, suportada por *big data*, e que permite a visualização de grandes volumes de dados.

### 1.3. METODOLOGIA DE TRABALHO

O primeiro passo deste trabalho foi buscar nas referências bibliográficas o emprego e definições de *big data*, além de verificar a existência de um modelo que possa lidar com a atual problemática relacionada ao volume de dados processados pelas organizações da área de segurança pública, principalmente no tocante à repressão criminal.

Em seguida, estudou-se a organização e o regimento interno de algumas instituições policiais brasileiras que lidam com investigação e, a partir de tal análise, buscou-se listar as fontes de informação mais importantes, além de apontar as principais demandas em relação à manipulação de dados como suporte às atividades investigativas. Procedeu-se, a partir dessas fontes de informação, sua aplicação ao modelo de informações e características do *big data*, no intuito de confirmar se uma plataforma de dados nesse ambiente é passível de suportar o trabalho investigativo desenvolvido pelas polícia judiciárias.

Posteriormente, realizou-se a descrição de cada insumo e seu respectivo formato de dados, permitindo conhecer mais detalhadamente as necessidades das polícias. Ressalta-se que foi dada ênfase à Polícia Federal tendo em vista seu protagonismo no contexto do combate aos mais diversos tipos de organizações criminosas.

A partir daí, buscou-se conhecer as tecnologias de banco de dados mais comumente utilizadas atualmente com o fito de propor, dentre os tipos de bancos de dados e soluções mais comuns existentes, quais são aqueles que podem compor uma plataforma de informações

baseadas em *big data* para uma organização policial. A partir desse estudo, buscou-se restringir as tecnologias mais adequadas para serem aplicadas no modelo.

Em seguida, por meio de um exemplo de investigação policial que se utilizou de quatro importantes tipos de informação, foi sugerida uma plataforma de processamento que seja capaz de congrega o armazenamento e recuperação de grandes volumes de dados, preservando a principal característica do trabalho investigativo policial, que é a capacidade de análise de vínculos entre investigados e eventos diversos (empresas, transferências bancárias, conversas telefônicas, encontros, etc.).

#### **1.4. RESULTADOS ESPERADOS**

Espera-se, a partir da definição de uma plataforma baseada em *big data* para a manipulação de grandes quantidades de informação de diferentes formatos, a sistematização e visualização de diversos tipos de dados atrelados a um caso e, em seguida, que seu uso seja difundido em investigações policiais cujo andamento dependa da descoberta de vínculos oriundos dessas diferentes fontes de informação, aumentando a eficácia e a eficiência dessas investigações.

#### **1.5. ORGANIZAÇÃO DO TEXTO**

No capítulo 2, é apresentada a revisão bibliográfica sobre o tema, listando os principais conceitos em *big data*. São descritos o contexto da explosão informacional atualmente vivido pela sociedade, a dificuldade que as tecnologias de bancos de dados tradicionais possuem em lidar com os novos desafios, e alguns projetos em segurança pública que obtiveram êxito na aplicação de conceitos e ferramentas baseadas em *big data*.

O capítulo 3 detalha a organização das polícias judiciárias no Brasil, com ênfase na Polícia Federal, buscando compreender suas necessidades em termos de informação e as diversas fontes de dados que fazem parte do processo de investigação policial.

O capítulo 4 analisa as principais soluções de bancos de dados NoSQL (*Not Only SQL*) existentes no mercado e sugere aquela que é mais adequadas para compor um modelo baseado em *big data* levando-se em consideração às necessidades da investigação policial. É apresentada também uma plataforma de armazenamento e processamento capaz de analisar e recuperar grandes volumes de informação, juntamente com seus componentes, e é descrito o

fluxo completo dos dados, iniciando-se pela entrada no órgão policial até a apresentação dos resultados.

O capítulo 5 apresenta um exemplo de investigação policial que se utiliza de parte dos insumos informacionais e descreve como é feita a inclusão de informações na plataforma de dados sugerida e sua respectiva visualização.

Finalmente, o capítulo 6 relata a conclusão do trabalho, explicando a contribuição alcançada e sugerindo tópicos de aprimoramento e de projetos futuros.



## 2. ESTUDO BIBLIOGRÁFICO – ESTADO DA ARTE

Um dos objetivos deste capítulo é apresentar os últimos avanços nos principais temas que envolvem o desenvolvimento desta pesquisa. Será possível também entender o contexto em que a proposta está inserida, quais são os trabalhos relacionados e qual a contribuição em construção.

Serão abordadas as características dos bancos de dados relacionais e dos repositórios tipo NoSQL, a fim de apontar qual das tecnologias melhor supera a problemática tratada no presente trabalho. A importância dos conceitos trazidos pelo sistema de arquivos HDFS e como esse paradigma poderá auxiliar em questões voltadas ao armazenamento e processamento de grandes volumes de dados também serão detalhados.

### 2.1. O TERMO *BIG DATA*

O volume de dados e a diversidade das informações geradas na sociedade estão se tornando cada vez mais extensos e complexos. Segundo Lyman e Varian (2002), antes de 2003, o mundo havia criado cerca de cinco exabytes<sup>1</sup> de dados em toda sua história. Esse é o mesmo quantitativo que é gerado atualmente a cada dois dias em todo o planeta (Breternitz & Silva, 2012).

Os bancos de dados relacionais não se mostram preparados para cumprir os requisitos necessários à implementação de soluções que lidam com enormes quantidades de dados, pois apresentam pouca flexibilidade em sua estrutura. Tais sistemas baseiam-se em quatro premissas, que os fazem eficientes e seguros: atomicidade, consistência, isolamento e durabilidade. O processamento não distribuído é o conceito dominante na maioria dessas soluções disponíveis, o que as tornam pouco adequadas para suportar o crescente fluxo de dados e a extração de informação exaustiva e diversificada. (Sadalage & Fowler, 2013)

Uma alternativa que pode auxiliar a superação das dificuldades expostas é a utilização das ferramentas do tipo *big data*. LI, Kuan-Ching et al. (2015) atrela o termo *big data* a conjuntos de dados grandes e complexos que são gerados por meio de diferentes fontes de

---

<sup>1</sup> 1 exabyte (EB) equivale a 10<sup>18</sup> bytes ou 10<sup>6</sup> terabytes.

informação, como mídias sociais, transações *online*, sensores e sistemas, dentre outros, de modo que é necessário um aparato especial capaz de lidar com o armazenamento, análise e processamento dessas informações.

Loshin (2013) discute o conceito focando o ambiente das organizações empresariais. Segundo o autor, *big data* é centralmente a aplicação de técnicas com boa razão custo/benefício voltadas para a resolução de problemas que ultrapassam a capacidade da organização de fazê-lo com seus atuais ativos tecnológicos.

Uddin & Gupta (2014) destacam o potencial de *big data* nas áreas de pesquisa e negócios, relativo ao potencial dessa técnica na solução de problemas do mundo real. Os autores destacam as 7 dimensões do termo, a seguir apresentadas: a) volume (o tamanho dos dados que são manipulados), b) velocidade (a rapidez com que os dados chegam às organizações), c) variedade (diversidade em relação à apresentação dos dados), d) veracidade (o grau de fidelidade dos dados), e) validade (a precisão que os dados se encontram para sua utilização), f) volatilidade (quanto tempo é necessário o armazenamento dos dados) e g) valor (o objetivo final, ou seja, aquilo que se deseja extrair dos dados).

O que é possível extrair das definições apresentadas é que *big data* não se resume a um sistema de gerenciamento de banco de dados robusto capaz de guardar enormes quantidades de dados; vai além, ao buscar relacionamentos e padrões nesse emaranhado desorganizado de dados a fim de produzir informações valiosas para setores da sociedade. *Big data* está inserido em um contexto no qual as pessoas buscam entender o colossal volume de dados às quais estão quase que diuturnamente submetidas.

## **2.2. A EXPLOSÃO INFORMACIONAL**

O contexto social vivido nas últimas décadas traz consigo o aumento do volume de informações produzido e disponibilizado por diferentes atividades sociais, o que acaba por dificultar sua identificação, acesso e posterior utilização desses ativos. É o chamado fenômeno da explosão informacional.

Saracevic (1996) descreve o fenômeno da explosão informacional, conceituando-o como sendo o “irreprimível crescimento exponencial da informação e de seus registros”. Bush (1945) destaca o impacto da criação de sistemas informatizados e o conseqüente aparecimento

do *boom* informacional no período pós Segunda Guerra Mundial, alicerçado principalmente nas atividades de ciência e tecnologia e a busca por maior conhecimento.

As tecnologias criadas para lidar com esse fenômeno, se por um lado ajudam a organização da informação e sua melhor pesquisa e utilização, por outro, abrem espaço para aumentar cada vez mais sua produção, retroalimentando esse ciclo. As redes sociais são exemplo de ferramenta que se popularizou em todo o mundo e abriu grandes possibilidades para a produção em massa de informações.

A previsão é que em 2020, pelas antenas de telefonia móvel do planeta, transitem 367 exabytes de dados, o que representa um aumento de quase 5 vezes ao que foi registrado em relação ao ano de 2015 (ComputerWorld, 2016).

As plataformas de redes sociais são websites ou aplicações que permitem pessoas interagir, criar, compartilhar e trocar informações, conforme detalha a Tabela 2.1. A publicação *Social Media Today* (2014) indica a existência 1,15 bilhão de usuários na rede social *Facebook*, os quais adicionam 350 milhões de fotografias diariamente. Do total de usuários, 751 milhões realizam o acesso por meio de telefones móveis. Há 500 milhões de usuários registrados no *Twitter*, que produzem um movimento diário de 400 milhões de publicações.

A rede social *Google Plus* possui 500 milhões de usuários. Existem quase 250 milhões de perfis no *LinkedIn*, que compartilham e registram informações a respeito de suas vidas profissionais, bem como detalham seus relacionamentos com empresas e outros tipos de organizações. Na base de dados do Instagram, estima-se a presença de 16 bilhões de fotografias, que geram mais de 1000 comentários por segundo por cerca de 130 milhões de usuários, que diariamente acrescentam mais de 5 milhões de fotografias. A estimativa é que 4,2 bilhões de pessoas acessam as redes sociais através de plataforma móveis em todo o planeta, o que representava, em 2014, cerca de 60% da população do planeta.

Tabela 2.1: Quantidade de Usuários nas Redes Sociais

Rede Social	Número estimado de usuários <sup>2</sup>	Proporção em relação ao número de habitantes do planeta <sup>3</sup>
Facebook	1.150	1 em cada 6,24
Twitter	500	1 em cada 14,36
Google Plus	500	1 em cada 14,36
LinkedIn	250	1 em cada 28,72
Instagram	130	1 em cada 55,25

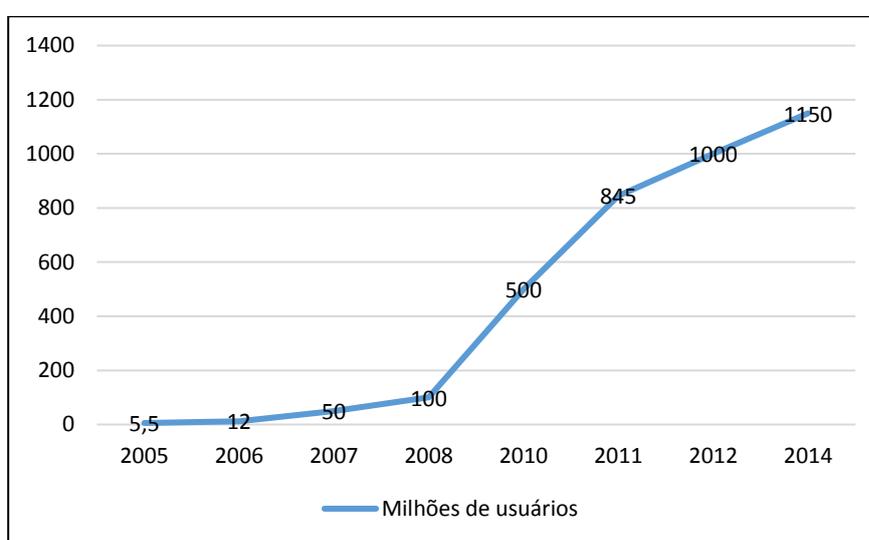


Figura 2.1: Quantidade de usuários do Facebook (2005 - 2014)

Segundo o IDC Iview (2012), entre 2005 e 2020, o universo digital vai crescer 300 vezes, partindo de 130 *exabytes* para 40.000 *exabytes* de dados, ou seja, haverá uma média de 5.200 gigabytes de dados para cada habitante da terra. As previsões ainda dão conta de que, a cada ano, esse universo digital dobrará de tamanho. Somente na área de segurança, englobando as atividades de investigação criminal e inteligência militar, estima-se que em 2020, cerca de 45% dos dados referentes ao tema poderão ser candidatos a utilização em *big data*.

<sup>2</sup> Em milhões de usuários

<sup>3</sup> Considerando população mundial de 7,183 bilhão, conforme o *site* [countrysmeters.info](http://countrysmeters.info)

O número de servidores de dados no mundo aumentará em 10 vezes e a quantidade de *datacenters* empresariais será 14 vezes maior. Até 2020, 40% do universo digital terá correlação com a computação em nuvem e cerca de 15% dele estará armazenada nesse tipo de plataforma. Outro fator impactante no volume de informações, segundo o mesmo relatório, além do aumento da computação em nuvem, é a explosão da chamada internet das coisas, proporcionando que bilhões de equipamentos de todos os tipos estejam conectados entre si trocando informações a todo momento.

A maioria da informação do universo digital (68%) será criada e utilizada por usuários finais (TV Digital, mídias sociais, envio e recebimento de imagens e vídeos através de telefones, dentre outros). Apenas 33% dos dados, até 2020, será capaz de ser analisado e a partir deles ser extraído alguma informação útil

Não se pode duvidar que o uso de mídias sociais para auxílio em investigações criminais é uma importante ferramenta na prevenção do crime, conforme pontuado por (Da Silva, 2014), o qual destaca o descrédito por parte dos usuários em seu monitoramento e a consequente sensação de liberdade quando estão navegando nesses ambientes. Assim, órgãos policiais podem se beneficiar na identificação de evidências e no apontamento de indivíduos associados com a atividade criminal, o que vem demandando a utilização de aparatos específicos de investigação.

O conjunto de descobertas advindos por meio da análise de grandes quantidades de dados disponíveis nas redes sociais passa pelo entendimento e descoberta como as pessoas pensam e agem. Para um processo de investigação, agregar dados existentes e registrados no mundo real àqueles que podem ser extraídos das redes sociais permite um ganho para os investigadores (Houses of Parliament, 2014). É o caso da pessoa suspeita de executar um crime que informa estar num lugar distante daquele onde ocorreu o delito, mas descobre-se nas redes sociais que ele estava nas imediações da ocorrência investigada.

Assim, percebe-se que o fenômeno da explosão informacional não é passado ou se encaminha para seu término; ao contrário, a tendência é que a criação de dados e o surgimento de novas tecnologias sejam feitas num ritmo bem maior a cada ano, o que exigirá habilidade, experiência e recursos para lidar com esse novo contexto. Além do volume, os dados vão tornar-se mais complexos, pela existência de diversos formatos, requerendo uma estrutura cada vez mais escalar, flexível e especializada de tecnologia da informação.

No entanto, ainda existem importantes dificuldades de cunho técnico. Originalmente, as informações contidas em redes sociais não foram criadas para o propósito de pesquisa, mas foram feitas para que sejam plataformas agradáveis de interação entre pessoas. Dessa forma, deve-se lançar mão de ferramentas de extração de dados automática, que não consegue a recuperação de 100% das informações, devido às restrições impostas por seus desenvolvedores (Russel, 2013). Ultrapassar barreiras como essa, que estão presente nos diversos *sites* de redes sociais, é justamente uma das funções de uma plataforma baseada em *big data*.

### **2.3. BANCO DE DADOS RELACIONAIS X NOSQL**

De plano, é necessário fazer uma definição do conceito de bancos de dados NoSQL. Segundo Tiwari (2011), NoSQL está relacionado a uma generalização dada pela literatura a respeito de bancos de dados que não seguem os pilares básicos dos Sistemas Gerenciadores de Banco de Dados Relacionais – SGBDRs. O termo, portanto, não se refere a uma tecnologia ou produto específicos; ele está vinculado a uma classe de soluções de manipulação de grandes volumes de dados em que se faz necessário a utilização de técnicas de escalabilidade.

Recentemente, diversos estudos foram guiados comparando o desempenho de tipos e modelos de base de dados. Os trabalhos e publicações, num modo geral, dividem as soluções existentes no mercado em dois grandes grupos, cujos princípios estão em oposição: banco de dados relacional e banco de dados NoSQL. Não se debatem apenas o desempenho ou a capacidade dos modelos; outras importantes características também são comparadas, tais como escalabilidade, segurança e flexibilidade de esquema.

Os SGBDRs foram concebidos na década de 1970 como solução à grande demanda de usuários para prover acesso facilitado às informações, tendo como alicerces as chamadas características ACID: atomicidade, consistência, isolamento e durabilidade.

Esses repositórios possuem bons desempenhos num ambiente específico de dados. Ao lidar com grandes volumes de informação em tempo real, típicos da internet e das novas aplicações, essas arquiteturas apresentam ineficiência, exigindo-se repensar o conceito de armazenamento e processamento voltado para *big data*. Segundo Tauro (2012), podem ser apontados três grandes problemas inerentes aos bancos de dados relacionais que os colocam em certa posição de fragilidade em relação a questões ligadas ao contexto tecnológico atual: a)

crescimento de dados advindos da internet; b) aumento da conectividade entre dispositivos e c) presença cada vez maior de dados semiestruturados.

Padhy (2011) corrobora com essa visão ao explicar que, além da necessidade do processamento de grandes fluxos de dados, outro fator que influenciou a criação de arquiteturas NoSQL foi a constante presença de tipos de dados complexos, semiestruturados ou não estruturados. A preocupação não é apenas em relação aos novos patamares de volume dos dados, mas é superar as dificuldades relacionadas à sua modelagem, que estão presentes em diversos domínios de aplicações, como internet, redes sociais, redes de sensores, dentre outros. Os bancos de dados relacionais são difíceis de lidar com aplicações voltadas para *data warehousing*, Grid<sup>4</sup>, Web 2.0 e nuvem.

Han et al. (2011) também apontam os grandes problemas e limitações dos bancos de dados relacionais no contexto de *big data*. São eles: a) leitura e escrita lentas, devido ao tamanho dos dados manipulados e a possibilidade maior de ocorrência de problemas relacionados ao controle de concorrência, o que pode acarretar queda de rendimento; b) limitação de capacidade, devido à própria dificuldade da arquitetura em lidar com um volume grande de dados em seu motor de busca e c) expansão limitada, principalmente devido ao mecanismo de inter-relacionamento de dados ocorrer entre tabelas e suas chaves estrangeiras que, em essência, não oferecem suporte adequado à escalabilidade.

Algumas tentativas, por parte dos desenvolvedores de SGBDs relacionais, vem sendo realizadas no intuito de se adequar às novas necessidades tecnológicas. Uma delas é agregar novas funcionalidades típicas de bancos de dados NoSQL, em alguns casos, por meio de encapsulamento de extensões do modelo objeto-relacional. Essa abordagem, segundo Vieira (2012), não causa grandes quebras de paradigma, pois as políticas de transação, replicação, segurança e gerenciamento dos dados continuam as mesmas, não afetando a estrutura de funcionamento do repositório de dados e tampouco as rotinas de *backup*, de auditoria, dentre outros. Adicionalmente, tal solução sofre críticas por não conseguir abarcar muitas outras funcionalidades necessárias ao *big data*, além de continuar preso à linguagem SQL que, por suas limitações naturais, não responde satisfatoriamente às necessidades das novas aplicações.

---

<sup>4</sup> Tipo particular de sistema de computação distribuída

Willke (2012) percebeu a necessidade de se construir arquiteturas capazes de analisar grandes volumes de dados e oferecer ferramentas que organize tais informações em grafos, de modo a facilitar a visualização e análise de informações.

## **2.4. O SISTEMA DE ARQUIVOS HDFS (HADOOP DISTRIBUTED FILE SYSTEM)**

Primeiramente, cabe ressaltar que HDFS não é um tipo de base ou repositório de dados, mas um sistema de arquivos distribuídos capaz de lidar com *petabytes* de informações. Segundo Kumar et al. (2014), o ambiente baseado em HDFS facilita a existência de processamento paralelo, criando inúmeros blocos de dados que podem ser distribuídos em máquinas na formação de *clusters* de processamento.

O HDFS é um sistema de arquivos distribuídos que possui algumas particularidades em relação a outros sistemas desse tipo. Segundo o projeto Apache Hadoop<sup>5</sup>, sua concepção é baseada em duas premissas principais: tolerância a falhas e utilização de hardware de baixo custo. Suas principais características são descritas a seguir:

1. Tolerância à falha de hardware, por meio da detecção e tratamento de falhas transparente ao usuário. Quando um nó falha, ele é rapidamente substituído por outro nó, sem que ocorra interrupção no processamento dos dados. Tudo é feito de maneira transparente pelo sistema, dispensando a interferência do usuário na resolução do problema;
2. Utilização de processamento por meio de tarefas (*jobs*) no lugar do processamento contínuo por meio de interações de usuários. A ênfase é permitir um *throughput* elevado, ao invés de focar na latência da aplicação;
3. Capacidade de lidar com conjunto de dados muito grandes (*gigabytes* ou *terabytes* de dados), por meio do balanceamento das tarefas entre centenas de nós computacionais;

---

<sup>5</sup> Apache Hadoop é um framework que desenvolveu seu próprio sistema de arquivos distribuídos, chamado de Hadoop Distributed File System – HDFS, baseado no GFS (Google File System). Disponível em <https://hadoop.apache.org>

4. Concepção de sistema de arquivo baseado na premissa de que é mais fácil movimentar os dados, distribuindo os blocos de dados entre as máquinas disponíveis, do que movimentar a infraestrutura de hardware;
5. Projeto foi realizado com a condição de ser portátil entre diversas plataformas de software e hardware, o que aumenta sua flexibilidade de uso dentro da organização, principalmente em relação ao aproveitamento do legado de máquinas e programas já existentes.

Goldman (2012) também destaca as vantagens da utilização do Hadoop em ambientes que necessitam de prover respostas rápidas, citando os seguintes aspectos: a) código aberto, o que permite melhorias constantes no código por meio de comunidade de desenvolvedores; b) economia de recursos, pois permite o processamento da massa de dados em computadores convencionais; c) robustez, tendo em vista a disponibilização de mecanismos de replicação de dados, dentre outros, garantindo que a aplicação continue seu processamento mesmo na ocorrência de falhas de algum recurso; d) escalabilidade, pois o acréscimo de máquinas no *cluster* é fácil e simples; e e) simplicidade, pois o sistema retira do desenvolvedor preocupações de gerenciar atividades como balanceamento de carga, tolerância a falhas e questões relativas à computação paralela.

## **2.5. O MODELO INFORMACIONAL BASEADO EM *BIG DATA***

Antes de propor aplicação de tecnologias mais adequadas para a implantação de *big data* em qualquer organização, faz-se necessário avaliar se realmente os insumos informacionais utilizados pela organização são condizentes com esse novo paradigma.

Para isso, este estudo recorre ao modelo informacional genérico proposto por (Soares, 2014), o qual se divide em 5 (cinco) grandes grupos de dados e insumos, conforme

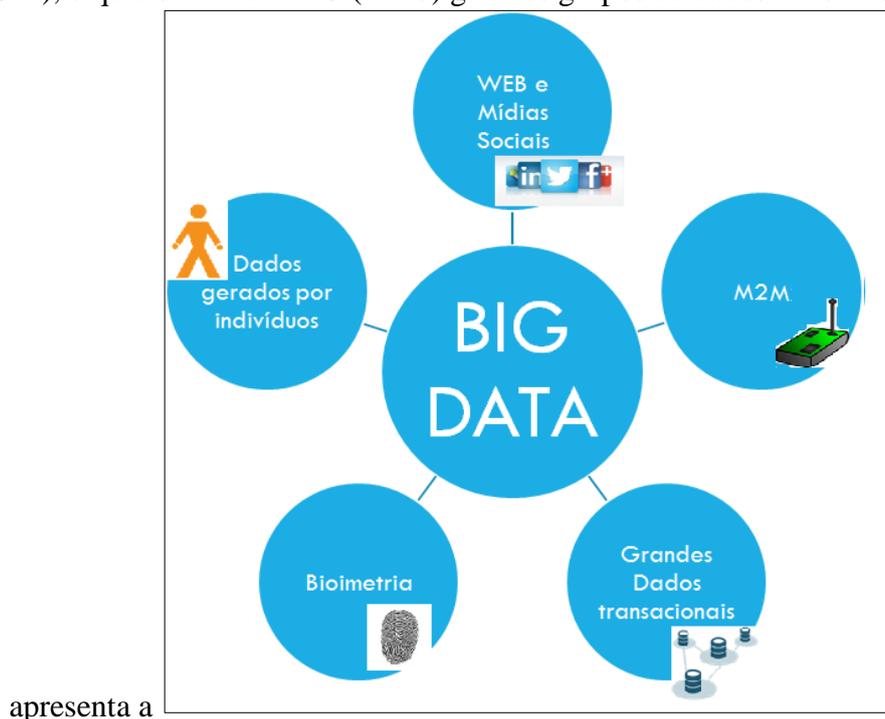


Figura 2.2: Modelo Informacional do *big data*

e o breve descritivo de cada um dos grupos:



Figura 2.2: Modelo Informacional do *big data*

### 1. Dados da web e mídias sociais

As informações provenientes de redes sociais, blogs e outros sítios de internet estão se tornando cada vez mais importantes pois estão em constante atualização. Dessa forma, elas tornam-se cada vez mais valiosas no processo de produção do conhecimento e uma demanda importante nas organizações refere-se à necessidade de realizar a integração desses dados com o repositório principal que se encontra na organização há algum tempo.

Cabe ressaltar que as políticas de acesso e utilização de informações produzidas nas redes sociais variam muito de acordo com o provedor, o que torna mais desafiador a utilização de dados provenientes da web e redes sociais. Nunca se deve estar distante dos limites legais, quando se fala em coleta e processamento de informações, sob pena de anulação do procedimento investigatório.

### 2. M2M (*MachineToMachine*),

M2M, ou a comunicação máquina a máquina, são tecnologias que utilizam dispositivos com ou sem fio conectados entre si, trocando informações. Esse tipo de tecnologia utiliza sensores ou medidores para capturar um evento que é encaminhado por meio de uma rede para uma aplicação (software) que traduz o referido evento em informação útil. Estima-se a existência em todo o mundo 50 bilhões de dispositivos, muitos deles interconectados através do protocolo IP e gerando informações diuturnamente. (Revista Exame, 2014).

### 3. Grandes dados transacionais

Os dados transacionais geralmente estão existentes na organização por meio de formatos estruturados ou não-estruturados e se relacionam diretamente a informações geradas por sistemas tradicionais da organização. Geralmente são informações que são tratadas há anos nos setores de TI e possuem pouca integração, paralelismo ou processamento distribuído.

### 4. Biometria

São informações oriundas de impressões digitais, reconhecimento de íris, retina, reconhecimento facial, dentre outros dispositivos de identificação por meio de características biológicas e/ou corporais.

#### 5. Dados gerados por indivíduos

São dados gerados por pessoas, geralmente materializados por meio de anotações, gravações de voz, fotografias, vídeos, documentos em papel, mensagens eletrônicas (*e-mails*) e que estão apresentados, geralmente, no formato não-estruturado. São informações valiosas, porém esbarram em dificuldades de integração com as bases de dados relacionais tradicionais.

Além dos grupos de dados e insumos, Soares (2012) destaca alguns aspectos importantes e que devem ser considerados ao implantar uma plataforma *big data* na organização. São eles:

1. Organização dos dados: diretrizes gerais de implementação, considerando o grau de integração das ferramentas *big data*, estrutura requerida, regras de acesso e responsabilidades;
2. Gerência dos metadados: relaciona-se à integração dos dados oriundos das ferramentas *big data* com os dados da organização;
3. Privacidade: relaciona-se à regulamentação de acesso aos dados, através de legislação governamental e termos de uso de redes sociais, por exemplo;
4. Qualidade dos dados: relaciona-se ao gerenciamento diferenciado dos dados oriundos de algumas fontes, pois elas devem ter especificidades diferentes;
5. Integração aos processos de negócios: no caso policial, deve voltar toda a sistemática e coleta de informações para subsidiar a atividade-fim, qual seja, a investigação;
6. Gerenciamento do ciclo de vida informacional: pode tornar-se um gargalo analisar todo o material disponível para coleta, principalmente aqueles produzidos por redes sociais, internet e fluxos eletrônicos, devido ao grande volume apresentado. Outro desafio refere-se em gerenciar qual informação deve ser arquivada ou descartada ao longo do tempo.

## 2.6. **BIG DATA NA SEGURANÇA PÚBLICA**

As agências de segurança pública e órgãos de defesa rotineiramente coletam massas de dados no curso de seu trabalho no intuito de prevenir e detectar crimes, bem como produzir importantes informações de inteligência. O desenvolvimento das comunicações eletrônicas propiciou o aumento do fluxo de dados e seus tipos, incrementando a disponibilidade de referências acerca de pessoas, seus relacionamentos e suas atividades rotineiras.

O acesso livre à internet e a possibilidade de atuar na rede de forma anônima transforma-se em outro desafio colocado frente aos órgãos responsáveis pela repressão à criminalidade. Buscar a atuação discreta e anônima é parte da estratégia das organizações criminosas, cujos objetivos, de modo geral, ao se conectarem digitalmente são: a) difundir informação; b) captar voluntários; c) realizar comunicação rápida entre pessoas; c) obter recursos; e, e) coordenar entes que possam ajustar suas ações para delinquir (Batista, 2004).

Diante desse novo cenário, a utilização de *big data* em segurança pública traz ganhos em vários processos relacionados à atividade policial e de segurança. Esse novo conceito pode ser utilizado nas seguintes atividades (Houses of Parliament, 2014):

1. Prevenção criminal: os dados de geolocalização criminal estão comumente sendo utilizados para encontrar padrões criminais relacionados à área de ocorrência, os quais podem dar suporte às equipes de investigação ao apontar os tipos criminais que ocorrem com mais regularidade em determinada área;
2. Predição criminal: certos tipos de crimes, como fraudes bancárias, terrorismo, roubo a bancos, dentre outros, possuem certos padrões de ocorrência. Algoritmos podem ser desenvolvidos para que tais elementos de previsibilidade sejam melhorados, permitindo criar rotinas de previsão que produzam informações importantes, buscando antever ações criminosas que coloquem a polícia em posição de vantagem em relação à ação delituosa;
3. Produção de informações estratégicas: a análise em *big data* pode ser utilizada para o planejamento estratégico das ações de órgãos de segurança pública. É o exemplo do projeto europeu *ePOOLICE*<sup>6</sup>, que busca relacionar dados internos da polícia com

---

<sup>6</sup> *Early Pursuit Against Organized Crime Using Environmental Scanning, the Law and Intelligence Systems*

- redes sociais, visando identificar vínculos entre indivíduos que praticam crimes cibernéticos, tráfico humano e tráfico de drogas em países da União Europeia;
4. Detecção criminal: alguns tipos de fraudes utilizam centralmente troca de dados e informações quando estão sendo cometidos. Ferramentas e softwares *big data* podem fazer com que tais processos se tornem mais eficientes e auxiliem a detectar padrões, como é o caso de crimes financeiros. Tais crimes possuem base em transações bancárias e operações financeiras diversas, o que gera grandes volumes de dados e dificulta sua descoberta;
  5. Segurança nacional: agências de segurança nacional coletam dados de inteligência envolvendo a investigação e o acompanhamento de indivíduos e organizações suspeitas de atividades danosas ao estado. Análise em *big data* fornece a capacidade de armazenar e processar grandes quantidades de comunicações eletrônicas de dados, a fim de identificar padrões ou conexões que podem indicar um comportamento ameaçador, como uma célula terrorista.

Os projetos e ações relacionados à *big data* na área de segurança pública já são realidade em algumas instituições policiais. Pesquisadores propuseram uma ferramenta capaz de mapear locais mais sensíveis às dinâmicas criminais prevendo, inclusive, pontos geográficos aos quais a polícia deve dar prioridade. Tal aplicação oferece elementos preditivos ao policiamento local, melhorando a capacidade dos agentes no combate à criminalidade ao otimizar o emprego de recursos humanos e materiais (State University of New Jersey, 2014).

A cidade americana de Detroit, considerada por cinco anos consecutivos uma das cinco cidades mais violentas dos Estados Unidos, adotou uma postura proativa na busca de soluções para fazer frente à criminalidade. Dentre o conjunto de estratégias, uma delas foi identificar indivíduos que estão engajados em atividades criminosas. A tarefa foi agregar *terabytes* de dados públicos e proprietários relacionados à criminalidade e seu *modus operandi*, de forma a entender o relacionamento de indivíduos, gangues e locais de atuação. A polícia necessitava de uma solução rápida e palatável para prover análises acerca de atividades criminosas em potencial. A ferramenta implementada permitiu a polícia local entender a dinâmica criminal da cidade e trabalhar de forma proativa, marcando presença em potenciais locais de crimes. (Datameer, 2014).

Outro caso que merece destaque, pelas cifras envolvidas, ocorreu na Inglaterra. As autoridades da aduana britânica investiram em um projeto do tipo *big data* para implementar

um sistema integrado de análise para o fisco daquele país. Com um investimento inicial de 45 milhões de libras e duração prevista de 5 anos, conseguiu-se a detecção de 2,6 bilhões de libras em fraude. O sistema utiliza 28 fontes de informação que processam 1 bilhão de registros internos e externos simultaneamente (transações comerciais, impostos recolhidos, contas bancárias, dados de empresas), ajudando a revelar relações que não seriam descobertas por um sistema tradicional de análise ou por um auditor fiscal (Capgemini Consulting, 2013).

O desenvolvimento da internet ocasionou o surgimento de conceitos que fazem parte do dia a dia dos cidadãos. A combinação da utilização da internet, sua massificação e a possibilidade de produção em massa de informações, inclusive a serviço da criminalidade, pugna pela mudança de paradigma nos processos de trabalho dos responsáveis por sua repressão. A título de exemplo, dada a dificuldade em lidar com essa temática, os Estados Unidos, por meio do FBI, implementaram a *Cyber Task Forces*<sup>7</sup>.

As mídias sociais podem ajudar a desenvolver novos modelos de polícia que são adaptáveis para o século 21. As forças de segurança pública podem retirar diversos benefícios no uso das redes sociais, até mesmo porque a área de negócios já vem lidando com os tais desafios há algum tempo. Em 2012, 67% dos usuários de internet utilizam mídias sociais. Entre 2011 e 2012, o tempo gasto pelos cidadãos americanos nas redes sociais aumentou 37%, passando de 88,4 bilhões de minutos para 121,1 bilhões de minutos. (U. S. DEPARTMENT OF JUSTICE, 2014)

As redes sociais estão cada vez mais presentes nos trabalhos da polícia. Técnicas de *Social Media Intelligence*, ou SOCMINT, podem ajudar em ameaças fornecendo indicadores sobre onde a violência está prestes a ocorrer. Cabe destacar que a regulação da utilização de dados das redes sociais ainda é muito controversa, surgindo dúvidas sobre como os dados pessoais podem ser armazenados, alterados, utilizados ou descobertos. (Omand, Bartlett & Miller, 2012).

---

<sup>7</sup>As Forças Tarefas Cibernéticas (tradução em português) tornam mais efetivo o combate à criminalidade digital, pois são estruturas que contam com agentes capacitados em informática forense, dotando-os de conhecimentos e ferramental apropriado para recuperar, preservar e analisar grandes volumes de evidências digitais (FBI, 2014).

Contudo, as soluções ainda são incipientes e as propostas ainda se limitam a trabalhar com a análise exaustiva de grandes massas de dados, em universos isolados de informações, como dados fiscais, dados geográficos e conteúdo proveniente de redes sociais. As limitações de ordem técnica devem-se à ausência de ferramentas capazes de integrar todo o volume e a variedade de dados disponíveis e dar uma resposta em tempo hábil às demandas existentes. A busca por modelos mais versáteis que trabalhem com estruturas de dados mais complexas e flexíveis e que possuam melhor capacidade em se adaptar à dinâmica tecnologia atual se tornam primordiais no que se refere ao futuro do trabalho investigativo.

Assim, o trabalho policial deve buscar novas tecnologias e abordagens metodológicas em coletar, processar e analisar grandes quantidades e tipos de dados, procurando abarcar todos os tipos de insumos informacionais que recebe.

O que se observa é que a utilização de técnicas de *big data* pode continuar a crescer e ser o principal protagonista no ambiente de informações volumosas e diversificadas ao qual as atividades de segurança pública e defesa nacional estão inseridas, sendo capaz de oferecer suporte a processos de análise vinculados à prevenção e predição criminal, produção de informações estratégicas, detecção criminal e segurança nacional.

Além do aspecto técnico, outro ponto que merece destaque quando o tema é segurança pública e investigação policial é a definição dos limites do uso das informações sigilosas manipuladas pelos governos frente à privacidade dos cidadãos. Caso implementado, um modelo analítico dessa envergadura permitirá o monitoramento de atividades diárias dos cidadãos de forma mais invasiva, por meio do cruzamento de dados de diversas fontes de informação, o que pode levar a uma sensação generalizada de vigilância estatal e a perda de credibilidade do trabalho investigativo.

### **3. A POLÍCIA INVESTIGATIVA NO BRASIL E SUAS FONTES DE INFORMAÇÃO**

Este capítulo tem por objetivo avaliar a estrutura organizacional dos principais órgãos policiais brasileiros que lidam com investigação criminal, a fim de constatar suas necessidades em termos informacionais e as diversas fontes de dados que fazem parte do processo investigativo. Também é possível constatar que uma plataforma de dados baseada em *big data* é de grande valia para a otimização dos trabalhos de investigação.

#### **3.1. ESTRUTURA E ORGANIZAÇÃO DAS POLÍCIAS JUDICIÁRIAS**

Para entender como uma plataforma de dados poderá auxiliar o processo de investigação em organismos policiais no Brasil, primeiramente deve-se compreender a estrutura dessas organizações e conhecer o tipo de informação que elas lidam para desempenhar sua finalidade social, qual seja, a elucidação de crimes por meio de atividades de coleta legalmente reconhecidas pelo estado.

A polícia judiciária tem como objetivo principal apurar as infrações penais e sua autoria por meio de atividades de investigação, os quais possuem o papel de subsidiar os atos seguintes sob responsabilidade do Ministério Público e da Justiça, que é promover o processo penal e realizar seu julgamento, respectivamente, buscando a condenação de criminosos e a absolvição dos inocentes. No Brasil, tais funções pré-processuais são exercidas pelas Polícias Cíveis estaduais e distrital, nas 27 (vinte e sete) unidades da federação, e pela Polícia Federal, que exerce a função exclusiva de polícia judiciária da União (Junior, 2006).

Este trabalho analisou a estrutura de 4 (quatro) departamentos de polícia civil de unidades federativas do Brasil, quais sejam, São Paulo, Rio de Janeiro, Minas Gerais e Distrito Federal, bem como a polícia judiciária da união, a Polícia Federal. Com as análises dos normativos listados na Tabela 3.1, buscou-se conhecer a estrutura e o detalhamento das competências de cada unidade e setores subordinados que compõe a estrutura de cada órgão policial.

Tabela 3.1: Legislação aplicada à organização das polícias judiciárias

Unidade Policial	Legislação
Polícia Federal	Regimento Interno do Departamento de Polícia Federal (Ministério da Justiça, 2011).
Polícia Civil do Estado de São Paulo	Reforma administrativa da Secretaria de Segurança Pública (Governo do Estado de São Paulo, 1969).
Polícia Civil do Estado do Rio de Janeiro	Estrutura básica da Polícia Civil do Rio de Janeiro (Governo do Estado do Rio de Janeiro, 2015).
Polícia Civil do Estado de Minas Gerais	Organização da Polícia Civil do Estado de Minas Gerais (Governo do Estado de Minas Gerais, 2008).
Polícia Civil do Distrito Federal	Regimento Interno da Polícia Civil Do Distrito Federal (Governo do Distrito Federal, 2009).

A escolha de parte – e não de todos os 27 departamentos de polícia civil brasileiros - foi necessária para viabilizar o estudo, optando-se pelos órgãos dos estados mais populosos do Brasil (São Paulo, Rio de Janeiro e Minas Gerais), da capital federal (Distrito Federal), além da Polícia Federal, que é reconhecidamente uma instituição em que se executa os maiores trabalhos policiais de repressão à criminalidade organizada no país, cujos resultados são constantemente publicados na mídia.

Durante a análise documental, num primeiro momento, percebe-se que existem muitas similaridades no que se refere à estrutura organizacional dessas instituições e sua organização interna, fazendo com que a base informacional e seu fluxo também possuam grandes coincidências entre si. O organograma exibido na Figura 3.1 representa, de forma genérica, as instituições analisadas.



Figura 3.1: Organograma genérico das instituições policiais estudadas

Geralmente, cada instituição policial possui uma estrutura principal que lida com as investigações tradicionais (roubos, furtos, fraudes diversas, homicídios, lesões corporais, dentre outros), através da instauração de inquéritos policiais, seu processamento e a execução de atividades correlatas, tais como a lavratura de depoimentos e relatos, solicitação de perícias, execução de medidas cautelares, tais como solicitação de prisões, buscas e apreensões, por exemplo. Um núcleo operacional está vinculado a essa estrutura objetivando coletar dados por meio das diligências, inclusive aquelas que se desenrolam nas ruas, realizadas por vigilâncias e registros, pelos agentes, das ações perpetradas pelos suspeitos ou investigados.

Além dessa estrutura, há em todas as organizações analisadas um departamento que lida com crimes cibernéticos ou de informática. Isso é um forte indicativo de que o ambiente virtual vem chamando a atenção da polícia no que se refere à presença de criminosos, e que os procedimentos para a repressão dessas práticas também estão se especializando, merecendo, inclusive, a criação de unidades para essa finalidade.

Os departamentos de criminalística, por sua vez, fornecem laudos periciais que vão subsidiar a investigação por meio da análise técnica de vestígios, examinando elementos e materiais que podem confirmar ou não a materialidade do ocorrido.

Os núcleos de identificação trabalham com a coleta e processamento de impressões digitais, além de serem responsáveis pela elaboração de retratos falados e outras técnicas de reconhecimento facial.

Os núcleos de inteligência policial são responsáveis por informações e operações sensíveis de inteligência, além de exercer atividades de contrainteligência nas instituições, objetivando, de modo geral, a proteção dos ativos de conhecimento do órgão.

O aumento da sofisticação criminal nos últimos anos levou o surgimento do conceito de crime organizado. Nesse modo de atuação percebem-se várias características que potencializam as ações delituosas, tais como hierarquia entre os integrantes, responsabilidade rígida entre entres, divisão territorial, alto poder de corrupção e grande poder econômico, desafiando, ainda mais, o poder de trabalho da polícia (Cernicchiaro, 1997).

Diante desse contexto, em todas as polícias analisadas constatou-se a presença de unidades ou departamentos específicos de combate ao crime organizado. Depreende-se, portanto, que nessas unidades, dada as características presentes nesse tipo de atuação criminal, é necessária a utilização de técnicas mais sofisticadas de investigação para prover o real combate a esses criminosos.

### **3.2. FONTES DE INFORMAÇÃO**

Após análise dos regimentos internos das polícias, suas atribuições e sua organização, em que pese haja algumas diferenças de nomenclaturas ou posicionamento hierárquico entre os setores, a essência de funcionamento é similar, ou seja, os órgãos trabalham com a base de informações bem parecida.

As subseções 3.2.1 a 3.2.8 agrupam esses insumos informacionais em tipos, cabendo ressaltar que a relação não é exaustiva, ou seja, é possível que apareçam novos tipos de informação à medida que a dinâmica criminal e a tecnologia também se encontram em constantes mudanças.

#### **3.2.1. Processo tradicional de investigação**

São dados provenientes de atos formalizados que estão diretamente relacionados à engrenagem de investigação da polícia judiciária e seu inquérito policial. As peças do inquérito são basicamente constituídas de documentos que relatam depoimentos, testemunhos,

informações produzidas por policiais, resultados de diligências, indiciamentos, respostas de ofícios encaminhados, análise de material apreendido, dentre outros (Silva et al., 2000).

### 3.2.2. Sistemas de informações transacionais

Tais informações possuem origem em duas categorias de bancos de dados. O primeiro tipo está relacionado à atividade finalística da polícia e lida com os dados de investigados, operações policiais, suspeitos, condenados, indiciados, testemunhas, vítimas, dentre outros; o segundo tipo está relacionado a outras atribuições polícia, não vinculadas diretamente à atividade criminal, tais como o sistema de registro de condutores de veículos e o sistema de identificação civil da população - que estão sob responsabilidade de algumas Polícias Cíveis - a base nacional do passaporte, o sistema controle de produtos químicos e o sistema de controle de armas - sob responsabilidade da Polícia Federal.<sup>8</sup>

### 3.2.3. Laudos periciais e relatórios de identificação

São as informações advindas da parte técnica da estrutura policial, exercida pelos peritos criminais e papiloscopistas ou datiloscopistas. Os laudos periciais são parte da investigação que analisam os vestígios presentes em um crime. São resultados de perícias em sistemas de informática, computadores, material biológico, entorpecentes, papel-moeda, arquivos contábeis, obras públicas, dentre outros materiais. Os relatórios de identificação lidam com as impressões digitais coletadas em locais de crime, buscando vinculá-las a algum indivíduo suspeito ou vítima.

### 3.2.4. Tabelas ou planilhas oriundas de medidas cautelares

Os extratos bancários, na maioria dos entes federados, são recebidos por meio de convênio entre o Banco Central e os órgãos de investigação. Conforme Manual de procedimentos do SIMBA<sup>9</sup>, são fornecidos às instituições policiais os seguintes dados: agências bancárias, contas correntes, titulares, extrato de movimentação e o fluxo (origem/destino) da movimentação. A instituição financeira, através de técnicas de criptografia, encaminha as

---

<sup>8</sup> O sítio [www.pf.gov.br/servicos](http://www.pf.gov.br/servicos) discrimina quais os serviços prestados pela Polícia Federal aos cidadãos.

<sup>9</sup> Disponível em [asspaweb.pgr.mpf.gov.br](http://asspaweb.pgr.mpf.gov.br)

informações à área de investigação policial que, através de técnicas de análise, posteriormente enviará relatórios de análise para a gestão da investigação ou operação policial.

Os extratos telefônicos são informações provenientes de uma operadora telefônica e contém as transações realizadas por terminais telefônicos dos usuários, tais como chamadas efetuadas, chamadas recebidas, data da chamada, horário da chamada, duração e localização da chamada (ERB utilizada).

Os dados recebidos por instituições bancárias e operadoras de telefonia possuem característica estruturada, com formato amigável e geralmente estão em tabelas ou planilhas, e são encaminhados à equipe de investigação sempre após autorização judicial.

#### 3.2.5. Interceptação telefônica

Essas informações são oriundas do conteúdo das ligações telefônicas. Algumas forças policiais utilizam sistemas que gerenciam o fluxo de conversas interceptadas e troca de mensagens tipo SMS. A Polícia Federal utiliza o Sistema SIS – Sistema de Interceptação de Sinais para fazer a gestão desse tipo de informação, mediante as autorizações legais pertinentes. (Jornal do Brasil, 2010).

#### 3.2.6. Interceptação telemática

Outra fonte valiosa de informações refere-se à troca de informações eletrônicas, como fluxo de dados na web, envio e recebimento de mensagens eletrônicas (*e-mails*), histórico de navegação, utilização de chats ou de chamadas telefônicas sobre IP (VoIP). O grande desafio é reproduzir essas informações, que estão em diversos formatos e protocolos diferentes para uma linguagem entendível ao analista. O provedor de acesso à internet do investigado repassa ao órgão policial o fluxo de dados que trafega na conta do usuário que sofreu a quebra do sigilo telemático. Os pacotes de dados devem ser remontados e interpretados para integrá-los à investigação e posterior análise (Peron, 2012).

#### 3.2.7. Vigilância a alvos

Esses dados são oriundos basicamente de ações de campo realizadas pelo serviço operacional da instituição policial. São filmagens, fotografias e outros relatos produzidos pela equipe de investigação através de vigilância aos alvos, que comprovam ações, encontros, viagens e contatos de investigados.

### 3.2.8. Fontes abertas

As fontes abertas propiciam uma gama de informações de domínio público e se compõe, por exemplo, de dados oriundos das diversas mídias existentes (jornais, revistas), dados públicos em geral (relatórios governamentais), redes sociais, sítios de internet, imagens de satélite, mapas, pesquisas científicas e trabalhos acadêmicos. As fontes abertas são valiosas para o trabalho policial, pois podem fornecer apoio para assuntos relacionados à inteligência e contrainteligência (Leite, 2013).

Percebe-se que aos atores que trabalham com a coleta e análise de dados nas entidades de segurança pública está posto o desafio de extrair significado de inúmeras fontes de informação, cujas características se encontram em estágios diversos em relação a quesitos como facilidade de obtenção, tipo de formato, confiabilidade, grau de digitalização, dentre outros aspectos.

Entender toda a massa de informações, relacionando-as na tentativa de determinar as relações de autoria e materialidade de um suposto delito é o salto de qualidade que se espera na aplicação de novos conceitos ou ferramentas. O processo de trabalho tradicional, executado de forma estanque e sem integração, é ineficaz e custoso. O analista policial não tem mais capacidade para vasculhar e interpretar manualmente o gigantesco volume de dados passível de ser utilizado em um caso.

### **3.3. APLICAÇÃO DAS FONTES DE DADOS NO MODELO INFORMACIONAL**

Esta seção busca associar o modelo de informações proposto por Soares (2014), o qual foi detalhado na seção 2.5, às fontes de dados das polícias judiciárias listadas na seção anterior. Assim, é possível perceber claramente que o conjunto de dados utilizados como insumos, se organizados e sistematizados para tal fim, pode ser submetido a uma plataforma *big data* e gerar ganhos de produtividade para a investigação.

A Tabela 3.2 relaciona grupos de dados teóricos, pertencente ao modelo informacional estudado, a exemplos de atividades que podem ser realizadas caso haja a implantação de novos conceitos e ferramentas *big data*:

Tabela 3.2: Exemplo de rotinas realizadas com implantação do *big data*

<b>Grupo de dados no <i>big data</i></b>	<b>Rotinas</b>
Dados da web e mídias sociais	<ul style="list-style-type: none"> <li>• Extração de informações sobre suspeitos em sítios de busca na internet de modo a inter-relacioná-las às bases de dados existentes;</li> <li>• Pesquisa de relacionamento entre pessoas no Facebook ou LinkedIn, provando vínculo ou contato entre indivíduos ou organizações;</li> <li>• Pesquisa fotográfica na rede social Instagram, indicando a existência de patrimônio de um investigado (veículos, joias ou imóveis, dentre outros);</li> <li>• Localização de pessoas por meio de <i>logins</i> realizados em redes sociais, permitindo indicar que um indivíduo está ou já esteve num determinado local, podendo provar, inclusive, encontros ou reuniões.</li> </ul>
M2M	<ul style="list-style-type: none"> <li>• Análise do padrão de deslocamento de um investigado, por meio da transmissão de sua localização em tempo real emitida por rastreador instalado em seu veículo;</li> <li>• Interceptação telemática da rede wi-fi doméstica de um investigado e cruzamento de informações com suas atividades financeiras;</li> </ul>
Grandes dados transacionais	<ul style="list-style-type: none"> <li>• Cruzamento de informações cartorárias existentes nas bases de dados policiais que contém informações sobre investigados, indiciados, suspeitos e interrogados com dados oriundos de extratos telefônicos e bancários;</li> <li>• Cruzamento de inquéritos policiais e processos judiciais com informações constantes em laudos periciais, extraídas de sistemas de criminalística;</li> <li>• Análise e relacionamento de informações constantes em bancos de dados geridos por órgão policiais (passaporte,</li> </ul>

	carteira de identidade, registro de condutores de veículos e registro de armas de fogo);
Biometria	<ul style="list-style-type: none"> <li>• Cruzamento de análises de sêmen, sangue e impressões digitais com outras bases de dados;</li> <li>• Processamento de dados de reconhecimento facial extraídos de sistemas de controle de acesso ou câmeras de segurança, e seu confronto com bases de dados de identificação;</li> </ul>
Dados gerados por indivíduos	<ul style="list-style-type: none"> <li>• Realização de análises em anotações, cartas, memorandos ou outros documentos coletados por meio de um mandado de busca e apreensão, procurando relacioná-las a outras bases de dados da investigação.</li> </ul>

Após verificar os insumos informacionais que as polícias judiciárias brasileiras trabalham, bem como constatar a crescente demanda por um novo modelo de análise dos dados que suporte a investigação policial, é necessário propor um modelo de arquitetura de análise que viabilize essa nova forma de atuar. Além do volume, a velocidade e a variedade de formato desses dados, agrega mais obstáculos no que se refere ao trato com as informações.

### 3.4. APLICAÇÃO DE CARACTERÍSTICAS DO MODELO INFORMACIONAL NA ESTRUTURA DE INVESTIGAÇÃO

Além da adequabilidade dos insumos informacionais, as características apontadas no modelo informacional do *big data*, também descritas na seção 2.5 devem ser observadas. São aspectos importantes, tais como a organização dos dados, metadados, privacidade dos dados, qualidade dos dados, integração dos dados aos processos de negócios e gerenciamento do ciclo de vida da informação. A Tabela 3.3 detalha e dá exemplos da aplicação desses conceitos na estrutura de investigação policial.

Tabela 3.3: Aplicação das características do modelo *big data* na investigação

Aspecto	Observações e exemplo na estrutura de investigação
Organização	Os dados estão em diversos setores da instituição policial e cada conjunto de informações possui sua política de acesso diferenciada. Não

	são todos os policiais que acessam todas as investigações, sendo necessário a devida segregação de acordo com o nível de permissão de cada usuário.
Metadados	A organização deve pensar o <i>big data</i> também como uma ferramenta integradora das diversas bases de dados existentes: se um cidadão vai até um posto de expedição do passaporte e solicita um novo documento, os dados do endereço fornecido, por exemplo, devem ser atualizados e disponíveis para o policial que trabalha na delegacia de repressão a entorpecentes e está há meses tentando localizar o indivíduo.
Privacidade	Dados protegidos por sigilo, tais como informações fiscais, bancárias e telemáticas, devem ter seu acesso autorizado judicialmente, sob pena de invalidação de todo o trabalho investigativo. Além do aspecto técnico, a definição dos limites do uso das informações sigilosas manipuladas deve respeitar a privacidade dos cidadãos.
Qualidade	A interceptação do fluxo telemático de dados deve possuir um critério de aceitabilidade menor que o preenchimento de dados oriundos de um cadastro de motorista, por conta do grande volume, variedade e velocidade que são características do primeiro tipo de dados, por exemplo. A plataforma de dados deve, em cada caso, pesar os critérios de qualidade e oportunidade de utilização do dado.
Integração aos processos de negócio	Todo o fluxo informacional deve ser idealizado para dar suporte à elucidação de crimes, procurando abastecer as equipes de investigação com as informações coletadas e tratadas, respeitando, obviamente, a compartimentação e sigilo.
Gerenciamento do ciclo de vida informacional	Quando um cidadão revalida sua carteira de motorista e fornece um novo endereço, esse dado deve ser disponibilizado para a organização; contudo, seu endereço antigo deve estar disponível, pois os locais onde o indivíduo já morou são informações importantes caso ele seja investigado em alguma operação policial.

### **3.5. O FORMATO DE DADOS DA INVESTIGAÇÃO NA POLÍCIA FEDERAL**

Conforme explanado no capítulo 1, um dos desafios da chamada era do *big data* é trabalhar com dados com grande variedade de formatos. Nos capítulos 2 e nas seções do capítulo 3 até agora apresentadas, ficou esclarecido que os órgãos policiais enfrentam essa dificuldade, pois a dinâmica da investigação requer inúmeros insumos informacionais dos mais diversos tipos. A equipe de investigação não pode descartar nenhuma prova ou indício, sob pena de comprometer o trabalho final e deixar criminosos livres, condenar um inocente ou até mesmo inocentar erroneamente um indivíduo.

Para alcançar uma proposta de modelo de *big data*, deve-se aprofundar os estudos das fontes de informação, de modo a conhecer a estrutura do dado contido em cada uma delas. Optou-se por detalhar os insumos informacionais utilizados pela Polícia Federal que, além de possuir abrangência nacional, suas fontes de informação são mais acessíveis, dada a visibilidade de seus trabalhos, principalmente no que se refere a suas técnicas operativas e de obtenção do conhecimento.

A Polícia Federal possui representação nas 27 unidades federativas, contando com mais de 200 unidades espalhadas em todas as regiões do país<sup>10</sup>, fazendo com que a análise dos tipos de dados que transitam em suas investigações seja ideal para o objetivo aqui proposto, uma vez que se consegue realizar a padronização dos procedimentos investigatórios em todo território nacional. Cabe ressaltar que as outras instituições de polícia judiciária existentes no país possuem abrangência estadual ou distrital.

A seguir, são apresentados o detalhamento dos tipos de dados mais utilizados nos trabalhos de investigação levados a cabo pelo Departamento de Polícia Federal:

#### **3.5.1. Informações cartorárias**

As informações cartorárias são aquelas que resultam da atividade-fim da polícia judiciária e está diretamente relacionada à formalização das investigações em documentos e procedimentos previstos em lei.

---

<sup>10</sup> As unidades da Polícia Federal estão listadas em <http://www.pf.gov.br/institucional/pf-pelo-brasil/>

A Polícia Federal, por exemplo, utiliza em todas as suas unidades descentralizadas (delegacias e postos) o Sistema SISCART – Sistema de Controle de Atividades Cartorárias, responsável em dar suporte à dinâmica da instauração e processamento do inquérito policial e seus diversos documentos que o compõe. Cada instância do SISCART contém informações valiosas acerca das investigações em curso, arquivadas ou encaminhadas à justiça para apreciação.

Quando um indivíduo é preso em flagrante, as informações relacionadas à sua qualificação (nome, nome dos pais, CPF, data de nascimento, endereços, dentre outros), bem como histórico do ocorrido (data, resumo), as circunstâncias da prisão, fotografia dos envolvidos, relacionamento existentes com outros criminosos e os objetos apreendidos com o suspeito (drogas, numerário, joias, veículos) são formalizadas em diversos documentos que são inseridos diretamente no SISCART.

Diariamente, o órgão processa as bases de dados locais do SISCART por meio do sistema ETL UNICART – União dos SISCART's do DPF, realizando a extração das informações cartorárias de cada instância local do SISCART e consolidando-as em uma base de dados unificada. Dessa forma, todas as informações inseridas e atualizadas nas delegacias do órgão, referentes a inquéritos policiais, apreensões, termos circunstanciados, partícipes de uma eventual ocorrência (presos, envolvidos ou testemunhas), guias de identificação, mandados de prisão, fotografias, dentre outros itens, encontram-se centralizadas em um único local, cuja implementação é um banco de dados MySQL. (DEPARTAMENTO DE POLÍCIA FEDERAL, 2015).

Segundo dados contidos no Relatório de Gestão do Departamento de Polícia Federal, referente ao exercício de 2013, foram instaurados naquele ano mais de 66 mil inquéritos policiais em todas as unidades do órgão espalhadas pelo Brasil. (MINISTÉRIO DA JUSTIÇA, 2014).

### 3.5.2. Base de dados de passaporte

Uma fonte de dados importante provém dos próprios serviços fornecidos pelos órgãos de segurança aos cidadãos, mesmo que eles não estejam diretamente relacionados à sua atividade fim. A Polícia Federal é responsável, no território nacional, pelo serviço de expedição do passaporte, documento que é utilizado pelo cidadão para realizar viagens e se identificar no exterior.

A base de dados do passaporte apresenta uma taxa de 2 milhões de documentos registrados por ano, apresentando razoável crescimento. Os dados, referentes ao período compreendido entre 2009 e 2013, confirmam a emissão de cerca de 9 milhões de documentos.<sup>11</sup>

O requerimento do passaporte é realizado no sítio eletrônico do Departamento de Polícia Federal ([www.dpf.gov.br](http://www.dpf.gov.br)), cujos formulários solicitam as seguintes informações: dados pessoais (nome, nome dos pais, raça, dentre outros), local de nascimento, documentos diversos (CPF, títulos de eleitor e certidão de casamento, dentre outros), endereço e telefone. Os dados gravados no documento incluem, além das informações pessoais, os dados biométricos (fotografia facial e duas impressões digitais) que são coletados quando o cidadão vai pessoalmente confirmar a solicitação do documento nos postos de atendimento.

Os serviços de expedição de passaporte, além de cumprir a função do estado no que se refere à expedição de documento de viagem, produzem uma base de dados de cidadãos bastante interessante, contendo informações pessoais detalhadas e dados biométricos do indivíduo.

### 3.5.3. Laudos periciais

Quando um delito ocorre e deixa vestígios, o perito criminal é o responsável por fazer análise do material relacionado, com o intuito de apontar a autoria do crime mediante análises técnicas. O resultado desse trabalho é formalizado em um laudo pericial. Se um agente público, por exemplo, adquire grandes somas em dinheiro e envia tais valores ao exterior, com o intuito de realizar sonegação fiscal e evasão de divisas, é o perito que vai apontar, através de relatórios contábeis, extratos de movimentação financeira e outros documentos, a materialidade dos fatos.

Outro exemplo ocorre quando um fazendeiro pratica o desmatamento de uma propriedade causando danos ambientais. O perito criminal da área de meio ambiente é designado para realizar as análises e elaborar o respectivo laudo que vai comprovar os prováveis danos ocorridos na região, apontando suas origens, extensões e consequências para o meio ambiente.

---

<sup>11</sup> Disponível em <http://g1.globo.com/brasil/noticia/2014/01/policia-federal-tem-emissao-recorde-de-passaportes-em-2013.html>

Segundo (Rodrigues et al., 2010), a base de laudos periciais é muito importante para a investigação, uma vez que esse tipo de documento descreve com detalhes a cena do crime, analisa e interpreta as evidências e a dinâmica dos fatos, lançando mão de recursos como fotografias, croquis, mapas e demais elementos que esclarecem a dinâmica dos fatos.

A Polícia Federal conta com o Sistema Nacional de Gestão de Atividades de Criminalística, que facilita a gestão da atividade de elaboração e expedição dos laudos periciais produzidos pelos peritos criminais federais.<sup>12</sup> O sistema gere uma base de conhecimentos técnicos a partir de trabalhos periciais realizados em todas as unidades federais de criminalística do país. O acervo eletrônico contém mais de 110.000 laudos periciais produzidos por quase 1.000 profissionais das seguintes áreas do conhecimento: documentoscopia, informática, laboratório, merceologia, balística, veículos, audiovisual e eletrônicos, contábil e econômica, meio ambiente, local de crime, engenharia, bombas e explosivos, biometria forense, medicina legal, patrimônio histórico, artístico e cultural e genética forense.

A base de laudos periciais é essencial para uma unidade de investigação porque oferece um histórico de análises de vestígios dos mais diversos tipos de crimes cometidos e que foram submetidos à perícia.

#### 3.5.4. Impressões digitais

Segundo (Garrido, 2009) a utilização da datiloscopia (identificação por meio de impressões digitais) ocupa importante lugar no contexto da identificação do indivíduo. Quando uma impressão digital é encontrada num local ou objeto de crime, seu processamento e análise podem associá-la a um indivíduo específico e, por consequência, agregar indícios importantes a uma determinada investigação.

A comparação do material coletado com uma base de dados de impressões digitais pode ser realizada de forma automatizada, aumentando a confiabilidade e produtividade dessa atividade. A Polícia Federal utiliza o sistema AFIS (*Automated Fingerprint Identification System*) para identificar de forma única os detentores de passaporte e cidadãos indiciados em procedimentos criminais, potencializando as formas de identificação civil e criminal de pessoas,

---

<sup>12</sup> Sistema Nacional de Gestão de Atividades de Criminalística do Departamento de Polícia Federal. Disponível em: <http://repositorio.enap.gov.br>

pois utiliza algoritmos de comparação e ferramentas capazes de processar imagens de impressões digitais e seus fragmentos<sup>13</sup>.

A Polícia Federal contava no ano de 2013 com um banco de dados de 13 milhões de impressões digitais e espera, nos próximos 10 anos, aumentar a capacidade de sua base de dados para 180 milhões de registros. (DPF, 2013)

Em relação ao formato de armazenamento de impressões digitais, existe o padrão ANSI/NIST que regula a conversão das imagens coletadas para o padrão binário, permitindo o intercâmbio dessas informações entre diversas bases de dados ou sistemas.

### 3.5.5. Informações bancárias

A quebra do sigilo bancário de um suspeito, após a devida autorização judicial, é processada por meio do SIMBA - Sistema de Investigação de Movimentação Bancária -, que permite o tráfego de informações entre as instituições financeiras e órgãos públicos solicitante, como a Polícia Federal. A partir de então, o investigado terá suas informações bancárias compiladas pelo banco ao qual possui conta vinculada e as informações de interesse da investigação em curso serão enviadas, via *hash* criptografado, à unidade de investigação. (PGR, 2012).

O Banco Central normatiza e padroniza o leiaute das informações a serem repassadas pelas instituições financeiras. A cada solicitação de quebra de sigilo bancário, o órgão policial deverá receber 5 (cinco) arquivos em formato texto, padrão ASCII, contendo as seguintes informações (Banco Central do Brasil, 2010):

1. Agência: dados da agência que mantêm as contas nas quais ocorreram as transações financeiras;
2. Contas: contas bancárias que tiveram o sigilo afastado pela decisão judicial;
3. Titulares: pessoas ou empresas titulares das contas bancárias e aquelas com as quais foram realizadas as transações;
4. Extrato: lançamentos realizados nas contas investigadas; e

---

<sup>13</sup> Monitoramento do Sistema Automatizado AFIS. Disponível em <http://portal2.tcu.gov.br/portal/pls/portal/docs/2639809.PDF>

5. Origem/destino: origem ou destino de recursos relacionados aos lançamentos discriminados no extrato.

#### 3.5.6. Interceptação telefônica

As chamadas telefônicas registradas entre dois indivíduos geram dois tipos de informações úteis a uma investigação. O primeiro deles, conforme explica o *site* Direitonet (2015), trata-se da interceptação telefônica, ou seja, o conteúdo em áudio da conversa pessoal, em que nenhum dos interlocutores tem ciência da ação. O segundo tipo refere-se às informações associadas à conversa, como números de telefone, titulares das linhas e o momento temporal em que a comunicação foi realizada. Tais ações devem obrigatoriamente ser precedidas de autorização judicial.

Assim, o formato dos dados relativos a essa prática investigativa divide-se em três tipos:

6. Arquivos de áudio, referente à gravação das conversas realizadas;
7. Dados estruturados repassados pela operadora telefônica, contendo informações similares a um extrato telefônico, como número de origem da ligação, número de destino, data e horário da chamada, duração da chamada, local da chamada (ERB utilizada);
8. Transcrições das conversas, conforme determina a Lei 9296/96<sup>14</sup>, que deve ser feita pela equipe responsável pela condução da investigação.

#### 3.5.7. Interceptação telemática

Os procedimentos de interceptação do fluxo telemático também estão regulados pela Lei 9296/96. A equipe de investigação solicita ao juiz autorização para proceder com a interceptação do fluxo de dados do investigado ao vislumbrar a possibilidade de coletar mais provas que, por esse método, permitirão complementar os procedimentos tradicionais de investigação.

---

<sup>14</sup> A Lei 9296/96, de 24 de julho de 1996, regulamenta a interceptação de comunicações telefônicas e telemáticas.

Segundo (Wendt, 2013), a interceptação do fluxo de comunicações telemáticas acontece de duas maneiras: a) quando o provedor de acesso do investigado grava o fluxo de comunicação e o repassa ao órgão investigador; ou b) quando ocorre a duplicação do fluxo de acesso do indivíduo, que é espelhado para outra máquina, permitindo que a polícia acesse o conteúdo dos dados trafegados. Geralmente, nos dois casos, ficam à disposição das equipes de investigação um fluxo de pacotes IP para serem remontados e analisados.

(Peron, 2012) propõe uma metodologia de análise do conteúdo interceptado, possibilitando a recuperação de diversos tipos de informações trocadas, dentre elas, objetos de webmail, chats diversos, conteúdo de páginas da internet, conteúdos de mensagens eletrônicas (*e-mails*) trocados, dentre outros artefatos.

### 3.5.8. Dados das redes sociais

A coleta de dados realizada em redes sociais usualmente é realizada por meio das APIs (*Application Programming Interface*), recursos que permitem a utilização de rotinas de programação pré-definidas que retornam determinadas informações de aplicações ou bases de dados. Os mantenedores de redes sociais disponibilizam as APIs como um ponto de interface entre aplicativos externos de consulta ou atualização e as informações armazenadas internamente.

Instagram é uma rede social para compartilhamento de imagens. A partir de um identificador único é possível, por meio de APIs existentes, obter informações dos indivíduos inscritos na plataforma, tais como usuários seguidos e seguidores, dados pessoais cadastrados na rede social e configurados como públicos por seu proprietário, imagens postadas e sua geolocalização associada, registros e comentários diversos. (Manikonda et al., 2014).

Na rede social Twitter, é possível coletar bem mais que as tradicionais publicações de até 140 caracteres, característica marcante do microblog. Além do conteúdo, as publicações vêm acompanhadas de dois conjuntos de dados: entidades e locais. As primeiras relacionam-se com *hashtags*, URL's e outras mídias que estão associadas à publicação; os segundos relacionam-se especificamente à localização onde a publicação foi realizada. (Russel, 2013).

Em relação à rede social Facebook, de maior difusão mundial, o conjunto de aplicativos existentes permite coletar uma série de informações vinculadas a um indivíduo, os quais são unicamente identificados por meio de um código (Id). A partir desse código é possível

expandir as informações sobre relação de amizade, interesses e dados pessoais configurados como público pelo usuário, além de publicações, locais visitados, dentre outras informações. (Santos, 2014).

A rede social LinkedIn é especializada em manter relações de trabalho, formação profissional e negócios. Os dados possíveis de serem coletados incluem o histórico educacional, histórico profissional, qualificações realizadas e relacionamento entre profissionais e organizações. (Russel, 2013).

A manipulação automatizada de informações contidas em redes por meio das APIs abre a possibilidade de obtenção de dados dos mais variados tipos e formatos, permitindo ganho na prospecção de conhecimento existentes nesses tipos de plataforma. A título de padronização, as informações capturadas podem ser gravadas em coleções de dados binários (*blobs*), indexada por meio de identificadores, permitindo uma formatação flexível e independente de linguagem, que pode ser utilizada para intercâmbio de dados numa eventual futura etapa de processamento ou armazenamento em repositório de dados.

### 3.5.9. Mineração na Web

A mineração de dados oriundos da internet objetiva a extração de informações úteis de suas páginas e repositório de dados, os quais lidam com enorme volume e variedade de informações. Algumas observações sobre essa atividade são feitas por (Liu & Chen-Chuan-Chang, 2004), a saber:

1. os dados existentes na internet são muito grandes e vem crescendo exponencialmente;
2. tais dados são amplos e diversos, podendo-se encontrar informações de praticamente todos os assuntos ou pessoas;
3. na web, existem dados de diversos tipos (estruturados, texto, dados multimídia, dentre outros);
4. a informação na web é heterogênea, ou seja, os conteúdos similares são apresentados de forma diversa por diferentes páginas;
5. muita da informação na web é ligada a outras informações através de links para páginas e objetos;
6. a web é composta não apenas por páginas superficiais, mas também por bases de dados que podem ser acessadas através de consultas especiais;

7. as informações na web são dinâmicas, ou seja, mudam constantemente.

Após a utilização de ferramenta própria para a mineração de dados, inserindo como parâmetro de busca o nome de investigados específicos, empresas, imóveis, eventos, notícias, dentre outras tantas características, é possível coletar informações relacionadas a indivíduos e armazená-las em unidades de objetos binários.

Uma alternativa para padronizar a coleta é a utilização do formato JSON (*JavaScript Object Notation*), amplamente empregado em diversas plataformas web, inclusive pelos mais populares websites de redes sociais. Esse padrão também foi escolhido por inúmeras empresas que desenvolvem soluções de banco de dados NoSQL. (ODM, 2014).

#### 3.5.10. Dados cadastrais diversos

Companhias de fornecimento de água, luz, gás, serviços de telefonia e TV à cabo, companhias aéreas, seguradoras e órgãos públicos são importantes fontes de informações cadastrais. Os dados provenientes dessas empresas, mediante autorização judicial, ajudam a robustecer o processo de investigação. É possível cruzar o histórico de endereços das pessoas, informações patrimoniais (imóveis) e outros de dados existentes nas bases de dados das empresas com as informações coletadas por outras fontes (interceptação telefônica, vigilância pessoal, dados oriundos de redes sociais, dentre outros).

Companhias de grande porte possuem geralmente um banco de dados relacional central que suportam as diversas operações da empresa, como estoque e suprimento, manufatura, finanças, vendas e distribuição, atividades de recursos humanos e módulos de serviços e relacionamento com clientes (Souza, 2000). As informações geralmente estão em formatos estruturados e podem ser enviados aos órgãos de investigação por meio de planilhas ou tabelas formatadas.

#### 3.5.11. Atividades operacionais

As atividades operacionais são compostas pelas diligências realizadas por agentes policiais e objetivam registrar encontros, acontecimentos ou fatos de interesse para uma determinada investigação. O resultado dessas diligências são geralmente registros fotográficos, filmagens e relatórios descritivo dos fatos.

A título de ilustração, quando uma operação em curso está lidando com crimes relacionados à corrupção, uma equipe policial é designada para coletar dados dos investigados, por meio de vigilância sobre os envolvidos e perseguições veladas, buscando conhecer a rotina, o comportamento e o *modus operandi* dos investigados frente à dinâmica das supostas práticas delituosas. Os registros fotográficos e filmagens podem ser executados pelos próprios investigadores ou são coletados por meio de câmeras de segurança de aeroportos, restaurantes, hotéis, *shopping centers* ou qualquer outro estabelecimento que possua equipamento de filmagem instalado.

O Portal de Notícias G1 detalha uma coleta de informações realizada pela Polícia Federal no bojo da Operação Lava Jato, ocasião em que o circuito interno do prédio onde funciona uma empresa registrou a chegada de investigados, os quais transportavam sacolas e documentos; fatos esses vinculados à investigação em tela<sup>15</sup>.

Ao final da diligência, o relatório que será anexado à investigação pode conter, dentre outros dados, os envolvidos identificados, o dia e horário dos fatos, sua localização, bem como os registros comprobatórios da ação coletados (arquivos de imagens ou gravações).

Após a apresentação e descrição dos insumos informacionais que podem compor uma investigação policial, a Tabela 3.4 sugere a utilização dos formatos dos dados relacionados a cada tipo descrito. Conforme visto anteriormente de forma geral e, nesta seção, especificamente na Polícia Federal, percebe-se que uma infraestrutura de *big data* voltada para as polícias judiciárias deve se atentar à grande diversidade de formatos existentes. A Tabela 3.4 busca sintetizar o que foi discutido neste capítulo.

Tabela 3.4: Formato dos dados dos insumos informacionais do DPF

<b>Insumo informacional</b>	<b>Formato de dados</b>
Informações cartorárias	<ul style="list-style-type: none"> <li>• Dados estruturados em base de dados relacionais</li> </ul>

<sup>15</sup> Disponível em <http://g1.globo.com/pr/norte-noroeste/noticia/2015/06/operacao-ferrari-prende-chefes-de-organizacao-criminosa-diz-pf.html>

	<ul style="list-style-type: none"> <li>• Dados em formato de imagens</li> </ul>
Base de dados de passaporte	<ul style="list-style-type: none"> <li>• Dados estruturas em base de dados relacionais</li> <li>• Dados em formato de imagem</li> <li>• Dados em formato de impressão digital</li> </ul>
Laudos periciais	<ul style="list-style-type: none"> <li>• Dados estruturados do sistema</li> <li>• Documentos em formato não estruturados (PDF)</li> </ul>
Impressões digitais	<ul style="list-style-type: none"> <li>• Dados em formato de impressão digital (Padrão NIST)</li> </ul>
Dados bancários	<ul style="list-style-type: none"> <li>• Arquivos texto em formato estruturado</li> </ul>
Interceptação telefônica	<ul style="list-style-type: none"> <li>• Dados estruturados do extrato telefônico</li> <li>• Dados em formato de áudio</li> </ul>
Interceptação telemática	<ul style="list-style-type: none"> <li>• Blocos de dados não estruturados</li> </ul>
Dados das redes sociais	<ul style="list-style-type: none"> <li>• Blocos de dados não estruturados</li> </ul>
Mineração na Web	<ul style="list-style-type: none"> <li>• Blocos de dados não estruturados</li> </ul>
Dados cadastrais diversos	<ul style="list-style-type: none"> <li>• Dados estruturados em bases relacionais</li> </ul>
Informações de atividades operacionais	<ul style="list-style-type: none"> <li>• Dados não estruturados</li> <li>• Documentos em formato de imagem</li> <li>• Documentos em formato de vídeo</li> </ul>

## 4. A PLATAFORMA *BIG DATA* x INVESTIGAÇÃO

Conforme descrito na Seção 2.3, os bancos de dados relacionais não são tecnologias capazes de superar o novo paradigma de armazenamento e processamento baseado em *big data*. Devido às características requeridas, tais tarefas deverão ficar a cargo de soluções baseadas em NoSQL.

Neste contexto, este capítulo visa analisar os tipos de bancos de dados NoSQL mais utilizados no mercado, levando-se em consideração fatores relacionados aos insumos informacionais, tipos de dados utilizados e a necessidade de novo paradigma tecnológico para suportar as atividades de investigação policial. Após a análise de empregabilidade dos tipos de repositório NoSQL, buscou-se realizar um comparativo entre modelos de modo a apontar qual o mais adequado a suprir as necessidades da investigação.

### 4.1. TIPOS DE BANCO DE DADOS NOSQL

Os bancos de dados NoSQL são uma alternativa às plataformas clássicas relacionais no que se refere a aplicações que manipulam grandes volumes de dados. A necessidade de escalar o servidor de dados e como tais bancos conseguem realizar essa tarefa com mais simplicidade e eficiência, colocam-nos como figuras centrais num modelo de *big data*.

Cattel (2011) examina os tipos mais comuns de banco de dados NoSQL, ressaltando que não existe uma solução que seja a melhor para suportar todas as necessidades do usuário. Segundo sua análise, deve haver uma priorização de modelo dependendo da aplicação requerida.

No mercado, existem cerca de 300 tipos de bancos de dados, dos quais quase 100 podem ser enquadradas como soluções NoSQL<sup>16</sup>. Esse panorama inviabiliza, para os propósitos deste trabalho, a análise de cada ferramenta individualmente. Dessa forma, optou-se por realizar estudos das bases de dados NoSQL disponíveis de acordo com a classificação consagrada pela literatura e corroborada por Sadalage & Fowler (2013): bancos de dados chave-valor, banco de dados de documentos, banco de dados de família de colunas e banco de dados de grafos. As

---

<sup>16</sup> Listagem disponível em [http://db-engines.com/en/ranking\\_categories](http://db-engines.com/en/ranking_categories)

subseções 4.2.1 a 4.2.4 abordam suas vantagens e desvantagens e apresenta, a título de ilustração, exemplos de registros de investigados num eventual trabalho policial.

#### 4.1.1. Banco de Dados Chave-Valor

Segundo Moniruzzaman & Hossain (2013), os bancos de dados desse tipo são considerados os mais fáceis de se utilizar, pois seu nível de abstração é bastante simples. Inserir, apagar ou consultar um conjunto de dados é realizado única e exclusivamente por meio de uma chave. O valor retornado nessas operações é um conjunto binário de dados que o depósito de dados apenas armazena, sem se preocupar em entender seu conteúdo. O detalhamento do que é manipulado, ou seja, o valor mapeado pela chave, é de inteira responsabilidade da aplicação.

A utilização dessa base de dados num modelo investigativo, levando-se em conta informações exemplificadas do registro de um investigado, construiria um modelo equivalente à Figura 4.1, em que a chave primária da tabela seria o CPF do indivíduo (é um atributo que identifica unicamente uma pessoa). O restante dos atributos (nome, número do RG, data de nascimento, nacionalidade e cor da pele) constituiria o valor retornado pela chave, visto pelo repositório apenas o compreende como um mero conjunto binário de dados.

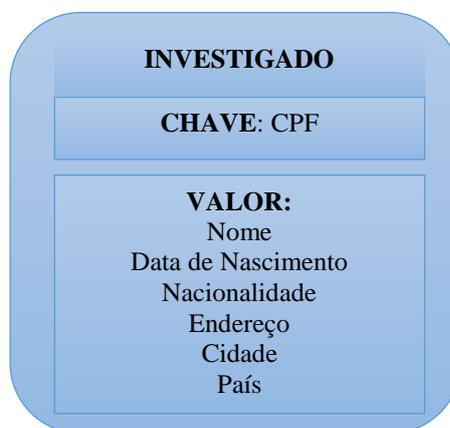


Figura 4.1: Exemplo de Registro no Banco de Dados de Chave-Valor

Claramente, vários obstáculos a um modelo investigativo seriam colocados caso esse modelo seja adotado. Não seria possível uma aplicação de usuário consultar todos os investigados que possuem como critério, por exemplo, um determinado nome ou parte dele. Da mesma forma, não se executariam pesquisas que levariam em consideração outros atributos como a data de nascimento de uma pessoa ou a cidade em que ela mora, pelo fato de que tais informações estarem na parte do registro cujo motor de busca e processamento do banco de

dados enxerga-o apenas como uma massa binária de dados. Outra decorrência dessa limitação é a inviabilidade de implantar relacionamento entre os diversos registros da base de dados, o que impede a realização de consultas mais complexas.

Sadalage and Fowler (2013) sugere alguns casos apropriados para uso dos bancos de dados chave-valor: armazenamento de informações de sessão do usuário, controle de preferências em perfis de usuário e dados de carrinhos de compras em sites de comércio eletrônico. Para tais implementações, o funcionamento de bancos de dados chave-valor realiza a manipulação de registros por meio de um identificador único, não sendo primordial a atualização ou consulta direta sobre qualquer um dos demais atributos de cada registro.

Os bancos de dados do tipo chave-valor até permitem, em alguns casos, a pesquisa por múltiplos atributos; contudo, tal solução exigiria a criação de um índice adicional mantido pelo lado do cliente, por meio de sua própria aplicação, o que não é o ideal quando existe a perspectiva da manipulação de grandes repositórios de informação com múltiplos atributos.

#### 4.1.2. Banco de Dados de Documento

Os bancos de dados de documentos armazenam e recuperam informações estruturadas em documentos, podendo-se valer de formatos tais como JSON (*JavaScript Object Notation*), XML (*Extensible Markup Language*) e outros. Esses bancos permitem a manipulação de estruturas de diferentes tipos, e seus registros podem ser pesquisados também por diferentes atributos. Esses repositórios também são desprovidos de esquema, ou seja, permitem flexibilidade na construção e modificação do modelo ao longo do tempo, bastando apenas a inclusão ou exclusão de atributos pertencentes a cada documento (Kaur et al., 2013).

O funcionamento do banco de dados de documento possui similaridades com o banco de dados de chave-valor, contudo, há uma diferença fundamental: naqueles, a parte armazenada pode ter seu valor examinado e manipulado pelo próprio repositório de dados, por meio de índices secundários mantidos pela própria base; nestes, conforme visto na subseção 4.1.1, os valores de atributos são vistos como uma grande massa binária. Outra vantagem dos primeiros é que parte dos dados do documento pode ser consultada sem a necessidade de recuperar todo o objeto para posterior análise.

Uma investigação policial poderia, dessa maneira, trabalhar como critério de pesquisa e manipulação todos os atributos contidos no esquema. Seria possível, por exemplo, pesquisar

indivíduos que possuem certa nacionalidade e moram em uma determinada cidade e que possuem uma profissão específica. Outros critérios como nome e data de nascimento também poderiam ser combinados, não estando os procedimentos de consulta e recuperação restritos apenas ao identificador da pessoa – o CPF. Um exemplo do registro de dados de um investigado modelo nesse tipo de repositório é apresentado na Figura 4.2.



Figura 4.2: Exemplo de Registro no Banco de Dados de Documentos

Abramova et al. (2014) sugere a utilização de bancos de dados orientados a documentos para a manipulação de documentos que podem ser armazenados em formatos de dados estruturados, como texto, e-mails e arquivos XML, por exemplo. A utilização de aplicações em que o foco não é relacionamento entre as entidades é preferível, pois cada documento é visto de forma estanque e o relacionamento entre eles é realizado por meio de migração de chaves sem qualquer garantia de integridade.

A princípio, os repositórios orientados a documentos apresentam características positivas para se adequarem ao modelo investigativo baseado em *big data*, tais como permitir a pesquisa por qualquer atributo de uma entidade e suportar boa flexibilidade de esquema. Contudo, a aplicação dessas soluções apresenta dificuldades quando o problema a ser resolvido é composto por estruturas ou entidades cujo relacionamento é basicamente não-hierárquico, pois vão exigir um custoso gerenciamento de chaves estrangeiras entre os diversos documentos do banco de dados (McCreary, 2013). Essa implementação não seria eficiente no processamento de consultas mais complexas, que envolvam a inter-relação de diversos tipos de entidade (documento) presentes numa investigação, tais como indivíduos, empresas, patrimônio, locais e eventos, dentre outros. À medida em que a base de dados cresce, a dificuldade de gerenciar

esses relacionamentos entre as entidades vai se tornando outro obstáculo, pois as chaves estrangeiras que se relacionam a outros documentos devem ser implementadas por meio da aplicação do usuário.

#### 4.1.3. Banco de Dados de Família de Colunas

O banco de dados de família de colunas possui um conceito um pouco parecido com os bancos de dados relacionais; porém, ao contrário desses, sua arquitetura é moldada para permitir o mapeamento de chaves para valores agrupados em colunas específicas. Cada família de colunas, portanto, é um mapeamento de dados que pode ser acessado por meio de uma chave.

A importância do armazenamento desse tipo deve-se ao fato de que, por muitas ocasiões, é necessário acessar informações específicas a respeito de um determinado registro. Por exemplo, a respeito de um investigado por crimes financeiros, é mais importante ter o acesso a determinadas características, tais como nome, data de nascimento, profissão e número do passaporte. Outros dados, como cor da pele e altura, não são tão necessários para os investigadores desse caso, mas podem ser primordiais para a resolução de um homicídio, em que as características físicas do suspeito são atributos muitas vezes confrontados com testemunhos, fotografias, retratos falados ou filmagens. A Figura 4.3 ilustra dois registros de investigados, sob a ótica de cada um dos casos descritos, num repositório do tipo família de colunas.

<b>INVESTIGADO</b>				
<b>CHAVE</b>	<b>COLUNA 1</b>	<b>COLUNA 2</b>	<b>COLUNA 3</b>	<b>COLUNA 4</b>
<b>CPF</b>	<b>NOME</b>	<b>PASSAPORTE</b>	<b>DATA NASCIMENTO</b>	<b>PROFISSÃO</b>
<b>CHAVE</b>	<b>COLUNA 1</b>	<b>COLUNA 3</b>	<b>COLUNA 5</b>	<b>COLUNA 6</b>
<b>CPF</b>	<b>NOME</b>	<b>DATA NASCIMENTO</b>	<b>COR DA PELE</b>	<b>ALTURA</b>

Figura 4.3: Exemplo de Registros no Banco de Dados de Família de Colunas

Os bancos de dados do tipo família de colunas, diferentemente do modelo orientado a documentos – que é mais adequado ao relacionamento hierárquico entre entidades – dispõem

de um esquema de partição mais eficiente, pois cada família de colunas pode ser comparada a um contêiner de linhas em uma tabela de banco de dados relacional, em que a chave identifica a linha que é constituída a partir de múltiplas colunas específicas. A diferença é que as linhas não necessitam, necessariamente, de ter as mesmas colunas, e novas colunas podem ser adicionadas a qualquer linha, a qualquer momento, sem a mesma obrigatoriedade de adicioná-las às outras linhas também. Assim, é possível agrupar colunas em família de colunas que são frequentemente manipuladas em conjunto.

Segundo Nayak et al. (2013), esses tipos de repositórios, pela forma como organizam os dados internamente, oferecem alto poder de escalabilidade, além de otimizar os processos internos de I/O, o que reflete numa ótima performance.

Contudo, Sullivan (2015) pontua que existem desvantagens na utilização do banco de dados de família de colunas principalmente em relação ao processamento de consultas que envolvam múltiplas entidades e requerem *queries* mais complexas. As informações sobre uma entidade são armazenadas, geralmente, de forma desnormalizada em uma única tabela, o que dificulta de sobremaneira a gestão dos relacionamentos.

#### 4.1.4. Banco de Dados de Grafo

Diferentemente dos três tipos de bancos de dados NoSQL apresentados nas subseções 4.1.1, 4.1.2 e 4.1.3, o banco de dados de grafo está diretamente relacionado a um modelo de dados estabelecido e bem conhecido: o modelo de grafos. Sua concepção é representar os dados e/ou o esquema dos dados como grafos dirigidos, ou como estruturas que permitem e generalizam a noção de grafos.

A opção para utilização do modelo de grafos é mais interessante quando pretende-se focar em informações sobre a interconectividade e que, muitas vezes, a própria topologia dos dados é mais ou tão importante para a aplicação quanto o conteúdo dos dados propriamente ditos. No lugar de tabelas com linhas e colunas e a rigidez própria da estrutura relacional, um modelo gráfico flexível é utilizado, permitindo, inclusive ser escalado por meio de várias máquinas (Angles & Gutierrez, 2008).

O modelo baseado em grafos possui três componentes básicos: a) as entidades, que são os vértices do grafo; b) os relacionamentos, que são as arestas e c) as propriedades, que abrigam

os atributos dos vértices e seus relacionamentos. A Figura 4.4 apresenta o relacionamento entre registros de investigados, aplicando a arquitetura de dados baseada em grafos:

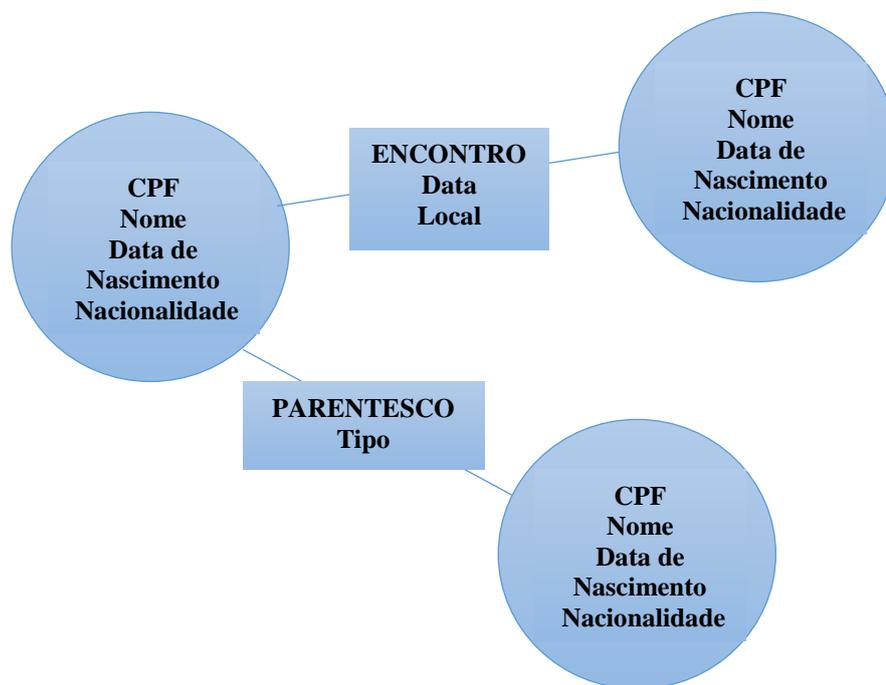


Figura 4.4: Exemplo de Registros no Banco de Dados Orientado a Grafos

Esse grafo representa a modelagem de duas relações entre três investigados num caso hipotético. A primeira relação é um encontro entre duas pessoas, cujos atributos são data e local. A segunda relação descreve um parentesco entre indivíduos, em que o tipo é descrito num atributo específico.

Os bancos de dados de grafos são especializados em manipular informações que apresentam relacionamentos complexos. O modelo relacional tem desempenho degradado à medida em que se aprofundam as inter-relações e exigem o processamento de operações entre tabelas (*joins*). A ênfase nos relacionamentos torna os bancos de dados de grafos diferentes de outras arquiteturas NoSQL; contudo, ressalta-se que eles também seguem os preceitos básicos dessa categoria, qual seja, a rejeição às características do modelo relacional.

O banco de dados de grafos apresenta a desvantagem nas situações em que deseja realizar consultas e atualizações de entidades de um mesmo tipo, pois essas demonstram-se serem mais custosas do que quando efetuadas em outros tipos de plataformas, uma vez que os vértices de um mesmo tipo (investigados, por exemplo) não estão fisicamente agrupados, como

num modelo relacional ou de família de colunas, onde os dados dos investigados estão presentes em uma única tabela.

McCreary (2013) pontua que, em termos de escalabilidade, os bancos de dados orientados a grafo possuem performance pior que outras soluções NoSQL, principalmente quando o grafo a ser trabalhado não está na memória RAM. Contudo, apenas com o aumento desse recurso, esse tipo de repositório tem capacidade de oferecer bom poder computacional, o que viabiliza sua utilização como parte de uma solução *big data*. Cabe ressaltar que, historicamente, os preços por *megabyte* de memória RAM vêm sofrendo grande queda – ficaram cerca de 50 vezes mais baratos ao longo dos últimos 10 anos<sup>17</sup> - tornando-se um recurso bastante acessível e relativamente barato para as organizações.

#### **4.2. A PLATAFORMA DE DADOS INVESTIGATIVA BASEADA EM NOSQL**

Após a análise de modelos de bancos de dados e a percepção de que o emprego do modelo relacional não é o mais adequado para uma plataforma integrada para investigações, por uma série de motivos já destacados na seção anterior, resta a avaliação de qual dos modelos de bases de dados NoSQL, dos quatro apresentados, poderia melhor suprir as necessidades de uma investigação policial levadas em consideração neste trabalho.

Primeiramente, ao resgatar pontos discutidos no capítulo 3, relativamente às fontes e insumos informacionais, percebe-se que um importante requisito para as atividades desempenhadas pela polícia é a necessidade de privilegiar o relacionamento dos dados, uma vez que é a partir desse procedimento que as análises e novas descobertas sobre um determinado caso são potencializadas e expandidas. Em suma, a investigação policial trabalha basicamente com o cruzamento de dados para produzir conhecimento, de modo a chegar à autoria de um crime ou, em determinadas circunstâncias, apontar novos direcionamentos em busca desse objetivo.

Na prática, não é difícil perceber esse dilema. Com o advento das redes sociais, a existência de diversas informações disponíveis com certa facilidade nas fontes abertas, a proliferação dos aparelhos de telefonia celular, a expansão da capacidade de gravação de

---

<sup>17</sup> Histórico disponível de 1957 a 2015 em <http://www.jcmit.com/memoryprice.htm>.

imagens e vídeos, a informatização, de um modo geral, das organizações públicas e privadas, fica cada vez mais difícil esconder vínculos entre uma pessoa e qualquer outra entidade, seja ela um outro indivíduo, uma empresa ou um patrimônio. O complicado é aglutinar essas fontes de informação que estão em formatos diversificados e se atualizam constantemente, no intuito de se permitir montar uma rede de vínculos que possa, de fato, provar um encontro, uma transferência de valores, uma aquisição patrimonial, ou uma série de outros eventos que constituem o conjunto probatório de uma investigação.

Dessa forma, a avaliação e a adequabilidade do tipo de base de dados deve focar em três aspectos principais, os quais estão presentes na essência do trabalho de concatenação das informações recebidas pela polícia: a) suporte ao relacionamento entre entidades, tendo em vista ser o produto final de uma investigação a prova de vínculos entre pessoas e entidades; b) bom desempenho na exploração e consulta entre esses relacionamentos, e c) bom nível de escalabilidade dos dados, uma vez que, conforme demonstrado neste trabalho, o aumento das possibilidades de coleta certamente obrigará a plataforma a armazenar volumes cada vez maiores de informação.

Cabe alertar que, assim como qualquer tecnologia ou sistema a ser definido em uma organização, a implantação de soluções baseadas em NoSQL não possui apenas vantagens e deve ser analisada com critério; a aplicação de conceitos como flexibilidade de esquema, escalabilidade, alta performance e disponibilidade, por outro lado, trazem desafios relacionados principalmente à manutenção das regras de consistência dos dados (De Souza & Dos Santos, 2015).

A Tabela 4.1 apresenta, de forma resumida, as principais vantagens e desvantagens dos quatro principais modelos de bancos de dados NoSQL existentes e que foram descritos nas subseções 4.2.1 a 4.2.4.

Tabela 4.1: Vantagens e desvantagens dos tipos de bases de dados NoSQL

Tipo de Repositório	Vantagens	Desvantagens
Chave-Valor	<ul style="list-style-type: none"> <li>• Simples e fácil de implementar;</li> <li>• Alta escalabilidade;</li> <li>• As pesquisas são rápidas.</li> </ul>	<ul style="list-style-type: none"> <li>• Não é possível consultar parte de um valor, o que restringe o critério de busca;</li> <li>• A linguagem de consulta é pobre e dificulta a pesquisa por relacionamentos.</li> </ul>
Documento	<ul style="list-style-type: none"> <li>• Permite inserção de registros incompletos;</li> <li>• Suportam consulta a todos os campos do documento;</li> <li>• Flexibilidade de esquema.</li> </ul>	<ul style="list-style-type: none"> <li>• A manutenção da relação entre as entidades é complexa.</li> </ul>
Família de Colunas	<ul style="list-style-type: none"> <li>• Fluxo de I/O otimizado;</li> <li>• Alta escalabilidade.</li> </ul>	<ul style="list-style-type: none"> <li>• Gestão entre relacionamento de entidades difícil;</li> <li>• As consultas que envolvem entidades múltiplas são complexas.</li> </ul>
Orientado a Grafos	<ul style="list-style-type: none"> <li>• Focado no relacionamento de entidades;</li> <li>• Processamento rápido de consultas em múltiplas entidades;</li> <li>• Esquema flexível;</li> <li>• Oferece suporte gráfico nativo.</li> </ul>	<ul style="list-style-type: none"> <li>• Escalabilidade é comprometida quando o grafo não pode ser armazenado em memória RAM;</li> <li>• Atualizações e consultas a uma mesma entidade apresentam custo maior.</li> </ul>

O estudo das necessidades informacionais da investigação policial e a possibilidade de adoção de um modelo baseado em *big data* como suporte a essas atividades possibilitam tecer as seguintes considerações em relação aos critérios vinculados ao relacionamento entre entidades, desempenhos das consultas e escalabilidade dos dados:

#### 4.2.1. Suporte ao relacionamento entre entidades

Os bancos de dados chave-valor não são projetados para trabalhar com relacionamento entre entidades; sua proposta de utilização é focada para modelos de dados simples.

Os repositórios do tipo família e colunas e orientados a documentos conseguem trabalhar com conceito de chaves estrangeiras e relacionamento entre tabelas, mas transfere a gestão dessas atividades para a aplicação do cliente.

O modelo orientado a grafos é o ideal para modelar relacionamentos complexos entre diversos tipos de entidade. A essência desses bancos de dados é a representação nativa de entidades (vértices) e seus relacionamentos (arestas).

#### 4.2.2. Desempenho na consulta das entidades e seus relacionamentos

Os bancos de dados chave-valor não suportam modelos de dados complexos e que necessitam utilizar consultas em múltiplas entidades. Como seus atributos são vistos como uma massa binária de dados, as consultas tornam-se extremamente limitadas: só é possível realizar operações sobre a chave do registro.

Os bancos de dados orientados a documentos e os repositórios do tipo família de colunas possuem a funcionalidade de permitir consultas com base nos atributos dos documentos e registros, respectivamente. Contudo, as pesquisas, quando necessitam a manipulação de diversos relacionamentos entre entidades, são avaliadas como complexas e muitas vezes inviáveis. O banco de dados orientado a documentos, por ser um modelo mais apropriado a relacionamentos aninhados, possui dificuldades no processamento de consultas entre entidades que se encontram num mesmo nível hierárquico. Já o repositório do tipo família de colunas prevê a desnormalização dos dados para seu melhor funcionamento, o que causa complexidade por parte das aplicações do usuário na execução de consultas complexas.

Os bancos de dados orientados a grafo conseguem processar com rapidez a relação entre entidades porque dispensam a execução de operações *joins*; é necessário tão somente percorrer os vértices e as arestas do grafo de acordo com o critério estabelecido pelo usuário. Esses repositórios apresentam outra vantagem, que é a execução de algoritmos específicos de modelagem de grafos – menor caminho entre duas entidades ou entidades que se interconectam por meio de outra entidade, dentre outros – o que traz dinamismo e tornam mais poderosas as consultas dos usuários.

### 4.2.3. Escalabilidade dos dados

Analisando-se a Figura 4.5 é possível depreender que, quanto maior a capacidade de flexibilização dos bancos de dados NoSQL, menor é seu poder de prover escalabilidade às suas bases de dados.

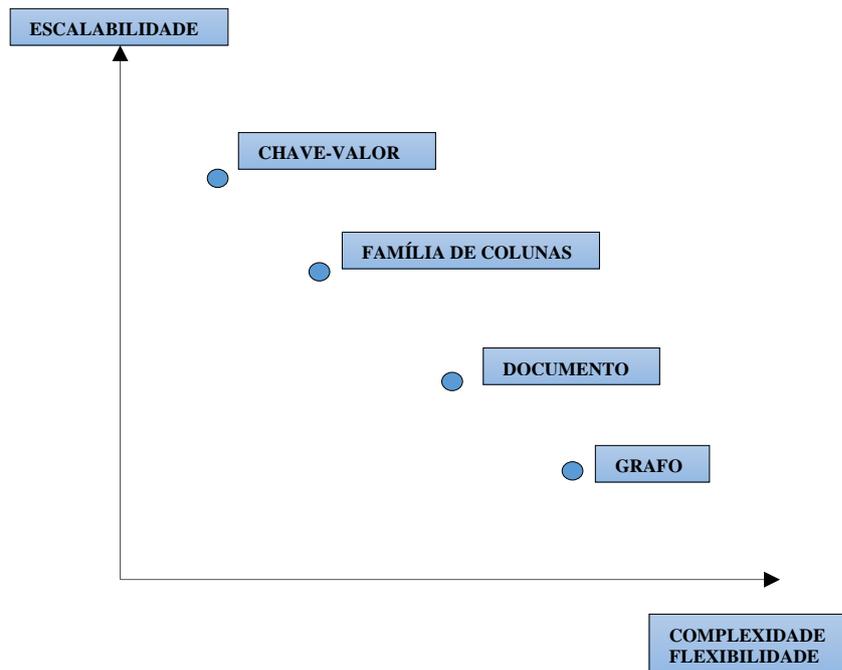


Figura 4.5: Escalabilidade x Complexidade e Flexibilidade dos bancos de dados NoSQL (adaptada de Schmutz (2013))

Os repositórios do tipo chave-valor, por ser o mais simples dos modelos apresentados, está mais preparado para lidar com o crescimento do volume de dados. Os bancos de dados do tipo família de colunas e orientado a documentos estão posicionados de forma intermediária se comparado com outros tipos de bases NoSQL.

O banco de dados orientado a grafos possui mais dificuldades em relação à sua escalabilidade, porém apresenta mais flexibilidade, como pode ser comprovado nas análises das vantagens e desvantagens dos bancos de dados NoSQL, apresentadas na Tabela 4.1. O processamento de grandes quantidades de vértices e arestas e o grau de crescimento do banco de dados pode ser potencializado, conforme visto, com o incremento de memória RAM

disponível no sistema. Kyrölä (2014) descreve detalhes sobre a performance dos grafos frente à disponibilidade de memória.

### **4.3. UTILIZAÇÃO DO BANCO DE DADOS DE GRAFO NA SOLUÇÃO**

Segundo (Angles & Gutierrez, 2008), a utilização da modelagem por meio de grafos apresenta vantagens nas seguintes aplicações, dentre outras: a) aplicações em que a complexidade das relações excede a capacidade de representação de outros modelos; b) soluções que exigem linguagem de consulta mais poderosa que as tradicionais linguagens SQLs; c) modelos que exigem representação mais elaborada e níveis de abstração mais adequados dos que os atualmente oferecidos por outros tipos de bancos de dados; d) aplicações que lidam com redes de informação, as quais, naturalmente, trabalham com grande inter-relação de dados.

No modelo *big data* que vem sendo proposto neste estudo, o banco de dados de grafo cumpriria a função de colocar a informação de forma amigável para a investigação, ao mesmo tempo em que realiza a organização e a estruturação dos dados, oferecendo uma análise gráfica e facilitada sobre itens que compõe um determinado caso, como investigados, pessoas vinculadas, informações patrimoniais, contas bancárias, aparelhos telefônicos, imagens e vídeos coletados por meio de atividades operacionais, extratos bancários, relatórios financeiros, interceptação telefônica, encontros, relações entre perfis de redes sociais, dentre outros.

Holzschuher & Peinl (2013) destacam em seus experimentos que, de maneira geral, o banco de dados de grafo pode ser utilizado como uma aplicação que oferecerá melhor performance em relação aos bancos de dados relacionais, especialmente quando se trabalhará com dados que estão interconectados, a exemplo de uma análise de vínculos em rede social.

Batra & Tyagi (2012) relataram que os bancos de dados de grafos são bem flexíveis, ao não exigir que se altere o esquema toda vez em que um relacionamento é adicionado ao modelo, ao contrário dos bancos de dados relacionais, cuja utilização se torna mais apropriada para modelos de dados mais estáticos, com poucos relacionamentos. O estudo afirma que esse tipo de repositório é mais rápido, pois lida diretamente apenas com os vértices específicos que estão diretamente à consulta, ao contrário do modelo relacional que deve percorrer toda a tabela de dados para encontrar aqueles os quais preenche um determinado critério.

Jaiswal & Agrawal (2013) corroborou tais informações, ao perceber em seus estudos que, à medida em que consultas que se utilizam de mais ligações entre as tabelas (*joins*) ou que manipulam múltiplas entidades relacionadas, há mais rapidez no processamento quando se utiliza banco de dados de grafos, em comparação com diversos outros modelos existentes.

Os insumos informacionais utilizados pelas polícias judiciárias detalhados neste trabalho são coletados por inúmeros aplicativos e fontes e estão presentes em sistemas de arquivos e bases de dados diversificados, geralmente de forma dispersa e, em muitas vezes, pouco estruturada. Tais informações estão presentes em sistemas operacionais, sistemas de arquivos e bancos de dados tradicionais, o que dificulta sua padronização e inter-relação.

Essa massa de dados deverá, então, ser submetida a outro recurso a fim de ser armazenada e recuperada, de forma a complementar as informações trabalhadas pelo banco de dados de grafo. Esse recurso é o HDFS, um sistema de arquivos distribuídos, que será detalhado na seção 4.4.

#### **4.4. UTILIZAÇÃO DO HDFS NA SOLUÇÃO**

A seção 2.4 apresentou os conceitos básicos do sistema de arquivos HDFS e do Projeto Hadoop. A utilização dessas ferramentas abre suporte a aplicações do tipo MapReduce<sup>18</sup>, que é um modelo de programação específico para a realização de tarefas em processamento paralelo, ao aderir seu mecanismo de funcionamento às características primordiais do Hadoop relacionadas à distribuição dos *jobs* e tolerância a falhas de hardware e software. MapReduce decompõe o trabalho em diversas tarefas menores que podem ser paralelizadas e executadas de forma independente umas das outras para, posteriormente, serem consolidadas e finalizadas.

Raste (2014) realizou simulações entre uma plataforma Hadoop, baseada em HDFS, e um banco de dados relacional e concluiu que o primeiro leva vantagem em aspectos como o carregamento de dados e o processamento de consultas. Nos experimentos, quanto maior o tamanho dos blocos de dados a ser processado, mais vantajosa é a utilização da plataforma Hadoop.

---

<sup>18</sup> É um modelo de programação paralela cujo funcionamento será apresentado e detalhado na subseção 4.6.4

Outra vantagem na utilização do HDFS, destacada por Sremack (2015), é a possibilidade que outros sistemas de bancos de dados operem sobre ele, permitindo o funcionamento de pacotes e mecanismos de *datawarehouse*, além de outros procedimentos especializados que viabilizam a análise de grandes volumes de dados. Ou seja, o HDFS abre possibilidades de desenvolvimento e posterior agregação de aplicações e funcionalidades na plataforma instalada, o que pode ser útil para pacotes ou aplicações específicas adquiridos posteriormente. A seção 4.5 detalha como o banco de dados de grafo e o sistema de arquivos HDFS trabalham na prática numa plataforma *big data*.

#### **4.5. A PLATAFORMA PROPOSTA**

Após a definição de qual a solução NoSQL é mais adequada ao modelo, bem como da necessidade de utilização de técnicas de processamento paralelo para lidar com volumes consideráveis de informação (grande parte não-estruturada), esta seção tratará de organizar como os componentes descritos ao longo deste estudo se organizam e interagem, no momento da chegada da informação no órgão policial até a apresentação de resultados ao analista.

A Figura 4.6 sintetiza a plataforma de dados que supre as necessidades do aparato investigativo, destacando o fluxo de dados composto pelos diversos tipos de insumos informacionais já examinados. A plataforma é composta por quatro componentes principais: insumos informacionais, banco de dados orientado a grafos, *cluster* Hadoop e resultados gerados ao analista.



Figura 4.6: Plataforma de big data para investigação policial

Como refletido na plataforma, a investigação deve permear as seguintes etapas, a saber: a) relação dos tipos de dados a trabalhar, por meio da definição dos insumos informacionais; b) utilização do banco de dados de grafo objetivando a formação dos metadados da investigação; c) inserção dos dados não catalogáveis no *cluster* Hadoop; e d) aplicação de operação em MapReduce. As subseções 4.5.1 a 4.5.4 detalham essas etapas.

#### 4.5.1. Definição dos Insumos Informacionais

Conforme listado na seção 3.2 e destacado na Figura 4.6, os insumos informacionais possíveis de serem utilizados na investigação policial são dados oriundos das seguintes fontes: a) processo tradicional de investigação; b) sistemas de informações transacionais; c) laudos periciais e relatórios de identificação; d) tabelas ou planilhas oriundas de medidas cautelares; e) interceptação telefônica; f) vigilância a alvos e g) fontes abertas.

Um passo importante que deve ser tomado pelos gestores do trabalho investigativo é definir quais desses insumos são relevantes para uma determinada apuração. Nem todos os casos policiais vão, necessariamente, trabalhar com todos os tipos de dados possíveis de serem coletados e analisados. A investigação de um homicídio difere de trabalhos que visam detectar e neutralizar uma quadrilha que pratica crimes por meio da internet; uma quadrilha que atua realizando o desmatamento da floresta e a venda ilegal de madeira possui um *modus operandi* completamente diferente de um grupo que se associa para corromper políticos. A forma da atuação, os tipos de vestígios possíveis que podem ser produzidos pelos criminosos, a conveniência e facilidade de obtenção da informação, dentre outros, são fatores que vão direcionar a coleta e utilização de fontes de informações.

Dessa maneira, a equipe de investigação deverá traçar estratégias e procurar definir o escopo dos dados iniciais a partir do qual o trabalho se desenvolverá. Nada impede, obviamente, que sejam incluídos novos tipos de informações à medida que a própria investigação vá avançando e novos suspeitos sejam revelados. Cabe ressaltar que o banco de dados proposto na plataforma deve possuir liberdade e flexibilidade de esquema, o que é positivo quando pode ocorrer acréscimo ou mudança de metadados ao longo dos trabalhos.

É a partir dos insumos informacionais que vai se construindo o catálogo de dados da investigação, o qual é aplicado sobre o banco de dados de grafo.

#### 4.5.2. Inserção dos Dados Catalogáveis no Banco de Dados de Grafos

O próximo passo é inserir essas informações em um banco de dados de grafo por meio de uma interface de usuário ou de um mecanismo automatizado, a depender do formato dessa entrada. Esses repositórios priorizam os relacionamentos das entidades às quais pretende-se trabalhar (investigar). Constituir uma base de dados contendo entidades, seus atributos e relacionamentos compõe a etapa da entrada da informação relacionada à investigação no banco de dados de grafos.

A utilização de banco de dados de grafo como componente da arquitetura que irá armazenar o catálogo (metadados) das entidades não se relaciona apenas com a possibilidade de proporcionar uma interface visual mais amigável para os investigadores ou analistas, mas, conforme visto na seção 4.3, está diretamente ligada à performance da aplicação quando se deseja trabalhar com inúmeras relações entre entidades. Alguns bancos de dados de grafo são capazes de armazenar mais de 30 bilhões de vértices e arestas<sup>19</sup>, permitindo a representação de um universo de entidades e relacionamentos bastante vasto.

Além das tradicionais operações matemáticas, lógicas, de ordenação e de comparação existentes em praticamente todos os tipos de sistemas de banco de dados, operações específicas podem ser aplicadas sobre um banco de dados orientados a grafo, conforme exemplos listados e a seguir, extraídos da documentação do banco de dados Neo4J<sup>20</sup>.

1. Seleção de vértices que satisfazem determinada condição. Esta operação pode ser útil numa investigação de crimes financeiros na qual deseja-se aprofundar os trabalhos em contas bancárias pertencentes a indivíduos de determinada família. No grafo, serão selecionados os vértices relacionados a contas correntes e indivíduos de uma mesma família.
2. Seleção de relacionamentos que satisfazem determinada função ou critério. Esta operação pode, dentre outras funcionalidades, oferecer um filtro entre chamadas telefônicas (relacionamentos) que ocorreram em determinado horário, partindo de

---

<sup>19</sup> Informações disponíveis em <http://neo4j.com/docs/stable/capabilities-capacity.html> e <http://franz.com/agraph/allegrograph/>, documentação dos bancos de dados Neo4J e AllegroDB, respectivamente.

<sup>20</sup> Idem

um aparelho telefônico pré-determinado (vértice). Seria útil no aprofundamento da resolução de um homicídio em que foi encontrado o *chip* de dados de um aparelho telefônico no local do crime.

3. Seleção de vértices que estabelecem um padrão de relacionamento. Seria possível selecionar indivíduos (vértices) que se comunicam diversas vezes por dia por meio de mensagens de e-mail, chamadas telefônicas ou redes sociais (relacionamentos no grafo) e provar que há uma relação entre eles, mesmo que tal fato seja negado por meio de um depoimento ou declarações posteriores.
4. Encontrar a quantidade de relacionamento entre dois vértices. Esta operação é útil, por exemplo, quando deseja-se medir o grau de intimidade entre dois investigados.
5. Encontrar o caminho com menor/maior custo entre dois vértices (algoritmo de Dijkstra). Esta operação poderia revelar o caminho de valores oriundos de propina que foram depositados para um investigado, mas que, no intuito de desviar o foco das investigações, ocorre a pulverização do depósito para diversas outras contas correntes.

Contudo, conforme já explicado na seção 3.5, os insumos informacionais estão acompanhados de outras informações extensas e não estruturadas. Uma interceptação telefônica contém a parte da informação estruturada (interlocutores, horário, duração, números telefônicos utilizados na chamada) que fazem parte do catálogo de dados e são inseridas no banco de dados; porém há também dados não estruturados, como o próprio arquivo de áudio. Outro exemplo é a interceptação telemática, cuja parte estruturada está relacionada aos dados de um investigado, seu endereço e informações do *link* de conexão à internet, os quais são acompanhados de grandes quantidades de informação não estruturadas (os pacotes interceptados tipo PCAP).

Em paralelo à sistematização no banco de dados de grafo, toda a informação anexa deve ser apartada no sistema de arquivo nativo. São *terabytes* e *petabytes* de informações que vão sendo armazenadas ao longo do tempo, compostos de documentos diversos, laudos periciais, relatórios de atividades operacionais, vídeos, áudios, informações extraídas de bases de dados internas e externas, fluxos interceptados de dados, e demais tipos de insumos informacionais que não podem ser tratados integralmente pela base de dados de grafo. O próximo passo é sua migração para um *cluster* Hadoop e a aplicação de rotinas de processamento, conforme explanado na subseção 4.5.3.

#### 4.5.3. Inserção de Dados Semiestruturados no *Cluster* Hadoop

De plano, cabe ressaltar que a utilização de bases de dados de grafo e HDFS como solução a questões voltadas à solução de gargalos de análise de informações ainda é incipiente. Gradoop é uma iniciativa ainda em estágios iniciais que propõe uma solução de análise de dados baseada em grafos, cujo suporte é Hadoop. O referido projeto ainda está em fase inicial de desenvolvimento, havendo necessidades de várias implementações de código e algumas camadas de sua arquitetura<sup>21</sup>.

Não basta apenas melhorar o grau de análise dos relacionamentos entre as entidades às quais pretende-se desenvolver uma investigação. Essa parte, considerada essencial, é suprida pela utilização de uma base de dados que possa catalogar esses dados, resolver consultas complexas sobre os relacionamentos existentes e oferecer ferramentas de visualização mais amigáveis.

Complementarmente, para que a plataforma de análise seja completa e não ficasse restrita aos dados catalogáveis, outro mecanismo de suma importância deve ser utilizado. A massa de dados semiestruturada pode conter informações valiosíssimas. Para ela, o HDFS oferecerá suporte para inúmeras funcionalidades relacionadas à pesquisa de seu conteúdo, por meio de aplicações do tipo MapReduce, que será abordado na subseção 4.5.4.

#### 4.5.4. Realização de Operações MapReduce

Segundo Lam (2010), o MapReduce é um modelo de processamento que possui grandes vantagens em termos de escalabilidade e suporte a inúmeros nós de processamento. Sobre tal modelo, seu cerne se compõe basicamente de dois tipos de tarefa, mapeadora (*Map*) e redutora (*Reduce*), que podem ser aplicadas sobre centenas de máquinas em um *cluster* de dados.

Silva (2011) destaca as importantes características técnicas do MapReduce: a função mapeadora (*Map*) objetiva distribuir o processamento de uma aplicação, tendo como entrada uma par de chave/valor e entregando como resultado uma função aplicada sobre essa entrada; a função redutora (*Reduce*) é responsável pelo refinamento do processamento realizado pela

---

<sup>21</sup> Para maiores detalhes, consultar [gradoop.com](http://gradoop.com)

aplicação da função mapeadora, consolidando seus resultados. O grande diferencial dessas funções é que elas podem ser executadas de forma totalmente independente, permitindo que o problema computacional seja paralelizado de forma transparente.

Como exemplo, a fim de verificar a incidência de palavras-chave numa determinada amostra de informações – como nomes de indivíduos ou localidade – num fluxo de *terabytes* de dados telemáticos interceptado na empresa de um suspeito, pode-se utilizar o processamento paralelo em MapReduce.

O fluxo interceptado será alocado nos blocos do HDFS divididos usualmente em pedaços de 64MB. A partir daí, a aplicação dispara várias tarefas *Maps*, que vão varrer a massa de dados e procurar *strings* de texto com as determinadas palavras-chave. A saída da tarefa *Map* é a quantidade de cada palavra-chave inspecionada na porção de dados.

Em seguida, cada tarefa *Reduce* do ambiente será responsável por somar as ocorrências encontradas pelas funções *Maps* e apresentar o resultado, que é uma lista com cada palavra-chave e sua ocorrência no fluxo telemático.

A Figura 4.7 mostra um exemplo do processamento, por meio do MapReduce, de busca das palavras-chaves “Fulano” e “Brasília” numa massa de dados.

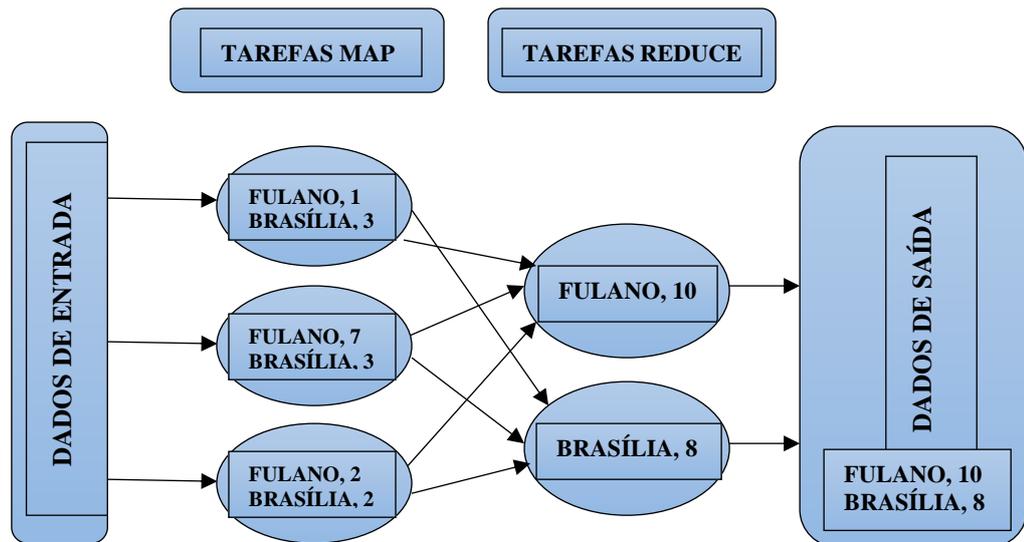


Figura 4.7: Exemplo de processamento MapReduce

O ambiente Hadoop divide os dados a serem inspecionados em 3 partes. Cada parte será submetida a uma tarefa *Map*, que em paralelo, fornecerá o resultado encontrado do número de ocorrências do termo “Fulano” e “Brasília”.

Em seguida, cada tarefa *Reduce* irá fazer o somatório de um termo, compilando os resultados provenientes da saída de cada tarefa *Map*. Finalmente, o resultado é apresentado: uma lista contendo cada termo (Fulano e Brasília) e, respectivamente, a quantidade de ocorrências em todo o fluxo de entrada, 10 e 8.

Assim, a parte da informação que não é amigavelmente inserida na base de dados de grafo será trabalhada pelo Hadoop, o qual proverá a capacidade de refinar *terabytes* de dados brutos proveniente das fontes informacionais menos estruturadas, nas quais ainda reside a grande dificuldade analítica.

## **5. APLICAÇÃO DA PLATAFORMA DE DADOS NUM EXEMPLO DE INVESTIGAÇÃO**

O capítulo 4 propõe a plataforma de dados *big data* voltada à investigação, definindo que ela deve estar apta a receber insumos informacionais os mais diversificados possíveis, a fim de permitir a análise e processamento de informações e revelar os vínculos de pessoas e ações perpetradas, tanto entre dados catalogáveis e estruturados quanto para arquivos extensos e sem estrutura formal.

Primeiramente, cabe esclarecer a dificuldade em se utilizar dados pertencentes a operações policiais reais, uma vez que os materiais relacionados a elas estão sob a guarda da justiça ou do próprio órgão policial, além de possuir, quase que em sua totalidade, conteúdo sigiloso relacionado a conversas ou registros íntimos entre pessoas. Expor casos reais pode ensejar infrações administrativas ou penais.

Dessa forma, objetivando a aplicação da plataforma, optou-se por utilizar um caso hipotético, composto por dados aleatórios que possibilitam descrever as seguintes etapas preconizadas no modelo proposto: a) definição dos insumos informacionais; b) inserção dos dados catalogáveis no banco de dados de grafo; c) inserção dos dados não estruturados no sistema de arquivos distribuídos e d) execução de operações MapReduce. As seções 5.1 a 5.4 descrevem pormenorizadamente cada uma dessas etapas.

### **5.1. DEFINIÇÃO DO INSUMO INFORMACIONAL**

Geralmente, a investigação policial é um processo incremental de junção de informações que aportam no órgão policial sob diversos canais e diferentes formas. A equipe de investigação é responsável, dentre outras tarefas, pela realização do planejamento das ações de captação desse conhecimento, à medida que os trabalhos vão se desenrolando e as condições para aquisição dos dados se tornem satisfatórias.

Alguns desses insumos, como a interceptação telemática ou telefônica, requerem a autorização judicial pertinente para seu emprego legal. Outros insumos dependem da aquisição de bens ou serviços por parte da polícia, como instalação de câmeras de vigilância, aquisição de equipamentos eletrônicos, investimentos em infraestrutura, compra de viaturas, dentre outros. O fato é que as informações nunca chegam de forma sincronizada e padronizada, pois

existem uma série de ações burocráticas e técnicas que geralmente se colocam como pré-requisitos para sua utilização.

O exemplo de uma investigação pode ilustrar melhor essa situação. Uma denúncia anônima chega ao órgão policial e descreve a atividade de uma quadrilha que está derrubando ilegalmente árvores numa determinada região, cuja preservação ambiental está prevista em lei. É comum que, num primeiro momento, uma equipe de investigadores seja designada para iniciar a apuração dos fatos supostamente criminosos e, para isso, policiais deslocam-se até a região para verificar a movimentação de máquinas e pessoas na área objeto da denúncia.

Ao encontrar esses indícios que a prática de crime está ocorrendo de fato, novas medidas podem ser tomadas para agregar dados aos trabalhos iniciais, como a solicitação de quebra do sigilo telefônico dos suspeitos, que objetivará alcançar toda a cadeia de produção, processamento e comercialização da madeira ilegal, desde os operários que trabalham no local até os comerciantes das toras, os intermediários, os proprietários de serralherias clandestinas e possíveis lojas de mobiliário que utilizam como matéria-prima a madeira ilegal.

Assim, novos horizontes podem ser explorados e, por consequência, outras ações também são colocadas em prática, como o afastamento do sigilo telemático dos envolvidos, cujo objetivo seria conhecer as mensagens eletrônicas (*e-mails*) que membros da quadrilha supostamente trocam entre si, descrevendo ações, valores, transações financeiras e inúmeros documentos tratados sobre suas atividades clandestinas.

Noutro momento, a investigação poderia alcançar o lado patrimonial dos criminosos e analisar documentos relacionados à quebra do sigilo bancário, desvendando a movimentação de valores entre pessoas e empresas do grupo, apontando os possíveis laranjas e confirmando o ganho financeiro ilegal relacionado à atividade ilícita.

A Figura 5.1 mostra a dinâmica da incorporação de fontes de informação à medida que a investigação avança:

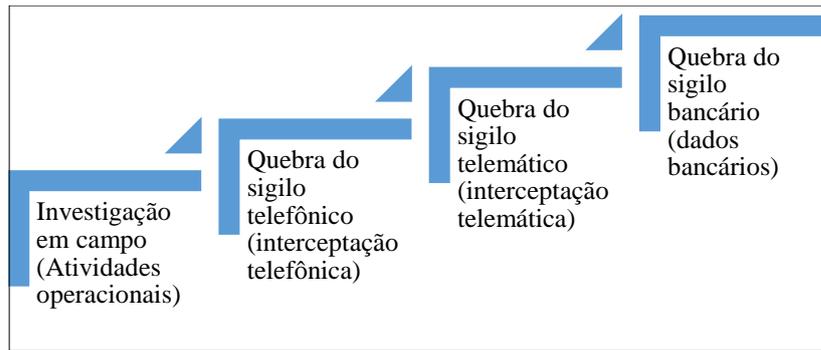


Figura 5.1: Processo de incorporação de fontes de informação numa investigação

Definir o que será utilizado na investigação também requer uma separação prévia dos tipos de dados que comporão o trabalho de apuração do caso. É preciso separar os insumos catalogáveis - que serão utilizados e inseridos no banco de dados de grafo - daqueles que se encontram em formato não estruturado - os quais deverão ser encaminhados posteriormente para o *cluster* Hadoop, para fins de pesquisa e processamento. Exemplo desses últimos, como visto no decorrer deste trabalho, são transcrições de vídeos e de áudios, conteúdos de pacotes de dados interceptados, documentos diversos não formatados, conteúdos da *web*, dentre outros.

As subseções 5.1.1 a 5.1.4 detalham os quatro tipos de insumos informacionais utilizados no exemplo aqui proposto: informações de atividades operacionais, interceptação telefônica, interceptação telemática e dados bancários.

#### 5.1.1. Atividades operacionais

Ao receber a denúncia e deslocar-se até a região onde supostamente está ocorrendo os delitos, investigadores devem registrar ações que podem corroborar com as informações fornecidas pelo denunciante. As atividades operacionais, portanto, são resultado de trabalho de campo, que geralmente descrevem encontros ou situações relevantes. Ao final da diligência, o policial deverá formalizar o acontecimento por meio de um relatório descritivo do que foi presenciado pela equipe.

O registro de um encontro, realizado por meio de uma atividade operacional de vigilância a investigados, é composto pelas pessoas envolvidas e detalhes sobre o acontecimento, tais como data, hora e local. Um arquivo de vídeo ou fotografia e um relatório descritivo podem estar atrelados ao registro principal.

Tabela 5.1: Registro de Atividade Operacional

<b>Atividade Operacional</b>			
Pessoa	Alvo5		
Pessoa	Alvo8		
Data Encontro	23/08/2015	Hora Encontro	08:15
Local Encontro	Altamira/PA		
Descrição: Alvo5 e Alvo8 são empresários de Brasília conhecidos do ramo de movelaria e são vistos jantando na região onde costumeiramente ocorrem ações de desmatamento ilegal, em Altamira/PA. Após o encontro, eles seguiram, em veículos separados, para seus respectivos hotéis.			
Caminho do arquivo de vídeo ou fotografia		"C:\operação1\filmagem1.mp4"	
Caminho do arquivo com a descrição do relatório		"C:\operação1\filmagem1.txt"	

O registro contido na Tabela 5.1 detalha o encontro de duas pessoas, Alvo5 e Alvo8, às 08:15h, em 23/08/2015 num restaurante de Altamira/PA. O registro do vídeo foi realizado e gravado no na pasta local da operação, no caminho "C:\operação1\filmagem1.mp4". A descrição do relatório encontra-se em arquivo, no caminho "C:\operação1\filmagem1.txt".

#### 5.1.2. Interceptação telefônica

Ao confirmar a existência de empresários de movelaria no local suspeito, a equipe de gestão da operação policial poderá decidir pela expansão da investigação e, diante dos indícios coletados, solicitar ao juiz a quebra do sigilo telefônico dos aparelhos de Alvo5 e Alvo8, para aprofundamento das investigações.

A interceptação telefônica utiliza-se de dados estruturados do extrato telefônico bem como dados em formato de áudio, que foram captados pela operadora e replicados para a unidade de investigação policial. O registro de interceptação telefônico é composto pelos interlocutores, seus telefones e detalhes sobre a chamada telefônica, como data, hora, duração e resumo da conversa. Um arquivo de áudio que é a própria conversa gravada estará atrelado a esse registro, juntamente com a transcrição do que foi tratado nela.

Tabela 5.2: Registro de Interceptação Telefônica

Interceptação Telefônica					
Pessoa origem	Alvo5			Telefone	(61)9615-0261
Pessoa destino	Alvo7			Telefone	(61)9636-7827
Data	25/08/2015	Hora	11:01	Duração	97
Resumo da conversa: Alvo5 solicita um carregamento de madeira que deverá ser entregue à empresa MM. Alvo7 informa que o valor do carregamento é R\$ 30 mil. Alvo7 diz que informará a Alvo8 seus dados bancários.					
Caminho da transcrição da conversa telefônica		"C:\operação1\chamada1.txt"			

A Tabela 5.2 detalha um registro de chamada telefônica realizada por Alvo5 a Alvo7, em 25/08/2015, às 11:01 h, com duração de 1 minuto e 37 segundos (97 segundos), ocasião em que foram realizadas tratativas acerca de um carregamento de madeira que deve ser entregue à empresa MM. O conteúdo transcrito da conversa foi armazenado no sistema de arquivos, na pasta "C:\operação1\chamada1.txt".

Cabe ressaltar que geralmente a inserção de dados relacionados à interceptação telefônica é incluída de forma automatizada na plataforma, uma vez que as operações policiais costumam lidar com milhares de registros dessa natureza, os quais geram inúmeros extratos telefônicos e gravações. Como exemplo, registra-se a existência de 45.594 páginas de extratos telefônicos e 4.468 horas de áudios coletados pelos investigadores em decorrência de grandes investigações - Operações Monte Carlo e Vegas<sup>22</sup> -, deflagradas pela Polícia Federal no ano de 2012.

### 5.1.3. Interceptação telemática

A interceptação telemática trabalha com dados semiestruturados que foram captados pelo provedor de internet do investigado e replicados para a unidade de investigação policial.

---

<sup>22</sup> Os dados constam no Relatório Final da CPMI Vegas e Monte Carlo, do Congresso Nacional, disponível em <http://www.senado.leg.br/atividade/rotinas/materia/getPDF.asp?t=118546&tp=1>

Geralmente, essa massa de dados chega aos investigadores por meio de arquivos com extensão PCAP (Packet Capture Data) ou similar.

O registro de interceptação telemática pode revelar inúmeros tipos de dados. Conforme descrito por Peron (2012), é possível entender o conteúdo dos arquivos por meio de aplicativos que interpretam os protocolos e remontam as informações. O autor demonstra que uma informação possível de ser compreendida é o tráfego de mensagens eletrônicas (*e-mails*) entre dois indivíduos, por meio da filtragem de conteúdo de protocolos POP e SMTP. É possível recuperar o endereço de origem, endereço de destino, assunto e conteúdo da mensagem. O *e-mail* remontado e seus anexos ficam atrelados ao registro.

Tabela 5.3: Registro de Interceptação Telefônica

Interceptação Telemática			
Endereço origem	alvo8@yyy.com.br		
Endereço destino	alvo7@yyy.com.br		
Data	27/08/2015	Hora	04:15
Assunto	Comprovante de Depósito		
Caminho do arquivo de e-mail e anexos	C:\operação1\comprovante.pdf		

A Tabela 5.3 detalha o envio de mensagem eletrônica oriunda da conta de e-mail alvo8@yyy.com.br à conta alvo7@yyy.com.br, cujo assunto é “Comprovante de depósito”, em 27/08/2015, às 04:15h. Em anexo, um arquivo PDF cujo nome é “comprovante.pdf” acompanha o conteúdo. O documento anexado foi armazenado no sistema de arquivos, na pasta “C:\operação1\comprovante.pdf”.

Ressalta-se que os dados brutos, ou seja, os pacotes interceptados pela investigação e que não são catalogados podem seguir, posteriormente, ao *cluster* Hadoop para processamento.

#### 5.1.4. Dados bancários

A fim de comprovar a transação bancária declarada nas conversas telefônicas, a equipe policial pode solicitar ao magistrado a quebra do sigilo bancário dos envolvidos, o que possibilita enriquecer o material probatório da investigação e descobrir o fluxo financeiro do esquema delituoso, inclusive com a possibilidade da descoberta de novos envolvidos no caso.

Os extratos bancários são recebidos por meio de sistema próprio, o SIMBA, e poderão ser tratados e inseridos na plataforma de forma automatizada. Existem casos que lidam com centenas de contas bancárias, o que representa um número bem maior de transações<sup>23</sup>. As principais informações são apresentadas na tabela a seguir:

Tabela 5.4: Registro de transação financeira

Transação Financeira					
Pessoa Origem	Alvo8				
Núm. Banco	Banco1	Nº Agência	7850	Nº. Conta	609444-4
Pessoa Destino	Alvo7				
Núm. Banco	Banco1	Nº Agência	3008	Nº. Conta	887965-3
Valor	R\$ 900.000,00	Data	26/08/2015		

A Tabela 5.4 detalha informações de correntistas, suas respectivas contas correntes, bom como valor e data de uma determinada transação. Nesse exemplo, em 26/08/2015, Alvo8, titular da conta corrente 609444-4, Banco1, depositou R\$ 900.000, 00 para Alvo7, titular da conta corrente 887965-3, Banco3.

## 5.2. INSERÇÃO DOS DADOS CATALOGÁVEIS NO BANCO DE DADOS ORIENTADO A GRAFOS

A fim de dar mais robustez ao exemplo proposto neste estudo, a massa de dados será complementada com mais informações relativas a investigados (alvos), linhas telefônicas e contas de *e-mail* interceptadas, contas bancárias, conversas telefônicas e transações financeiras.

Faz-se necessário o carregamento dos dados da investigação em nível macro num banco de dados orientado a grafos, com vistas a superar as limitações oferecidas por arquiteturas relacionais em relação ao processamento e melhor exploração do relacionamento entre entidades, além de possibilitar a interação com interfaces gráficas e visuais.

As seguintes informações aleatórias foram inseridas e os respectivos procedimentos encontram-se detalhados nas seções 5.2.1 a 5.2.8.

---

<sup>23</sup> Como exemplo, a perícia do Caso Banestado utilizou-se de 455 contas correntes em suas análises. Disponível em <http://dvl.ccn.ufsc.br/congresso/anais/5CCF/20140425134553.pdf>.

1. 15 investigados;
2. 20 linhas telefônicas;
3. 15 endereços de *e-mail*;
4. 27 contas bancárias;
5. 18 encontros registrados entre investigados;
6. 250 chamadas telefônicas interceptadas;
7. 500 transações financeiras;
8. 13 mensagens eletrônicas interceptadas;

Os dados foram inseridos no Neo4j, banco de dados que possui casos de utilização em diversos seguimentos, como gerenciamento de rede, softwares analíticos, pesquisa científica, roteamento, gestão organizacional e de projeto, sistemas de recomendações, análises de redes sociais, dentre outros (Neo4j, 2016).

A escolha do Neo4j como banco de dados capaz de implementar este trabalho também está atrelada ao fato de ele ser o repositório orientado a grafos mais popular em meios acadêmicos e empresariais, se comparado a outros concorrentes da mesma categoria, segundo critérios como citação, pesquisa e utilização, conforme classificação elaborada pelo *site* especializado *DB-engines*<sup>24</sup>.

O Neo4j implementa a propriedade do modelo grafo, oferecendo características ACID e suporte a *cluster*. Esse repositório trabalha com armazenamento compacto e cache de memória para grafos, o que resulta em ganho de escala e eficiência no processamento de bilhões de nós em uma base de dados.

O Neo4j utiliza Cypher, linguagem declarativa específica para realizar operações em grafo. Por meio dela, é possível caminhar ao longo de todos os elementos do grafo (vértices e arestas), sem que o tamanho do grafo e dos dados impacte no desempenho do processamento de consultas e atualizações no grafo (Omand, Bartlett & Miller, 2012).

---

<sup>24</sup> A classificação de popularidade dos bancos de dados orientados a grafo, bem como a metodologia utilizada, está disponível em <http://db-engines.com/en/ranking/graph+dbms>

As subseções 5.2.1 a 5.2.8 descrevem o mecanismo de inserção dos dados catalogáveis pela operação policial, inclusive seus comandos em linguagem Cypher.

### 5.2.1. Inserção dos investigados

A Tabela 5.5 detalha informações de 15 indivíduos investigados na operação exemplo.

Tabela 5.5: Informações de Investigados

Nome	Data de Nascimento	CPF
Alvo1	01/06/1956	135.626.530-41
Alvo2	13/09/1941	010.041.791-76
Alvo3	02/08/1966	645.103.203-05
Alvo4	22/04/1977	433.042.041-59
Alvo5	28/03/1949	399.048.978-97
Alvo6	26/08/1946	551.495.346-39
Alvo7	11/11/1980	564.113.699-74
Alvo8	01/03/1977	840.000.505-03
Alvo9	08/09/1952	724.259.346-32
Alvo10	31/03/1955	510.392.202-55
Alvo11	12/12/1971	683.629.050-56
Alvo12	11/03/1970	516.500.670-86
Alvo13	30/11/1949	247.534.749-92
Alvo14	01/06/1955	180.173.396-72
Alvo15	03/08/1991	831.631.712-38

A Figura 5.2 indica os comandos de inserção de dados no Neo4j referentes a 3 investigados.

```
CREATE (p1:Pessoa { nome : 'Alvo1', dataNascimento : '01/06/1956' , CPF : '135.626.530-41'})
CREATE (p2:Pessoa { nome : 'Alvo2', dataNascimento : '13/09/1941' , CPF : '010.041.791-76'})
CREATE (p3:Pessoa { nome : 'Alvo3', dataNascimento : '02/08/1966' , CPF : '645.103.203-05'})
```

Figura 5.2: Inserção de investigados

### 5.2.2. Inserção dos aparelhos telefônicos e sua vinculação com os investigados

A Tabela 5.6 detalha informações de 20 linhas telefônicas interceptadas na operação exemplo.

Tabela 5.6: Informações de linhas telefônicas

Pessoa	Número	Operadora
--------	--------	-----------

Alvo1	(61)9934-1023	Op1
Alvo1	(61)9619-9651	Op1
Alvo2	(61)8132-1530	Op2
Alvo2	(61)9799-8985	Op1
Alvo3	(51)9658-4361	Op3
Alvo4	(61)9495-4487	Op3
Alvo5	(61)9615-0261	Op3
Alvo5	(61)9775-6938	Op4
Alvo6	(31)8467-9923	Op2
Alvo7	(61)9636-7827	Op1
Alvo7	(61)8977-2201	Op3
Alvo8	(61)9459-5966	Op3
Alvo9	(31)8696-5241	Op3
Alvo10	(61)9539-7580	Op4
Alvo11	(61)9299-5160	Op4
Alvo12	(61)8506-1381	Op4
Alvo12	(61)8379-1871	Op5
Alvo13	(61)9925-4888	Op1
Alvo14	(61)9970-4491	Op3
Alvo15	(91)9565-1231	Op5

A Figura 5.3 indica os comandos de inserção de dados no Neo4j referentes a 3 linhas telefônicas e a vinculação com seus respectivos alvos.

```
CREATE (t1:Telefone { numero : '(61)9934-1023', operadora : 'Op1' })
CREATE (t2:Telefone { numero : '(61)8132-1530', operadora : 'Op1' })
CREATE (t3:Telefone { numero : '(51)9658-4361', operadora : 'Op2' })
CREATE (p1)-[rpt1:POSSUI_TEL]->(t1)
CREATE (p2)-[rpt2:POSSUI_TEL]->(t2)
CREATE (p3)-[rpt3:POSSUI_TEL]->(t3)
```

Figura 5.3: Inserção de linhas telefônicas

### 5.2.3. Inserção das contas de *e-mail* e sua vinculação com os investigados

A Tabela 5.7 detalha informações de 15 contas de *e-mail* interceptadas na operação exemplo.

Tabela 5.7: Informações de contas de *e-mail*

Pessoa	Conta
Alvo1	alvo1@xxx.com.br
Alvo2	alvo2@yyy.com.br
Alvo3	alvo3@zzz.com.br
Alvo4	alvo4@xxx.com.br

Alvo5	alvo5@xxx.com.br
Alvo6	alvo6@xxx.com.br
Alvo7	alvo7@yyy.com.br
Alvo8	alvo8@yyy.com.br
Alvo9	alvo9@zzz.com.br
Alvo10	alvo10@zzz.com.br
Alvo11	alvo11@xxx.com.br
Alvo12	alvo12@yyy.com.br
Alvo13	alvo13@yyy.com.br
Alvo14	alvo14@xxx.com.br
Alvo15	alvo15@xxx.com.br

A Figura 5.4 indica os comandos de inserção de dados no Neo4j referentes a 3 contas de *e-mail* e vinculação com seus respectivos alvos.

```
CREATE (e1:Email { conta : 'alvo1@xxx.com.br' })
CREATE (e2:Email { conta : 'alvo2@yyy.com.br' })
CREATE (e3:Email { conta : 'alvo3@zzz.com.br' })
CREATE (p1)-[rpm1:POSSUI_MAIL]->(e1)
CREATE (p2)-[rpm2:POSSUI_MAIL]->(e2)
CREATE (p3)-[rpm3:POSSUI_MAIL]->(e3)
```

Figura 5.4: Inserção de contas de e-mail

#### 5.2.4. Inserção das contas bancárias

A Tabela 5.8 detalha informações de 27 contas correntes, cujo sigilo bancário foi quebrado na operação exemplo.

Tabela 5.8: Informações de contas bancária

Pessoa	Banco	Agência	Número
Alvo1	Banco1	2912	514503-0
Alvo1	Banco1	2355	527895-5
Alvo1	Banco2	7556	69873-3
Alvo2	Banco2	9954	591565-3
Alvo2	Banco1	0644	591807-6
Alvo2	Banco1	3506	898329-3
Alvo3	Banco1	3455	205086-5
Alvo3	Banco4	0405	98289-6
Alvo4	Banco1	5889	317870-7
Alvo4	Banco4	1332	232673-4

Alvo5	Banco3	3146	339463-9
Alvo5	Banco1	5383	46826-3
Alvo6	Banco5	6515	134624-8
Alvo6	Banco3	9648	582276-2
Alvo7	Banco1	3008	887965-3
Alvo7	Banco3	2241	442126-2
Alvo8	Banco1	7850	609444-4
Alvo8	Banco1	1910	567186-5
Alvo9	Banco1	0391	90088-5
Alvo9	Banco1	7496	461039-2
Alvo10	Banco2	8493	423923-9
Alvo10	Banco1	8525	348247-1
Alvo11	Banco2	5490	943716-7
Alvo12	Banco2	7229	956010-7
Alvo13	Banco2	2231	104918-2
Alvo14	Banco1	1272	921758-2
Alvo15	Banco1	9940	414401-2

A Figura 5.5 indica os comandos de inserção de dados no Neo4j referentes a 3 contas bancárias e seus respectivos titulares investigados.

```
CREATE (c1:Conta { nomeBanco : 'Banco1', agencia : '2912', numero : '514503-0' })
CREATE (c2:Conta { nomeBanco : 'Banco2', agencia : '9954', numero : '591565-3' })
CREATE (c3:Conta { nomeBanco : 'Banco1', agencia : '3455', numero : '205086-5' })
CREATE (p1)-[rt1:TITULAR]->(c1)
CREATE (p2)-[rt2:TITULAR]->(c2)
CREATE (p3)-[rt3:TITULAR]->(c3)
```

Figura 5.5: Inserção de contas de contas bancárias

#### 5.2.5. Inserção de encontros entre investigados filmados pela equipe operacional

A Tabela 5.9 detalha informações acerca de 18 encontros entre investigados registrados na operação exemplo.

Tabela 5.9: Informações sobre encontros

<b>Pessoa</b>	<b>Pessoa</b>	<b>Data</b>	<b>Horário</b>	<b>Local</b>
Alvo5	Alvo8	23/08/2015	08:15	Altamira/PA
Alvo3	Alvo2	24/08/2015	11:12	Altamira/PA
Alvo1	Alvo2	26/08/2015	10:07	Altamira/PA
Alvo4	Alvo2	26/08/2015	09:01	Altamira/PA
Alvo4	Alvo5	26/08/2015	23:55	Altamira/PA
Alvo5	Alvo1	26/08/2015	06:00	Altamira/PA

Alvo1	Alvo3	30/08/2015	19:15	Belém/PA
Alvo1	Alvo2	04/09/2015	22:34	Altamira/PA
Alvo6	Alvo10	04/09/2015	13:15	Brasília/DF
Alvo2	Alvo3	05/09/2015	18:04	Altamira/PA
Alvo2	Alvo1	10/09/2015	17:01	Brasília/DF
Alvo2	Alvo3	11/09/2015	16:22	Altamira/PA
Alvo3	Alvo1	15/09/2015	15:49	Altamira/PA
Alvo7	Alvo13	15/09/2015	16:22	Brasília/DF
Alvo8	Alvo10	17/09/2015	15:43	Brasília/DF
Alvo7	Alvo13	26/09/2015	14:01	Brasília/DF
Alvo9	Alvo5	26/09/2015	11:59	Brasília/DF
Alvo6	Alvo7	30/09/2015	12:04	Altamira/PA

A Figura 5.6 indica os comandos de inserção de dados no Neo4j referentes a 3 encontros entre investigados.

```
CREATE (p1)-[re1:ENCONTROU {data : '26/08/2015', hora : '10:07', local : 'Altamira/PA',
caminho : 'c:\\operação\\encontro1.mp3' } ] -(p2)
CREATE (p1)-[re2:ENCONTROU {data : '30/08/2015', hora : '19:15', local : 'Belém/PA',
caminho : 'c:\\operação\\encontro2.mp3' } ] -(p3)
CREATE (p1)-[re3:ENCONTROU {data : '04/09/2015', hora : '22:34', local : 'Altamira/PA',
caminho : 'c:\\operação\\encontro3.mp3' } ] -(p2)
```

Figura 5.6: Inserção de um encontro de investigados

#### 5.2.6. Inserção das chamadas telefônicas oriunda da quebra do sigilo telefônico

A Tabela 5.10 detalha uma amostra de informações sobre 250 chamadas telefônicas relativas a conversa entre investigados na operação exemplo. A tabela completa encontra-se no Apêndice A.

Tabela 5.10: Informações sobre chamadas telefônicas

<b>Origem Chamada</b>	<b>Destino Chamada</b>	<b>Data</b>	<b>Hora</b>	<b>Duração (em segundos)</b>
(31)8467-9923	(51)9658-4361	24/08/2015	12:03	90
(61)9615-0261	(51)9658-4361	24/08/2015	15:16	501
(31)8467-9923	(51)9658-4361	24/08/2015	10:38	266
(61)9934-1023	(51)9658-4361	24/08/2015	10:17	252
(61)9615-0261	(61)9934-1023	24/08/2015	12:36	561
(61)9636-7827	(61)8132-1530	24/08/2015	17:53	292
(51)9658-4361	(61)9636-7827	24/08/2015	9:47	313
...				

A Figura 5.7 indica os comandos de inserção de dados no Neo4j referentes a 3 chamadas telefônicas entre investigados.

```
CREATE (t6)-[rc1:CHAMOU { data : '24/08/2015' , hora : '12:03' , duração : 90 ,
caminho : 'c:\\operação\\chamada1.mp3' } ] ->(t3)
CREATE (t5)-[rc2:CHAMOU { data : '24/08/2015' , hora : '15:16' , duração : 501 ,
caminho : 'c:\\operação\\chamada2.mp3' } ] ->(t3)
CREATE (t6)-[rc3:CHAMOU { data : '24/08/2015' , hora : '10:38' , duração : 266 ,
caminho : 'c:\\operação\\chamada3.mp3' } ] ->(t3)
```

Figura 5.7: Inserção de chamadas telefônicas

5.2.7. Inserção da movimentação financeira do grupo oriunda da quebra do sigilo bancário:

A Tabela 5.11 detalha uma amostra de informações sobre 500 movimentações bancárias relativas às contas correntes dos investigados utilizadas na operação exemplo. A tabela completa encontra-se no Apêndice B.

Tabela 5.11: Informações sobre movimentação financeira

<b>C/C Origem</b>	<b>C/C Destino</b>	<b>Data</b>	<b>Valor</b>
514503-0	887965-3	24/08/2015	R\$ 150.000,00
591565-3	591807-6	08/09/2015	R\$ 150.000,00
317870-7	461039-2	24/08/2015	R\$ 790,00
317870-7	90088-5	24/08/2015	R\$ 980,00
317870-7	339463-9	24/08/2015	R\$ 2.500,00
134624-8	921758-2	24/08/2015	R\$ 3.600,00
887965-3	514503-0	24/08/2015	R\$ 80.000,00
104918-2	46826-3	24/08/2015	R\$ 6.800,00
104918-2	887965-3	24/08/2015	R\$ 35.000,00
...			

A Figura 5.8 indica os comandos de inserção de dados no Neo4j referentes a 3 movimentações bancárias entre contas-correntes de investigados.

```
CREATE(c1)-[rd22:DEPOSITOU { data : '24/08/2015' , valor :26000 }] -> (c7)
CREATE(c2)-[rd36:DEPOSITOU { data : '24/08/2015' , valor :59000 }] -> (c17)
CREATE(c4)-[rd96:DEPOSITOU { data : '24/08/2015' , valor :790 }] -> (c24)
```

Figura 5.8: Inserção da movimentação bancária

### 5.2.8. Inserção de mensagem eletrônica oriunda da interceptação telemática:

A Tabela 5.12: Informações de mensagens eletrônicas (*e-mails*) Tabela 5.12 detalha informações sobre 13 trocas de mensagens eletrônicas (*e-mails*) entre investigados.

Tabela 5.12: Informações de mensagens eletrônicas (*e-mails*)

Pessoa	Pessoa	Data	Horário	Assunto
Alvo1	Alvo2	27/08/2015	12:16	Encontro de ontem
Alvo8	Alvo7	27/08/2015	04:15	Depósito bancário
Alvo4	Alvo5	27/08/2015	14:44	Boletos para pagamento
Alvo5	Alvo1	27/08/2015	16:51	Depósitos
Alvo4	Alvo2	28/08/2015	06:52	Comissão
Alvo3	Alvo2	31/08/2015	14:21	Depósito efetuado
Alvo1	Alvo3	01/09/2015	21:26	Viagem para Brasília
Alvo1	Alvo2	06/09/2015	00:43	Bilhetes aéreos
Alvo2	Alvo3	10/09/2015	19:34	Segue assunto discutido semana passada
Alvo2	Alvo1	15/09/2015	12:10	Conversa de ontem
Alvo5	Alvo8	29/09/2015	18:06	Conta corrente
Alvo3	Alvo1	30/09/2015	19:58	Local para encaminhamento da mercadoria
Alvo8	Alvo1	30/09/2015	12:13	Altamira

A Figura 5.9 indica os comandos de inserção de dados no Neo4j referentes a 3 mensagens trocadas entre contas de *e-mail* de investigados.

```
CREATE (e1)-[rm1:ESCREVEU {data : '27/08/2015', hora : '12:16', assunto: 'Encontro de ontem',
caminho : 'c:\\operação\\email1.eml' }] ->(p2)
CREATE (e1)-[rm1:ESCREVEU {data : '01/09/2015', hora : '21:26', assunto: 'viagem para Brasília',
caminho : 'c:\\operação\\email2.eml' }] ->(p3)
CREATE (e1)-[rm2:ESCREVEU {data : '06/09/2015', hora : '00:43', assunto: 'Bilhetes aéreos',
caminho : 'c:\\operação\\email3.eml' }] ->(p2)
```

Figura 5.9: Inserção da troca de *e-mails* entre investigados

Após a inserção dos dados, o banco de dados Neo4j é capaz de fornecer os resultados da forma como apresentada na Figura 5.10. Nota-se que é possível perceber, a priori, o

relacionamento entre indivíduos (círculo azul), os telefones (círculos rosas), contas de *e-mail* (círculos laranjas) e contas bancárias (*círculos vermelhos*). A visão inicial dos dados, em que pese pareça apenas um emaranhado de entidades (círculos) e vínculos (traços e setas), já permite verificar algum padrão de relacionamento que, oportunamente, pode ser explorado pelos analistas do caso.

As informações contidas em tabelas ou estruturas resultantes de consultas em bases de dados relacionais não trariam à investigação resultados que, após refinados e submetidos a outros tipos operações, abram caminho para o apontamento de rumos e expansão dos trabalhos investigativos de forma gráfica e intuitiva, sem auxílio de ferramentas específicas ou *softwares* de análise de dados.

A facilidade de visualização das informações com ênfase em seus relacionamentos e tipos de entidade é uma vantagem considerável quando se utiliza o banco de dados orientado a grafos. Outro ganho a ser percebido é forma como a consulta é elaborada – em linguagem declarativa – o que evita longas construções e junções entre tabelas (*joins*). A seção 5.2.9 tratará de exemplos de exploração dessas informações e como é possível tirar vantagens ao optar pelo modelo aqui proposto.

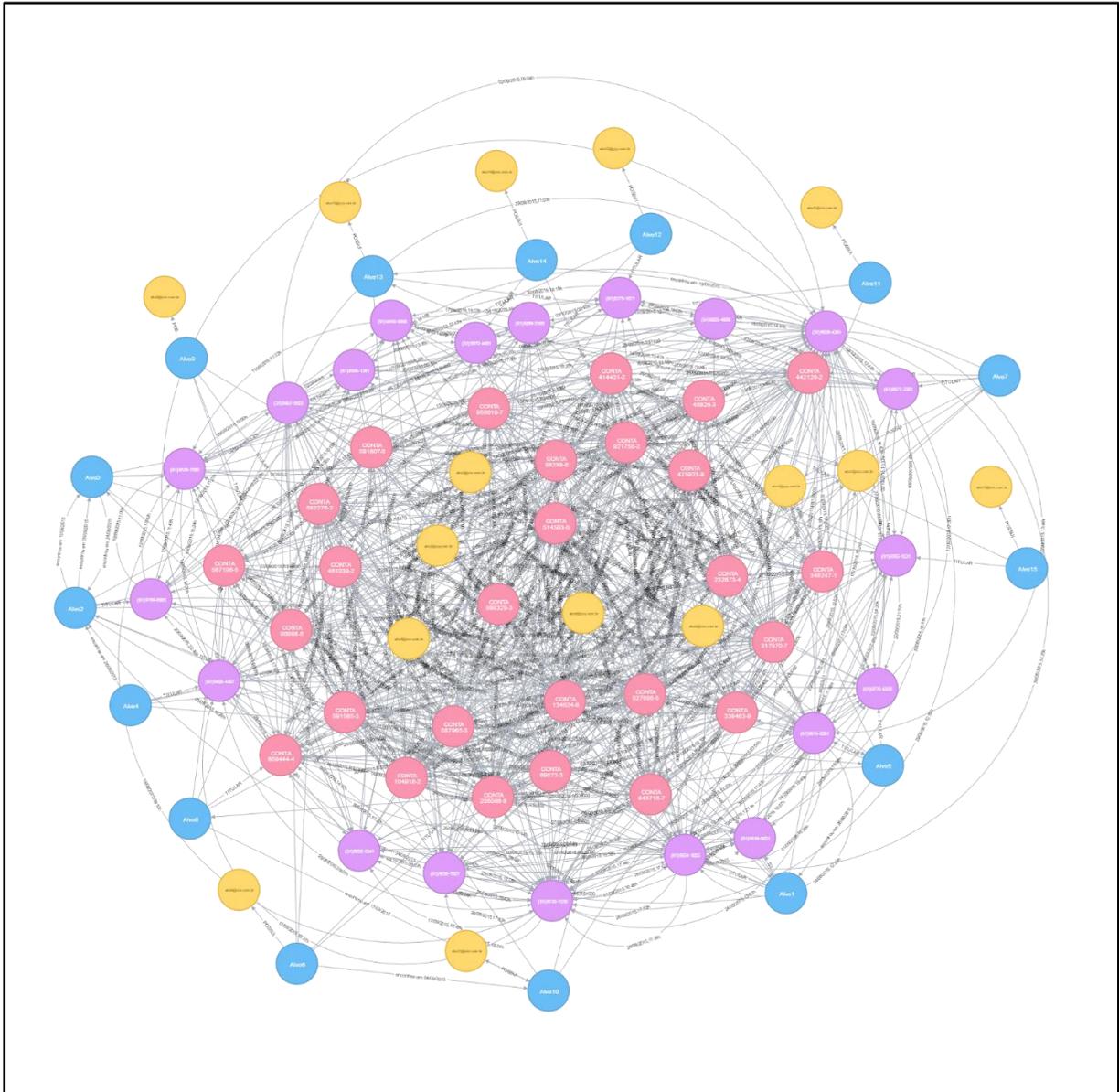


Figura 5.10: Diagrama de relacionamento entre investigadores, contas de *e-mail*, contas bancárias e telefones da operação policial

### 5.2.9. Exploração da Base de Dados

Esta subseção irá exemplificar quatro situações nas quais o conjunto de dados da investigação é manipulado por meio de diversos tipos de pesquisas e análises, buscando mostrar o quão prático é lidar com essas atividades num banco de dados de grafo face aos resultados possíveis de serem alcançados numa base relacional.

- a) Rede de encontros entre investigadores.

A consulta e seu resultado, retratados nas Figuras Figura 5.11 e Figura 5.12 mostram a dinâmica de relacionamento pessoal entre investigados, resultado de uma pesquisa no banco de dados em que foram relacionados todos os encontros registrados na operação. Nela é possível perceber com clareza quais os alvos possuem contato presencial com outras pessoas, bem como as datas e horários em que os fatos ocorrerem. É possível descobrir pessoas que se destacam em relação ao grupo e desenhar o alcance da organização investigada.

```
MATCH (PessoaOrigem)-[:ENCONTROU]->(PessoaDestino)
RETURN PessoaOrigem, PessoaDestino
```

Figura 5.11: Consulta referente à rede de encontros entre investigados

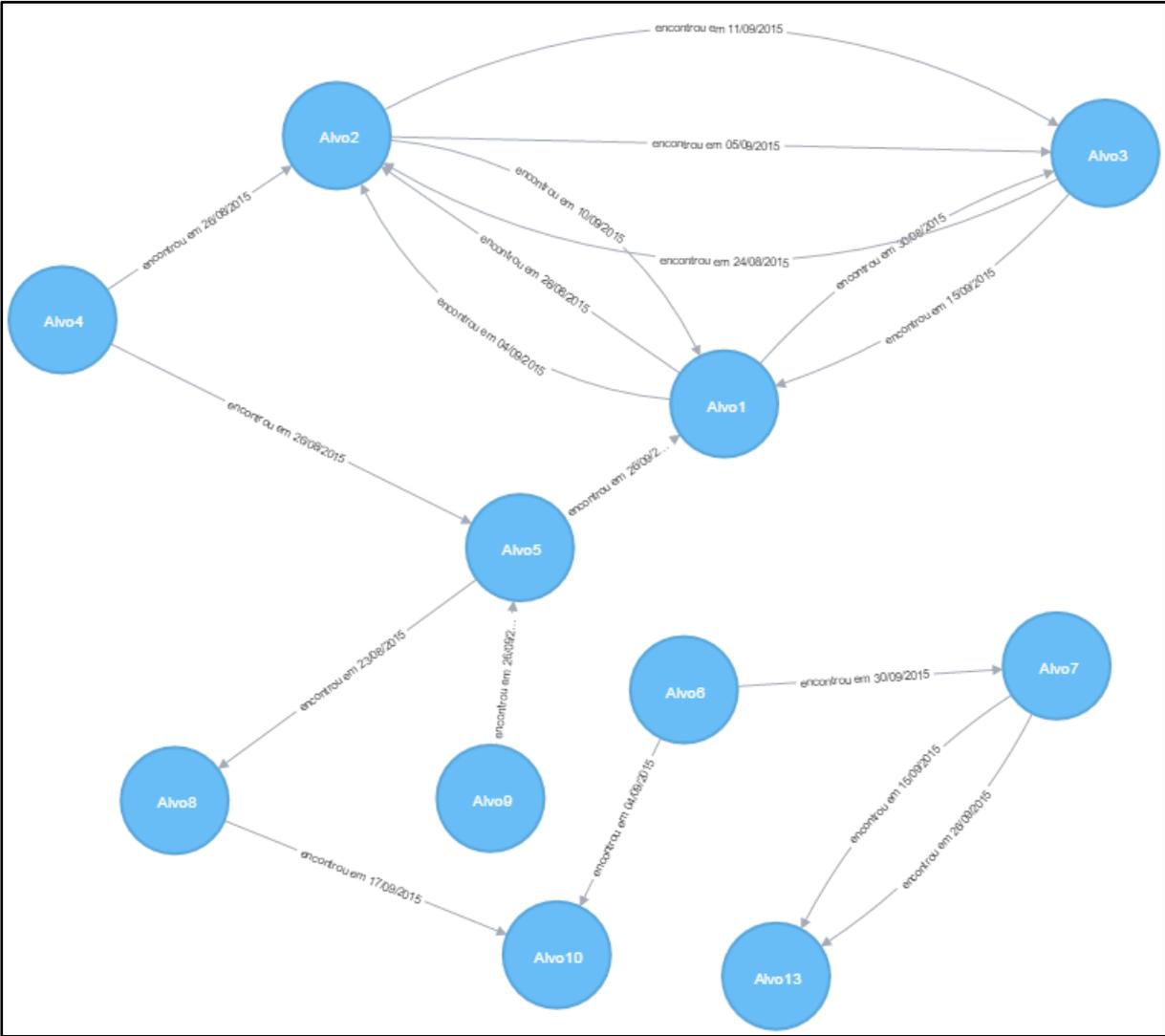


Figura 5.12: Rede de encontro entre investigados

Percebe-se claramente que o Alvo5 é o elemento central que une dois grupos de investigados: o primeiro, composto pelos Alvo1, Alvo2, Alvo3 e Alvo4, e o segundo, composto pelos Alvo6, Alvo7, Alvo8, Alvo9, Alvo10 e Alvo13.

No período da investigação, é possível também constatar a intensidade dos encontros entre os investigados Alvo1, Alvo2 e Alvo 3, os quais formam um núcleo sólido à parte.

b) Chamadas telefônicas realizadas por um determinado alvo

Ao analisar o grafo produzido no item a, percebe-se que Alvo5 possui papel de destaque na organização, uma vez que é o nó central responsável por ligar dois grupos de investigados.

A consulta e seu resultado apresentados nas Figuras Figura 5.13 e Figura 5.14 expandem os contatos pessoais realizados, ao detalhar também as ligações telefônicas e troca de mensagens eletrônicas (*e-mails*) realizadas entre os investigados.

```
MATCH (PessoaOrigem)-[:POSSUI_TEL]->(TelefoneOrigem)-[Chamou]->(TelefoneDestino)<-[:POSSUI_TEL]-(PessoaDestino)
MATCH (PessoaOrigem)-[:POSSUI_MAIL]->(EmailOrigem)-[Escreveu]->(EmailDestino)<-[:POSSUI_MAIL]-(PessoaDestino)
WHERE PessoaOrigem.nome = 'Alvo5'
RETURN PessoaOrigem, Pessoa Destino, TelefoneOrigem, TelefoneDestino, Emailorigem, EmailDestino
```

Figura 5.13: Consulta referente a chamadas telefônicas, e-mails e encontros de investigados

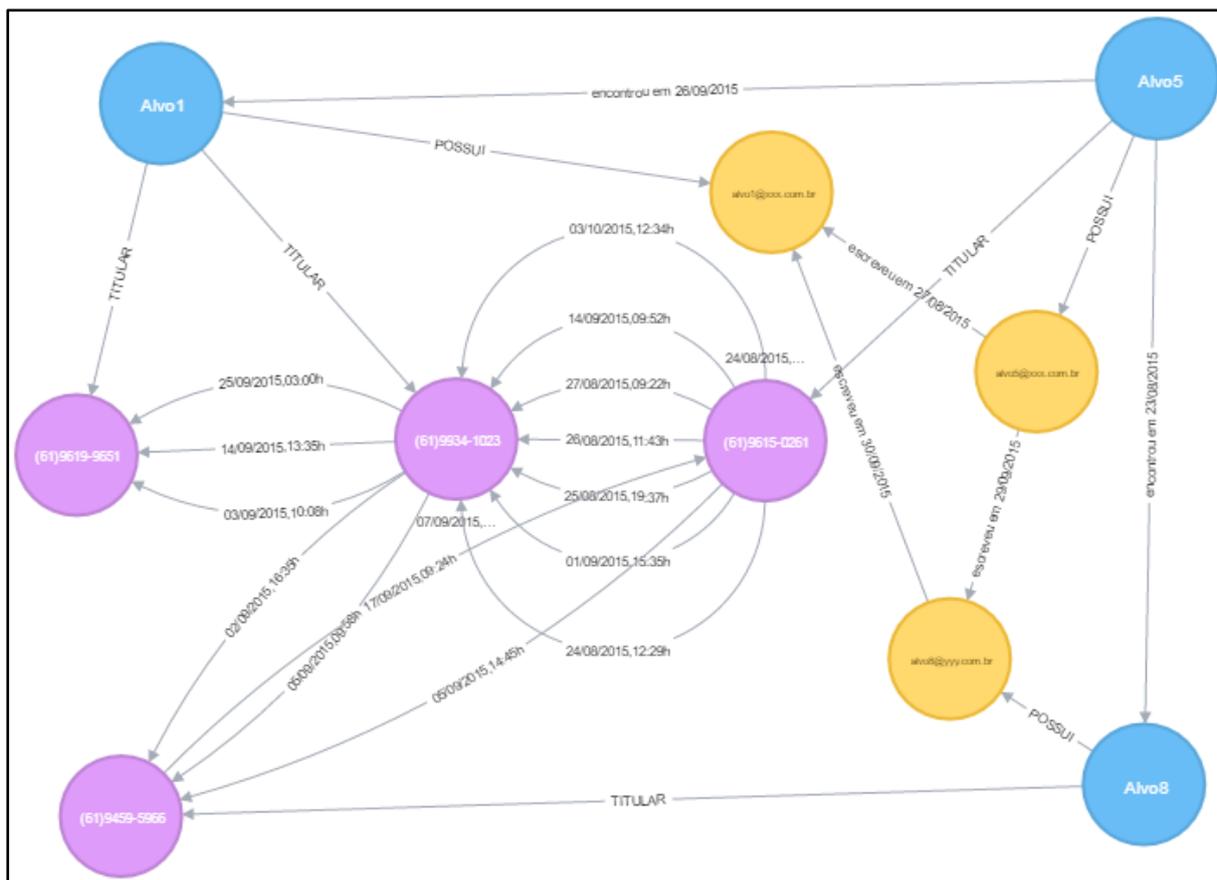


Figura 5.14: Chamadas telefônicas, e-mails e encontros de investigados

O grafo mostra que Alvo5 mantém conversas telefônicas com Alvo8 e Alvo1, além de estar presente em um encontro com cada um deles, nas datas 29/09/2015 e 26/09/2015, respectivamente. Também é possível perceber contatos entre esses dois últimos investigados, por meio de conversas telefônicas e mensagens de e-mail.

É possível constatar com clareza quais os terminais telefônicos mais utilizados por cada investigado. No presente caso, Alvo1 é proprietário de duas linhas telefônicas sendo que apenas uma delas, (61) 9934-1023, é que realiza e recebe chamadas para os aparelhos de Alvo5 e Alvo8.

Outro ponto positivo na utilização deste modelo – banco de dados de grafo – é a possibilidade de visualizar com facilidade diversos tipos de entidade de uma só vez (pessoas, contas de e-mail, telefones), o que permite uma análise ampla de acontecimentos. É possível constatar o período e nível de intensidade entre esses contatos, permitindo um enfoque num intervalo temporal mais específico que depois pode ser melhor trabalhado.

c) Transações financeiras

No conjunto de dados da operação, realizou-se consulta objetivando filtrar as transações bancárias superiores a R\$ 100.000,00. Optou-se, também, por combinar tais dados com os encontros realizados entre os investigados. O processamento da pesquisa e seu resultado, ilustrados nas Figuras Figura 5.15 e Figura 5.16, produziram um interessante grafo, no qual foi possível verificar a relação de contatos pessoais entre os alvos e o fluxo financeiro mais importante entre as pessoas.

```
MATCH (PessoaOrigem)-[e:ENCONTROU]->(PessoaDestino)
MATCH (PessoaOrigem)-[t:TITULAR]->(ContaOrigem)-[Depositou]->(ContaDestino)-[:TITULAR]-(PessoaDestino)
WHERE Depositou.valor >= 100000
RETURN PessoaOrigem, PessoaDestino, e, t, Depositou
```

Figura 5.15: Consulta referente a informações bancárias de investigados

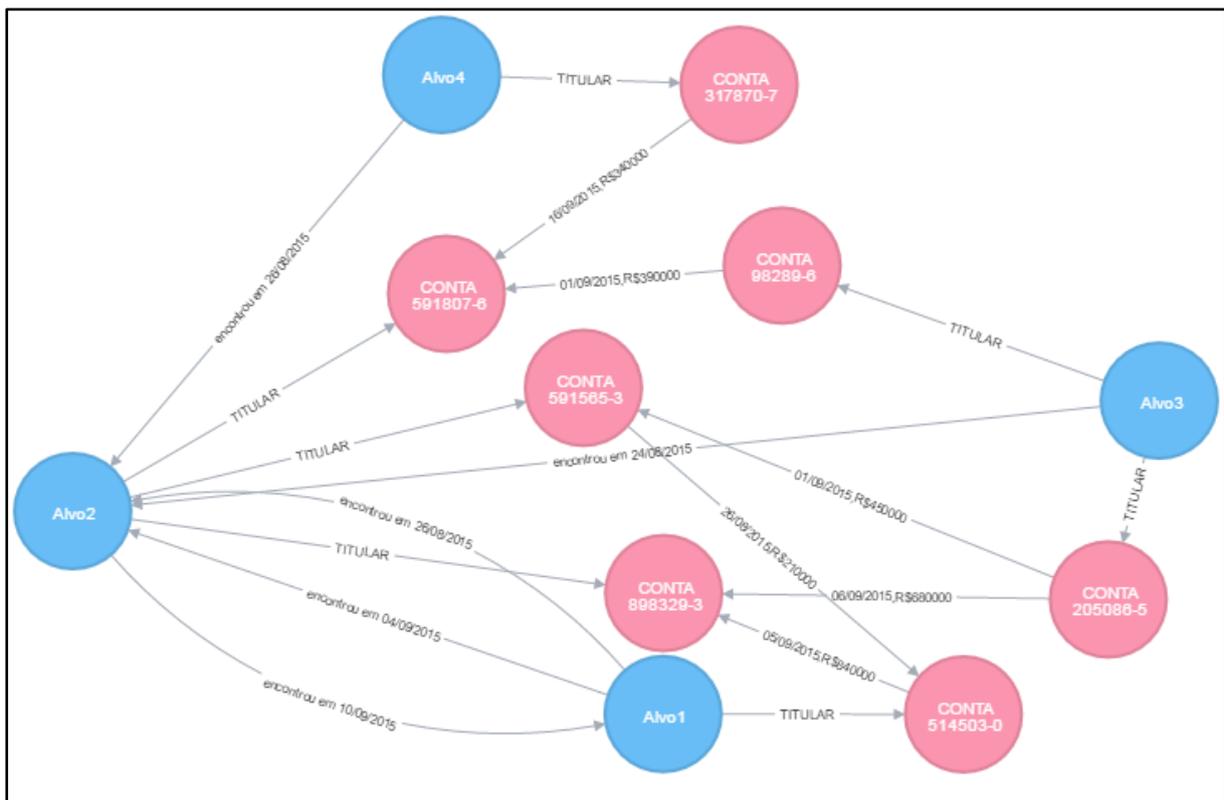


Figura 5.16: Informações bancárias de investigados

Na Figura 5.16 também é possível perceber que Alvo2 realizou depósito no valor de R\$ 210.000,00 na conta corrente 514503-0, cujo titular é Alvo1, na mesma data em que se

encontraram, em 26/08/2015. Por sua vez, Alvo1 depositou R\$ 84.000,00 para Alvo2, um dia após se encontrarem, em 04/09/2015.

Análises desse tipo são importantes, pois podem abrir a possibilidade de esmiuçar outras fontes de dados, como redes sociais, interceptação telefônica, interceptação telemática, dentre outras, no período em que encontros e transferências bancárias foram registrados com mais intensidade, objetivando coletar conjunto probatório mais diversificado e expandir o leque de investigados e eventos durante um certo período de tempo.

d) O “caminho” do dinheiro

Muitas vezes, objetivando dissimular a comprovação de pagamento de valores, criminosos fazem diversos repasses entre contas correntes de forma sistemática, a fim de tentar ludibriar os órgãos de controle governamental e esconder o pagamento de uma propina ou a retribuição financeira para o cometimento de crimes.

As Figura 5.17 e Figura 5.18 mostram, por meio de uma única consulta na base de dados, como o fluxo financeiro de uma conta bancária até outro beneficiário pode ser rastreado.

```
MATCH fFluxo =(:Conta { numero: "609444-4" })-[:DEPOSITOU*0..10]->(:Conta { numero: "887965-3" })
MATCH (Pessoa)-[:TITULAR]->(Conta)
RETURN Pessoa, Conta, fFluxo
```

Figura 5.17: Consulta referente ao fluxo financeiro ente contas bancárias

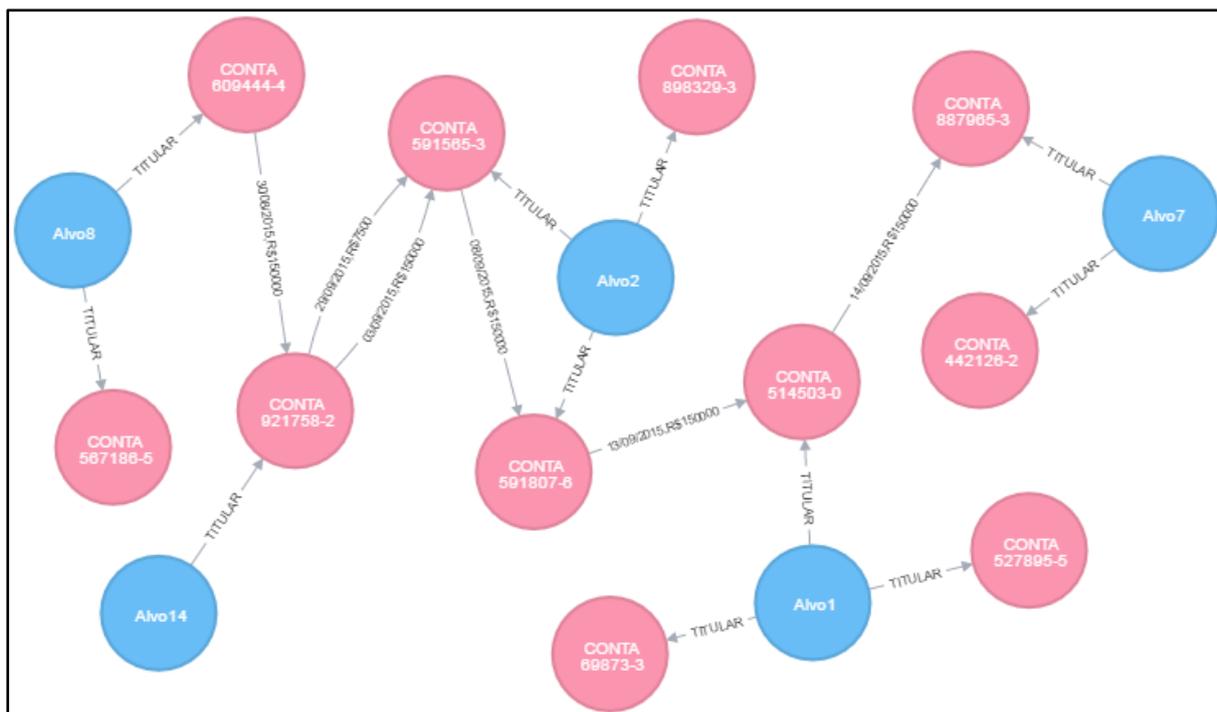


Figura 5.18: Fluxo financeiro ente contas bancárias

No exemplo, o valor de R\$ 150.000,00 depositado inicialmente por Alvo8 ao Alvo14, em 30/08/2015, é repassado por meio de 5 outros depósitos bancários, e finalmente chega em 14/09/2015 na conta bancária 887965-3, pertencente a Alvo7. São utilizados como intermediárias as contas correntes de mais 2 investigados –Alvo2 e Alvo1 – até o crédito na conta do destinatário final.

É possível também “caminhar” ao longo do grafo, percebendo o percurso que os valores entram e saem das contas bancárias, facilitando o trabalho do analista policial em perceber o fluxo financeiro da organização, inclusive, possibilitando a identificação de “laranjas” - pessoas que não possuem patrimônio suficiente para justificar sua movimentação financeira.

O resultado dessa consulta demonstra um típico caso em que funções específicas de banco de dados de grafo, relativas a operações sobre vértices e arestas, não presente em bases de dados tradicionais, revelam um ganho analítico importante.

Ante os exemplos apresentados, vê-se que o universo de possibilidades de realização de consultas é bastante amplo. Reforça-se a facilidade com que o investigador tem acesso à interface gráfica produzida automaticamente pelos resultados das pesquisas no banco de dados, o que permite obter um panorama favorável quando se trabalha com bases de dados de milhares de registros e seus relacionamentos diversos.

A utilização de bancos de dados relacionais leva à necessidade aplicativos e ferramentas acessórias de análise, uma vez que comparando os resultados de consultas – em forma de tabelas – não é possível ter uma visão ampla e estruturada do universo de relação entre os diversos tipos de entidades. Seria necessária também a construção de consultas complexas (*joins*) entre as entidades (tabelas), o que pode limitar o trabalho e impactar em seu tempo de resposta.

O banco de dados orientado a grafo é intuitivo e permite o aprofundamento e exploração das entidades e seus vínculos; ações simples (chamadas telefônicas ou encontros), por exemplo, podem ser vistos num contexto amplo e tem sua importância reavaliada a todo momento no contexto em que é apresentado. A seção 5.3 lida com a inserção de dados não estruturados no cluster Hadoop.

### **5.3. INSERÇÃO DOS DADOS SEMIESTRUTURADOS NO *CLUSTER* HADOOP**

A fim de exemplificar a manipulação de dados semiestruturados, optou-se por utilizar o Apache Hadoop 2.3.0 para o processamento de tarefas do tipo MapReduce. É uma distribuição de código aberto, que suporta HDFS, além de ser amplamente utilizada no mercado<sup>25</sup>.

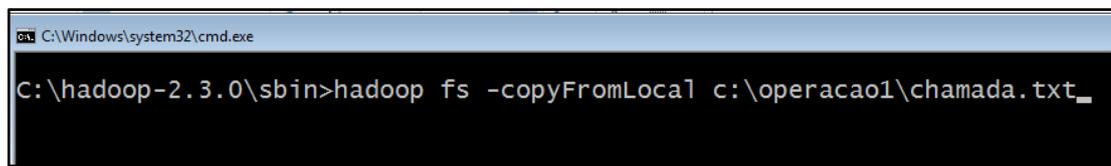
Os dados semiestruturados devem seguir para o sistema de arquivos do Hadoop – o HDFS – e, a partir de lá, serem submetidos a processamentos específicos, conforme já detalhado na seção 4.5.

A tarefa de inserir dados para o sistema de arquivos HDFS é simples, bastando para isso a execução de uma rotina automatizada que importe os dados provenientes do sistema de

---

<sup>25</sup> Instalação e documentação disponível em <https://archive.apache.org/dist/hadoop/core/hadoop-2.3.0/>

arquivos nativo. A Figura 5.19 mostra a execução desse comando no Hadoop versão 2.5 referente a uma transcrição de chamada telefônica:



```
C:\Windows\system32\cmd.exe
C:\hadoop-2.3.0\sbin>hadoop fs -copyFromLocal c:\operacao1\chamada.txt_
```

Figura 5.19: Transferência de arquivo para o HDFS

Após a inserção da massa de arquivos não estruturados no *cluster*, a próxima etapa é executar operações no conjunto de dados.

#### 5.4. REALIZAÇÃO DE OPERAÇÕES MAPREDUCE

Goldman (2012) propõe inúmeras aplicações construídas utilizando-se MapReduce como suporte a conjuntos de dados não estruturados. Uma delas é o contador de palavras (*WordCount*) em arquivos textos. Considerando a operação policial que serve de exemplo a este trabalho, pode-se utilizar arquivos textos oriundos de transcrições de interceptação telefônica e submetê-los ao procedimento *WordCount*<sup>26</sup>, a fim de descobrir quais são os termos mais citados nas conversas dos suspeitos. O resultado pode revelar informações importantes para a investigação, como referências a empresas, pessoas, endereços e valores que mais comumente estão presentes na investigação.

Para simular o processamento, criou-se um conteúdo para arquivos de transcrição de chamadas telefônicas “c:\operacao1\chamada.txt”, “c:\operacao1\chamada2.txt” e “c:\operacao1\chamada3.txt”, utilizados na operação modelo, conforme abaixo:

*“Alvo5 solicita um carregamento de madeira que deverá ser entregue à empresa MM. Alvo7 informa que o valor do carregamento é R\$ 30 mil. Alvo7 diz que ligará à tarde para Maria e a repassará seus dados bancários.”*

---

<sup>26</sup> O código do procedimento *WordCount* pode ser acessado em <http://disciplinas.stoa.usp.br/pluginfile.php/97549/course/section/29436/JAI2012-Hadoop.pdf>

*“Alvo7 liga para Alvo8, que trabalha na empresa MM, em Brasília, e informa seus dados bancários. Alvo7 informa que acabara de falar com Alvo5 e confirma a negociação. Alvo8 pergunta sobre o estado de saúde de Alvo9, esposa de Alvo7.”*

*“Alvo7 liga para Alvo5 em 29/08/2015, às 18:00h. Luiz agradece a João por esse ter intermediado todo o negócio, informa que já realizou o pagamento de 10% de sua comissão e que a madeira foi encaminhada à empresa MM. Alvo fala que Alvo10, seu sócio, está em Altamira/PA aguardando-o.”*

A descrição da filmagem realizada pela equipe operacional também utilizada no exemplo foi armazenada no arquivo texto “c:\operacao\filmagem.txt”. Seu conteúdo é:

*“Alvo5 e Alvo8 são empresários de Brasília conhecidos do ramo de movelaria e são vistos jantando na região onde costumeiramente ocorrem ações de desmatamento ilegal, em Altamira/PA. Após o encontro, eles seguiram, em veículos separados, para seus respectivos hotéis.”*

Após o carregamento dos arquivos no *cluster* Hadoop, eles foram submetidos à rotina *WordCount*. O fluxo de processamento sobre a contagem de palavras utilizando o paradigma de mapeamento e redução é mostrado na Figura 5.20.



Figura 5.20: Fluxo de processamento de arquivos não estruturados

Nota-se que o resultado do processamento pode revelar aspectos interessantes à investigação. Excluindo-se as palavras que são irrelevantes para um trabalho investigativo – preposições, conjunções, verbos e artigos, como “de”, “que”, “com”, “da”, “e”, “é”, dentre outras – aparecem com razoável frequência nomes de pessoas (“Alvo5”, “Alvo8”, “Alvo9”) e o nome de uma empresa (“MM”). Também foram citadas localidades (“Brasília” e “Altamira”), além de palavras que podem levar diretamente às atividades criminosas (“madeira”, “carregamento”).

Outros documentos podem ser agregados ao modelo, enriquecendo ainda mais as análises, tais como depoimentos, notícias, arquivos em fontes abertas e conteúdos em redes sociais, de forma a permitir a análise do conteúdo dessa massa de dados.

Holmes (2012), Lam (2010), Tao et al.(2013) e Rajaraman & Ullman (2012) descrevem várias funcionalidades que permitem ser programadas sob a forma de tarefas MapReduce e que podem auxiliar a análise de informações, a saber:

1. Ordenar valores (realiza-se a divisão do grupo em partes, sua ordenação interna e a junção de cada partição ordenada);
2. Encontrar palavras-chaves e ranqueá-las com base em ocorrência numa determinada massa de dados;
3. Identificar conteúdo léxicos similares entre milhares de documentos;
4. Identificar impressões digitais similares, por meio de comparação de características em fragmentos coletados;
5. Contar elementos diferentes num determinado fluxo de dados;
6. Identificar pontos geográficos que se localizam próximos uns dos outros;

Aplicando-se essas funcionalidades aos tipos de insumos informacionais utilizados pelas polícias judiciárias – discutidos na seção 3.2 –, a utilização desse modelo de programação na análise policial pode gerar resultados expressivos, conforme listado na Tabela 5.13. Nesse ambiente, as operações MapReduce se tornam protagonistas para auxiliar o analista na descoberta de novos padrões e relacionamento entre entidades que estão presentes em conjuntos de dados volumosos e não estruturados e que, por sua própria natureza, são difíceis de terem seu conteúdo pesquisado sem o auxílio de uma ferramenta.

Tabela 5.13: Operações MapReduce passíveis de serem aplicadas em fluxos de dados não estruturados.

<b>Insumo informacional</b>	<b>Operações MapReduce passíveis</b>
Processo tradicional de investigação	<p>Pesquisar por palavra-chave (nome de investigado, localidades, empresas) citados nos seguintes documentos: inquéritos policiais, apreensões, termos circunstanciados, partícipes de uma eventual ocorrência (presos, envolvidos ou testemunhas), guias de identificação, mandados de prisão;</p> <p>Ordenar as localidades mais citadas nos seguintes documentos: inquéritos policiais, apreensões, termos circunstanciados, partícipes de uma eventual ocorrência (presos, envolvidos ou testemunhas), guias de identificação, mandados de prisão.</p>
Laudos periciais e relatórios de identificação	<p>Pesquisa por palavra-chave (nome de investigado, localidades, empresas) contidas na base de laudos periciais e relatórios de identificação;</p> <p>Listar localidades por tipo de exame pericial, a fim de agrupar a ocorrência de determinado tipo de crime por região;</p> <p>Agrupar impressões digitais suspeitas e que possuem similaridade.</p>
Interceptação telefônica	<p>Pesquisa por palavra-chave (nome de investigado e empresas) nas transcrições realizadas dos áudios interceptados;</p> <p>Listagem de pessoas e empresas mais citadas nas conversas telefônicas.</p>

Interceptação telemática	Listagem de empresas e indivíduos mais citados no fluxo de dados interceptado;  Páginas de internet mais visitadas pelo investigado, num determinado período;  Identificar a existência de documentos similares em fluxos diferentes interceptados;
Vigilância a alvos	Pesquisa por palavra-chave (nome de investigado e empresas) em toda base de relatórios de vigilância operacional.
Fontes abertas	Pesquisa por localidades comuns visitadas por investigados e que foram postadas em redes sociais;  Pesquisa por palavra-chave em postagens, blogs e demais conteúdos abertos.

O que se percebe é que o ambiente HDFS vai dar a completude ao banco de dados de grafo na forma como a informação é relacionada e apresentada. Enquanto este foca no relacionamento entre as entidades e se apoia num importante recurso gráfico que reproduz a complexa rede de relacionamentos entre investigados e demais entidades, aquele vai realizar o processamento da massa bruta de informações, composta principalmente por dados semiestruturados e não estruturados que estão associados às entidades.

Realizar a junção dos tipos de entrada de dados, permitindo que as análises se complementem, é de fato superar um desafio grande e necessário à maioria das aplicações que necessitam de análises informacionais, o que não é diferente das atividades aqui relatadas.

Dessa forma, dados provenientes da interceptação telemática, documentos em variados formatos, *e-mails*, conteúdo de páginas da internet e redes sociais, dentre outros, poderão ser objeto de inúmeras operações MapReduce e oferecer novas informações que poderão compor o portfólio de conhecimento da investigação, expandindo os horizontes de trabalho.

## 6. CONCLUSÕES

O presente trabalho buscou encontrar alternativas para facilitar as tarefas de análise de vínculos, principalmente em relação à visualização de grandes volumes de dados, a fim de melhorar o processo de investigação das polícias judiciárias brasileiras. Propõe-se a utilização de uma plataforma de armazenamento e análise de dados desvinculada dos sistemas tradicionais - baseados em bancos de dados relacionais -, os quais não estão preparados para lidar com dados volumosos e não estruturados. Obteve-se resultado satisfatório com a utilização de banco de dados orientado a grafos com suporte a *big data*.

Apresenta-se também um estudo das fontes de informação utilizadas pelas polícias judiciárias no Brasil e a viabilidade de sua aplicação a um modelo de *big data*, cujos princípios vem paulatinamente sendo empregados por instituições de segurança pública, defesa nacional e combate à criminalidade em diversos países.

Buscando a definição dos componentes da plataforma de investigação, foram realizadas análises comparativas dos principais tipos de bases de dados NoSQL, levando-se em conta critérios essenciais para o trabalho de policial, a saber: representação amigável ao relacionamento entre entidades, bom desempenho na consulta de entidades e seus relacionamentos e suporte à escalabilidade dos dados. Dentre os tipos de repositório analisados, verificou-se que bancos de dados orientados a grafos oferecem a melhor resposta aos desafios existentes no contexto informacional vivido pelos órgãos policiais, permitindo ganhos consideráveis em relação à visualização de informações e análise de vínculos entre entidades diversas.

Em relação aos dados não estruturados, a adoção de um cluster Hadoop traz as seguintes vantagens à manipulação de grandes volumes de informação: tolerância a falhas de hardware, balanceamento de tarefas entre nós computacionais, portabilidade entre plataformas de software e hardware, escalabilidade e fácil gerenciamento dos componentes. A presença do HDFS na plataforma proposta justifica-se pela necessidade de complementar o trabalho feito pela base de grafos, ao permitir o processamento de grandes massas de dados que não são catalogáveis de forma estruturada, tais como documentos em formato texto, laudos periciais, relatórios de atividades operacionais, descrição de vídeos, transcrição de áudios, fluxos de interceptação de dados.

Quanto à realização de testes na plataforma proposta, devido à dificuldade relatada em se utilizar dados reais de investigações policiais, cujos materiais encontram-se em arquivos da justiça e polícia – o que pode causar constrangimentos às pessoas envolvidas -, optou-se por elaborar uma operação policial modelo, que contou com insumos informacionais aleatórios dos seguintes tipos: atividades operacionais, interceptação telemática, interceptação telefônica e dados bancários.

As funcionalidades oferecidas pela plataforma, tais como criação de grafos de relacionamento, realização de consultas por meio de linguagem declarativa, possibilidade de manipulação e visualização de centenas de entidades e de relacionamentos de forma mais amigável, processamento de informações oriundas de fontes diversificadas e possibilidade de realizar operações mais complexas em dados não estruturados, dentre outras, trazem ganhos analíticos aos policiais e peritos que lidam com processos investigativos complexos, havendo, inclusive, maior agilidade na descoberta de novos indícios.

Esta pesquisa proporciona diretrizes para o desenvolvimento de um ferramental que melhorará todo o processo investigativo, principalmente naquilo que tange a produção de provas, resultando em celeridade e qualidade da apuração das infrações criminais, trazendo benefícios para os cidadãos no contexto em que cada vez mais são produzidos, disponibilizados e acessíveis dados e informações na sociedade.

## **6.1. TRABALHOS FUTUROS**

Sugere-se que trabalhos futuros realizem a sedimentação da tecnologia proposta, principalmente em termos de interface e facilidade em seu uso corriqueiro nas investigações e perícias levadas a cabo pelos órgãos policiais. O desenvolvimento de um protótipo também poderá viabilizar os ganhos vislumbrados neste estudo.

Testes podem ser realizados a fim de avaliar o desempenho do sistema no processamento dos grafos gerados pelo modelo que, conforme visto, pode ter problemas de escalabilidade à medida que o número de vértices e arestas aumentam de forma considerável.

Em relação à programação em MapReduce, existe ainda um campo não muito explorado sobre o desenvolvimento e o refino de rotinas específicas e customizadas para o trabalho policial; tarefas essas que podem trazer benefícios em relação à operacionalidade e

funcionalidade da plataforma, ampliando o horizonte de processamento e construção do conhecimento para os analistas.

Noutra vertente, o poder de concentrar informação e extrair análises de forma fácil e completa exige que trabalhos voltados para a camada de segurança da plataforma sejam desenvolvidos, visando a proteção dos dados manipulados e permitindo sua utilização prática nos órgãos policiais.

## REFERÊNCIAS BIBLIOGRÁFICAS

Abramova, V., Bernardino, J. & Furtado, P. (2014). Which NoSQL Database? A Performance Overview.

Angles, R. & Gutierrez, C. (2008). Survey of graph database models. *ACM Computing Surveys (CSUR)*, v. 40, n. 1, p. 1.

ANP/DPF. (2009). *Investigação policial criminal*. Brasília: Academia Nacional de Polícia.

Banco Central do Brasil. (2010). Carta-Circular 3.454, 14 jun. 2010. Acesso em 27 jul. 2015, disponível em <[http://www.bcb.gov.br/pre/normativos/busca/normativo.asp?tipo=c\\_circ&ano=2010&numero=3454](http://www.bcb.gov.br/pre/normativos/busca/normativo.asp?tipo=c_circ&ano=2010&numero=3454)>. Acesso em 27 jul. 2015.

Batista, G. et al. (2004). Ciberterrorismo: a nova forma de crime do sec. XXI, como combatê-lo?. *Revista da Academia Militar, Lisboa*, v. 6, n. 1, p. 29-60.

Batra, S., & Tyagi, C. (2012). Comparative analysis of relational and graph databases. *International Journal of Soft Computing and Engineering (IJSCE)*, 2(2), 509-512.

Breternitz, V. J., & Silva, L. A. (2013). Big data: Um novo conceito gerando oportunidades e desafios. *Revista Eletrônica de Tecnologia e Cultura*, 2(2).

Bush, V. (1945). Science: The endless frontier. *Transactions of the Kansas Academy of Science (1903-)*, v. 48, n. 3, p. 231-264.

Capgemini Consulting. (2013). *An Interview with Mike Hainey. Head of Data Analytics at HMCR*.

Cattel, R. Scalable SQL and NoSQL data stores.(2011). *ACM SIGMOD Record*, v. 39, n. 4, p. 12-27.

Cernicchiaro, L. V. (1997). Crime organizado. *Revista CEJ* 1.2: 98-100.

Computerworld. (2016). Cisco projeta expansão intensa dos dados móveis (afetando custos de TI). Acesso em 23 fev. 2016, disponível em: <<http://computerworld.com.br/cisco-projeta-expansao-intensa-dos-dados-moveis-afetando-custos-de-ti>>.

Da Silva, I. J. Inteligência Policial e Redes Sociais: a Polícia Federal em busca de uma política constitucionalmente sustentável (2014). Revista Brasileira de Ciências Policiais, v. 4, n. 1, p. 93-124.

Datameer. (2014). Detroit law enforcement's secret weapon: big data analytics. Acesso em 08 Fev 2015, disponível em <<http://www.datameer.com/blog/customer-stories/detroit-law-enforcements-secret-weapon-big-data-analytics.html>>.

De Souza, V. C. O. & Dos Santos, M. V. C. (2015). Maturing, Consolidation and Performance of NoSQL Databases-Comparative Study. In: Proceedings of the annual conference on Brazilian Symposium on Information Systems: Information Systems: A Computer Socio-Technical Perspective-Volume 1. Brazilian Computer Society.

Departamento de Polícia Federal. (2015). ETL-UNICART. Manual do Usuário. Versão 1.01. 13 Fev. 2015.

Direitonet. (2005). Interceptação telefônica e quebra de sigilo de dados. Acesso em 05 Ago. 2015, disponível em <<http://www.direitonet.com.br/artigos/exibir/2346/Interceptacao-telefonica-e-quebra-de-sigilo-de-dados>>.

FBI (2014). Cyber Task Forces. (2014). Acesso em 20 mar. 2015, disponível em: <<http://www.fbi.gov/about-us/investigate/cyber/cyber-task-forces-building-alliances-to-improve-the-nations-cybersecurity-1>>.

Garrido, R. G. (2009). Evolução dos Processos de Identificação Humana: das características antropométricas ao DNA. Genética na Escola, v. 5, p. 38-40.

Goldman, A. (2012). Apache Hadoop: conceitos teóricos e práticos, evolução e novas possibilidades. XXXI Jornadas de atualizações em informática, p. 88-136.

Governo do Distrito Federal (2009). Decreto nº 30.490, de 22 de junho de 2009. Acesso em 04 Mar. 2016, disponível em <<http://sileg.sga.df.gov.br>>.

Governo do Estado de Minas Gerais (2008). Decreto nº 43.852, de 11 de agosto de 2008, e suas alterações. Acesso em 04 Mar. 2016, disponível em <<http://www.almg.gov.br>>.

Governo do Estado de São Paulo (1969). Decreto nº 52.213, de 24 de julho de 1969, e suas alterações. Acesso em 04 Mar. 2016, disponível em <[al.sp.gov.br](http://al.sp.gov.br)>.

Governo do Estado do Rio de Janeiro (2015). Decreto nº 45.222, de 16 de abril de 2015. Disponível em <[ilep.planejamento.rj.gov.br](http://ilep.planejamento.rj.gov.br)>.

Han, J. et al. (2011). Survey on NoSQL database. In: Pervasive computing and applications (ICPCA), 6th international conference on. IEEE, 2011. p. 363-366.

Holmes, A. (2012). Hadoop in practice. Manning Publications Co.

Holzschuher, F. & Peinl, R. (2013). Performance of graph query languages: comparison of cypher, gremlin and native access in Neo4j. In Proceedings of the Joint EDBT/ICDT 2013 Workshops (pp. 195-204). ACM.

Houses of Parliament. (2014). Parliamentary Office of Science e Technology. Post Note Number 470.

IDC Iview. (2012). The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East.

Jaiswal, G., & Agrawal, A. P. (2013). Comparative analysis of Relational and Graph databases. IOSR Journal of Engineering (IOSRJEN).

Jornal do Brasil (2010). Sistema de Interceptação de Sinais revolucionária investigação. Publicado em 23 de maio de 2010. Acesso em 23 de Abr 2015, disponível em <<http://www.jb.com.br/pais/noticias/2010/05/23/sistema-de-interceptacao-de-sinais-revolucionaria-investigacao/>>.

Junior, N. N., & de Andrade Nery, R. M. (2006). *Constituição Federal comentada: e legislação constitucional: atualizada até 10 de abril de 2006*. Editora Revista dos Tribunais.

Kaur, H. et al. (2013). A Review Of Non Relational Databases, Their Types, Advantages And Disadvantages. *International Journal of Engineering Research and Technology (IJERT)*, v. 2.

Kumar, R., Parashar, B. B., Gupta, S., Sharma, Y. & Gupta, N. (2014). Apache Hadoop, NoSQL and NewSQL Solutions of Big Data. *International Journal of Advance Foundation and Research in Science & Engineering (IJAFRSE)*, 1(6), 28-36.

Kyrola, A., & Guestrin, C. (2014). GraphChi-DB: Simple Design for a Scalable Graph Database System--on Just a PC. arXiv preprint arXiv:1403.0701.

Lam, C. (2010). *Hadoop in action*. Manning Publications Co.

Leite, S. S. (2014). O Emprego das Fontes Abertas no Âmbito da Atividade de Inteligência Policial. *Revista Brasileira de Ciências Policiais*, 5(1), 11-45.

Li, K. C., Jiang, H., Yang, L. T., & Cuzzocrea, A. (Eds.). (2015). *Big Data: Algorithms, Analytics, and Applications*. CRC Press.

Liu, B., & Chen-Chuan-Chang, K. (2004). Editorial: special issue on web content mining. *Acm Sigkdd explorations newsletter*, 6(2), 1-4.

Loshin, D. (2013). *Big data analytics: from strategic planning to enterprise integration with tools, techniques, NoSQL, and graph*. Elsevier.

Lyman, P. & Varian, H. R.; *How Much Information 2003?* University of California, Berkeley, School of Information Management and Systems. Acesso 23 mai. 2014, disponível em: <<http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/execsum.htm>>.

Manikonda, L., Hu, Y. & Kambhampati, S. (2014). Analyzing user activities, demographics, social network structure and user-generated content on instagram. arXiv preprint arXiv:1410.8099.MANYIKA, James et al. Big data: The next frontier for innovation, competition, and productivity. 2011.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity.

McCreary, D., & Kelly, A. (2013). Making sense of NoSQL. Greenwich, Conn.: Manning Publications.

Ministério da Justiça (2011). Portaria nº 2.877, de 30 de dezembro de 2011. Acesso em 04 Mar. 2016, disponível em <<http://www.pf.gov.br>>.

Ministério da Justiça. (2014). Prestação de Contas Ordinárias Anual. Relatório de Gestão do Acesso em 02 Jun. 2015, disponível em <<https://contas.tcu.gov.br/econtrole/ObterDocumentoSisdoc?codArqCatalogado=7266953&seAbrirDocNoBrowser=1>>

Moniruzzaman, A. B. M. & Hossain, S. A. (2013). Nosql database: New era of databases for big data analytics-classification, characteristics and comparison. arXiv preprint arXiv:1307.0191.Nascimento, A. C. A., & Winter, A (2008). On the Oblivious-Transfer Capacity of Noisy Resources. IEEE Transactions on Information Theory, v. 54, p. 2572-2581.

Nayak, A., Poriya, A., & Poojary, D. (2013). Type of NOSQL databases and its comparison with relational databases. International Journal of Applied Information Systems, 5(4), 16-19.

Omand, D., Bartlett, J., & Miller, C. (2012). Introducing social media intelligence (SOCMINT). Intelligence and National Security, 27(6), 801-823.

Oracle Data Mining (ODM) Blog. (2014). How to import JSON data to Data Mining for Data. 08 Dez. 2014. Acesso em 01 Jul. 2014, disponível em <[https://blogs.oracle.com/datamining/entry/how\\_to\\_import\\_json\\_data](https://blogs.oracle.com/datamining/entry/how_to_import_json_data)>.

Padhy, R. P., Patra, M. R., & Satapathy, S. C. (2011). RDBMS to NoSQL: Reviewing some next-generation non-relational databases. *International Journal of Advanced Engineering Science and Technologies*, 11(1), 15-30.

Peron, A. (2012). SIT e CLIT: ferramentas e metodologia para aprimoramento de investigações criminais utilizando interceptações de conexão à internet.

PGR. (2012). Memorando de Instução – ASSPA/PGR. Versão 2.1. Acesso em 27 Jul. 2015, disponível em <<https://asspaweb.pgr.mpf.gov.br/site/index.php/sistemas/sigilo-bancario>>. Acesso em 27 jul. 2015.

Rajaraman, A., & Ullman, J. D. (2012). *Mining of massive datasets* (Vol. 1). Cambridge: Cambridge University Press.

Raste, K. S. (2014). *Big Data Analytics-Hadoop Performance Anaylsis*.

Revista Exame. (2014). 50 bi de dispositivos estarão conectados à internet até 2020, 4 mar. 2014. Acesso em 10 fev. 2016, disponível em: <http://exame.abril.com.br/tecnologia/noticias/50-bi-de-dispositivos-estarao-conectados-a-internet-ate-2020>>.

Rodrigues, C. V., da Silva, M. T., & Truzzi, O. M. S. (2010). Perícia criminal: uma abordagem de serviços. *Gest. Prod.*, São Carlos, 17(4), 843-857.

Russell, M. A. (2013). *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More*. " O'Reilly Media, Inc."

Sadalage, P. J., & Fowler, M. (2013). *NoSQL Essencial: um guia conciso para o mundo emergente da persistência poliglota*. Novatec, São Paulo.

Santos, D. C. (2014). *Coleta automatizada e análise de dados em fan pages do facebook*.

Saracevic, T. (2008). *Ciência da informação: origem, evolução e relações*. *Perspectivas em ciência da informação*, 1(1).

Silva, J. G. D., & Geraldo, J. (2000). O inquérito policial e a polícia judiciária. Campinas: Bookseller, 21.

Silva, M. A. D. (2011). Utilização da Computação Distribuída para o Armazenamento e Indexação de Dados Forenses. Dissertação de Mestrado, Publicação PPGENE.DM – 091/12, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 138p.

Soares, S. (2012). Big Data Governance. Information Asset, LLC.

Social Media Today. (2014). Top 5 Social Media Predictions for 2015. Acesso em 12 Abr 2015, disponível em <<http://www.socialmediatoday.com/content/top-5-social-media-predictions-2015>>.

Souza, C. A. D. & Zwicker, R. (2000). Sistemas integrados de gestão empresarial: estudos de casos de implementação de sistemas ERP. Faculdade de Economia, Administração e Contabilidade.

Sremack, J. (2015). Big Data Forensics–Learning Hadoop Investigations. Packt Publishing Ltd.

State University of New Jersey Office. (2014). Predicting crime with big data, affordably. ScienceDaily. Acesso em 18 Mar. 2014, disponível em: <<http://www.sciencedaily.com/releases/2014/01/140110142212.htm>>.

Sullivan, D. (2015). NoSQL for Mere Mortals. Addison-Wesley Professional.

Tao, Y., Lin, W., & Xiao, X. (2013). Minimal mapreduce algorithms. In Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data (pp. 529-540). ACM.

Tauro, C. J., Aravindh, S. & Shreeharsha, A. B. (2012). Comparative study of the new generation, agile, scalable, high performance NOSQL databases. International Journal of Computer Applications, 48(20), 1-4.

Tiwari, S. (2011). Professional NoSQL. John Wiley & Sons.

U. S. Department Of Justice. (2014). Office of Justice Programs. Nacional Institute of Justice. New Perspectives in Policing.

Uddin, M. F., & Gupta, N. (2014, April). Seven V's of Big Data understanding Big Data to extract value. In American Society for Engineering Education (ASEE Zone 1), 2014 Zone 1 Conference of the (pp. 1-5). IEEE.

Vieira, M. R., Figueiredo, J. M. D., Liberatti, G. & Viebrantz, A. F. M. (2012). Bancos de Dados NoSQL: conceitos, ferramentas, linguagens e estudos de casos no contexto de Big Data. Simpósio Brasileiro de Bancos de Dados.

Wendt, E., & Jorge, H. V. N. (2013). Crimes cibernéticos: Ameaças e procedimentos de investigação. Brasport.

Willke, T. L., Jain, N., & Gu, H. (2012). GraphBuilder—A Scalable Graph Construction Library for Apache™ Hadoop™. Big Learning WS at NIPS.

## **APÊNDICES**

**A – DADOS DE CHAMADAS TELEFÔNICAS DA OPERAÇÃO EXEMPLO**

<b>Origem Chamada</b>	<b>Destino Chamada</b>	<b>Data</b>	<b>Hora</b>	<b>Duração (em segundos)</b>
(31)8467-9923	(51)9658-4361	24/08/2015	12:03	90
(61)9615-0261	(51)9658-4361	24/08/2015	15:16	501
(31)8467-9923	(51)9658-4361	24/08/2015	10:38	266
(61)9934-1023	(51)9658-4361	24/08/2015	10:17	252
(61)9615-0261	(61)9934-1023	24/08/2015	12:36	561
(61)9636-7827	(61)8132-1530	24/08/2015	17:53	292
(51)9658-4361	(61)9636-7827	24/08/2015	9:47	313
(51)9658-4361	(61)9934-1023	24/08/2015	9:59	269
(61)9615-0261	(61)8132-1530	24/08/2015	13:01	388
(31)8467-9923	(51)9658-4361	24/08/2015	18:28	427
(61)9615-0261	(61)9934-1023	24/08/2015	12:29	309
(61)9636-7827	(61)8132-1530	25/08/2015	19:42	401
(31)8467-9923	(61)8132-1530	25/08/2015	13:15	473
(31)8467-9923	(51)9658-4361	25/08/2015	9:11	463
(61)9495-4487	(61)8132-1530	25/08/2015	9:01	264
(61)9615-0261	(51)9658-4361	25/08/2015	10:01	270
(61)9615-0261	(61)8132-1530	25/08/2015	19:15	114
(61)9615-0261	(61)9636-7827	25/08/2015	11:01	97
(61)9615-0261	(61)8132-1530	25/08/2015	19:40	279
(61)9495-4487	(51)9658-4361	25/08/2015	18:55	278
(61)8132-1530	(61)9934-1023	25/08/2015	13:39	335
(61)8132-1530	(61)9459-5966	25/08/2015	17:17	369
(61)9615-0261	(61)9934-1023	25/08/2015	19:37	219
(61)8132-1530	(61)9934-1023	25/08/2015	11:44	43
(61)9636-7827	(61)9934-1023	25/08/2015	18:49	499
(61)8132-1530	(51)9658-4361	25/08/2015	13:23	358
(61)9934-1023	(61)8132-1530	26/08/2015	17:52	384
(61)9495-4487	(61)8132-1530	26/08/2015	16:35	146
(61)9495-4487	(51)9658-4361	26/08/2015	14:06	497
(61)9615-0261	(61)9934-1023	26/08/2015	11:43	88
(61)9934-1023	(61)8132-1530	26/08/2015	11:35	565
(61)8132-1530	(31)8696-5241	27/08/2015	13:41	51
(61)9636-7827	(61)8132-1530	27/08/2015	9:54	436
(61)9934-1023	(51)9658-4361	27/08/2015	16:50	465
(61)9495-4487	(61)8132-1530	27/08/2015	19:51	485
(61)9615-0261	(61)9934-1023	27/08/2015	9:22	28
(61)9636-7827	(51)9658-4361	27/08/2015	11:10	566
(61)8132-1530	(61)9636-7827	27/08/2015	18:24	312
(61)9636-7827	(61)9934-1023	28/08/2015	16:48	165

(51)9658-4361	(61)9934-1023	28/08/2015	12:53	492
(61)8132-1530	(61)9615-0261	29/08/2015	9:35	420
(31)8467-9923	(61)9934-1023	29/08/2015	18:19	530
(51)9658-4361	(61)8132-1530	29/08/2015	12:55	27
(51)9658-4361	(61)8132-1530	29/08/2015	15:08	108
(31)8467-9923	(51)9658-4361	29/08/2015	11:03	256
(61)9636-7827	(61)8132-1530	29/08/2015	12:08	266
(51)9658-4361	(61)9934-1023	30/08/2015	16:51	358
(31)8467-9923	(51)9658-4361	30/08/2015	19:15	405
(31)8467-9923	(61)9934-1023	30/08/2015	11:47	237
(31)8467-9923	(61)8132-1530	30/08/2015	19:19	101
(61)9934-1023	(51)9658-4361	30/08/2015	10:48	453
(51)9658-4361	(61)9934-1023	30/08/2015	14:25	537
(61)9495-4487	(61)8132-1530	30/08/2015	10:22	74
(61)9615-0261	(51)9658-4361	30/08/2015	15:31	467
(61)9495-4487	(61)8132-1530	31/08/2015	19:23	484
(31)8467-9923	(61)8132-1530	01/09/2015	11:32	511
(61)9615-0261	(61)9934-1023	01/09/2015	15:35	55
(61)8132-1530	(51)9658-4361	01/09/2015	17:55	44
(61)9495-4487	(51)9658-4361	01/09/2015	13:58	50
(31)8467-9923	(61)8132-1530	01/09/2015	10:17	599
(61)9615-0261	(51)9658-4361	01/09/2015	10:33	403
(51)9658-4361	(61)9934-1023	01/09/2015	12:21	308
(31)8467-9923	(61)8132-1530	01/09/2015	19:34	599
(51)9658-4361	(61)8132-1530	01/09/2015	15:20	44
(61)8132-1530	(61)9934-1023	01/09/2015	16:48	399
(61)8132-1530	(51)9658-4361	01/09/2015	15:28	431
(61)9636-7827	(51)9658-4361	01/09/2015	17:30	110
(61)8132-1530	(61)9934-1023	01/09/2015	17:44	352
(31)8467-9923	(51)9658-4361	02/09/2015	9:04	590
(61)8132-1530	(51)9658-4361	02/09/2015	11:24	31
(61)9615-0261	(51)9658-4361	02/09/2015	17:59	30
(31)8467-9923	(61)8506-1381	02/09/2015	10:38	312
(61)9934-1023	(61)9459-5966	02/09/2015	16:35	225
(61)9495-4487	(91)9565-1231	02/09/2015	11:50	486
(61)9636-7827	(91)9565-1231	02/09/2015	11:18	400
(61)9636-7827	(61)8132-1530	02/09/2015	10:14	357
(31)8467-9923	(61)9934-1023	02/09/2015	15:52	558
(31)8467-9923	(61)9619-9651	02/09/2015	18:46	326
(31)8467-9923	(61)9799-8985	02/09/2015	16:54	165
(51)9658-4361	(61)9615-0261	03/09/2015	13:44	170
(51)9658-4361	(61)8379-1871	03/09/2015	10:32	296

(61)9615-0261	(61)9970-4491	03/09/2015	13:37	175
(51)9658-4361	(61)9539-7580	03/09/2015	15:11	227
(51)9658-4361	(61)9970-4491	03/09/2015	12:27	454
(61)9934-1023	(31)8696-5241	03/09/2015	15:16	36
(61)9495-4487	(61)8132-1530	03/09/2015	10:29	557
(61)8132-1530	(91)9565-1231	03/09/2015	17:14	121
(61)9495-4487	(61)8506-1381	03/09/2015	12:55	528
(61)9615-0261	(61)8506-1381	03/09/2015	14:52	451
(61)9934-1023	(61)9619-9651	03/09/2015	10:08	74
(61)9495-4487	(31)8696-5241	03/09/2015	12:26	571
(31)8467-9923	(61)9539-7580	04/09/2015	12:47	529
(61)9934-1023	(61)9775-6938	04/09/2015	18:49	336
(61)9934-1023	(61)9970-4491	04/09/2015	11:53	402
(61)9495-4487	(61)9299-5160	04/09/2015	15:09	530
(31)8467-9923	(61)9925-4888	04/09/2015	17:38	261
(61)9495-4487	(51)9658-4361	04/09/2015	18:47	546
(51)9658-4361	(91)9565-1231	05/09/2015	11:45	357
(61)9615-0261	(91)9565-1231	05/09/2015	13:9	385
(61)9615-0261	(61)9925-4888	05/09/2015	16:53	147
(61)8132-1530	(61)8379-1871	05/09/2015	18:31	78
(61)9495-4487	(61)9539-7580	05/09/2015	14:07	324
(61)9934-1023	(31)8467-9923	05/09/2015	10:04	399
(51)9658-4361	(61)9934-1023	05/09/2015	15:55	249
(61)9934-1023	(61)9459-5966	05/09/2015	9:58	398
(31)8467-9923	(31)8696-5241	05/09/2015	9:34	352
(61)9615-0261	(61)9459-5966	05/09/2015	14:45	579
(31)8467-9923	(61)8379-1871	07/09/2015	17:17	71
(61)9495-4487	(61)9619-9651	07/09/2015	11:08	463
(61)9934-1023	(61)8379-1871	07/09/2015	19:45	529
(61)8132-1530	(31)8696-5241	07/09/2015	9:04	276
(61)9495-4487	(61)9970-4491	07/09/2015	13:11	507
(31)8467-9923	(61)9299-5160	07/09/2015	9:22	249
(61)9636-7827	(61)9459-5966	07/09/2015	14:24	592
(51)9658-4361	(61)9775-6938	07/09/2015	17:22	451
(61)9636-7827	(61)9299-5160	07/09/2015	16:31	61
(61)9495-4487	(61)9925-4888	07/09/2015	14:45	48
(31)8467-9923	(61)9615-0261	07/09/2015	9:40	217
(61)8132-1530	(31)8696-5241	07/09/2015	12:42	224
(61)9495-4487	(61)8132-1530	07/09/2015	11:12	500
(61)8132-1530	(61)9299-5160	07/09/2015	19:21	289
(61)9934-1023	(61)9615-0261	07/09/2015	17:01	21
(61)9619-9651	(31)8467-9923	08/09/2015	19:20	78

(61)8379-1871	(61)8132-1530	08/09/2015	11:52	516
(61)8977-2201	(91)9565-1231	08/09/2015	9:18	503
(91)9565-1231	(61)9636-7827	08/09/2015	12:53	75
(61)9495-4487	(51)9658-4361	08/09/2015	14:10	391
(61)8977-2201	(61)9934-1023	08/09/2015	14:16	245
(61)9799-8985	(61)9619-9651	09/09/2015	9:44	381
(31)8467-9923	(61)9539-7580	09/09/2015	19:00	263
(61)9925-4888	(61)8506-1381	09/09/2015	9:07	85
(61)9636-7827	(31)8467-9923	09/09/2015	12:56	346
(61)8379-1871	(61)9636-7827	09/09/2015	13:56	472
(61)8379-1871	(61)9925-4888	09/09/2015	18:55	339
(61)9636-7827	(61)9495-4487	09/09/2015	14:57	307
(61)9539-7580	(31)8696-5241	10/09/2015	17:27	575
(31)8696-5241	(61)9299-5160	11/09/2015	17:33	176
(31)8696-5241	(61)9775-6938	11/09/2015	15:24	588
(31)8696-5241	(61)9775-6938	11/09/2015	15:45	120
(61)9636-7827	(61)9925-4888	11/09/2015	18:36	101
(61)9459-5966	(61)9539-7580	11/09/2015	17:22	375
(61)9619-9651	(61)9775-6938	11/09/2015	11:20	236
(61)9299-5160	(61)9495-4487	12/09/2015	10:08	409
(61)9459-5966	(61)9970-4491	12/09/2015	14:13	415
(91)9565-1231	(31)8696-5241	12/09/2015	10:08	581
(61)8977-2201	(61)9299-5160	12/09/2015	19:29	149
(61)9539-7580	(51)9658-4361	12/09/2015	18:14	136
(61)9925-4888	(61)9775-6938	12/09/2015	10:05	396
(61)8506-1381	(61)8977-2201	12/09/2015	14:38	81
(61)8977-2201	(61)9970-4491	12/09/2015	11:24	233
(61)8132-1530	(61)9495-4487	12/09/2015	17:47	157
(61)9799-8985	(61)9539-7580	12/09/2015	10:42	593
(61)9495-4487	(51)9658-4361	12/09/2015	17:44	224
(31)8696-5241	(61)8506-1381	13/09/2015	15:28	159
(61)8132-1530	(61)9970-4491	13/09/2015	9:41	181
(31)8467-9923	(61)9636-7827	13/09/2015	19:48	193
(61)9636-7827	(91)9565-1231	13/09/2015	12:13	319
(61)9619-9651	(91)9565-1231	13/09/2015	11:08	288
(31)8696-5241	(61)9775-6938	13/09/2015	9:45	354
(61)8977-2201	(61)9459-5966	14/09/2015	15:47	158
(61)9615-0261	(61)9934-1023	14/09/2015	9:52	497
(61)9539-7580	(61)9799-8985	14/09/2015	15:24	412
(61)9934-1023	(61)9619-9651	14/09/2015	13:35	64
(61)9539-7580	(61)9495-4487	14/09/2015	12:31	347
(61)9799-8985	(61)9539-7580	15/09/2015	11:48	368

(61)9799-8985	(61)9299-5160	17/09/2015	10:44	572
(61)8132-1530	(61)9495-4487	17/09/2015	19:37	463
(61)9459-5966	(61)9615-0261	17/09/2015	9:24	272
(31)8467-9923	(61)9799-8985	17/09/2015	11:27	393
(61)9299-5160	(61)9459-5966	17/09/2015	15:11	552
(61)9636-7827	(91)9565-1231	17/09/2015	19:16	417
(51)9658-4361	(61)9615-0261	17/09/2015	17:30	216
(61)8132-1530	(61)9615-0261	17/09/2015	12:02	280
(61)8132-1530	(31)8696-5241	17/09/2015	12:47	220
(61)9636-7827	(61)9299-5160	17/09/2015	13:34	406
(61)8506-1381	(61)9636-7827	18/09/2015	18:57	171
(51)9658-4361	(61)9970-4491	18/09/2015	11:55	573
(61)9934-1023	(61)9799-8985	18/09/2015	14:05	237
(61)9636-7827	(61)9619-9651	18/09/2015	15:58	338
(61)9299-5160	(61)9539-7580	18/09/2015	9:24	307
(61)8132-1530	(31)8467-9923	18/09/2015	9:10	337
(61)9799-8985	(61)9539-7580	18/09/2015	11:00	390
(61)9299-5160	(61)9495-4487	18/09/2015	10:48	513
(61)9925-4888	(51)9658-4361	18/09/2015	14:44	251
(61)9970-4491	(61)9799-8985	19/09/2015	14:26	505
(51)9658-4361	(91)9565-1231	19/09/2015	16:42	166
(61)9925-4888	(61)8977-2201	19/09/2015	11:49	469
(61)9495-4487	(61)9799-8985	20/09/2015	22:40	451
(51)9658-4361	(61)9636-7827	20/09/2015	13:02	449
(61)9970-4491	(61)9775-6938	20/09/2015	15:53	76
(61)9299-5160	(61)9636-7827	20/09/2015	16:52	487
(61)8506-1381	(61)9970-4491	20/09/2015	13:36	364
(61)9970-4491	(61)9539-7580	21/09/2015	0:11	297
(91)9565-1231	(61)8506-1381	21/09/2015	15:01	268
(61)9925-4888	(61)9495-4487	21/09/2015	0:47	241
(31)8696-5241	(61)9925-4888	21/09/2015	20:21	226
(61)9299-5160	(61)9934-1023	21/09/2015	3:23	557
(61)8977-2201	(61)8379-1871	22/09/2015	7:34	581
(91)9565-1231	(61)9775-6938	22/09/2015	4:20	427
(61)9775-6938	(61)8132-1530	22/09/2015	13:20	546
(91)9565-1231	(61)9775-6938	22/09/2015	21:51	115
(61)9619-9651	(31)8696-5241	22/09/2015	20:54	467
(31)8696-5241	(31)8467-9923	22/09/2015	9:00	173
(61)9299-5160	(61)9934-1023	23/09/2015	12:18	152
(61)9459-5966	(61)9299-5160	23/09/2015	6:56	595
(61)8132-1530	(91)9565-1231	23/09/2015	3:54	547
(31)8696-5241	(61)9299-5160	25/09/2015	18:44	525

(61)9970-4491	(61)8506-1381	25/09/2015	0:17	222
(61)9934-1023	(61)9619-9651	25/09/2015	3:00	135
(61)9619-9651	(51)9658-4361	25/09/2015	20:07	89
(61)9775-6938	(61)9619-9651	25/09/2015	18:06	73
(31)8467-9923	(61)8132-1530	25/09/2015	22:37	216
(51)9658-4361	(61)8132-1530	26/09/2015	5:14	553
(61)8506-1381	(31)8696-5241	26/09/2015	9:39	496
(61)8132-1530	(61)9619-9651	26/09/2015	17:07	424
(61)9775-6938	(61)9970-4491	26/09/2015	5:13	94
(31)8467-9923	(61)9775-6938	26/09/2015	11:00	298
(61)9775-6938	(51)9658-4361	27/09/2015	23:33	567
(61)9539-7580	(61)9459-5966	27/09/2015	2:36	242
(61)9775-6938	(61)9934-1023	27/09/2015	15:26	109
(61)9799-8985	(61)8379-1871	27/09/2015	21:11	445
(61)9539-7580	(61)9459-5966	28/09/2015	1:47	192
(61)9799-8985	(61)8506-1381	30/09/2015	2:12	268
(61)9775-6938	(61)9495-4487	01/10/2015	9:54	547
(61)9615-0261	(91)9565-1231	01/10/2015	7:51	172
(61)9615-0261	(61)9970-4491	01/10/2015	15:46	279
(61)9775-6938	(61)8977-2201	01/10/2015	18:43	116
(91)9565-1231	(61)9459-5966	01/10/2015	22:49	429
(31)8467-9923	(61)8506-1381	01/10/2015	14:48	378
(31)8696-5241	(61)9970-4491	01/10/2015	17:08	358
(91)9565-1231	(61)8977-2201	02/10/2015	10:37	520
(51)9658-4361	(61)8977-2201	02/10/2015	13:46	514
(61)9925-4888	(61)9615-0261	02/10/2015	13:45	125
(61)9799-8985	(61)9934-1023	02/10/2015	22:45	486
(61)9299-5160	(61)8379-1871	02/10/2015	0:40	310
(61)9615-0261	(61)9934-1023	03/10/2015	12:34	30
(31)8467-9923	(61)9619-9651	03/10/2015	7:57	216
(61)9495-4487	(51)9658-4361	03/10/2015	10:34	232
(61)9539-7580	(61)9799-8985	03/10/2015	4:55	206
(61)9636-7827	(31)8696-5241	03/10/2015	9:07	176
(61)8379-1871	(61)9539-7580	04/10/2015	5:33	576
(61)9459-5966	(61)8379-1871	04/10/2015	19:19	306
(31)8467-9923	(61)8379-1871	04/10/2015	23:58	57
(61)9799-8985	(61)9615-0261	04/10/2015	22:42	88
(61)8977-2201	(61)9615-0261	05/10/2015	10:24	493
(61)9539-7580	(91)9565-1231	05/10/2015	17:29	576
(61)8379-1871	(61)8132-1530	05/10/2015	21:26	50
(61)9615-0261	(61)9539-7580	06/09/2016	10:31	574
(61)8132-1530	(61)9495-4487	06/09/2016	17:42	570

(61)9934-1023	(61)9299-5160	06/09/2016	15:11	117
---------------	---------------	------------	-------	-----

**B – DADOS DE MOVIMENTAÇÕES BANCÁRIAS DA OPERAÇÃO EXEMPLO**

<b>C/C Origem</b>	<b>C/C Destino</b>	<b>Data</b>	<b>Valor</b>
514503-0	887965-3	24/08/2015	R\$ 150.000,00
591565-3	591807-6	08/09/2015	R\$ 150.000,00
317870-7	461039-2	24/08/2015	R\$ 790,00
317870-7	90088-5	24/08/2015	R\$ 980,00
317870-7	339463-9	24/08/2015	R\$ 2.500,00
134624-8	921758-2	24/08/2015	R\$ 3.600,00
887965-3	514503-0	24/08/2015	R\$ 80.000,00
104918-2	46826-3	24/08/2015	R\$ 6.800,00
104918-2	887965-3	24/08/2015	R\$ 35.000,00
921758-2	567186-5	24/08/2015	R\$ 540,00
527895-5	943716-7	24/08/2015	R\$ 310.000,00
232673-4	205086-5	24/08/2015	R\$ 990,00
46826-3	414401-2	24/08/2015	R\$ 3.200,00
461039-2	134624-8	24/08/2015	R\$ 760,00
514503-0	69873-3	25/08/2015	R\$ 1.000,00
90088-5	527895-5	25/08/2015	R\$ 500,00
104918-2	943716-7	25/08/2015	R\$ 3.900,00
414401-2	46826-3	25/08/2015	R\$ 4.000,00
527895-5	134624-8	25/08/2015	R\$ 6.600,00
582276-2	414401-2	25/08/2015	R\$ 1.600,00
582276-2	90088-5	25/08/2015	R\$ 27.000,00
582276-2	609444-4	25/08/2015	R\$ 8.800,00
442126-2	956010-7	25/08/2015	R\$ 450.000,00
514503-0	921758-2	26/08/2015	R\$ 9.500,00
591565-3	514503-0	26/08/2015	R\$ 210.000,00
205086-5	943716-7	26/08/2015	R\$ 39.000,00
205086-5	514503-0	26/08/2015	R\$ 6.700,00
317870-7	527895-5	26/08/2015	R\$ 52.000,00
339463-9	591565-3	26/08/2015	R\$ 22.000,00
887965-3	69873-3	26/08/2015	R\$ 7.000,00
423923-9	348247-1	26/08/2015	R\$ 8.300,00
423923-9	414401-2	26/08/2015	R\$ 630.000,00
104918-2	609444-4	26/08/2015	R\$ 2.300,00
921758-2	232673-4	26/08/2015	R\$ 160.000,00
591807-6	567186-5	26/08/2015	R\$ 2.600,00
442126-2	205086-5	26/08/2015	R\$ 83.000,00
591565-3	442126-2	27/08/2015	R\$ 32.000,00
591565-3	514503-0	27/08/2015	R\$ 97.000,00
205086-5	943716-7	27/08/2015	R\$ 3.600,00

317870-7	609444-4	27/08/2015	R\$ 9.600,00
134624-8	232673-4	27/08/2015	R\$ 7.200,00
609444-4	921758-2	27/08/2015	R\$ 1.000,00
527895-5	591565-3	27/08/2015	R\$ 720,00
591807-6	956010-7	27/08/2015	R\$ 8.500,00
98289-6	898329-3	27/08/2015	R\$ 3.300,00
232673-4	527895-5	27/08/2015	R\$ 780,00
582276-2	898329-3	27/08/2015	R\$ 8.000,00
582276-2	69873-3	27/08/2015	R\$ 39.000,00
567186-5	46826-3	27/08/2015	R\$ 6.600,00
348247-1	232673-4	27/08/2015	R\$ 350,00
898329-3	591565-3	27/08/2015	R\$ 8.100,00
514503-0	527895-5	28/08/2015	R\$ 58.000,00
591565-3	609444-4	28/08/2015	R\$ 7.400,00
205086-5	442126-2	28/08/2015	R\$ 35.000,00
205086-5	943716-7	28/08/2015	R\$ 950,00
339463-9	898329-3	28/08/2015	R\$ 670.000,00
339463-9	348247-1	28/08/2015	R\$ 85.000,00
956010-7	348247-1	28/08/2015	R\$ 840,00
956010-7	514503-0	28/08/2015	R\$ 4.300,00
46826-3	98289-6	28/08/2015	R\$ 7.200,00
567186-5	921758-2	28/08/2015	R\$ 860.000,00
567186-5	90088-5	28/08/2015	R\$ 4.900,00
348247-1	423923-9	28/08/2015	R\$ 6.200,00
69873-3	582276-2	28/08/2015	R\$ 85.000,00
514503-0	442126-2	29/08/2015	R\$ 600,00
317870-7	414401-2	29/08/2015	R\$ 250,00
317870-7	90088-5	29/08/2015	R\$ 98.000,00
339463-9	90088-5	29/08/2015	R\$ 520,00
887965-3	567186-5	29/08/2015	R\$ 26.000,00
90088-5	414401-2	29/08/2015	R\$ 8.300,00
104918-2	582276-2	29/08/2015	R\$ 1.200,00
527895-5	104918-2	29/08/2015	R\$ 1.400,00
582276-2	414401-2	29/08/2015	R\$ 96.000,00
461039-2	339463-9	29/08/2015	R\$ 65.000,00
69873-3	104918-2	29/08/2015	R\$ 3.700,00
591565-3	134624-8	30/08/2015	R\$ 100.000,00
317870-7	205086-5	30/08/2015	R\$ 84.000,00
134624-8	591565-3	30/08/2015	R\$ 77.000,00
887965-3	104918-2	30/08/2015	R\$ 4.800,00
887965-3	609444-4	30/08/2015	R\$ 32.000,00
90088-5	414401-2	30/08/2015	R\$ 6.000,00

943716-7	527895-5	30/08/2015	R\$ 390.000,00
414401-2	567186-5	30/08/2015	R\$ 9.400,00
414401-2	921758-2	30/08/2015	R\$ 5.300,00
582276-2	423923-9	30/08/2015	R\$ 44.000,00
442126-2	591565-3	30/08/2015	R\$ 1.200,00
514503-0	134624-8	31/08/2015	R\$ 440.000,00
591565-3	90088-5	31/08/2015	R\$ 82.000,00
591565-3	887965-3	31/08/2015	R\$ 4.700,00
591565-3	339463-9	31/08/2015	R\$ 57.000,00
205086-5	527895-5	31/08/2015	R\$ 7.800,00
205086-5	591565-3	31/08/2015	R\$ 630,00
609444-4	423923-9	31/08/2015	R\$ 16.000,00
90088-5	591565-3	31/08/2015	R\$ 3.700,00
104918-2	921758-2	31/08/2015	R\$ 6.800,00
591807-6	461039-2	31/08/2015	R\$ 530.000,00
98289-6	609444-4	31/08/2015	R\$ 1.000,00
46826-3	339463-9	31/08/2015	R\$ 210.000,00
461039-2	317870-7	31/08/2015	R\$ 6.700,00
591565-3	582276-2	01/09/2015	R\$ 500.000,00
205086-5	98289-6	01/09/2015	R\$ 6.200,00
205086-5	591565-3	01/09/2015	R\$ 450.000,00
317870-7	514503-0	01/09/2015	R\$ 7.200,00
90088-5	567186-5	01/09/2015	R\$ 94.000,00
921758-2	898329-3	01/09/2015	R\$ 4.400,00
98289-6	591807-6	01/09/2015	R\$ 390.000,00
46826-3	69873-3	01/09/2015	R\$ 990,00
582276-2	514503-0	01/09/2015	R\$ 6.100,00
442126-2	46826-3	01/09/2015	R\$ 640,00
461039-2	339463-9	01/09/2015	R\$ 9.000,00
348247-1	943716-7	01/09/2015	R\$ 8.400,00
317870-7	46826-3	02/09/2015	R\$ 3.700,00
609444-4	461039-2	02/09/2015	R\$ 69.000,00
90088-5	461039-2	02/09/2015	R\$ 200.000,00
943716-7	69873-3	02/09/2015	R\$ 930,00
956010-7	442126-2	02/09/2015	R\$ 640,00
921758-2	567186-5	02/09/2015	R\$ 35.000,00
414401-2	591565-3	02/09/2015	R\$ 350,00
527895-5	134624-8	02/09/2015	R\$ 880,00
98289-6	943716-7	02/09/2015	R\$ 920,00
609444-4	887965-3	02/09/2015	R\$ 900.000,00
582276-2	134624-8	02/09/2015	R\$ 2.800,00
514503-0	90088-5	03/09/2015	R\$ 1.100,00

317870-7	423923-9	03/09/2015	R\$ 140,00
609444-4	887965-3	03/09/2015	R\$ 58.000,00
414401-2	442126-2	03/09/2015	R\$ 280.000,00
232673-4	414401-2	03/09/2015	R\$ 79.000,00
232673-4	609444-4	03/09/2015	R\$ 920,00
46826-3	527895-5	03/09/2015	R\$ 10.000,00
46826-3	134624-8	03/09/2015	R\$ 500,00
567186-5	461039-2	03/09/2015	R\$ 11.000,00
567186-5	582276-2	03/09/2015	R\$ 67.000,00
69873-3	104918-2	03/09/2015	R\$ 9.700,00
514503-0	423923-9	04/09/2015	R\$ 70.000,00
591565-3	582276-2	04/09/2015	R\$ 8.700,00
205086-5	232673-4	04/09/2015	R\$ 1.100,00
205086-5	527895-5	04/09/2015	R\$ 1.700,00
609444-4	921758-2	04/09/2015	R\$ 61.000,00
609444-4	134624-8	04/09/2015	R\$ 10.000,00
104918-2	514503-0	04/09/2015	R\$ 5.000,00
921758-2	943716-7	04/09/2015	R\$ 730.000,00
414401-2	567186-5	04/09/2015	R\$ 3.300,00
442126-2	582276-2	04/09/2015	R\$ 5.700,00
442126-2	943716-7	04/09/2015	R\$ 370,00
348247-1	898329-3	04/09/2015	R\$ 1.200,00
69873-3	46826-3	04/09/2015	R\$ 540.000,00
69873-3	591565-3	04/09/2015	R\$ 970,00
69873-3	514503-0	04/09/2015	R\$ 410,00
514503-0	898329-3	05/09/2015	R\$ 840.000,00
591565-3	921758-2	05/09/2015	R\$ 5.300,00
591565-3	887965-3	05/09/2015	R\$ 60.000,00
609444-4	134624-8	05/09/2015	R\$ 67.000,00
90088-5	98289-6	05/09/2015	R\$ 390.000,00
956010-7	423923-9	05/09/2015	R\$ 88.000,00
921758-2	887965-3	05/09/2015	R\$ 350,00
921758-2	514503-0	05/09/2015	R\$ 11.000,00
527895-5	887965-3	05/09/2015	R\$ 310,00
591807-6	317870-7	05/09/2015	R\$ 86.000,00
98289-6	414401-2	05/09/2015	R\$ 540.000,00
232673-4	527895-5	05/09/2015	R\$ 32.000,00
46826-3	232673-4	05/09/2015	R\$ 400.000,00
442126-2	582276-2	05/09/2015	R\$ 79.000,00
461039-2	898329-3	05/09/2015	R\$ 950,00
514503-0	90088-5	06/09/2015	R\$ 6.900,00
205086-5	898329-3	06/09/2015	R\$ 680.000,00

423923-9	442126-2	06/09/2015	R\$ 940,00
423923-9	514503-0	06/09/2015	R\$ 930,00
921758-2	956010-7	06/09/2015	R\$ 1.600,00
527895-5	317870-7	06/09/2015	R\$ 10.000,00
591807-6	461039-2	06/09/2015	R\$ 94.000,00
98289-6	348247-1	06/09/2015	R\$ 860.000,00
442126-2	591807-6	06/09/2015	R\$ 4.600,00
69873-3	461039-2	06/09/2015	R\$ 210,00
69873-3	921758-2	06/09/2015	R\$ 3.000,00
591565-3	943716-7	07/09/2015	R\$ 890,00
591565-3	134624-8	07/09/2015	R\$ 1.000.000,00
887965-3	527895-5	07/09/2015	R\$ 5.300,00
609444-4	943716-7	07/09/2015	R\$ 7.100,00
104918-2	134624-8	07/09/2015	R\$ 10.000,00
921758-2	591565-3	03/09/2015	R\$ 150.000,00
98289-6	591807-6	07/09/2015	R\$ 900,00
98289-6	104918-2	07/09/2015	R\$ 68.000,00
582276-2	339463-9	07/09/2015	R\$ 8.500,00
442126-2	943716-7	07/09/2015	R\$ 7.200,00
442126-2	317870-7	07/09/2015	R\$ 170.000,00
69873-3	104918-2	07/09/2015	R\$ 88.000,00
317870-7	414401-2	08/09/2015	R\$ 10.000,00
134624-8	582276-2	08/09/2015	R\$ 74.000,00
134624-8	414401-2	08/09/2015	R\$ 480.000,00
609444-4	69873-3	08/09/2015	R\$ 13.000,00
423923-9	591807-6	08/09/2015	R\$ 680.000,00
943716-7	591807-6	08/09/2015	R\$ 520,00
956010-7	423923-9	08/09/2015	R\$ 7.100,00
527895-5	98289-6	08/09/2015	R\$ 4.300,00
232673-4	921758-2	08/09/2015	R\$ 6.800,00
567186-5	609444-4	08/09/2015	R\$ 770,00
348247-1	98289-6	08/09/2015	R\$ 870.000,00
591565-3	69873-3	09/09/2015	R\$ 7.400,00
591565-3	348247-1	09/09/2015	R\$ 840,00
205086-5	609444-4	09/09/2015	R\$ 130,00
339463-9	887965-3	09/09/2015	R\$ 26.000,00
134624-8	414401-2	09/09/2015	R\$ 540,00
887965-3	339463-9	09/09/2015	R\$ 7.600,00
609444-4	591807-6	09/09/2015	R\$ 2.700,00
423923-9	90088-5	09/09/2015	R\$ 9.900,00
98289-6	232673-4	09/09/2015	R\$ 1.200,00
232673-4	317870-7	09/09/2015	R\$ 940,00

461039-2	134624-8	09/09/2015	R\$ 66.000,00
69873-3	205086-5	09/09/2015	R\$ 75.000,00
591565-3	609444-4	10/09/2015	R\$ 150,00
591565-3	609444-4	10/09/2015	R\$ 210.000,00
591565-3	317870-7	10/09/2015	R\$ 7.900,00
205086-5	514503-0	10/09/2015	R\$ 1.300,00
90088-5	414401-2	10/09/2015	R\$ 770,00
956010-7	898329-3	10/09/2015	R\$ 6.800,00
104918-2	591807-6	10/09/2015	R\$ 94.000,00
921758-2	461039-2	10/09/2015	R\$ 6.100,00
921758-2	582276-2	10/09/2015	R\$ 23.000,00
527895-5	514503-0	10/09/2015	R\$ 1.300,00
98289-6	339463-9	10/09/2015	R\$ 44.000,00
232673-4	887965-3	10/09/2015	R\$ 720.000,00
46826-3	104918-2	10/09/2015	R\$ 280,00
46826-3	317870-7	10/09/2015	R\$ 8.100,00
69873-3	956010-7	10/09/2015	R\$ 100.000,00
514503-0	69873-3	11/09/2015	R\$ 1.100,00
514503-0	527895-5	11/09/2015	R\$ 670,00
317870-7	609444-4	11/09/2015	R\$ 60.000,00
339463-9	921758-2	11/09/2015	R\$ 79.000,00
339463-9	514503-0	11/09/2015	R\$ 5.700,00
134624-8	69873-3	11/09/2015	R\$ 77.000,00
423923-9	348247-1	11/09/2015	R\$ 100.000,00
956010-7	317870-7	11/09/2015	R\$ 2.100,00
414401-2	591565-3	11/09/2015	R\$ 460,00
46826-3	527895-5	11/09/2015	R\$ 6.900,00
442126-2	205086-5	11/09/2015	R\$ 3.000,00
69873-3	339463-9	11/09/2015	R\$ 8.700,00
591565-3	104918-2	12/09/2015	R\$ 1.100,00
205086-5	461039-2	12/09/2015	R\$ 7.500,00
317870-7	887965-3	12/09/2015	R\$ 4.800,00
134624-8	232673-4	12/09/2015	R\$ 310.000,00
921758-2	582276-2	12/09/2015	R\$ 1.100,00
921758-2	591807-6	12/09/2015	R\$ 8.700,00
527895-5	104918-2	12/09/2015	R\$ 6.300,00
582276-2	205086-5	12/09/2015	R\$ 540,00
442126-2	317870-7	12/09/2015	R\$ 100.000,00
69873-3	348247-1	12/09/2015	R\$ 6.400,00
317870-7	134624-8	13/09/2015	R\$ 300.000,00
887965-3	134624-8	13/09/2015	R\$ 7.100,00
956010-7	348247-1	13/09/2015	R\$ 61.000,00

104918-2	567186-5	13/09/2015	R\$ 650,00
921758-2	339463-9	13/09/2015	R\$ 190,00
232673-4	69873-3	13/09/2015	R\$ 25.000,00
232673-4	591565-3	13/09/2015	R\$ 6.700,00
46826-3	582276-2	13/09/2015	R\$ 650.000,00
46826-3	591807-6	13/09/2015	R\$ 9.800,00
582276-2	921758-2	13/09/2015	R\$ 12.000,00
461039-2	898329-3	13/09/2015	R\$ 9.100,00
514503-0	591565-3	14/09/2015	R\$ 4.300,00
205086-5	943716-7	14/09/2015	R\$ 390.000,00
317870-7	514503-0	14/09/2015	R\$ 2.000,00
339463-9	527895-5	14/09/2015	R\$ 980,00
609444-4	591807-6	14/09/2015	R\$ 66.000,00
921758-2	442126-2	14/09/2015	R\$ 450,00
591807-6	98289-6	14/09/2015	R\$ 89.000,00
442126-2	134624-8	14/09/2015	R\$ 8.600,00
591565-3	609444-4	15/09/2015	R\$ 860.000,00
134624-8	414401-2	15/09/2015	R\$ 600,00
921758-2	348247-1	15/09/2015	R\$ 280,00
921758-2	348247-1	15/09/2015	R\$ 35.000,00
527895-5	348247-1	15/09/2015	R\$ 8.400,00
442126-2	514503-0	15/09/2015	R\$ 540,00
898329-3	956010-7	15/09/2015	R\$ 8.700,00
591565-3	514503-0	16/09/2015	R\$ 5.400,00
205086-5	46826-3	16/09/2015	R\$ 5.300,00
317870-7	591807-6	16/09/2015	R\$ 340.000,00
339463-9	898329-3	16/09/2015	R\$ 140,00
339463-9	582276-2	16/09/2015	R\$ 290.000,00
134624-8	527895-5	16/09/2015	R\$ 2.800,00
887965-3	414401-2	16/09/2015	R\$ 62.000,00
423923-9	527895-5	16/09/2015	R\$ 420.000,00
423923-9	514503-0	16/09/2015	R\$ 17.000,00
956010-7	46826-3	16/09/2015	R\$ 930,00
414401-2	339463-9	16/09/2015	R\$ 6.000,00
527895-5	582276-2	16/09/2015	R\$ 9.800,00
232673-4	514503-0	16/09/2015	R\$ 1.600,00
205086-5	423923-9	17/09/2015	R\$ 94.000,00
943716-7	317870-7	17/09/2015	R\$ 490,00
104918-2	887965-3	17/09/2015	R\$ 88.000,00
591807-6	514503-0	13/09/2015	R\$ 150.000,00
98289-6	414401-2	17/09/2015	R\$ 330.000,00
582276-2	527895-5	17/09/2015	R\$ 5.600,00

582276-2	205086-5	17/09/2015	R\$ 770.000,00
442126-2	423923-9	17/09/2015	R\$ 660.000,00
348247-1	921758-2	17/09/2015	R\$ 9.500,00
348247-1	956010-7	17/09/2015	R\$ 640.000,00
514503-0	943716-7	18/09/2015	R\$ 480,00
591565-3	461039-2	18/09/2015	R\$ 63.000,00
205086-5	527895-5	18/09/2015	R\$ 3.600,00
339463-9	582276-2	18/09/2015	R\$ 95.000,00
90088-5	134624-8	18/09/2015	R\$ 790,00
423923-9	98289-6	18/09/2015	R\$ 4.000,00
423923-9	943716-7	18/09/2015	R\$ 9.000,00
943716-7	609444-4	18/09/2015	R\$ 530.000,00
956010-7	461039-2	18/09/2015	R\$ 4.500,00
104918-2	956010-7	18/09/2015	R\$ 810,00
582276-2	921758-2	18/09/2015	R\$ 88.000,00
514503-0	414401-2	19/09/2015	R\$ 770,00
205086-5	442126-2	19/09/2015	R\$ 37.000,00
134624-8	898329-3	19/09/2015	R\$ 550,00
134624-8	591807-6	19/09/2015	R\$ 2.900,00
134624-8	887965-3	19/09/2015	R\$ 75.000,00
423923-9	104918-2	19/09/2015	R\$ 8.600,00
104918-2	943716-7	19/09/2015	R\$ 9.700,00
921758-2	232673-4	19/09/2015	R\$ 6.200,00
414401-2	514503-0	19/09/2015	R\$ 870.000,00
46826-3	339463-9	19/09/2015	R\$ 410.000,00
582276-2	514503-0	19/09/2015	R\$ 10.000,00
567186-5	205086-5	19/09/2015	R\$ 11.000,00
461039-2	98289-6	19/09/2015	R\$ 5.200,00
591565-3	943716-7	20/09/2015	R\$ 5.500,00
339463-9	104918-2	20/09/2015	R\$ 3.600,00
887965-3	90088-5	20/09/2015	R\$ 4.300,00
887965-3	205086-5	20/09/2015	R\$ 240.000,00
609444-4	898329-3	20/09/2015	R\$ 52.000,00
423923-9	442126-2	20/09/2015	R\$ 4.200,00
104918-2	461039-2	20/09/2015	R\$ 800.000,00
104918-2	205086-5	20/09/2015	R\$ 510,00
414401-2	317870-7	20/09/2015	R\$ 25.000,00
591807-6	527895-5	20/09/2015	R\$ 980,00
98289-6	921758-2	20/09/2015	R\$ 200.000,00
98289-6	104918-2	20/09/2015	R\$ 400,00
582276-2	414401-2	20/09/2015	R\$ 550,00
514503-0	943716-7	21/09/2015	R\$ 80.000,00

591565-3	461039-2	21/09/2015	R\$ 72.000,00
591565-3	339463-9	21/09/2015	R\$ 3.900,00
317870-7	69873-3	21/09/2015	R\$ 19.000,00
317870-7	514503-0	21/09/2015	R\$ 310,00
134624-8	98289-6	21/09/2015	R\$ 1.000,00
921758-2	442126-2	21/09/2015	R\$ 45.000,00
527895-5	205086-5	21/09/2015	R\$ 1.700,00
591807-6	527895-5	21/09/2015	R\$ 94.000,00
232673-4	134624-8	21/09/2015	R\$ 4.200,00
514503-0	339463-9	22/09/2015	R\$ 28.000,00
205086-5	442126-2	22/09/2015	R\$ 76.000,00
317870-7	461039-2	22/09/2015	R\$ 20.000,00
887965-3	339463-9	22/09/2015	R\$ 8.300,00
956010-7	232673-4	22/09/2015	R\$ 4.800,00
104918-2	527895-5	22/09/2015	R\$ 380.000,00
414401-2	348247-1	22/09/2015	R\$ 85.000,00
46826-3	591807-6	22/09/2015	R\$ 9.900,00
591565-3	46826-3	23/09/2015	R\$ 53.000,00
591565-3	90088-5	23/09/2015	R\$ 10.000,00
90088-5	898329-3	23/09/2015	R\$ 7.400,00
943716-7	46826-3	23/09/2015	R\$ 220,00
943716-7	609444-4	23/09/2015	R\$ 560.000,00
414401-2	514503-0	23/09/2015	R\$ 210.000,00
591807-6	461039-2	23/09/2015	R\$ 3.200,00
582276-2	887965-3	23/09/2015	R\$ 19.000,00
514503-0	567186-5	24/09/2015	R\$ 470,00
514503-0	591565-3	24/09/2015	R\$ 9.700,00
205086-5	104918-2	24/09/2015	R\$ 660,00
134624-8	591565-3	24/09/2015	R\$ 3.000,00
887965-3	956010-7	24/09/2015	R\$ 7.400,00
956010-7	104918-2	24/09/2015	R\$ 350,00
921758-2	527895-5	24/09/2015	R\$ 15.000,00
414401-2	582276-2	24/09/2015	R\$ 510.000,00
527895-5	943716-7	24/09/2015	R\$ 190,00
98289-6	46826-3	24/09/2015	R\$ 1.900,00
582276-2	104918-2	24/09/2015	R\$ 180,00
442126-2	69873-3	24/09/2015	R\$ 930,00
348247-1	514503-0	24/09/2015	R\$ 3.300,00
69873-3	442126-2	24/09/2015	R\$ 84.000,00
898329-3	956010-7	24/09/2015	R\$ 7.300,00
591565-3	887965-3	25/09/2015	R\$ 450.000,00
339463-9	898329-3	25/09/2015	R\$ 5.700,00

609444-4	567186-5	25/09/2015	R\$ 25.000,00
943716-7	898329-3	25/09/2015	R\$ 240.000,00
943716-7	527895-5	25/09/2015	R\$ 27.000,00
104918-2	90088-5	25/09/2015	R\$ 880.000,00
104918-2	317870-7	25/09/2015	R\$ 7.700,00
921758-2	46826-3	25/09/2015	R\$ 2.200,00
527895-5	591565-3	25/09/2015	R\$ 86.000,00
46826-3	104918-2	25/09/2015	R\$ 3.300,00
442126-2	98289-6	25/09/2015	R\$ 230,00
442126-2	943716-7	25/09/2015	R\$ 68.000,00
567186-5	609444-4	25/09/2015	R\$ 9.800,00
461039-2	46826-3	25/09/2015	R\$ 390,00
898329-3	591565-3	25/09/2015	R\$ 93.000,00
514503-0	69873-3	26/09/2015	R\$ 8.800,00
317870-7	956010-7	26/09/2015	R\$ 920,00
317870-7	514503-0	26/09/2015	R\$ 82.000,00
887965-3	898329-3	26/09/2015	R\$ 610,00
887965-3	348247-1	26/09/2015	R\$ 5.300,00
609444-4	514503-0	26/09/2015	R\$ 6.100,00
943716-7	609444-4	26/09/2015	R\$ 3.600,00
943716-7	317870-7	26/09/2015	R\$ 57.000,00
104918-2	461039-2	26/09/2015	R\$ 1.700,00
104918-2	317870-7	26/09/2015	R\$ 4.400,00
414401-2	887965-3	26/09/2015	R\$ 990.000,00
591807-6	461039-2	26/09/2015	R\$ 920.000,00
514503-0	567186-5	27/09/2015	R\$ 1.900,00
514503-0	98289-6	27/09/2015	R\$ 1.500,00
591565-3	887965-3	27/09/2015	R\$ 440,00
205086-5	69873-3	27/09/2015	R\$ 1.800,00
339463-9	348247-1	27/09/2015	R\$ 650,00
423923-9	317870-7	27/09/2015	R\$ 250.000,00
921758-2	104918-2	27/09/2015	R\$ 220.000,00
414401-2	898329-3	27/09/2015	R\$ 87.000,00
527895-5	582276-2	27/09/2015	R\$ 8.200,00
232673-4	134624-8	27/09/2015	R\$ 2.600,00
591565-3	348247-1	28/09/2015	R\$ 7.100,00
591565-3	205086-5	28/09/2015	R\$ 4.300,00
591565-3	514503-0	28/09/2015	R\$ 85.000,00
205086-5	956010-7	28/09/2015	R\$ 540.000,00
134624-8	461039-2	28/09/2015	R\$ 940.000,00
90088-5	46826-3	28/09/2015	R\$ 70.000,00
956010-7	442126-2	28/09/2015	R\$ 7.400,00

527895-5	232673-4	28/09/2015	R\$ 230.000,00
591807-6	205086-5	28/09/2015	R\$ 3.000,00
442126-2	591565-3	28/09/2015	R\$ 2.400,00
567186-5	339463-9	28/09/2015	R\$ 690,00
205086-5	567186-5	29/09/2015	R\$ 4.800,00
205086-5	943716-7	29/09/2015	R\$ 5.800,00
339463-9	898329-3	29/09/2015	R\$ 3.600,00
339463-9	46826-3	29/09/2015	R\$ 67.000,00
609444-4	943716-7	29/09/2015	R\$ 990,00
90088-5	317870-7	29/09/2015	R\$ 9.500,00
90088-5	514503-0	29/09/2015	R\$ 6.600,00
956010-7	317870-7	29/09/2015	R\$ 88.000,00
956010-7	514503-0	29/09/2015	R\$ 630,00
921758-2	591565-3	29/09/2015	R\$ 7.500,00
414401-2	348247-1	29/09/2015	R\$ 6.800,00
414401-2	232673-4	29/09/2015	R\$ 930,00
414401-2	232673-4	29/09/2015	R\$ 35.000,00
442126-2	423923-9	29/09/2015	R\$ 75.000,00
442126-2	134624-8	29/09/2015	R\$ 870.000,00
461039-2	317870-7	29/09/2015	R\$ 100,00
348247-1	956010-7	29/09/2015	R\$ 89.000,00
898329-3	348247-1	29/09/2015	R\$ 7.000,00
514503-0	46826-3	30/09/2015	R\$ 5.800,00
317870-7	205086-5	30/09/2015	R\$ 5.700,00
134624-8	567186-5	30/09/2015	R\$ 9.200,00
887965-3	461039-2	30/09/2015	R\$ 98.000,00
956010-7	69873-3	30/09/2015	R\$ 920,00
921758-2	104918-2	30/09/2015	R\$ 590.000,00
414401-2	205086-5	30/09/2015	R\$ 3.000,00
414401-2	514503-0	30/09/2015	R\$ 2.900,00
514503-0	527895-5	01/10/2015	R\$ 1.300,00
514503-0	943716-7	01/10/2015	R\$ 470,00
205086-5	232673-4	01/10/2015	R\$ 1.600,00
317870-7	423923-9	01/10/2015	R\$ 2.800,00
423923-9	582276-2	01/10/2015	R\$ 130,00
943716-7	205086-5	01/10/2015	R\$ 22.000,00
956010-7	348247-1	01/10/2015	R\$ 6.900,00
956010-7	232673-4	01/10/2015	R\$ 49.000,00
527895-5	90088-5	01/10/2015	R\$ 480.000,00
98289-6	232673-4	01/10/2015	R\$ 68.000,00
582276-2	104918-2	01/10/2015	R\$ 5.200,00
461039-2	514503-0	01/10/2015	R\$ 90.000,00

348247-1	898329-3	01/10/2015	R\$ 65.000,00
591565-3	104918-2	02/10/2015	R\$ 690.000,00
205086-5	956010-7	02/10/2015	R\$ 46.000,00
205086-5	423923-9	02/10/2015	R\$ 320,00
205086-5	514503-0	02/10/2015	R\$ 49.000,00
339463-9	887965-3	02/10/2015	R\$ 8.500,00
887965-3	582276-2	02/10/2015	R\$ 7.400,00
943716-7	339463-9	02/10/2015	R\$ 4.000,00
104918-2	567186-5	02/10/2015	R\$ 580,00
348247-1	956010-7	02/10/2015	R\$ 9.500,00
69873-3	134624-8	02/10/2015	R\$ 3.400,00
514503-0	442126-2	03/10/2015	R\$ 160,00
205086-5	582276-2	03/10/2015	R\$ 44.000,00
205086-5	46826-3	03/10/2015	R\$ 46.000,00
205086-5	317870-7	03/10/2015	R\$ 2.400,00
317870-7	567186-5	03/10/2015	R\$ 41.000,00
134624-8	527895-5	03/10/2015	R\$ 82.000,00
887965-3	339463-9	03/10/2015	R\$ 3.500,00
423923-9	527895-5	03/10/2015	R\$ 980,00
943716-7	348247-1	03/10/2015	R\$ 7.800,00
956010-7	205086-5	03/10/2015	R\$ 410.000,00
98289-6	69873-3	03/10/2015	R\$ 1.800,00
582276-2	567186-5	03/10/2015	R\$ 2.300,00
567186-5	527895-5	03/10/2015	R\$ 520.000,00
69873-3	943716-7	03/10/2015	R\$ 8.100,00
69873-3	205086-5	03/10/2015	R\$ 97.000,00
943716-7	205086-5	04/10/2015	R\$ 1.500,00
104918-2	232673-4	04/10/2015	R\$ 4.800,00
104918-2	527895-5	04/10/2015	R\$ 5.700,00
527895-5	414401-2	04/10/2015	R\$ 21.000,00
46826-3	134624-8	04/10/2015	R\$ 7.100,00
567186-5	591807-6	04/10/2015	R\$ 30.000,00
348247-1	317870-7	04/10/2015	R\$ 1.200,00
317870-7	527895-5	05/10/2015	R\$ 780,00
317870-7	104918-2	05/10/2015	R\$ 790.000,00
339463-9	205086-5	05/10/2015	R\$ 540,00
887965-3	921758-2	05/10/2015	R\$ 370.000,00
609444-4	921758-2	30/08/2015	R\$ 150.000,00
956010-7	414401-2	05/10/2015	R\$ 560,00
527895-5	582276-2	05/10/2015	R\$ 850,00
582276-2	609444-4	05/10/2015	R\$ 67.000,00