



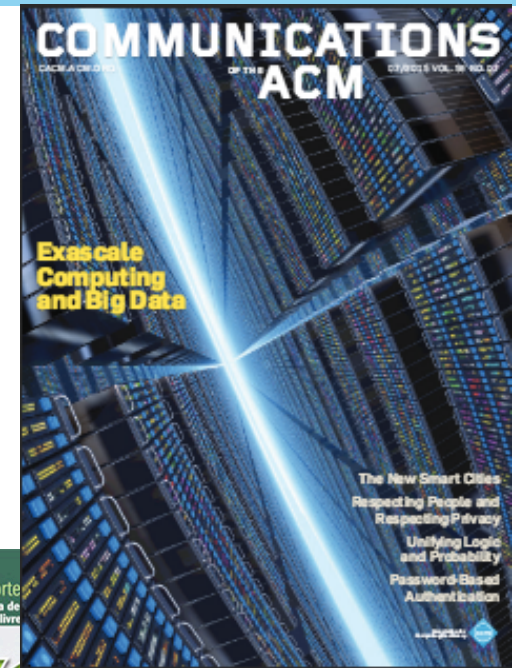
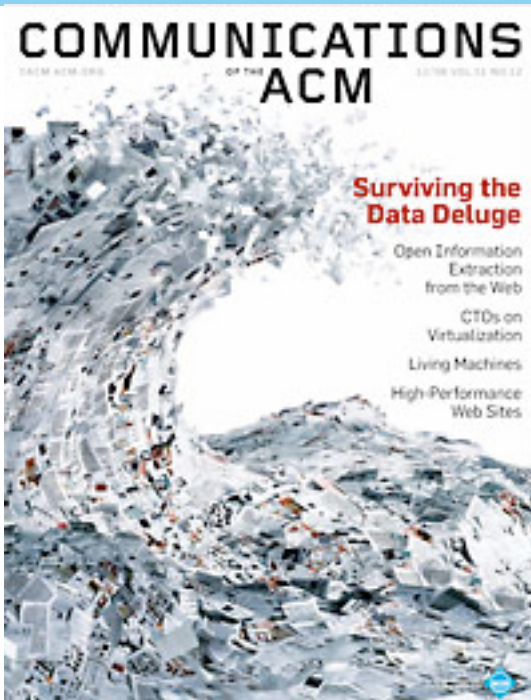
**Centro Federal de Educação Tecnológica  
Celso Suckow da Fonseca  
CEFET/RJ**

## Big Data & IoT

**Eduardo Ogasawara**  
**eogasawara@ieee.org**

**<http://eic.cefet-rj.br/~eogasawara>**

# Data Deluge



# *A Era do Big Data*

- Aumento crescente nos dados produzidos no mundo
- Em 2011 a estimativa era de 1.8ZB  $\approx 10^{21}$ B
- Presente nos eixos
  - Indústria
    - Serviços (Google, Facebook, etc)
    - Finanças
    - *e-commerce*
  - Governo
    - Mobilidade urbana
  - Ciência
    - Levantamentos astronômicos (LST)
    - *e-Science*

# O fenômeno do Big Data

## The Phenomenon of Big Data

**1.8ZB**



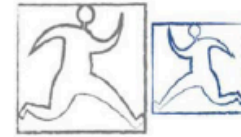
Data generated during 2 days in 2011 (larger than the accumulated amount of data generated from the origin of civilization to 2003)

**750 million**

The amount of pictures uploaded to Facebook



**966PB**



In 2009, the storage capacity of American manufacturing industry

**209 billion**

The number of RFID tags in 2021 (12 million in 2011)



**200+TB**



Data downloaded during a computer geek's 2450 thousand hours

**200PB**

The amount of data generated by a smart urban project in China



**800 billion dollars**



Personal location data in 10 years

**300 billion dollars**

Medical expense saving by big data analysis in America



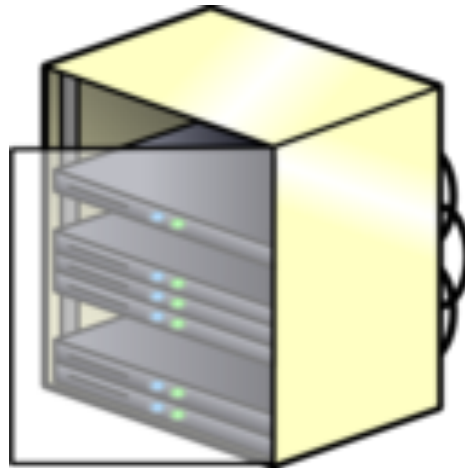
**\$32+B**



The purchase amount of the 4 big companies since 2010

# *O que é Big Data?*

- Datasets que não podem ser capturados, gerenciados ou processados por um único computador convencional



# *Big Data: Características*



# Big Data: Desafios na Gerência de Dados

## ■ Coleta e armazenamento de dados



- Transferência de dados
  - Intrasite
  - Intersites
- Pré-processamento
  - Limpeza de dados
  - Remoção de redundância
- Carga (ETL)

Declínio do princípio de *one size fits all*

# *Big Data: Armazenamento e Processamento*

- Sistema de arquivos
- Tecnologia de Banco de Dados
- Processamento de fluxos de dados



# *Big Data: Armazenamento de Dados*

- Raw Data or Flat Files
  - Direct Attached Storage (DAS)
  - Network Attached Storage (NAS)
  - Storage Area Network (SAN)
- Aspectos de armazenamento distribuído
  - Consistência
  - Disponibilidade
  - Particionamento com redundância

# *Big Data: Armazenamento e Consulta a Dados*

- SGBD Relacional Centralizado
  - Amostragem
- SGBD Relacional Paralelo
  - Dados homogêneos
  - Dados normalizados
  - Carga de dados
  - Controle de concorrência
- Sistemas No-SQL
  - Chave-valor
  - Colunas
  - Documento
  - Paralelos / UDF paralelas

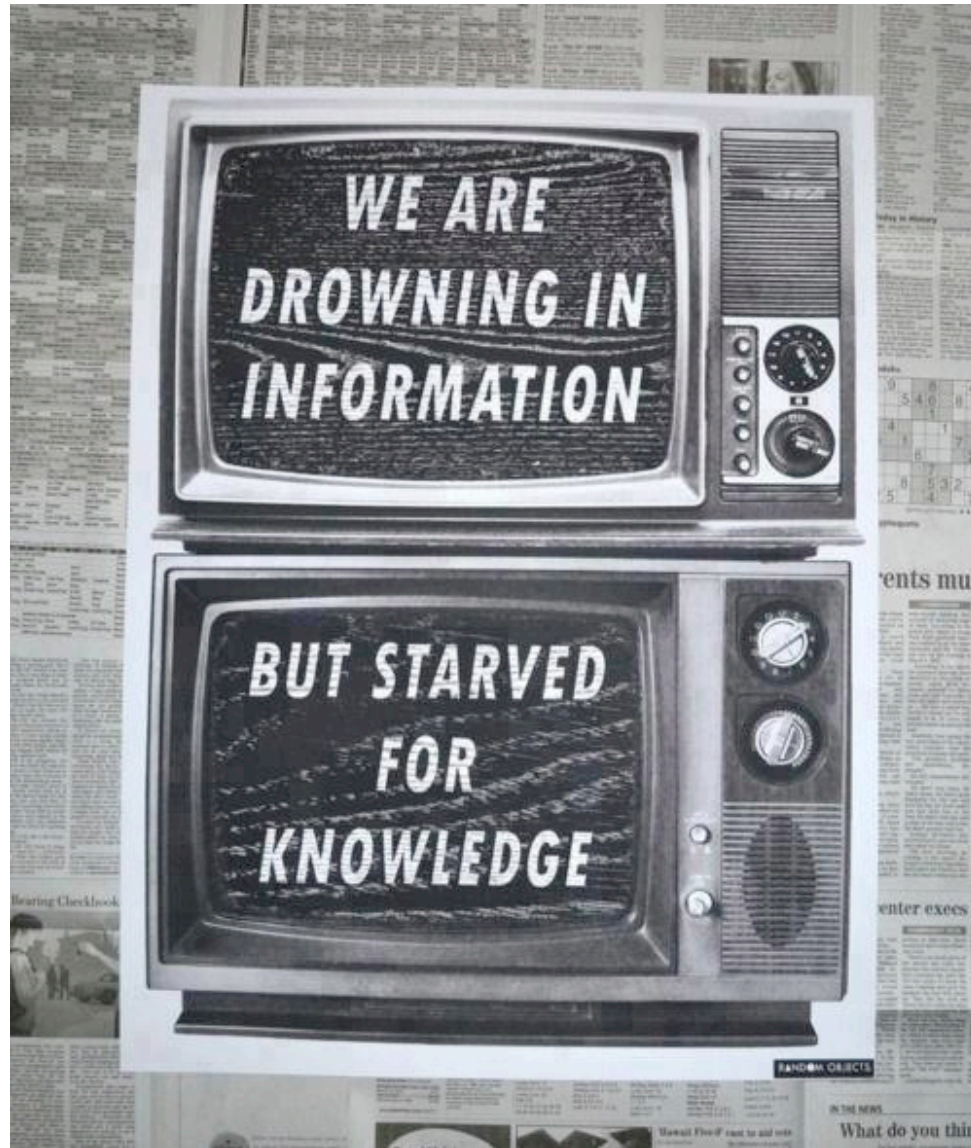
## *Big Data: Processamento de Fluxos*

- Uso de UDFs
- Modelo de processamento paralelo de *Message Passage Interface (MPI)* e *Open Multiprocessing (OpenMP)* podem não ser suficientes
- Map Reduce
  - Hadoop
  - Spark
- PigLatin

# Big Data – Aspectos Arquiteturais e de Negócios

<h3>Vertical Apps</h3>	<h3>Ad/Media Apps</h3>	<h3>Business Intelligence</h3>	<h3>Analytics and Visualization</h3>
<h3>Log Data Apps</h3>		<h3>Data As A Service</h3>	
<h3>Analytics Infrastructure</h3>	<h3>Operational Infrastructure</h3>	<h3>Infrastructure As A Service</h3>	<h3>Structured Databases</h3>
<h3>Technologies</h3>			

# *Análise de Dados*



## *Tipos de análise de dados*

- Tempo real
- Off-line
- Em memória
- Análise centrada em Inteligência de Negócios
- Análise Massiva de Dados (Hadoop)

# Ciência de Dados



# Ciência de dados: desafio para a ciência, indústria e governo

OEI | Ciencia | Red | Formación | Contactar |



iberoamérica **divulga**

red iberoamericana  
de comunicación  
y divulgación científica

Inicio | Opinión | Reportajes | Noticias | Entrevistas | Multimedia

Salud | Comunidad

## Ciência de dados: desafio para a ciência, indústria e governo

11 de agosto de 2015

Por Artur Ziviani, Fábio Porto e Eduardo Ogasawara. LABJOR

O tratamento do dilúvio de dados sendo produzido pelas ciências e por bilhões de usuários de serviços globais de internet se apresenta como um dos grandes desafios para a atual sociedade do conhecimento. Apresentado de forma geral como um vetor de múltiplas facetas, o fenômeno ainda está sendo interpretado pelos cientistas e tem impulsionado iniciativas em diversas áreas. Nas ciências, o dilúvio aparece como a expressão de uma nova maneira de investigação, incitando biólogos, astrônomos, bioquímicos e demais pesquisadores em diversas áreas científicas a enfrentar problemas computacionais na chamada e-ciência, que se tornam barreiras para as suas descobertas.

Na indústria, o dilúvio de dados aparece em sintonia com o ambiente de computação em nuvem, provendo escalabilidade e tolerância a falhas, em ambientes computacionais cada vez mais complexos e de tamanho proporcional ao desafio abordado. No setor governamental, há oportunidades de se debruçar sobre imensas bases de dados do setor público com vistas a tornar o planejamento mais eficiente, bem como novos serviços que possam melhorar o atendimento ao cidadão. Novas profissões especializadas no trato e, principalmente, na análise e interpretação de grandes volumes de dados, surgiram, trazendo o método científico para o setor empresarial.

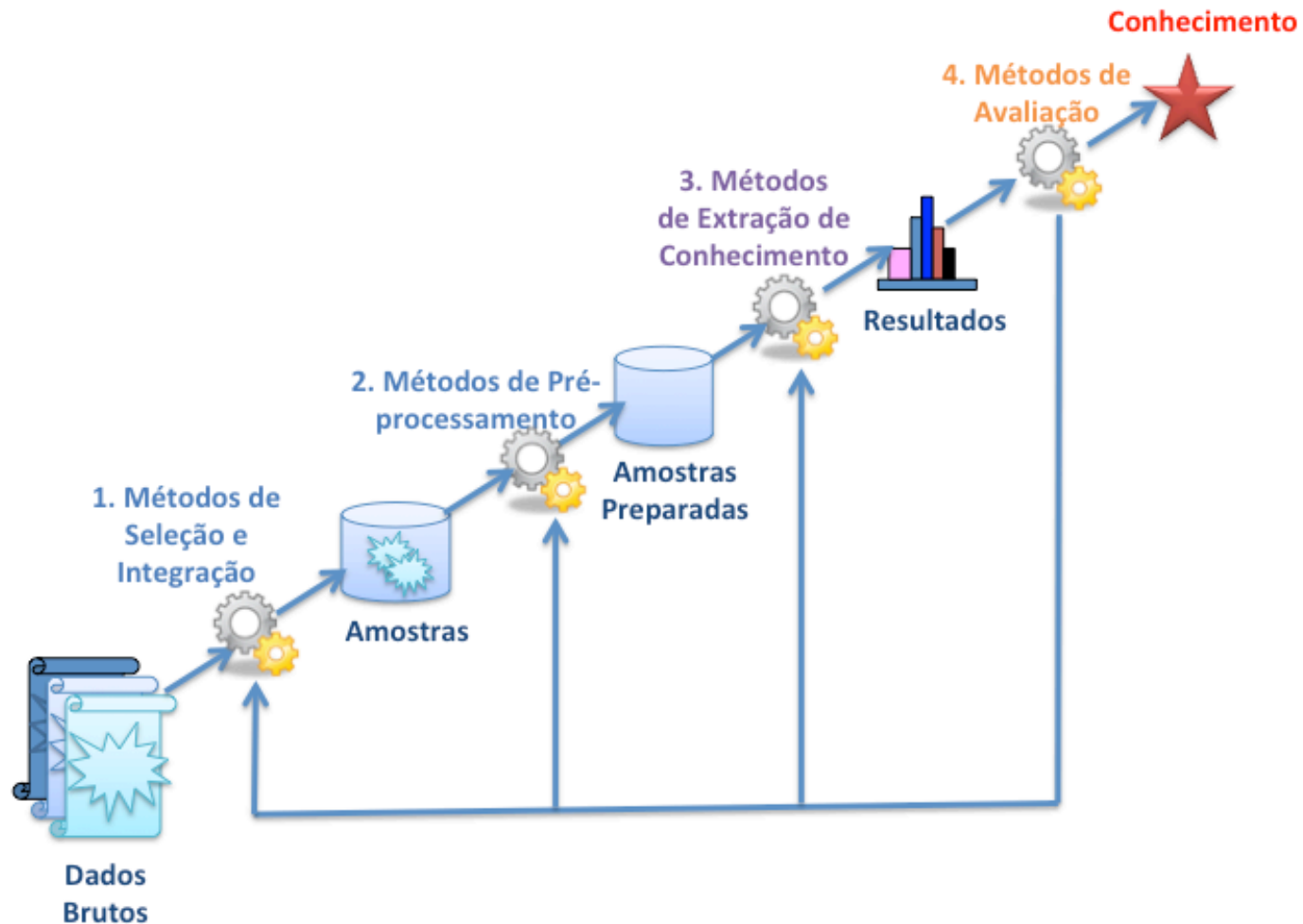


Twitter

facebook



# Processo Geral de Extração de Conhecimento



Classificação, Padrões Frequentes, Agrupamento

# *Como o meio acadêmico aborda a Ciência de Dados*

- No mundo
  - Criação de fóruns específicos (conferências, revistas)
  - Programas de Pós-graduação
    - Mestrados (23 nos EUA)
    - Especializações
- No Brasil
  - INCT em Ciência de Dados (LNCC, CEFET/RJ, IME, UFPR, ...)
  - Programa de Pós-Graduação em Ciência de Dados (PPGCD) do CEFET/RJ
  - Especializações

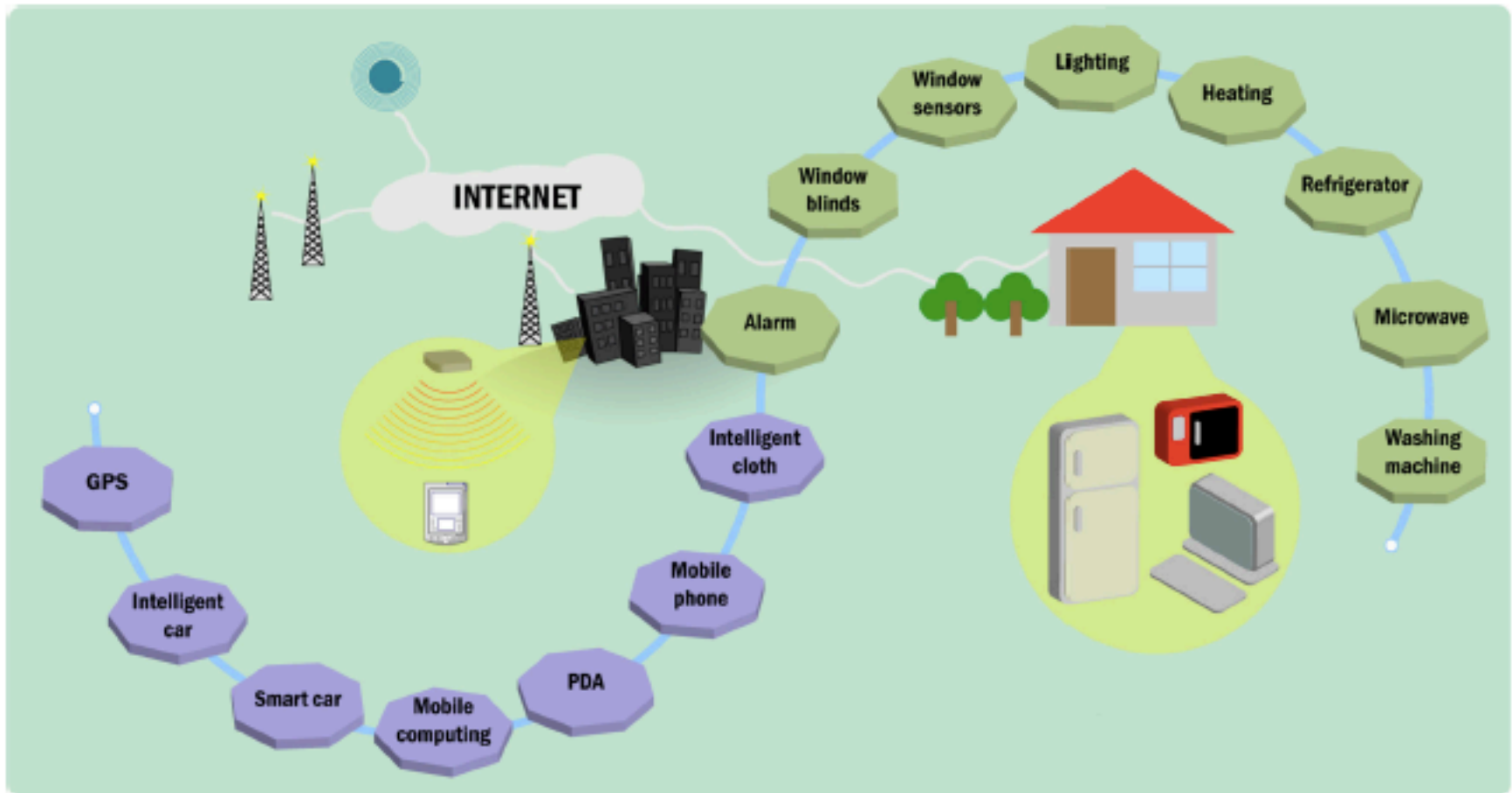
- IoT (Internet das Coisas)
  - Cada objeto passa a ser identificado e conectado a rede
  - Objetos funcionam como sensores
  - Aspectos temporais e espaciais



# *IoT: Aplicações*

- **Logística**
  - monitoramento de veículos
  - otimização dinâmica de rotas
- **Cidades inteligentes**
  - gerenciamento de recursos hídricos
  - redução de congestionamento
  - identificação de acidentes
  - vagas para estacionamento de carros
  - aumento de segurança
- **Medicina**
  - acompanhamento de pacientes
- **Casas inteligentes**
  - uso de recursos e cômodos
  - controle de temperatura
  - reconhecimento de faces e identificação de humor
  - alerta de eventos raros

# Ligação entre IoT e Big Data



Hoje não é o principal cenário de Big Data, mas em 2030...

# *Big Data em IoT: Um novo desafio à nossa porta*

*Fabio Porto, Eduardo Ogasawara, Artur Ziviani*

- Mercado de 4 trilhões de dólares (2020)
- 25 bilhões de dispositivos
- 50 trilhões de GB de dados
- Fontes de dados de alta frequência de transmissão
- Coleta e armazenamento desses dados



# *Big Data em IoT: Desafios*

- Gerência de Dados
  - Amostragem
  - Agregação
  - Armazenamento espaço-temporal
  - Processamento de grandes volumes de transações em memória
- Análise de Dados
  - As abordagens de KDD são centralizadas
  - No cenário de IoT os algoritmos precisam ser paralelizados
  - Redução de dimensionalidade e numerosidade
  - Método de Mineração: Classificação, Agrupamento, Padrões Frequentes

## *Considerações Finais*

- Oportunidade de desenvolver artefatos inteligentes
  - Potencial de produção de patentes
- Cenário muito rico para as pesquisas em:
  - Banco de Dados
  - Mineração de Dados
  - Aprendizado de máquina
  - Ciência de Dados
- Pressão crescente pela formação dos Cientistas de Dados
  - Habilidades diferenciadas



# Referências

- [1] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, 2013, Internet of Things (IoT): A Vision, Architectural Elements, and Future Directions, *Future Gener. Comput. Syst.*, v. 29, n. 7 (Sep.), p. 1645–1660.
- [2] D. Miorandi, S. Sicari, F. De Pellegrini, and I. Chlamtac, 2012, Internet of things: Vision, applications and research challenges, *Ad Hoc Networks*, v. 10, n. 7 (Sep.), p. 1497–1516.
- [3] L. Atzori, A. Iera, and G. Morabito, 2010, The Internet of Things: A survey, *Computer Networks*, v. 54, n. 15 (Oct.), p. 2787–2805.
- [4] F. Chen, P. Deng, J. Wan, D. Zhang, A.V. Vasilakos, X. Rong, F. Chen, P. Deng, J. Wan, et al., 2015, Data Mining for the Internet of Things: Literature Review and Challenges, Data Mining for the Internet of Things: Literature Review and Challenges, *International Journal of Distributed Sensor Networks*, *International Journal of Distributed Sensor Networks*, v. 2015, 2015 (Aug.), p. e431047.
- [5] C.-W. Tsai, C.-F. Lai, M.-C. Chiang, and L.T. Yang, 2014, Data Mining for Internet of Things: A Survey, *IEEE Communications Surveys Tutorials*, v. 16, n. 1 (First.), p. 77–97.
- [6] C. Perera, C.H. Liu, S. Jayawardena, and M. Chen, 2014, A Survey on Internet of Things From Industrial Market Perspective, *IEEE Access*, v. 2, p. 1660–1679.
- [7] M. Chen, S. Mao, and Y. Liu, 2014, Big Data: A Survey, *Mobile Networks and Applications*, v. 19, n. 2 (Jan.), p. 171–209.
- [8] R.-I. Ciobanu, V. Cristea, C. Dobre, and F. Pop, 2014, "Big Data Platforms for the Internet of Things", In: N. Bessis and C. Dobre, eds., *Big Data and Internet of Things: A Roadmap for Smart Environments*, Springer International Publishing, p. 3–34.