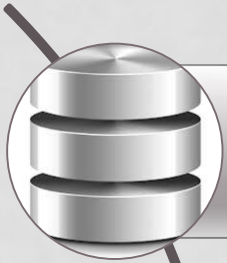


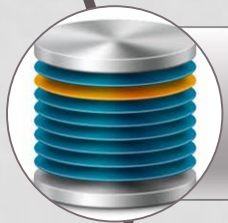
BIG DATA

Jorge Rady de Almeida Jr.
jorgerady@usp.br
Escola Politécnica da USP

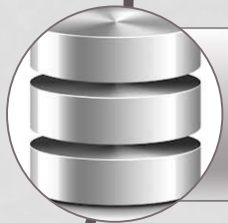
BIG DATA - MOTIVAÇÃO



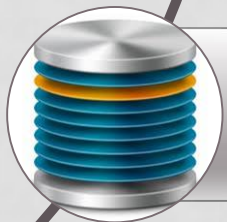
Aumento da geração e armazenamento digital de dados



Aumento da capacidade de processamento em geral: Computadores, Tablets, Smartphones, ...



Aumento da capacidade de armazenamento



Internet das Coisas (Internet of Things)

BIG DATA - MOTIVAÇÃO



Dados contêm
informação/conhecimento
significativo e que pelo seu
custo e tempo de
obtenção merecem
tratamento adequado

Compreende:
análise, captura,
tratamento,
armazenamento,
compartilhamento,
consulta e
visualização

Problema: tomar
grandes/complexos
conjuntos de dados
que os aplicativos
tradicionais não
consigam processar
em tempo
adequado

BIG DATA - DEFINIÇÃO

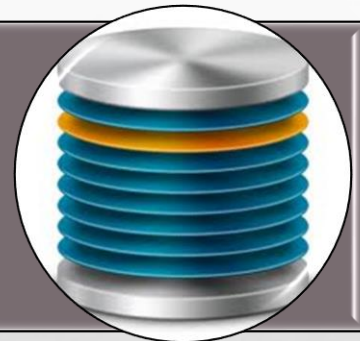
Uma (nova) geração de tecnologias e arquiteturas construídas para economicamente extrair grandes volumes de uma grande variedade de dados, permitindo captura em alta velocidade, descoberta e/ou análise



BIG DATA – 5 V's

- Quantidade crescente de pessoas e dispositivos conectados à Internet
- Volume de dados em constante aumento: terabytes → petabytes → exabytes → zetabytes . . .
- Estatísticas?

Volume: quantidade de dados



BIG DATA - 5 V's

Velocidade: geração de dados cada vez mais acelerada e necessidade de processamento mais rápido/ágil

Batch →
periódicos →
near real-time
→ real-time

Data streaming:
transferência
de dados a
uma alta taxa
de transmissão/
recepção

Bolsas de
valores,
veículos
modernos
(dezenas de
sensores), ...

BIG DATA - 5 V's

Variedade: estruturados (bases de dados), semi-estruturados (XML) e não estruturados (documentos, imagens, redes sociais, ...)

Veracidade: confiança no significado e no conteúdo dos dados - Qualidade dos dados e seu custo

Valor: mostrar grau aceitável/considerável de benefícios



BIG DATA - DESAFIOS



BIG DATA X DATA WAREHOUSE

- Análise de questões importantes ao negócio da empresa
- Dados estruturados e não estruturados x Dados estruturados
- Consistência “média” x alta consistência



BIG DATA ANALYTICS

- Processo de examinar grandes conjuntos de dados para obter padrões, correlações, tendências “ocultas”
- Data mining, text analytics, análise estatística e preditiva



BIG DATA – BDSSG

- The Big Data Senior Steering Group (**BDSSG**)
 - Criação em 2011 sob o programa **NITRD**
(Networking and Information Technology R&D)
 - Identificar pesquisas/desenvolvimentos em Big Data
- The White House Big Data R&D Initiative
 - Criação em 2012
 - Abrangeu o BDSSG

BIG DATA - BDSSG - ÁREAS PRIORITÁRIAS

Saúde

Ambiente e sustentabilidade

Resposta a emergências

Resiliência a Desastres

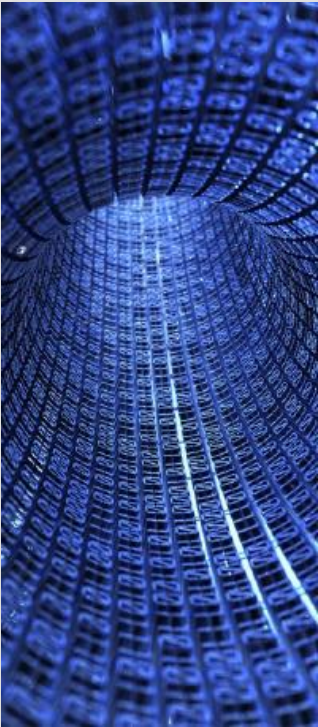
Ciências

Educação

Cyber espaço seguro

Transporte e energia

Manufatura e robótica





THE FEDERAL BIG DATA RESEARCH AND DEVELOPMENT STRATEGIC PLAN

THE NETWORKING AND INFORMATION
TECHNOLOGY RESEARCH AND
DEVELOPMENT PROGRAM



MAY 2016

BIG DATA – BDSSG

- 7 estratégias importantes para a pesquisa e desenvolvimento em Big Data
 - Alavancar nova geração de técnicas/tecnologias para obtenção de dados
 - Alavancar nova geração de técnicas/tecnologias para análise de dados
 - Aperfeiçoar a cyber-estrutura para apoio a agência governamentais
 - Aperfeiçoar políticas de compartilhamento de dados
 - Promover uso ético e seguro dos dados
 - Promover o aperfeiçoamento de cientistas de dados
 - Incentivar a Rede de Colaboração do Big Data

BIG DATA - SEGURANÇA

Problemas de Segurança

- Também são “big”
- Número grande de usuários
- Volume, variedade, velocidade, . . .



BIG DATA - SEGURANÇA

✓ Problemas de Segurança

- ❖ Informações públicas
- ❖ Proteger a privacidade
- ❖ Gerente de informações/Conhecimento: políticas de segurança
- ❖ Controle de acesso
- ❖ Conformidade com Leis/Normas
- ❖ Criptografia



BIG DATA - SEGURANÇA

✓ Problemas de Segurança

- ❖ Computação distribuída é mais problemática
- ❖ Necessidade de auditoria constante
- ❖ Medidas de segurança devem acompanhar o desenvolvimento da tecnologia
- ❖ A origem dos dados deve ser verificada e certificada



BIG DATA - SEGURANÇA

**NORMA
BRASILEIRA**

**ABNT NBR
ISO/IEC
27002**

Segunda edição
08.11.2013

Válida a partir de
08.12.2013

**Tecnologia da informação — Técnicas de
segurança — Código de prática para controles
de segurança da informação**

*Information technology — Security techniques — Code of practice for
information security controls*



BIG DATA - SEGURANÇA



Políticas de Segurança da Informação

Organização Interna

Segurança em Recursos Humanos

Gestão de Ativos

Controle de Acesso

Criptografia

Segurança Física e do Ambiente

BIG DATA - SEGURANÇA



Segurança nas Operações

Segurança nas Comunicações

Aquisição, Desenv. e Manutenção de Sistemas

Relacionamento na Cadeia de Suprimento

Gestão de Incidentes de Segurança da Informação

Continuidade da Segurança da Informação

Conformidade

BIG DATA



Hadoop: projeto *open source* coordenado pelo *Apache Software Foundation*.



Framework para ambientes distribuídos usado principalmente para a análise de grandes volumes de dados



Dois serviços chave (computação paralela)

- HDFS (*Hadoop Distributed File System*) → processamento visando a distribuição de dados
- MapReduce: identificação e compactação de dados

GRATO PELA ATENÇÃO