

Big Data Integration

Desafios e Oportunidades

Aluna: Rayelle Vera Cruz (rivcs)

Professores: Ana Carolina Salgado e Fernando Fonseca

Motivação



Big Data



**Oportunidades
em BDI**



**Integração
de Dados**



**Desafios em
BDI**



Conclusões



2,500,000,000,
000,000,000

(2.5 Quintilhões) **BYTES de DADOS**

90%

Gerados somente nos últimos 2 anos

3 BILHÕES de pessoas tem acesso à internet hoje em dia

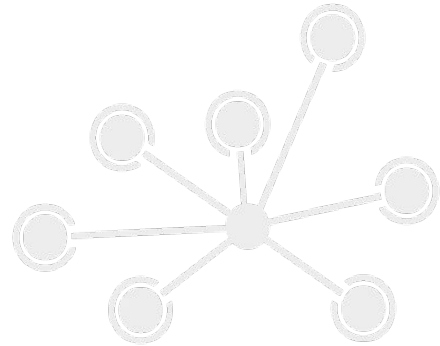


1990

A world map in shades of blue and teal, with five white location pins placed on the continents of North America, South America, Europe, Africa, and Australia. The year '1990' is written in large white font in the center of the map.

Mas o que fazer com todos esses dados?

Imagens, videos, posts, logs, audios, sms, mms, textos, gifs, doc. Xml, banco de dados SQL e NoSQL, transações...



1.

Integração de Dados

O que é?

INTEGRAÇÃO DE DADOS

Agrupar em uma única fonte ou framework de consulta fontes autônomas e heterogêneas.

Problemas/Desafios

Heterogeneidade

Esquemas diferentes, diferenças de hardware, software, SGBDs utilizados, diferenças sintáticas e semânticas

Ambiguidade Semântica

Mesmas informações podem ser modeladas de formas diferentes em fontes distintas.

Ambiguidade na Representação de Instâncias

Diferentes modos de representar uma mesma instância.
Número x Cadeia de Caracteres

Inconsistência

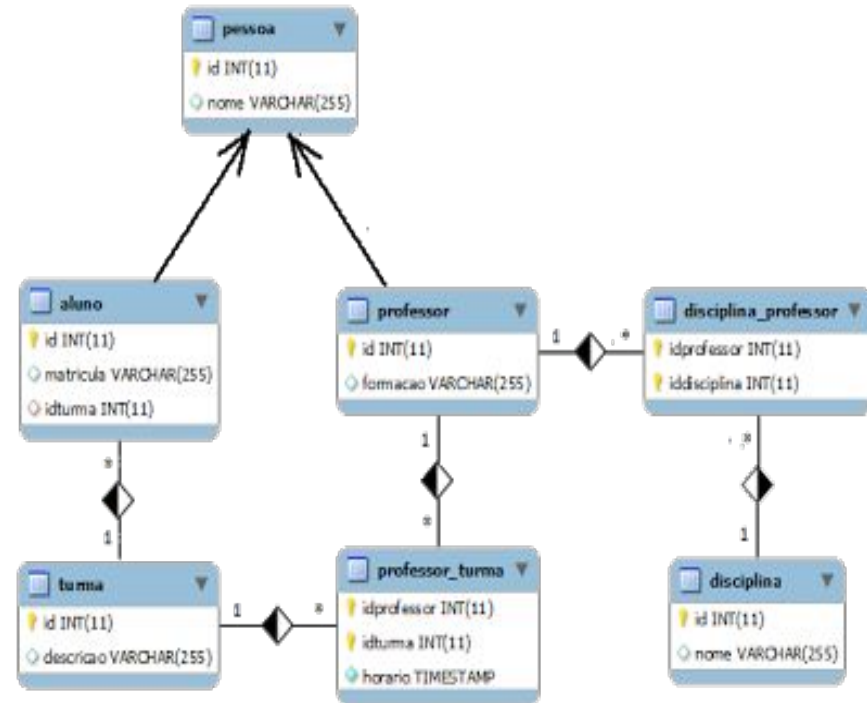
Diz respeito à qual fonte contém a informação correta e atualizada.

Arquitetura / Etapas da Integração de Dados Tradicional



Alinhamento de Esquemas

Mapeamento de instâncias e estruturas que contenham informações correspondentes. De forma a contornar o problema de fontes com esquemas distintos.



Resolução de Entidades

Correspondência de entidades que referenciam o mesmo objeto no mundo real. Esta etapa está ligada diretamente ao desafio da ambiguidade na representação de instâncias

Agrupamento de registros correspondentes, decidindo qual dado contém o valor correto.



VALOR

X

Qualidade



2. Big Data

O que é?

Big Data é o termo que descreve o imenso volume de dados – estruturados e não estruturados – que impactam os negócios no dia a dia. Mas o importante não é a quantidade de dados. E sim o que as empresas fazem com os dados que realmente importam. Big Data pode ser analisado para a obtenção de insights que levam a melhores decisões e direções estratégicas de negócio.



1997

Cientistas da NASA

2008

Cientistas Americanos

Computing Community Consortium

V's

Volume, Velocidade, Variedade, Veracidade e Valor

Big Data

BIG DATA INTEGRATIO N

Big Data + Data Integration?

Desafios e Oportunidades - BDI

Desafios

- ▶ Volume
- ▶ Velocidade
- ▶ Variedade
- ▶ Veracidade

Uma perspectiva sobre as dimensões
V's do Big Data

VOLUME

Como integrar essa quantidade toda?

VELOCIDADE

Como integrar dados que mudam em fração de minutos?

VARIEDADE

Como integrar dezenas de milhares de fontes heterogêneas?

VERACIDADE

Como integrar dados cuja veracidade difere entre fontes?

Oportunidades

- ▶ Redundância
- ▶ Plataformas de Big Data
- ▶ Vantagem Competitiva

Uma perspectiva sobre como confrontar os desafios de Big Data com as oportunidades geradas por ele.

REDUNDÂNCIA

Como a redundância de dados é uma oportunidade?

Oportunidades

PLATAFORMA

S

Quais plataformas são capazes de trabalhar com tão grande volume de dados, ao mesmo tempo?

VANTAGEM COMPETITIVA

Como ganhar vantagem competitiva
diante desse cenário?

3. Conclusões

CONCLUSÕES

- ▶ Integração de Dados
 - ▶ Arquitetura / Etapas
 - ◆ Alinhamento de Esquemas
 - ◆ Resolução de Entidades
 - ◆ Fusão de Dados

CONCLUSÕES

- ▶ Big Data
 - ▶ Desafios em BDI
 - ◆ Volume
 - ◆ Velocidade
 - ◆ Variedade
 - ◆ Veracidade

CONCLUSÕES

- ▶ Oportunidades em BDI
 - ◆ Redundância
 - ◆ Plataformas
 - ◆ Valor

1. BRYANT, R. E.; KATZ, R. H.; LAZOWSKA, E. D. **Big-Data Computing: Creating revolutionary breakthroughs in commerce, science, and society**, 2008. 7.
2. CHEN, M.; MAO, S.; LIU, Y. Big Data: A Survey, New York, 22 Janeiro 2014. 39.
3. CODD, E. F. **A Relational Model of Data for Large Shared Data Banks**. IBM Research Laboratory. San Jose, p. 11. 1970.
4. COX, M.; ELLSWORTH, D. Application-Controlled Demand Paging for Out-of-Core Visualization. **8th Conference on Visualization**, Phoenix, 18 Outubro 1997. 10.
5. DOAN, A.; HALEVY, A.; IVES, Z. Data Integration on the Web. In: _____ **Principles of Data Integration**. 1^a. ed. San Francisco: Morgan Kaufmann Publishers, 2012. Cap. 15, p. 520.
6. DONG, X. L.; SAHA, B.; SRIVASTAVA, D. Less is More: Selecting Sources Wisely for Integration. **VLDV'13**, 2013. 12.
7. DONG, X. L.; SRIVASTAVA, D. **Big Data Integration**. [S.l.]: Morgan & Claypool Publishers, 2015.
8. SALGADO, A. C.; LÓSCIO, B. F. Integração de Dados na Web, 2001.

REFERÊNCIAS

9. VOUCHERCLOUD. Every Day Big Data Statistics - 2.5 Quintillion Bytes of Data Created Daily. **VCloudNews**, 2015. Disponível em: <<http://www.vcloudnews.com/every-day-big-data-statistics-2-5-quintillion-bytes-of-data-created-daily/>>. Acesso em: 19 Novembro 2016.
10. ZICARI, R. V. Big Data: Challenges and Opportunities. In: AKERKAR, R. **Big Data Computing**. [S.l.]: [s.n.], 2013. Cap. 3, p. 564.

Perguntas?



OBRIGADA!