

Big Data

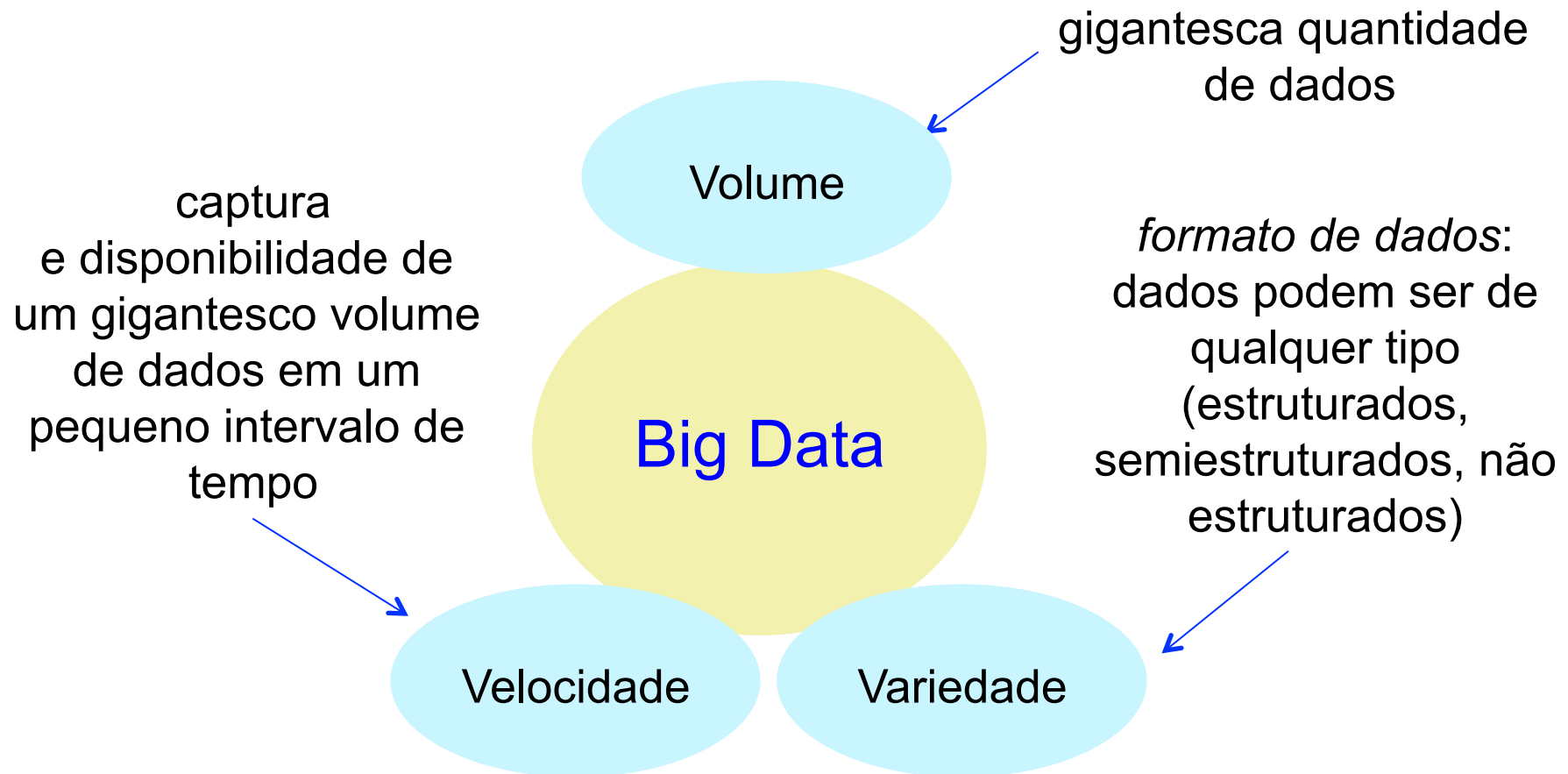
Processamento Analítico de Dados
Profa. Dra. Cristina Dutra de Aguiar Ciferri

Big Data



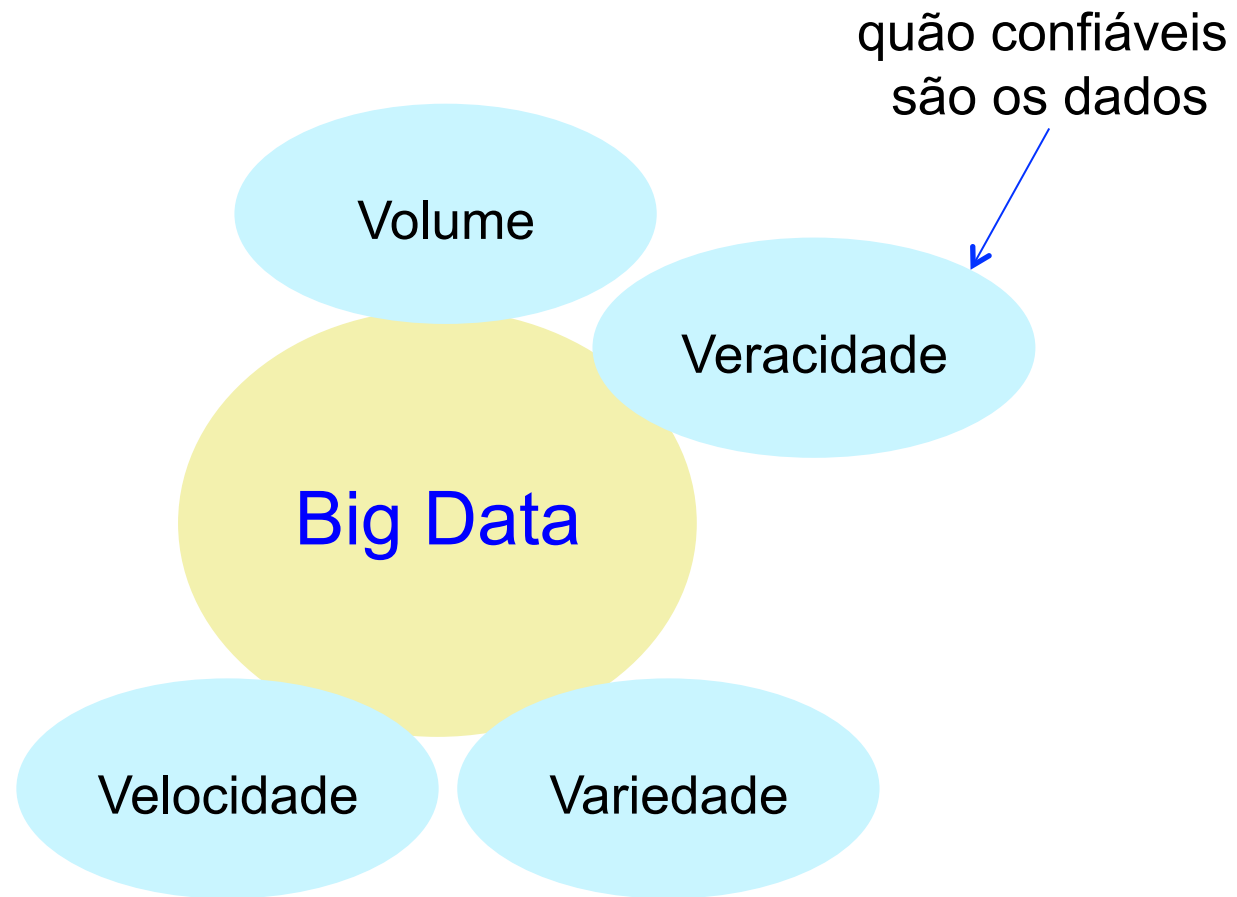
Imagen: https://students.washington.edu/bits/club/wp-content/uploads/2014/05/Big-data.Red_.jpg

Definição: Modelo de 3Vs



Chen, M.; Mao, S.; Liu, Y. Big Data: A Survey. *Mobile Network Applications*, v. 19, p. 171-209, 2014.

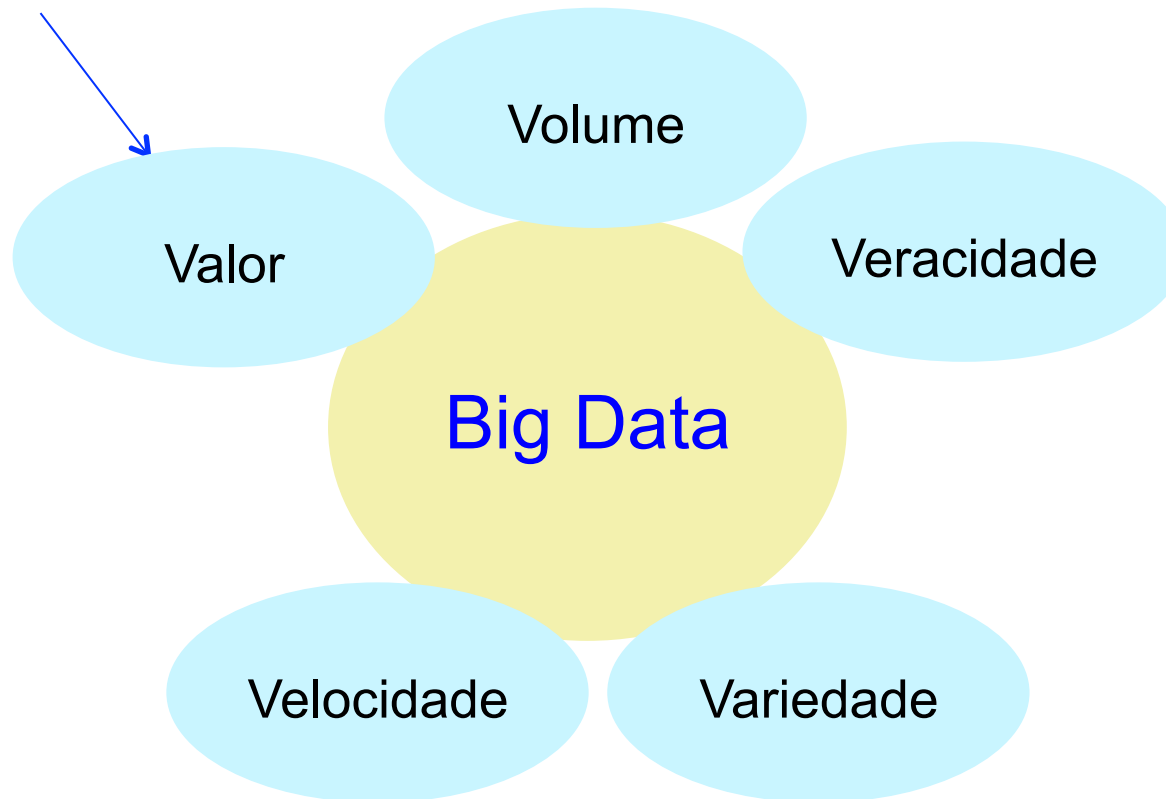
Definição: Modelo de 4Vs



Dong, X. L.; Srivastava, D. Big Data Integration. *Proceedings of the VLDB Endowment*, v. 6, n. 11, p. 1188-1189, 2013.

Definição: Modelo de 5Vs

os dados devem ter importância

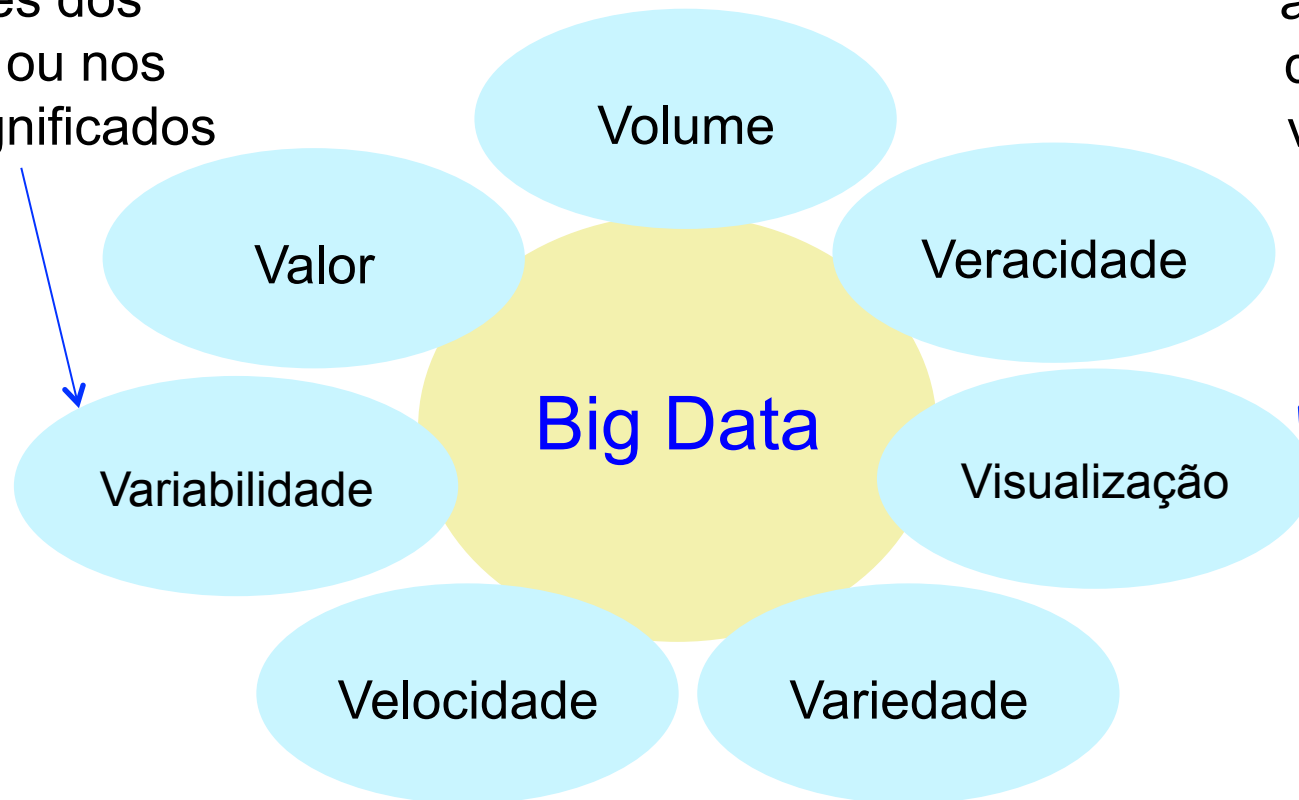


Sharma, S.; Mangat, V. Technology and Trends to Handle Big Data: Survey. In: *Proceedings of the Fifth International Conference on Advanced Computing & Communication Technologies*, p. 266-271, 2015.

Definição: Modelo de 7Vs

alterações nos valores dos dados ou nos seus significados

exibição apropriada dos dados volumosos



Wrembel, R. Novel Big Data Integration Techniques: Painel Discussion at BigNovelTI 2017@ADBIS2017

Big Data

- Dados possuem **informações valiosas**
- Exemplos
 - análise de navegação e *reviews* de usuários
 - recomendação personalizada de produtos
 - análise de mensagens de redes sociais



Observatório da Dengue

- Monitoramento de mensagens de redes sociais
 - desenvolvido por pesquisadores da UFMG

Aumento de mensagens relacionadas ao termo “dengue” em uma região pode indicar um possível surto

observatôriodadengue

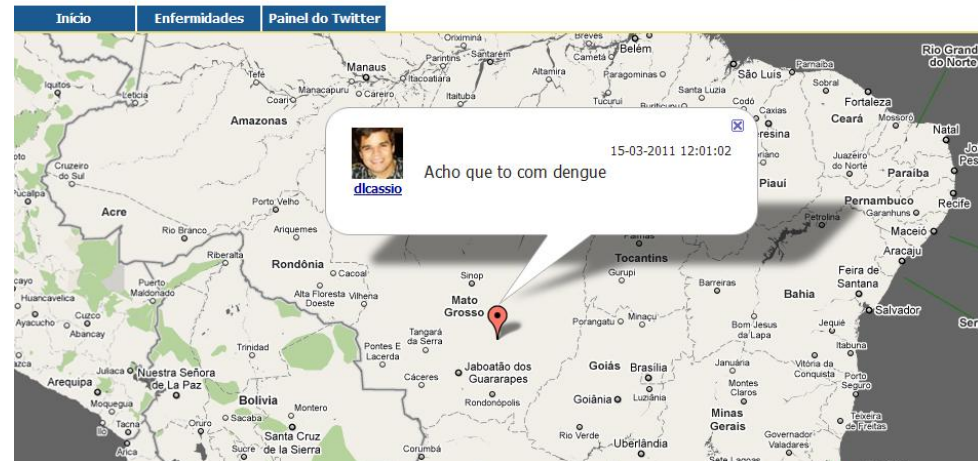


Imagem: http://www.ufmg.br/online/arquivos/anexos/Portal_Observatorio_Dengue.JPG

Desafios

- Uso de **ambientes computacionais** com grande capacidade de armazenamento e processamento
 - clusters de computadores
 - computação em nuvem
- Uso de **sistemas de arquivos distribuídos**
 - HDFS

Desafios

- Emprego de **paradigmas de programação paralela e distribuída**
 - *framework* MapReduce
 - *framework* Spark
- Uso de **bancos de dados NoSQL**
 - diferentes modelos de dados