

Big Data

Prof .Maiquel de Brito

`maiquel.b@ufsc.br`

BLU3024

UFSC Blumenau

Contexto

Qual é o **disco** de rock mais ouvido da história?

Contexto

Qual é o **disco** de rock mais ouvido da história?

Back in black

AC/DC, 1980 d.C

51.000.000 cópias



Contexto

Qual é o **disco** de rock mais ouvido da história?

Back in black

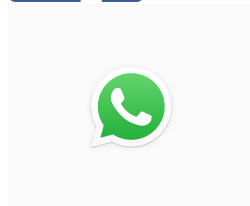
AC/DC, 1980 d.C

51.000.000 cópias



Mas, pode-se dizer que esse foi o disco mais **ouvido**?





Contexto

Aplicações atuais

produzem **muitos dados**

existem porque dispõem de muitos dados

produção massiva e rápida de informações

- tweets

- cliques em webpages

- trajetórias de veículos

- dados de sensores cardíacos

- ⋮

O que é *Big Data*?

Big data são ativos informacionais com alto volume, alta velocidade e/ou alta variedade, que requerem novas formas de processamento para possibilitar tomadas de decisão aprimoradas, descobertas de informações e otimização de processos. (Gartner, 2012)

De onde vêm os *Big Data*?

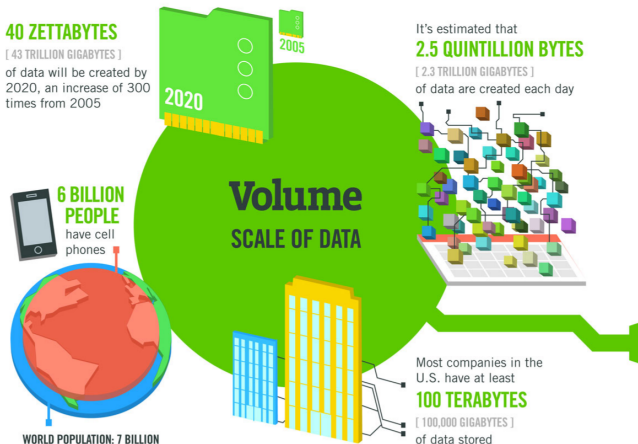
Algumas fontes de *big data* :

- Redes sociais, multimídia e websites
todos nós produzimos dados
- Instrumentos científicos (ex. CERN - 1PB/sec)
coletas de dados diversos
- Dispositivos móveis
rastreamento contínuo de objetos
- Sensores
coletando medidas de todo tipo

Características do *Big Data*

- Volume
- Variedade
- Velocidade
- Veracidade

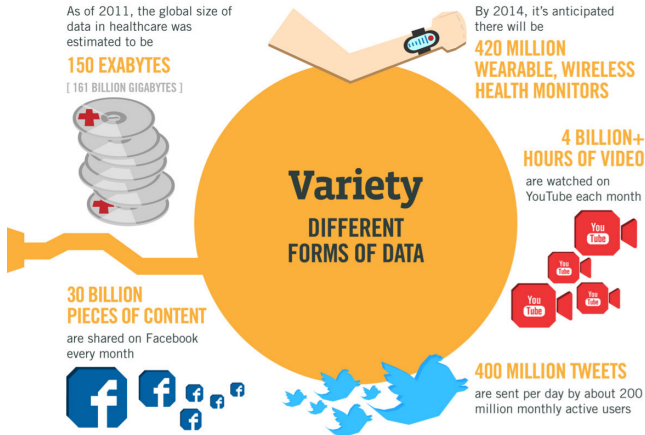
Características do *Big Data* - Volume



Adaptado de <http://www.ibmbigdatahub.com/infographic/four-vs-big-data>

- O volume dos dados cresce exponencialmente
- Até mesmo pequenas informações podem gerar *big data*

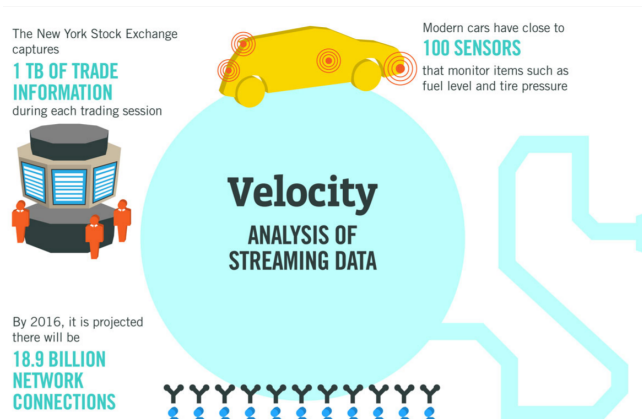
Características do *Big Data* - Variedade



Adaptado de <http://www.ibmbigdatahub.com/infographic/four-vs-big-data>

- Vários formatos, tipos e estruturas
XML estruturados a dados multimídia

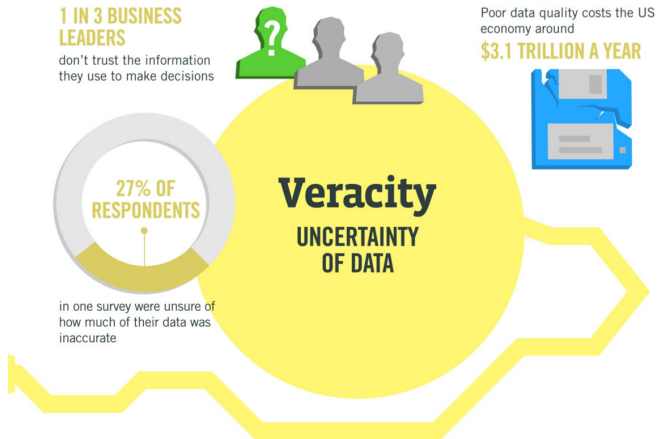
Características do *Big Data* - Velocidade



Adaptado de <http://www.ibmbigdatahub.com/infographic/four-vs-big-data>

Dados são gerados rapidamente e precisam de processamento rápido

Características do *Big Data* - Veracidade



Adaptado de <http://www.ibmbigdatahub.com/infographic/four-vs-big-data>

Possibilidade de inconsistências, incompletude, latência, ambiguidades, aproximações, etc.