WHAT IS
BIG DATA?

WHAT IS
BIG DATA?

HADOOP

RDBMS

Petabytes

Gigabytes

Batch

Linear Scale

TRADITIONAL SOLUTIONS

RDBMS

Cost

IS THERE REALLY A USE CASE?

HADOOP - A GOOD SOLUTION
- Support Huge Volume
- Storage Efficiency
- Good Data Recovery Solution
- Horizontal Scaling
- Cost Effective
- Easy For Programmers & Non Programmers

WHAT IS BIG DATA?
Many Terabytes, Petabytes, Exabytes...

3Vs - Volume, Velocity, Variety

BIG DATA - CHALLENGES
Storage
Computational Efficiency
Data Loss
Cost

WHAT IS
BIG DATA?

WHAT IS BIG DATA?

Many Terabytes, Petabytes, Exabytes...

Name	Abbr.	Size
Kilo	K	1,024
Mega	M	1,048,576
Giga	G	1,073,741,824
Tera	T	1,099,511,627,776
Peta	P	1,125,899,906,842,624
Exa	E	1,152,921,504,606,846,976
Zetta	Z	1,180,591,620,717,411,303,424
Yotta	Y	1,208,925,819,614,629,174,706,176

3Vs - Volume Velocity Variety

Volume

Velocity

Variety
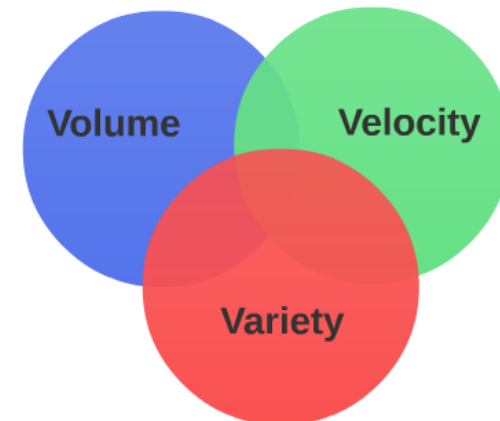
IS THERE REALLY A USE CASE?

Science

- Large Hadron Collider - 1 Petabyte every second
- NASA - 1.73 Gigabyte every hour

Government

- NSA - Utah Data Center - Yottabyte Capacity
- Big Data Research and Development Initiative
- Barack Obama's successful 2012 re-election campaign

Private

- eBay - 40PB Hadoop cluster for search, consumer recommendations, and merchandising
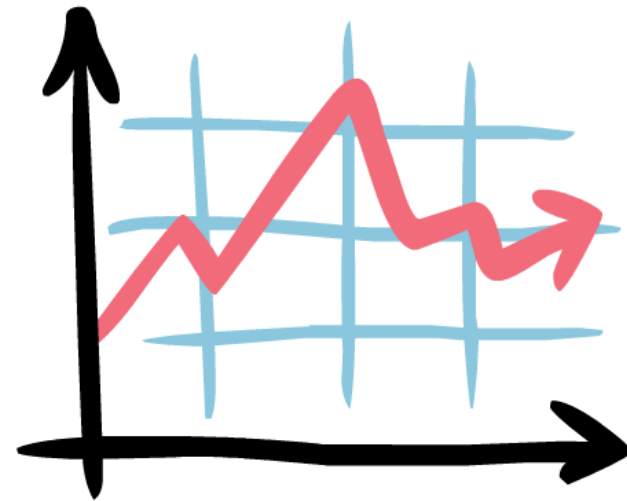- Facebook - 30 PB Hadoop cluster. 50 billion photos. 130TB of logs every day.

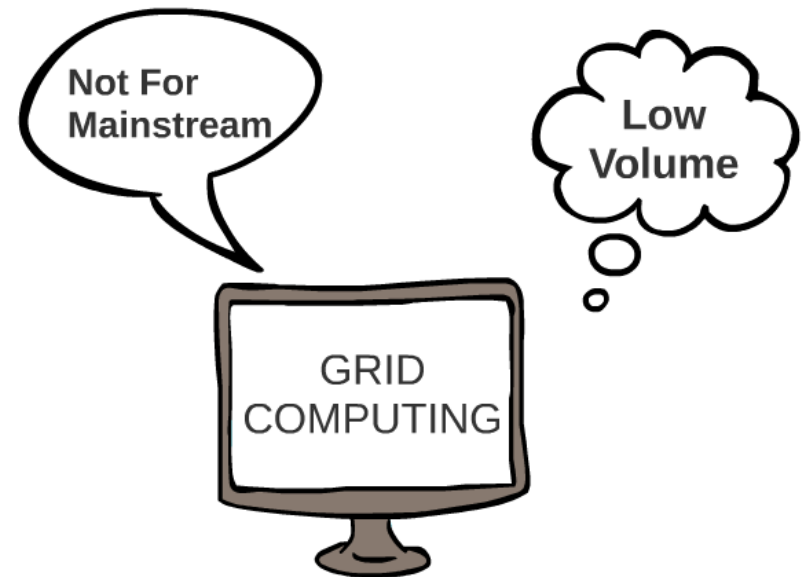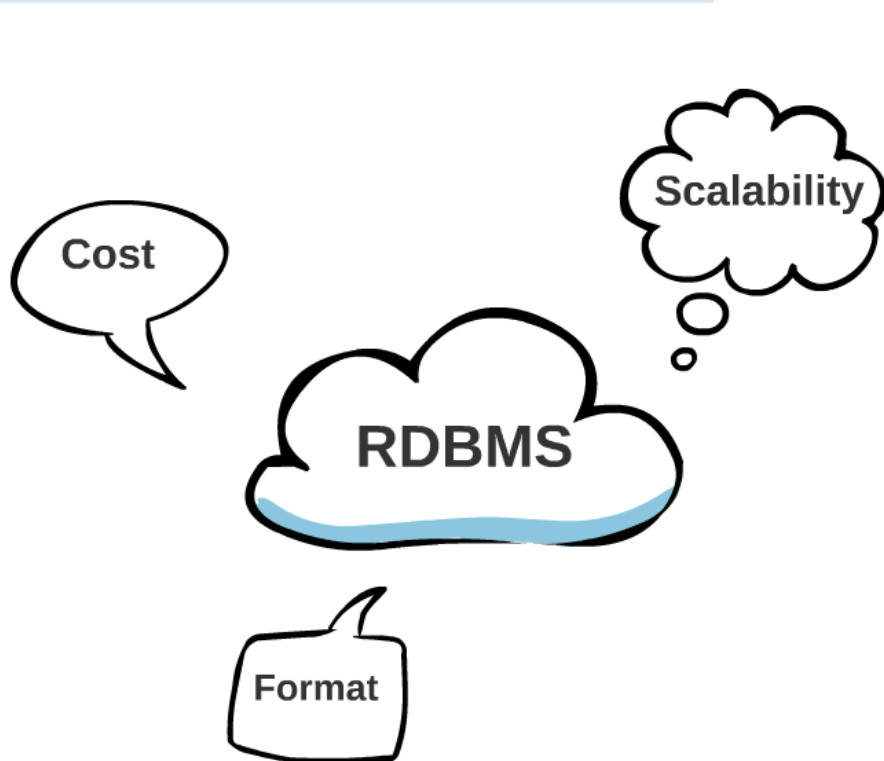# BIG DATA - CHALLENGES

Storage

Computational Efficiency

Data Loss

Cost

HADOOP - A GOOD SOLUTION

✓ Support Huge Volume

✓ Storage Efficiency

✓ Good Data Recovery Solution

✓ Horizontal Scaling

✓ Cost Effective

✓ Easy For Programmers & Non Programmers

HADOOP

RDBMS

Dynamic Schema
Linear Scale
Batch
Petabytes
Write Once, Read Many Times

Read Write Many times
Nonlinear Scale
Interactive and Batch
Static Schema
Gigabytes