

Cluster, Grid, Clouds – Julien Dhallenne

**Large data storage (examples, volumes,
challenges)**

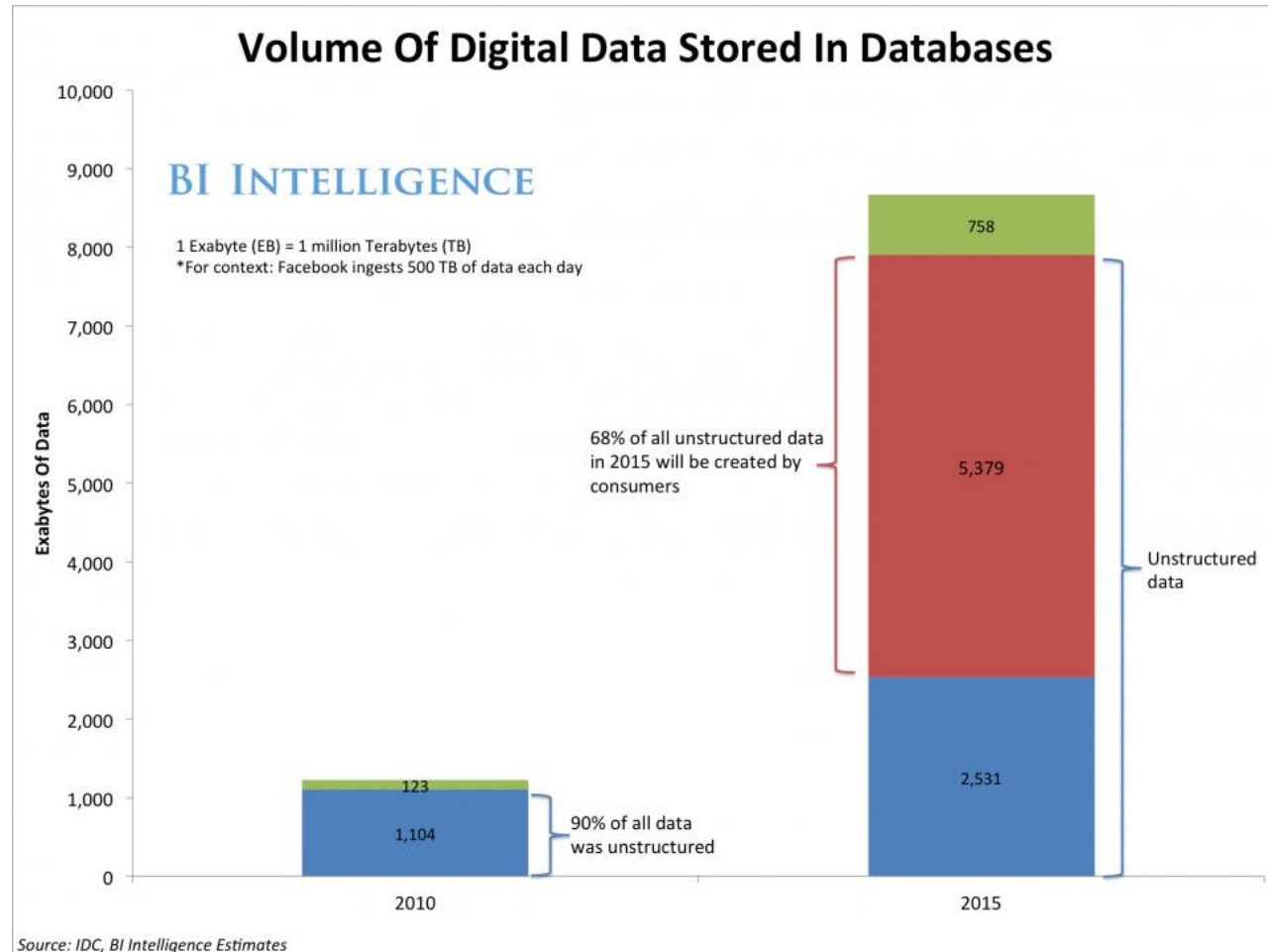
Introduction

- More and more data generated
 - Social content
 - Scientific projects
 - IoT
- Current data storage > 10^{24} bytes (Yottabyte)
- How to store this data?

Outline

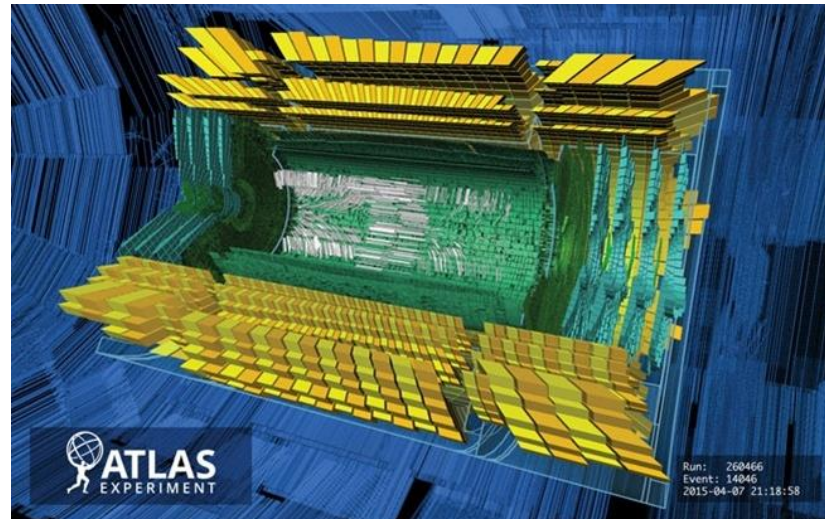
- Examples
 - Social Media
 - Scientific measurements
 - IoT – sensors data
- Volumes
- Challenges

Examples 1/3 – Social Medias

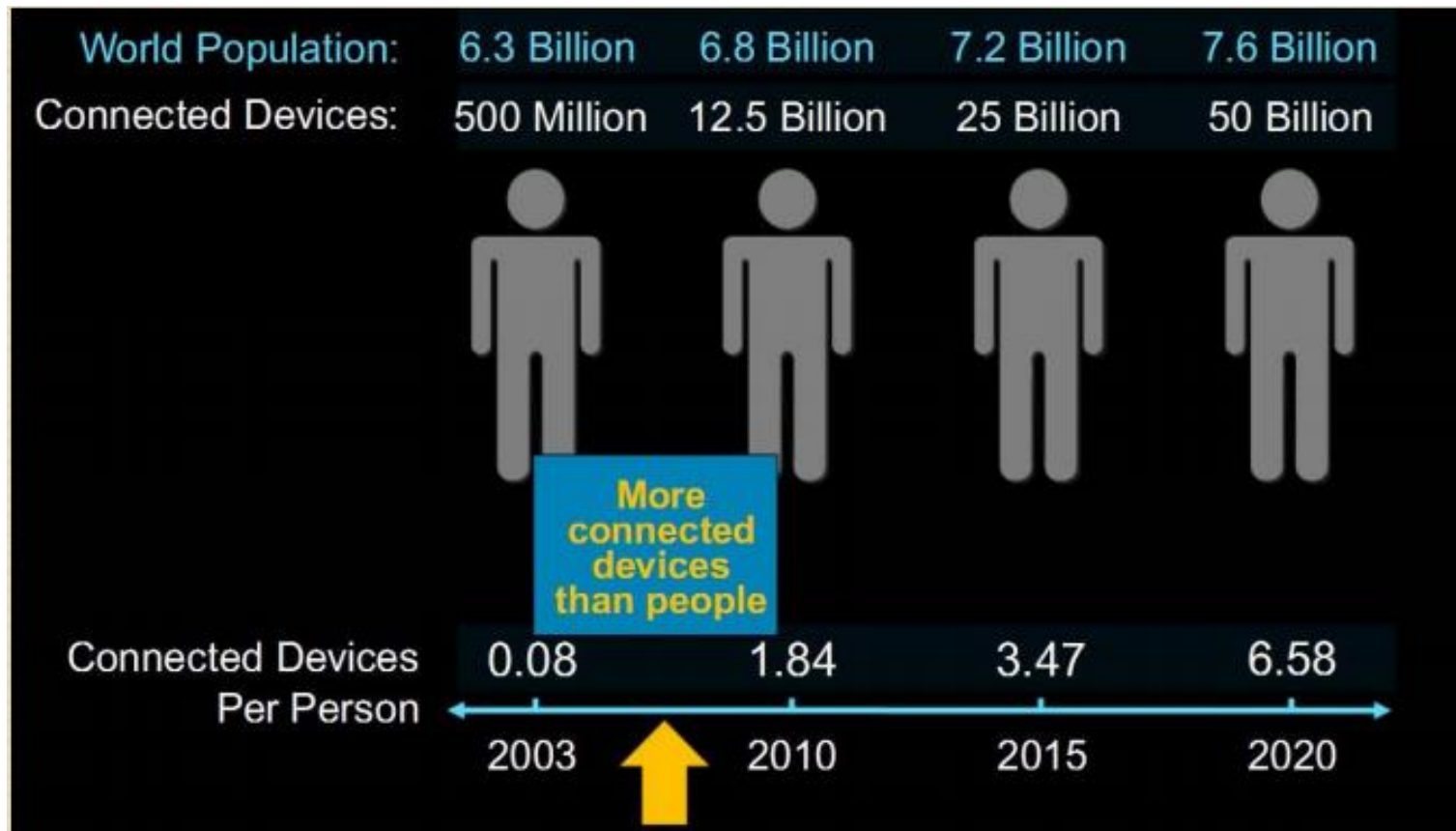


Examples 2/3 – Scientific measurements

- The Hadron Collider at CERN (~27km) produces so much data that scientists must discard most of it, hoping they haven't thrown away anything useful



Examples 3/3 – IoT sensors data



Outline

- Examples
 - Social Media
 - Scientific measurements
 - IoT – sensor data
- **Volumes**
- Challenges

Volumes - IoT

Information from the Internet of Things:

We have gone beyond the decimal system

Today data scientist uses **Yottabytes** to describe how much government data the NSA or FBI have on people altogether.

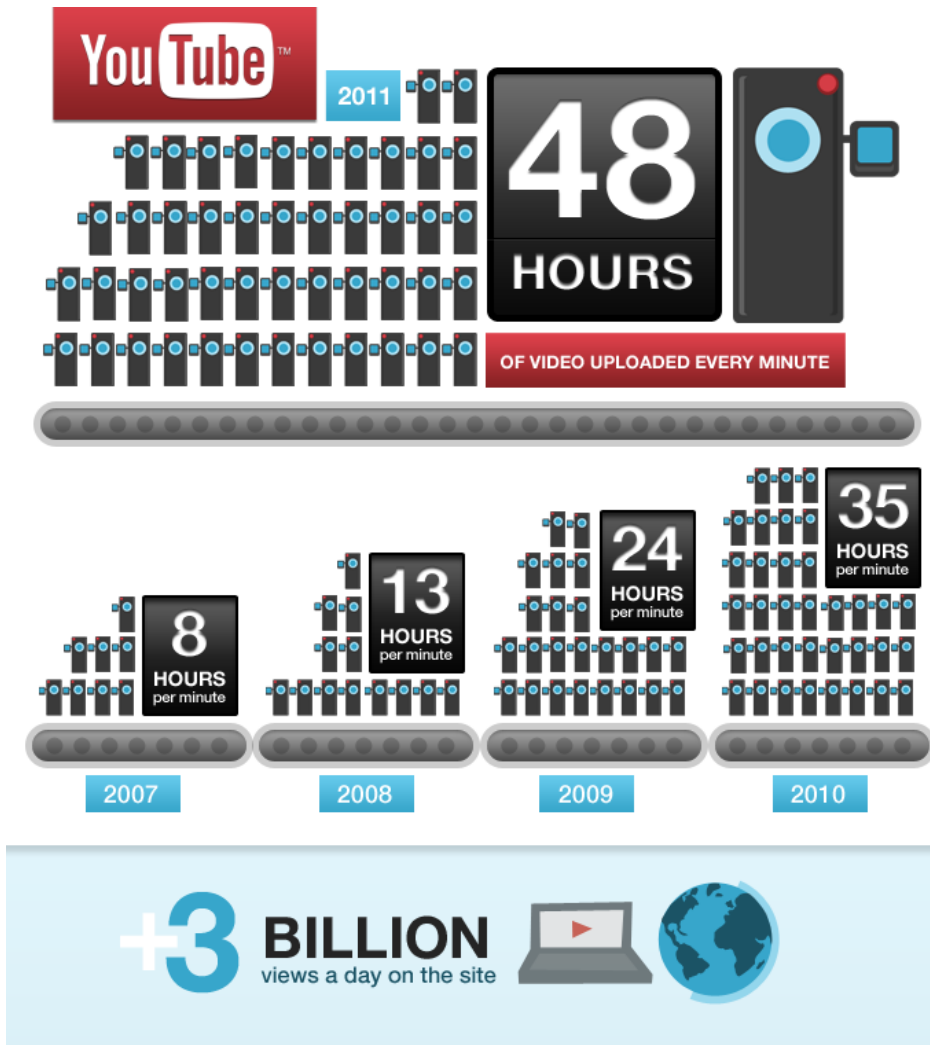
In the near future, **Brontobyte** will be the measurement to describe the type of sensor data that will be generated from the IoT (Internet of Things)



<http://tinyurl.com/ahytzdf>

Volumes – Youtube data volume

- Consequences :
 - Huge data storage
 - Numerous I/O operations



Challenges – current technologies

Featured Drives for Cloud Storage

Terascale HDD



Low cost, low-power, high capacity storage for scale-out 3.5-inch multi-drive environments.

1TB-4TB

SATA 6Gb/s, SAS 12Gb/s

[LEARN MORE >](#)

Enterprise Capacity 3.5 HDD



World's fastest 6TB nearline SED hard drive for bulk data storage.

500GB-6TB

SATA 6Gb/s, SAS 12Gb/s, SAS 6Gb/s

[LEARN MORE >](#)

1200 SSD



Server virtualisation, online transaction processing, high performance computing and heavy data analytics applications.

200GB-800GB

SAS 12Gb/s

[LEARN MORE >](#)

Source : Seagate's website, <http://www.seagate.com/gb/en/solutions/cloud/data-center-cloud/products/>, 1.6.2015

Challenges – adapted file systems

- Quantcast File System . High-performance, fault-tolerant, distributed file system
- HDFS . Distributed file system providing high-throughput access
- Ceph . Unified, distributed storage system
- Lustre . File system for computer clusters
- GlusterFS . Scale-out NAS file system
- PVFS . Designed to scale to petabytes of storage

Source : Seagate's website, <http://www.seagate.com/gb/en/solutions/cloud/data-center-cloud/products/>, 1.6.2015

Thank you

Questions?