

**Big Data generated by
People:**

**The Unstructured
Challenge**

After this video you will be able to..

- Describe how people are generating data at a faster pace than ever before
- Explain what unstructured data is
- List challenges for extracting value from this massive unstructured data

facebook



Documents



Twitter



Internet searches



facebook

YouTube



Blogger™



**Text-heavy
Unstructured**

A huge growth and volume of data!



Daily **facebook** data

>

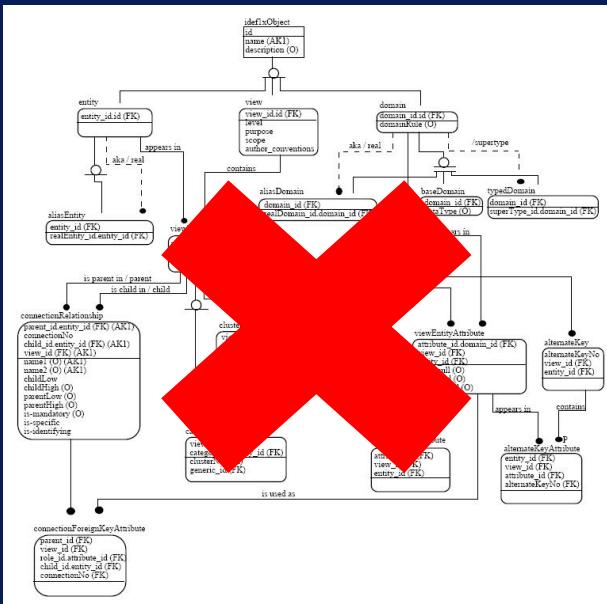
All US Academic
Libraries

2 PBs vs. 30+ PBs

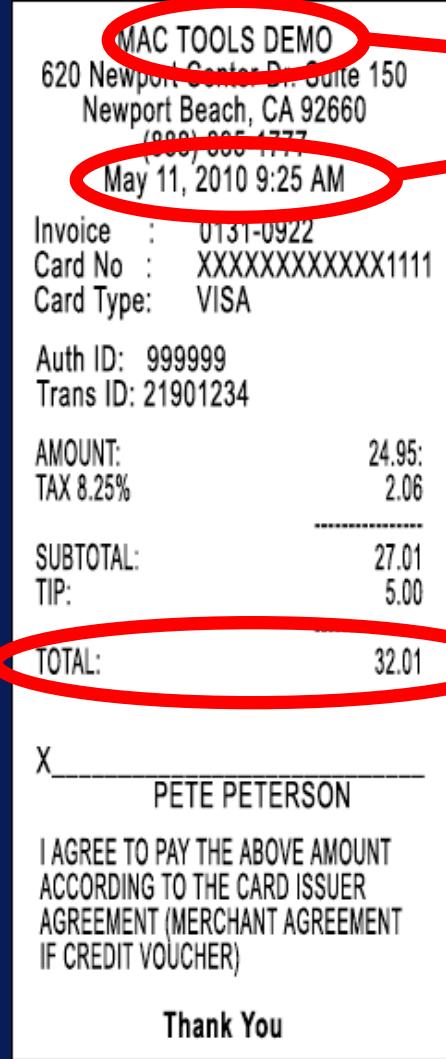
Company	Data Processed Daily
eBay	100 Petabytes (PB)
Google	100 PB
Facebook	30+ PB
Twitter	100 Terabytes(=.1PB)
Spotify	64 Terabytes

The Unstructured Data Challenge

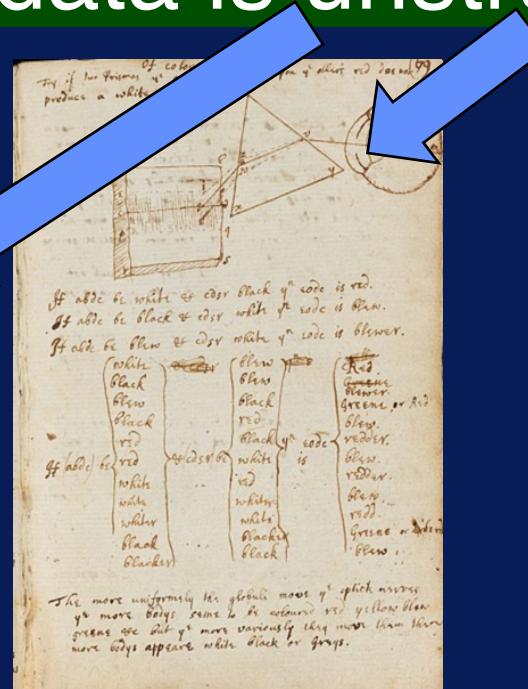
The Unstructured Data Challenge

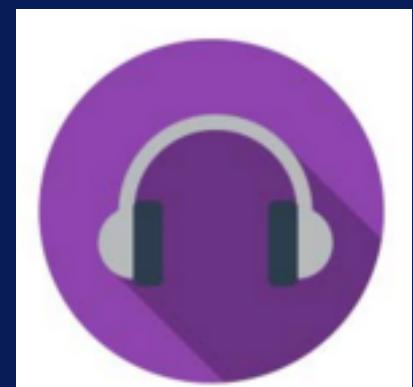
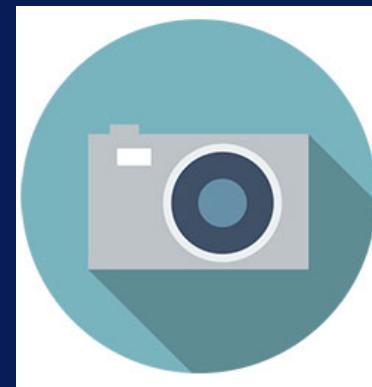
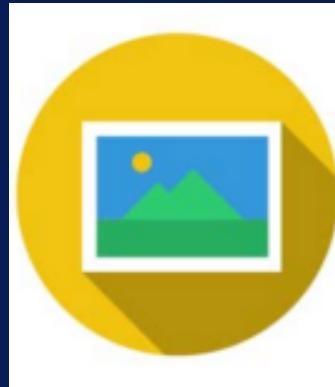


Structure



80%-90% of entire data is unstructured!





In every minute




100.000
tweets are send



284.500
content elements are shared



48 hours
of video is uploaded



2 Million
searches are made



3.600
Photos are shared



48.000
Apps are downloaded



571
websites are created



210.000€
is spent in webshops



2.083
check-ins to locations



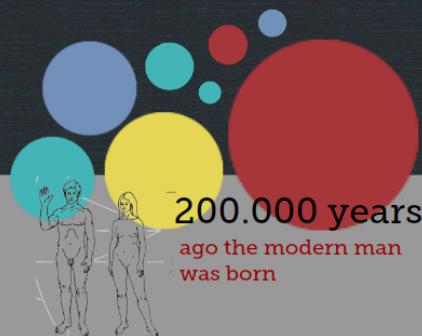
480
new people enters to Internet



20.000 years
equals the time people spend in Facebook every day



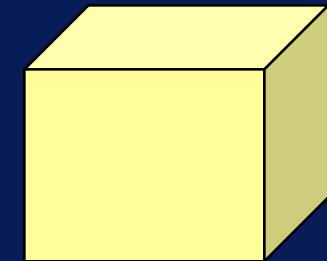
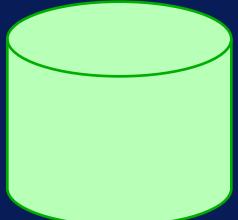
3 Billion hours
of video is watched every month. That is
342.239 years



Licenced under CC by Jarmo Röksä, 12/2012

Velocity

$$V = \frac{\Delta x}{\Delta t}$$

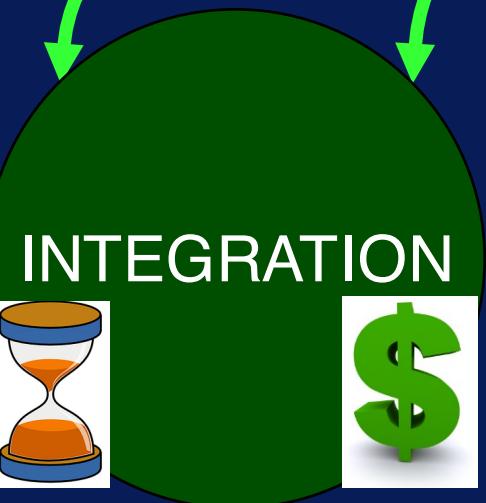


TRANSACTIONS

TWEETS

LIKES

BLOGS



Data Acquisition



Storage

Retrieval

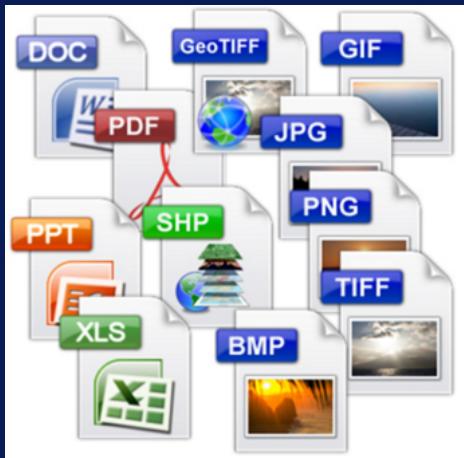
Cleaning

A 3D-style icon representing data retrieval. It shows a stack of four white cylinders on the left, with a large blue upward-pointing arrow positioned below them. To the right of the cylinders is a magnifying glass with a circular frame, focused on one of the cylinders, symbolizing the examination or filtering of data.

Processing

A
Z

A 3D-style icon representing data processing. It consists of a white rectangular box with a black downward-pointing arrow in the center. The letter 'A' is positioned at the top edge of the box, and the letter 'Z' is positioned at the bottom edge, suggesting a range or sequence of operations.



Tools

Data

Skilled People

?



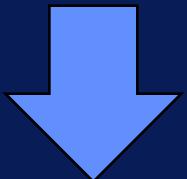
Value

?

People



Unstructured Data



Challenges



$$V = \frac{\Delta x}{\Delta t}$$

