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

EMAIL
facebook
Twitter
YouTube
Internet searches
Blogger
Documents

Text-heavy

Unstructured

A huge growth and volume of data!

Daily facebook data

>

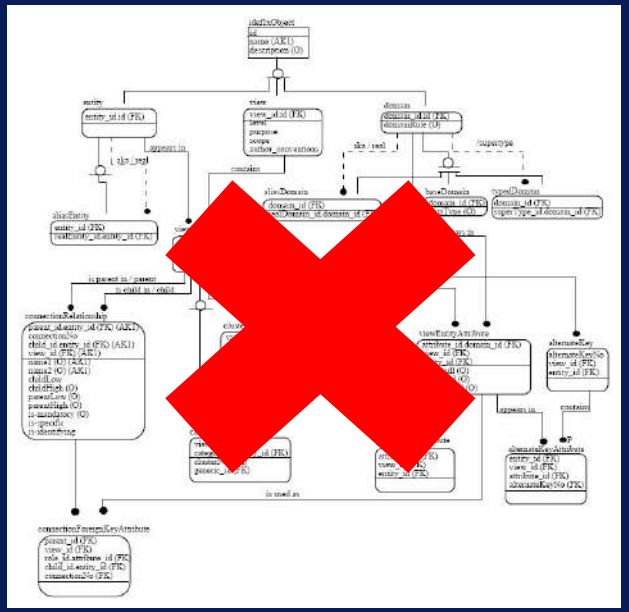
All US Academic Libraries

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



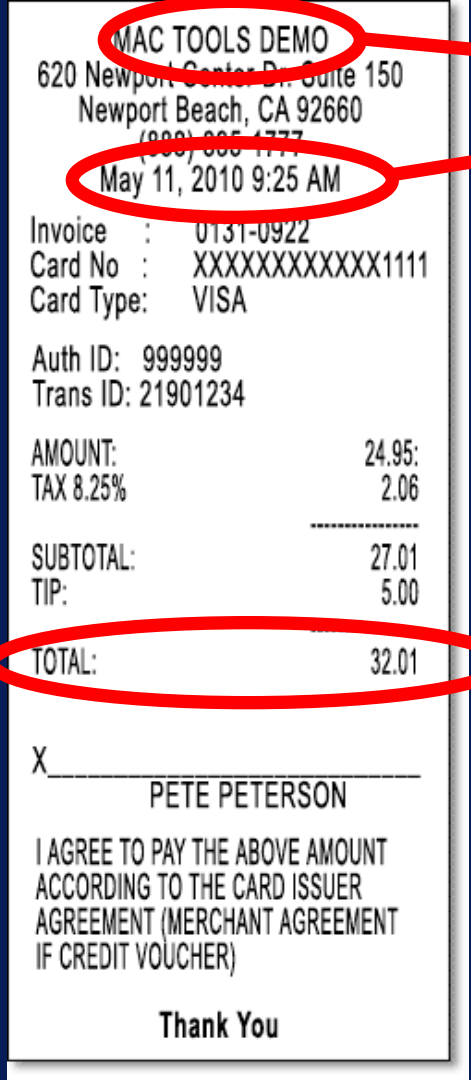
Structure

SALES RECEIPT

Date: _____

Qty.	Description	Price	Amount
		Subtotal	
		Tax	
		Total	

Sale Made with :
[ ] Cash
[ ] Credit Card
[ ] Check, No. _____
[ ] Other _____



MAC TOOLS DEMO
620 Newport Center Dr. Suite 150
Newport Beach, CA 92660
(800) 855-1777
May 11, 2010 9:25 AM

Invoice : 0131-0922
Card No : XXXXXXXXXXX1111
Card Type: VISA

Auth ID: 999999
Trans ID: 21901234

AMOUNT: 24.95:
TAX 8.25% 2.06

SUBTOTAL: 27.01
TIP: 5.00

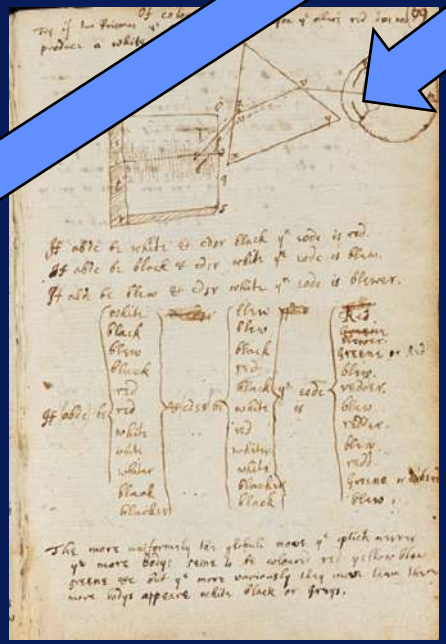
TOTAL: 32.01

X_____
PETE PETERSON

I AGREE TO PAY THE ABOVE AMOUNT ACCORDING TO THE CARD ISSUER AGREEMENT (MERCHANT AGREEMENT IF CREDIT VOUCHER)

Thank You

80%-90% of entire data is unstructured!



HTML

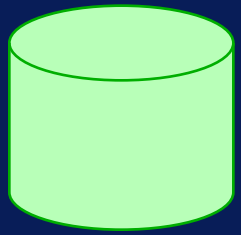
PPT

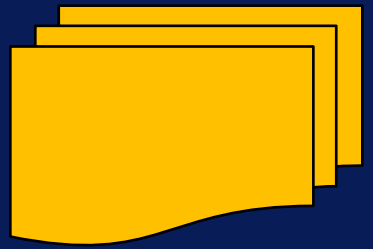
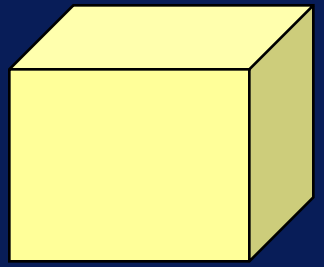
XML

PDF

Velocity

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



TRANSACTIONS

TWEETS

LIKES

BLOGS

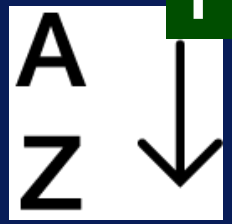
INTEGRATION

Data Acquisition

Storage

Retrieval

Cleaning



Processing

+

+

Tools

Data

Skilled People

?

?

Value

People

INTEGRATION

Unstructured Data

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

Challenges