

# Lesson 1

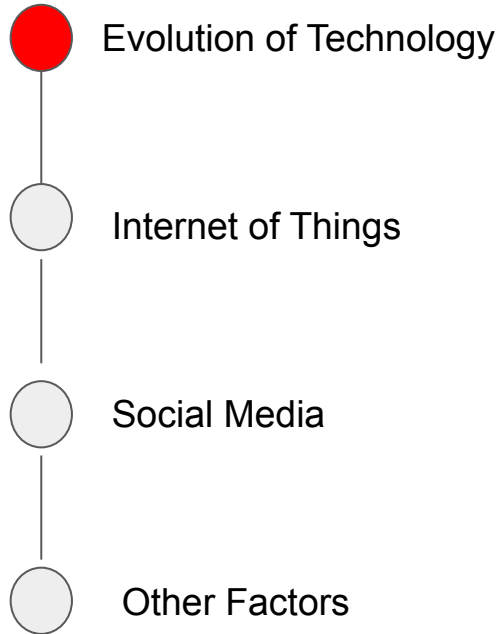
Data

“We are drowning in data, but  
starving for knowledge!”

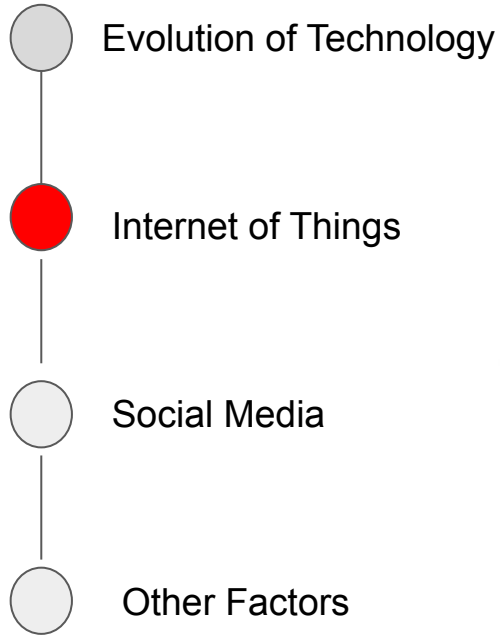
(John Naisbitt, 1982)



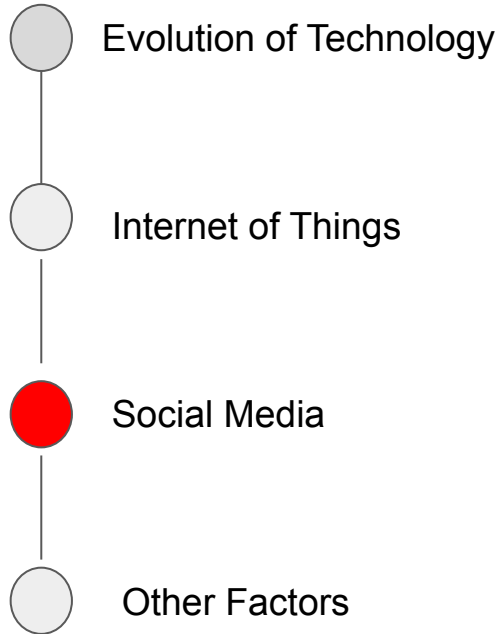
# Data Sources



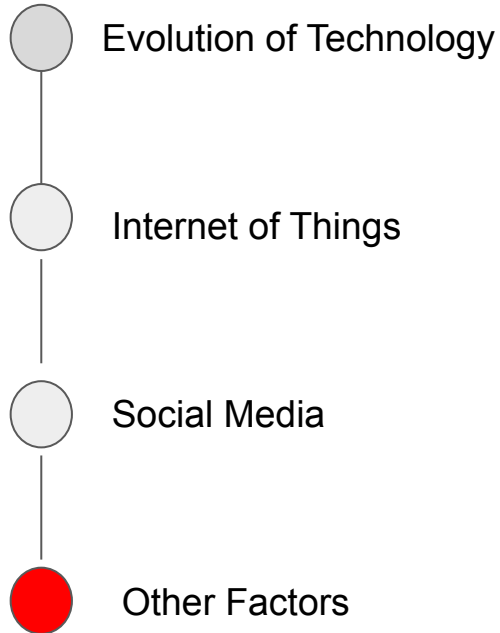
# Data Sources



# Data Sources



# Data Sources



Structured data

Databases

Semi-structured data

XML / JSON data

Email

Web pages

Unstructured data

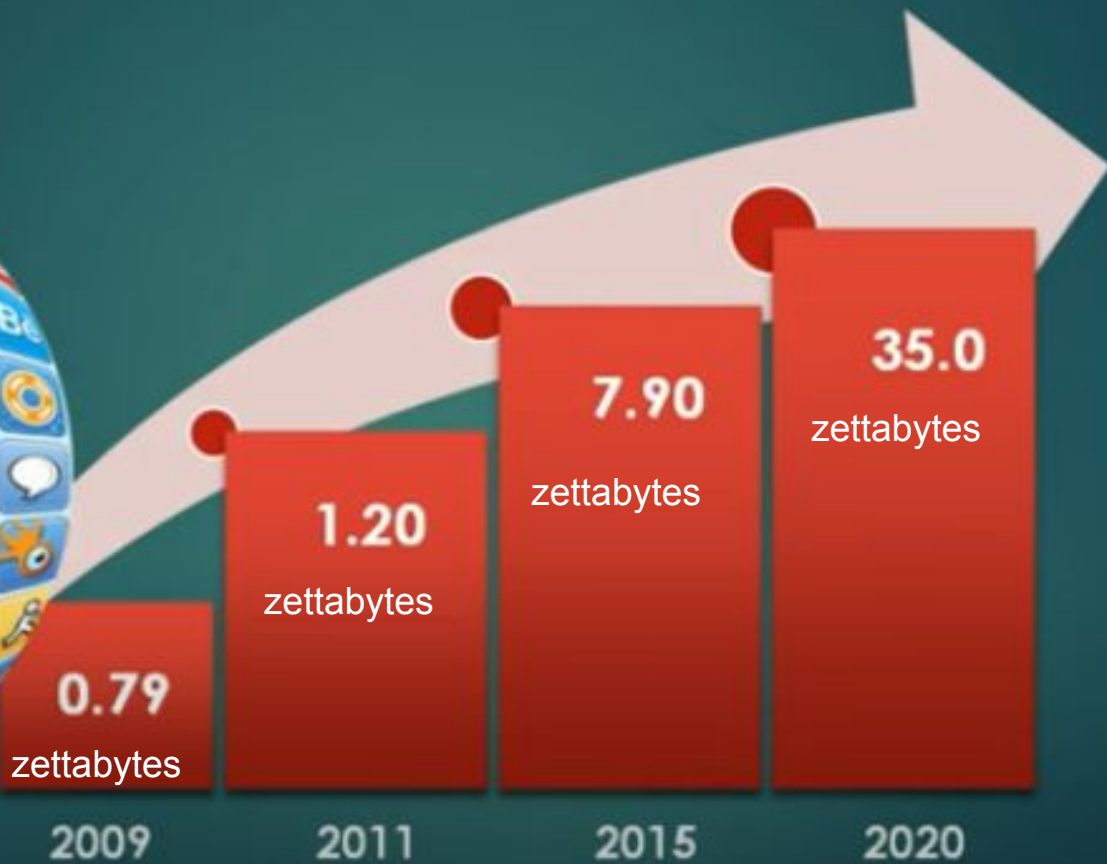
Audio

Video

Image data

Natural language

Documents





# What is Data Mining?





Walmart takes in approximately  
**\$36 million**  
from across 4300 US  
stores every day.

Walmart collects  
**2.5 petabytes**  
of unstructured data from  
1 million customers every  
hour.



# Big Data @ Walmart



## Walmart Big Data Facts and Figures



**245 million**  
customers visiting  
10,900 stores and  
10 active websites  
across the  
globe-Walmart is  
a name to reckon  
in the retail sector.

Walmart sees close to  
**300,000**  
social mentions every week.



# Why Walmart Always Stocks Up On Strawberry Pop-Tarts Before a Hurricane

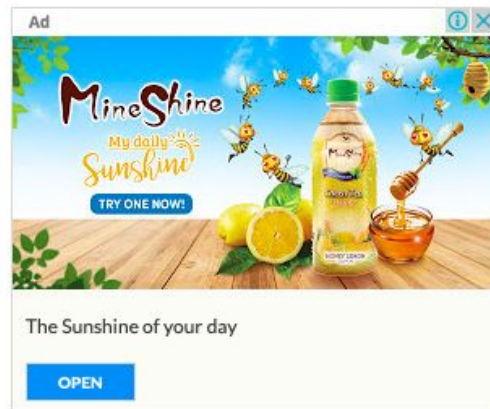
There's a practical reason the toaster pastries are so popular.



BY [MARIA CARTER](#) AUG 25, 2017



KELLOGG'S + GETTY IMAGES



# NEED FOR MACHINE LEARNING



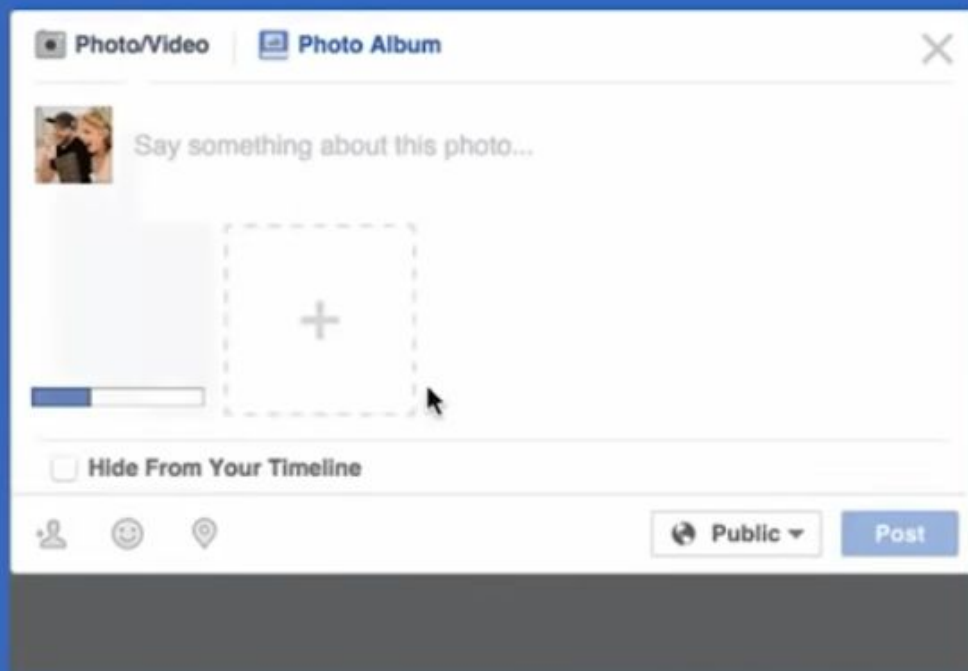
"Over 2.5 quintillion bytes of data are created every single day, and it's only going to grow from there. By 2020, it's estimated that 1.7MB of data will be created every second for every person on earth."

# How Netflix used big data and analytics to generate billions





# Facebook Tags

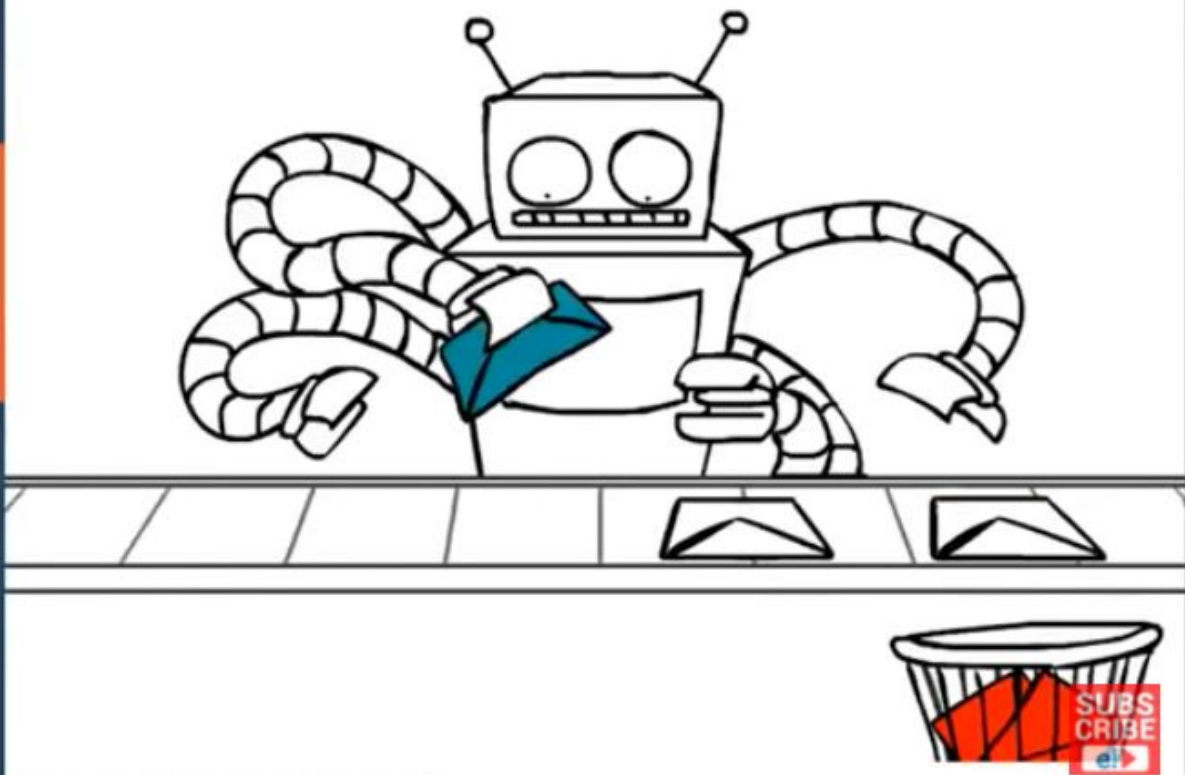




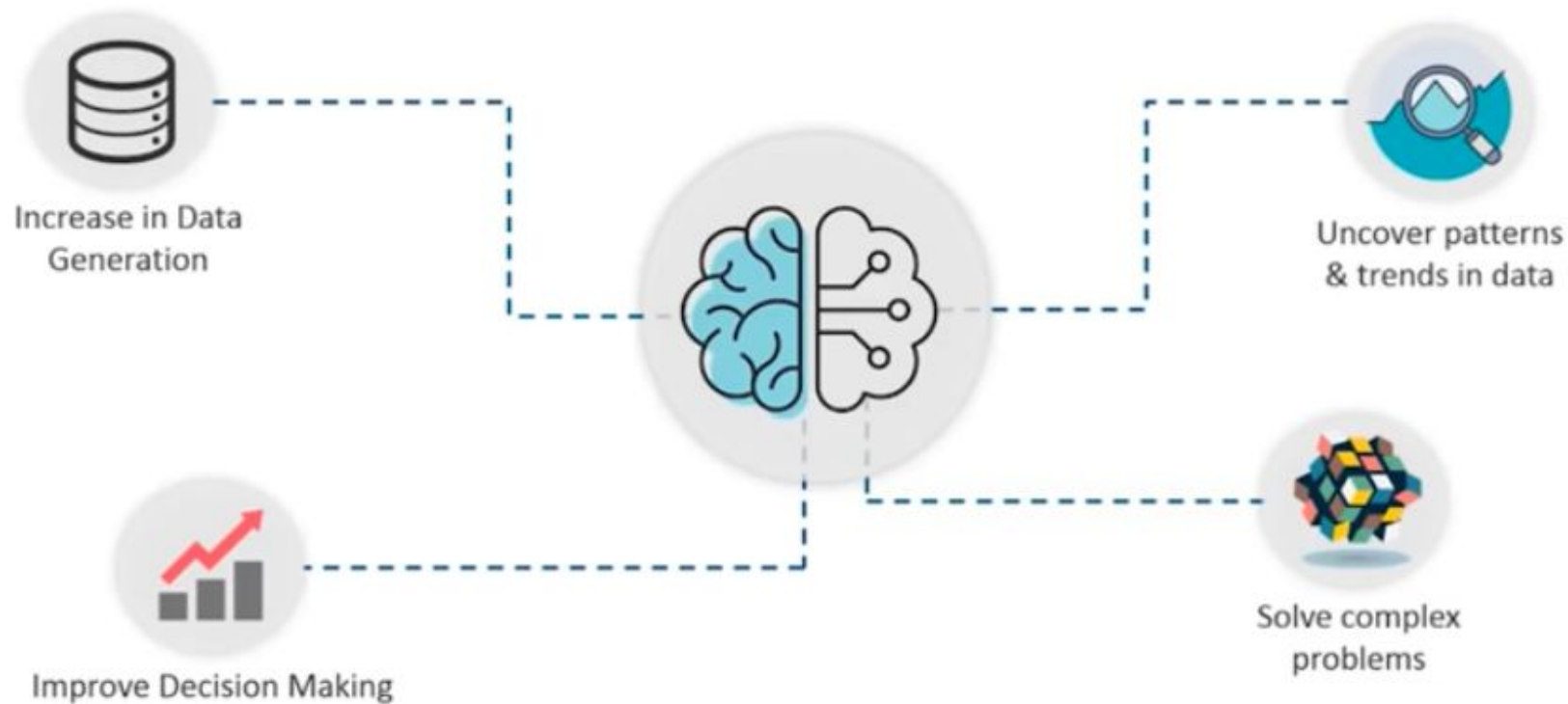
# Amazon Alexa



# Spam Filtering







### Unstructured data

The university has 5600 students.

John's ID is number 1, he is 18 years old and already holds a B.Sc. degree.

David's ID is number 2, he is 31 years old and holds a Ph.D. degree. Robert's ID is number 3, he is 51 years old and also holds the same degree as David, a Ph.D. degree.

### Semi-structured data

```
<University>
  <Student ID="1">
    <Name>John</Name>
    <Age>18</Age>
    <Degree>B.Sc.</Degree>
  </Student>
  <Student ID="2">
    <Name>David</Name>
    <Age>31</Age>
    <Degree>Ph.D. </Degree>
  </Student>
  ....
</University>
```

### Structured data

ID	Name	Age	Degree
1	John	18	B.Sc.
2	David	31	Ph.D.
3	Robert	51	Ph.D.
4	Rick	26	M.Sc.
5	Michael	19	B.Sc.