https://github.com/JyotinderSingh

WHAT LAUNCHED THE BIG DATA ERA

- A growing torrent of data
  - Billions of mobile devices
  - 30 billion pieces of content shared on facebook every month
- Cloud computing
  - On demand computing
    - dynamic & scalable data analysis

APPLICATIONS :

Big Data → Better Models → Precise Predictions

- Recommender Systems
- Personalized Marketting
  - eg. walmart
- Sentiment Analysis
  - eg. Product reviews
  - NLP
  - also called opinion mining

- mobile Advertising
  - GPS can allow location targetting
  - Home Depot coupons (on your phone) when you're near their outlet will make you visit them

https://github.com/JyotinderSingh

→ Data analysis to schedule flights

→ Biomedical Applications

↳ 2 billion Human genomes sequenced by 2025

↳ upto 40 exabytes of storage

↳ Personalized cancer treatment

↳ customize treatment & drug

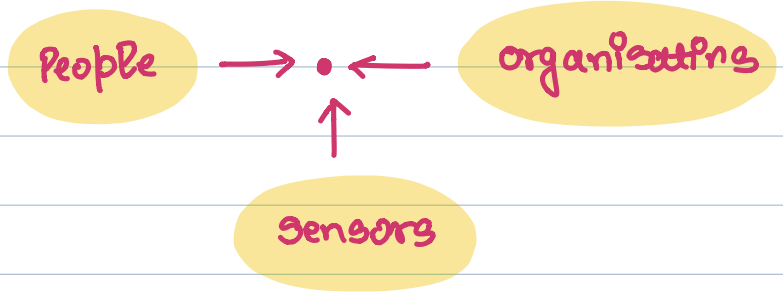
Problem → How to integrate data from different sources

→ Smart Cities

→ lower energy costs, pollution

→ Improve services, traffic, safety

HOW CAN BIG DATA HELP



* Stop Wildfires

* Precision Medicine

Area of medicine targetted towards an individual, by analysing their genetics, environment, & daily activities to help detect a problem early on, provide the right drug at the right dose that is suitable just for her

https://github.com/JyotinderSingh

THREE MAJOR SOURCES OF DATA :

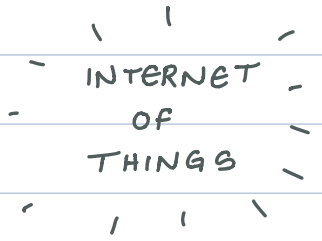
- MACHINES
- PEOPLE
- ORGANISATIONS

WHAT MAKES A DEVICE SMART?

→ connect to other devices/networks

→ collect & analyse data autonomously

→ Provide environmental context



increasing the number of machines that sense **X** data collected by each device

= MACHINES
↓
Biggest Source

ADVANTAGES OF MACHINE GENERATED DATA :

RDBMS

Data moved to computational space

IN-SITU

Bring computation to data

real time notification
+
real time action

https://github.com/JyotinderSingh

SCADA : Supervisory control and Data Acquisation

→ for remote monitoring + control of industrial processes

→ reduce waste, improve efficiency

→ identify trends, patterns, & anomalies.

BIG DATA GENERATED BY PEOPLE :

* Text Heavy
* unstructured → doesn't conform to a data model

VELOCITY : VOLUME AND FAST GENERATION OF DATA

PREDICT CUSTOMER BEHAVIOUR

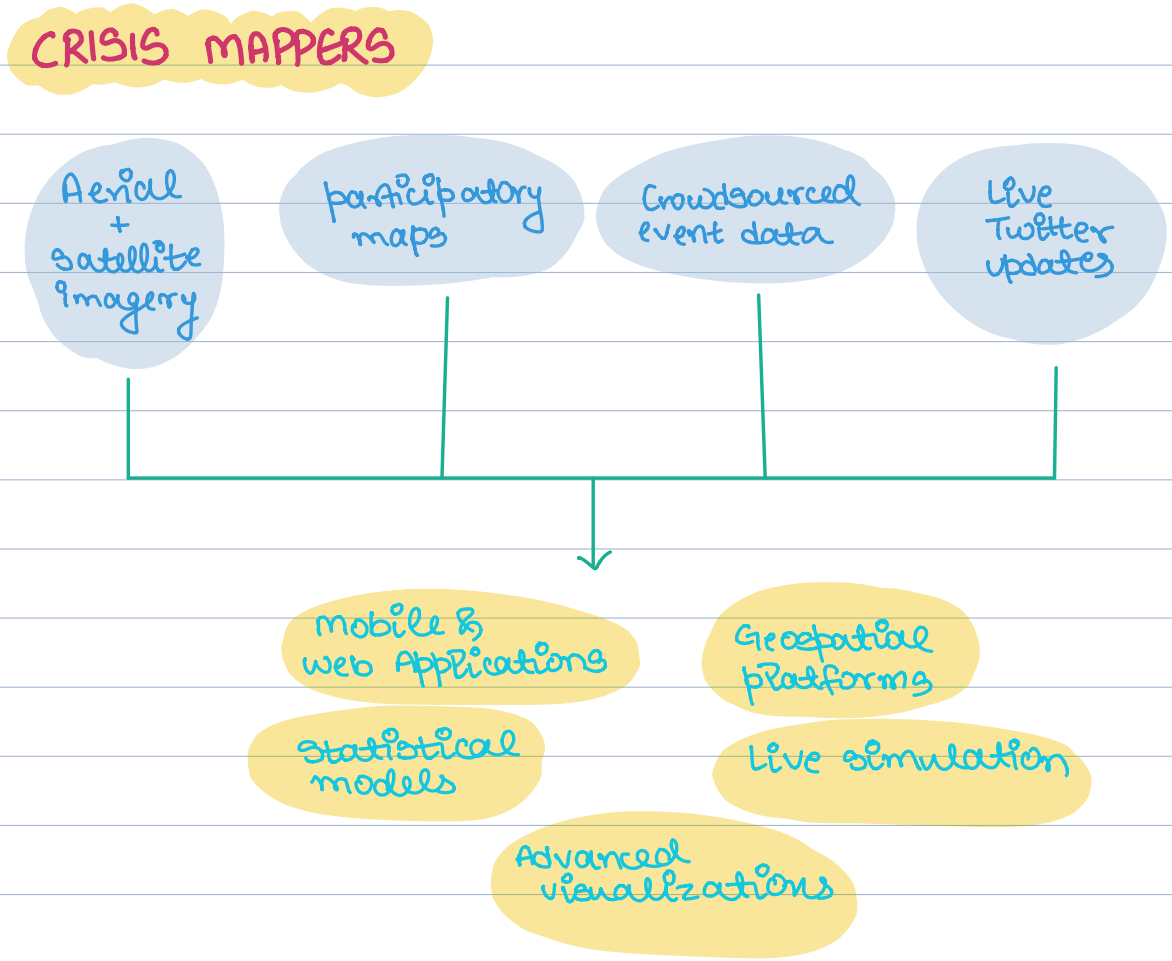
Personalized recommendations
↓
happier customers
↓
higher profits

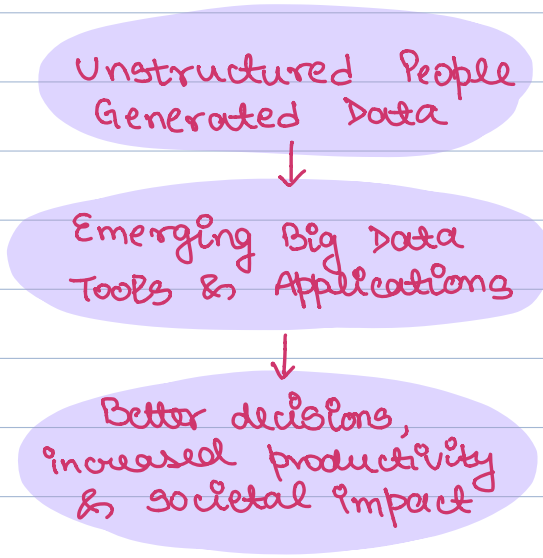
COLLECTIVE DISASTER RESPONSE

Predict & evacuate

find best escape routes

https://github.com/JyotinderSingh





https://github.com/JyotinderSingh

DATA SILOS

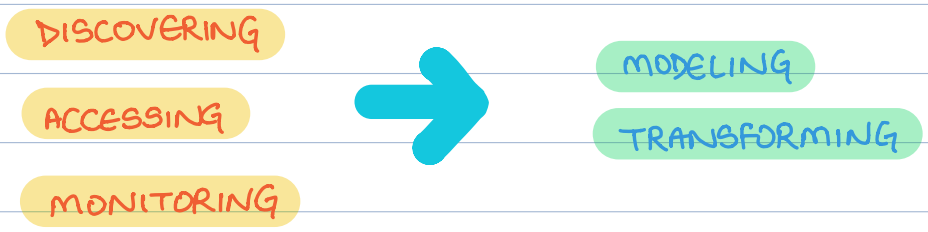
- Hindered opportunity generation
  - hindered growth of large scale pattern recognition
  - as no one in the system has access to all that data that the organisation owns
- unconnected islands of information

Each dataset is compartmentalized. If such silos are left untouched, organisations risk having outdated, unsynchronized, & even invisible datasets.

DATA INTEGRATION:

Turning complex data into something useful.

DATA INTEGRATION PROCESS



https://github.com/JyotinderSingh

WHY DO WE NEED DATA INTEGRATION?

→ Data comes in all shapes & sizes

→ reduces data complexity

→ Increase data availability

→ unify your data system

Increase data collaboration

↓

Add value to your big data

SPATIAL DATA

+

NON SPATIAL DATA

↓

more accurate simulations