ORACLE
Certified Professional

Oracle Certified Professional Java Programmer

Microsoft Certified Professional

sohailimran@yahoo.com

Sohail IMRAN سهيل عمران

Intro

Ref: Web

what?

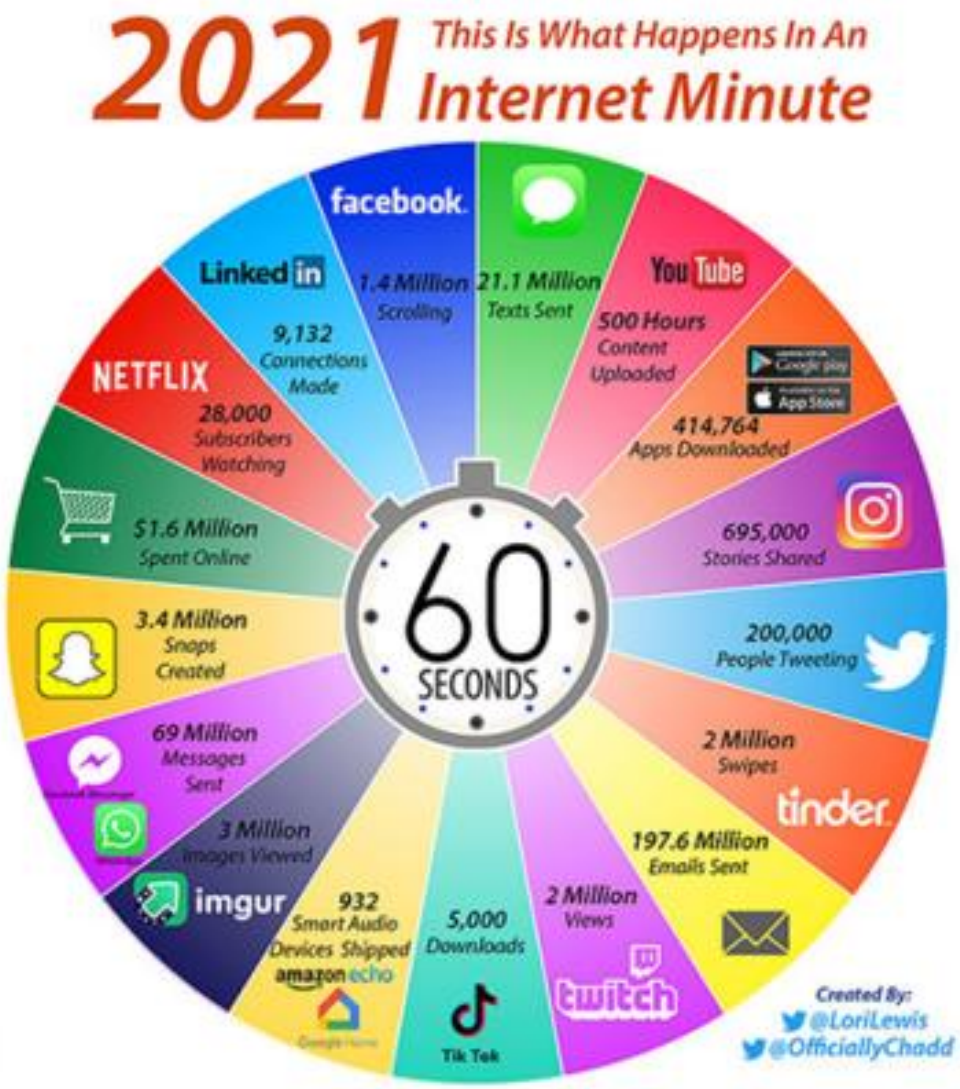
No single standard definition…

"Big Data" is data whose scale, diversity, and complexity require new architecture, techniques, algorithms, and analytics to manage it and extract value and hidden knowledge from it…

Big data is the term for a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications.

per min?


2020 This Is What Happens In An Internet Minute
60 SECONDS
facebook
1.3 Million Logging In
19 Million Texts Sent
YouTube
4.7 Million Videos Viewed
400,000 Apps Downloaded
694,444 Scrolling Instagram
194,444 People Tweeting
1.6 Million Swipes
190 Million Emails Sent
1.2 Million Views
1,400 Downloads
305 Smart Speakers Shipped
2.5 Million Images Viewed
59 Million Messages Sent
2.5 Million Snaps Created
$1.1 Million Spent Online
764,000 Hours Watched
4.1 Million Search Queries
Google
NETFLIX
imgur
amazon echo
Tik Tok
twitch
tinder
Created By:
@LoriLewis
@OfficiallyChadd


2021 This Is What Happens In An Internet Minute
60 SECONDS
facebook
1.4 Million Scrolling
21.1 Million Texts Sent
YouTube
500 Hours Content Uploaded
414,764 Apps Downloaded
695,000 Stories Shared
200,000 People Tweeting
2 Million Swipes
197.6 Million Emails Sent
2 Million Views
5,000 Downloads
932 Smart Audio Devices Shipped
3 Million Images Viewed
69 Million Messages Sent
3.4 Million Snaps Created
$1.6 Million Spent Online
28,000 Subscribers Watching
9,132 Connections Made
LinkedIn
NETFLIX
imgur
amazon echo
Tik Tok
twitch
tinder
Created By:
@LoriLewis
@OfficiallyChadd

Why Big Data is a Big Deal

Convergence of Factors : Data, Technology, and Thinking

Machine Learning

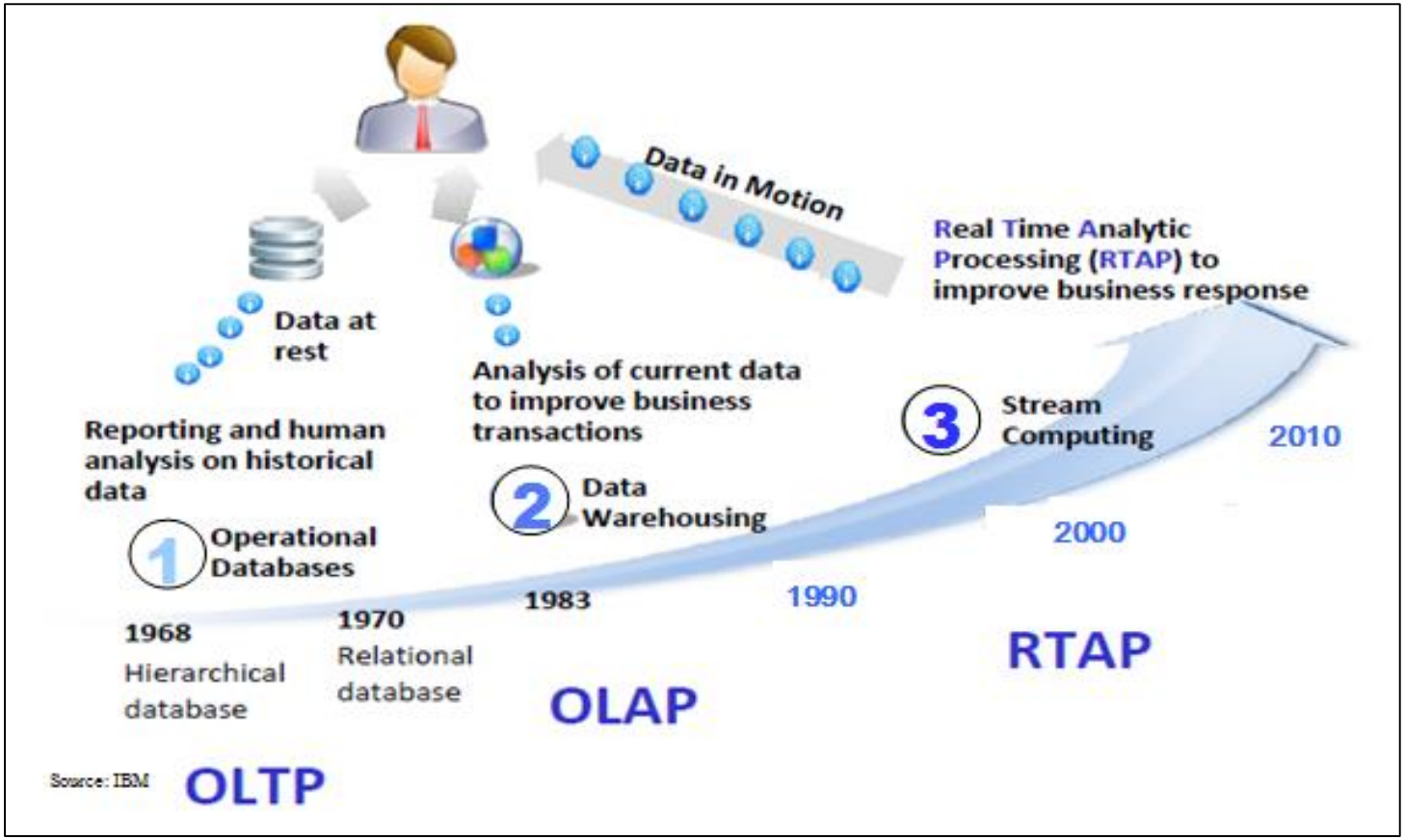
Deluge of Data

Compute Power

Data-based decisions

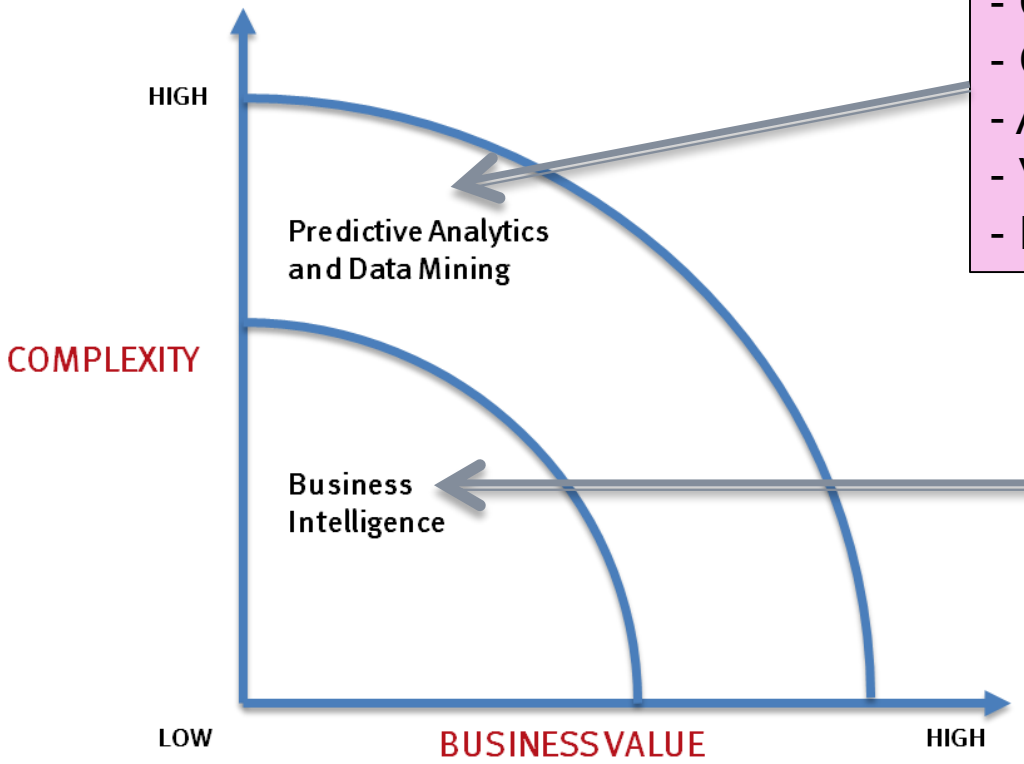
Characteristics

harnessing big data



- **OLTP:** Online Transaction Processing (DBMSs)
- **OLAP:** Online Analytical Processing (Data Warehousing)
- **RTAP:** Real-Time Analytics Processing (Big Data Architecture & technology)

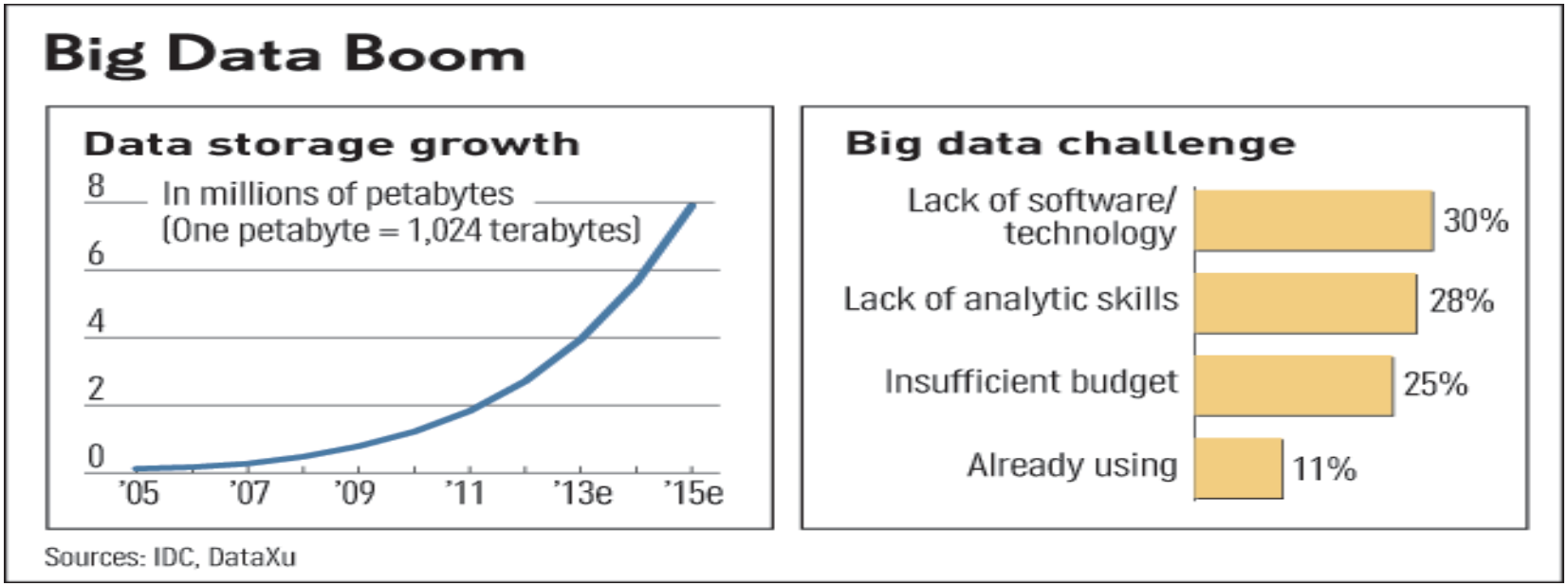
what's driving big data



- Optimizations and predictive analytics
- Complex statistical analysis
- All types of data, and many sources
- Very large datasets
- More of a real-time

- Ad-hoc querying and reporting
- Data mining techniques
- Structured data, typical sources
- Small to mid-size datasets

Challenges



- **The Bottleneck is in technology**
  - New architecture, algorithms, techniques are needed
  - Data management and analysis
- **Also in technical skills**
  - Experts in using the new technology and dealing with big data

large scale data processing

- **Traditional approach**
  - Database and data warehousing systems
  - Well-defined structure
  - Small enough data

- **Big data**
  - Data sets not suitable for databases
    - E.g., Internet data crawled by Google, Yahoo!, Facebook, …
  - May need near real-time (streaming) analysis
    - Different from data warehousing
  - Different programming paradigm

application

tools


STORAGE
splunk>hunk
hadoop
mongoDB
cassandra
RainStor
ANALYTICS
BLOCKCHAIN
kafka
Spark
KNIME
MINING
presto
rapidminer
elasticsearch
VISUALIZATION
+ableau
plotly
BIG DATA

benefits


Cost reduction
BIG DATA ANALYTICS
Faster, better decision making
New products & services