PULCHRA
Participatory Urban Learning Community Hubs through Research and Activation

OPEN UNIVERSITY OF CYPRUS
www.ouc.ac.cy

Science in the City

Building Participatory Urban Learning Community Hubs through Research and Activation

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824466

PULCHRA
Participatory Urban Learning Community Hubs through Research and Activation

OPEN UNIVERSITY OF CYPRUS
www.ouc.ac.cy

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824466

INTRODUCTION

- Big Data may well be the Next Big Thing in the IT world.
- Big data burst upon the scene in the first decade of the 21st century.
- The first organizations to embrace it were online and startup firms. Firms like Google, eBay, LinkedIn, and Facebook were built around big data from the beginning.
- Like many new information technologies, big data can bring about dramatic cost reductions, substantial improvements in the time required to perform a computing task, or new product and service offerings.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824466

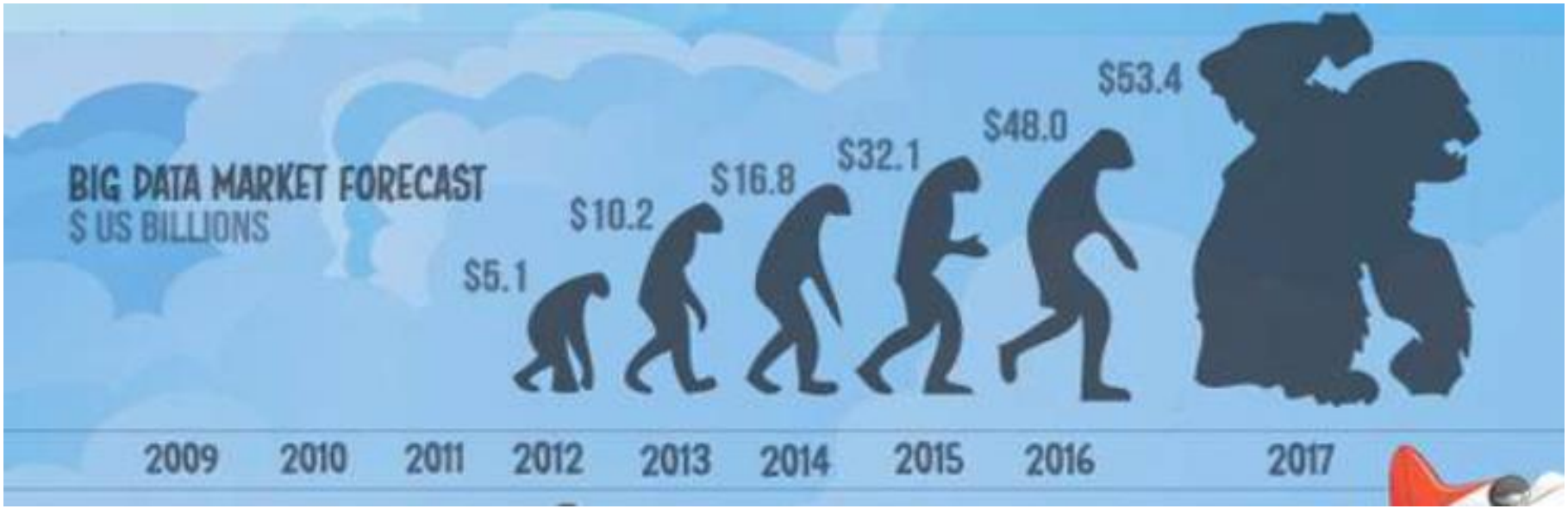
WHAT IS BIG DATA?

- ‘Big Data’ is similar to ‘small data’, but bigger in size
- but having data bigger it requires different approaches: Techniques, tools and architecture
- an aim to solve new problems or old problems in a better way
- Big Data generates value from the storage and processing of very large quantities of digital information that cannot be analyzed with traditional computing techniques.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 824466

WHAT IS BIG DATA

- Walmart handles more than 1 million customer transactions every hour.
- Facebook handles 40 billion photos from its user base.
- Decoding the human genome originally took 10years to process; now it can be achieved in one week.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824466

THREE CHARACTERISTICS OF BIG DATA V3S

Volume
- Data quantity

Velocity
- Data Speed

Variety
- Data Types

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824466

1st Character of Big Data

Volume

- A typical PC might have had 10 gigabytes of storage in 2000.
- Today, Facebook ingests 500 terabytes of new data every day.
- Boeing 737 will generate 240 terabytes of flight data during a single flight across the US.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824466

2ND CHARACTER OF BIG DATA

VELOCITY

- Clickstreams and ad impressions capture user behaviour at millions of events per second
- high-frequency stock trading algorithms reflect market changes within microseconds
- machine to machine processes exchange data between billions of devices
- infrastructure and sensors generate massive log data in real-time
- on-line gaming systems support millions of concurrent users, each producing multiple inputs per second.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824466

3rd Character of Big Data

Variety

- Big Data isn't just numbers, dates, and strings. Big Data is also geospatial data, 3D data, audio and video, and unstructured text, including log files and social media.
- Traditional database systems were designed to address smaller volumes of structured data, fewer updates or a predictable, consistent data structure.
- Big Data analysis includes different types of data

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824466

STORING BIG DATA

- **Analyzing your data characteristics**
- Selecting data sources for analysis
- Eliminating redundant data
- Establishing the role of NoSQL

- **Overview of Big Data stores**
- Data models: key value, graph, document, column-family
- Hadoop Distributed File System
- HBase
- Hive

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824466

Processing Big Data

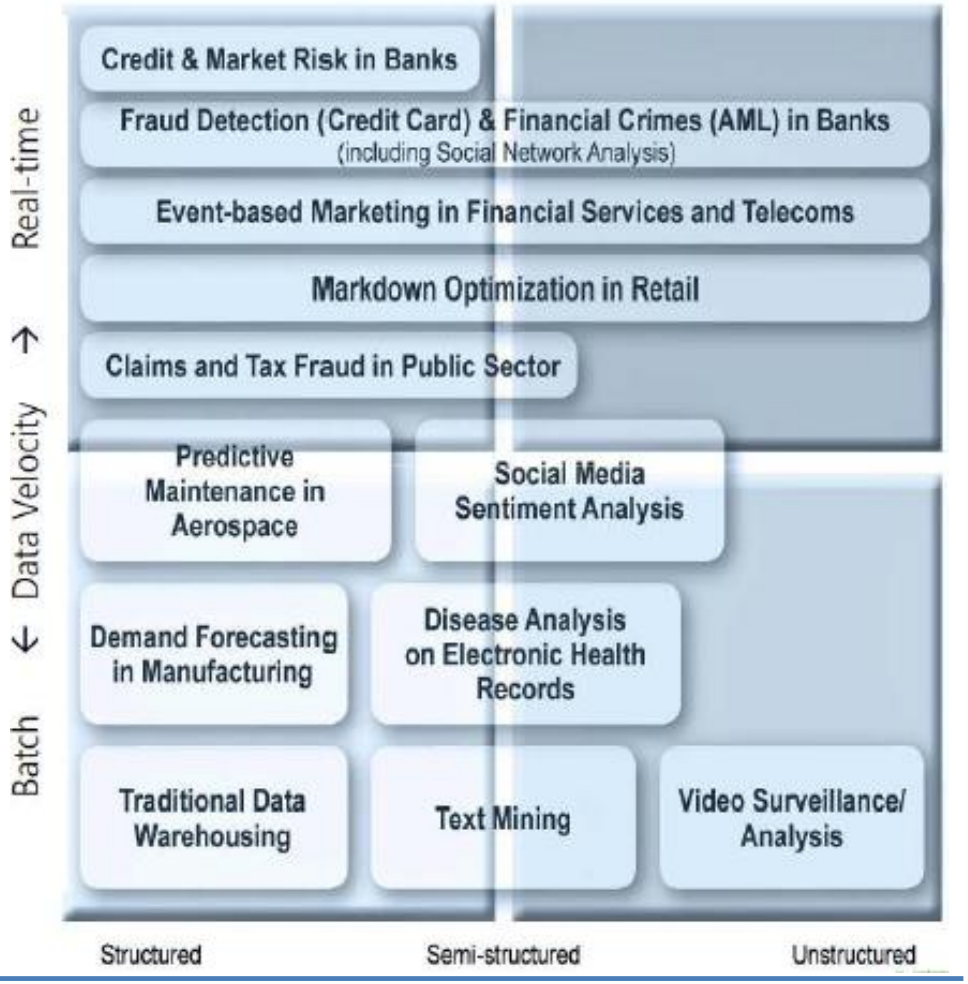
- **Integrating disparate data stores**
  - Mapping data to the programming framework
  - Connecting and extracting data from storage
  - Transforming data for processing
  - Subdividing data in preparation for Hadoop MapReduce

- **Employing Hadoop MapReduce**
  - Creating the components of Hadoop MapReduce jobs
  - Distributing data processing across server farms
  - Executing Hadoop MapReduce jobs
  - Monitoring the progress of job flows

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824466

THE STRUCTURE OF BIG DATA

- Structured
  - Most traditional data sources
- Semi-structured
  - Many sources of big data
- Unstructured
  - Video data, audio data



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824466

WHY BIG DATA

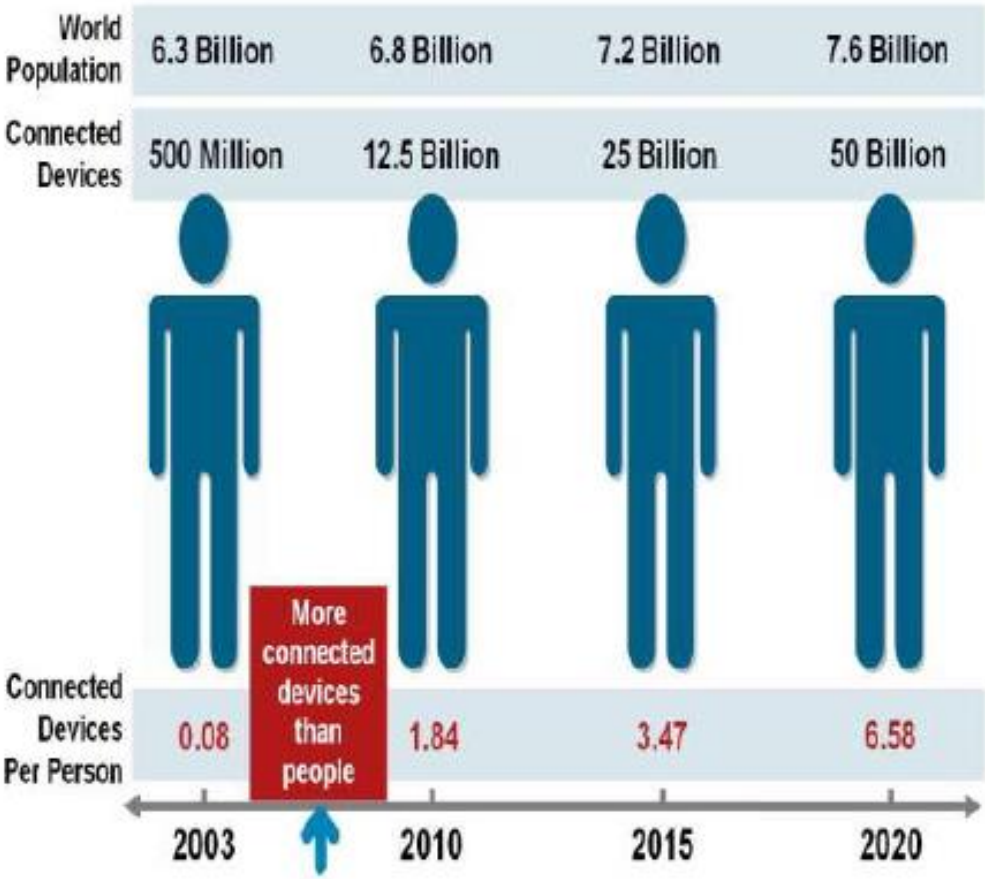
- Growth of Big Data is needed
  - Increase of storage capacities
  - Increase of processing power
  - Availability of data(different data types)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824466

WHY BIG DATA

- FB generates 10TB daily
- Twitter generates 7TB of data Daily
- IBM claims 90% of today's stored data was generated in just the last two years.

Figure 1. The Internet of Things Was "Born" Between 2008 and 2009



Source: Cisco IBSG, April 2011

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824466

How Is Big Data Different?

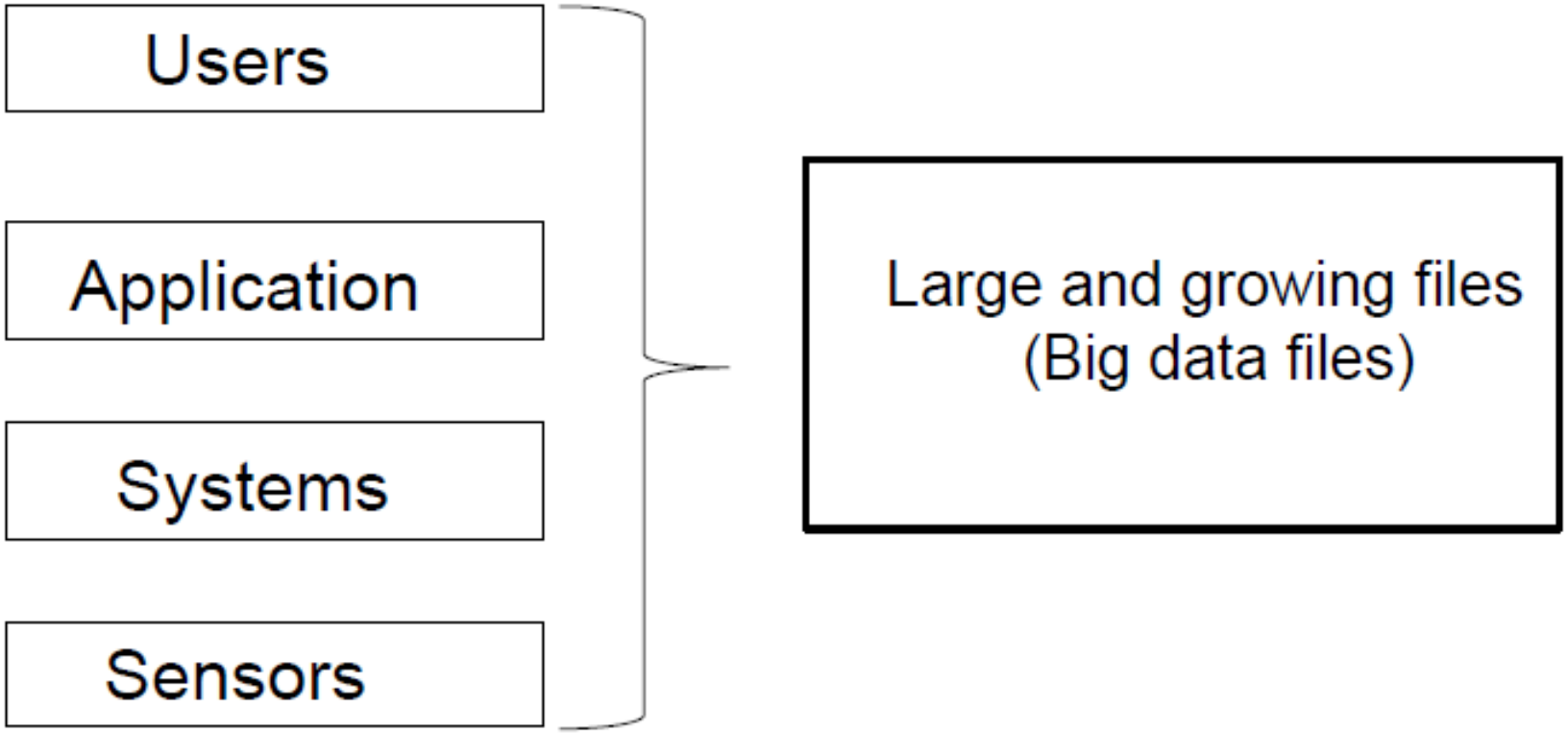
1. Automatically generated by a machine (e.g. Sensor embedded in an engine)

2. Typically an entirely new source of data (e.g. Use of the internet)

3. Not designed to be friendly (e.g. Text streams)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824466

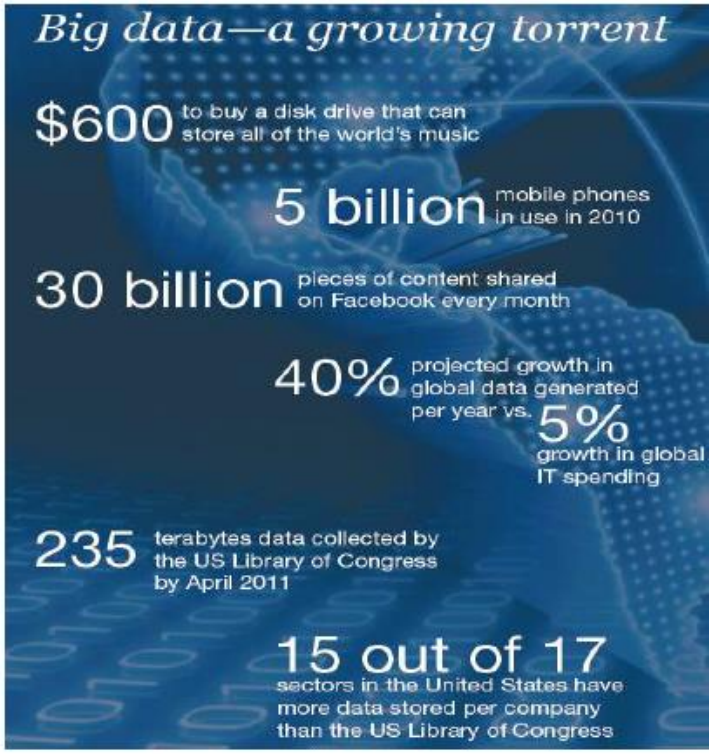
BIG DATA SOURCES



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824466

DATA GENERATION POINTS EXAMPLES

- Mobile Devices
- Microphones
- Readers/Scanners
- Science facilities
- Programs/ Software
- Social Media
- Cameras



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824466

Big Data Analytics

- Examining large amount of data
- Appropriate information (about data)
- Identification of hidden patterns, unknown correlations
- Better business decisions: strategic and operational
- Effective marketing, customer satisfaction, increased revenue

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824466

Types of tools used in Big-Data

- Where processing is **hosted**?
  - Distributed Servers / Cloud (e.g. Amazon EC2)

- Where data is **stored**?
  - Distributed Storage (e.g. Amazon S3)

- What is the **programming model**?
  - Distributed Processing (e.g. MapReduce)

- How data is **stored & indexed**?
  - High-performance schema-free databases (e.g. MongoDB)

- What operations are performed on data?
  - Analytic / Semantic Processing

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824466

Application Of Big Data analytics

Smarter Healthcare

Multi-channel sales

Homeland Security

Telecom

Traffic Control

Trading Analytics

Manufacturing

Search Quality

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824466

Risks of Big Data

- Will be so overwhelmed
- Need the right people and solve the right problems
- Costs escalate too fast
- Isn't necessary to capture 100%
- Many sources of big data is privacy
- self-regulation (data compression)
- Legal regulation

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824466

How Big data impacts on IT

- Big data is a troublesome force presenting opportunities with challenges to IT organizations.
  - By 2015 4.4 million IT jobs in Big Data ; 1.9 million is in US itself
  - In 2017, Data scientist's was No. 1 Job in the Harvard's ranking.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824466

Benefits of Big Data

- Real-time big data isn't just a process for storing petabytes or exabytes of data in a data warehouse, It's about the ability to make better decisions and take meaningful actions at the right time.

- Fast forward to the present and technologies like Hadoop give you the scale and flexibility to store data before you know how you are going to process it.

- Technologies such as MapReduce, Hive and Impala enable you to run queries without changing the data structures underneath.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824466