Big Data and Smart Cities Applications

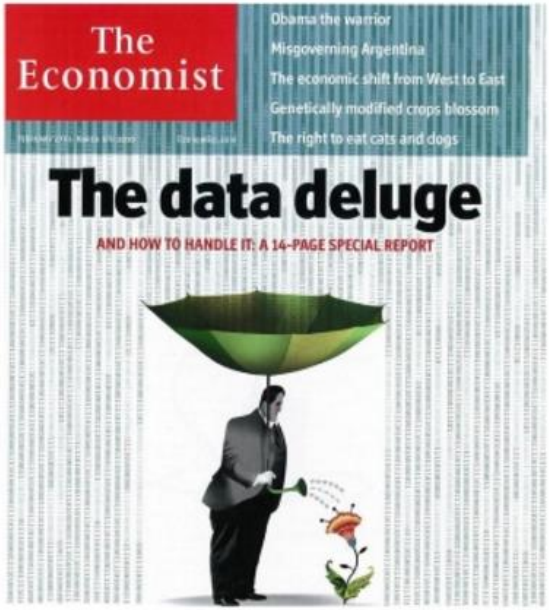
Distributed Computing Architectures And Environmental Science Applications

University of Ferrara - Department of Economics and Management

Maria Cristina Vistoli, Andrea Ferraro, Barbara Martelli
Istituto Nazionale Fisica Nucleare, CNAF Bologna

June 6th 2016

Big Data is going BIGGER !!!



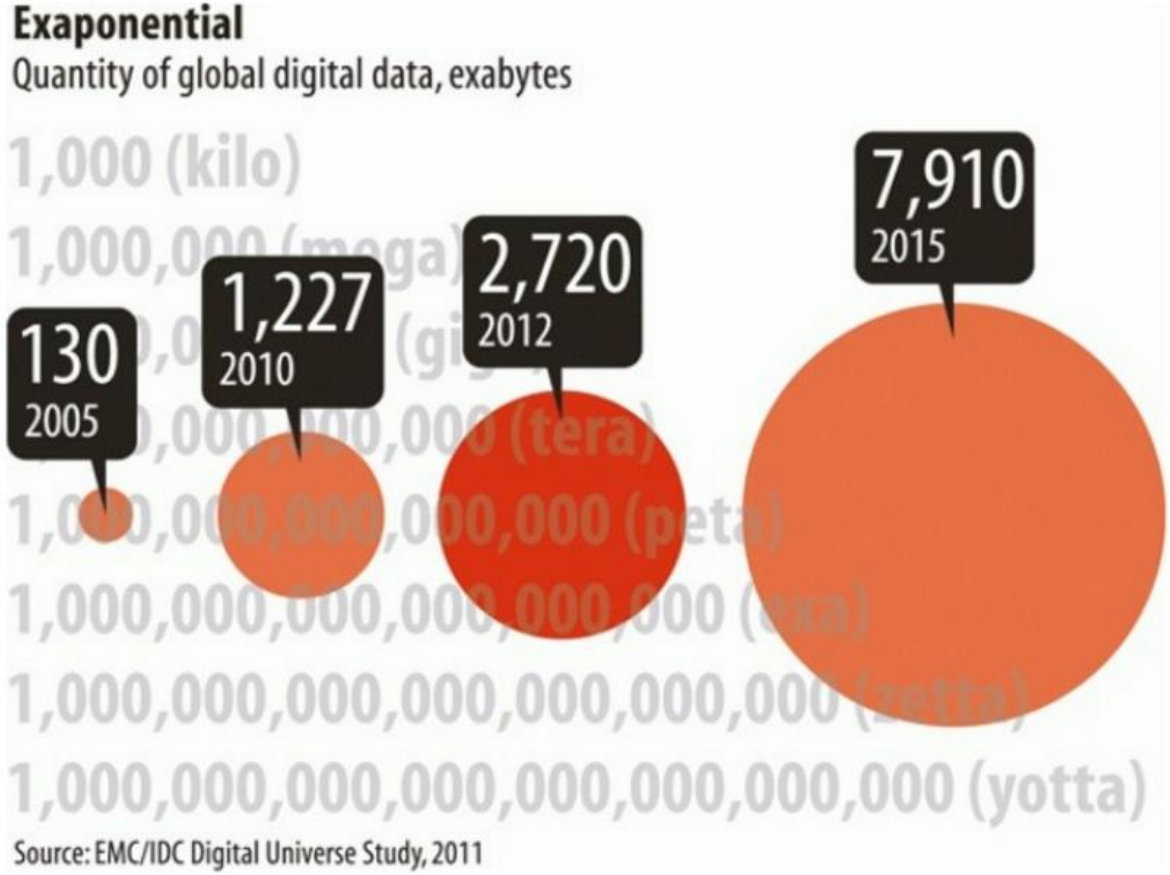
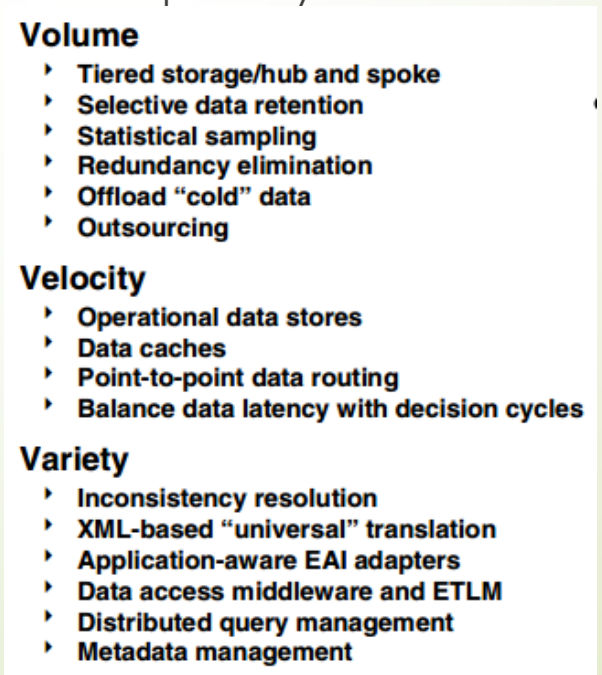


Image courtesy: the Economist

Big Data (definition)

- International Journal of Internet Science:
  - **Big Data** is a loosely defined term used to describe data sets so large and complex that they become awkward to work with using standard statistical software. (International Journal of Internet Science, 2012, 7(1),1-5)
- Wikipedia:
  - **Big Data** is a term for data sets that are so large or complex that traditional data processing applications are inadequate. Challenges include analysis, capture, data curation, search, sharing, storage, transfer, visualization, querying, updating and information privacy.
- **Big Data** = the 5Vs:



The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015
4.4 MILLION IT JOBS
will be created globally to support big data, with 1.9 million in the United States

Volume
SCALE OF DATA

40 ZETTABYTES
[40 TRILLION GIGABYTES]
of data will be created by 2020, an increase of 300 times from 2005

It's estimated that
2.5 QUINTILLION BYTES
[2.3 TRILLION GIGABYTES]
of data are created each day

6 BILLION PEOPLE
have cell phones

Most companies in the U.S. have at least
100 TERABYTES
[100,000 GIGABYTES]
of data stored

Variety
DIFFERENT FORMS OF DATA

As of 2011, the global size of data in healthcare was estimated to be
150 EXABYTES
[161 BILLION GIGABYTES]

By 2014, it's anticipated there will be
420 MILLION WEARABLE, WIRELESS HEALTH MONITORS

4 BILLION+ HOURS OF VIDEO
are watched on YouTube each month

30 BILLION PIECES OF CONTENT
are shared on Facebook every month

400 MILLION TWEETS
are sent per day by about 200 million monthly active users

Velocity
ANALYSIS OF STREAMING DATA

The New York Stock Exchange captures
1 TB OF TRADE INFORMATION
during each trading session

Modern cars have close to
100 SENSORS
that monitor items such as fuel level and tire pressure

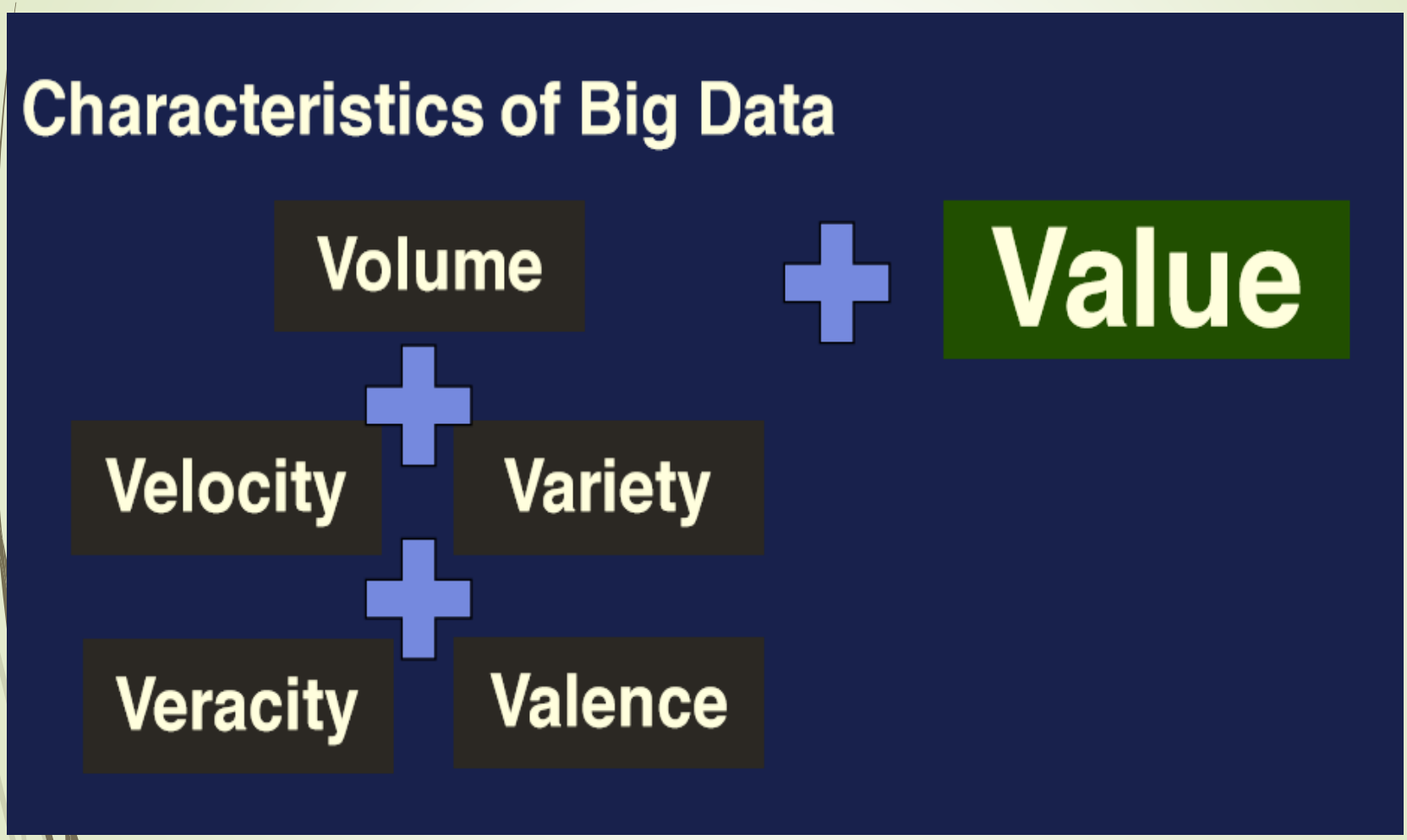
By 2016, it is projected there will be
18.9 BILLION NETWORK CONNECTIONS
– almost 2.5 connections per person on earth

Veracity
UNCERTAINTY OF DATA

1 IN 3 BUSINESS LEADERS
don't trust the information they use to make decisions

Poor data quality costs the US economy around
$3.1 TRILLION A YEAR

in one survey were unsure of how much of their data was inaccurate


Characteristics of Big Data
Volume
+
Value
Velocity
Variety
Veracity
Valence

Data Torrent

Computing
Anytime, Anywhere

Big Data Era

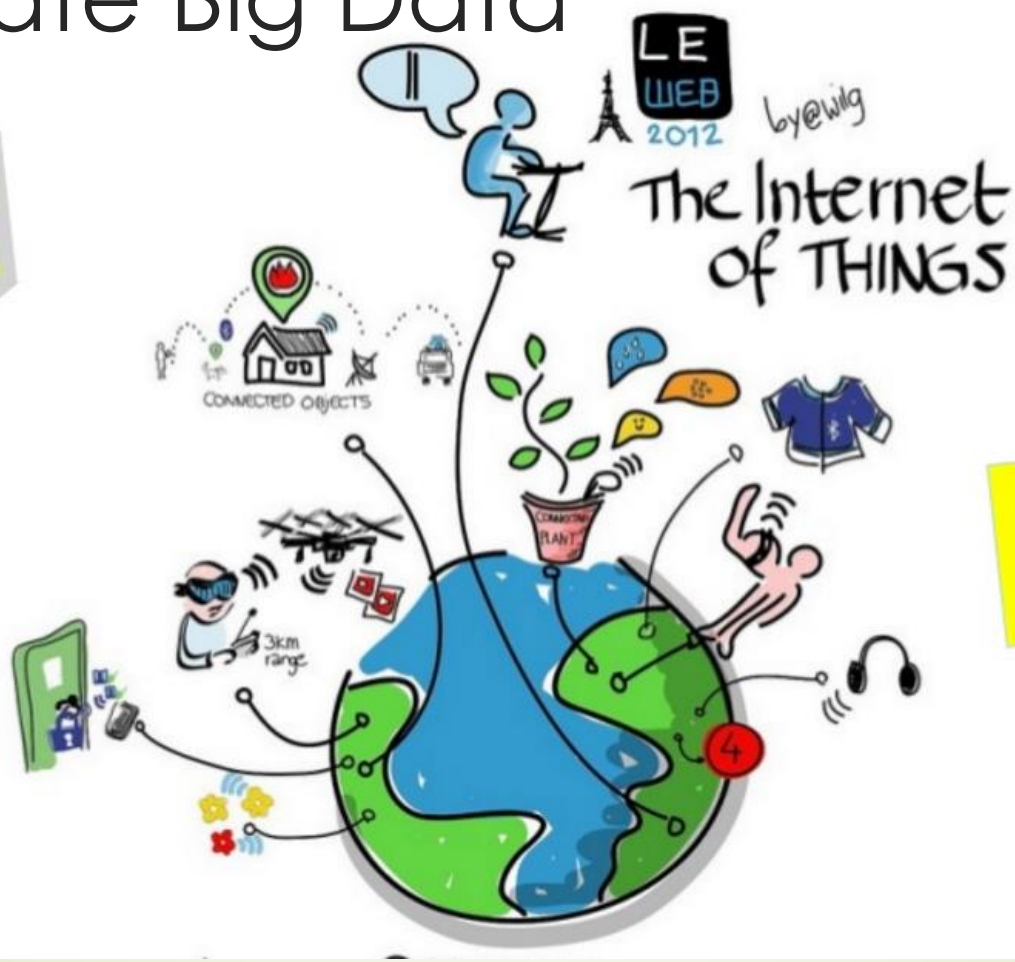
Smart cities (definition)

- Wikipedia:
  - A **Smart City** is an urban development vision to integrate multiple ICT solutions in a secure fashion to manage all city's assets
  - The goal of building a **Smart City** is to improve quality of life by using technology to improve the efficiency of services and meet residents' needs.
  - **Smart City** allows city officials to interact directly with the community and the city infrastructure and to monitor what is happening in the city, how the city is evolving, and how to enable a better quality of life.
  - A **Smart City,** use sensors integrated with real-time monitoring systems. Data are collected from citizens and devices - then processed and analyzed.

Smart cities (IoT and people) generate Big Data

Data

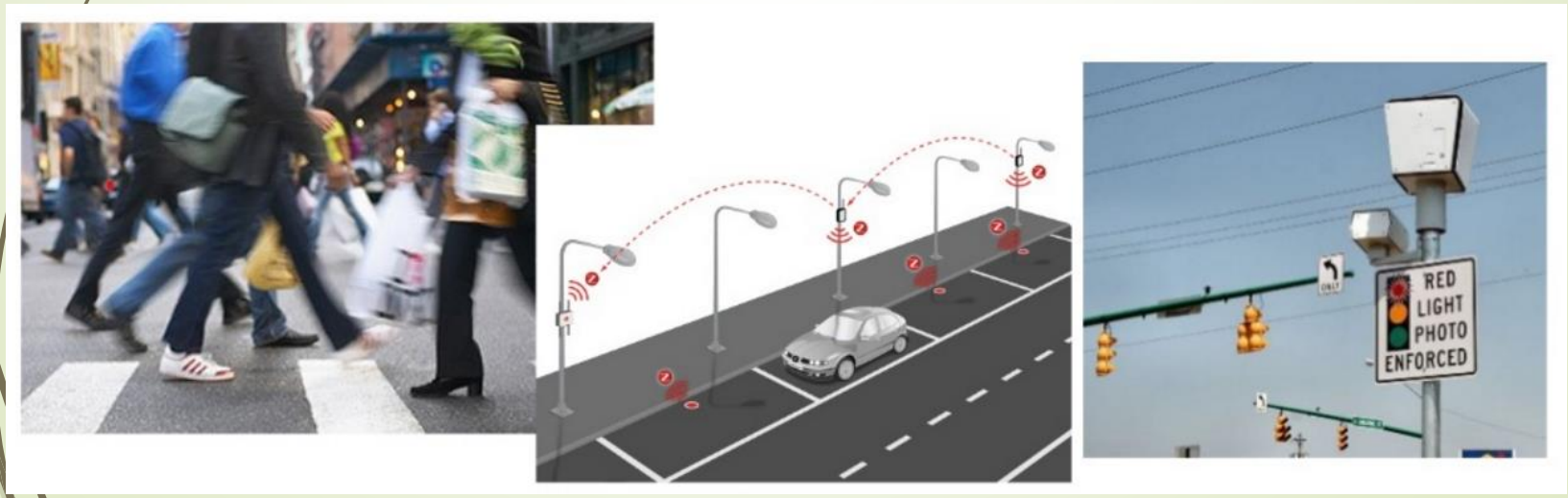
Semantics



Social networks

Data generated by Smart Cities

- Real-time data
- Heterogeneous/distributed data
- Re-usable, modular, flexible datasets
- Different type sources of data (from IoT or persons)
- Dynamicity: require context-aware solutions



Big Data for Smart Cities

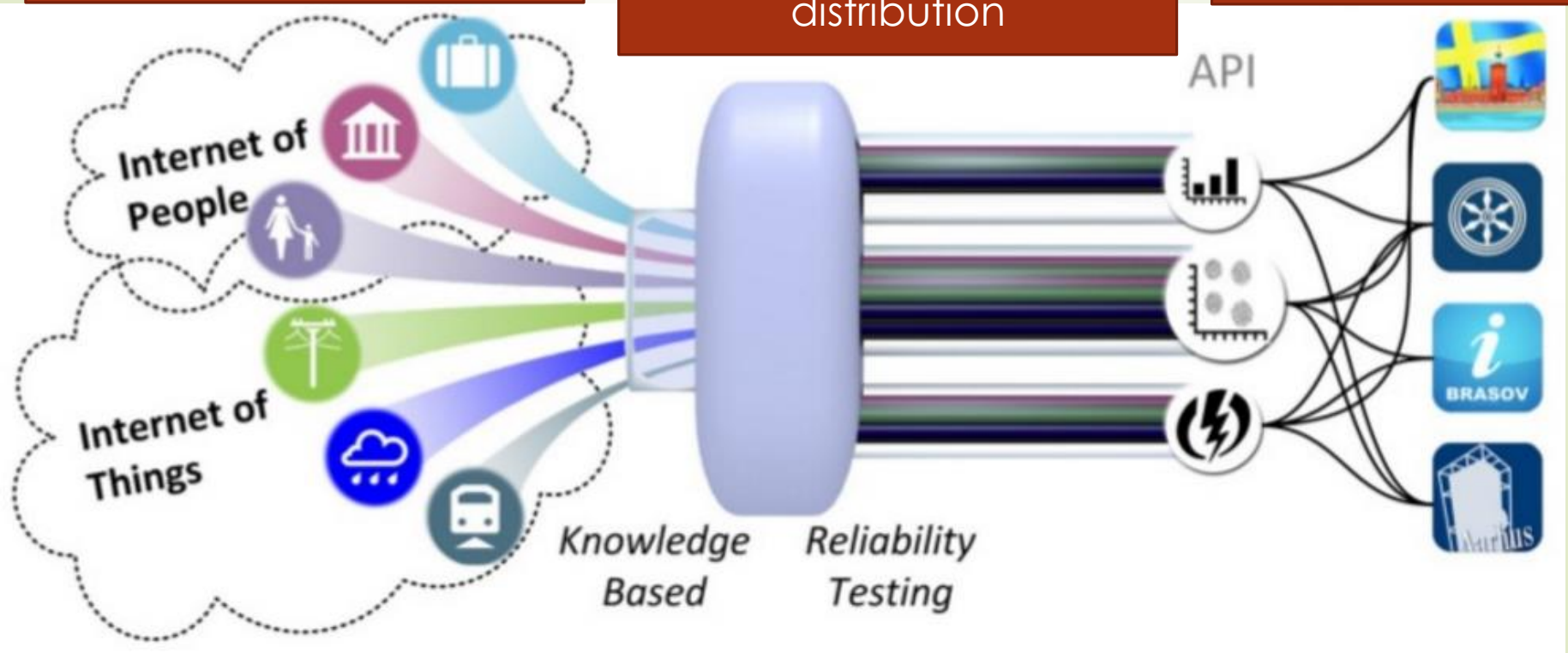
- Empower citizen
- Provide new business for SMEs
- Better governance and public services
- Smart centralized monitoring and control
- Better energy efficiency and eco-sustainables models
- New healthcare, eldery-care
- Intelligent transport
- Intelligent/coordinated cooling/heating of city buildings

Big Data for Smart Cities Applications

Big Data producers

Big Data gathering, processing, analysis, distribution

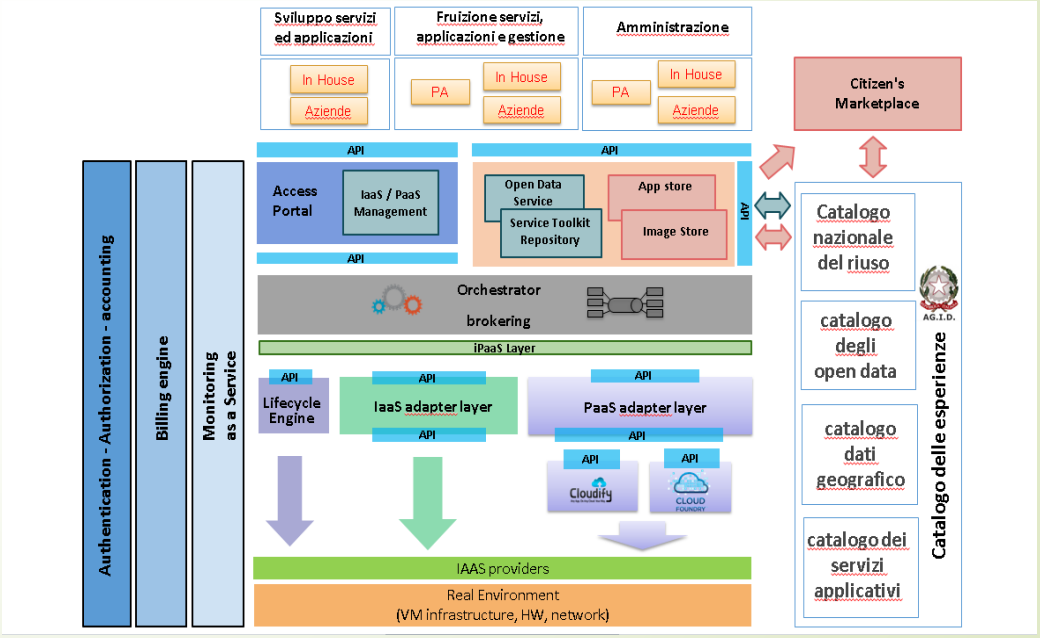
Smart Cities Applications



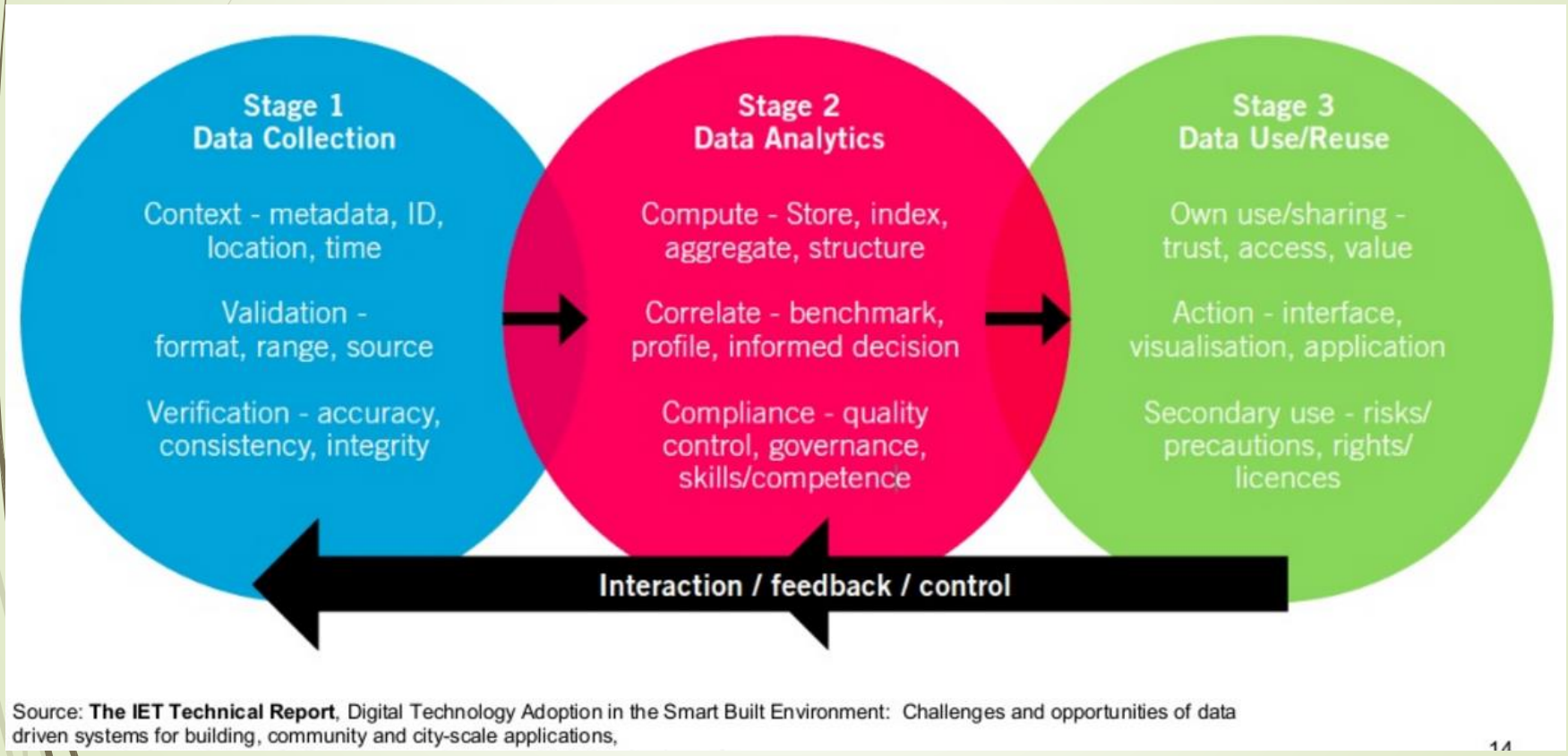
Big data for smart cities: Open City Platform Project

- The public sector needs to reduce IT costs while increasing services to the citizens, their efficiency, velocity and completeness. One of the most efficient way to obtain such results is to invest in new technologies models, such as cloud computing, platform for Intelligent Information Management and sharing the available data
- Open City Platform is a MIUR founded project which aims to develop a cloud platform for the public administration, based on open source software.

- One of the application-level components of OCP is the big-data analysis engine
- Based on open source technologies like:
  - Hadoop
  - Mesos
  - Openstack
- Main topics:
  - Health care
  - Cultural heritage
  - Transport/traffic
  - Public services
  - Education
  - Environment

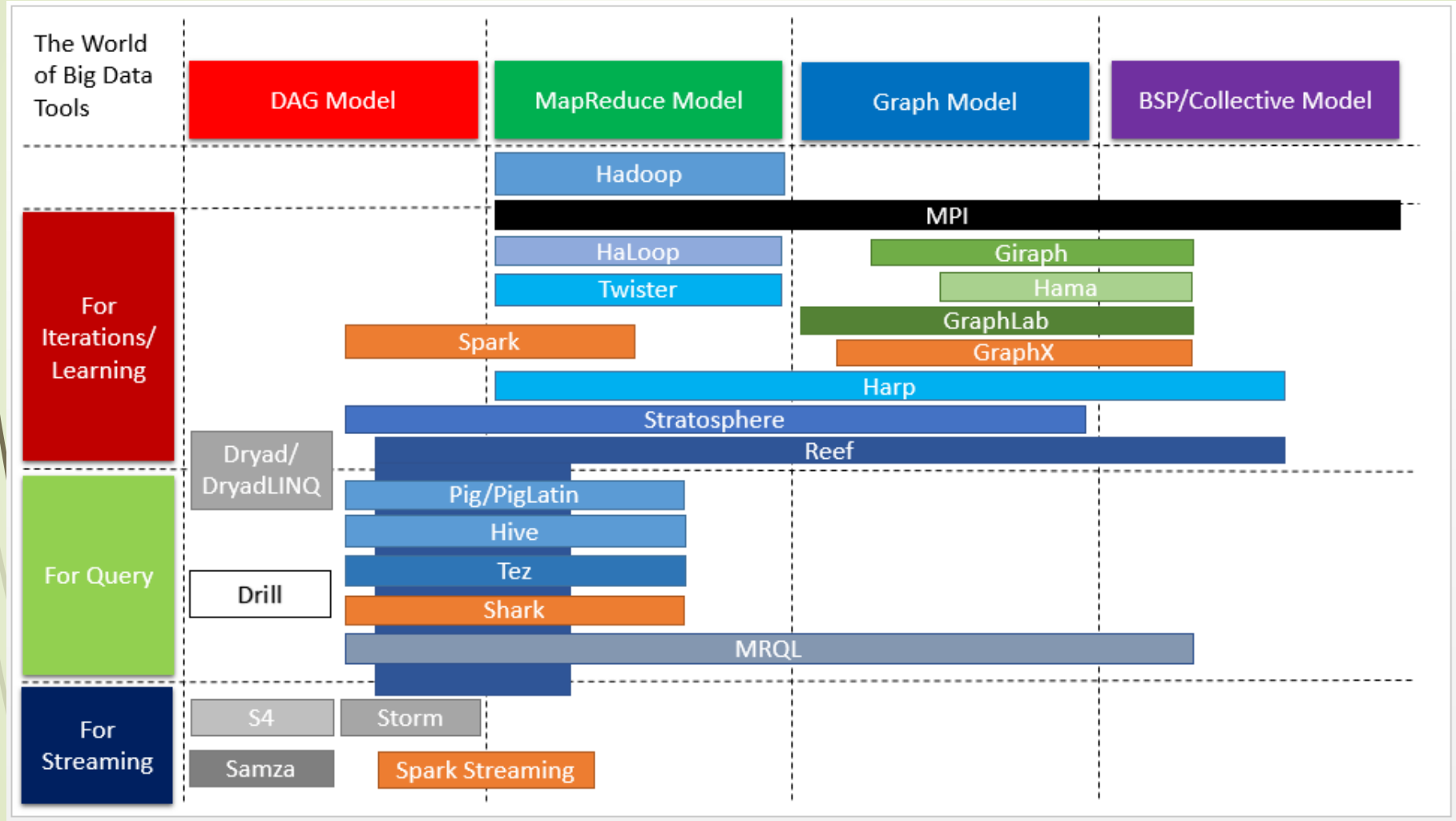


Data lifecycle

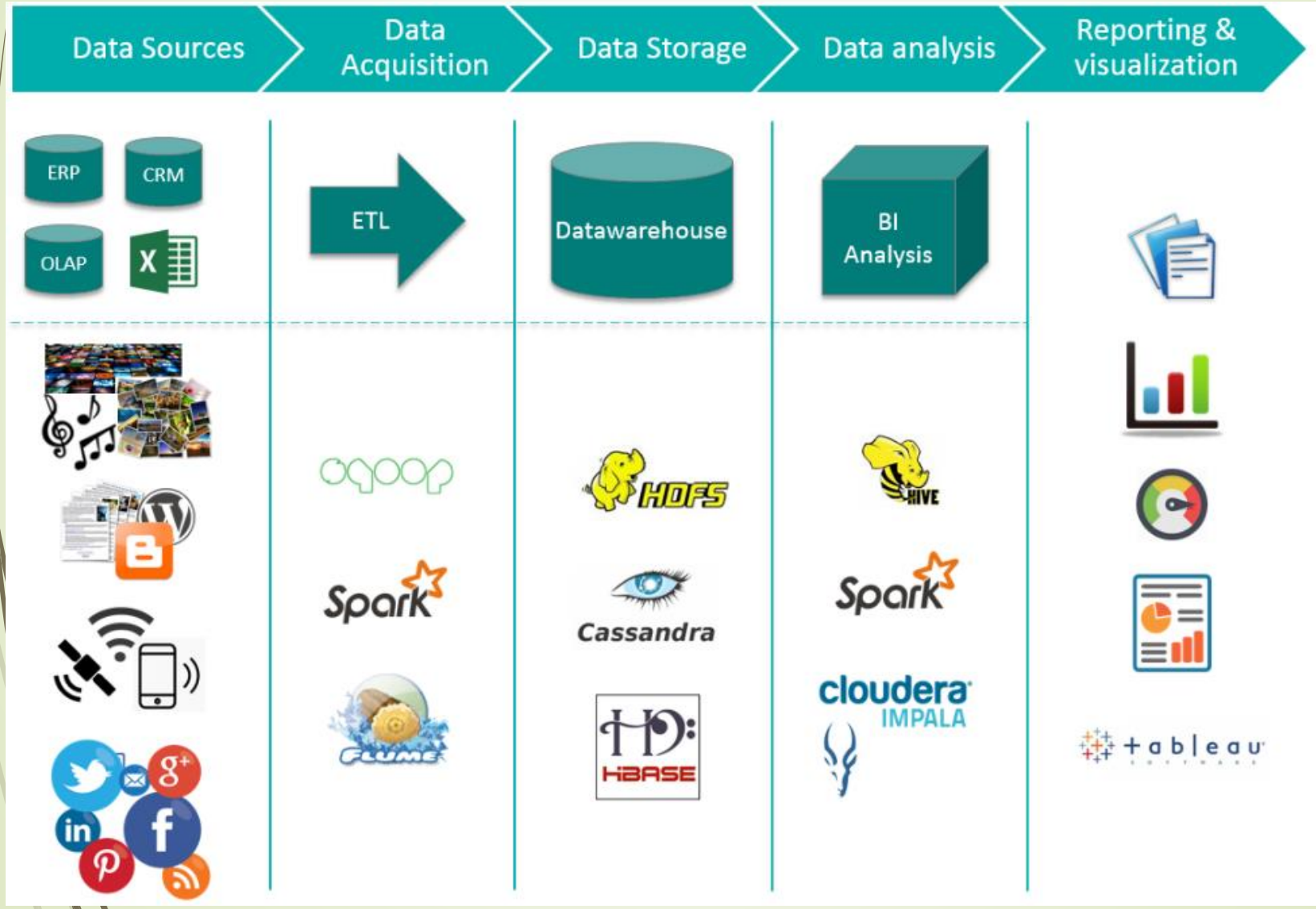


Source: **The IET Technical Report**, Digital Technology Adoption in the Smart Built Environment: Challenges and opportunities of data driven systems for building, community and city-scale applications,

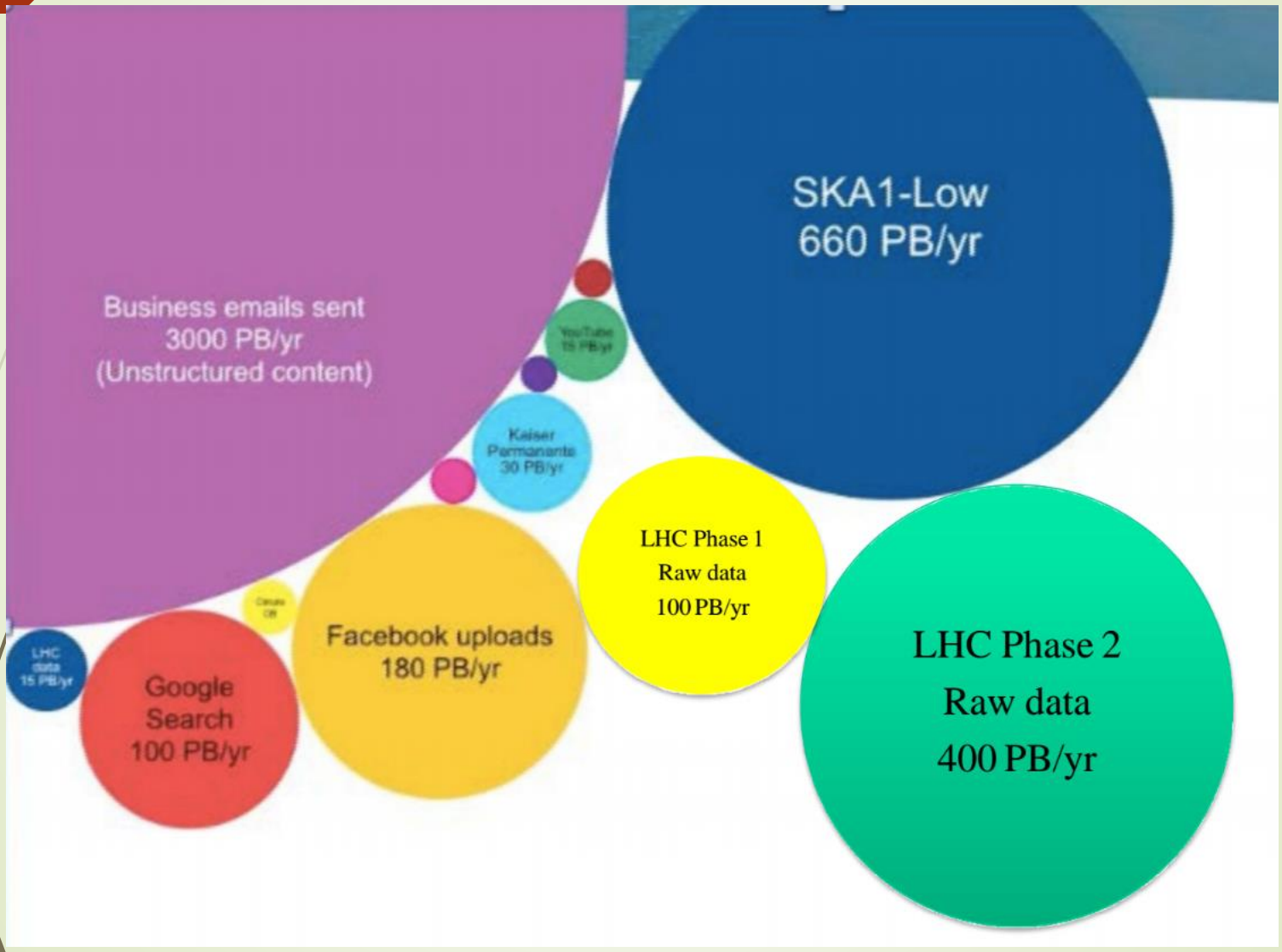
Not one technology, but a mix of technologies



Choosing the right tool for the right task



INFN expertize to Storage Big Data



SCIENTIFIC BIGDATA ITALIAN HUBs

HTC@INFN (30PB)

HPC@CINECA (10PF)

GARR High-speed network

CNR datasets

IMPORTANT PROJECT OF COMMON EUROPEAN INTEREST (IPCEI)

ON
HIGH PERFORMANCE COMPUTING
AND
BIG DATA ENABLED APPLICATIONS
(IPCEI-HPC-BDA)

European Strategic Positioning Paper

On-the-edge for Italian scientific big-data and smartcities applications

- The main INFN Data Center to collect LHC data is located in Bologna at INFN CNAF
- Several scientific communities focused in Big Data are located in Emilia-Romagna
- INFN is strongly involved
  - Smart Cities Projects (Open City Project)
  - Big Data Projects (Emilia-Romagna BiGData Community)
  - Data Cloud Computing, next talk by Davide Salomoni.
- Emilia-Romagna Region is very active in supporting Big Data and Smart Cities activities to improve and support the economic growth of the enterprises and local manifactures