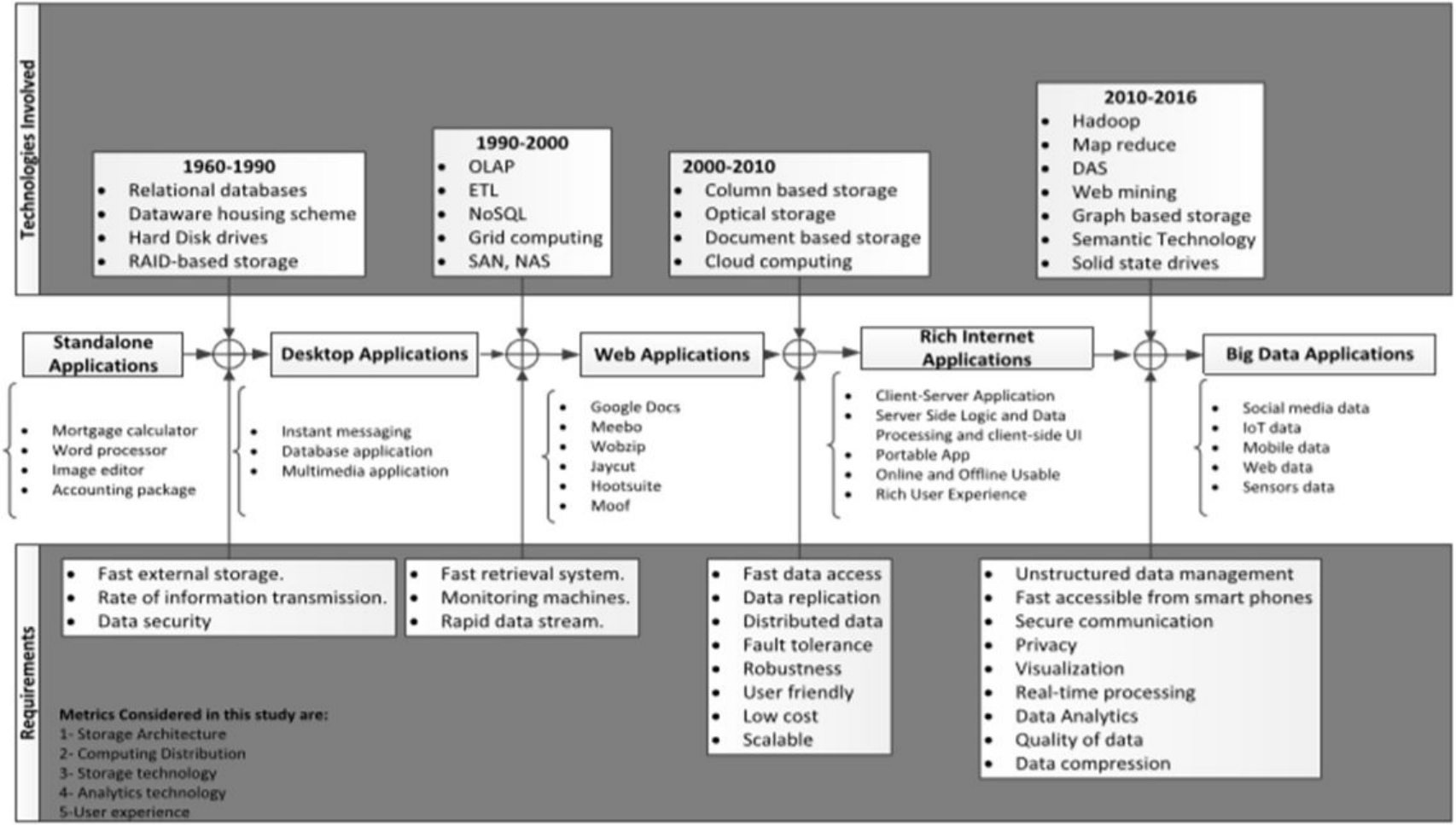
BIG DATA

FROM BEGINNING TO FUTURE

ORIGINS OF BIG DATA APPLICATIONS

- Consideration during the research:
  - Application architecture
  - Chronological development
  - Gradual evaluation
- Five metrics:
  - Storage architecture
  - Computing distribution
  - Storage technology
  - Analytics technology
  - User Experience


Technologies Involved
1960-1990
Relational databases
Dataware housing scheme
Hard Disk drives
RAID-based storage
1990-2000
OLAP
ETL
NoSQL
Grid computing
SAN, NAS
2000-2010
Column based storage
Optical storage
Document based storage
Cloud computing
2010-2016
Hadoop
Map reduce
DAS
Web mining
Graph based storage
Semantic Technology
Solid state drives
Standalone Applications
Mortgage calculator
Word processor
Image editor
Accounting package
Desktop Applications
Instant messaging
Database application
Multimedia application
Web Applications
Google Docs
Meebo
Wobzip
Jaycut
Hootsuite
Moof
Rich Internet Applications
Client-Server Application
Server Side Logic and Data Processing and client-side UI
Portable App
Online and Offline Usable
Rich User Experience
Big Data Applications
Social media data
IoT data
Mobile data
Web data
Sensors data
Requirements
Fast external storage.
Rate of information transmission.
Data security
Fast retrieval system.
Monitoring machines.
Rapid data stream.
Fast data access
Data replication
Distributed data
Fault tolerance
Robustness
User friendly
Low cost
Scalable
Unstructured data management
Fast accessible from smart phones
Secure communication
Privacy
Visualization
Real-time processing
Data Analytics
Quality of data
Data compression
Metrics Considered in this study are:
1- Storage Architecture
2- Computing Distribution
3- Storage technology
4- Analytics technology
5-User experience

TRENDS

- 90% of the data has been generated the last two years (Science Daily, 2016)
- Increase in demand for skills in Big data.
- Annual growth rate during the 1990s / annual growth rate in 1998.
- Growth rate of data (fig on next slide).

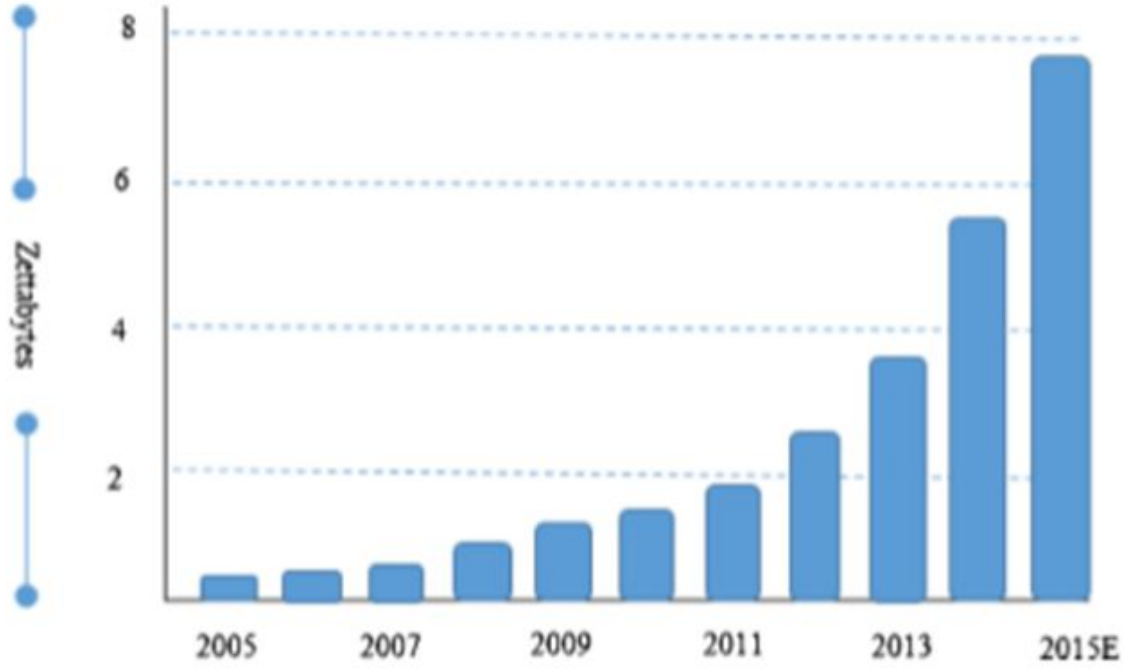


Fig. 2. The rapid growth rate of data in Zettabytes.

SOURCES OF BIG DATA

- E-mail services
- Internet of Things (GPS, mobiles, intelligent clothing, alarms, blinds, window sensors etc)
- Self-Quantified
- Multimedia
- Social media

BIG DATA HANDLING

- BATCHED BASED
  - Apache Hadoop
  - Skytree Server
  - Talend Open Studio
  - Jaspersoft
  - Dryad
  - Tableau
  - Karmasphere
- STREAM PROCESSING
  - Storm
  - S4
  - SQL Stream
  - Splunk
  - Apache Kafka
  - SAP Hana

BIG DATA METHODS

- Hashing
- Bloom Filter
- Indexing
- Parallell computing

AREAS TO BE EXPLORED

- Graph processing
- Heterogenous computing
- Hybrid computing
- In memory processing

BIG DATA TECHNIQUES

- Data Mining
- Web Mining
  - Web Content Mining
  - Web Structure Mining
- Visualization Methods
- Machine Learning
  - Supervised
  - Unsupervised
- Optimization Methods
- Social Network Analysis (SNA)

AREAS TO BE EXPLORED

- Distributed Mining
- Scalable Machine Learning
- Time Variable data
- Mining from Sparse data

OPEN RESEARCH CHALLENGES

- NoSQL
- High-performance computing systems
- Big Data indexing schemes
- Analytics
- Data quality
- Visualization
- Big data security

Future Technologies	Potentially Marginalized Technologies	Brief Description	References
Cloud computing	– Virtualization – Software-defined networking	Provides on demand data storage service.	Hashem et al. (2015), Aboifazli et al. (2015), Yousafzai et al. (2016a,b)
Granular computing	– Discretization – Type-2 fuzzy sets and systems	Divides data into smaller modules, and aggregate all the modules after completion of the specific task.	Pedrycz (2013)
Software-defined storage	– Storage Virtualization – Storage resource Management	Separates the hardware from the software and makes flexible data processing.	Rouse (2014) and Akhunzada et al. (2015)
Stream computing	– Object-oriented programming – Smalltalk library standard	Delivers real-time analytic processing on constantly changing data in motion.	Bayoumi et al. (2009)
Artificial intelligence	– Optimization – Neural networks – Big data mining – IoT mining	Help to make intelligent devices.	Charniak et al. (2014)
Parallel computing	– Distributed computing	Makes process execution fast.	Darriba et al. (2012)
Bio-inspired computing	– Immune systems – Lindermayer systems – Membrane computers	Provides high-efficiency by incorporating several new factors such as robustness, scalability and flexibility in the computational tools.	Castillo and Melin (2012)
Fourth generation optical disks	– 3D optical data storage. – Holographic data storage	Provides efficient data storage.	Hamann et al. (2006)
Quantum computing	– Electronic computing – Optical computing. – Quantum clock	Much faster computing, for some kinds of problems, chemical modeling, new materials with programmed properties, Hypothetical of high-temperature superconductivity and superfluidity.	Finch et al. (2014)
Smart grid computing	– Image processing	Provides access to resources (systems, data, applications, and services) via the Internet.	Fang et al. (2012)
Optical computing	– Laser – Transistor	Allows a higher bandwidth than the electrons used in conventional computers.	Woods and Naughton (2012)
Quantum cryptography	– Public-key encryption – Signature schemes	Helps in performing cryptographic tasks.	Gilbert and Weinstein (2014)
Semantic web	– SPARQL – Notation3 – Web ontology language	Enables users to find, share and combine information more easily.	Berners-Lee and Hendler (2001)
Edge Computing	– Fog computing – Mobile edge computing – Cloudlet	Facilitates the users by bringing computation down towards the edge of the network.	Ahmed and Ahmed (2016), Jararweh et al. (2016) and Satyanarayanan et al. (2015)

BIG DATA

FROM BEGINNING TO FUTURE