Big Data trends and challenges
Valentin Cristea

New Era of Big Data

- Internet became an immense producer and consumer of huge volumes of data

- Companies need Big Data in collaborative partnerships in the market and across the globe, and in creating business advantage for clients; e.g.:
  - Enabling rolling plan, forecasting and budgeting
  - Making risk-aware decisions
  - Leveraging customer sentiment analysis
  - etc.

New Era of Big Data (2)

- Health applications coordinate doctors, pharmacies, laboratories, insurers by using
  - electronic health records
  - doctors collaboration on diagnosis and treatment plan
  - real time reference patient information (such as imaging investigations)

- Smart Transportation
  - connects various types of data (toll booths, traffic patterns, meters, eye witnesses, video cameras)
  - and use them efficiently to avoid traffic jams in the context of continuously growing of the population

New Era of Big Data (3)

- Governments, agency administration, and social programs
    - address the management of citizen's contacts, an improved way of taxes collection, social services for unemployment, pensions, child benefits, insurance, etc.

- Web
    - Textual comprehension of Web content for Web search
    - Sentiment analysis
    - Skill search
    - Trust computing
    - Detection of social interaction structures in directed graphs

Big Data Benefits

- Healthcare: 20% decrease in patient mortality by analyzing streaming patient data

- Telco: 92% decrease in processing time by analyzing networking and call data

- Utilities: 99% improved accuracy in placing power generation resources by analyzing 2.8 petabytes of untapped data

"Big Data as differentiator" presentation by Harry van Dorenmalen, Chairman IBM Europe at Big Data Summit, Paris May 2013

Big Data = Big Issues ?

- Big Challenge: extract useful knowledge from the huge volumes (petabytes) of data that embodies new characteristics:

- Come from diverse sources: companies, customers, administration, sensors, social networks, open sources

- Are uncertain – need managing data reliability and predictability for veracity

- Have various forms – structured, unstructured, text, video, images, audio

- Are time sensitive, data in motion – streaming data analysis allows decisions in fractions of a second

Future research in Big Data

- Big Data Storage
  - Very high capacity, extensible data stores
  - Various data models – relational, key-value, document, graph, time series

- High performance Big Data analysis, processing, mining
  - Parallelization
  - Distributed computing
  - In-memory Big Data
  - MapReduce
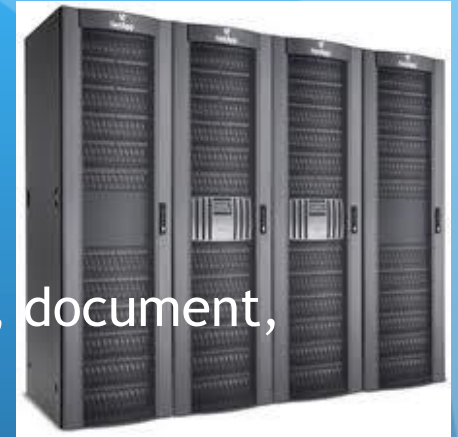
Future research in Big Data (2)

- Eliminate «information noise»
  - Context-aware data tailoring
  - Provide quality evaluation and cleaning tools
  - Find more useful data outside, rather than inside (e.g., Open Data)
  - Knowledge-based processing and Ontological reasoning

- Facilitate collaborative work (and collaborative decision making)
  - People with different backgrounds should be able to interact to gain insight
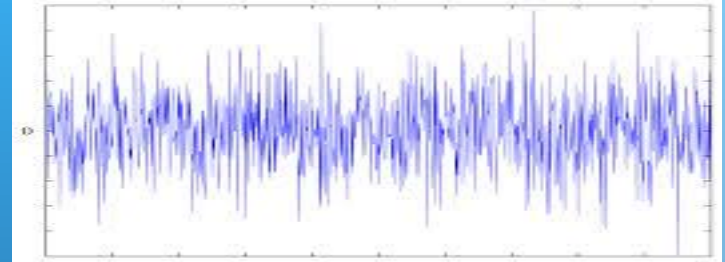
Big Data Platforms

- Automate as much as possible the management of data
  - Develop specialized approaches for data management (e.g., based on the multidimensional paradigm) to improve performance

- Easy access to Big Data
  - from various locations and devices (including mobile)
  - by non-expert end users without specific knowledge about the source or technology underneath
  - and companies of various sizes (big companies, SMEs …)

Big Data Platforms (2)

- New storage and compute components will be needed to allow global data availability and access over different types of networks, for various cooperative user communities, which
  - exploit different facets of Big Data
  - in application with different service level requirements (batch, online, near real-time, collaborative, etc.)
  - for various (often unpredictable) time periods
  - with different security needs going up to highly trusted and secure environments

Future platform features

- Support massive growth of storage capacity

- Offer Big Data services on demand

- Guarantee the performance agreed with customers (SLA)

- Facilitate faster applications deployment

- Support various Big Data representations and access patterns

- Enhance intelligent, autonomous, pro-active behavior based on context data integration with applications and infrastructure

- Cloud Computing might be a solution!