

TAMING THE BEASTS OF BIG AND SMALL DATA



D. M. G. STROO, PhD
Director of Invisi BV,
Netherlands

Invisi BV, Netherlands

Topics

- Types of data that flood this modern world
- Being selective with data
- Varied data and varied storage
- Data warehousing in the Big Data era
- Belarus and the data revolution

A Tsunami of Data Floods our World



Invisi

Invisi

Not All Data Is Created Equal



Personal (chat) data



Business (application) data

Not All Data Is Created Equal



Machine-generated data like web server logs, Internet of Things

Invisi

Invisi

Internet of Things

- Devices that collect and exchange data
- Smart homes
- Health data (e.g. Apple's Health Kit)



Not All Data Is Created Equal



Personal health data collected by Apple, Nike, etc

Types of Data

- Structured (as in: stored in a relational database)
 - Repetitive
- Unstructured (Big Data)
 - Repetitive
 - Non repetitive (varied)
 - Repetitive varied (based on templates)
 - Truly varied
 - images
 - text documents
 - e-mails



Organisational relevance

- Many data have no organisational relevance
- Put time and money into irrelevant data?
- Structured (business) data is always relevant
- Unstructured repetitive data mostly is irrelevant
- Unstructured varied data is often relevant, but hard to analyse



Types of Data per Branche

- We have a lot to gain in some branches by learning to analyse unstructured varied data



Collect (Big) Data Selectively

- A lot of data has limited or no business relevance
- There is more and more data digitally available
- Most data is noise, we have to understand and look for the signals in the noise
- Collecting more digital data means that the haystack to find the needle in only gets bigger, or: finding meaningful data gets harder

Which Data Do You Have And What Is Its Value?

- Which applications do you use?
- How and where is data of those applications stored?
- Do you use e-mail?
- Through which other digital ways do you communicate externally and internally?
- Which data is out there that is relevant for you?

Useful Facts

- Data are a collection of facts
- Only when a fact is true and useful, is it a signal
- A fact is useful when it:
 - informs
 - is relevant
 - deserves a response

Two Types of Signals

- Exploratory signals
- Signals of change
 - Significant changes in patterns
 - Significant changes in magnitude
- The appearance of a known and significant pattern

Invisi

Data Stewardship

- Get to know the land
- Watch over the land



Invisi

Different Types of Data Requires Different Storage

- Traditionally we store data in relational databases
- Or we store documents on a file system
- Now many options:
 - relational, NoSQL, Hadoop
 - storage or in-memory
 - local or cloud
 - row or column oriented
 - open source or closed source, free or paid

Invisi

Join Data From Different Storage Types

- Classical ETL loads data from flat files and relational databases into the data warehouse relational database
- Now your data sources are web services, Hadoop, JSON, etc. and your data warehouse may be relational, Hadoop or cloud
- Replicate and store data or virtualise

Invisi

Rise of The Data Lake

- A data lake is a single store of all data in the enterprise
- Opposite of the data mart, that represents cleansed and clearly structured information for easy consumption
- A great way of storing all your data to answer all your questions or a means to postpone the hard question of which data provides you the information that answers the questions you have?
- A Data Lake does not replace Data Marts or similar structures to work on data sense making

Invisi

Data Warehousing Lives

- Structured business data living in relational databases still holds significant business value for many organisations
- Integrating and restructuring such data to deliver information that can lead to actionable insights is still in demand
- The reason why many data warehousing and business intelligence initiatives fail is through a lack of skilled professionals understanding the best practices and methods
- The reasons why many organisations fail to start such initiatives is lack of funds, lack of interest from vendors and professionals, to name a few

Invisi

Useful or Jargon?

- Use the type of information application and storage that suits your needs, not what is in fashion



Invisi

Data Chasm

- Big corporations like Google and Facebook process and own huge amounts of data from billions of people, often ignoring the rules of privacy
- Many organisations big and small are still struggling to deal with their own internal structured data from core business applications to provide insight about their products, services, customers, finances, etc
- Tools to collect, store and visualise data are still too expensive, too complex, too labour intensive or too hard to find for many businesses and organisations

Invisi

Standard Models and Dashboards

- Many small and medium organisations have similar business needs for analytics and insight
- Core financial insight, retail insight including online, HR insight, business services all can benefit from standard data models as foundation for standard dashboards
- Such standard data models and dashboards can be offered as a cloud based, low cost subscription model connected to the organisations' varied systems

Invisi

Bring The BI Process to the Masses

- Apart from limited access to technical BI, data warehousing and Big Data technology, also access to consulting of the BI and Analytics process is limited
- We need to look to new models, services and offers to assist smaller companies with real analytical needs to unlock their full business potential
- The Analytics consulting area is one that will stay open for enterprising individuals and companies, as its business model is out of scope for huge corporations like Google

Invisi

Big Data on Small Devices

- Mobility of Business Intelligence and Analytics is still limited
- Vendors like Tableau are making steps with apps like Vizable, but there is room for improvement and competition
- The many web based tools like Cognos and Yellowfin can make quick wins by adopting responsive design practices



Invisi

Belarus Data Revolution?

- Belarus originated companies Wargaming and Viber Media work heavily with their large amounts of user data
- What about other Belarus companies? Belarus banks, Belarus retail, Belarus telecom, Belarus municipalities?
- Do Belarus companies and organisations benefit and get knowledge from international cooperation?
- Do Belarus companies have access to tools, methods and skilled people to get insight from their data?

Invisi

Belarus Data Revolution?

- Will highly skilled and affordable Belarus workers deliver accessible and easily usable tools to companies and organisations worldwide to turn their data into actionable insight?
- Will highly skilled and affordable English speaking Belarus experts deliver high quality services and support to companies and organisations worldwide to have them get the most out of the information collected from their data?
- Will these highly skilled Belarus workers and experts deliver the same tools and services on their internal market?

Invisi

Belarus Data Revolution - Possibilities

- Increased emphasis on the role of data in applications
- Focus on modern ways of collecting, manipulating, storing and analysing data
- Contributing to open source initiatives

Invisi