



CONL
722

Big Data Challenges and Opportunities

1.1.1: Business Intelligence and Data Science

Hello and Welcome to the first session for CONL722, Big Data challenges and opportunities..

Objectives



Explore business intelligence
and data science

This session will briefly introduce you to data science and big data as well as provide an overview of business intelligence supporting decision making.

Computerised Systems

- Since 1970s, computerised systems provided automation of business processes
- Accumulation of growing amounts of data in operational databases.
- Use operational data to support decision-making
- Gaining competitive advantage.
 - Raw data can support operations what it is intended for
 - When processed will become knowledge
 - Discovering patterns/trends/threats etc.

Since the 1970s, computerised systems provided organizations with a competitive advantage through the automation of business processes to offer more efficient and cost-effective services to the customer.

This has resulted in the accumulation of growing amounts of data in operational databases.

Now the focus is to use operational data to support decision-making, as a means **of gaining competitive** advantage.

The systems were never designed to support such business activities.

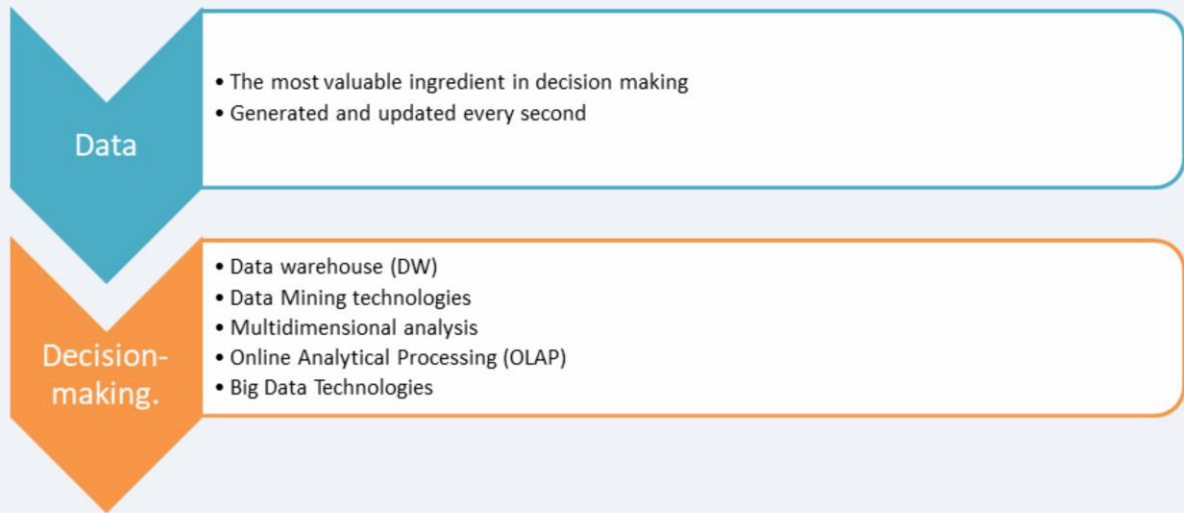
Organisations typically have numerous systems with overlapping and sometimes contradictory definitions.

Raw data can support operations what it is intended for

When processed will become knowledge

Discovering patterns/trends/threats etc.

Computerised Systems



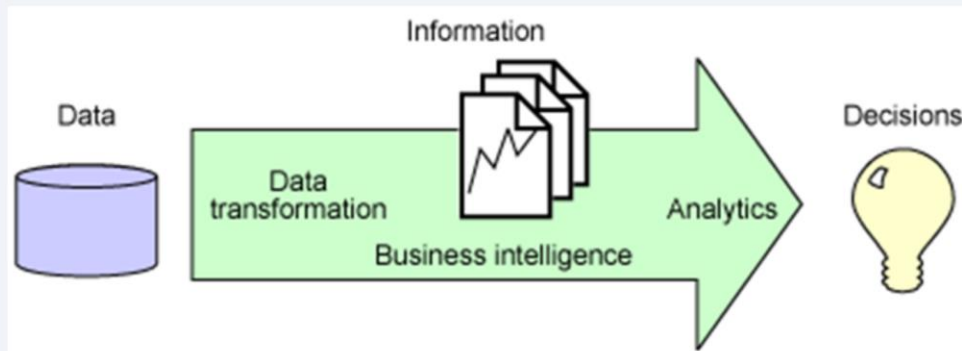
Recognising that the data is the **most** valuable ingredient in decision making

Generated and updated every second

Organizations need to turn their archives of data into a source of knowledge so that a single integrated/consolidated view of the organization's data is presented to the user.

Big Data Technologies, **Data warehouse (DW)** and Data Mining, multidimensional and OLAP technologies have deemed the solution to meet the requirements of a system capable of supporting decision-making.

Business Intelligence



Gathering and analysis of vast amounts of data to gain insights that drive strategic and tactical business decisions. It encompasses the technologies listed in the previous slide allowing to gather, store, access, and analyse data to improve the business decision-making capabilities.

Business Intelligence can be defined as the collection of the processes and technologies which transform data into information. It encompasses Data warehousing, Multidimensional analysis, Online Analytical Processing (OLAP) Data mining, Data visualization, Simple queries Analytical tools for reporting and Big Data technologies.

These technologies allow business users to gather, store, access and analyse data to improve the business decision-making capabilities.

This can be applied to any organisation and every organisation is heavily relying and government systems are heavily relying on getting the correct information at the right time in the right format.

Transaction Processing Systems (TPS/OLTP)

- Can generate huge volumes of data (Raw material)
 - A telephone company may generate 200 million call records per day
- Characterized by:
 - Mission-critical (if it goes down, so does the business)
 - Small number of typically repetitive transaction types
 - Large numbers of clients (typically clerical, or on-line users)
 - High throughput (transactions/second)
 - Throughput demands may vary throughout the day, week, year
 - Short, predictable response times (1-3 seconds, 95%-100% of the time)
 - Extremely reliable and available
 - Highly secure
 - Data must be kept consistent, up-to-date, and correct

Transaction processing systems (also known as TPS) can generate huge volumes of data (Raw material)

A telephone company may generate 200 million call records per day

TPS systems are mission-critical (if it goes down, so does the business).

They have a small number of typically repetitive transaction types and support large numbers of clients (typically clerical, or on-line users).

The system should give High throughput (transactions/second), which may vary throughout the day, week, year.

TPS support short, predictable response times (1-3 seconds, 95%-100% of the time) and must be extremely reliable and available.

The system must be highly secure and data must be kept consistent, up-to-date, and correct

On-Line Transaction Processing

OLTP systems

Transaction
Server

Remote
procedures on
the server

Customized code
written for both
client and server

Minimum communication overhead

Short structured
exchanges
between client
and server

Request/reply
pair as opposed
to multiple SQL
messages

OLTP Systems (interactive TPS) typically require:

- The Use of a Transaction Server and remote procedures on the server

- Customized code must be written for both client and server,

- The system must have minimum communication overhead by initialising short structured exchanges between client and server, usually a request/reply pair as opposed to multiple SQL messages

The Problem

Department-owned systems

Creating islands of inaccessible and incompatible data

Corporate decision makers cannot easily extract essential information

Organisational memory is fragmented

Different systems

Different database technologies

Different locations

Disadvantages

An underused intelligence system

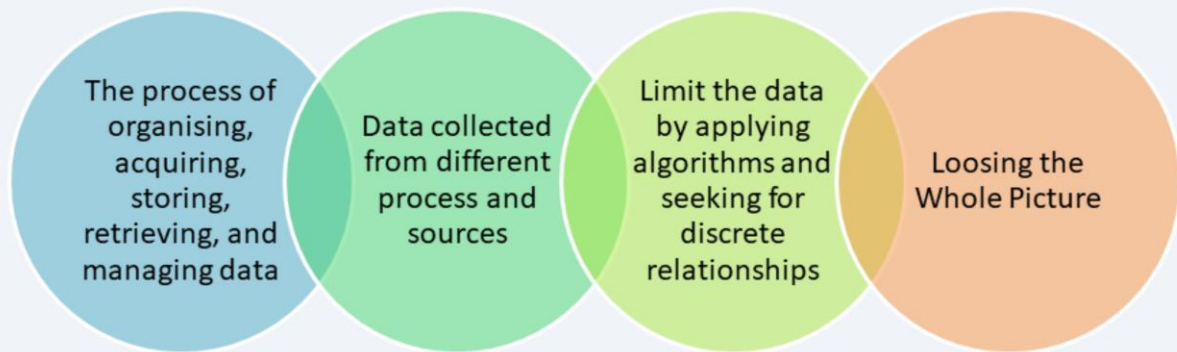
Undetected key facts about transactions, customers etc.

Creating TPS to support organizational tasks and responsibilities encourage the trend towards department-owned systems, creating islands of inaccessible and incompatible data. Corporate decision makers cannot easily extract essential information.

Organisational memory is fragmented through the use of different systems, different database technologies in Different locations.

This will have many disadvantages like underused intelligence system containing undetected key facts about transactions, customers etc.

Data Management



Different types of systems are continuously generating and storing data in a different format.

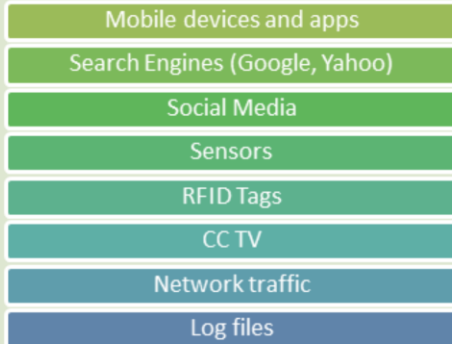
Data management is the process of organising, acquiring, storing, retrieving, and managing data

Data is collected from different process and sources.

Traditional data management try to limit the data in the process of modelling by applying algorithms and always seeking for discrete relationships. However, this will result in loosing the Whole Picture

Data Management

Data is also generated from various other sources



Significantly increasing



For example, data is generated from
Transaction processing system

- Sensors
- RFID Tags
- CC TV
- Mobile devices and apps
- Search Engines (Google, Yahoo)
- Social Media
- Network traffic
- Log files

Volume, Variety and Velocity of data have increased significantly in the last few decades.

Data Science

Data Science is an evolving discipline

Extract knowledge from large amount of data

A platform with a complex blending of technology, algorithm development, and data interference

Methodology by which actionable insights can be inferred from data

As we have seen the data is generated and stored in the systems and processed and analysed appropriately this data is capable of providing useful insights and information to the decision-makers.

Data Science is an evolving discipline and it assists in extracting knowledge from a large amount of data.

It is a platform with a complex blending of technology, algorithm development, and data interference.

Data science is a methodology by which actionable insights are inferred from data.

Data Science

Data managers now need to support Organizational intelligence technologies

Data science team required right tools, process and people with right skills.

The diversity of data, data analysis tasks and data analysis approaches poses many challenges....

Data analysis technologies are adhering to the ever-changing landscape of data collection and applications.

Data managers now need to support Organizational intelligence technologies.

Data science team required the right tools, process and people with the right skills.

The diversity of data, data analysis tasks and data analysis approaches pose many challenges....

Data Analysis an Example

Success of a Movie is analysed based on the revenue generated by

Initially the ticket Sales (TPS)

Later CDs, DVDs, Blue Rays etc. Sales (TPS)

Then the merchandise across multiple channels (TPS)

Population demographics, sentiments/feelings, reviews, feedback were not a visible part of the decision making

Review can be recorded and used as score/rating

Not including the comments

Take an example of analysing the success of a movie.

The success of a Movie is analysed based on the revenue generated by

Initially

Ticket Sales (transaction processing system)

Later

CDs, DVDs, Blue Rays etc. Sales (transaction processing systems)

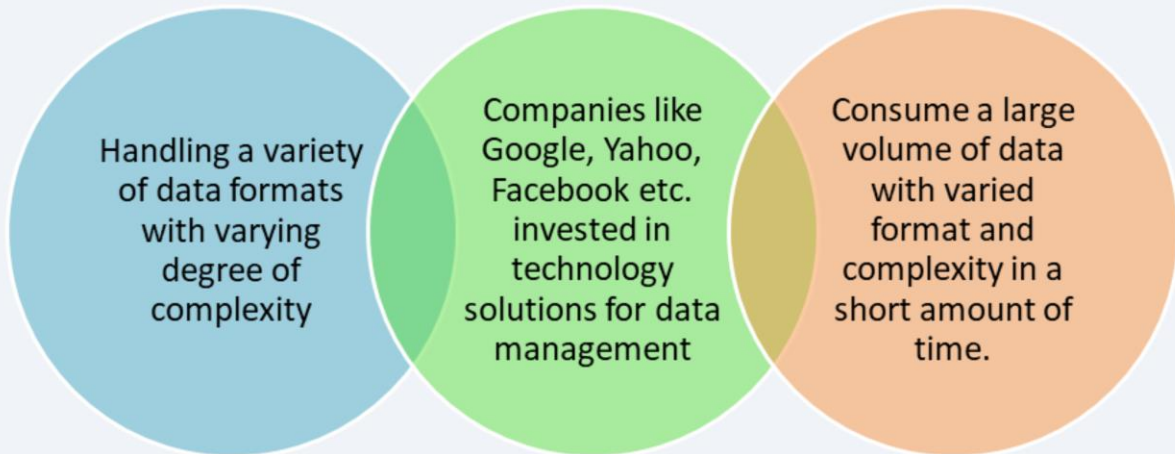
Then

Merchandise across multiple channels (transaction processing systems)

But the population demographics, sentiments/feelings, reviews, feedback were not a visible part of the decision making.

The review can be recorded and used as a score/rating, but not the comments.

Data Analysis



Considering transactional data, mainly the sales details, viewership, rating information along with the non-transactional data such as sentiments. Comments etc. require handling a variety of data formats with varying degree of complexity. Companies like Google, Yahoo, Facebook etc. invested in technology solutions for data management.

The systems are designed to consume a large volume of data with varied format and complexity in a short amount of time.

Big Data

Biggest phenomenon of the modern computing industry (?)

Buzz word?

Must have technology (?)

The fundamental reason for its popularity (?)

In the light of the 21st century data analysis is Big data is the biggest phenomenon of the modern computing industry (?)

Or is it juts a Buzz word?

Do you think it is a technology that every organisation must have?

While exploring the answers for the above questions, don't forget that the fundamental reason for its popularity is, it is a technology platform that provides the capability to process DATA of MULTIPLE formats and structures without the constraints of the traditional systems and database platforms.

Big Data

What is Big Data ?

- Large Volume of Data with Varying degree of formats, complexities, and ambiguities (Variety), generated at different Velocities, which cannot be processed using traditional technologies, methods, algorithms or any off-the-shelf solutions.
- Source include: Sensor networks, click stream data, scanning devices, machine log, airplane engines, consumer driven data from social media, transactional data

Why Big Data ?

- Provides the ability to access large volume of data to gain critical and useful insights
- Learning process is machine managed with minimum human intervention, hence the analysis is simpler and error free .

To conclude from what we have discussed so far, we can say Big Data is Large Volume of Data with Varying degree of formats, complexities, and ambiguities (Variety), generated at different Velocities, which cannot be processed using traditional technologies, methods, algorithms or any off-the-shelf solutions.


Data can be from different sources such as Sensor networks, clickstream data, scanning devices, machine log, aeroplane engines, consumer-driven data from social media, and traditional transactional data.

Why we use Big Data? This can be answered briefly highlighting two key points. It provides the ability to access a large volume of data to gain critical and useful insights. The learning process is machine managed with minimum human intervention, hence the analysis is simpler and error-free.

Following weeks will find more about big data well as explore the opportunities and challenges. .

Next

Big Data



Fundamentals and Architecture

The next video will be discussing the fundamentals and architecture of big data.