MONASH University

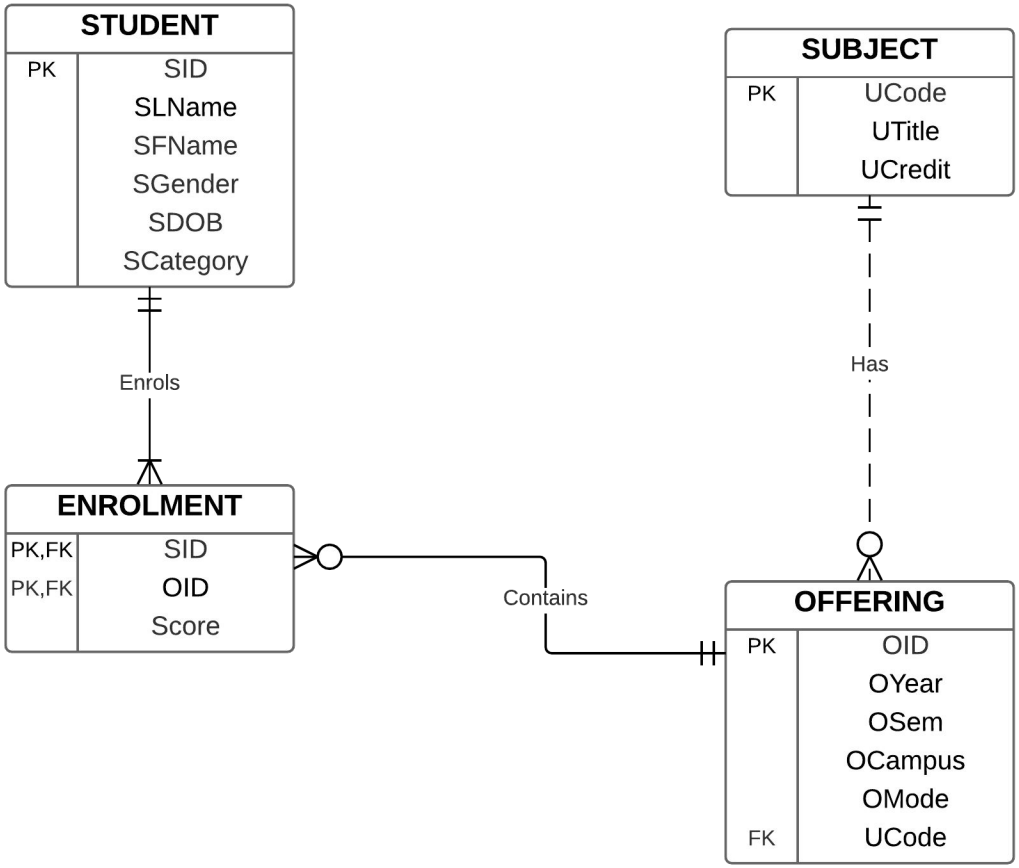
MONASH
INFORMATION
TECHNOLOGY

Where To?
Scheduled Final Assessment Preparation

FIT2094-FIT3171

GROUP OF EIGHT AUSTRALIA

Operational Database - the unit's focus



Application Development

- Web based front ends
  - Wide range of approaches: PHP, ASP.NET, etc
  - Very Rudimentary (requires VPN)

```php
<?php
//SQL query statement
$query = "SELECT studid,
 rtrim(studfname) || ' ' || rtrim(studlname) as sname,
 to_char(studdob,'dd-Mon-yyyy') as sbdate,
 studemail
 FROM uni.student
 ORDER BY studid";
```

- PL/SQL
  - backend development
  - Triggers, functions, procedures and packages
  - Procedure to change employee departments: move_employee (empno, new dept)
    - **`move_employee (101, 2)`**

FIT2104 - Web database interface

Learning outcomes

Expand all

On successful completion of this unit, you should be able to:

1. Explain the need and importance for application developers to have skills in this area of IT applications;
2. Describe and compare the key basic technologies which underly the development of web database applications;
3. Evaluate and assess the key technological issues confronting developers when building applications of this type;
4. Implement the key features of programming languages which are commonly used for developing web database applications;
5. Analyse, design, develop and implement a web database application using a commonly used programming language;
6. Evaluate and critique proposed web database solutions to a business problem.

Usage of database

- Example of a supermarket
- Decision making
  - Operational level
    - How often do we need to re-stock X-item?
  - Strategic and tactical level
    - Is there any branch that performs worse than the state average?
    - What is the total sales made by each state each year and across a number of years?

Operational Data vs. Decision Support Data

- Operational data
  - Mostly stored in relational database
  - Optimized to support transactions representing daily operations
  - Example:
    - How many students enrolled in FIT2094?
- Decision support data differs from operational data in three main areas:
  - Time span
  - Granularity
  - Dimensionality
  - Example:
    - What is the total number of students in the foundation units in each year (subtotal of the two semesters numbers) and the total across years, across a single unit.

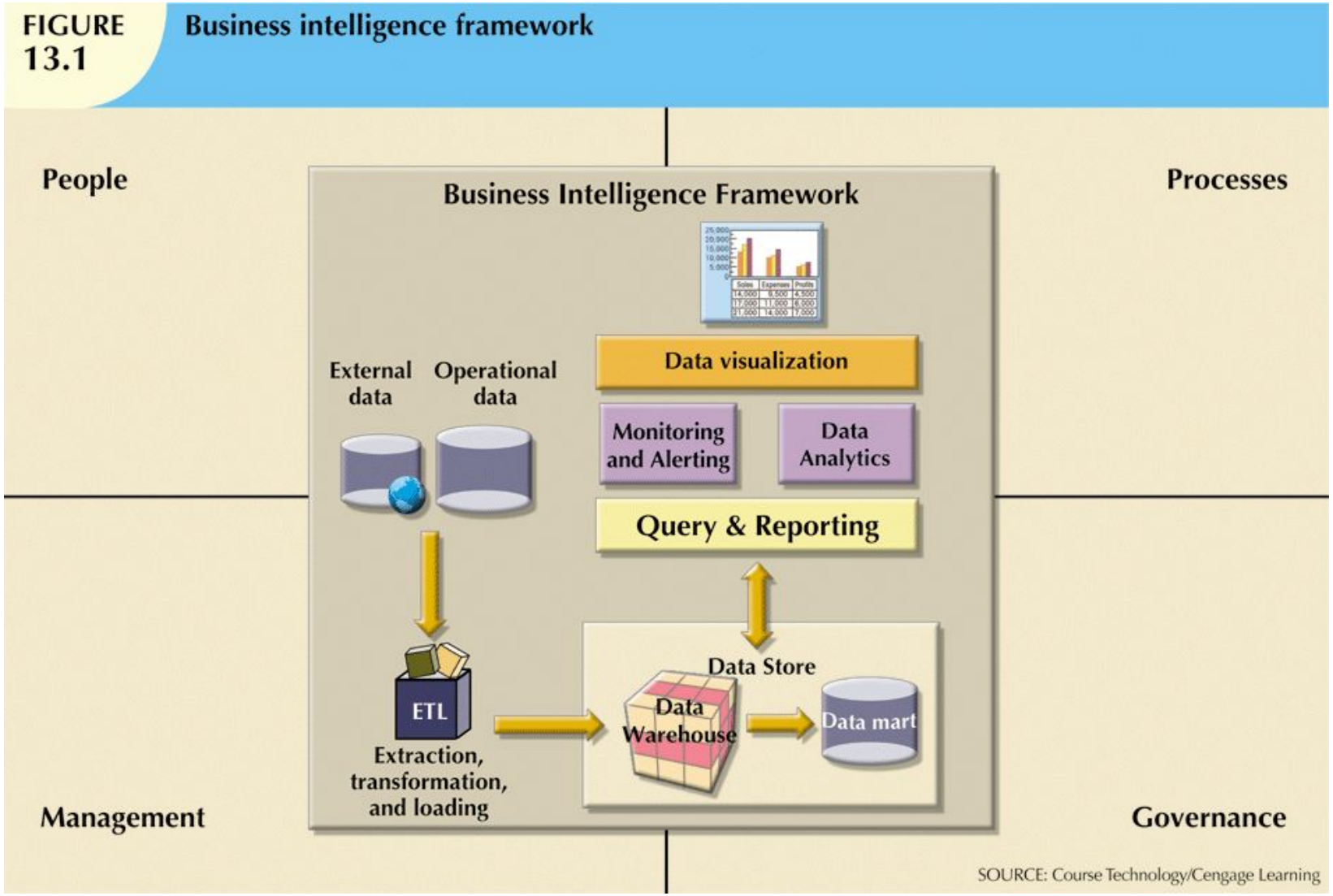


FIGURE 13.1
Business intelligence framework
People
Processes
Business Intelligence Framework
External data
Operational data
Data visualization
Monitoring and Alerting
Data Analytics
Query & Reporting
ETL
Extraction, transformation, and loading
Data Store
Data Warehouse
Data mart
Management
Governance
SOURCE: Course Technology/Cengage Learning

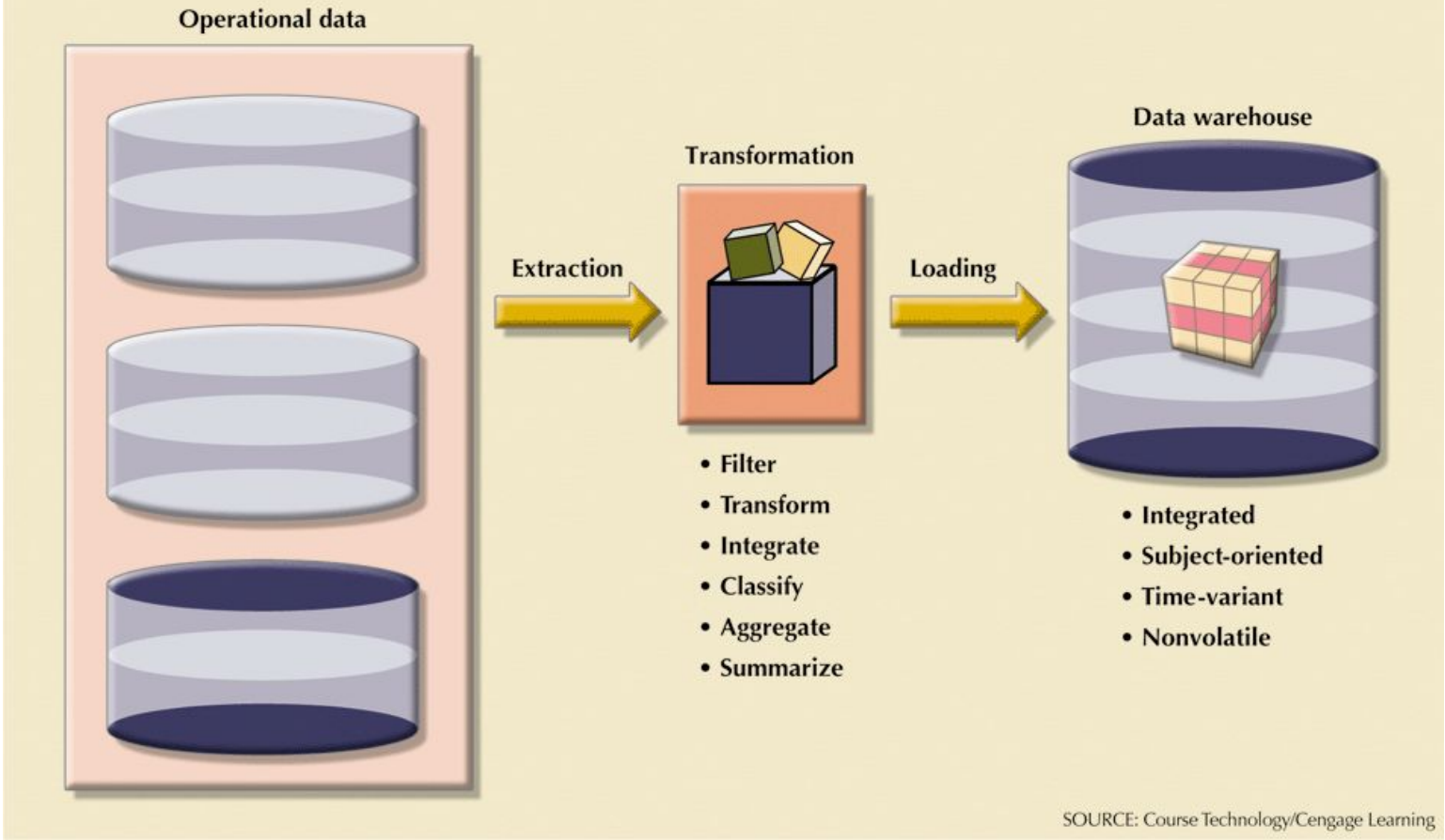
TABLE 13.5 Contrasting Operational and Decision Support Data Characteristics

CHARACTERISTIC	OPERATIONAL DATA	DECISION SUPPORT DATA
Data currency	Current operations Real-time data	Historic data Snapshot of company data Time component (week/month/year)
Granularity	Atomic-detailed data	Summarized data
Summarization level	Low; some aggregate yields	High; many aggregation levels
Data model	Highly normalized Mostly relational DBMSs	Non-normalized Complex structures Some relational, but mostly multidimensional DBMSs
Transaction type	Mostly updates	Mostly query
Transaction volumes	High update volumes	Periodic loads and summary calculations
Transaction speed	Updates are critical	Retrievals are critical
Query activity	Low to medium	High
Query scope	Narrow range	Broad range
Query complexity	Simple to medium	Very complex
Data volumes	Hundreds of gigabytes	Terabytes to petabytes

The Data Warehouse

- Database size
  - DBMS must support very large databases (VLDBs)
- Integrated, subject-oriented, time-variant, and nonvolatile collection of data
  - Provides support for decision making
- Usually a read-only database optimized for data analysis and query processing
- Requires time, money, and considerable managerial effort to create

FIGURE 13.4 The ETL process



SOURCE: Course Technology/Cengage Learning

FIT3003 - Business intelligence and data warehousing

Learning outcomes

Expand all

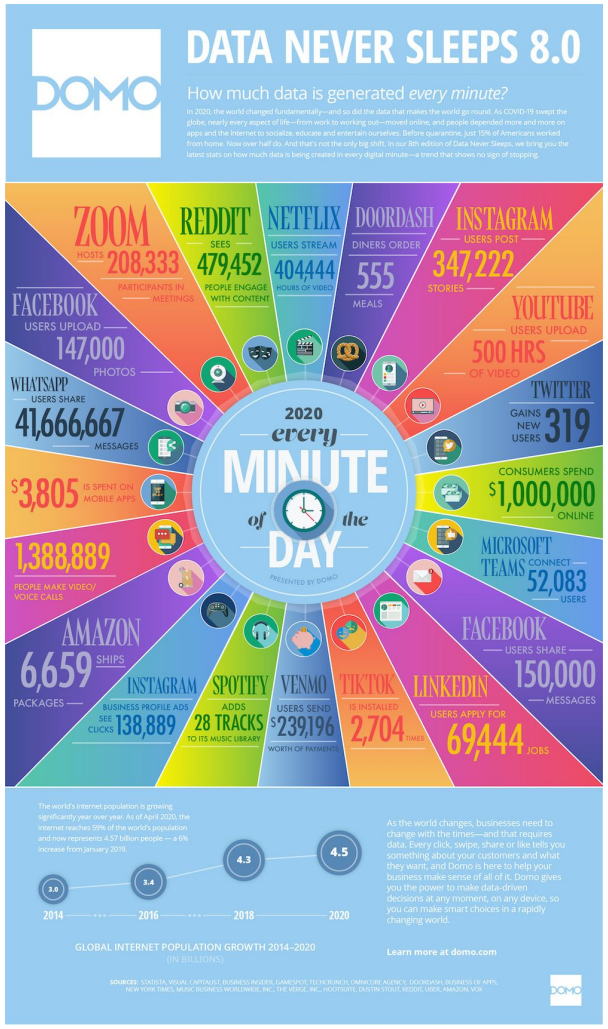
On successful completion of this unit, you should be able to:

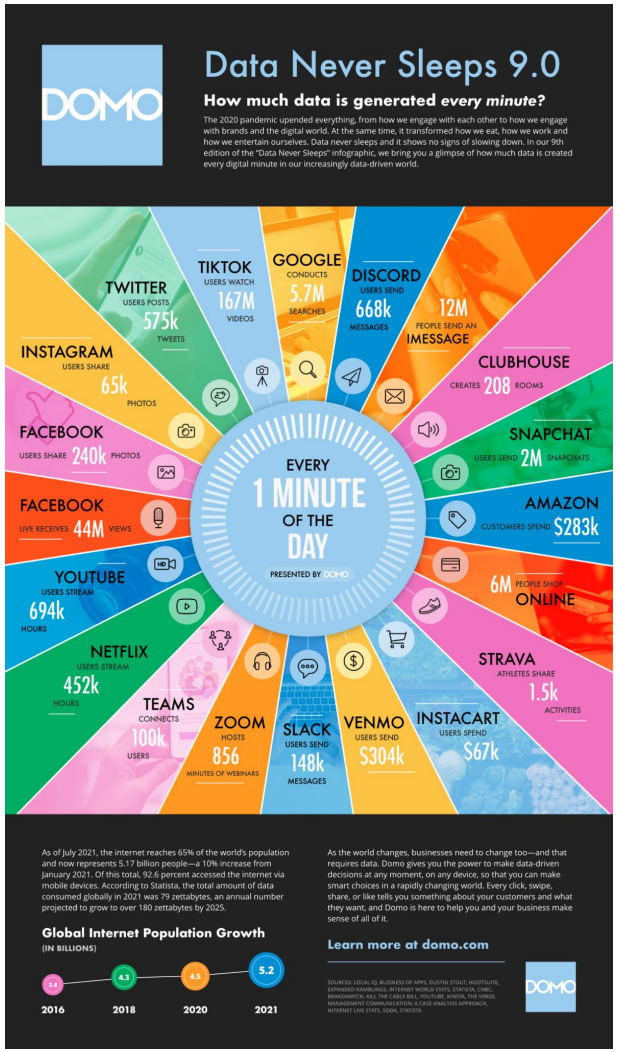
1. Design multi-dimensional databases and data warehouses;
2. Use fact and dimensional modelling;
3. Implement online analytical processing (OLAP) queries;
4. Explain the roles of data warehousing architecture and the concepts of granularity in data warehousing;
5. Create business intelligence reports using data warehouses and OLAP.

IoT - the explosion - Data, Data, Data …..

https://youtu.be/bUnOo10z42k

Data Growth 2020 - 2021



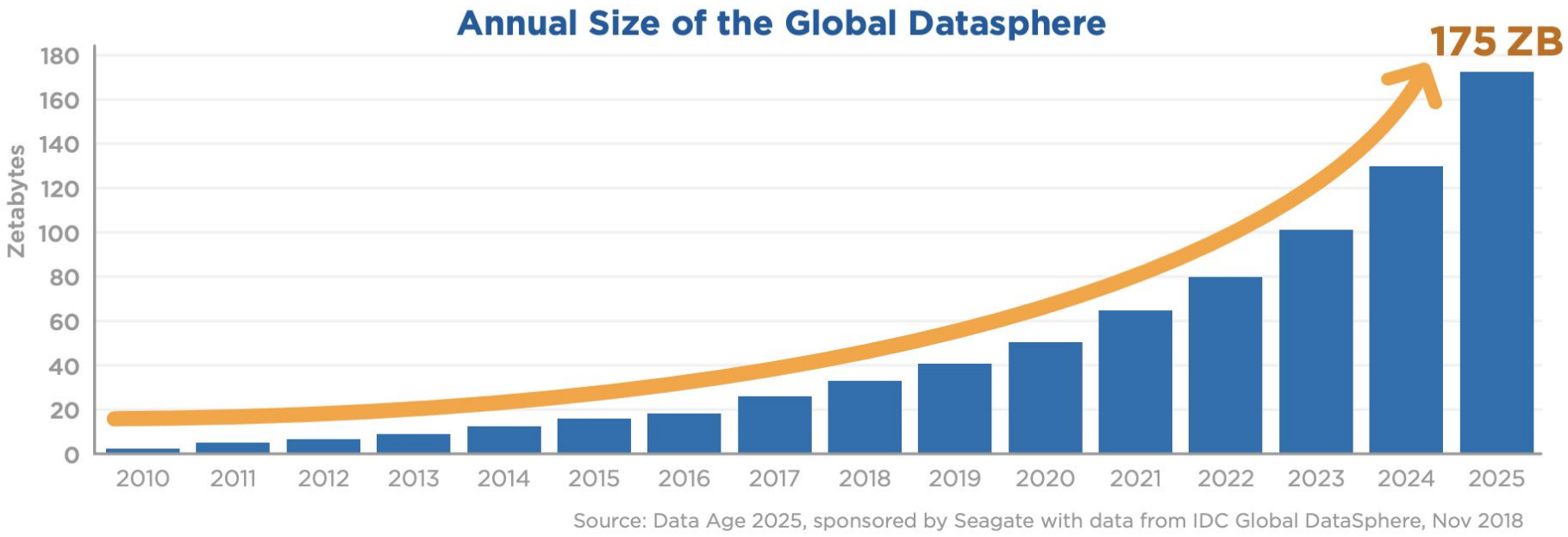


Source: https://www.domo.com/learn/data-never-sleeps-8 https://www.domo.com/learn/infographic/data-never-sleeps-9

MONASH University

Data Growth

Figure 1 – Annual Size of the Global Datasphere



Source: https://www.seagate.com/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf

Big Data Characteristics

- Volume
  - The quantity of data to be stored
- Velocity
  - The speed at which data enters the system and must be processed
- Variety
  - Variations in the structure of the data to be stored

FIGURE 14.1 ORIGINAL VIEW OF BIG DATA

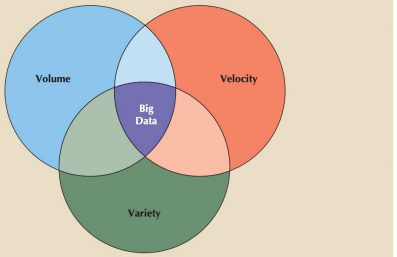
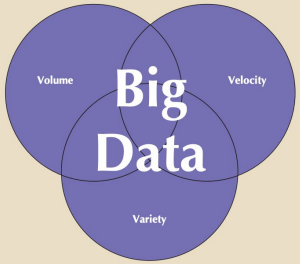


FIGURE 14.2 CURRENT VIEW OF BIG DATA



Big Data Characteristics: Volume

TABLE 14.1

STORAGE CAPACITY UNITS

TERM	CAPACITY	ABBREVIATION
Bit	0 or 1 value	b
Byte	8 bits	B
Kilobyte	1024* bytes	KB
Megabyte	1024 KB	MB
Gigabyte	1024 MB	GB
Terabyte	1024 GB	TB
Petabyte	1024 TB	PB
Exabyte	1024 PB	EB
Zettabyte	1024 EB	ZB
Yottabyte	1024 ZB	YB

* Note that because bits are binary in nature and are the basis on which all other storage values are based, all values for data storage units are defined in terms of powers of 2. For example, the prefix *kilo* typically means 1000; however, in data storage, a kilobyte = 2^{10} = 1024 bytes.

- Scaling up: keeping the same number of systems but migrating each one to a larger system
- Scaling out: when the workload exceeds server capacity, it is spread out across a number of servers

Scaling

- How do we scale current relational systems? SQL designed for database as a single physical entity
  - Purchase bigger "boxes": costly and has real limits
  - Increase the number of processors, yielding parallel computation/database with complex issues to handle
  - Distribute database – challenges to maintain ACID transaction principles and issues of availability/consistency
- The rise of OO programming in the 80's also highlighted a problem known as the "Impedance Mismatch"
  - The program treats items as objects, but they need to be mapped to relational tables ("de aggregating" the object)
  - Also issues about "private" vs "public" (relational about need, OO absolute characteristic of data)

Scaling continued

- Big players, notably Google and Amazon chose a different path
  - Lots and lots of smaller boxes ("commodity" servers)
  - Non relational structure
  - Google: Bigtable
    - https://research.google/pubs/pub27898/
    - https://cloud.google.com/bigtable/docs/overview
    - Used for wide range of apps Gmail, Google Earth, YouTube
  - Amazon: Dynamo
    - http://www.read.seas.harvard.edu/~kohler/class/cs239-w08/decandia07dynamo.pdf
    - Based on Dynamo: https://aws.amazon.com/dynamodb/

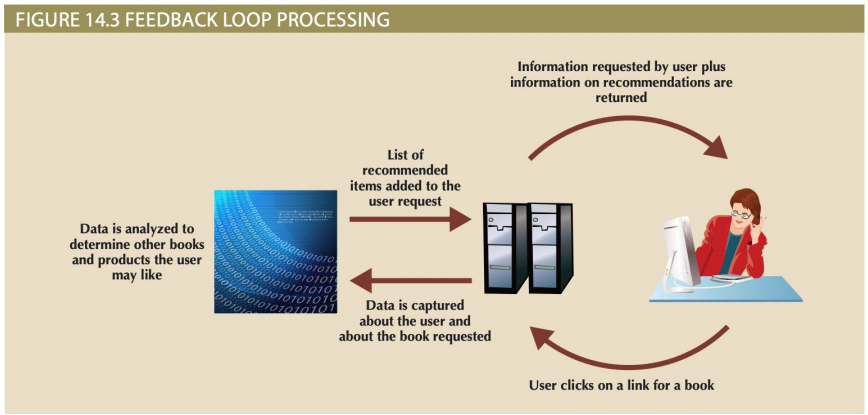
Scaling continued

- Term "NoSQL" coined by John Oskarsson in 2009 after calling a ..."free meetup about “open source, distributed, non relational databases” or NOSQL for short"…
  - http://blog.oskarsson.nu/post/22996139456/nosql-meetup
- Characteristics
  - Non relational,
  - mostly open source,
  - distributed (cluster friendly),
  - schema-less (no fixed storage schema)

Big Data Characteristics: Velocity

- Stream processing: focuses on input processing and requires analysis of data stream as it enters the system
  - CERN Large Hadron Collider 600TB per second ⇨ 1 GB per second
- Feedback loop processing: analysis of data to produce actionable results

FIGURE 14.3 FEEDBACK LOOP PROCESSING



Fast Data Processing

- Computer systems
  - –Parallel computer: A single machine with massive number of CPUs.
  - –Cluster of computers: Multiple machines connected via network; Commodity computer.
- Database structure
  - –Non-relational database (NoSQL)
    - •No update, append only. Optimised for a ’main’ operation
    - •Examples: MongoDB, Cassandra
  - –Distributed File Systems
    - •HDFS (Hadoop File Systems) / Parquee File Systems
- Parallel data processing
  - –Hadoop / Spark
- In Memory database

Big Data Characteristics: Variety

- Structured data: fits into a predefined data model
  - Relational databases
  - Incoming data decomposed under normalisation rules to fit the data model
- Unstructured data: does not fit into a predefined model
  - Big Data requires that the data is captured in its natural format as generated without imposing a data model on it
- Semi structured data: combines elements of both

TABLE 14.2

ADDITIONAL Vs OF BIG DATA

CHARACTERISTIC	DESCRIPTION
Variability	Data meaning changes based on context.
Veracity	Data is correct.
Value (Viability)	Data can provide meaningful information.
Visualization	Data can be presented in such a way as to make it understandable.

FIT3176 - Advanced database design

Learning outcomes

Expand all

On successful completion of this unit, you should be able to:

1. Describe various non-relational database systems, including NoSQL;
2. Compare and contrast between relational and non-relational database design and model;
3. Design database systems using document-store and column-store design techniques;
4. Explain transactions systems in non-relational systems;
5. Implement and manipulate document-store and column-store database systems;
6. Investigate graph model in database systems;
7. Demonstrate graph query processing.

FIT3182 - Big data management and processing

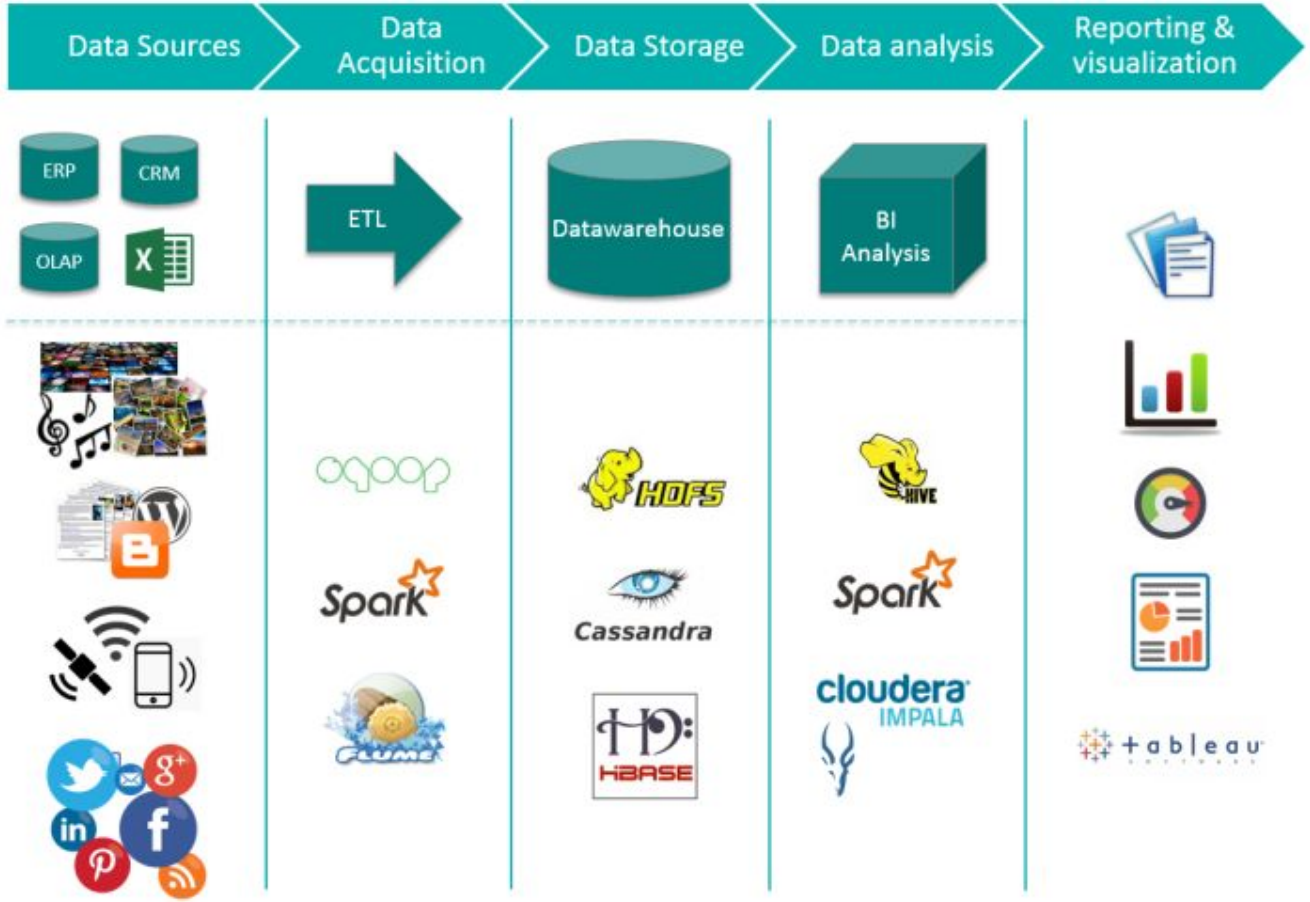
Learning outcomes

Collapse all

On successful completion of this unit, you should be able to:

1. identify big data concepts and technologies;
2. write and interpret parallel database processing algorithms and methods;
3. use big data processing frameworks and technologies;
4. describe and compare NoSQL technologies;
5. use big data streaming technologies.

Data Processing Ecosystem



http://www.clearpeaks.com/blog/big-data/big-data-ecosystem-spark-and-tableau

"Horses for Courses"

- Conventional RDBMS will continue play an important and significant role in OLTP (Online Transactions Processing)
- Increasingly now a *range* of database products are available, need to select appropriate product/model for task at hand.

FIT2094-FIT3171
Scheduled Final Assessment

2021 S2 Scheduled Final Assessment

- **Timed: 2 hours 10 minutes (reading time included)**
- **eExam platform: https://eassessment.monash.edu/**
- **Close book (no cheat sheets), online supervision, no aids (e.g. dictionary).**
- **Standard stationary allowed (e.g. pencil, pen, ruler, eraser)**
- **Learn more here: https://www.monash.edu/exams/electronic-exams**
  - **eExams with online supervision**
  - For the modelling question makes use of a hybrid question
    - write answer on paper (***indicate*** **answered on the question**)
      - 2 working pages, **submit one page with complete diagram**
    - photograph with phone and upload via QR code (after paper has been completed) - important that you practice this process by completing General Knowledge 2021 practice exam
- **Note that the scheduled final assessment is a time-pressured test**
  - **manage your time wisely**

2021 S2 Final Assessment Structure

- 100 marks 50% of your final mark in FIT2094-FIT3171.
  - Minimum to pass the unit overall:
    - **45%** in-semester, **45%** final assessment and **50%** overall
  - *Assignment 2B marking will not be finalised before you sit your Final Assessment, we will provide you with sample solution couple of days before the final assessment*
- Questions:
  - Cover theory and application
  - Timing is crucial - 100 marks, 120 mins - **1 mark/minute target**
  - SQL/PHP/PL-SQL Case Study will be released on Moodle the day before the scheduled final assessment day. **Do not bring the case study** to the Final Assessment **or discuss it on Ed forums**.
  - No tables/data provided on FITUGDB, no access to SQL Developer, LucidChart and other softwares
- Final Assessment when? It's your responsibility to check the date/time

2021 Final Assessment

- The teaching staff does NOT have any authority on the eExam environment. If you have any problem during the exam, please contact the eExam team (see Getting help and support). Emailing teaching staff or posting on the forum will NOT solve the problem.

- All content specified during your semester of study in FIT2094 - FIT3171 is examinable, including but not limited to:
  - Pre-reading (weekly Coronel & Morris chapters)
  - Workshop/Workshop Q&A Slides and Videos
  - Tutorial Notes, and
  - all other Moodle Materials (except where explicitly stated NOT EXAMINABLE).

2021 Mock Final Assessment

- Serves to provide an overview of the general structure of the final assessment paper only.
- Available Thursday, 21st October 2021 12 Noon
  - Sample solution available Wednesday, 27th October 9AM (please attempt the mock paper before accessing the sample solution)
  - Closes 6 hours before the Scheduled Final Assessment
  - Forums also closes 6 hrs before Scheduled Final Assessment
- Link on Moodle "Scheduled Final Assessments" block/page
- To protect the integrity of the paper: NO ACTUAL FINAL ASSESSMENT PAPER QUESTIONS are included; and the COMPOSITION OF THE SUB QUESTIONS and MARK DISTRIBUTION in each part are SUBJECT TO CHANGE.
- Suggest you leave it until you can attempt it under time limit of 2 hours 10 minutes (**time control is critical**)
- Allows multiple attempts

Workshop Week 2 and 5 – Data Modelling

INCLUDING BUT NOT LIMITED TO THESE TOPICS...

- Conceptual vs Logical Level
- Entity
  - –Strong vs weak
  - –Associative entity
- Types of attributes
- Relationship
  - –Type : one-to-one, one-to-many, many-to-many
  - –Cardinality and Participation
  - –Identifying vs Non-identifying.
- Mapping from Conceptual to Logical
  - –e.g. Mapping many-to-many
- FIT3171 - UML

Workshop Week 3 – Relational Model

INCLUDING BUT NOT LIMITED TO THESE TOPICS...

- Relational model properties.
- Keys
  - Superkey, Candidate Key, Primary Key
  - Foreign Key
- Data Integrity
  - Entity integrity
  - Referential Integrity
- Relational Algebra
  - Understanding of efficiency

Workshop Week 4 – Normalisation

INCLUDING BUT NOT LIMITED TO THESE TOPICS...

- UNF to 3NF
  - –Mapping form to UNF
  - –UNF to 1NF – remove repeating group
  - –1NF to 2NF – remove partial dependency (***general*** definition)
  - –2NF to 3NF – remove transitive dependency.
- Dependency diagrams
  - –Use the general definition (partial dependency based on all CKs)
  - –Partial in 1NF, Transitive in 2NF, Full in 3NF
    - •use this notation: cust_id → cust_name, cust_address
- Be careful in choosing the PK!
- Mapping a set of 3NF relations to a logical model

Workshop Week 6 and Week 8– DDL and DML

INCLUDING BUT NOT LIMITED TO THESE TOPICS...

- DDL
  - CREATE TABLE statements
    - Primary key definition
    - Foreign key definition
    - Other Constraints
  - ALTER
  - DROP
- DML
  - INSERT
    - Adherence to referential integrity constraints and the order of insertion
  - Oracle Sequence
  - UPDATE (DML)
  - DELETE (DML)

Workshop Week 7, 9 and 11 – SQL

INCLUDING BUT NOT LIMITED TO THESE TOPICS...

- Single table retrieval with predicate
- Join
  - Inner join
  - Outer join
- Aggregate functions
- Set Operators
- Subquery
- Oracle functions

  TO_CHAR, TO_DATE, NVL, UPPER, LOWER, ROUND, LPAD, RPAD

DB Connectivity, Web Technology and PL-SQL

INCLUDING BUT NOT LIMITED TO THESE TOPICS...

- Web database connectivity
  - –Understanding of the principles and ALL core concepts:
    - •Database middleware
    - •Web to database middleware
  - –Using PHP to communicate with databases
    - •must understand and able to write php code which relates to database (i.e. PHP related workshop and tutorial materials are examinable)
  - –Database design frameworks
    - •modern frameworks
    - •ORM
  - –Security → SQL Injection, SQL Injection Prevention

- FIT3171 - TRIGGER
  - –maintain data integrity
  - –update value of a column

Workshop Week 8 – Transaction Management

INCLUDING BUT NOT LIMITED TO THESE TOPICS...

- Transactions
  - –transaction boundaries (start and end)
  - –use of commit/rollback
- ACID properties
- Transaction problems (serial vs interleaved)
- Transaction management with locks
- Wait for graphs
- Restart and recovery using transaction log

Workshop Week 12

- The content of week 12's workshop
  - –Database Trends
  - –Future directions

Is NOT examinable (questions relate to this week's new content will not appear on the exam)

Consultations for Scheduled Final Assessment

- Online consultation sessions will be provided
  - Details to be posted on Moodle

- Don’t come to consultations in the hope of obtaining some 'extra' information about the paper
  - Session intended to clear up any issues *YOU* find as you prepare for the Scheduled Final Assessment

http://blog.proqc.com/administrative-professionals-quality-thank-you/