MONASH University

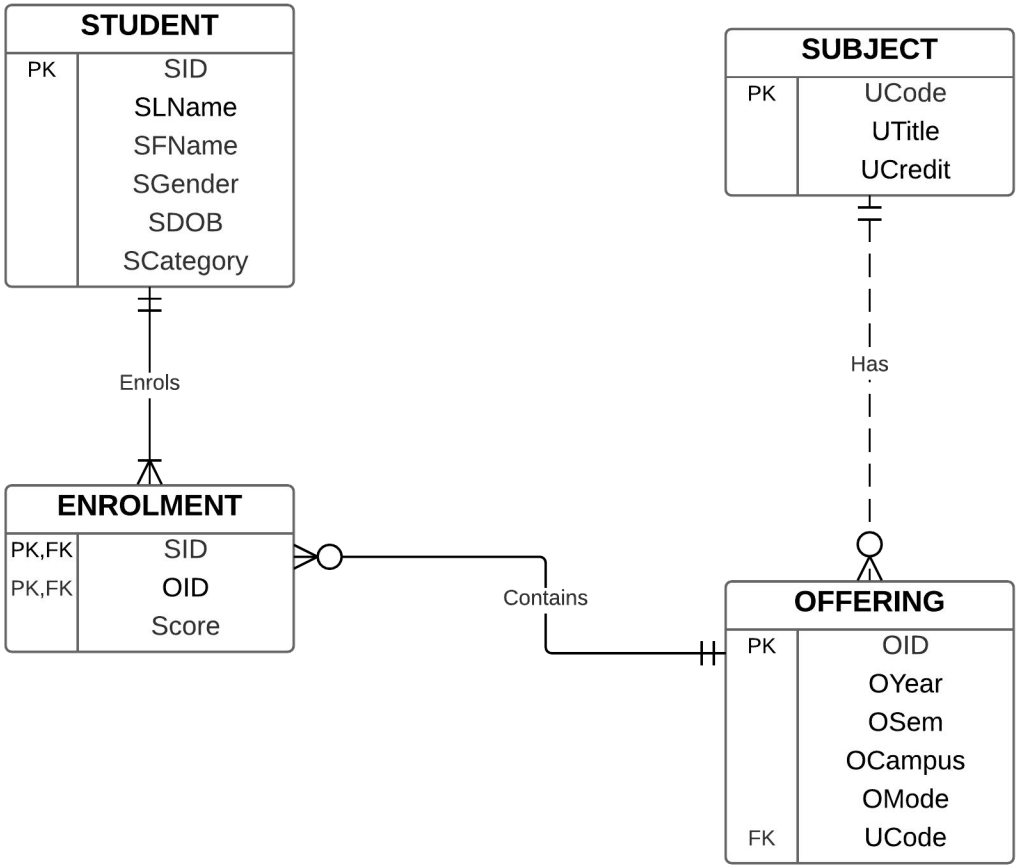
MONASH
INFORMATION
TECHNOLOGY

Database Current Trends
Exam Preparation

FIT9132

GROUP OF EIGHT AUSTRALIA

Operational Database



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 FIT9132?
- 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.

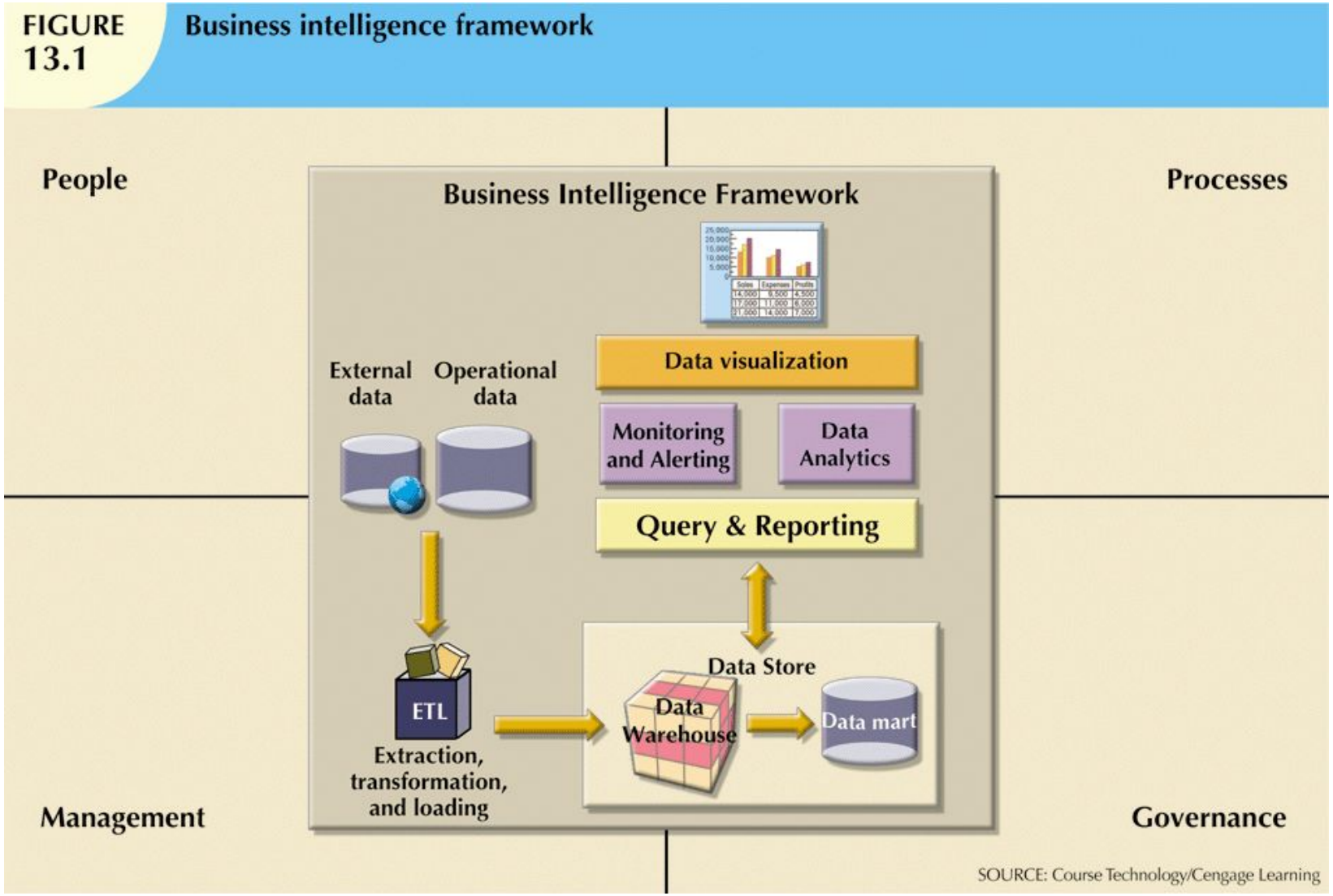


FIGURE 13.1
Business intelligence framework
People
Processes
Business Intelligence Framework
External data
Operational data
Data visualization
Monitoring and Alerting
Data Analytics
Query & Reporting
ETL
Extraction, transformation, and loading
Data Store
Data Warehouse
Data mart
Management
Governance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

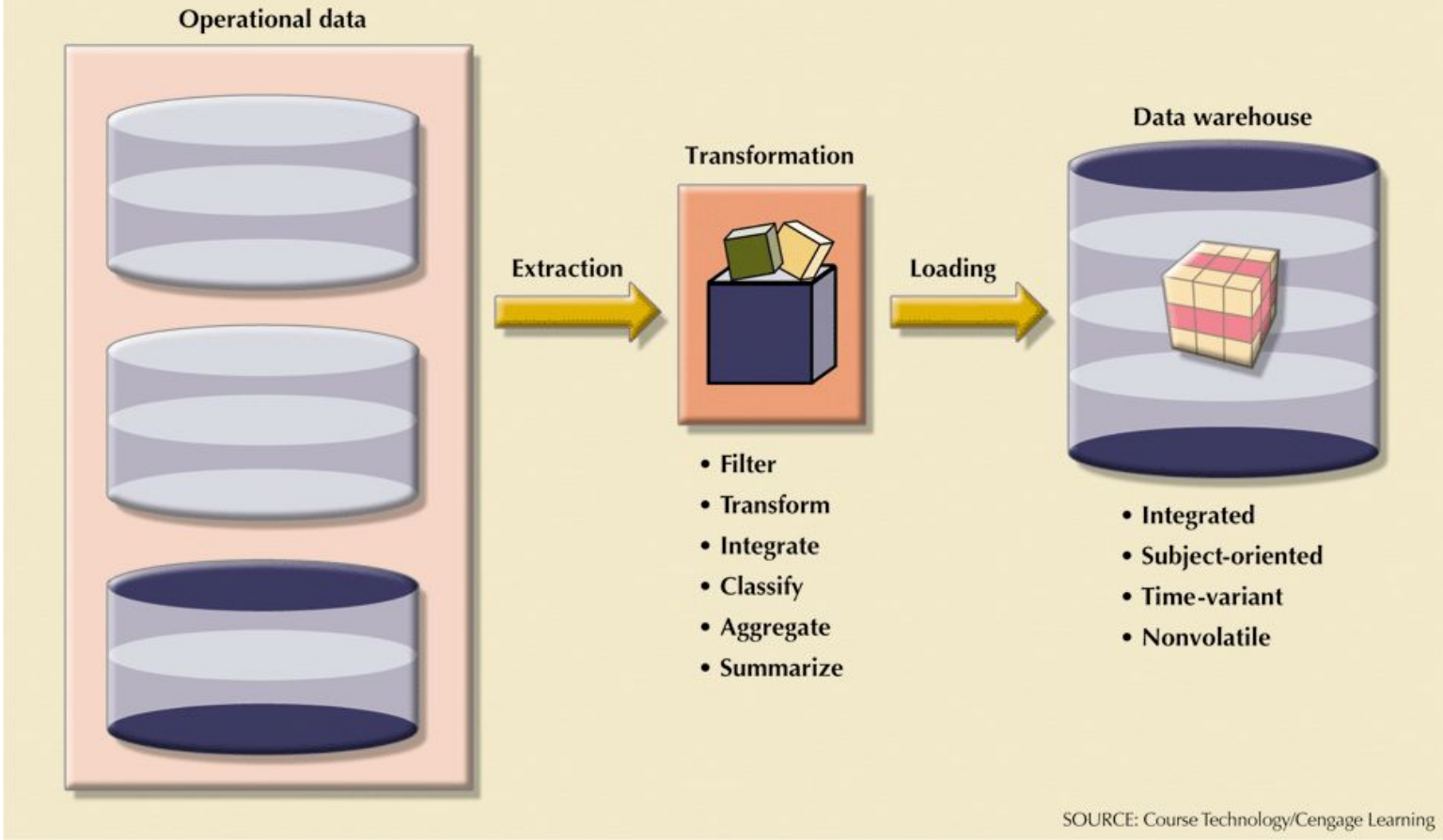
Decision Support Database Requirements

- Specialized DBMS tailored to provide fast answers to complex queries
- Three main requirements
  - Database schema
  - Data extraction and loading
  - Database size
- Database schema
  - Complex data representations
  - Aggregated and summarized data
  - Queries extract multidimensional time slices
- Data extraction and filtering
  - Supports different data sources
    - Flat files
    - Hierarchical, network, and relational databases
    - Multiple vendors
  - Checking for inconsistent data

The Data Warehouse (FIT5195, FIT5137)

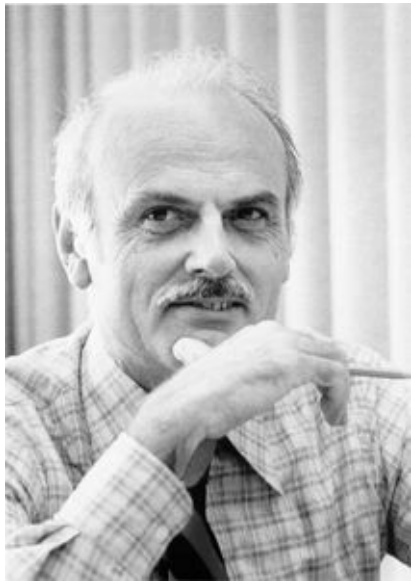
- Database size
  - In 2013, eBay had around 50 Petabytes of data in its data warehouses (50,000 Terabytes)
  - DBMS must support very large databases (VLDBs)
- Integrated, subject-oriented, time-variant, and nonvolatile collection of data
  - Provides support for decision making
- Usually a read-only 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

Database Hall of Fame



E.F “Ted” Codd

C.J Date

Peter Chen

Michael Stonebraker

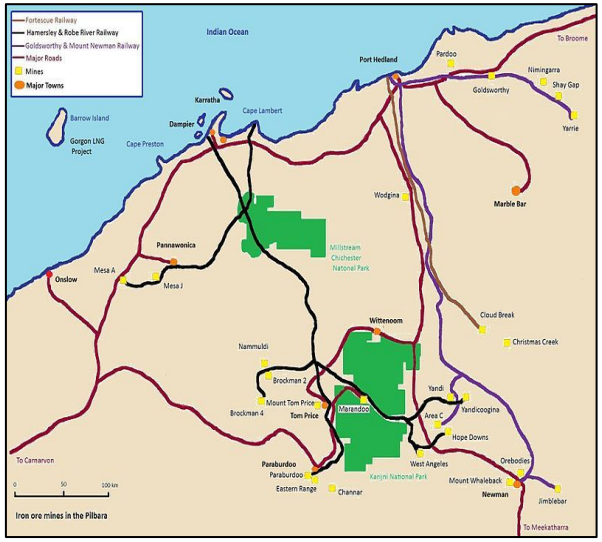
Internet of Things (IoT)

What is happening with data?

1 ZB = 10^{21} bytes = 1 billion terabytes = 1 trillion gigabytes

http://www.emc.com/collateral/analyst-reports/idc-digital-universe-2014.pdf
https://www.emc.com/leadership/digital-universe/index.htm

Railway In Mining



- Pilbara region, WA
- Trains perform round trips from the mining site to the port
- Loaded minerals and ores

- Length: > 2KM
- Load: > 10 Ton/car
- Speed: 5-10 Km/hr

- Instrumented Ore Car (IOC)
- Expensive Sensors
- Trained Professionals to maintain the sensors

Challenges

(1)
Expensive sensors that require professionals to maintain

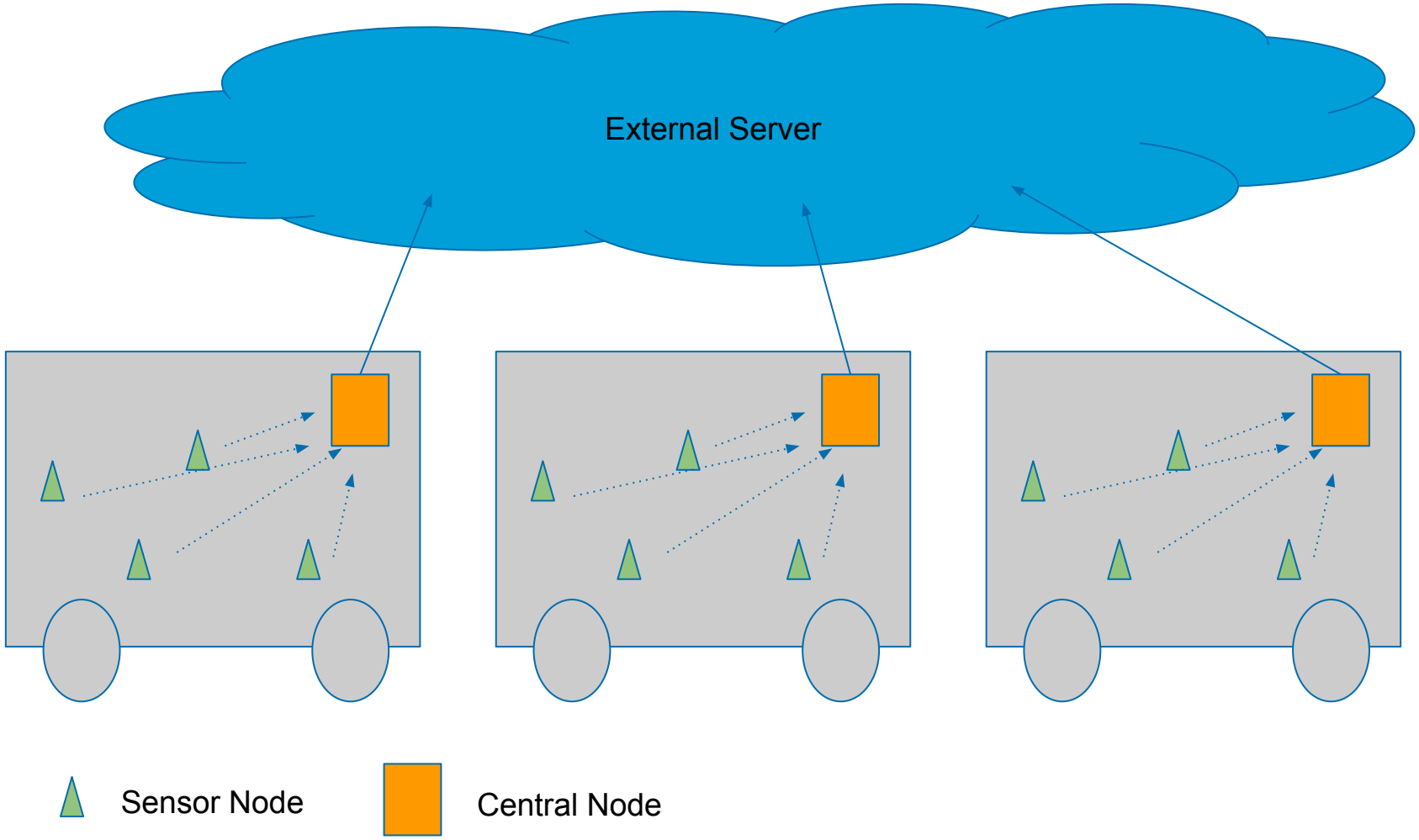
(2)
Large volume of data generated by the sensors

→ **Cheap, self-configured, massive array of sensors**

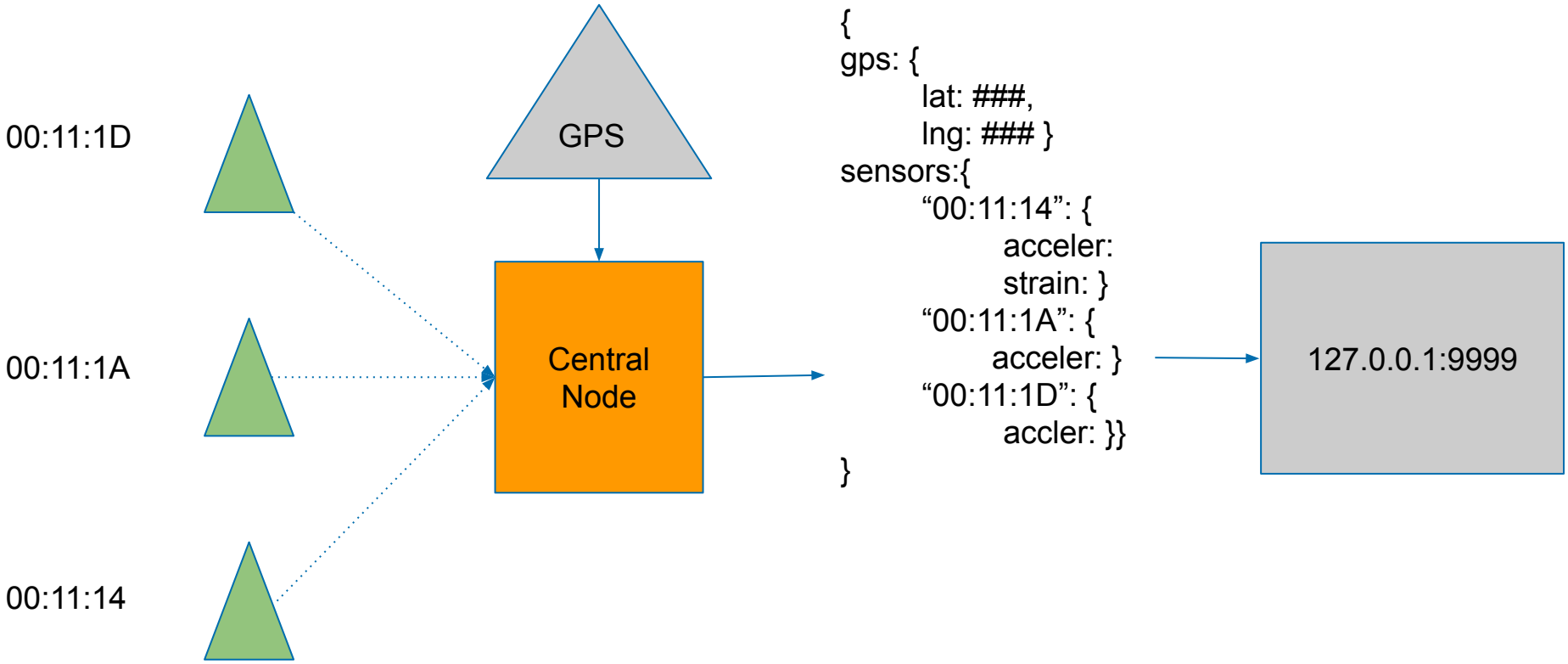
→ **Fast data processing and retrieval**

Needs expertise from Eng and IT

Network Structure



Central Node Process



How Big is the Data?

Quantity	Data Returned
Timestamp	12-Jun-2015; 09:35:15
Geo-location	N35°43.57518,W078°49.78314
Direction	ToMine
Acceleration	0.285g
Pressure	65psi
Ambient temperature	73 degrees F
Surface temperature	78 degrees F
Humidity	35%

- 16 Sensors
- 200 Ore Cars
- 25 Records Per Second

16 * 200 * 25 = 80,000 records/sec

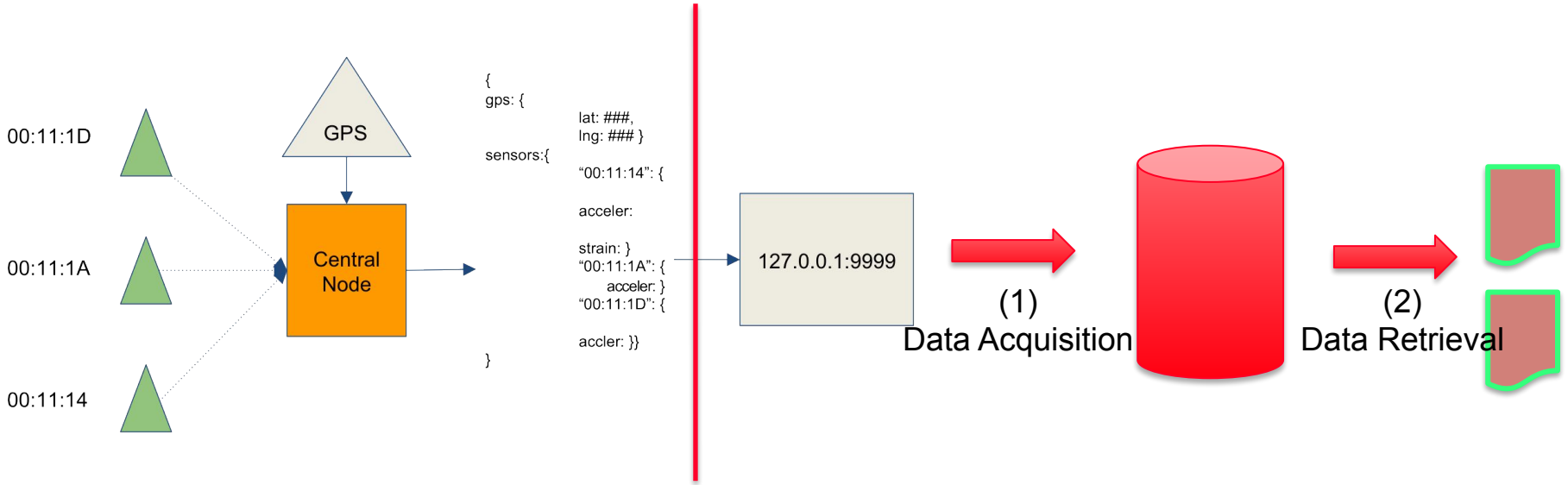
```
Welcome to Ubuntu 14.04.3 LTS (GNU/Linux 3.13.0-46-generic x86_64)

 * Documentation:  https://help.ubuntu.com/
ubuntu@master:~$ mongo
MongoDB shell version: 3.0.4
connecting to: test
2015-11-06T11:49:56.337+1100 I CONTROL  [initandlisten]
2015-11-06T11:49:56.337+1100 I CONTROL  [initandlisten] ** WARNING:
/sys/kernel/mm/transparent_hugepage/defrag is 'always'.
2015-11-06T11:49:56.337+1100 I CONTROL  [initandlisten] **        We suggest setting it to
'never'
2015-11-06T11:49:56.337+1100 I CONTROL  [initandlisten]
>  Use IRT
> db.sensordata.find().pretty()


{
        "_id" : ObjectId("5663ce2ce4b099b72ceca8c2"),
         “gps”: { "GPSLat" : -21.63893238,"GPSLon" : 116.70659242},
        "SomatTime" : 74711,
        "CarOrient" : 30.2,
        "EorL" : 1,
        "Direction" : "ToPort    ",
        "minSND" : 0,
        "iSegment" : 5876,
        "maxSND" : 0,
        "PipeA" : 0,
        "maxCFB" : 0,
        "minCFB" : 0,
        "Bounce" : 0,
        "minCFA" : 0,
        "maxCFA" : 0,
        "kmh" : 30.2,
        "PipeB" : 0,
        "Rock" : 0,
        "accR3" : 0,
        "accR4" : 0,
        "maxBounce" : 0,
        "LATACC" : 0
}

Type "it" for more
>
```

Big Data Processing



Two main problems:

1. How to receive data … massive amount of data
2. How to retrieve data … very fast

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

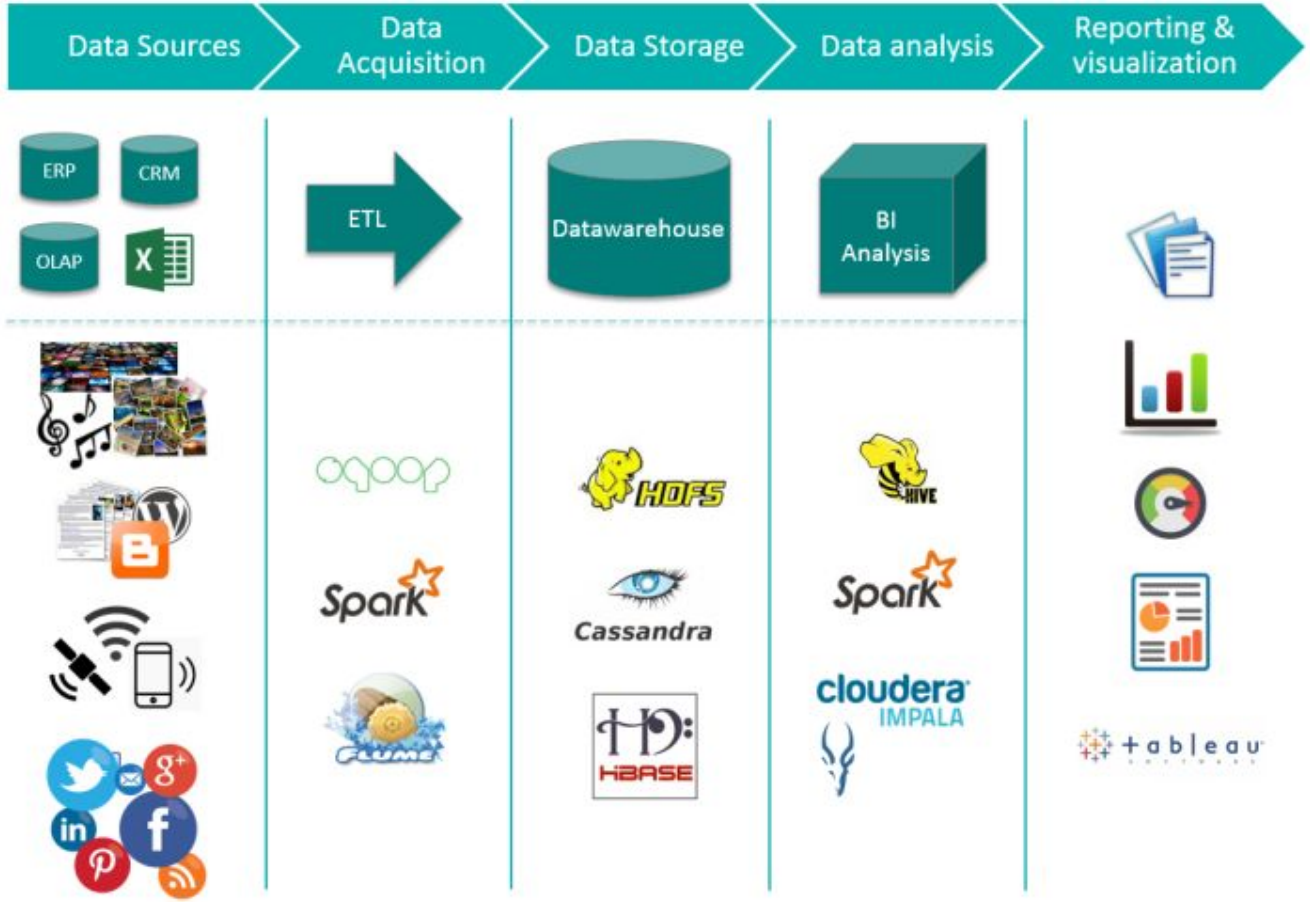
Scaling continued

- Big players, notably Google and Amazon chose a different path
  - Lots and lots of smaller boxes ("commodity" servers)
  - Non relational structure
  - Google: Bigtable
    - http://static.googleusercontent.com/media/research.google.com/en//archive/bigtable-osdi06.pdf
  - Amazon: Dyanmo
    - http://www.read.seas.harvard.edu/~kohler/class/cs239-w08/decandia07dynamo.pdf
- 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)
  - See MongoDB https://www.mongodb.com/nosql-explained

Fast Data Processing (FIT5202, FIT5148)

- 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

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.

FIT9132 Exam

2019 Exam Format

- **2 HOUR writing**
- **10 minutes reading**
  - **Question Booklet (do not write on)**
  - **Answer Booklet**
- 100 marks 50% of your final mark in FIT9132.
  - Minimum to pass FIT9132 overall:
    - 40% non-exam, 40% exam and 50% overall
- Questions:
  - 5 questions – theory and application
  - Sample paper on Moodle (Exam block).

MONASH University

Office Use Only

Semester Two 2019
Examination Period

Faculty of Information Technology

EXAM CODES: FIT9132

TITLE OF PAPER: Introduction to databases – Question Booklet

EXAM DURATION: 2 hours writing time

READING TIME: 10 minutes

THIS PAPER IS FOR STUDENTS STUDYING AT: (tick where applicable)

☒ Caulfield ☐ Clayton ☐ Parkville ☐ Peninsula
☐ Monash Extension ☐ Off Campus Learning ☒ Malaysia ☐ Sth Africa
☐ Other (specify)

During an exam, you must not have in your possession any item/material that has not been authorised for your exam. This includes books, notes, paper, electronic device/s, mobile phone, smart watch/device, calculator, pencil case, or writing on any part of your body. Any authorised items are listed below. Items/materials on your desk, chair, in your clothing or otherwise on your person will be deemed to be in your possession.

No examination materials are to be removed from the room. This includes retaining, copying, memorising or noting down content of exam material for personal use or to share with any other person by any means following your exam.
Failure to comply with the above instructions, or attempting to cheat or cheating in an exam is a discipline offence under Part 7 of the Monash University (Council) Regulations, or a breach of instructions under Part 3 of the Monash University (Academic Board) Regulations.

AUTHORISED MATERIALS

OPEN BOOK	☐ YES	☒ NO
CALCULATORS	☐ YES	☒ NO
SPECIFICALLY PERMITTED ITEMS	☐ YES	☒ NO

if yes, items permitted are:

Any written work placed on this booklet will NOT BE MARKED

MONASH University

Office Use Only

Semester Two 2019
Examination Period

Faculty of Information Technology

EXAM CODES: FIT9132

TITLE OF PAPER: Introduction to databases – Answer Booklet

EXAM DURATION: 2 hours writing time

READING TIME: 10 minutes

THIS PAPER IS FOR STUDENTS STUDYING AT: (tick where applicable)

☒ Caulfield ☐ Clayton ☐ Parkville ☐ Peninsula
☐ Monash Extension ☐ Off Campus Learning ☒ Malaysia ☐ Sth Africa
☐ Other (specify)

During an exam, you must not have in your possession any item/material that has not been authorised for your exam. This includes books, notes, paper, electronic device/s, mobile phone, smart watch/device, calculator, pencil case, or writing on any part of your body. Any authorised items are listed below. Items/materials on your desk, chair, in your clothing or otherwise on your person will be deemed to be in your possession.

No examination materials are to be removed from the room. This includes retaining, copying, memorising or noting down content of exam material for personal use or to share with any other person by any means following your exam. Failure to comply with the above instructions, or attempting to cheat or cheating in an exam is a discipline offence under Part 7 of the Monash University (Council) Regulations, or a breach of instructions under Part 3 of the Monash University (Academic Board) Regulations.

AUTHORISED MATERIALS

OPEN BOOK	☐ YES	☒ NO
CALCULATORS	☐ YES	☒ NO
SPECIFICALLY PERMITTED ITEMS	☐ YES	☒ NO

if yes, items permitted are:

ALL ANSWERS WRITTEN IN THIS BOOKLET MUST BE IN BLUE or BLACK PEN, *PENCIL MAY BE USED FOR DIAGRAMS ONLY*

Candidates must complete this section if required to write answers within this paper

STUDENT ID: _ _ _ _ _ _ _ _ DESK NUMBER: _ _ _ _ _

Q1 (10)	Q2 (20)	Q3 (10)	Q4 (50)	Q5 (10)	Total (100)

Lecture Week 2 and 4 – Data Modelling

- 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

Lecture Week 3 – Relational Model

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

Lecture Week 5 – Normalisation

- UNF to 3 NF
  - Mapping form to UNF
  - UNF to 1 NF – remove repeating group.
  - 1NF to 2 NF – remove partial dependency.
  - 2NF to 3NF – remove transitive dependency.
- Dependency diagrams
- Be careful in choosing the PK!
- Mapping a set of 3NF relations to a logical model

Lecture Week 6 – Data Definition Language

- CREATE TABLE statements
  - –Primary key definition
  - –Foreign key definition
  - –Other Constraints
- ALTER
- INSERT
  - –Adherence to referential integrity constraints
    - •Order of insertion
- Oracle Sequence
- UPDATE (DML)
- DELETE (DML)

Lecture Week 7, 9 and 11 – SQL

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

Lecture Week 8 – Transaction Management

- Transaction.
- ACID properties.
- Transaction problems.
- Transaction management with locks.
- Wait For Graphs
- Restart and Recovery using Transaction Log.

Lecture Week 10 – PL/SQL

- Triggers
  - –Statement vs row triggers
  - –Correlation names (old/new)
  - –Mutating table restrictions
  - –Coding triggers
    - •IF THEN ELSE structure
    - •SELECT … INTO
    - •Stopping an action via raise_application_error

Lecture Week 12 – Database Trends

- The content of week 12's lecture
  - –Database Trends
  - –Big Data

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

Consultations for Final Exam

- The week before the exam, several consultation sessions will be provided.
  Tuesday, Wednesday, Thursday & Friday
  Details to be posted on Moodle

- Don't come to consultations in a hope to squeeze some useful information about final exam
  - Session intended to clear up any issues *YOU* find as you prepare for the exam

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