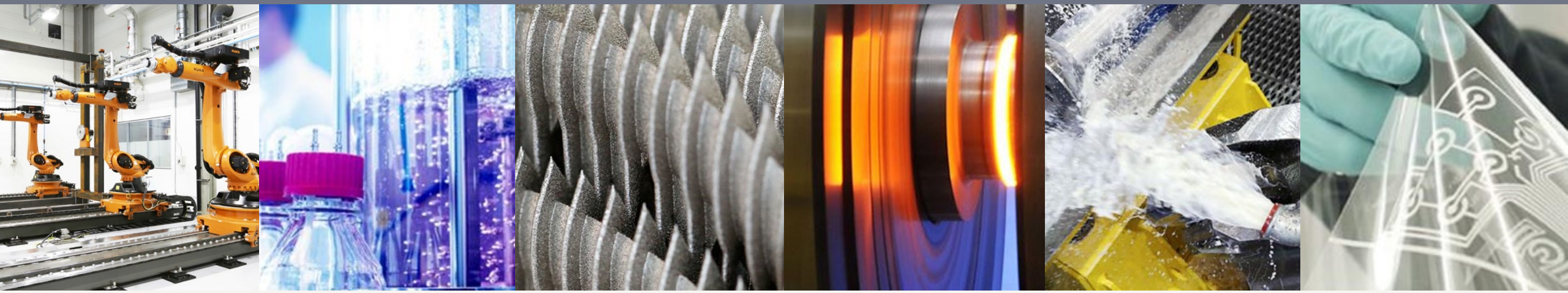
Emerging Skills Programme

Open Government Licence Guidelines

Emma Carroll-Walsh
Programme Exec
27/05/2021



What Is Open Government Licence (OGL)

Basic Guidelines

- User is free to copy, publish and transmit the information
- User is free to adapt the information
- User is free to exploit commercially and non-commercially
  - Combine it with other information
  - Include their own product or application
- User must acknowledge source of information by linking to any attribution statement specified by the provider
- Cannot use material in any way that suggests official status or that the provider endorses you or your use of the information
- Information is licensed ‘as-is’ so has no cover of warranties, obligations or liabilities
- Must follow UK Government Licensing Framework (UKGLF)

This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Introduction to Data Analysis and Machine Learning

This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Unit 2: Data Management in Manufacturing

This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Learning Objectives

1) Understanding practical data analysis: Processes and tools
   a. Understanding of data science, data modelling and data management tools as part of digital manufacturing; ☑
   b. Understand what makes collectable data valuable and improve manufacturing processes for industry and industrial ☑
   c. Application of suitable data protection protocols including regulatory requirements, data encryption, and data
2) Understanding the role of data ethics, regulation and data protection
   a. Basic knowledge of ethics and data protection in manufacturing industry;
   b. Understanding principles, tools and techniques of ethical data use and data protection implications.
3) Knowledge of machine learning
   a. Knowledge of taxonomy and principles of machine learning; ☑
   b. Understanding principle of machine learning and the difference between supervised and unsupervised learning. ☑
4) Ability to use machine learning algorithm and tools for digital manufacturing process
   a. Understanding machine learning tools and programming; ☑
   b. Be able to manipulate a dataset through RStudio, Python or others.

OGL National Composite Centre 2022
This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Unit Objectives

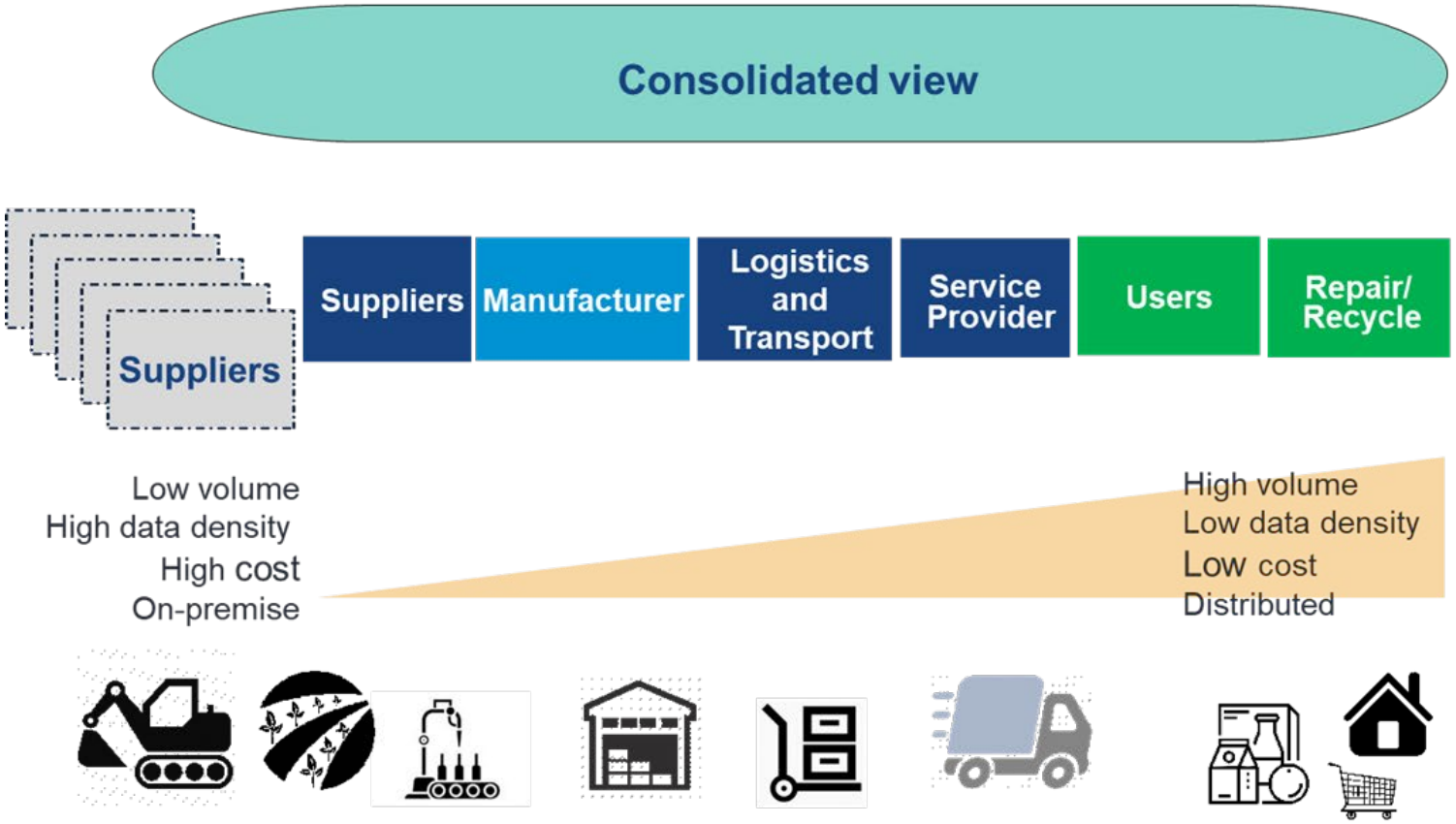
- ***Understand*** the ***value of data*** in the manufacturing supply chain.
- ***Understand*** and ***remember*** the concept of big data and the complexity of managing it.
- ***Understand*** the ***concepts*** of relational and non-relational (NoSQL) databases and that different types of data require different storage solutions.
- ***Understand*** the ***concept of data modelling*** and how to use it.
- ***Understand*** the ***difference*** between a data warehouse and a data lake and ***when to use*** each.
- ***Understand*** the ***concepts*** of data preparation, tidying and transforming data.

OGL National Composite Centre 2022
This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Data Used in Manufacturing

This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Industry 4.0 Requires Consolidated Data Across Supply Chain



Key I4.0 Data Challenges

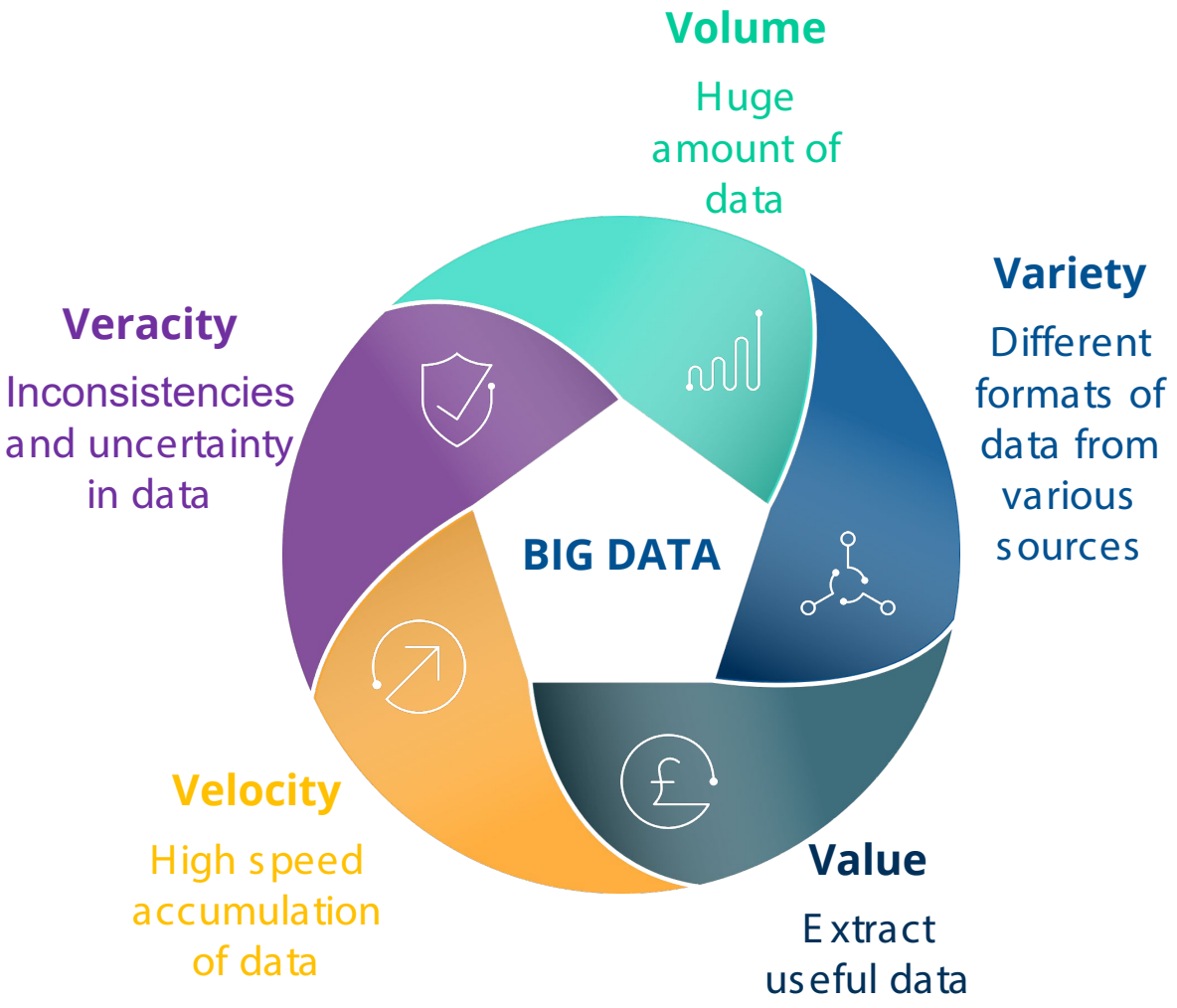
- The inability to extract knowledge from the data (value)
- The unstructured format of collected data (variety)
- The massive data to manage, store, and process (volume)
- Insufficient quality of the collected data (veracity)
- Insufficient data processing power (velocity)

Industry 4.0 Implementation Challenges and Opportunities: A Managerial Perspective

This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Big Data

- Big data is a term that describes the large volume of data – both structured and unstructured – that inundates the manufacturing supply chain on a day-to-day basis.
- It is not the amount of data that's important, it's what organisations do with the data that matters. Big data can be analysed for insights that lead to better decisions and strategic imperatives.



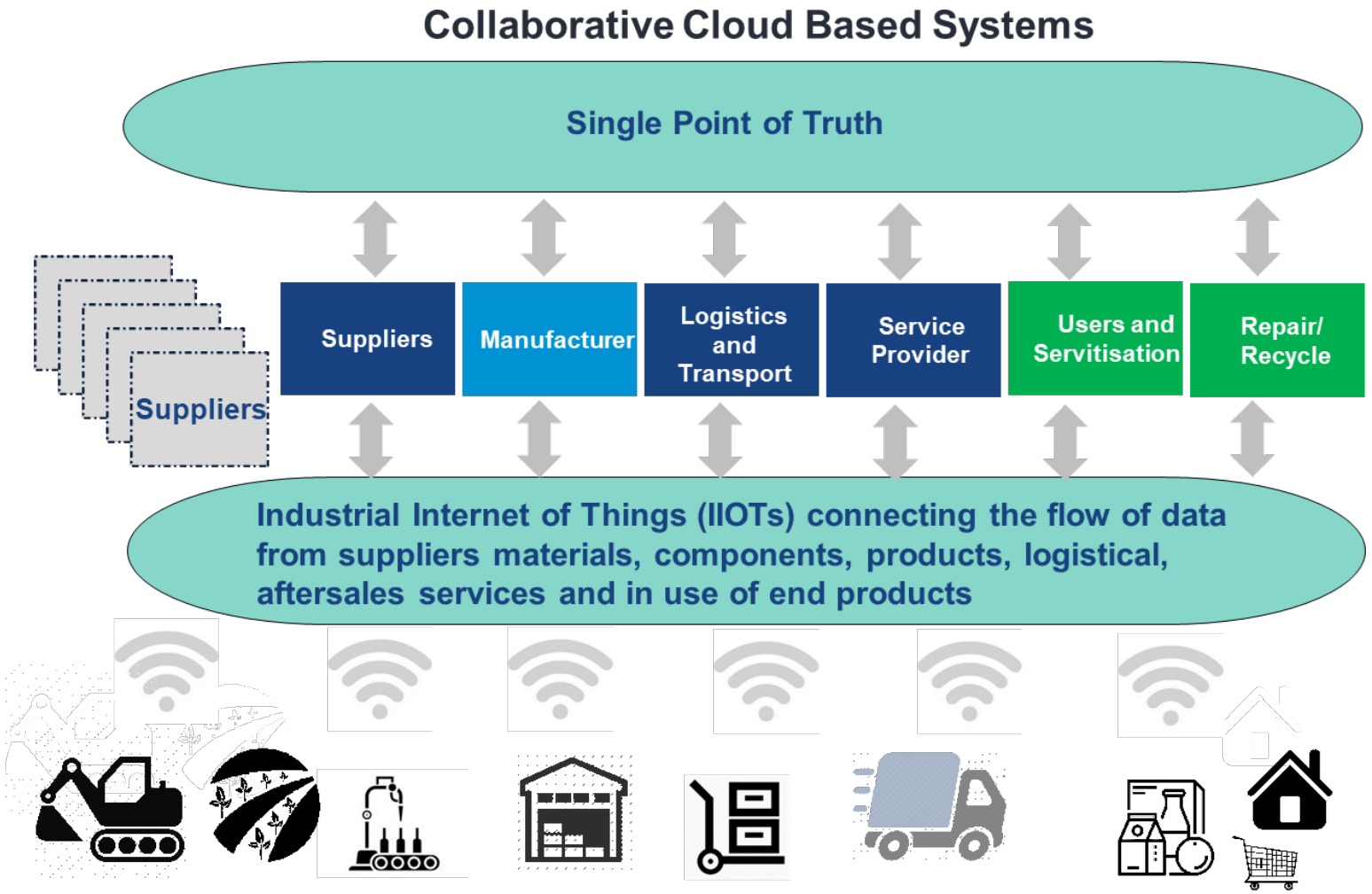
This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Current Reality

Siloed data is the norm in manufacturing environments.

This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Possible Solution to Data Consolidation Goal?



This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

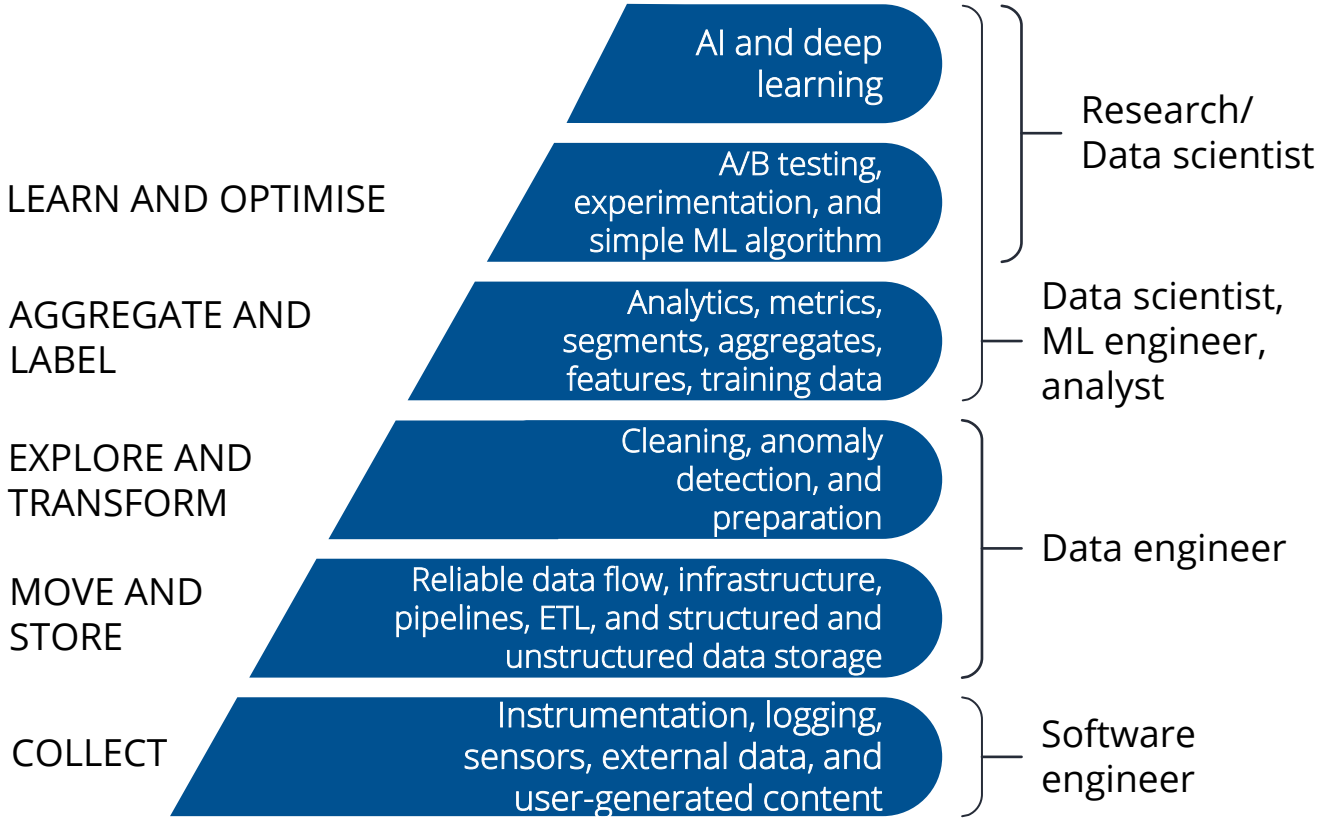
How do we Create a Single View of Supply Chain Data?

- Data management is complex…
  - How do we extract value from the data?
  - How do we deal with unstructured data?
  - How do we store and process massive amounts of data?
  - How do we ensure high data quality?
  - How do we deal with real-time information?
  - How do we manage different data formats?
  - How do we deal with security and privacy?
- In order to produce a solution, we will introduce some relevant technical concepts and consider how they might apply to the problem.

OGL National Composite Centre 2022
This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Plan to Solve the Silo Problem – People and Processes

- Solve connectivity issues e.g., comms, software APIs, etc.
- Metadata! Context is key. What sensor? History, calibration, etc.
- Transform data into standardized format (on/offline timestamps, numerical, etc.).
- Ensure accessibility, integrity, security.
- Combine data insights from various sources and contextualize data access: Find the right information at the right time.
- Create data insights: Refine, post-process, calculate inferred measurements, visualize, etc.
- Turn data insights into actions within and across projects, functions, business units, supply chain.



This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Plan to Solve the Silo Problem - Systems

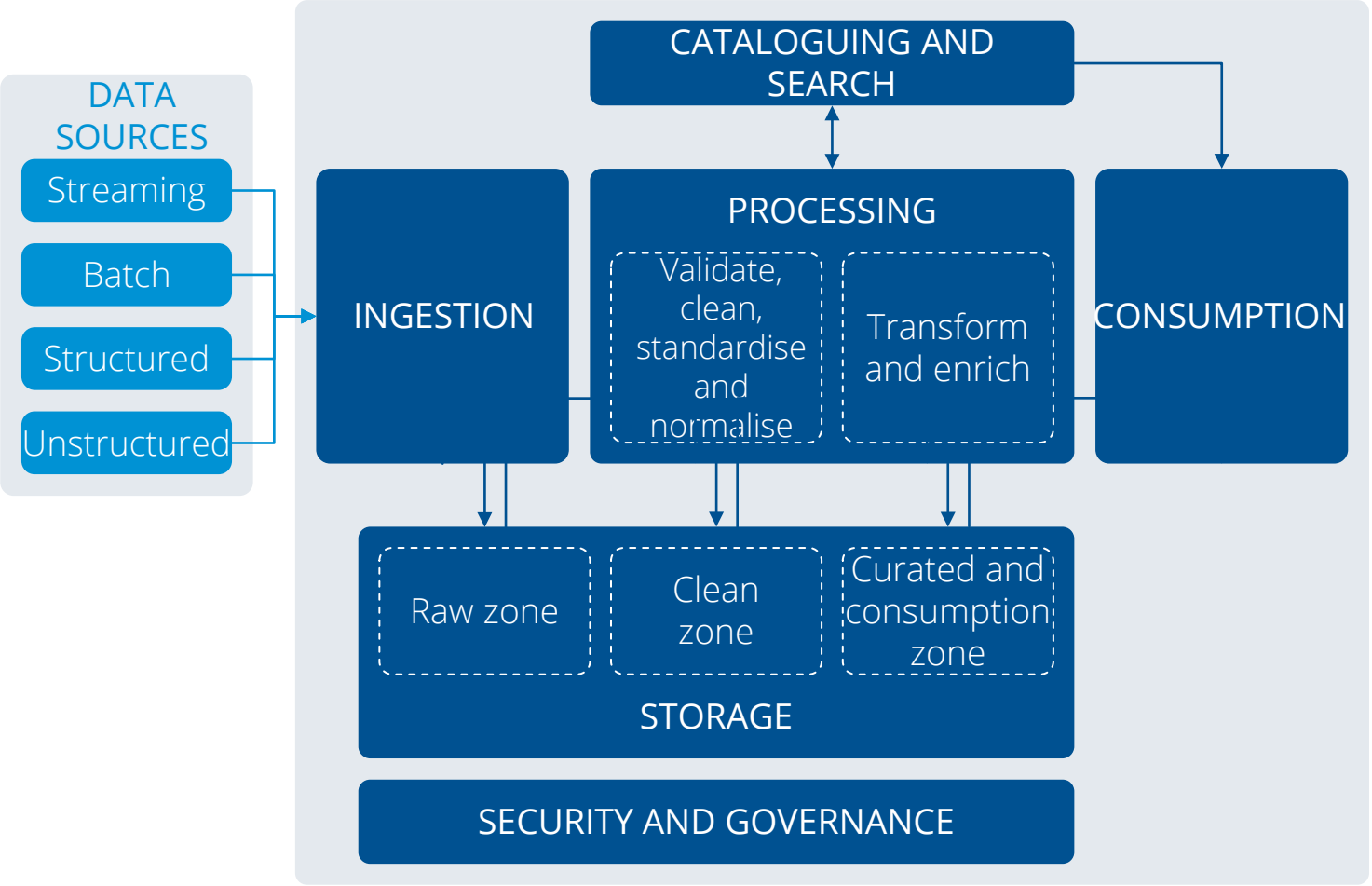
Architecture of typical cloud based advanced analytics framework.

Various types of data as input:

- Batch and real-time / streaming.
- Structured and unstructured.

Three stages of data transformation:

- Ingestion (curate, stage).
- Validate, clean, standardise, and normalise.
- Transform and enrich.



This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

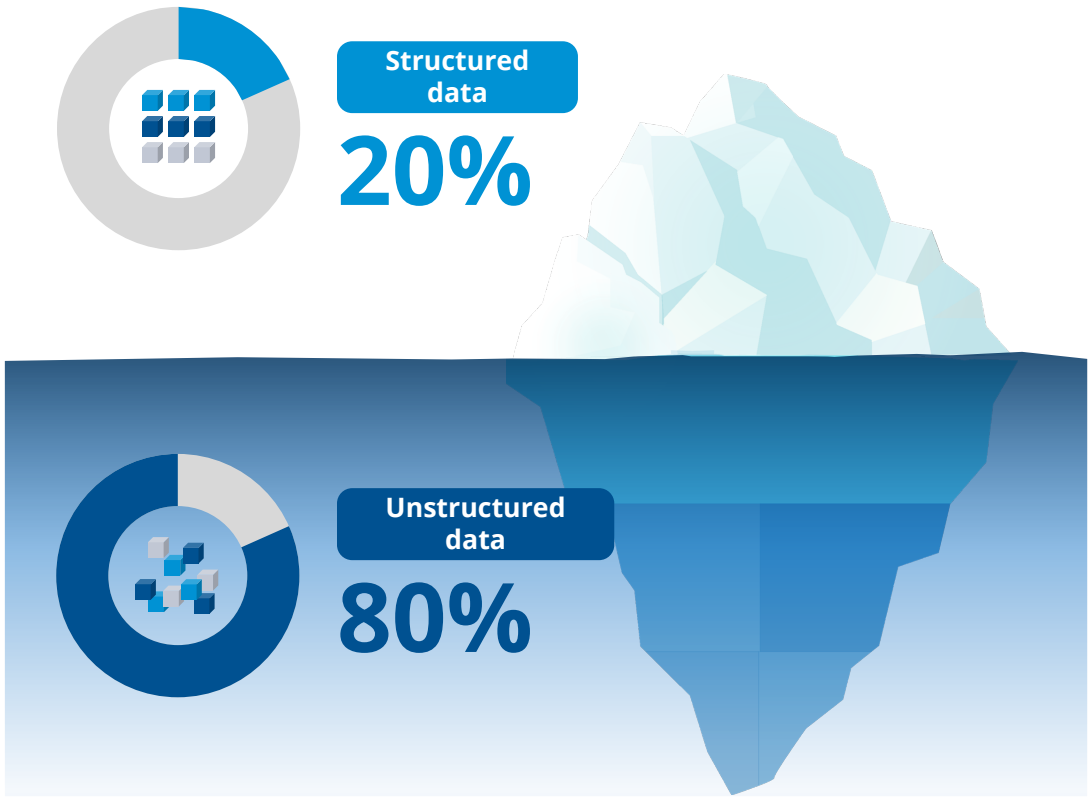
Structured Versus Unstructured Data

Structured data

- Can be displayed in rows, columns and relational databases i.e. tabular.
- Numbers, dates and strings.
- Estimated 20% of enterprise data.
- Requires less storage.
- Easier to manage and protect with legacy systems.

Unstructured data

- Cannot be displayed in rows, columns and relational databases i.e. non-tabular.
- Sensor data, images, audio, video, emails, spreadsheets.
- Estimated 80% of enterprise data.
- Requires more storage.
- More difficult to manage and protect with legacy systems.



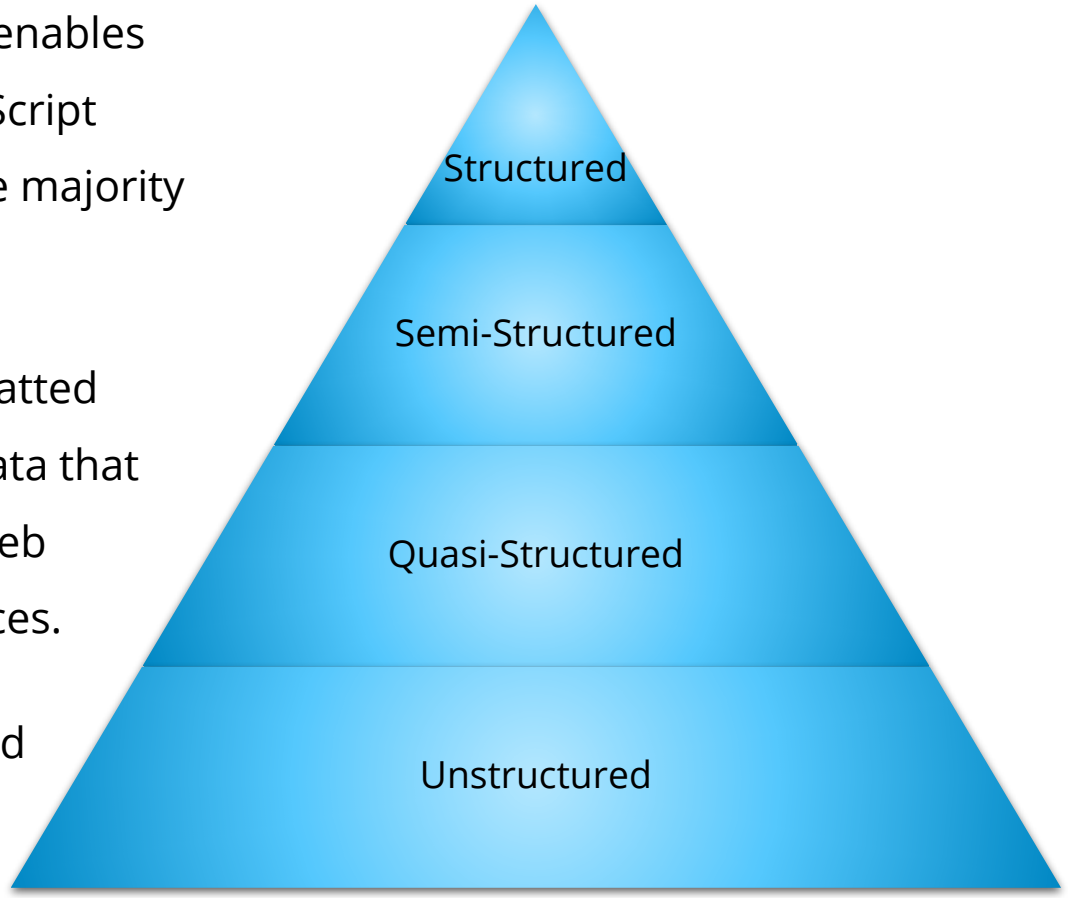
This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Further Classification

- **Semi-structured:** Text files with a discernible pattern that enables parsing such as Extensible Markup Language (XML) or JavaScript Object Notation (JSON) files. JSON is the format used for the majority of Internet of Things (IoT) implementations.

- **Quasi-structured:** Text files with formats that can be formatted with effort, tools, and time (for instance, web clickstream data that may contain inconsistencies in data values and formats). Web clickstreams are frequently used by cloud-based web-services.

Manufacturing data from IoT devices are therefore often stored in semi-structured format e.g., JSON.



This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Databases

- A database is a systematic collection of data which supports electronic storage and manipulation of data. There are two main types of databases:
  - Relational.
  - Non-relational or “NoSQL” (*means same thing*)(SQL- Structured Query Language).
- Relational databases are the most common and store structured data in a tabular format.
- NoSQL databases (aka "not only SQL") are non-tabular databases optimised for performance/ reliability and unstructured data.
- A Database Management System (DBMS) enable users to access databases on the systems, popular DBMSs are:
  - Relational (RDBMS).
    - IBM DB2, MySQL, Oracle Database, Teradata, PostgreSQL, Microsoft SQL Server, SQLite.
  - NoSQL.
    - Apache Cassandra (partition row store), MongoDB (document store), DynamoDB (key-value store), Apache HBase (wide column store).

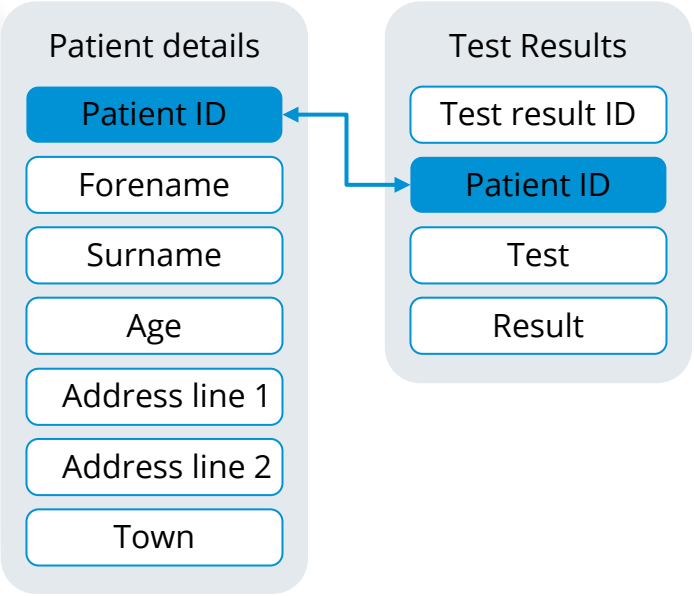
This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Relational Databases

- Invented by Edgar Codd (1970).
- Database / schema.
- Collection of tables.
- Tables / entities.
- Table contains records (rows) and fields (columns).
- Relationships between tables (entities) using primary and foreign keys.

Patient ID	Forename	Surname	Age	Address line 1	Address line 2	Town
1234	John	Freeman	43	123 Main Road	Townsend	Heighton
5678	Anna	Kline	55	89a Orchard Lane	Old Oak	Heighton

Test result ID	Patient ID	Test	Result
000053	1234	Pulse	75
000054	1234	BP	105/76
000055	1234	Respiration	45
000056	5678	Pulse	81
000057	5678	Respiration	50



This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

SQL

- SQL stands for Structured Query Language.
- Is used to manipulate data stored in relational databases via Relational Database Management Systems (RDBMSs).
- Became a standard of the American National Standards Institute (ANSI) in 1986, and of the International Organization for Standardization (ISO) in 1987.

Typical operations

- Create / insert, read (select), update, delete (CRUD).

Create

```
CREATE DATABASE MyDatabase;
```

```
CREATE INDEX IndexName
ON TableName(col1);
```

```
CREATE TABLE OurTable (
            id int,
            name varchar(12)
);
```

Delete

```
DROP DATABASE OurDatabase;
```

```
DROP TABLE OurTable;
```

Update table

```
UPDATE OurTable
SET col1 = 56
WHERE col2 = 'something';
```

Delete records

```
DELETE FROM OurTable
WHERE col1 = 'something';
```

Add/remove column

```
ALTER TABLE OurTable
ADD col5 int;
```

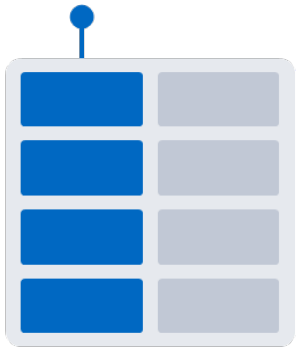
```
ALTER TABLE OurTable
DROP COLUMN col5;
```

```
SELECT
            col1,
            AVG(col2) * AVG(col3)
AS total
FROM tbl
GROUP BY col1
HAVING total > 2
```

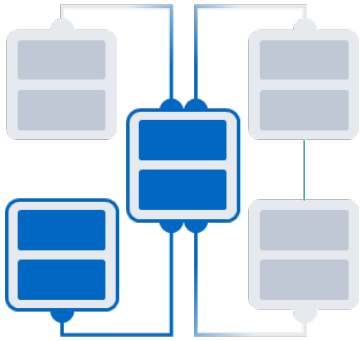

This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

NoSQL Databases

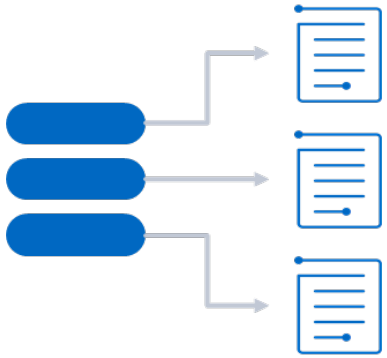
- Were designed to deal with large volumes of data in structured, semi-structured and unstructured format.
- Performance scales horizontally or “out”, enabling it to adapt to big data application requirements.
- Are non-relational and schema-free and requires programming skills to use.
- Typically, don’t enforce ACID properties, but BASE properties (eventually consistent).



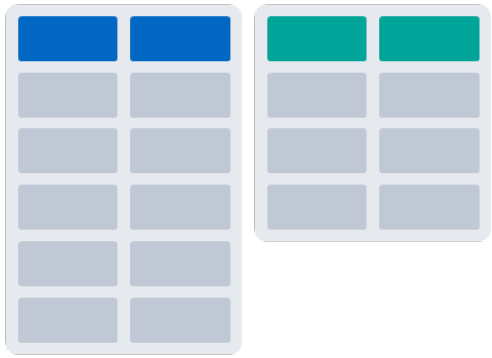
Key-value



Graph



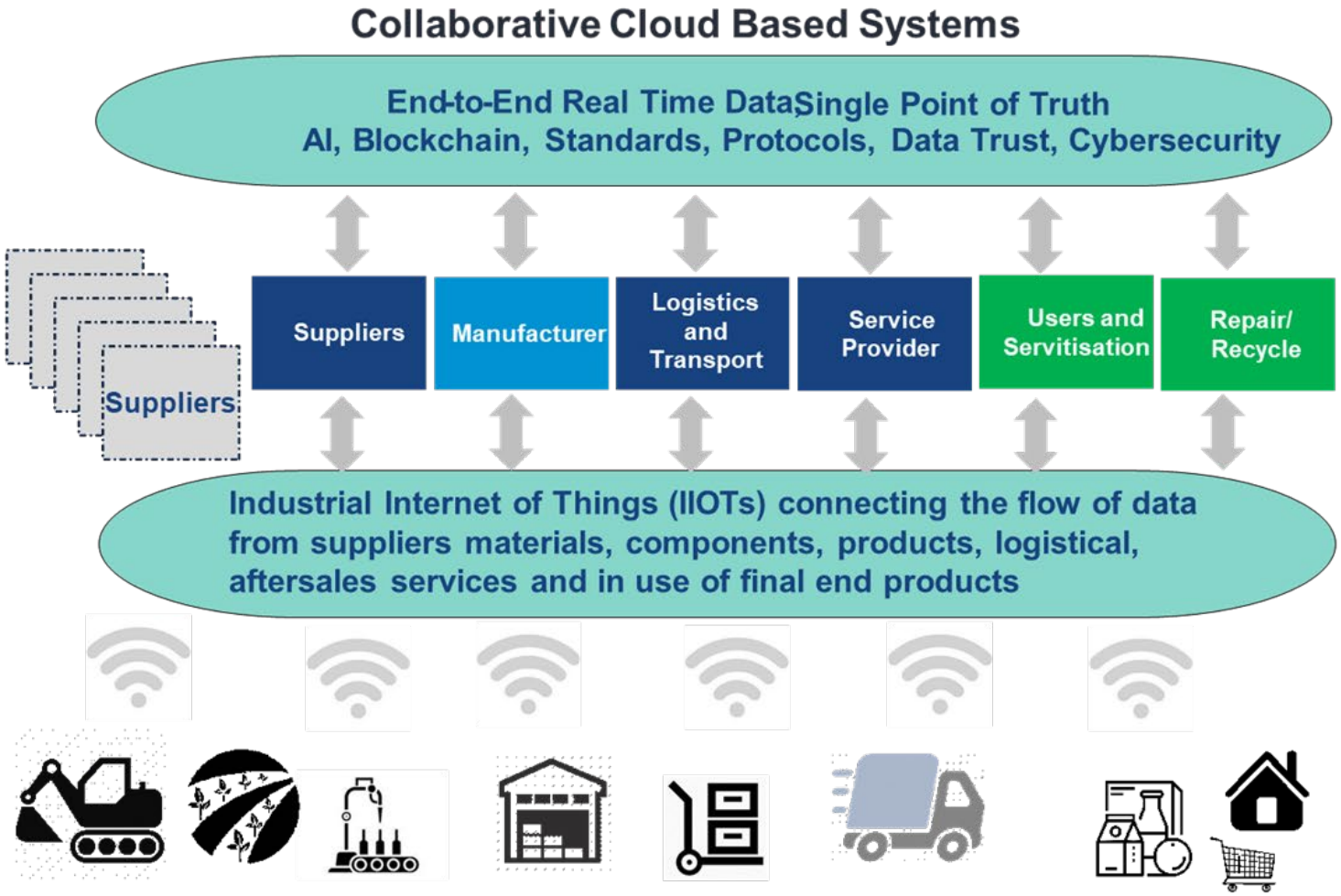
Document



Custom store

This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Solution to Consolidation Problem?



Technological features:

- Automation and robotics/ sensors
- Internet of Things (IoT)
- Cloud-based systems
- Artificial Intelligence
- Blockchain and DLT
- Cyber security
- Servitisation
- New business models

This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Reflection and Discussion

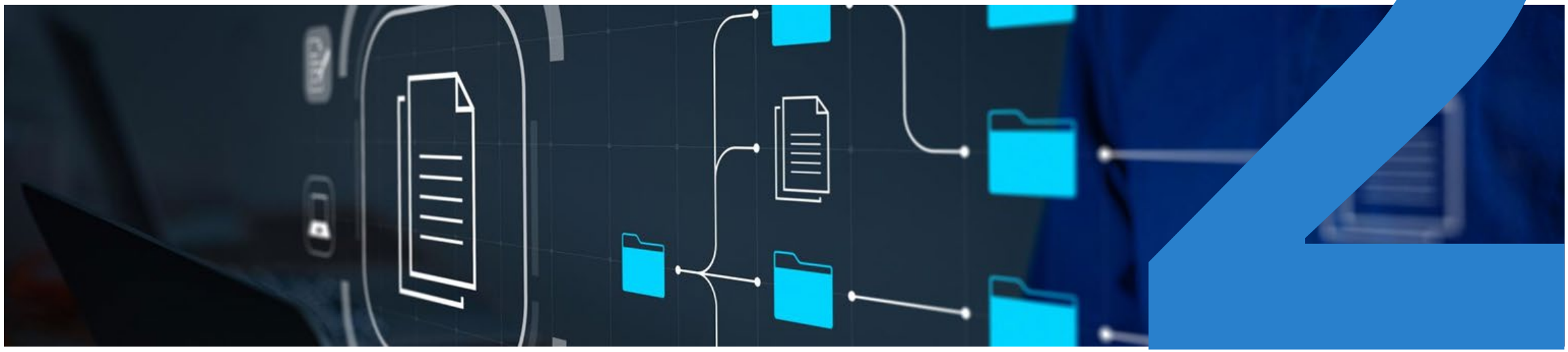
Do you have challenges that could be solved by Data Science in your workplace?

Make a note of the following point:

- Idea(s).
- Advantages of implementing your ideas (e.g., increased yields, increase resource efficiency,...).
- Concepts/ideas from the previous section that would help you.
- Potential barriers to implementing the ideas.

OGL National Composite Centre 2022
This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Data Management in Manufacturing



OGL National Composite Centre 2022
This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Data Management Plan For Industry 4.0

Data architecture
Data architecture is constructed by a data architect. The deliverable is a data management plan for an organisation.

Data models
Data models define how to link datasets within your data storage.

Databases
A management system to store and retrieve data efficiently.

Data warehouses
A centralised storage system for datasets from multiple sources. Can only store structured data.

Data lakes
A centralised repository for datasets from any source, where data can be both structured and unstructured.

Data governance
Data governance defines policies to ensure data is managed according to best practice. Is the process outlined by a business to police their data.

source: DAMA.Org

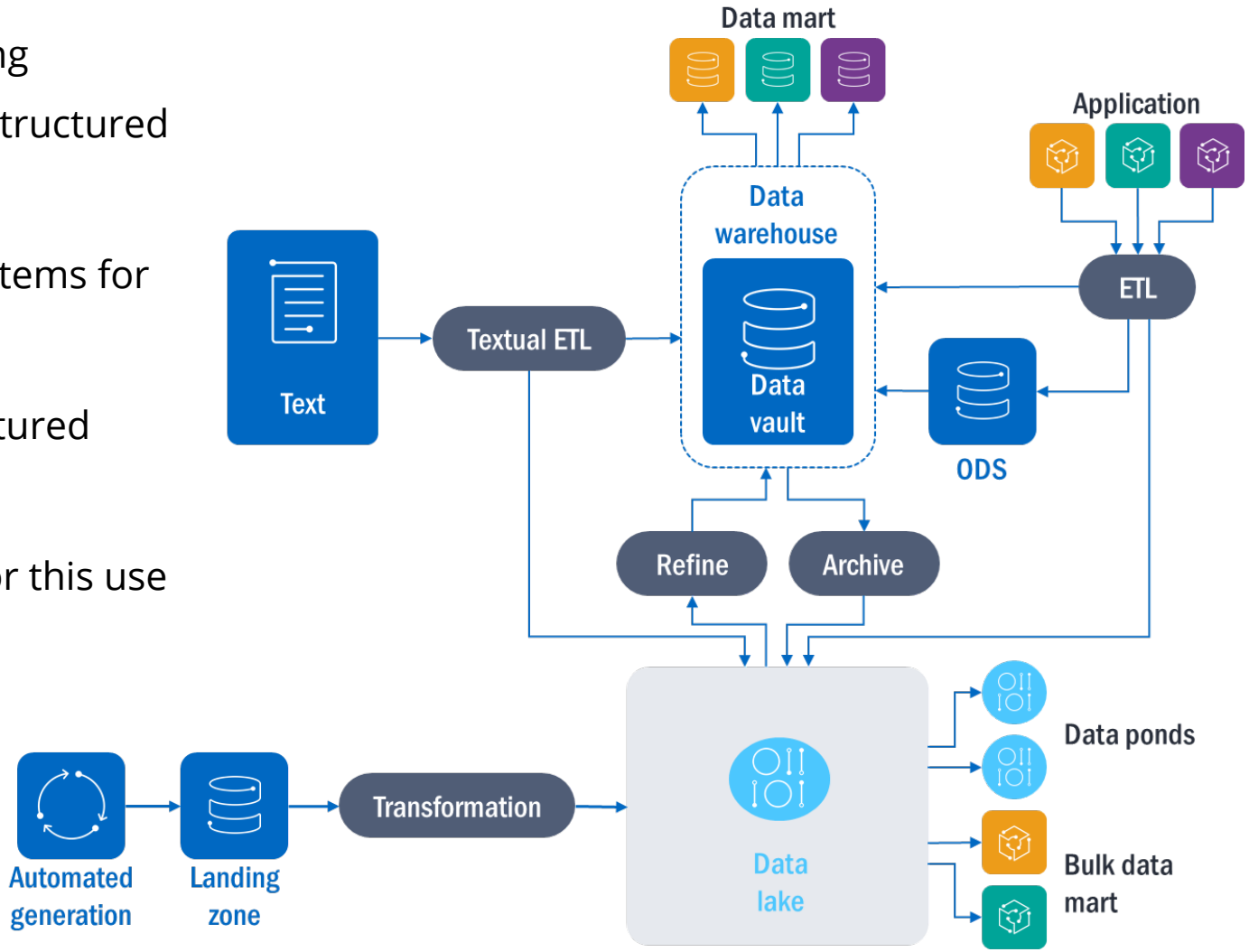
OGL National Composite Centre 2022
This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Data Architecture

Data architecture for end-to-end manufacturing problems will require both structured and unstructured processing capabilities.

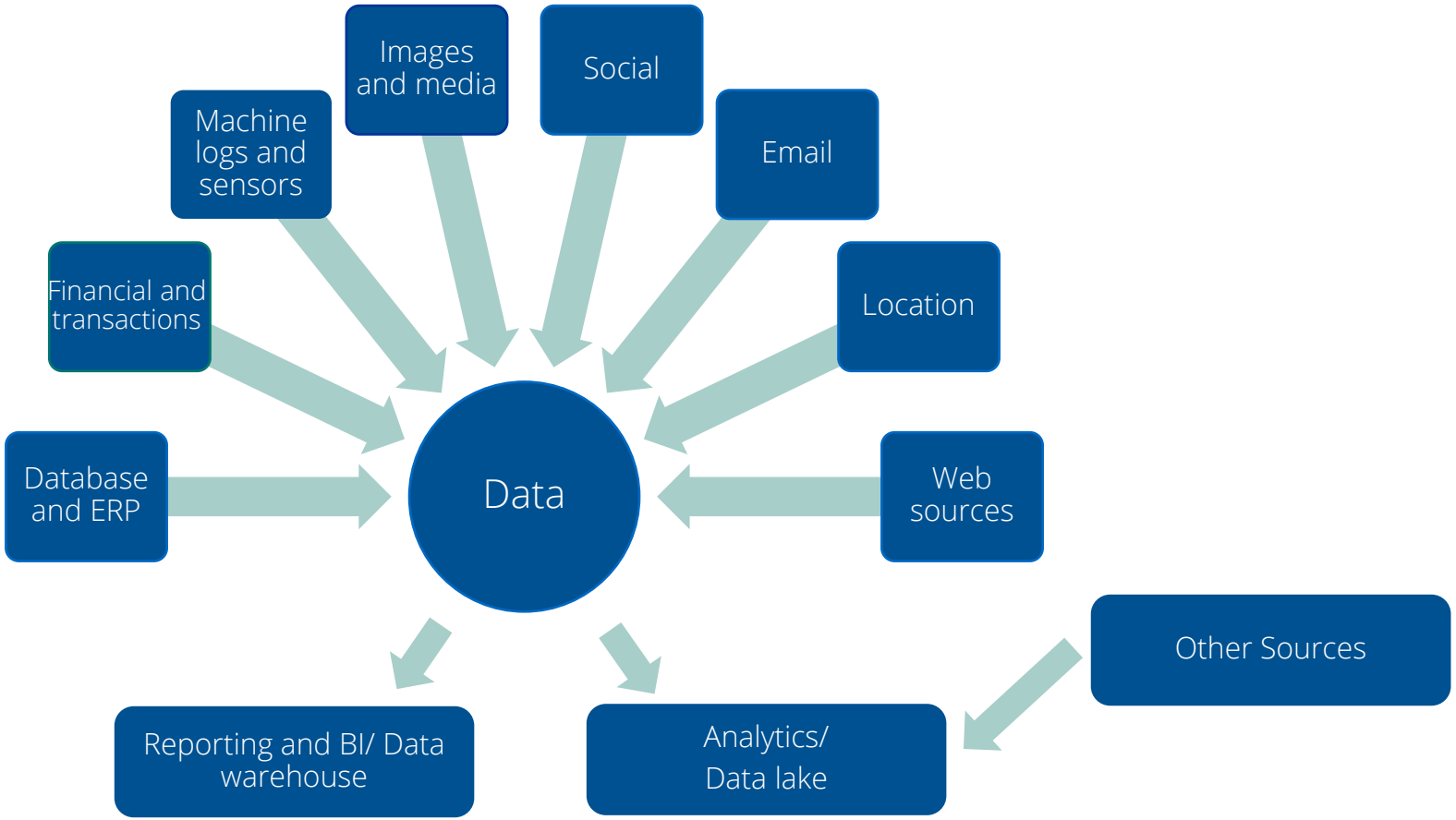
- Data warehouses are tried and tested systems for hosting structured data.
- Data lakes can store and process unstructured data.
- It makes sense to use both approaches for this use case.

There is no one right architecture, depending on resources available and use cases one can implement different approaches, if best practices are followed.



This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

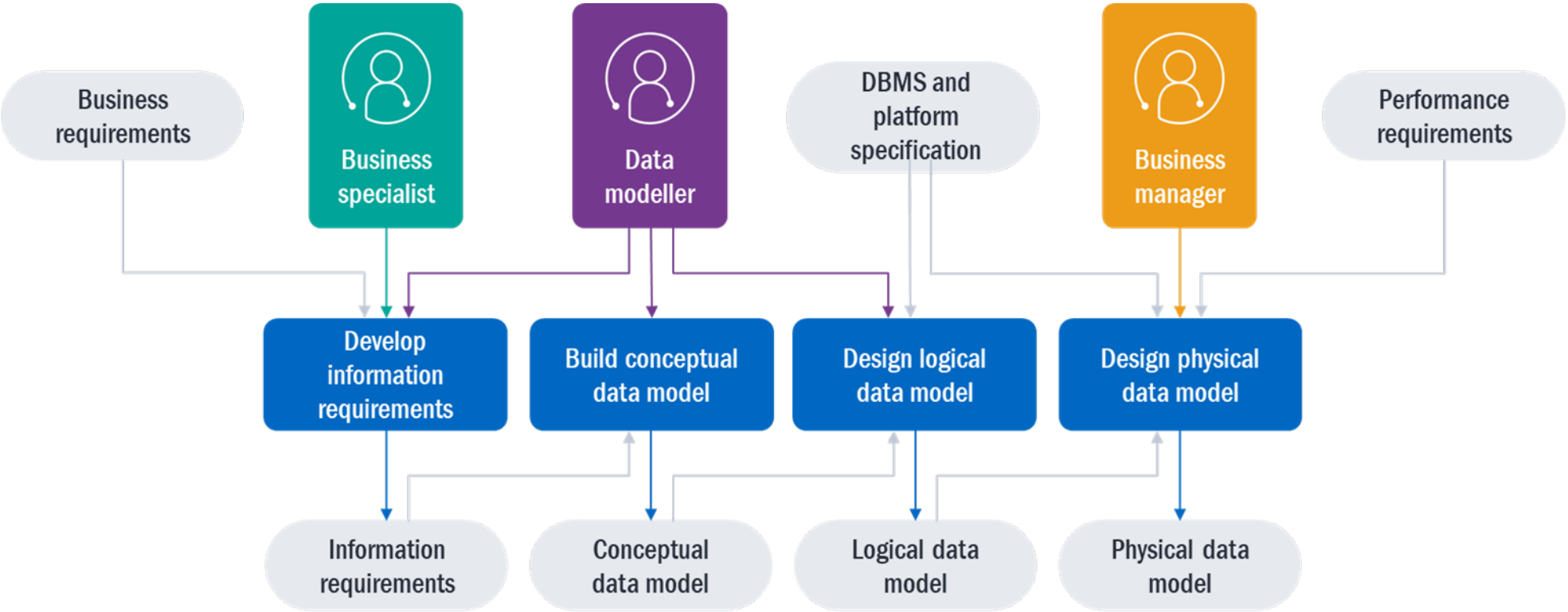
Data Warehouse vs Data Lake – Which One?



This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Data Modelling

- Data models help us to think about the problem that we are solving.
- One of the most important components of building a data management system, often overlooked.
- We will next provide a few examples of outputs from the process.



This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Conceptual Model – Supply Chain

Consists of 3 main activities:

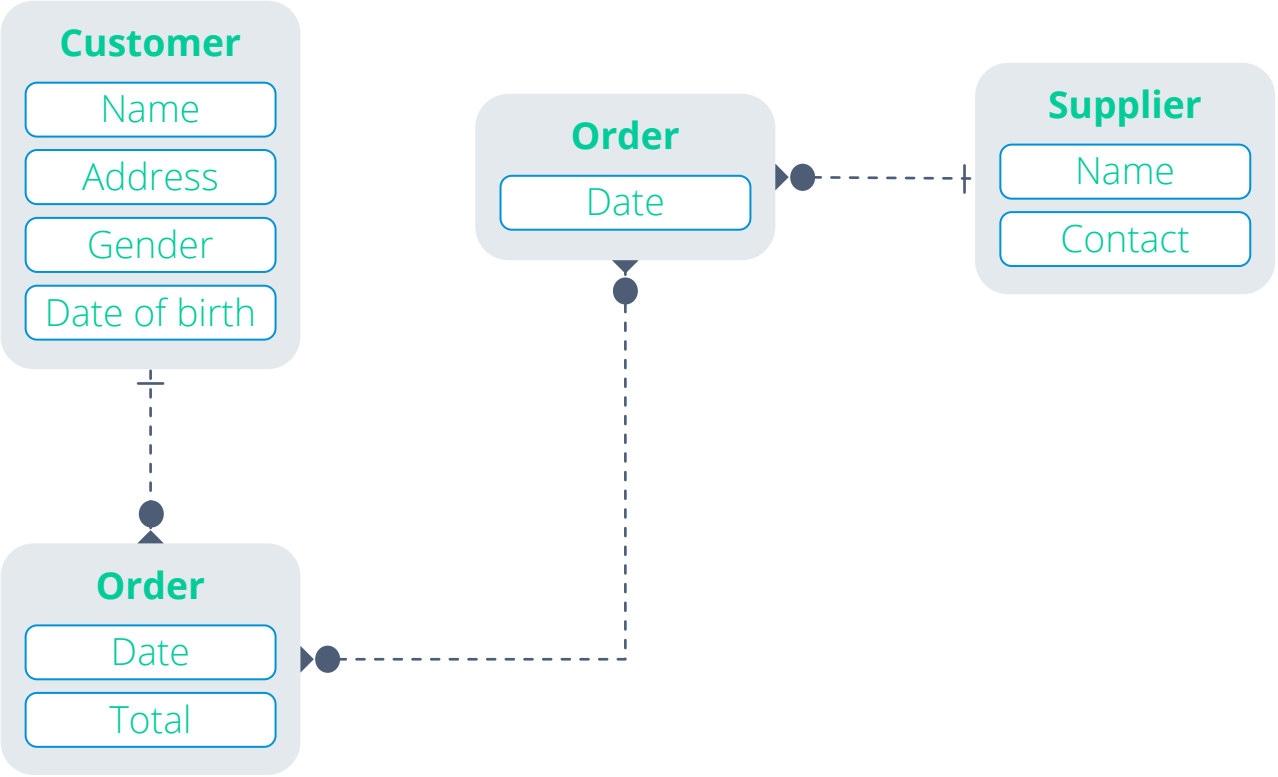
- Elicitation of requirements.
- Design of solution.
- Evaluation of solution.

- Define entities and relationships.
- Define hierarchies.
- DAMA has many resources for. further research: www.dama.org

This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Logical Model – Supply Chain

- Decide on type of database - not specific DBMS.
- Typically logical model is for relational database, but can be adapted for others.
- CASE tools can generate logical from conceptual.
- Some typical steps for this stage are:
- Implement relational relationships between tables based on conceptual model.
- Implement attributes.
- Specify primary and foreign key.



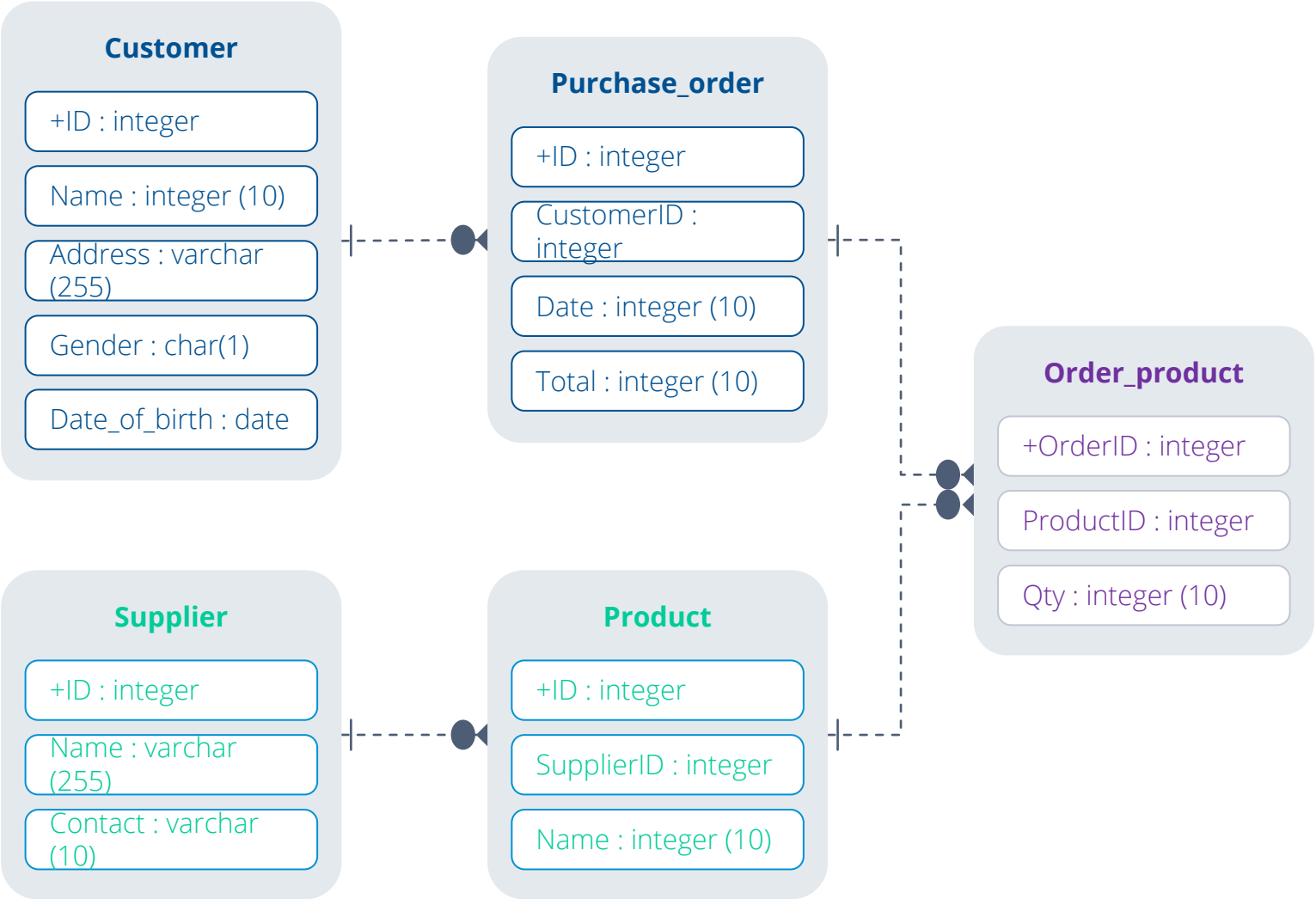
This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Physical Model – Supply Chain

Make design decisions for specific DBMS e.g. Apache Cassandra or PostgreSQL.

Design

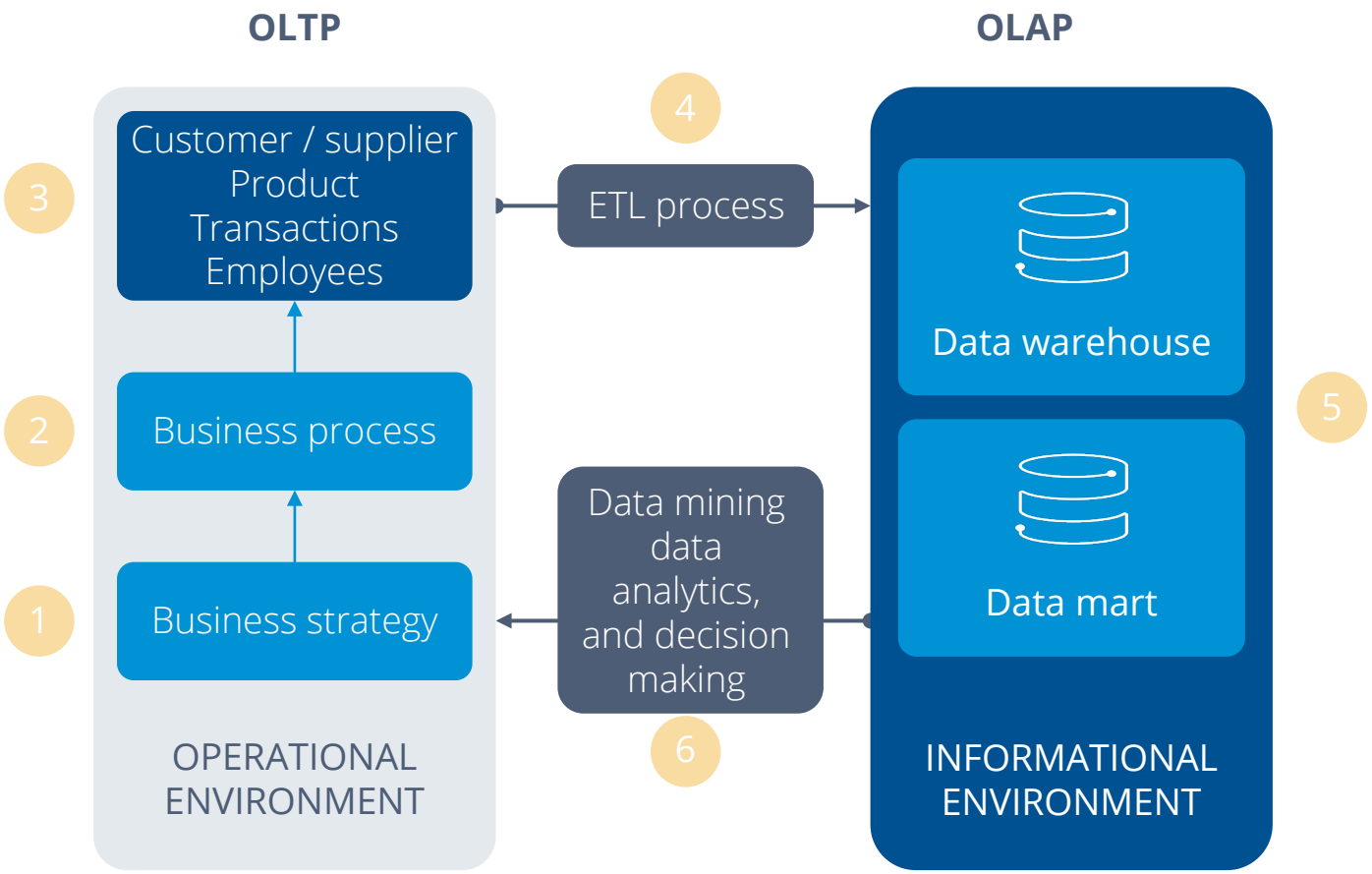
- Tables, columns, and views.
- Indexes.
- Physical storage configuration.
- Data types.



This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

OLTP Versus OLAP (Relational Data)

- Online transactional processing (OLTP) is an online transaction system.
- Online analytic processing (OLAP) is an online data retrieval and analysis system.
- OLTP's main operations are insert, update and delete (faster than OLAP).
- OLAP's main operation is to extract multidimensional data for analysis (faster than OLTP).
- The tables in OLTP database must be normalized (3NF).
- Tables in OLAP database may not be normalized.



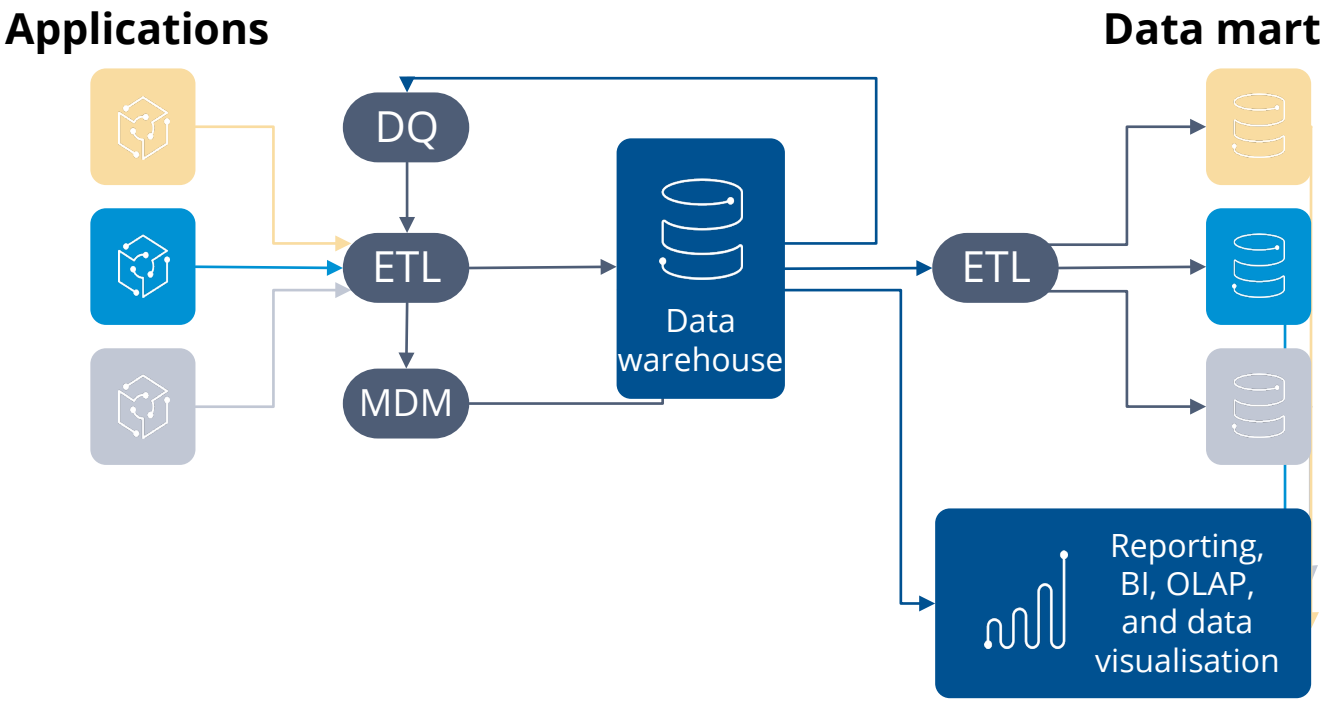
This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Data Warehouse

- Around since 1990s, intention to store all the data about an enterprise and make it available for analytics.
- *"A copy of transaction data specifically structured for query and analysis"* – Kimball.

Many technologies derive from data warehouse implementations including:

- Extract, transform, load (**ETL**).
- Data quality (**DQ**) and profiling tools.
- Data modelling tools.
- Business glossaries.
- Metadata repositories.
- Data governance tools, data lineage.
- Master data management (**MDM**) systems.
- Enterprise information integration (EII), data federation, and data virtualization tools.
- Can only store structured data.



OGL National Composite Centre 2022
This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

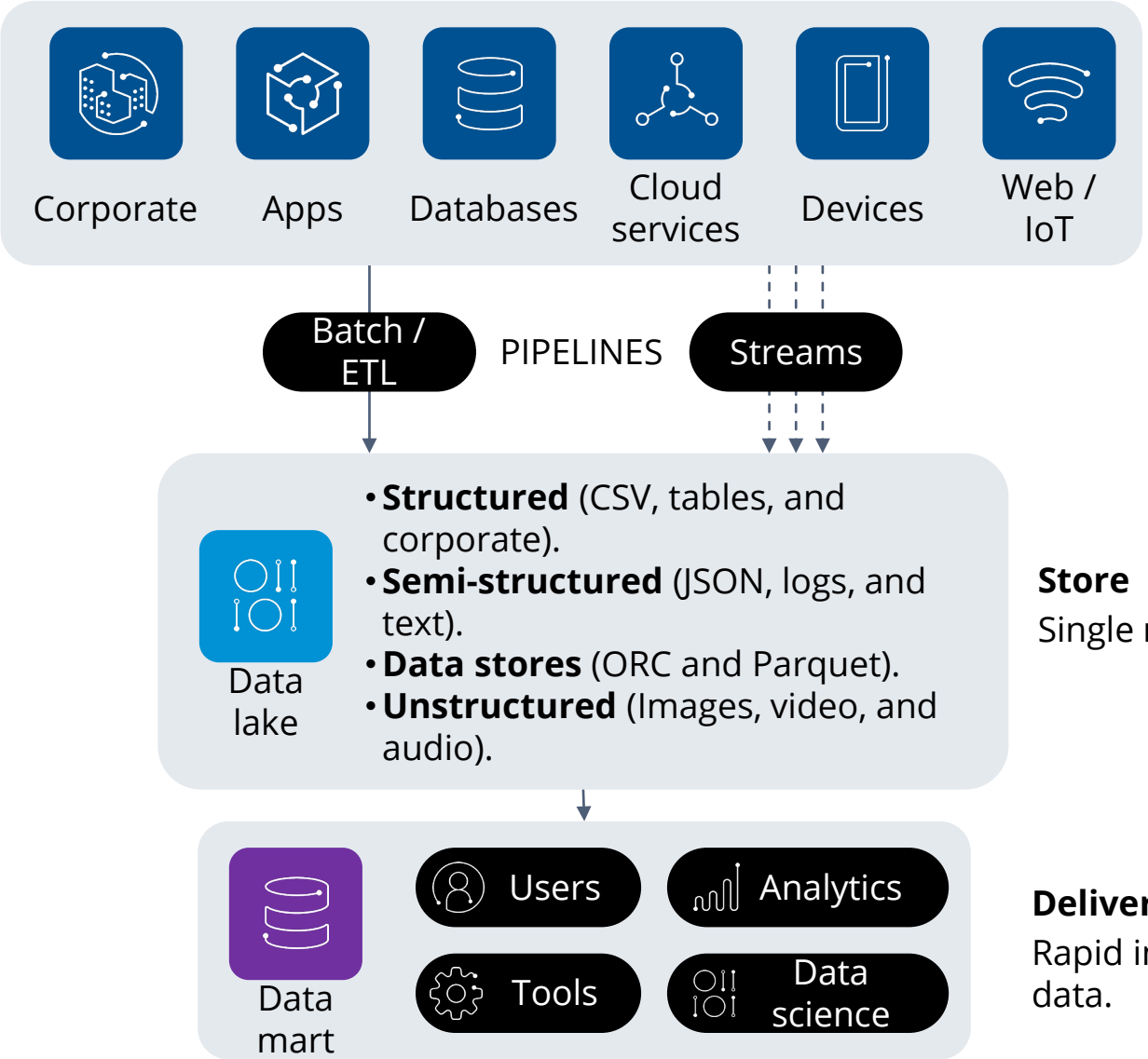
Data Lake

A data lake is a set of centralised repositories containing vast amounts of raw data (either structured or unstructured), described by metadata, organized into identifiable datasets, and available on demand.

A data lake is a collection of data such that:

- The data have no fixed schema and was consolidated from various sources.
- All data formats should be possible (structured and unstructured).
- The data have not been transformed i.e., is raw.
- The data are conceptually present in one single place, but can be physically distributed.
- The data are used by one or multiple experts in data science.
- The data must be associated with a metadata catalogue.
- The data must be associated with rules and methods for their governance, data quality, data security and data lifecycle must be maintained.

OGL National Composite Centre 2022
This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0



Capture
Batch or streaming data from many sources.

Store
Single repository for all formats of data.

Deliver
Rapid insights and business results from all data.

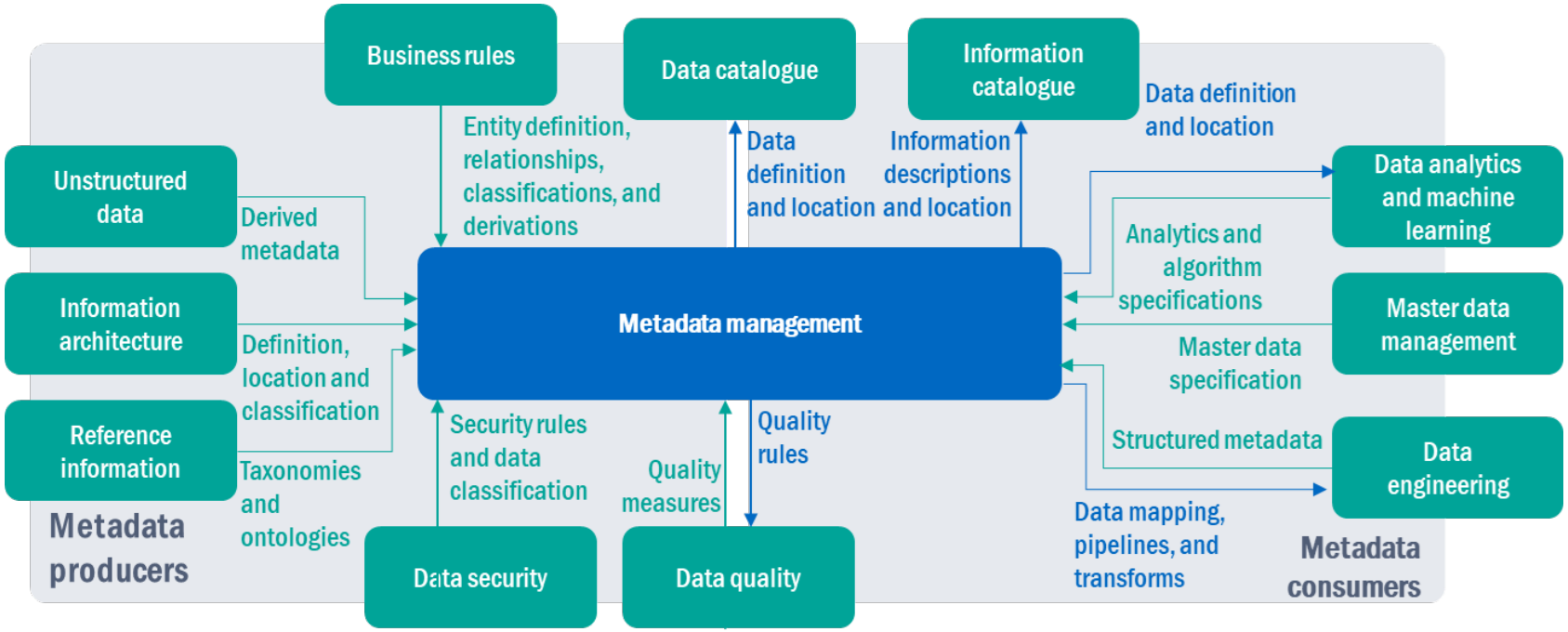
This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

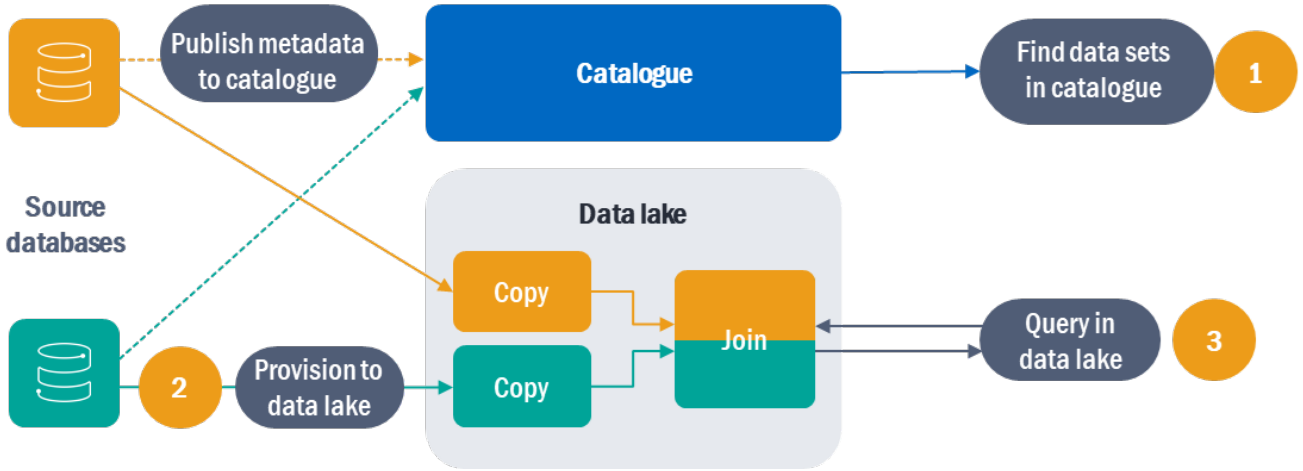
Metadata Management

Metadata is key to the data lake architecture

- The metadata catalogue contains all the information a data lake needs to reconstitute information from the raw data stored in the lake.
- There are many different types of metadata required to reconstitute the information. The image to the right lists the different metadata categories typically stored.
- The catalogue is typically built when data is loaded into the data lake.
- The data is only transformed when queried, hence why the process is called Extract/ Load/ Transform or ELT as opposed to Extract/ Transform/ Load or ETL which data warehouses use.

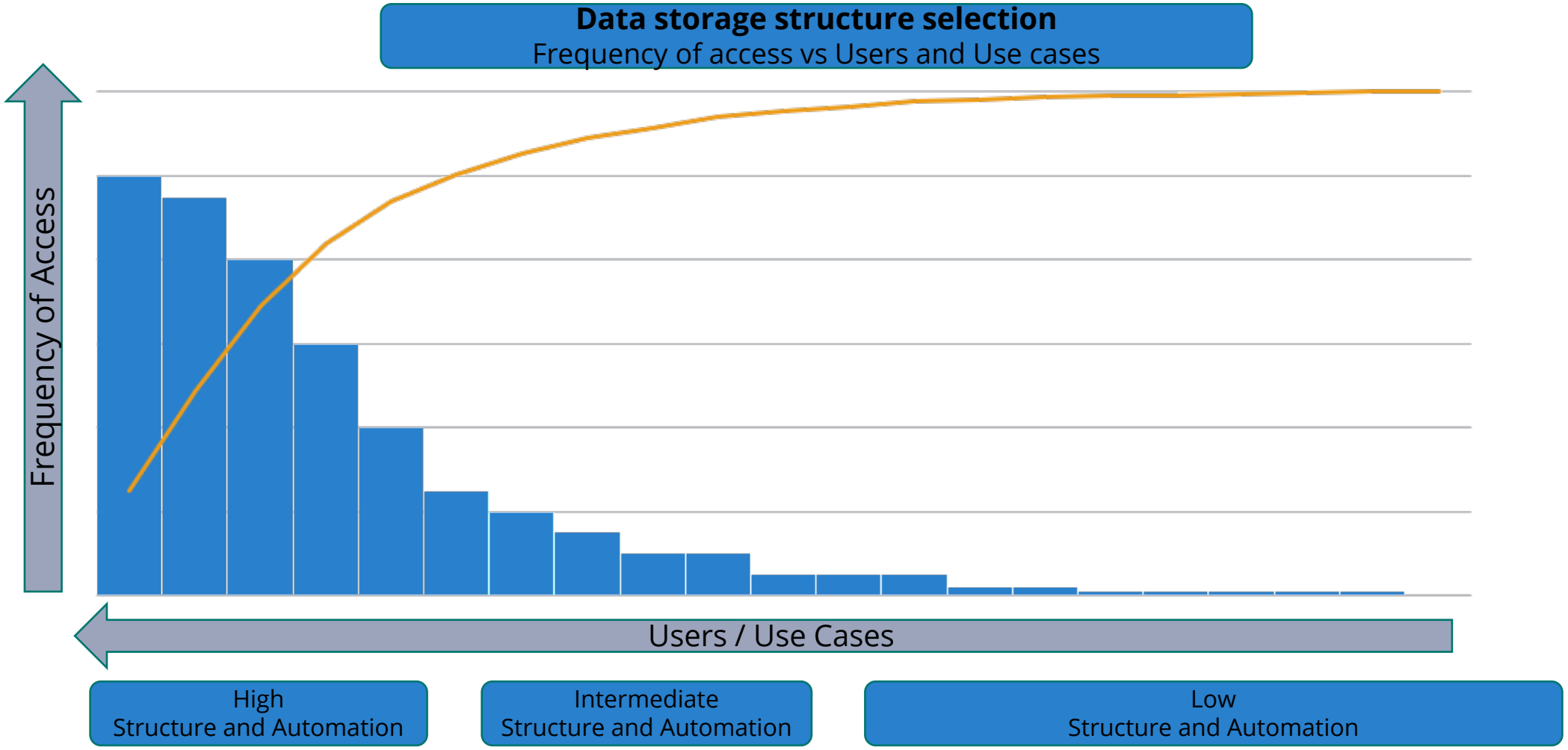
This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0





This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Spreadsheet, Database, Data Warehouse or Data Lake?



This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Data Management Platform Selection

- Understand your use case e.g.
  - Small departmental license tracking system.
  - Organisation wide reporting requirement.
- Understand your pain points e.g.
  - Spreadsheets and spread marts.
  - Lack of data integration e.g., manufacturing supply chain nodes.
- Understand different analytic requirements:
  - Operational reporting/ data extracts, ad-hoc analyses, advanced analytics (predictive modelling, ML, DL), BI dashboarding.

	Database	Data Warehouse	Data Lake
Data	Structured	Structured	Raw and Unstructured
Processing	Schema on write	Schema on write	Schema on read
Cost	Free or $	$$$	$
Agility	Varies	Minimal	Maximum
Security	Immature	Mature	Immature
Users	Anyone	IT/ business users	Data scientists
Use cases	Reporting, analysis and automation	Machine learning	Data science and research

This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Reflection and Discussion

Do you have challenges that could be solved by Data Science in your workplace?

Make a note of the following point:

- Idea(s).
- Advantages of implementing your ideas (e.g., increased yields, increase resource efficiency,...).
- Concepts/ideas from the previous section that would help you.
- Potential barriers to implementing the ideas.

This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Data Preparation

This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

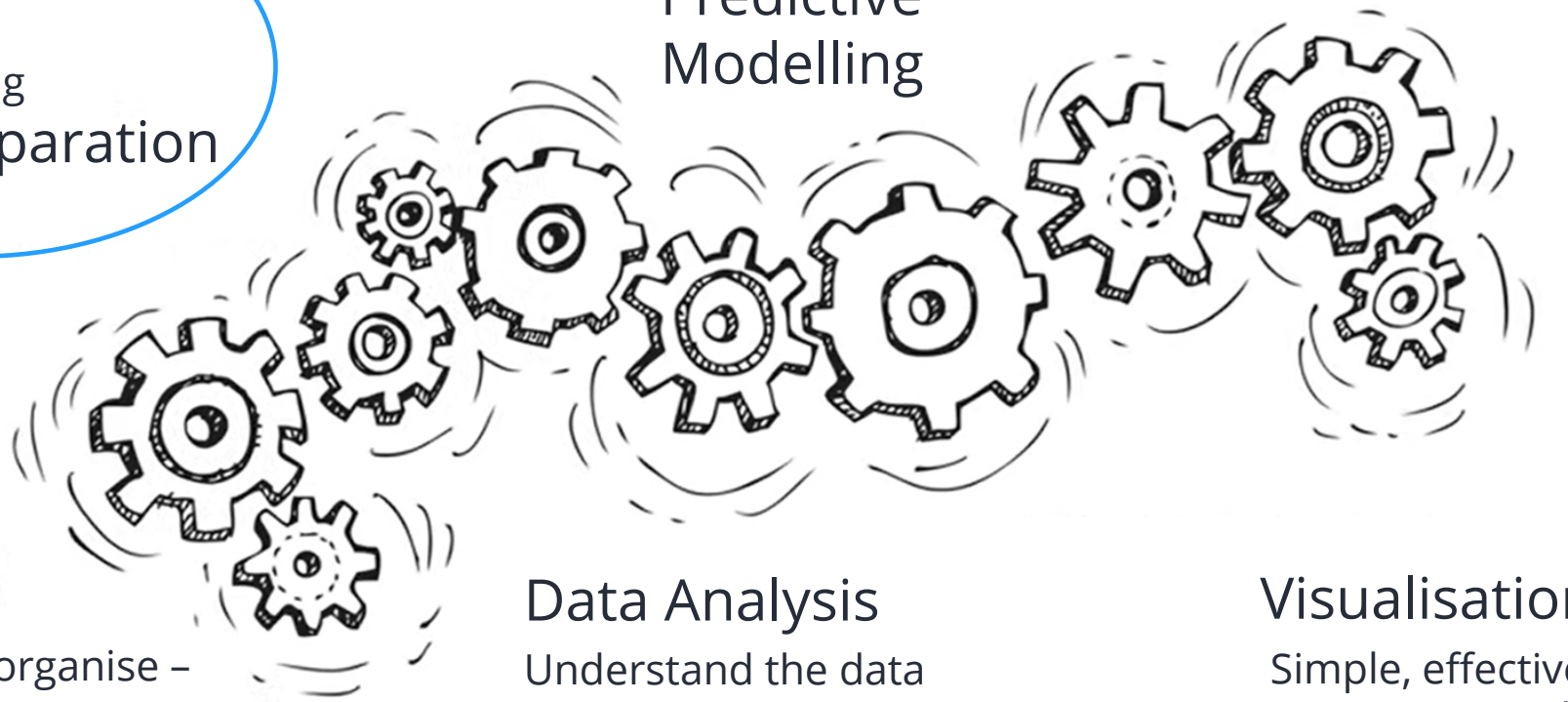
Recall the Typical Analytic Workflow From Unit 1?

We now focus on data preparation

Cleaning
Formatting
Transforming
Data Preparation

Iterative application of algorithms to identify best model fitting the problem.
Validate predictions.

Predictive Modelling



Data collection

Find relevant data and organise – often via automation / programming.

Data Analysis

Understand the data
Variation, bias, outliers.

Visualisation

Simple, effective manner to communicate the results.

This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Considerations Before Using a New Dataset

- Data governance questions:
  - Who is the data owner?
  - Are there data governance policies in place for the dataset?
  - Any relevant regulation and legislation that should be considered?
- Data documentation:
  - Is there metadata or a data dictionary available?
  - Is there a change management process in place for the data?
  - What was the context of collection for the data?
- Understanding the data:
  - Complete and accurate (independence, distribution, bias).
  - Data distribution – need for standardisation?
  - Dealing with errors and outliers.
  - Data quality.

KDNuggets: Don't touch a dataset without asking these 10 questions

This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Data Used For Machine Learning

Most machine learning algorithms require data to be numbers in one of the following formats:

Continuous: These are data elements that are defined on an interval that can be limited or unlimited. Examples include flow rates, tank levels or pressures.

Categorical:

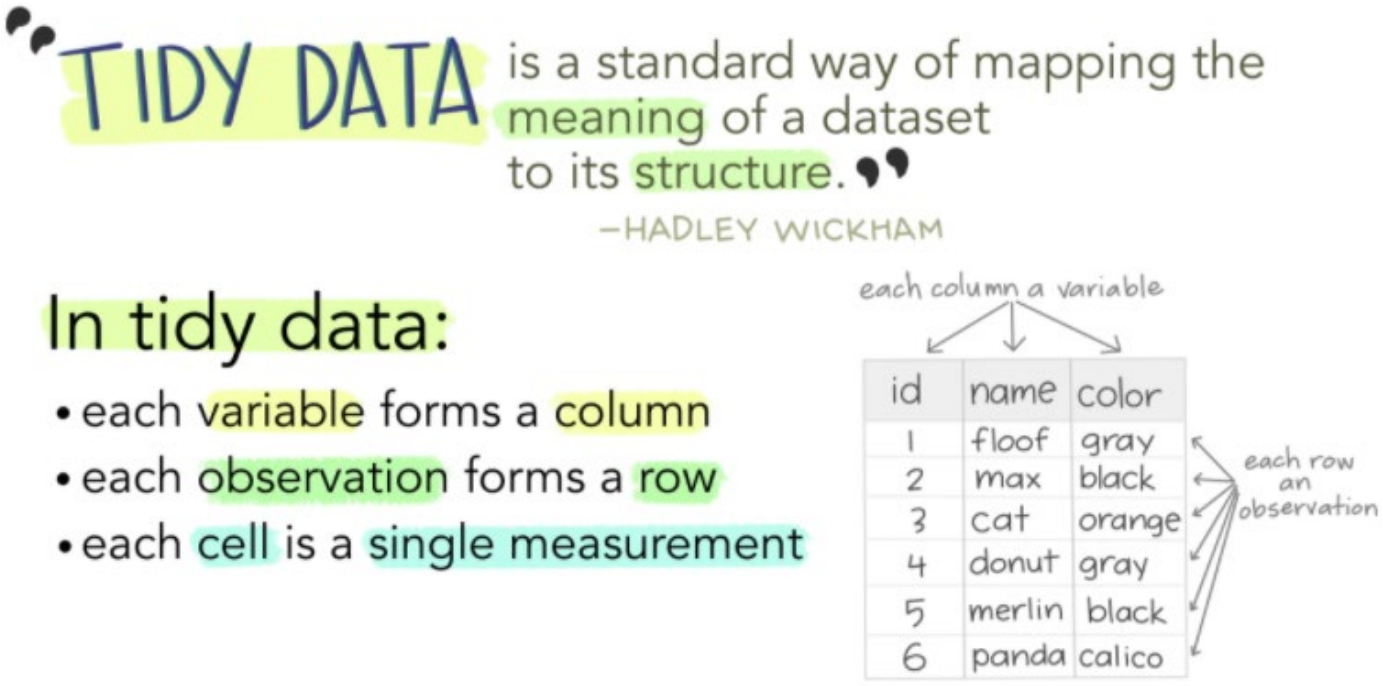
- **Nominal**: These are data elements that can only take on a limited set of values with no meaningful ordering in between. Examples include type of product, purpose of loan.
- **Ordinal**: These are data elements that can only take on a limited set of values with a meaningful ordering in between. Examples include tensile strength rating, age coded as young, middle aged, and old.
- **Binary**: These are data elements that can only take on two values. Examples include gender, true or false.

Unstructured data therefore typically requires pre-processing in order to be useful for ML and AI.

OGL National Composite Centre 2022
This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Data Preparation – Why?

- Tidy messy data.
- Machine learning algorithms require data to be numbers.
- Some machine learning algorithms impose requirements on the data.
- Statistical noise and errors in the data may need to be corrected.
- Complex nonlinear relationships may be teased out of the data.



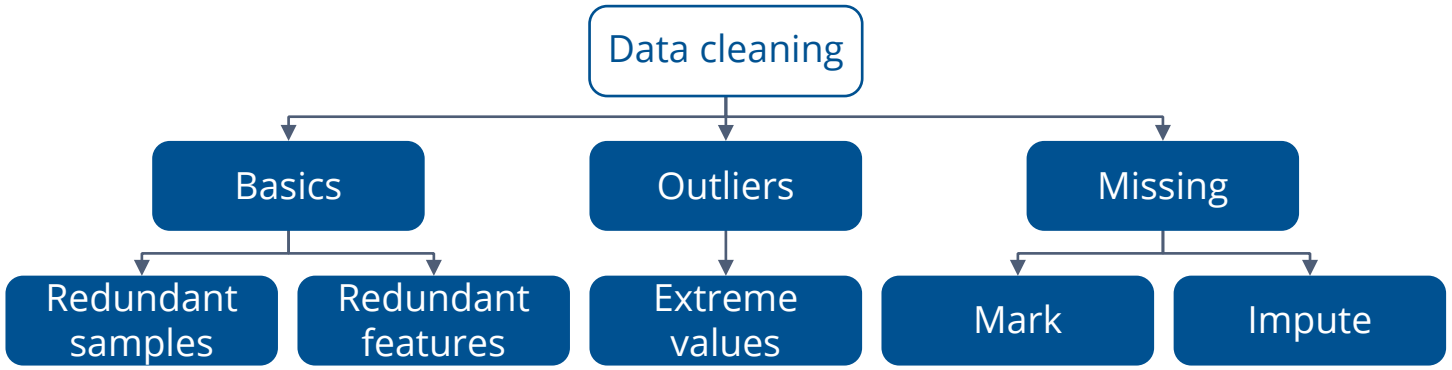
Wickham, H. (2014). Tidy Data. Journal of Statistical Software 59 (10). DOI: 10.18637/jss.v059.i10

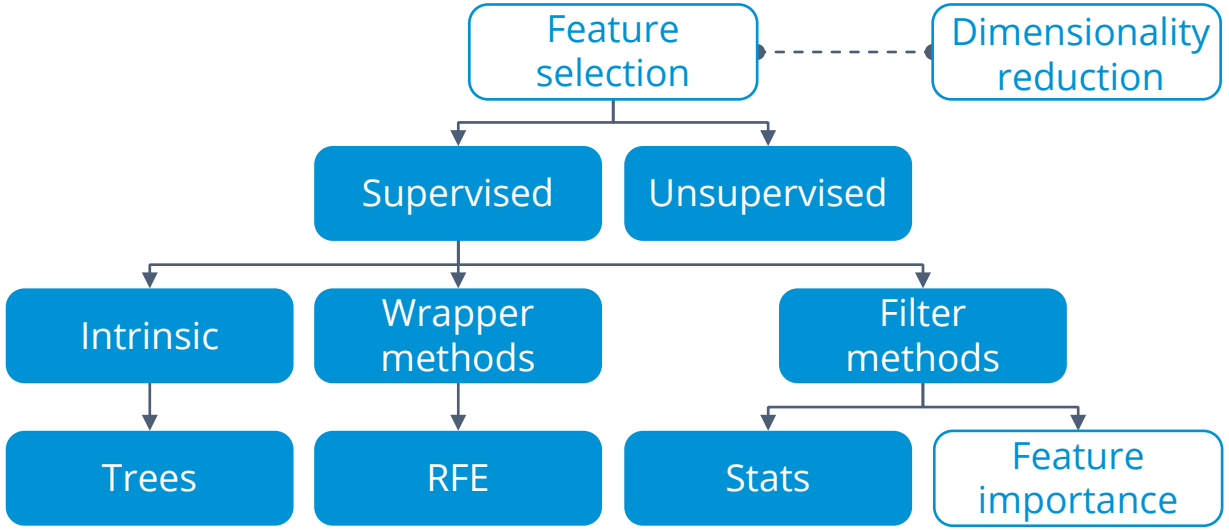
Tidy Data

This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Data Preparation – How?

- **Data cleaning**
  - Identifying and correcting mistakes or errors in the data.
- **Feature selection**
  - Identifying those input variables that are most relevant to the task.
- **Data transforms**
  - Changing the scale or distribution of variables.
- **Feature engineering**
  - Deriving new variables from available data.
- **Dimensionality reduction**
  - Creating compact projections of the data.





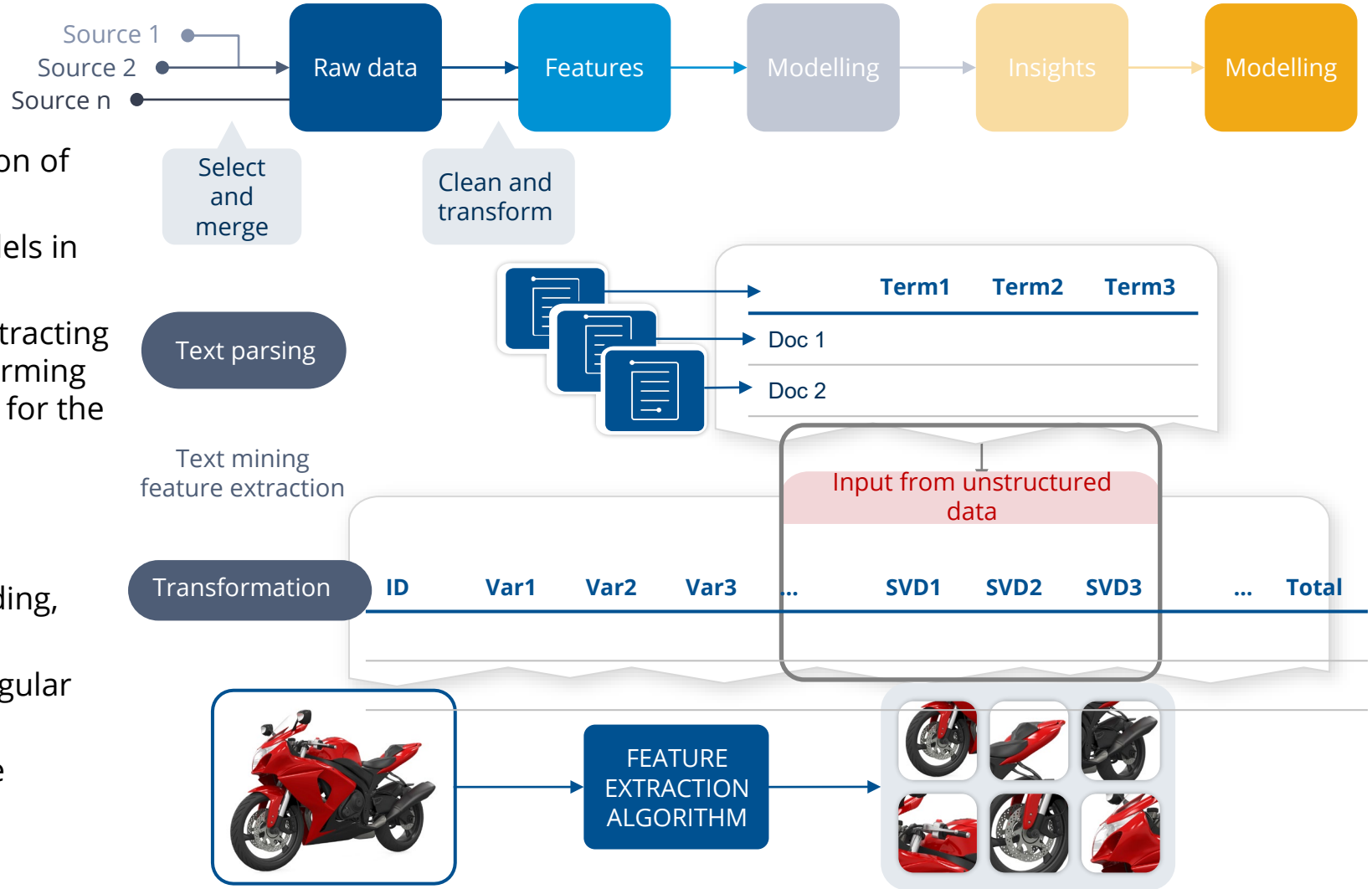
This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Feature Extraction

- A feature is a numeric representation of an aspect of raw data.
- Features sit between data and models in the machine learning pipeline.
- Feature engineering is the act of extracting features from raw data and transforming them into formats that are suitable for the machine learning model.

Feature extraction is different for different types of data

- Structured data (e.g. one-hot-encoding, principal component analysis).
- Text (e.g. parsing, bag of words, singular value decomposition).
- Images (e.g. neural network feature extraction).



This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Reflection and Discussion

Do you have challenges that could be solved by Data Science in your workplace?

Make a note of the following point:

- Idea(s).
- Advantages of implementing your ideas (e.g., increased yields, increase resource efficiency,...).
- Concepts/ideas from the previous section that would help you.
- Potential barriers to implementing the ideas.

This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Discussion

This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Data Preparation Discussion

List a few data sources in your workplace:

- List possible use cases?
- How critical is the integrity of the listed data sources?
- What tools do you currently use to manage and access the data?
- What risks do you need to communicate to your stakeholders?
- Can you think of examples where data and data definitions may have different meanings to different stakeholders?

This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Q&A

This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

References

1. Bajic, B., Rikalovic, A., Suzic, N., & Piuri, V. (2021). Industry 4.0 Implementation Challenges and Opportunities: A Managerial Perspective. *IEEE Systems Journal*, *15*(1), 546–559.
2. Kleppmann, M. (2017). Designing Data-Intensive Applications: The Big Ideas behind Reliable, Scalable, and Maintainable Systems. In *O'Reilly Media, Inc.* O'Reilly.
3. Gorelik, A. (2019). *The Enterprise Big Data Lake*. O'Reilly.
4. Inmon, W. H., Linstedt, D. (2015). *Data Architecture: A Primer for the Data Scientist -- Big Data, Data Warehouse and Data Vault*. Morgan Kaufmann.
5. Simsion, G. C., & Witt, G. C. (2005). *Data modeling essentials* (3rd ed.). Amsterdam; Boston.
6. Zheng, A., & Casari, A. (2018). *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists*. O'Reilly.
7. Torres, M. T., Hart, G., & Emery, T. (2019). *The Dstl Biscuit Book Artificial Intelligence, Data Science and (mostly) Machine Learning*. 1–40.

This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

What Next?

List of possible further courses

Course	Topic	Level
Become a Data Engineer – Udacity	Learn to design data models, build data warehouses and data lakes, automate data pipelines, and work with massive datasets. At the end of the program, you'll combine your new skills by completing a capstone project.	Intermediate
IBM Data Engineering Professional Certificate - Coursera	You will learn to use the Python programming language and Linux/UNIX shell scripts to extract, transform and load (ETL) data. You will work with Relational Databases (RDBMS) and query data using SQL statements. You will use NoSQL databases and unstructured data. You will be introduced to Big Data and work with Big Data engines like Hadoop and Spark. You will gain experience with creating Data Warehouses and utilize Business Intelligence tools to analyse and extract insights.	Beginner

This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0

Copyright Utilised

Reference	Copyright
The DSTL Biscuit Book	http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3
Tidy Data	https://creativecommons.org/licenses/by/4.0/

OGL National Composite Centre 2022
This IDT403 Introduction to Data Analysis and Machine Learning content is licensed under the Open Government Licence 3.0