Data Management

Organize Data & bring value to the organization

Data Management

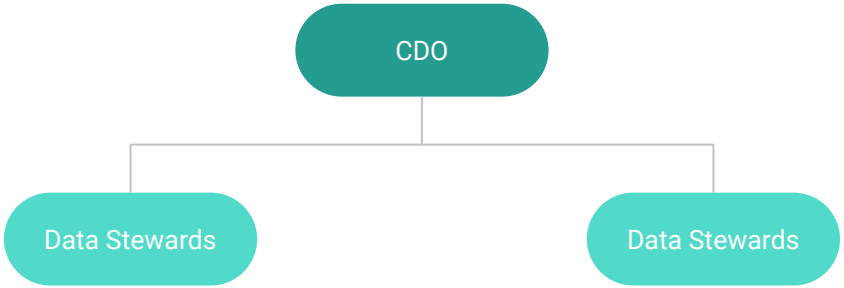
What is Data Management

Data Governance

Role

- Define **policies and procedures** according to the business needs and regulation (GDPR)

Organisational Chart



Activities of Data Governance

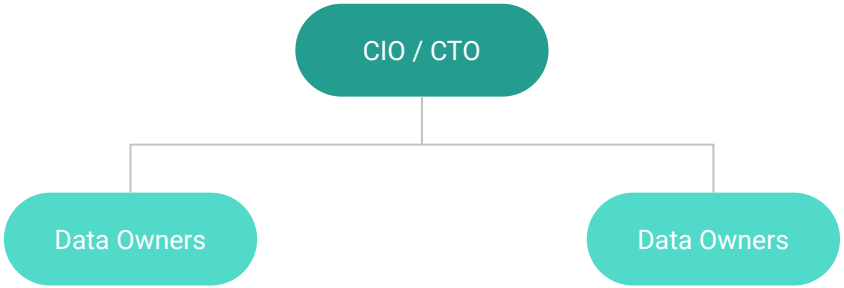
- Defining a standard of client data (currencies in €, temperature in celsius)
- Define regulations the organisation needs to follow (GDPR, ISO)
- Audit Data to check whether the standards are met

Data Management

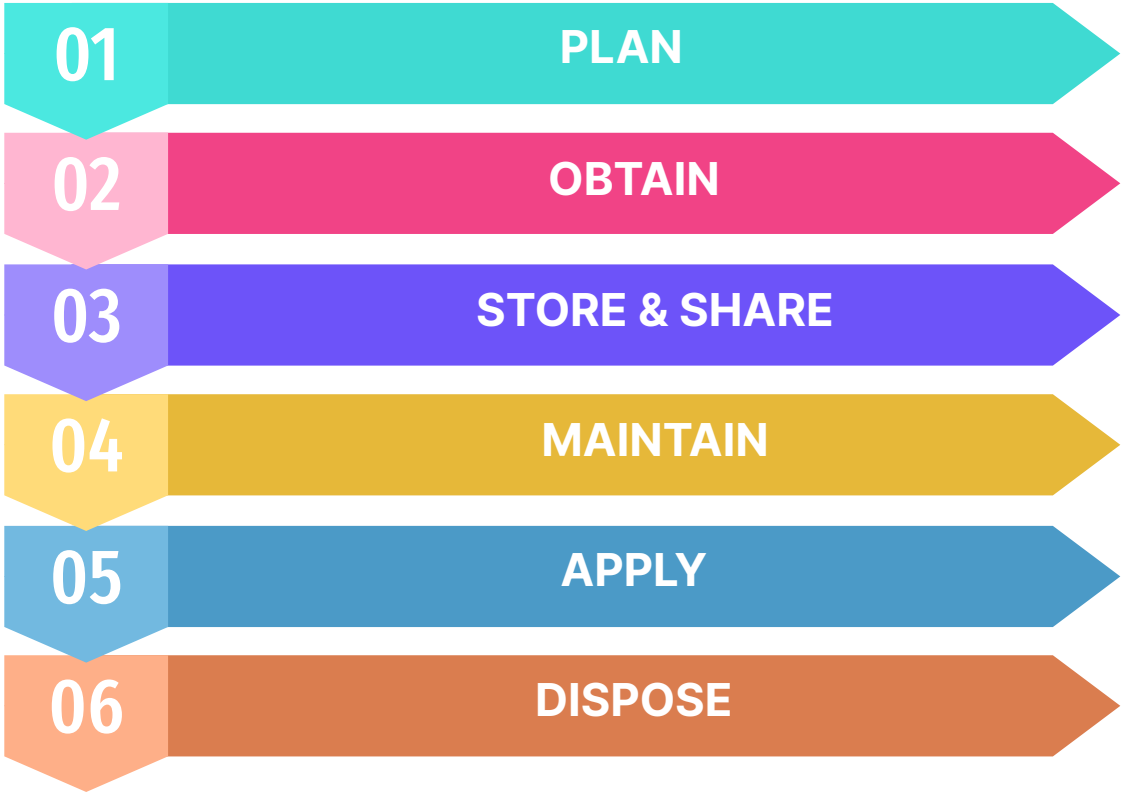
Role

- **Implement** policies and procedures from the Data Governance
- **Data Quality is an important** part of Data Management

Organisational Chart



Information lifecycle - POSMAD



Activities of Data Management

- **Data Profiling** - Assess the quality of the data
    - *Should be done across a specific dimension*
    - *i.e “Completeness” | “Accuracy” | “Consistency” | “Lineage”*
- **Data Remediation** - Cleaning the data
- Improve **Data Quality**

Focus on Data Quality

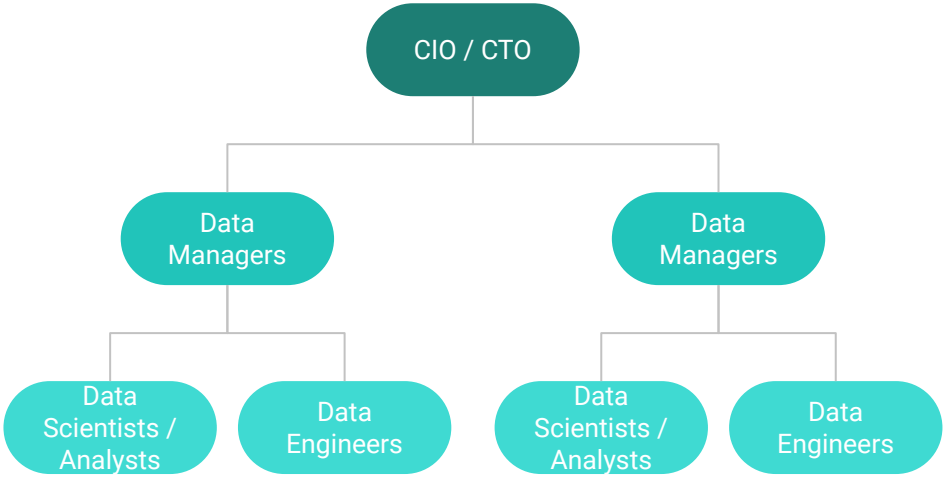
- **Transactional Data**
  - *Sales*
  - *Order*
  - *Receipts*
- **Master Data**
  - *Few selected customer, business data that needs to be perfect*
  - *There needs to **be one copy** of this data —> **One source of truth***
- **Reference Data**
  - *Any data that are used as reference (List of countries, Zip codes, Currency conversion table)*
- **MetaData**
  - *Data that explains the data*
    - *Business metadata (sensitivity of the data, allowed users | teams)*
    - *Technical metadata (String, Integer)*

Data Analytics

Role

- Leverage company data to build a **Data product** (Dashboard, ML Algorithm)

Organisational Chart

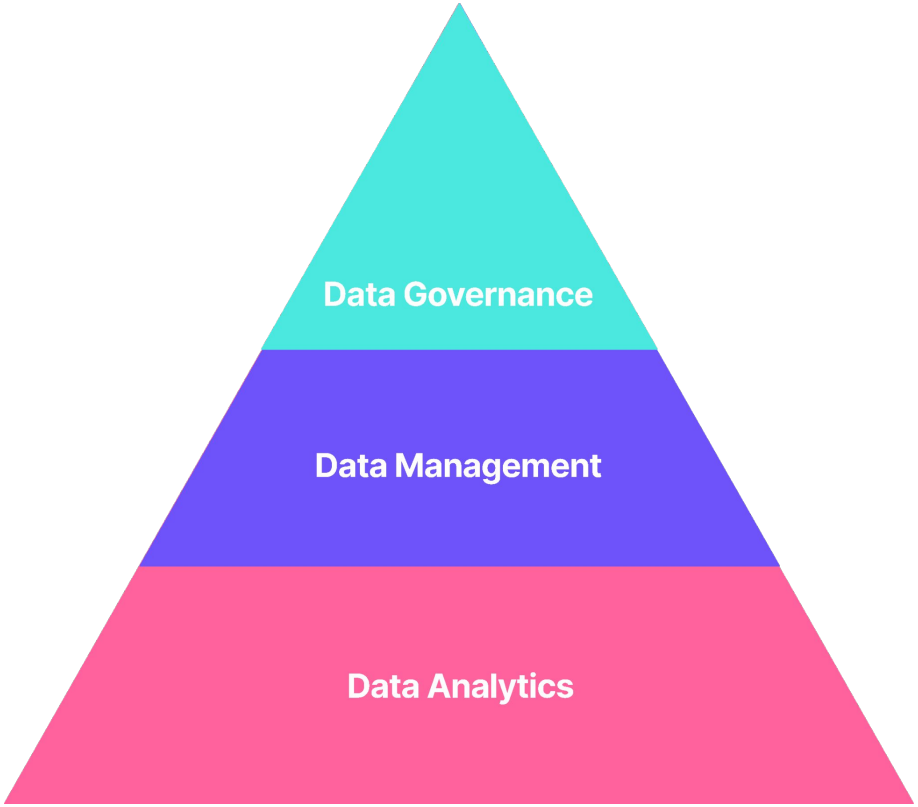


Best Practices in DM & DG

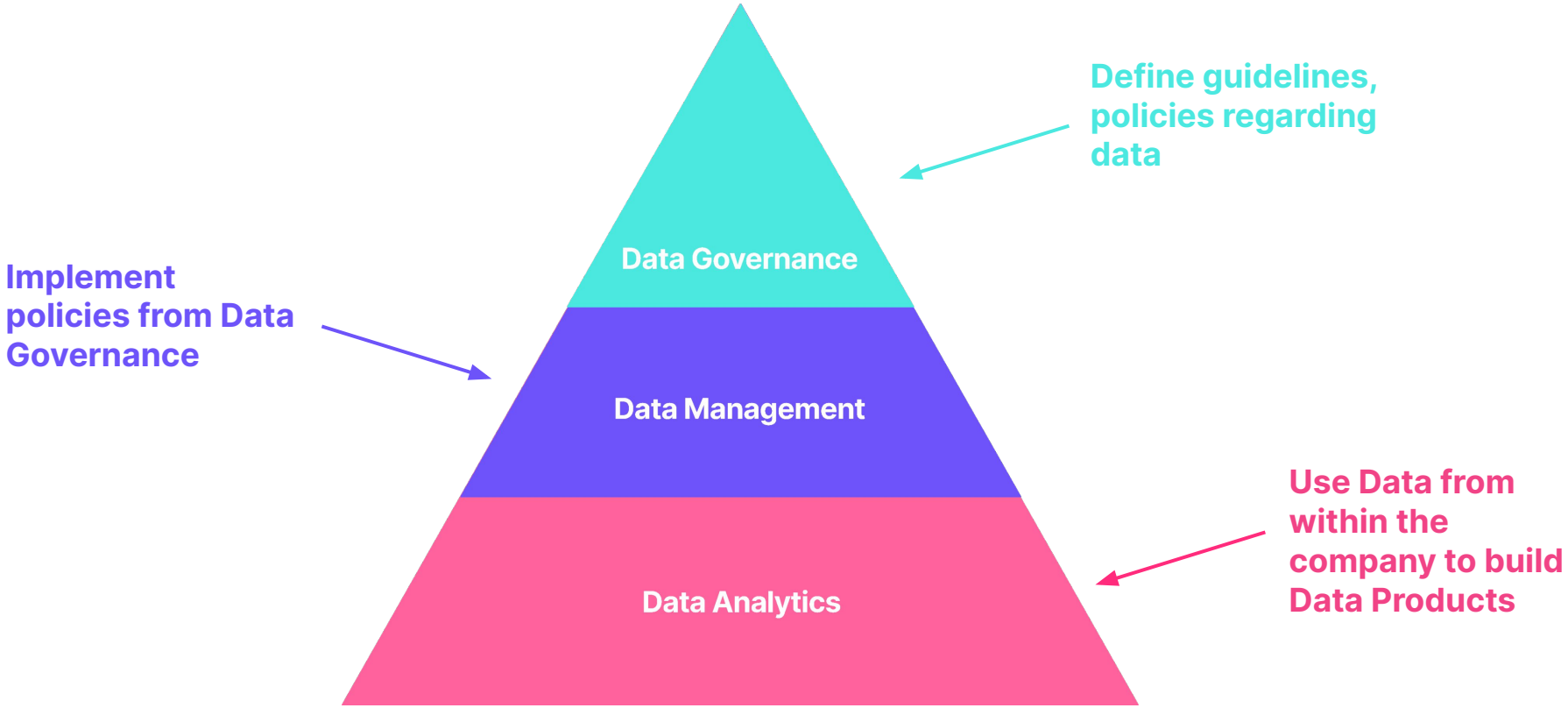
- DM & DG should be done **by different people**
    - *DG should audit whether DM is implementing what has been defined the right way.*
    - *It can be done internally or externally.*
- DM & DG should start as a **separate** departments **then melt into** the other departments.
- IT & Data should be separate
    - *Data is more volatile than IT (that needs to always be exact).*
    - *Counter-intuitively IT might be the most reluctant when dealing with Data projects*

Data in the Organization

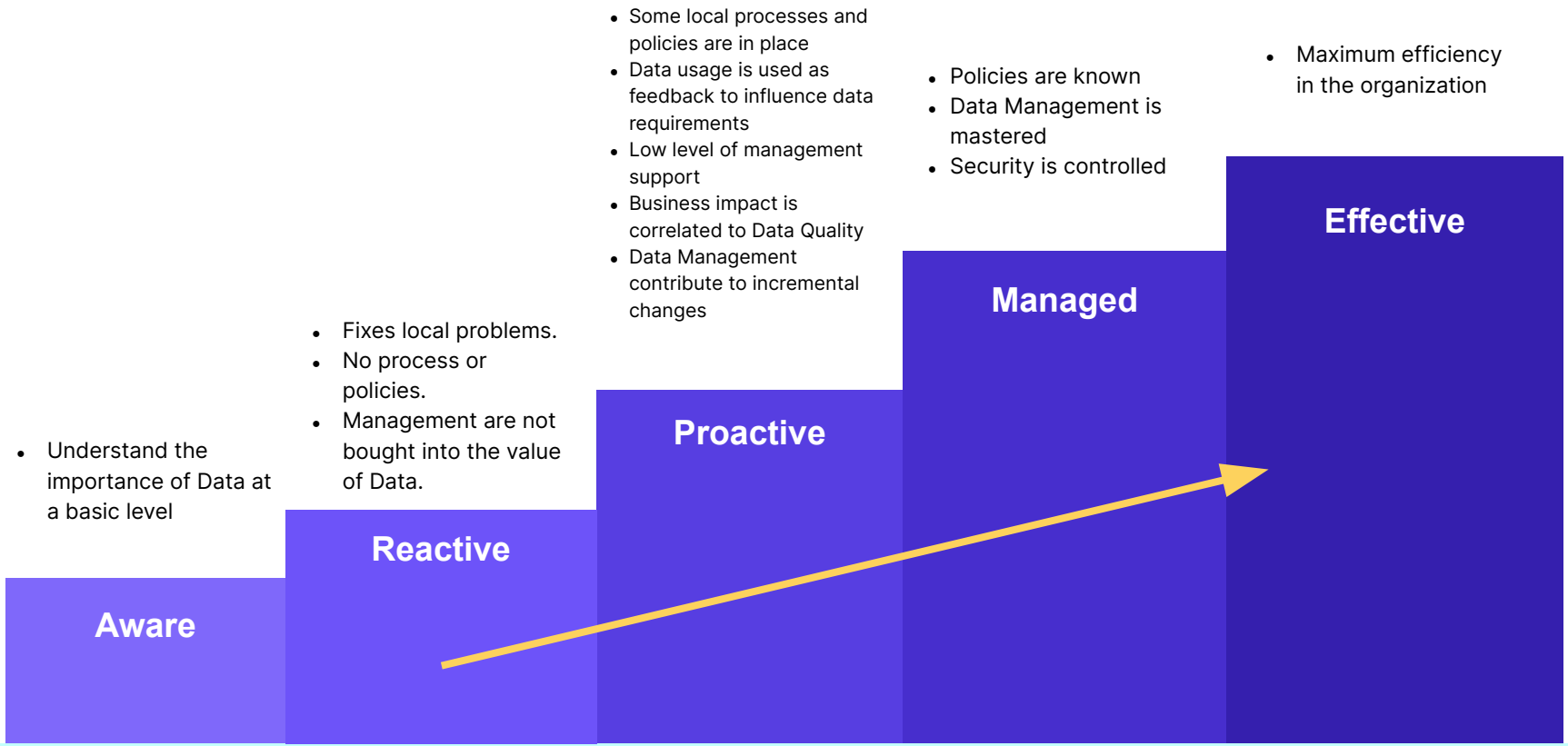
Organisation of Data



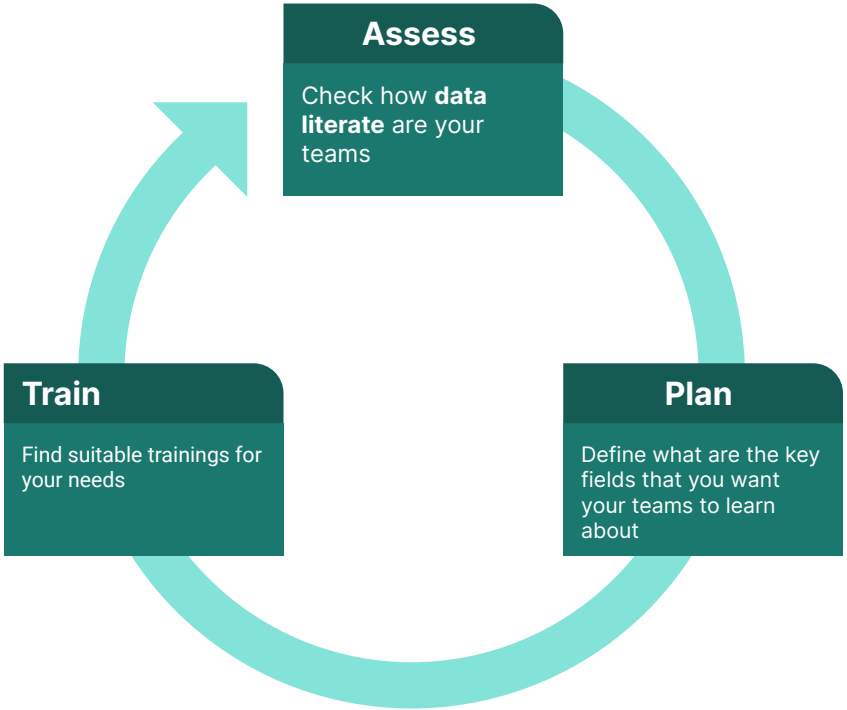
Organisation of Data



Data maturity stages



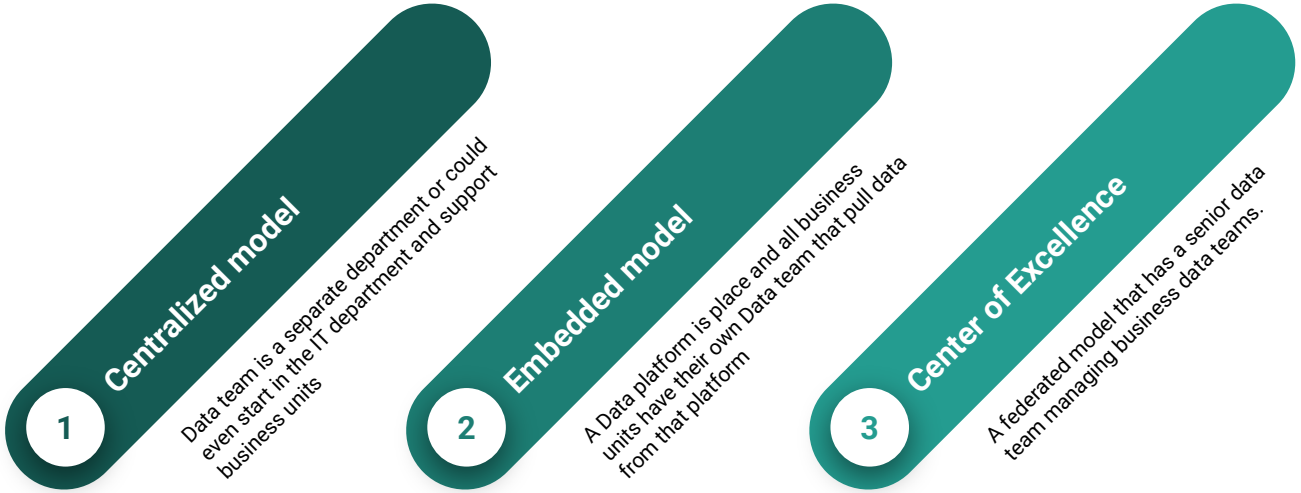
How to cultivate your organisation to become data driven?



Are your teams data-literate?

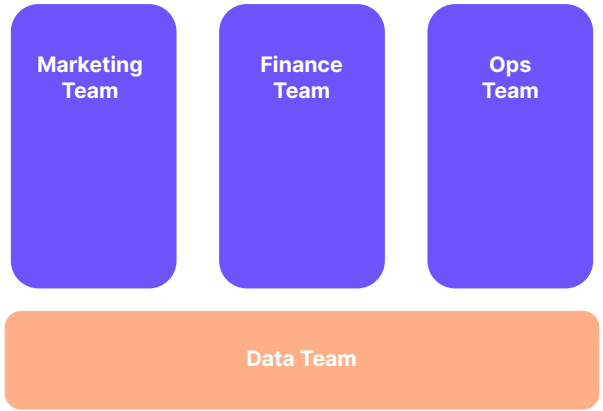
- Data Literacy assessment example

Data Organisation based on company maturity



Data Organisation based on company maturity

Centralized model



Advantages

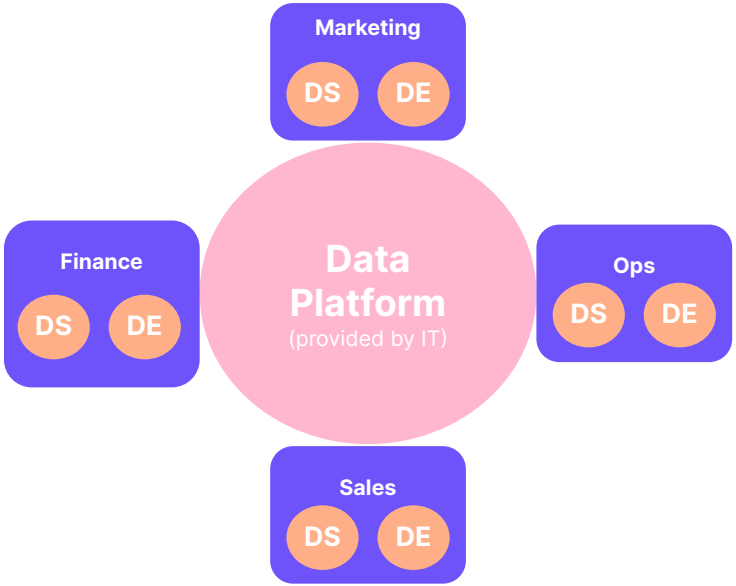
- Simple to implement
- Broader variety of projects
- Easy to prioritise projects
- Clear career growth

Flaws

- Another “support” function - Business function won’t take ownership of the data
- Disconnection between business & Data
- Business functions data needs will become faster than what the data team can handle

Data Organisation based on company maturity

Embedded model



Advantages

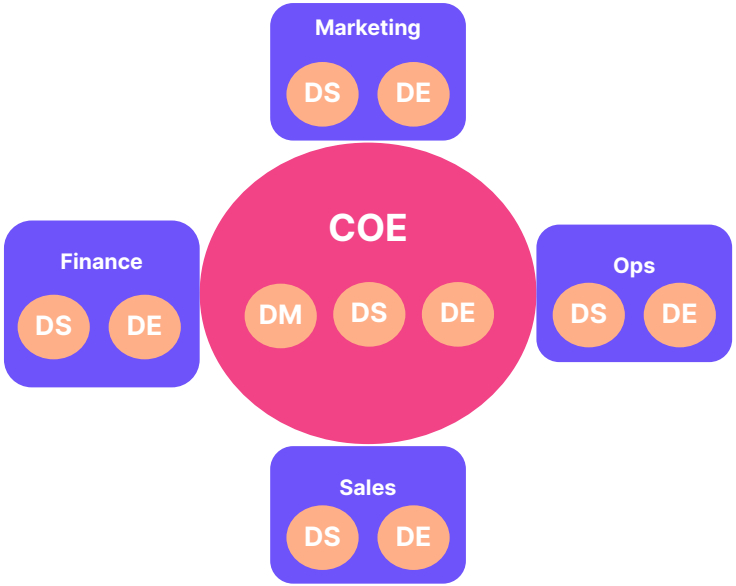
- Each data team is agile
- Business & Data is closely related
- Each data team is more specialized

Flaws

- Lack of one source of truth
- Creation of silos
- Hard for business people to manage them as they might not have the technical background

Data Organisation based on company maturity

Center of Excellence



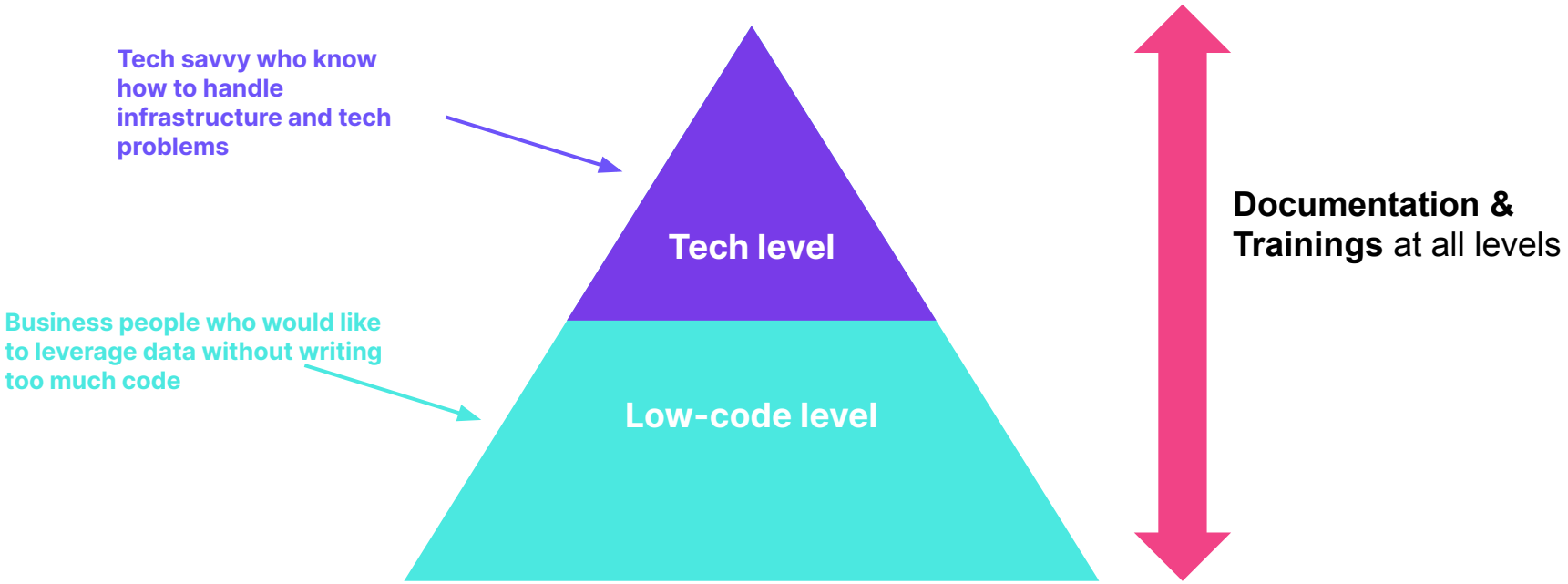
Advantages

- All the advantages of Centralized + Embedded

Flaws

- Not suited for small to mid-sized businesses
- Need an additional layer of coordination

Best practices when building your Data platform



No-code / Low-code tools VS Full-code tools

Purpose	No-code	Full-code
Machine Learning	Dataiku, Preligens	Python (sklearn, tensorflow)
Storage	Airtable	SQL, NoSQL
Analytics	Tableau, PowerBI	Python (pandas)
Automation	Zapier, Make	Python (API)
Documentation	Notion	Github Page

Data Monetization

Direct Selling

- Database selling (leads)
- Data Sharing

Advantages

- Simple to implement

Flaws

- Hard to create new data
- Not necessarily compliant to regulations

Indirect selling

- APIs
- Tiers
- Online Dashboards

Advantages

- Very valuable
- Reliable revenu stream

Flaws

- Hard to implement

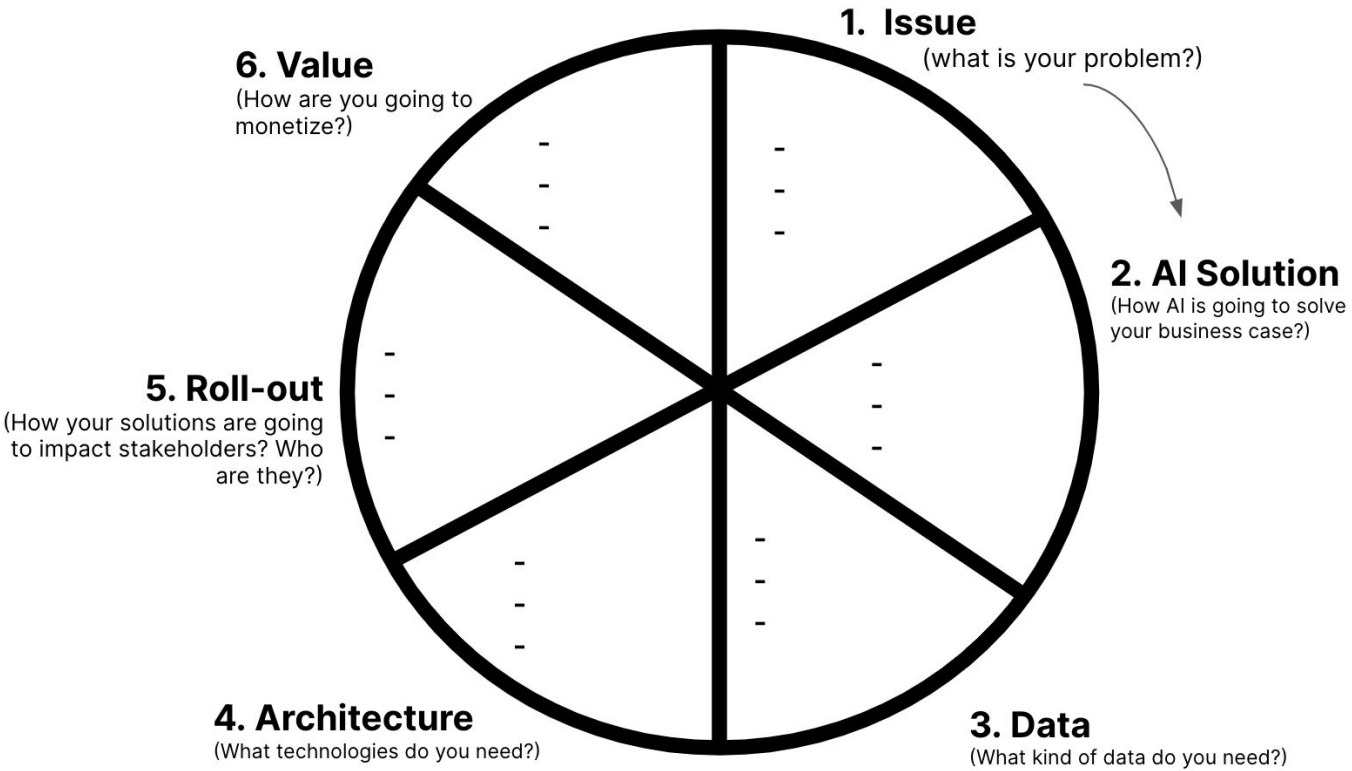
Data Regulation

Use Data with Caution

- Anonymize datasets

Practice

What do you want to accomplish?



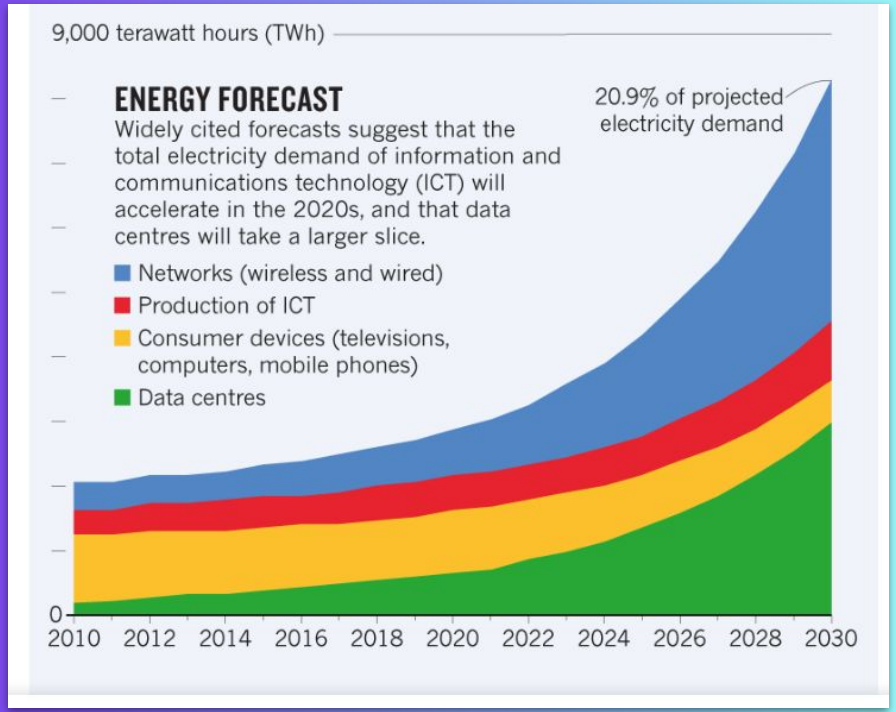
Data Ethics

Training AI with Biais

- Carefully choose train sets
- Make sure your teams are diverse!

Sustainable Big Data

- Don’t be fooled Data Centers are expensive for the planet



Thanks!

See you in the next course