Importance of a Common Data Layer for AI and BI

August 2024
Metin Sarıkaya

About Me

Head of DWH, BI and Big Data Technologies at Akbank, Turkey

15 years of experience in Data & Analytics
Strong knowledge of the finance and telecommunication sectors
Passionate about fostering a data-driven culture and improving data strategies

BS, Industrial Engineering, Istanbul Technical University
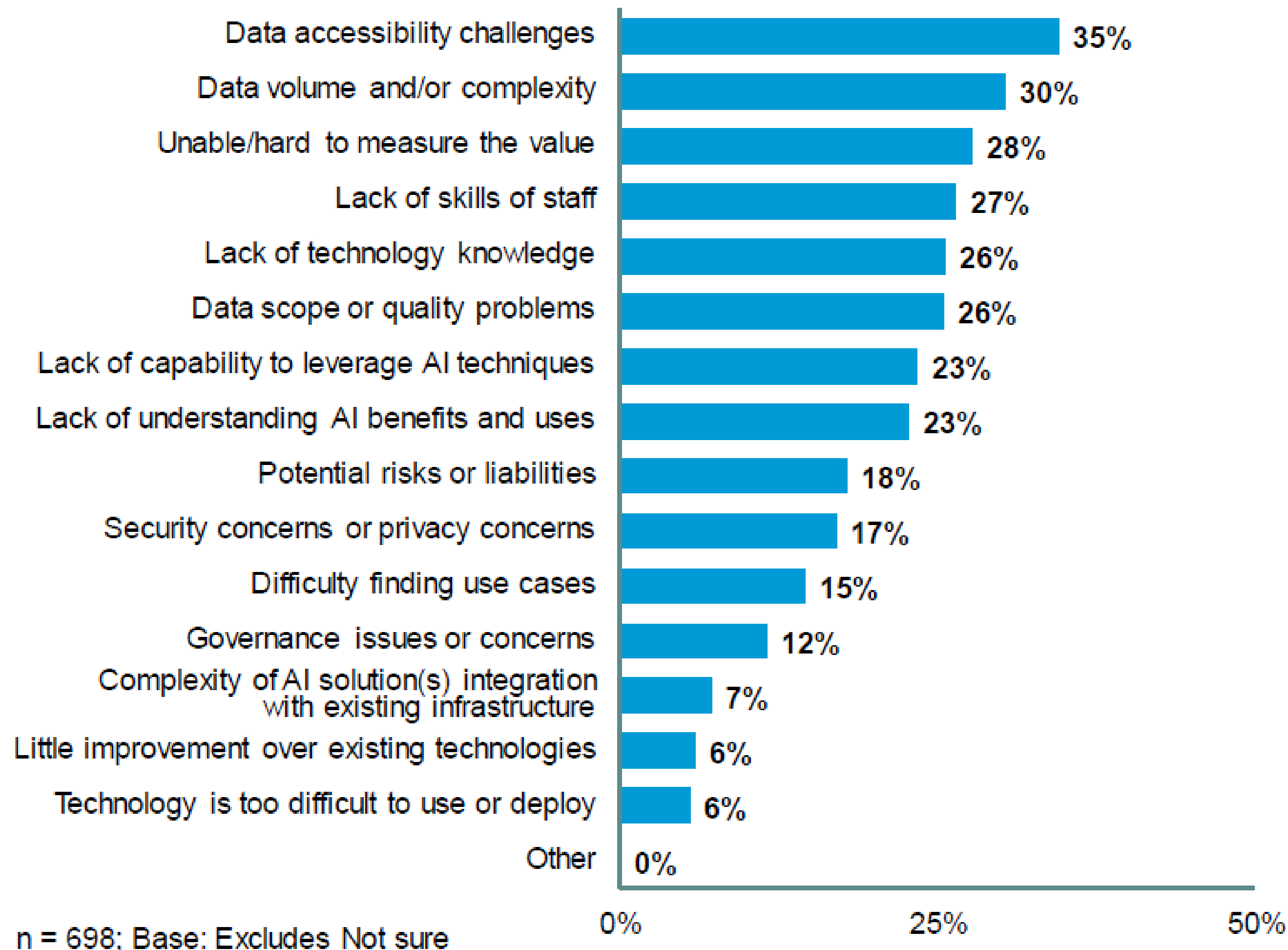MS, Information Technology, Istanbul Technical University
MS, Big Data & Analytics, Sabanci University

Contents

2 of AI's top 3 barriers are related to «Data»

A Gartner survey shows that the top barriers to AI implementation reveal Data as the 1st challenge.

Barrier	%
Data accessibility challenges	35%
Data volume and/or complexity	30%
Unable/hard to measure the value	28%
Lack of skills of staff	27%
Lack of technology knowledge	26%
Data scope or quality problems	26%
Lack of capability to leverage AI techniques	23%
Lack of understanding AI benefits and uses	23%
Potential risks or liabilities	18%
Security concerns or privacy concerns	17%
Difficulty finding use cases	15%
Governance issues or concerns	12%
Complexity of AI solution(s) integration with existing infrastructure	7%
Little improvement over existing technologies	6%
Technology is too difficult to use or deploy	6%
Other	0%

n = 698; Base: Excludes Not sure

01

Data Accessibility Challenges

Limited data accessibility is a critical obstacle preventing AI implementation's full potential and restricting model accuracy.

This also hinders BI efforts, leading to incomplete insights and undermining strategic decision-making.
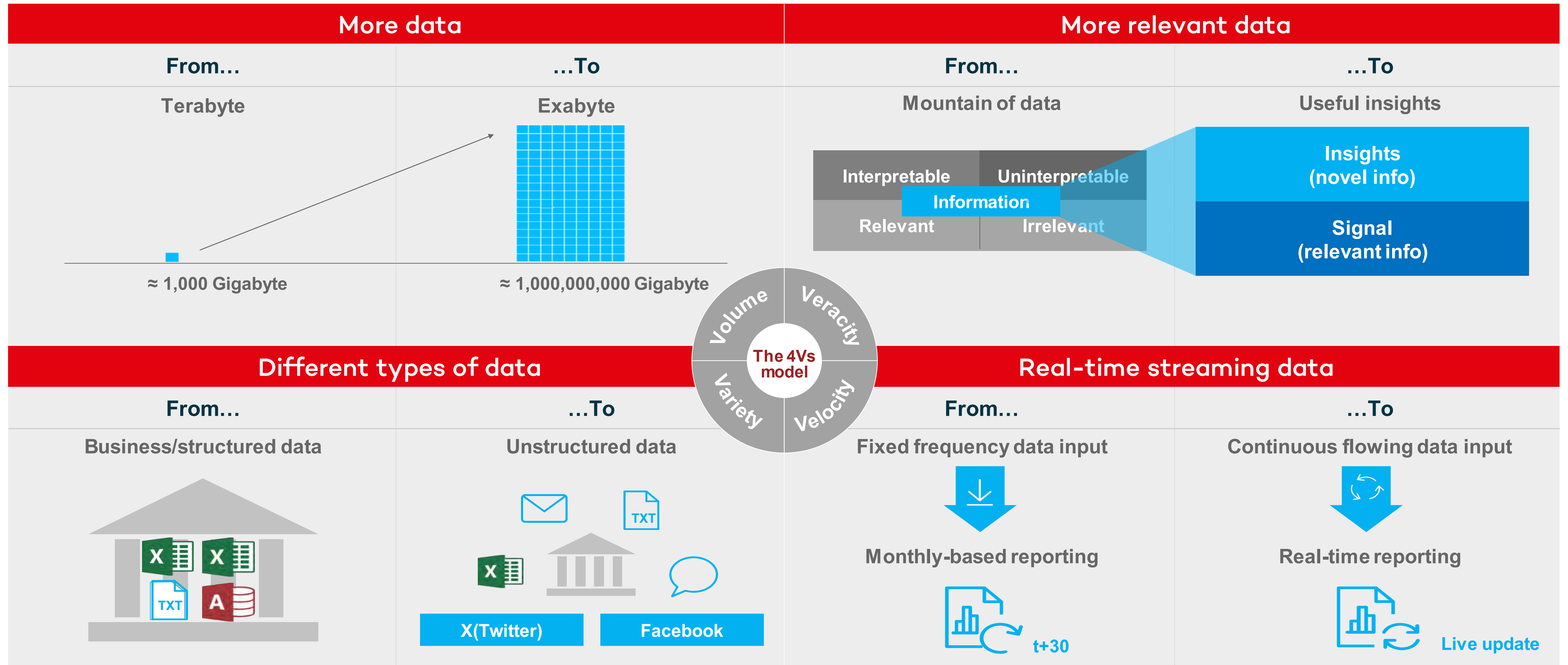
02

Data Volume and/or Complexity

Massive data volume and complexity overwhelm AI and BI systems, making it difficult to process information and generate accurate insights.

Source: Gartner Data & Analytics Conference, 2023, «Develop an AI Governance Strategy to Balance Value and Risk», Jim Hare

Data Revolution: What does it really mean?

Data revolution means access to large volumes of new data. We need to capture the enormous value hidden in this data.

More data		More relevant data	
From…	**…To**	**From…**	**…To**
Terabyte	Exabyte	Mountain of data	Useful insights

≈ 1,000 Gigabyte ≈ 1,000,000,000 Gigabyte

Interpretable · Uninterpretable
Information
Relevant · Irrelevant

Insights (novel info)
Signal (relevant info)

The 4Vs model
Volume · Veracity · Velocity · Variety

Different types of data		Real-time streaming data	
From…	**…To**	**From…**	**…To**
Business/structured data	Unstructured data	Fixed frequency data input	Continuous flowing data input
	X(Twitter) · Facebook	Monthly-based reporting	Real-time reporting
		t+30	Live update

The main challenges of «Data» for organisations

Organisations need help in efficient data management and strategic decision making.

Business Challenges
- Lack of data dictionary/business glossary
- No single point of contact for data needs
- Need for more maturity in data governance processes.

Technology Challenges
- Lack of well-structured data architecture
- Siloed and fragmented data management
- Repetitive data in different data marts
- Data quality problems

Organisational Challenges
- Lack of CDO / CAO roles
- Lack of Data/Reporting/AI-related roles
- Unclear definition of roles and responsibilities

| QUALITY | TIMELINESS | REPORTING | DATA COMPLEXITY | BUSINESS NEEDS REALISATION | AI & ML MODELS |
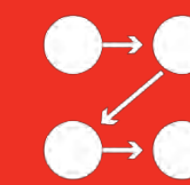
Common Data Marts Layer: Is it magic?

CDMs with high-quality, well-defined, easy-to-use, and flexible data are crucial for AI and BI activities.

Common Dimensions																			Data Marts
Date					Customer Org. Unit				Account Org. Unit				Cust	Product Hierarchy					
Day	Week	Month	Quarter	Year	RM / PM	Business Unit	Branch	Organisation	RM / PM	Business Unit	Branch	Organisation	Customer	Prd. Name	Prd. Group	Prd. Sub Group	Prd. Category		

Facts / KPIs

- Cross-Sell by Customer based
  - Common data mart
  - Marketing view
  - Finance view
- Cross-Sell by Portfolio based
  - Common data mart
  - Portfolio view
- Loan for Credit Risk
  - Common data mart
  - Credit Risk view
- Loan for Portfolio Mng.
  - Common data mart
 - Portfolio view

CDM Layer consists of:

- Fact tables, organised around common dimensions
- Calculated dimensional attributes
- Flat hierarchies of core dimensions
- Very detailed measures to promote future
- re-uses
- Features(input variables) for use in AI and ML models
- Star schema structures for high-performance BI queries

CDM Layer has:

- Business owners
- High-quality data
- Well-defined data model
- Business glossary for definitions
- Flexible structure that makes it easy to add new measures in the future
- An intuitive user experience that allows for effortless usage

Benefits of CDM Layer for BI

CDMs provide many advantages to BI activities, such as reliability, consistency, ease of use, and simplicity.
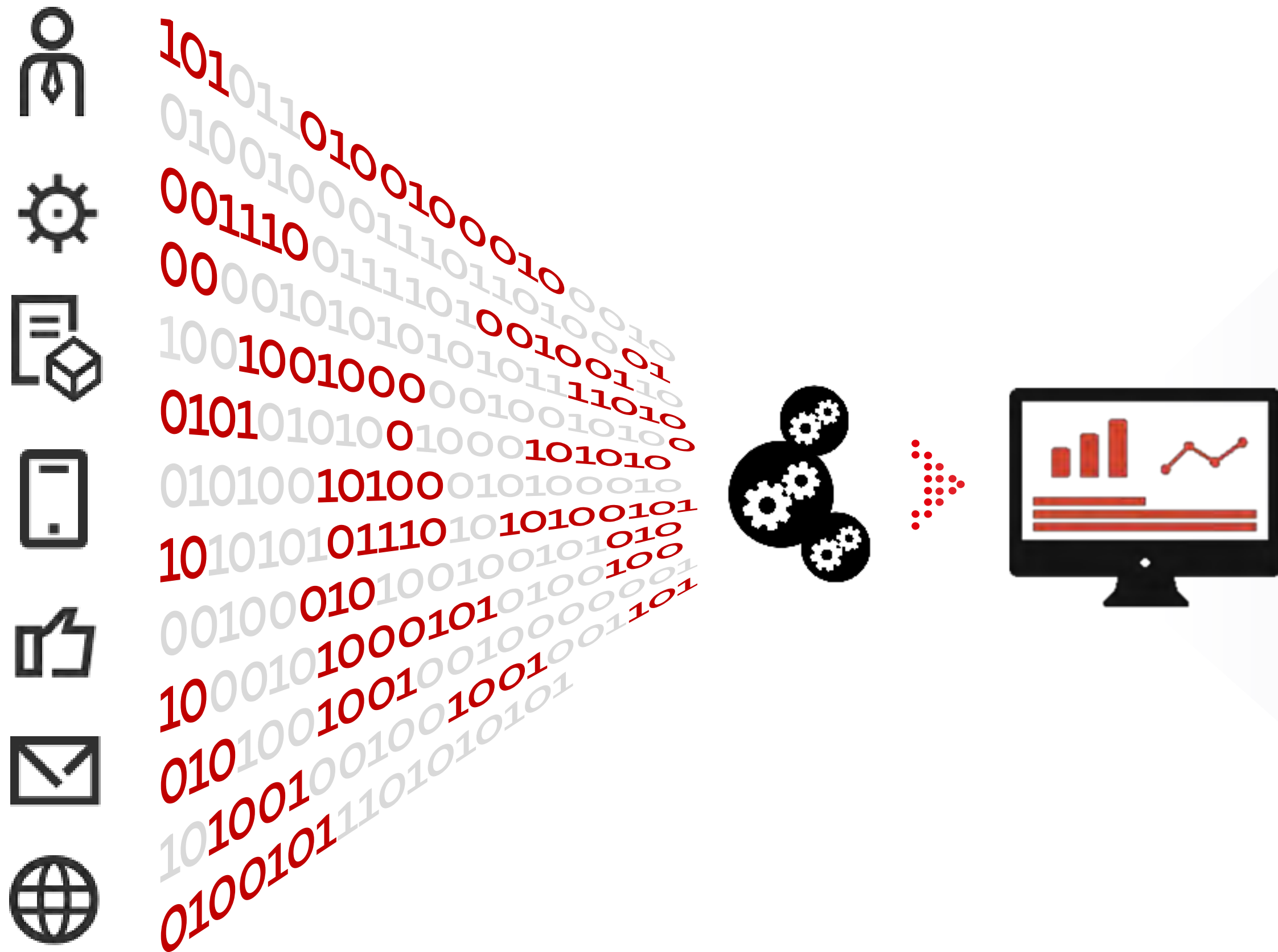
The reports produced are **reliable** and **consistent** with each other.

Different reports prepared on the same topic yield the **same results**.

It is very useful for **data democratisation**.

Self-service BI activities can be carried out much more **easily** and **quickly**.

Thanks to the dimensional-modelled data structure, report **queries** are **simpler**, and **reports** run **faster**.

When supported by **data governance**, it provides significant benefits in reporting and analysis activities across the organisation.

Benefits of CDM Layer for AI

A Feature Store is a critical component in the AI ecosystem, designed to manage and serve features consistently across different stages of the AI lifecycle.
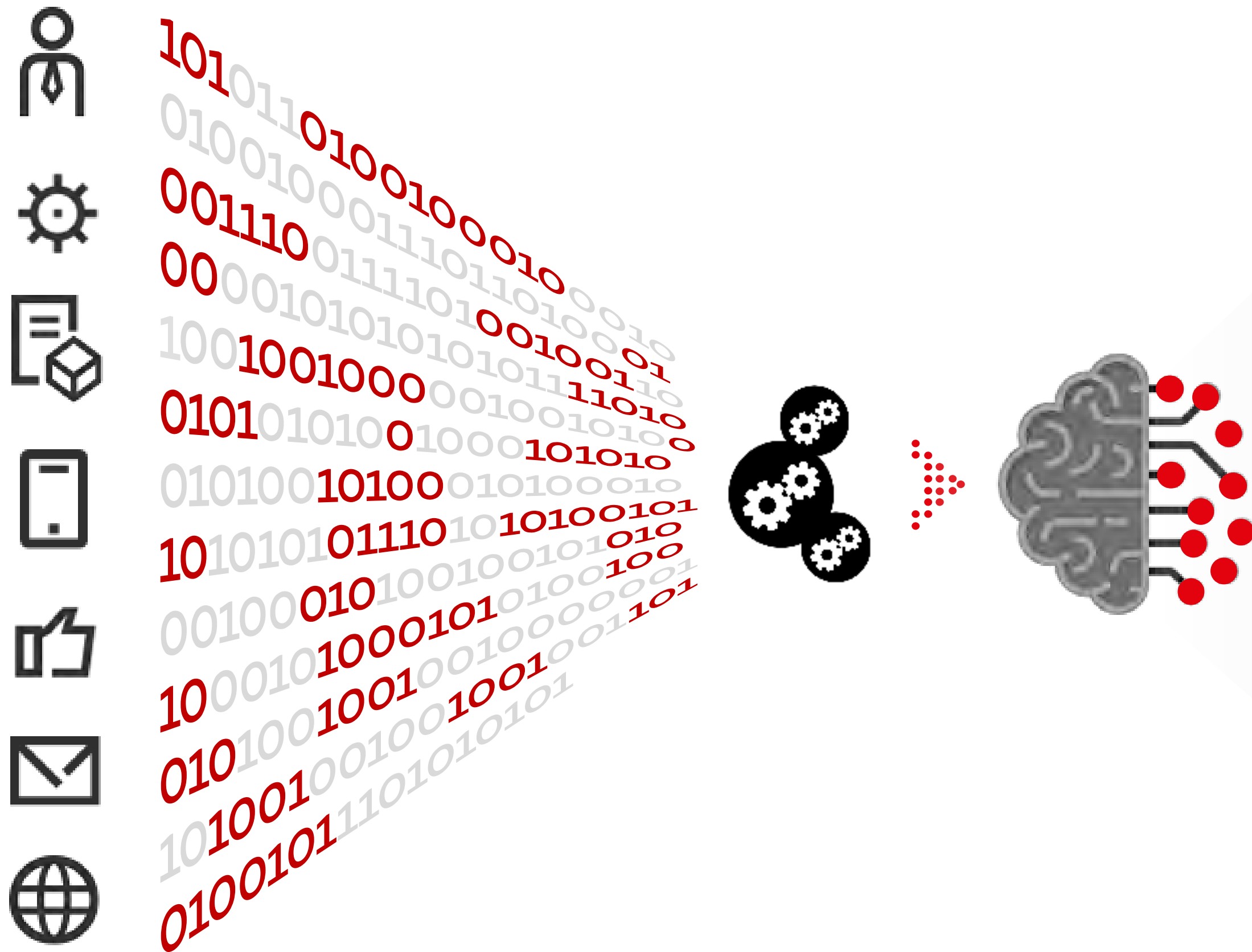
The CDM layer can be used as a **feature store**, making it easy to create new features using existing features. The **measures** and **attributes** in CDMs are basically **input variables**.

As organisations scale their AI efforts, managing features across **multiple models** and teams can become **challenging**. Feature stores **simplify** this by centralising feature mng.

By **reusing features** and reducing duplication of effort, feature stores **speed up** the model development process.
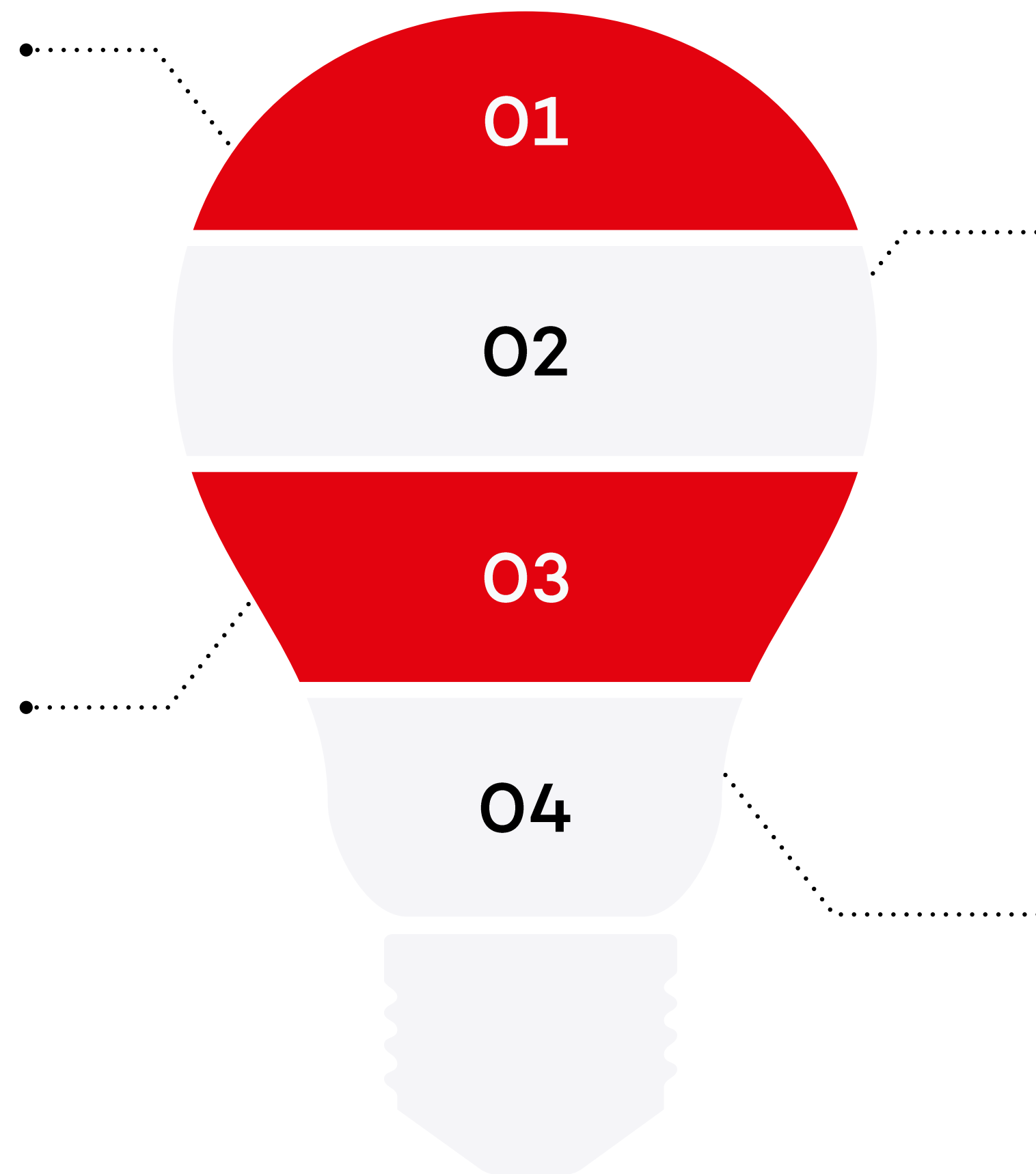
With **centralised** feature management, it is easier to enforce **data governance** policies, track feature usage, and ensure compliance.

Conclusion

The future of data-driven Success depends on how effectively organisations utilise the CDM Layer to unlock the full potential of their data.

The Common Data Marts(CDM) layer is essential for both AI and BI, providing a foundation for reliable, consistent, and efficient data management.

The CDM Layer functions as a crucial feature store, supporting AI initiatives by ensuring high-quality, reusable, and well-managed data across the AI lifecycle.

The Data Revolution requires a strategic approach to data management, with the CDM Layer playing a central role in overcoming challenges like vast data volumes, complex structures, and the need for real-time insights.

Leveraging the CDM Layer is not just a technical solution but a strategic necessity for driving innovation and achieving meaningful, impactful outcomes in AI and BI.

O1
O2
O3
O4

Thank you