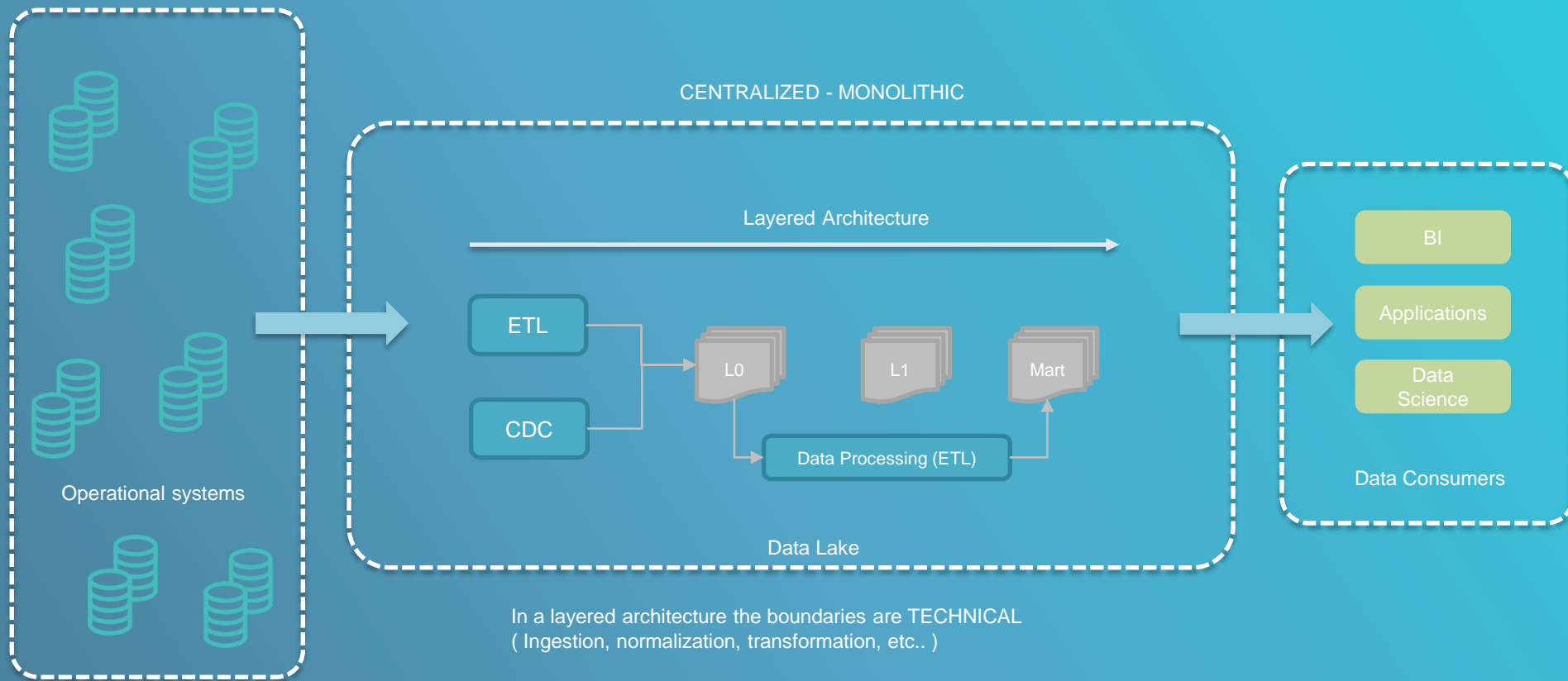




agilelab

Data Lake is dead  
long live the Data Mesh

# Data Lake – overall process



# Data Lake – Org perspective



- They are not aware about the process of data warehousing
- They don't know about data consumer needs
- They are not focused on providing quality on their data because it is not on their purpose.

**Disengaged in creating added value for the organization**

- They suffer about data quality lack and change management driven by the source domains.
- They need to know all the domains – too hard.
- They get requirements from data consumers.

**This team is becoming a bottleneck.**

- They don't understand why is so hard and time consuming to have new requirements implemented.
- they have to wait the datalake team before to have a new dataset available.
- They don't understand if problems are on data source or on data lake

**No trust on data**



Source systems



ETL

CDC

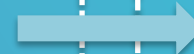
L0

L1

Mart

Data Processing (ETL)

Data Lake



BI

Applications

Data  
Science

Data Consumers



# What is Data Mesh?

- ***Serving*** over *ingesting*
- ***Discovering*** and ***using*** over *extracting* and *loading*
- ***Publishing events as streams*** over *flowing data around* via centralized pipelines
- ***Ecosystem of data products*** over *centralized data platform*

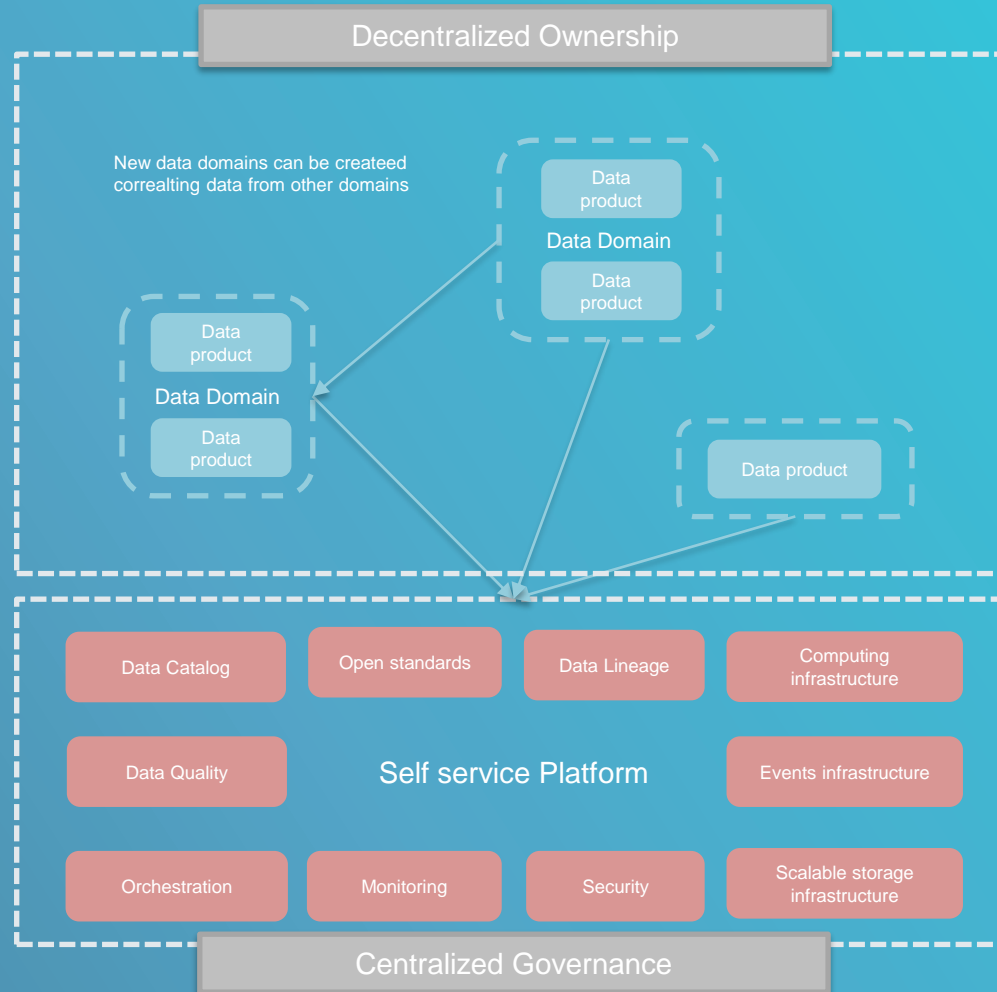
# Data Mesh concept



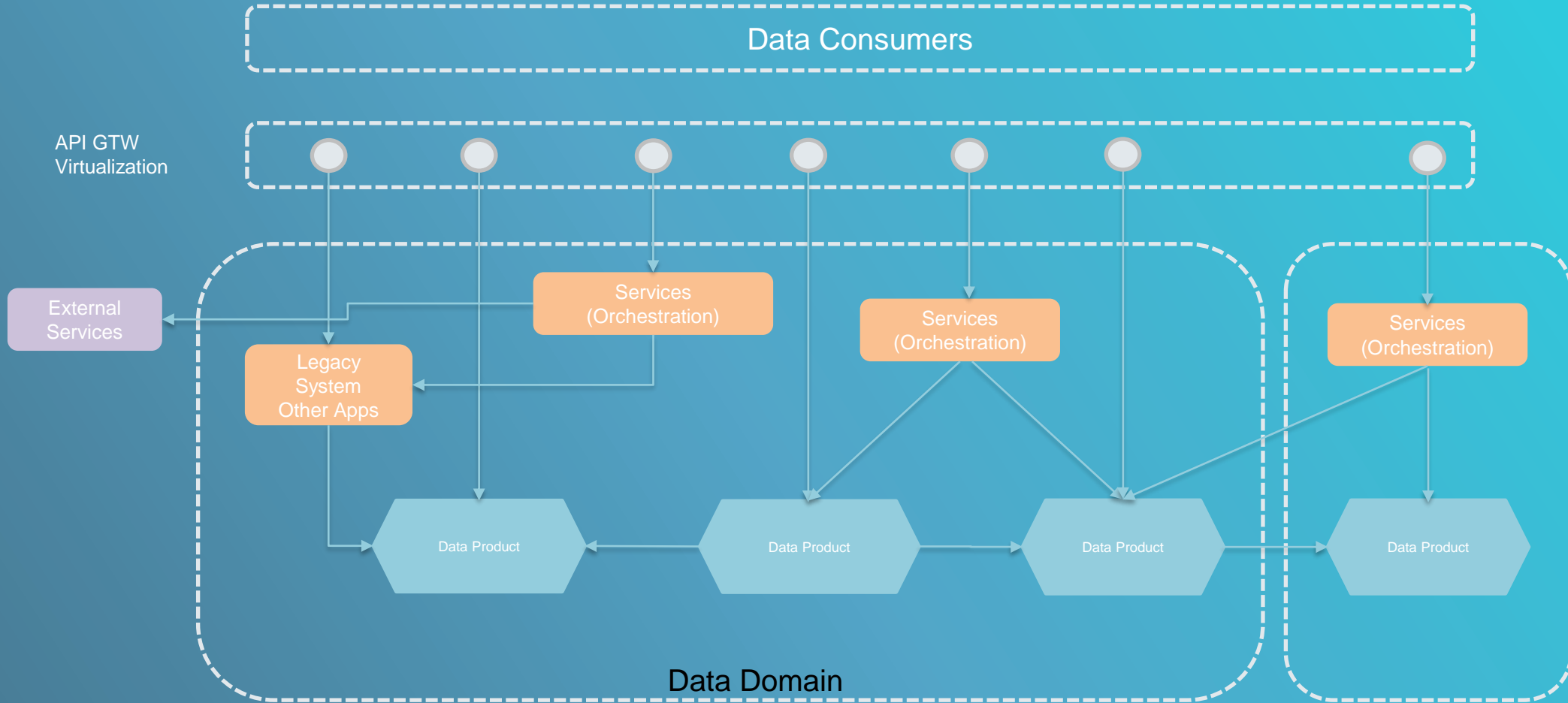
## Paradigm shift

- Data as a product
- Cross Functional domain-data teams
- Ecosystem of data products
- Decentralized ownership
- Data product is polyglot ( OLTP, OLAP, Events )

Decentralized ownership is enabling faster results and no team bottleneck ( like the datalake team is )



# Data Mesh overview



# Data Product



This is the **architectural quanta**. An independent component with **high functional cohesion**, that can be built, deployed and operated by a **single cross-functional team** without complexity explosion.

**It is easier to scale and parallelize work across multiple teams.**

## Input Ports

- Operational systems
- Other Data Products
- External services



## Control Ports



Observability API

Descriptive API

Data Pipeline

Data Access

Stream Processing

Internal processes (GDPR, DQ, etc )

Data + Metadata  
(syntax+semantic, expected behaviour, access control )

Infrastructure

## Output Ports

- Events
- SQLView
- Raw

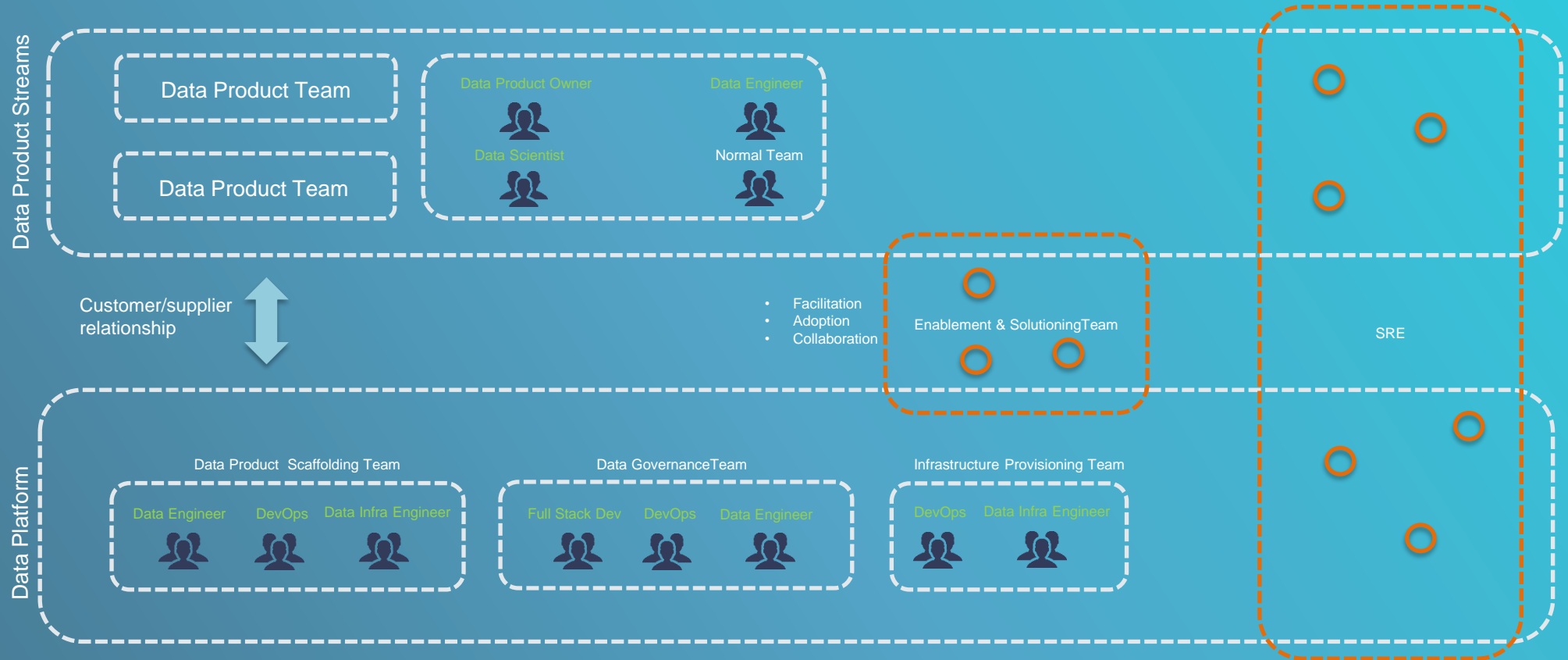


- Discoverable
- Addressable
- Trustworthy and truthful (SLO)
- Self Describing
- Interoperable
- Secure
- Independent ( provision and deployment )





# Team Organization







# Suggestions for adoption

- **Start from the sources and build data products to reduce the footprint of data lake**
- **Blind everything with data virtualization to smooth the data consumer experience during the migration**
- **Plan it carefully**
- **Plan it to handle exceptions**
- **Initial use case & tech assessment is mandatory**



# Benefits

- Data products centralize ownership for privacy, data lifecycle, scaling, workload isolation and security
- Self-contained data products are enabling composition over inheritance
- No a-priori data duplication
- DWH Virtualization reduces costs ( CAPEX & OPEX )

# Thank you!



×

## Next steps?

×



[www.agilelab.it](http://www.agilelab.it)



[info@agilelab.it](mailto:info@agilelab.it)



[www.linkedin.com/company/agile-lab](https://www.linkedin.com/company/agile-lab)