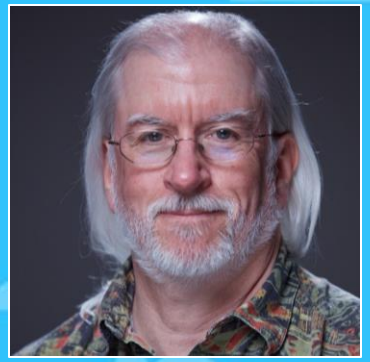
USING DATAOPS TO BUILD AND MANAGE DATA PRODUCTS AND DATA MESH ON SNOWFLAKE

The Dream Team



Kent Graziano

Non-Executive Advisor

DataOps.live

Omar Khawaja

Head BI & Analytics

Roche

Paul Rankin

Head of Data Management & Architecture

Roche

Guy Adams

Chief Technology Officer

DataOps.live

2

THE VISION

Innovating Diagnostics, shaping healthcare, changing lives

Doing now what patients need next

*Maximize outcomes for **customers**, **patients** and **Roche** through innovative **data & analytics products***

Legacy infrastructure & platforms

High level view of the legacy data infrastructure, limited and non scalable

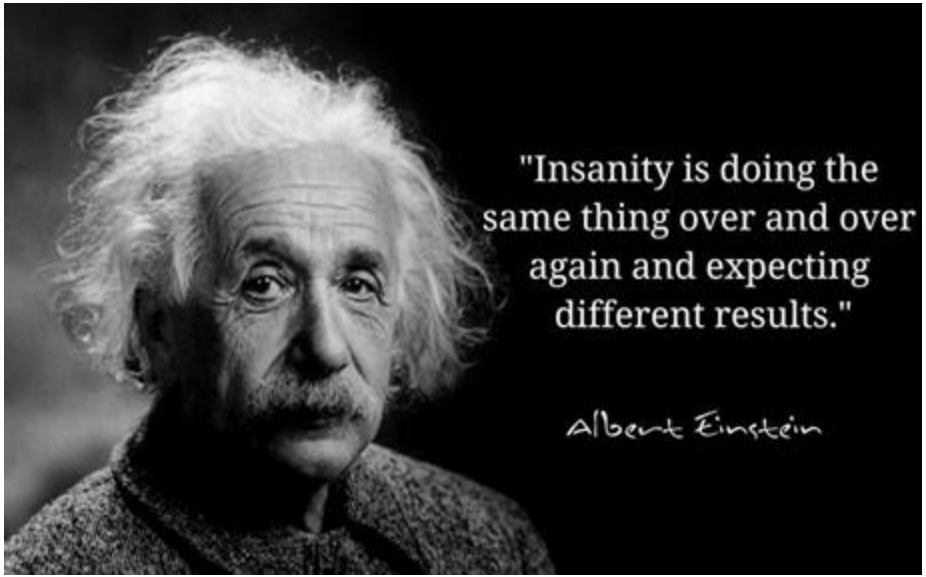
Classic BI Setup: Monolithic & hard to change

All BI infrastructure & data sources are on-premise (data sources, excl. Salesforce)

Multiple physical & virtual servers, hard to maintain and slow to scale-up

Pain Points

3 months	Lead time to scale up / increase compute
3-4 months	Average release cycle / time to market
2 - 3	Major incidents per year
4 days	Average time to implement a Hotfix
M (medium)	Reliability satisfaction
L (low)	Performance satisfaction



"If you want to bring a fundamental change in people's belief and behavior...you need to create a community around them, where those new beliefs can be practiced and expressed and nurtured."

- Malcolm Gladwell

Data Mesh Approach

4 principles of Data Mesh to handle the traditional modes of failures

Domain Oriented Decentralization

Soul

The ***domain oriented approach*** as the organizing principle for data allows us to combine domain expertise with the technological capabilities (of the self-serve data infrastructure) necessary to generate business value

Data as a Product

Heart

Viewed through the lens of the domain, ***data*** can become a portfolio of discrete ***products***

Self-Service Data & Analytics Infrastructure as Platform

Body

Self-service data infrastructure enables the product teams to create data products

Federated Computation Governance

Mind

One of the core features of the data mesh is its ***federated governance*** model that achieves interoperability through standardization. Only with interoperable data can analyses involving multiple data products lead to valuable insights and action

SINGLE DEPARTMENT

How do we get from Centralized to Distributed

Central IT Ownership

Organizationally, it shifts from centralized ownership of data by specialists who run the data platform technologies to a decentralized data ownership model pushing ownership and accountability of the data back to the business domains where data is produced from or is used.

Domain Ownership

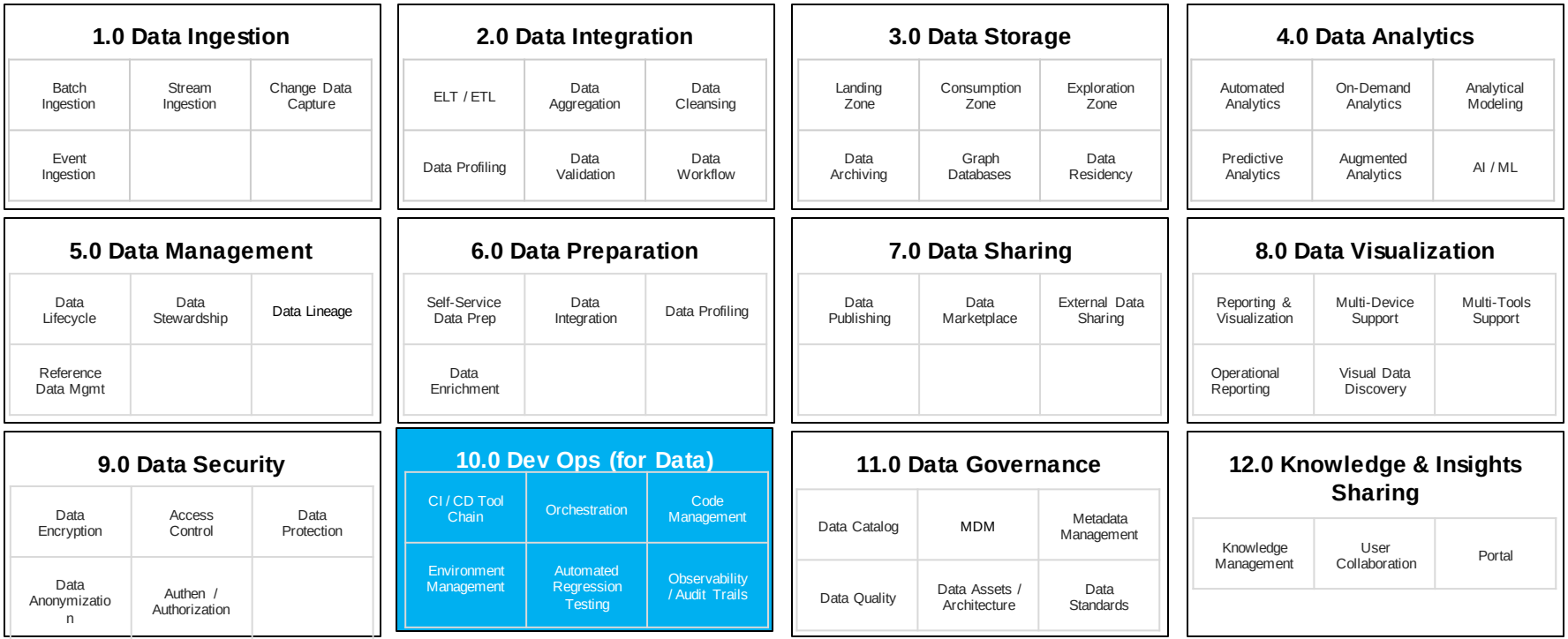
Monolithic

Architecturally, it shifts from collecting data in monolithic warehouses and lakes to connecting data through a distributed mesh of data products accessed through standardized protocols.

Distributed

Self-Service Data & Analytics Infrastructure as Platform

Designing the platform with data product teams as end customers based on capabilities



Single Domain Requirements

Developer Agility

Same agility & speed as standard Software Development processes

Orchestration

Orchestration of other tools in the pipeline: Talend, Alteryx, Python,

Automated Testing

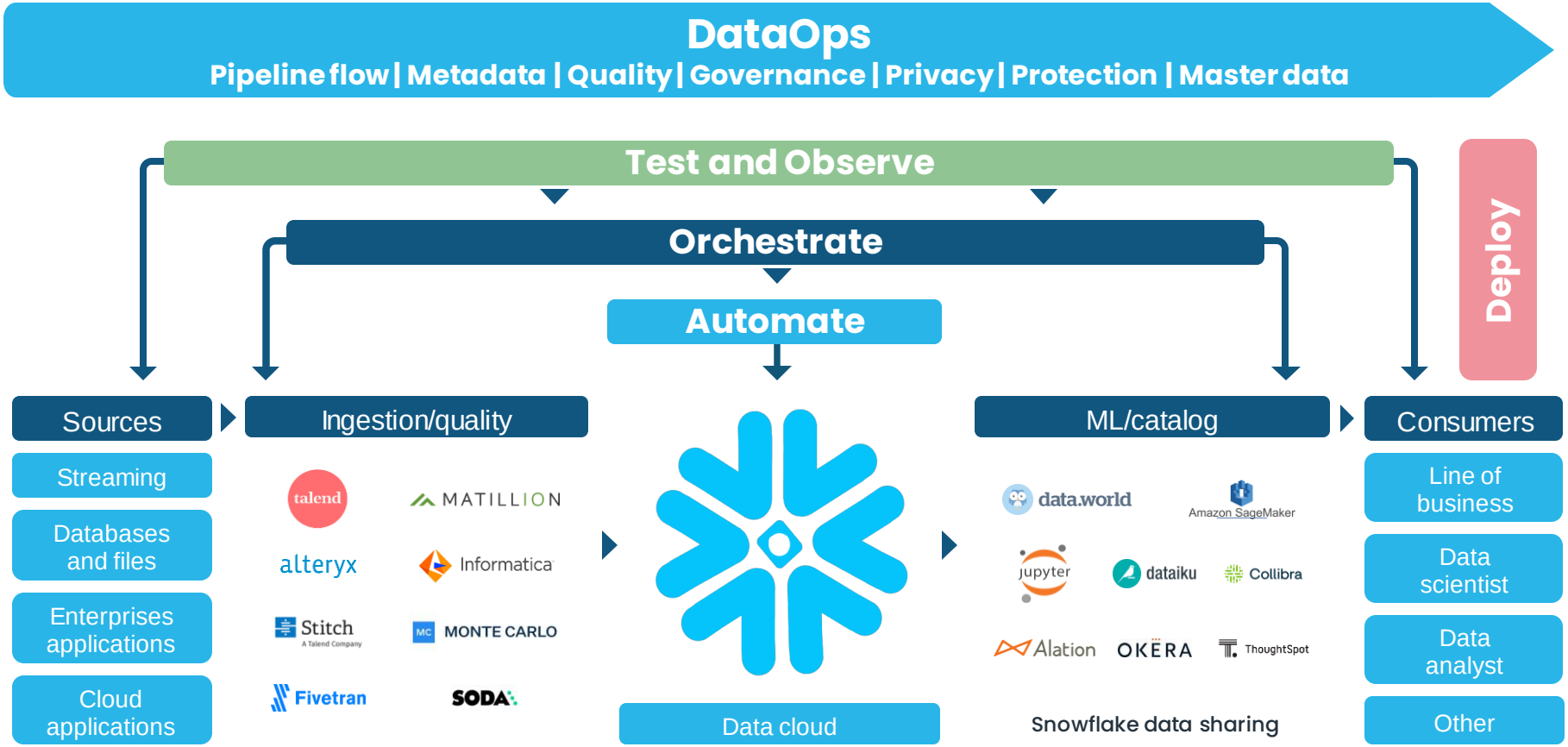
Commencement and Orchestration of Automated Data Regression Testing

Clear Directions

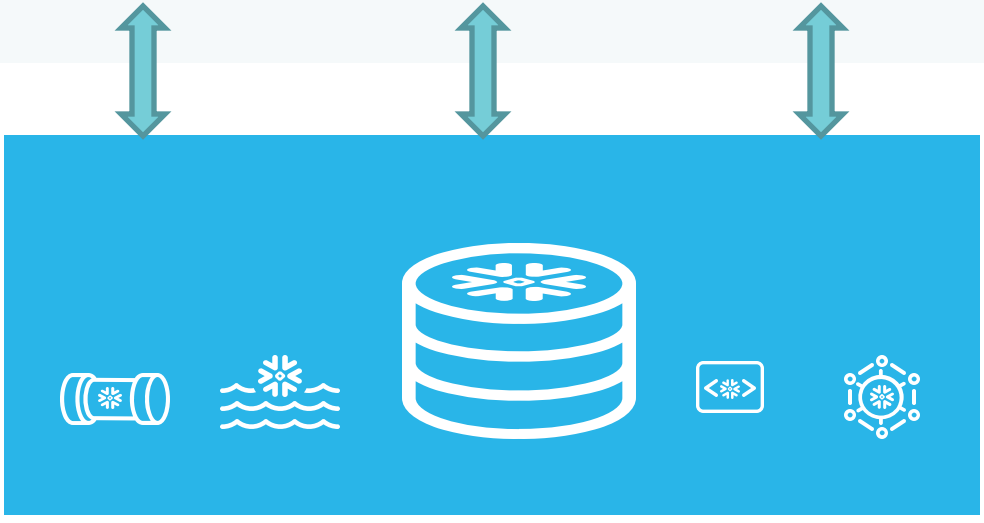
Simple clear directions on how to build, test and deploy data products in Roche

SINGLE DEPARTMENT IN DATAOPS

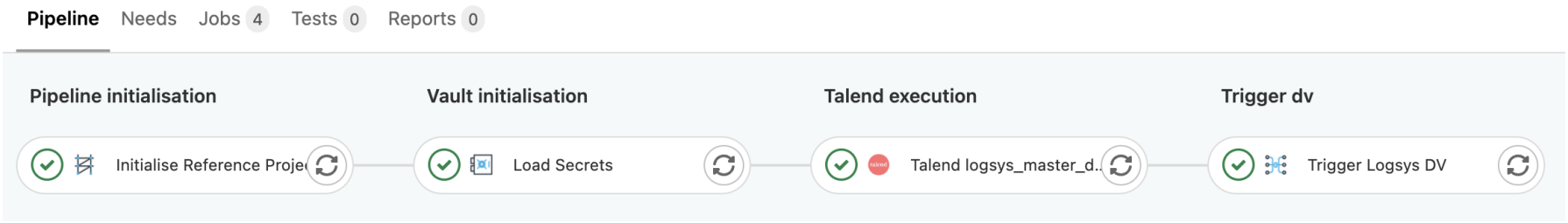
DataOps

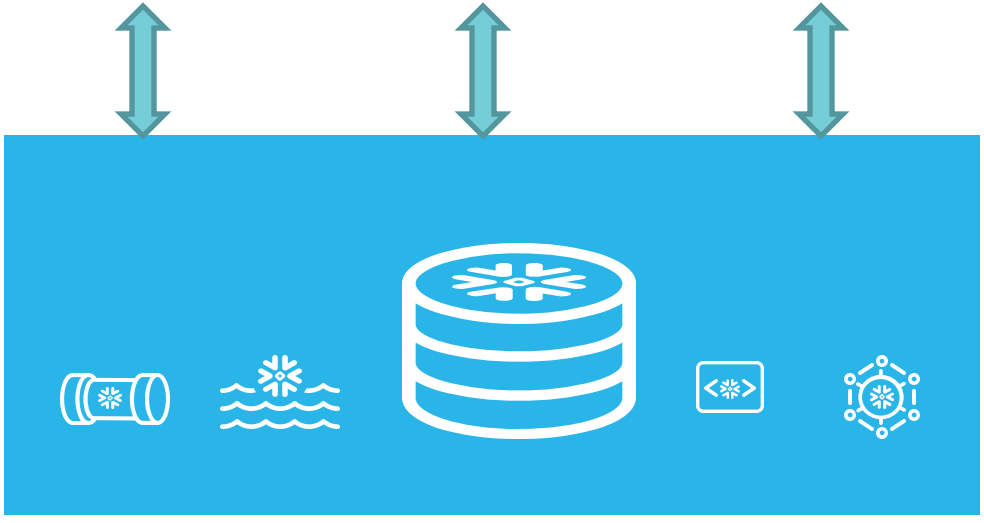


AUTOMATE

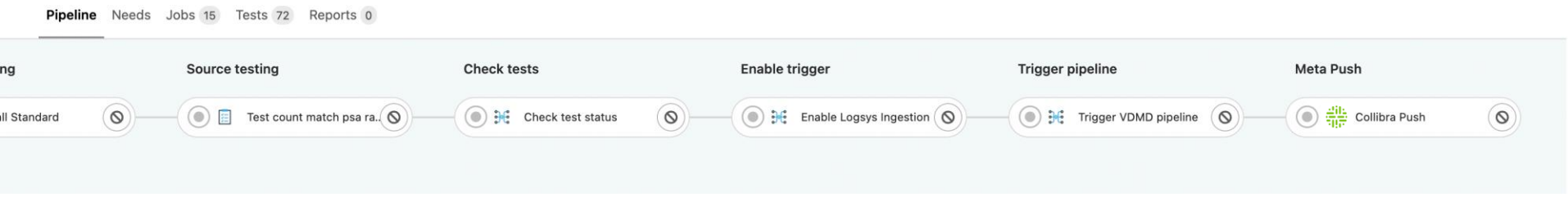


ORCHESTRATE

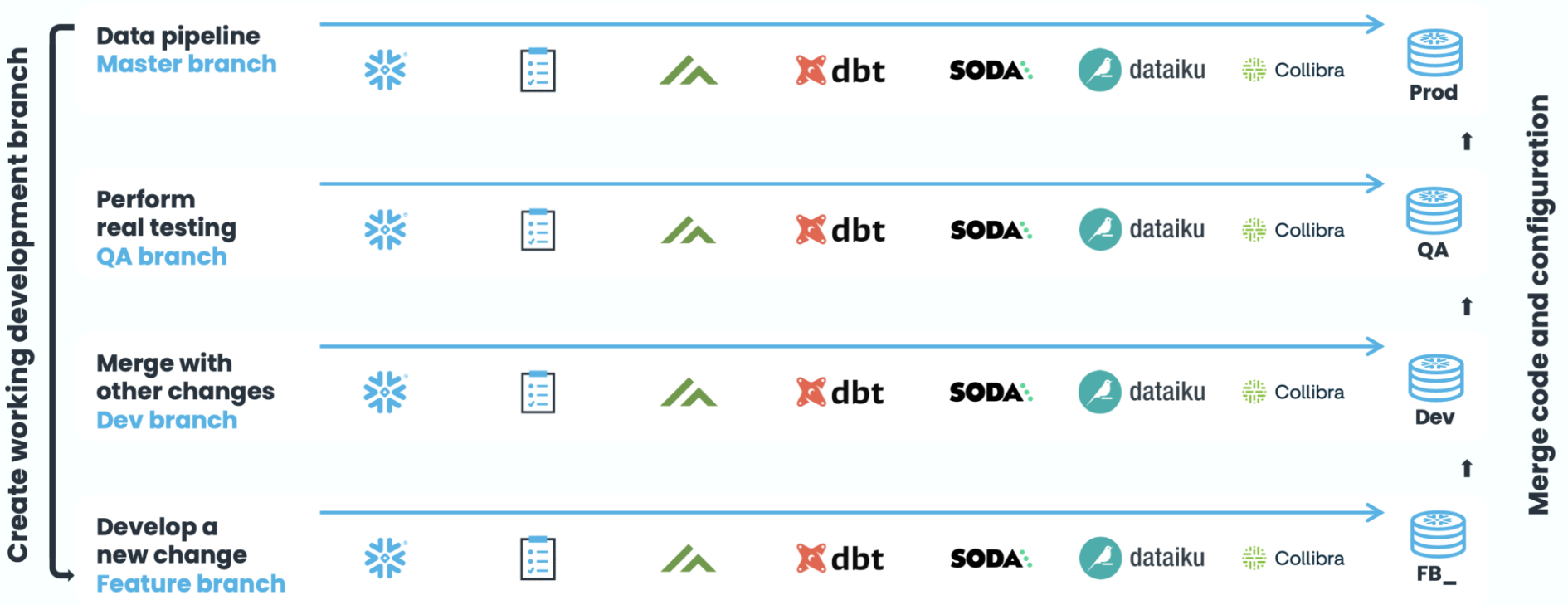




TEST AND OBSERVE



DEVELOP AND DEPLOY



FULL MULTI-DOMAIN DATA MESH

Multi-domain requirements

Interoperability

… of data products across multiple domains and the enterprise

Federated Governance

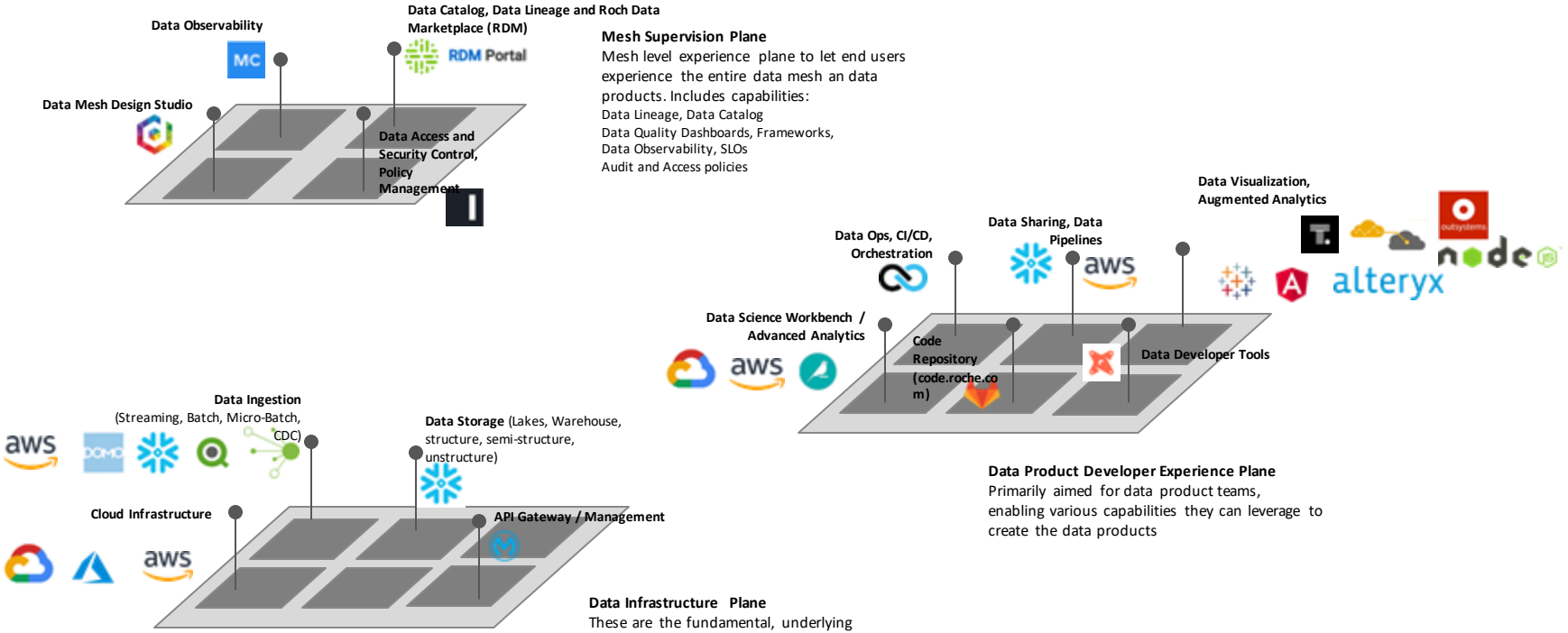
Governance, Data Security, Observability, Auditability & Policy Enforcement

Enterprise Agility

All domain teams enabled to move forward faster … but consistently

Self-Service Data & Analytics Infrastructure as Platform

Platform setup towards the data mesh experience

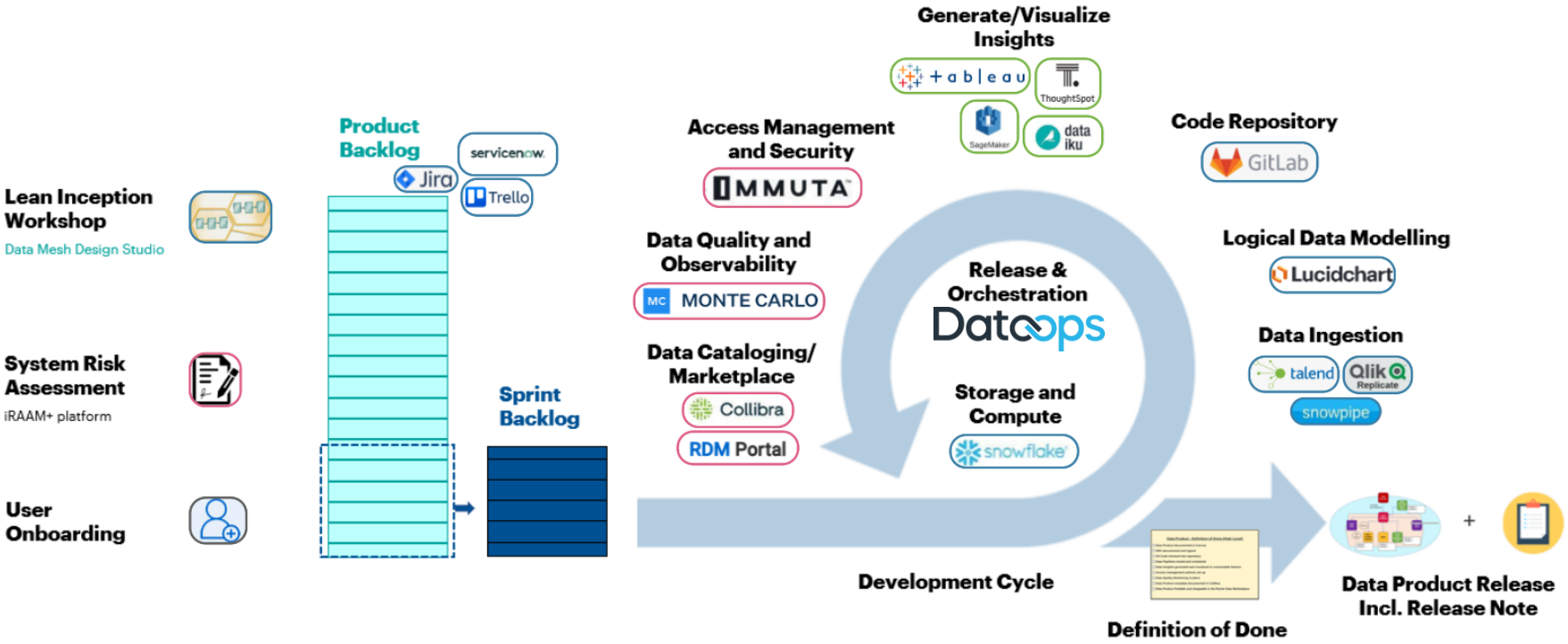


Mesh Supervision Plane
Mesh level experience plane to let end users experience the entire data mesh an data products. Includes capabilities:
Data Lineage, Data Catalog
Data Quality Dashboards, Frameworks,
Data Observability, SLOs
Audit and Access policies

Data Product Developer Experience Plane
Primarily aimed for data product teams, enabling various capabilities they can leverage to create the data products

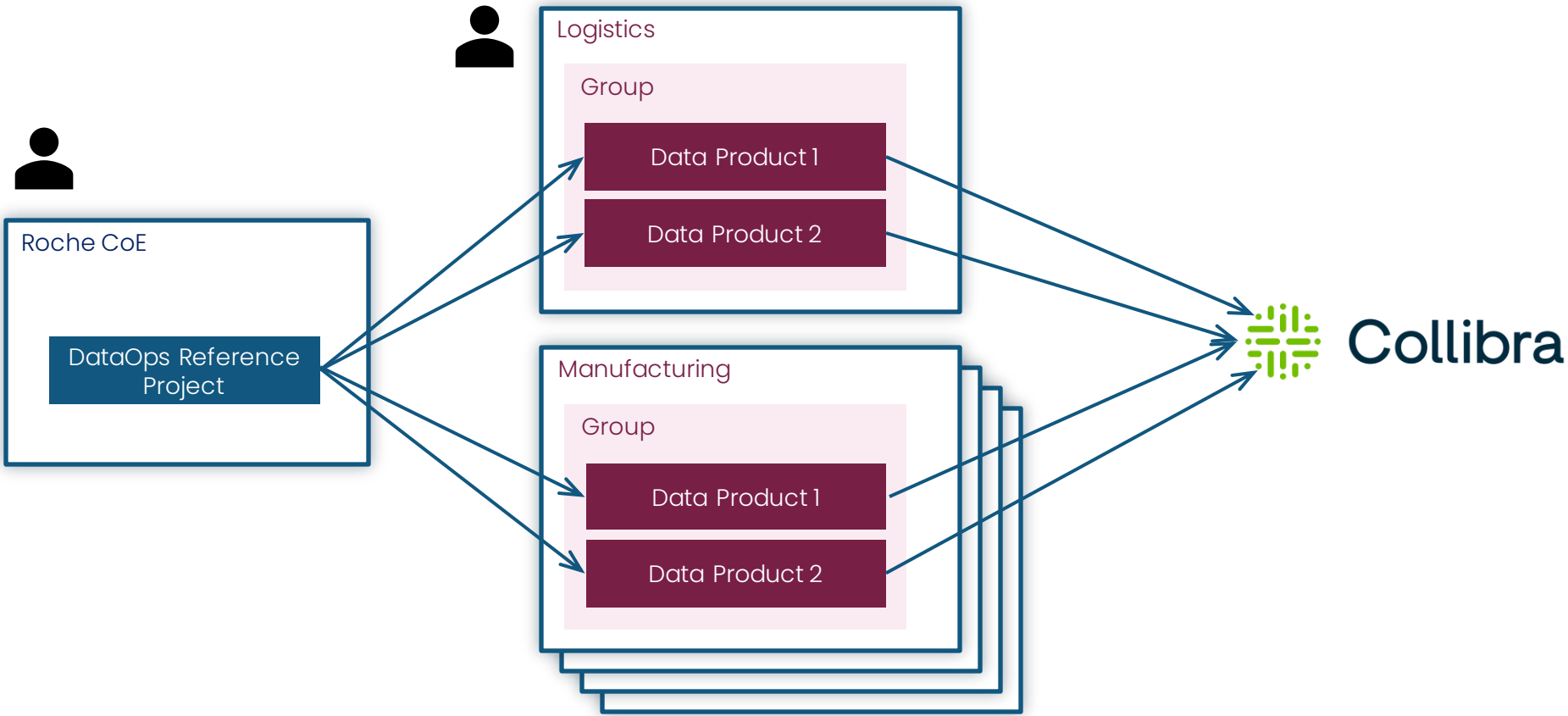
Data Infrastructure Plane
These are the fundamental, underlying capabilities for data & analytics infrastructure.

DataOps moved to the heart of everything



FULL MULTI-DOMAIN DATA MESH IN DATAOPS

REFERENCE PROJECT INHERITANCE



CENTRE OF ENABLEMENT

- **Blueprint includes**
- Data Mesh Design patterns
- Best practices
- Options
- Deployment advice
- Team design
- Training advice
- Etc.

- (Free for all enterprise customers)

Where are we now?

In Summary

Roche

Roche Mesh Experience

What we have achieved to date

2021 Feb Platform Build Started

1300
Users, Devs, Consumers
onboarded since start

180 TB
of data stored as of today
>20% increase over last month

6-8 weeks
average MVP time
Reduced from 6 months

40
Data Product Teams onboarded
Increasing 2 per month

50
Data products created
according to the Data Mesh Principles

120
Releases last month
In comparison to 1 release every 3 months

> 15
additional capabilities & partners integrated
#IMMUTA #DataOps #MonteCarlo #Collibra and many more

BEST PRACTICES AND LESSONS LEARNED

BEST PRACTICES AND LESSONS LEARNED

Team Make Up

Differentiate …..
DataOps Engineer vs Data Engineer vs ‘Casual’ User

Ensure each team has a strong DataOps Engineer (or 1 DataOps to 2 teams)

DataOps Engineer owns the process

Centre of ‘Excellence’

Start roll out when ‘good enough’ – don’t wait

Support with strong documentation

Weekly cadence calls x3 – about 30 minutes’

Enablement breeds Excellence

Product Stack Harmony

Choose Products that naturally work in harmony with each other

Integrations already built – no reinvention.

Best products within their specialist domains

THANK YOU.

Dataops