



ADC - C4

Enabling a Modern Data Architecture on AWS

Mark Hodnett

Senior Data Architect
fourTheorem

Lodewyk Van Der Merwe

IT Lead - Platform Architecture
RenaissanceRe

“RenRe”

30 Year Track
Record

Leading Stand-
Alone Property &
Casualty Reinsurer

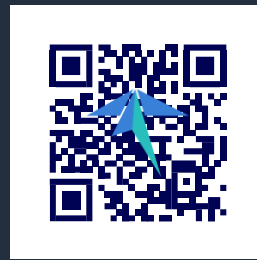
RenaissanceRe

718 Employees
(Feb '23)

11 Global
Offices

\$9.2BN Gross
Written Premium
2022

\$17BN Total
Managed Capital
2022



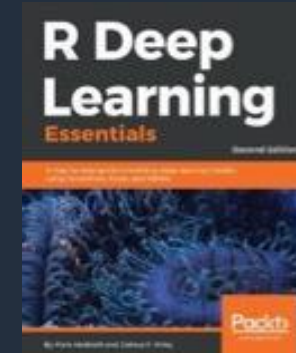
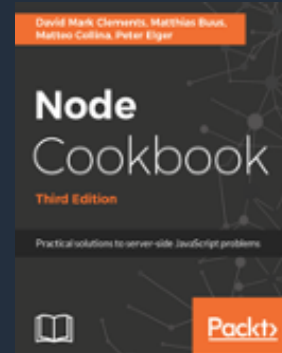
We are a pioneering technology consultancy focused on **AWS** and **serverless**

We can help with:

- Migrations & Application Modernization
- Accelerated High Performance Computing
- Event-Driven Architectures
- Data engineering, Analytics & AI



SLIC Watch



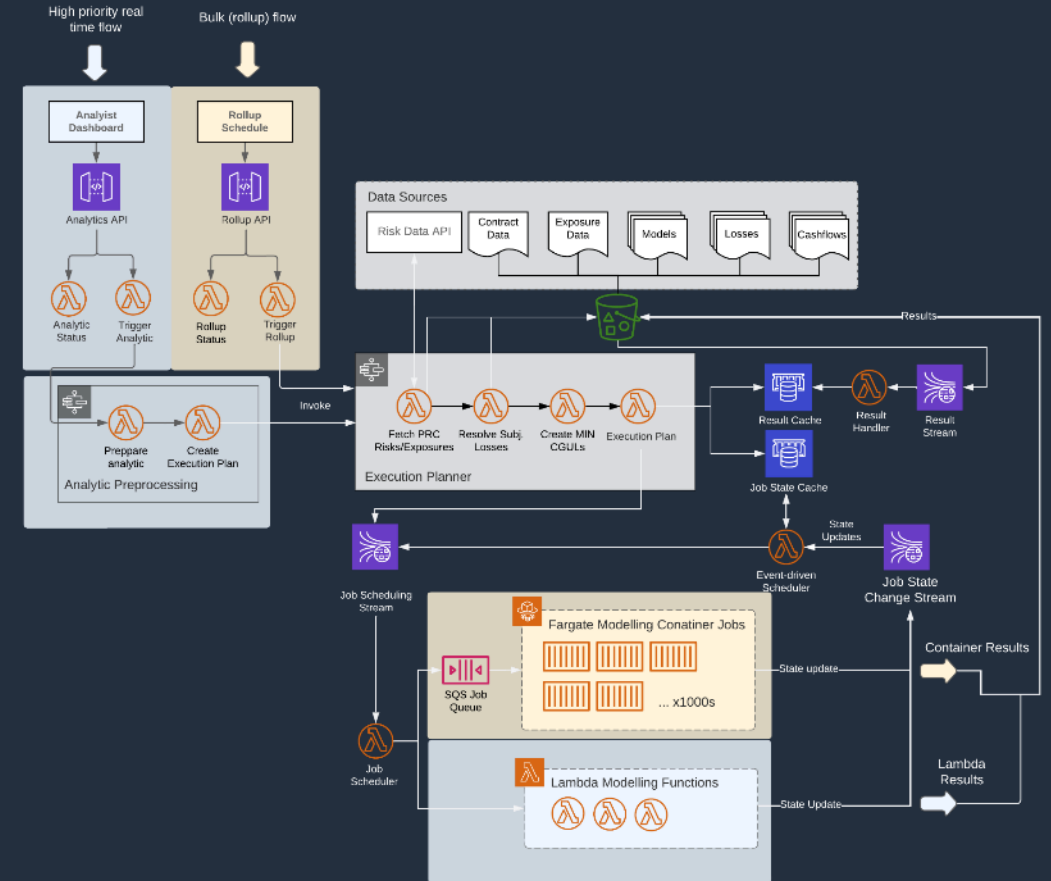
RenRe – Our First Workload

- Risk Rollup
 - Uses financial modeling to understand our portfolio of risk
 - Run internal custom-built risk model on all reinsurance deals
 - HPC (High-Performance Computing) workload
 - ~45TB data processed, ~600GB produced for analytics
 - 10/12-hour runtime
- Goal
 - Re-Imagine for at least 10x performance improvement

RenRe – Our First Workload

Results

- Rollup executing in ~ 1 hour
- Removed constraints on the number of runs
- Now supporting data volumes up to 5x
- Huge business impact
- Carbon footprint reduction



- AWS HPC Blog: [A serverless Architecture for high-performance financial modelling](#)

Our Software Solution Estate

- Built over the past 30 years to support a very successful business
 - Data driven business
- Has some challenges
 - Disparate data and storage technology
 - Inconsistent data models across the organization
 - Inconsistent integration patterns
 - Implicit dependencies at the data layer
 - Inconsistent use of APIs and Message Queues
 - Lots of ETL

Modernization Goals

- Elevate our competitive advantage through technology
- Decentralize and democratize data access
- Support significant business scale-up and enhanced efficiency
- Enable business agility through composability
- Rapid development and deployment of business capabilities

Architectural Frames of Reference

- Adopt Data Mesh principles
 - A focus on data as a key value generator
- Adopt event driven, evolutionary architecture
 - The system must be agile and flex with the business
- Domain Driven Design at the macro level
 - Enable devolved responsibility through bounded contexts

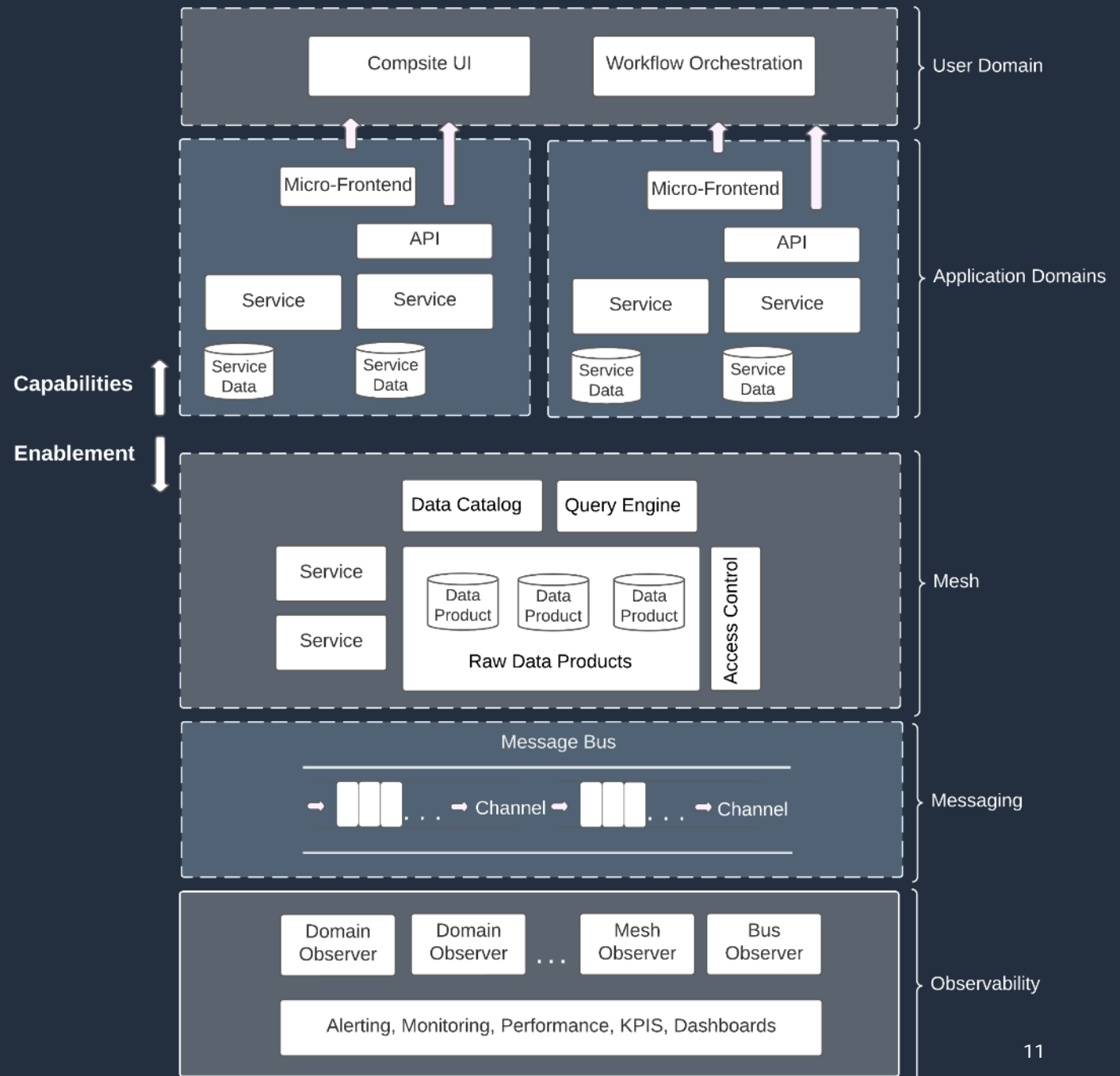
Iris

- Iris is the project name for the development platform within RenRe
- Ancient Greek goddess.
- Iris is associated with communication, messages, the rainbow, and new endeavours.

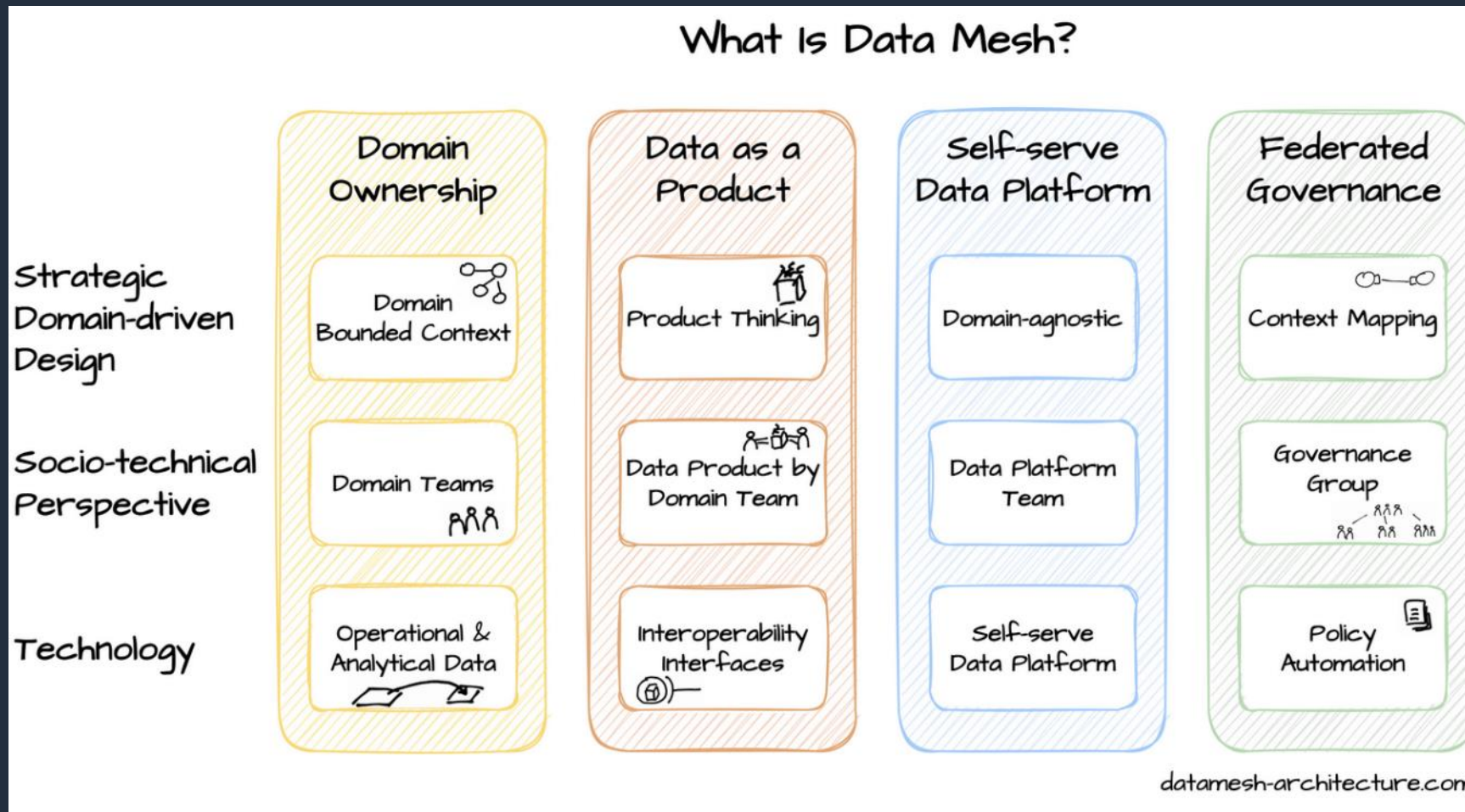


Conceptual Overview

- Serverless-first platform
- Decentralized domains
 - Business aligned
 - Bounded context
 - Autonomous teams
- Enabling layers
 - Data mesh
 - Message / Event Bus
 - Observability

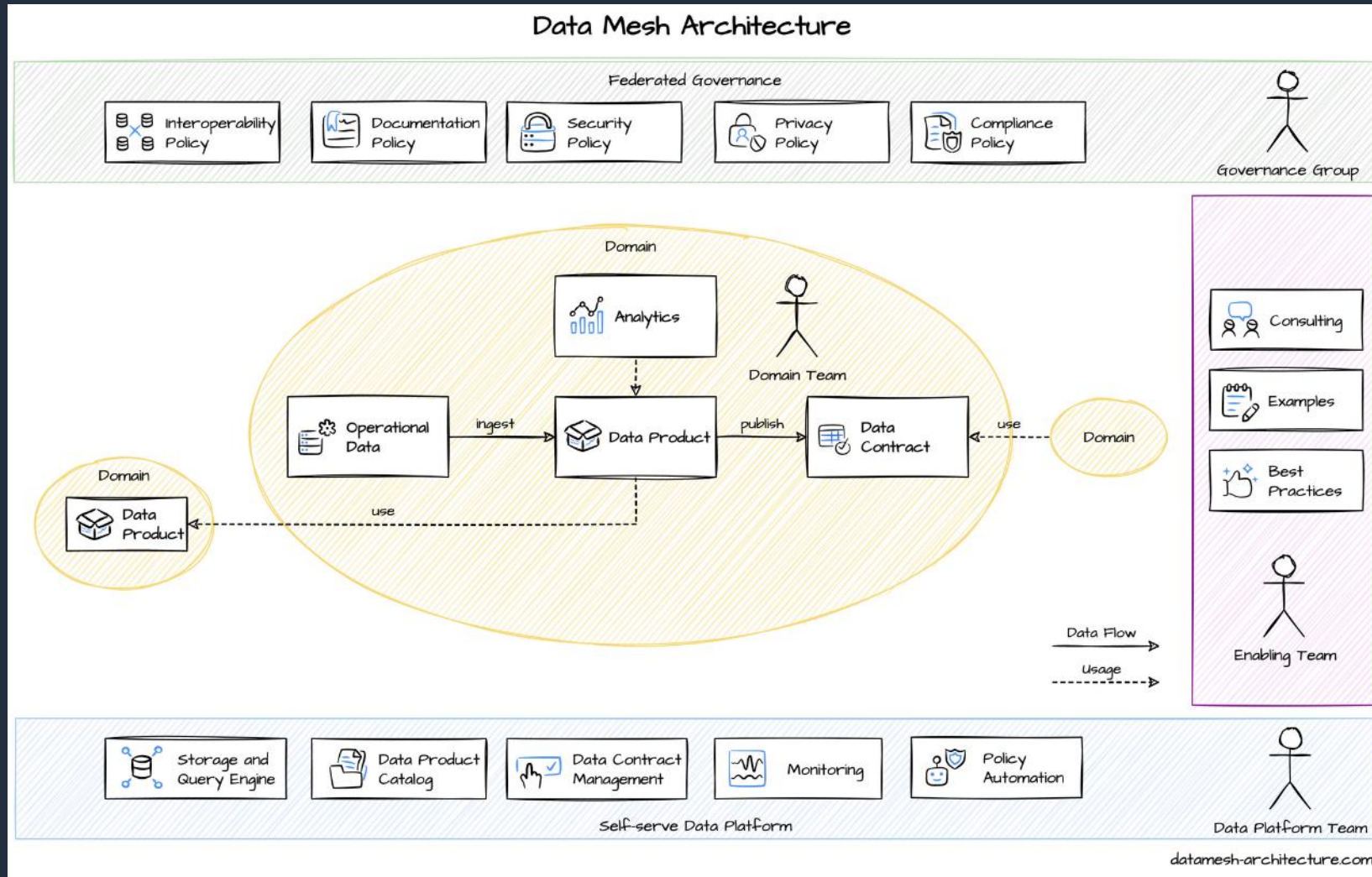


Mesh Principles



Source: <https://www.datamesh-architecture.com>

Canonical Architecture



Source: <https://www.datamesh-architecture.com>



Data Consumers

The different data consumption patterns in RenRe

- Machine-to-Machine
- Power users that access prod systems using SQL / notebooks
- Underwriters / business analysts – typically use derived data
- Report builders and consumers

How do we solve for these use cases on our development platform?

Data Contracts

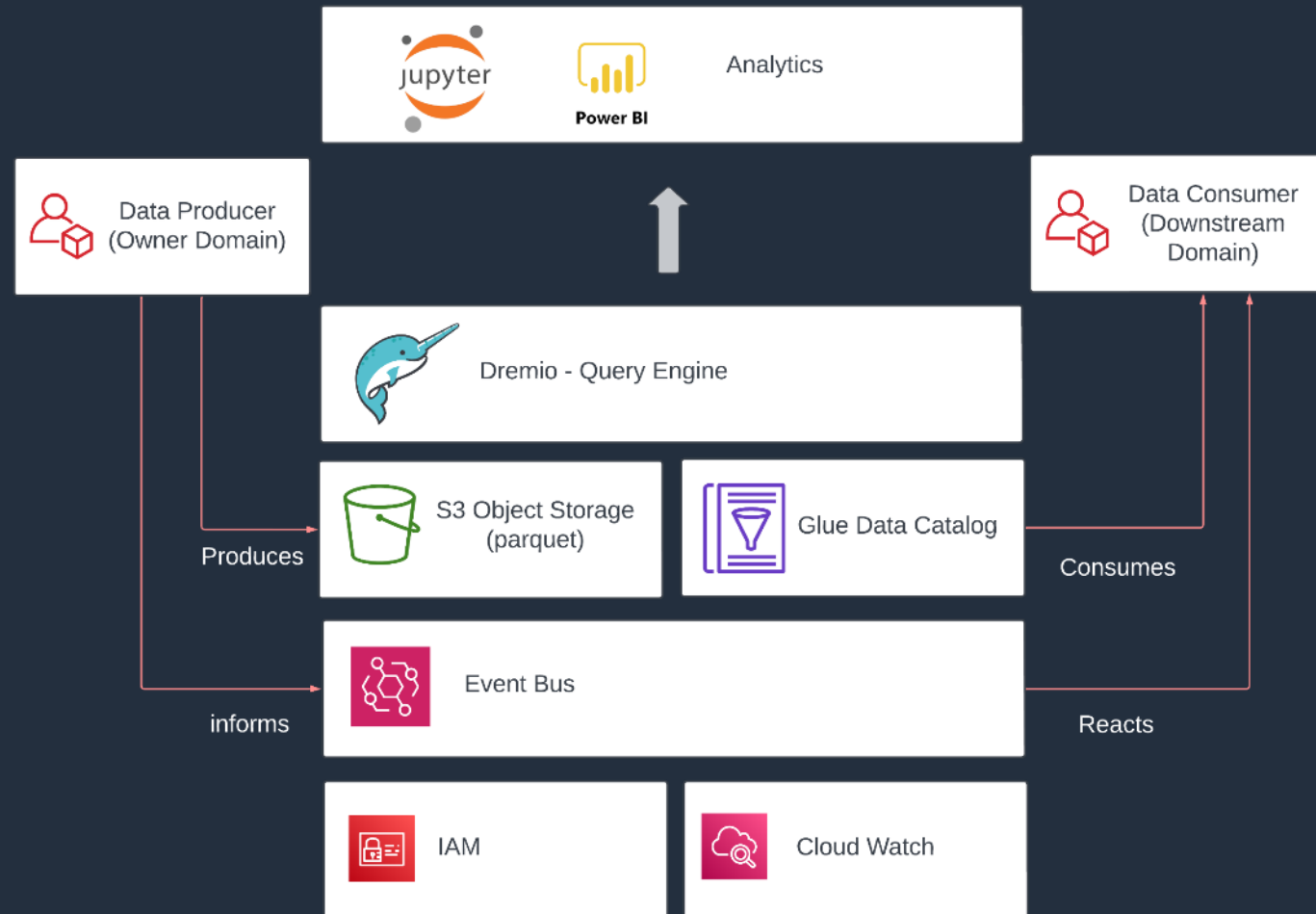
Similar to an API contract

- Schema versioning, documentation, etc
- Data Catalog – relational model
- Teams publishing products have a responsibility to honor the contract
- Contracts can evolve with new schema versions
- Use of data quality tools (e.g. Soda) to enforce constraints on data

```
checks for customer:  
  - schema:  
    fail:  
      when required column missing:  
        - name  
  - invalid_count(source) = 0:  
    valid values: ['online', 'retail']  
  - min(probability_churn) >= 0  
  - max(probability_churn) < 1.0
```

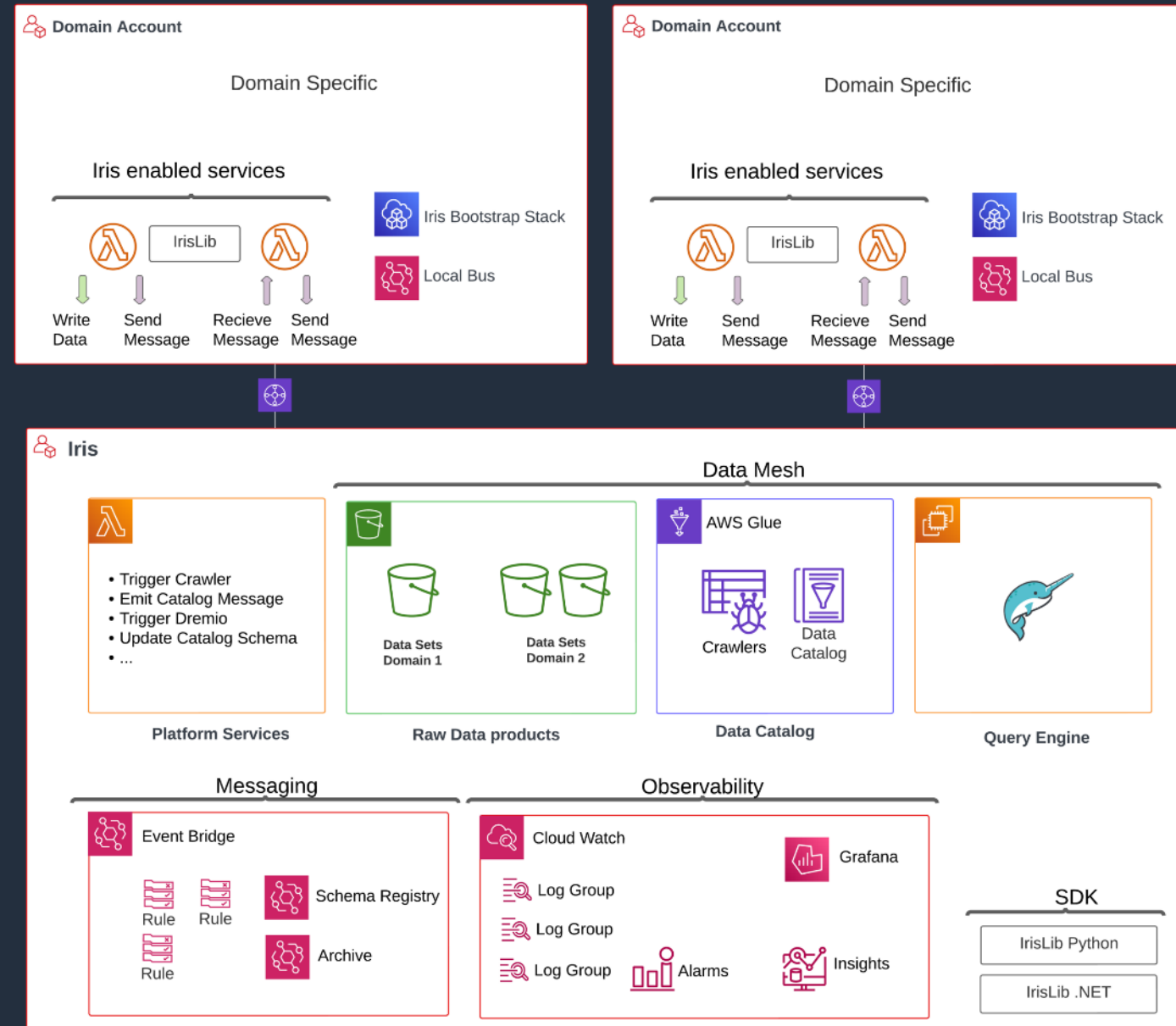
Mesh Technology Stack

- **Storage:** S3 (Parquet)
- **Catalog:** AWS Glue
- **Analytics engine:** Dremio
- **Access:** IAM/SSO
- **Observability:** CloudWatch



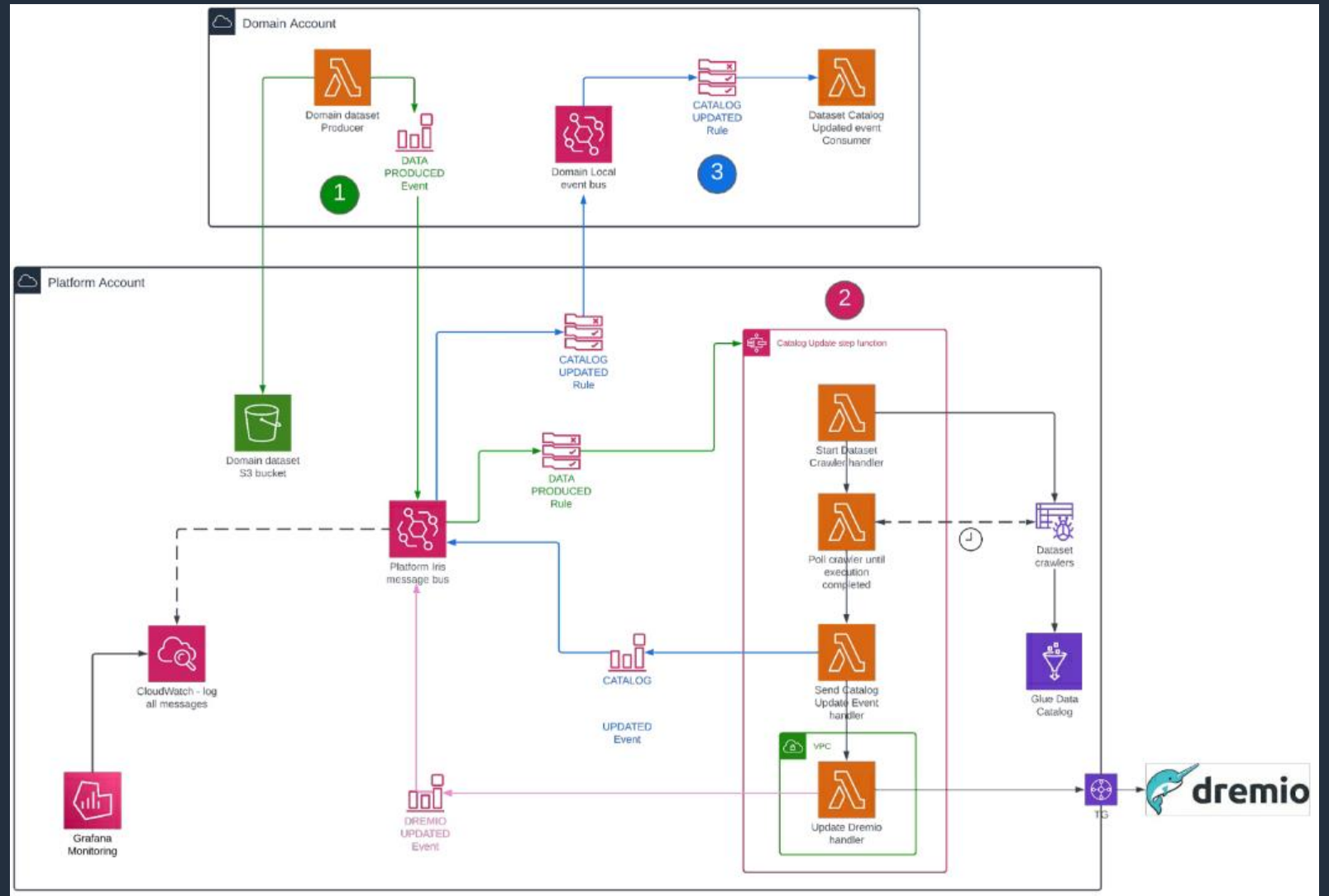
Realization on AWS

- Domains map to AWS accounts
- Iris provides enabling layers
 - Data Mesh
 - Common messaging
 - Common observability
 - Shared SDK



Publishing a Data Set

1. Producer creates a dataset and sends an event notification
2. Iris responds to the event and updates the Glue catalog (Crawler or schema API)
3. Iris emits event to announce data is ready to be consumed



Consuming a Data set

Dremio

- Dremio is an open data lakehouse, providing self-service analytics through an SQL interface.
- This is intended for reporting and analytical workloads for human users.
- A "reflection" is an optimized materialization of source data or a query, similar to a materialized view.
- Data sets exist in layers and can be built using other data sets (e.g. raw/business/reporting)

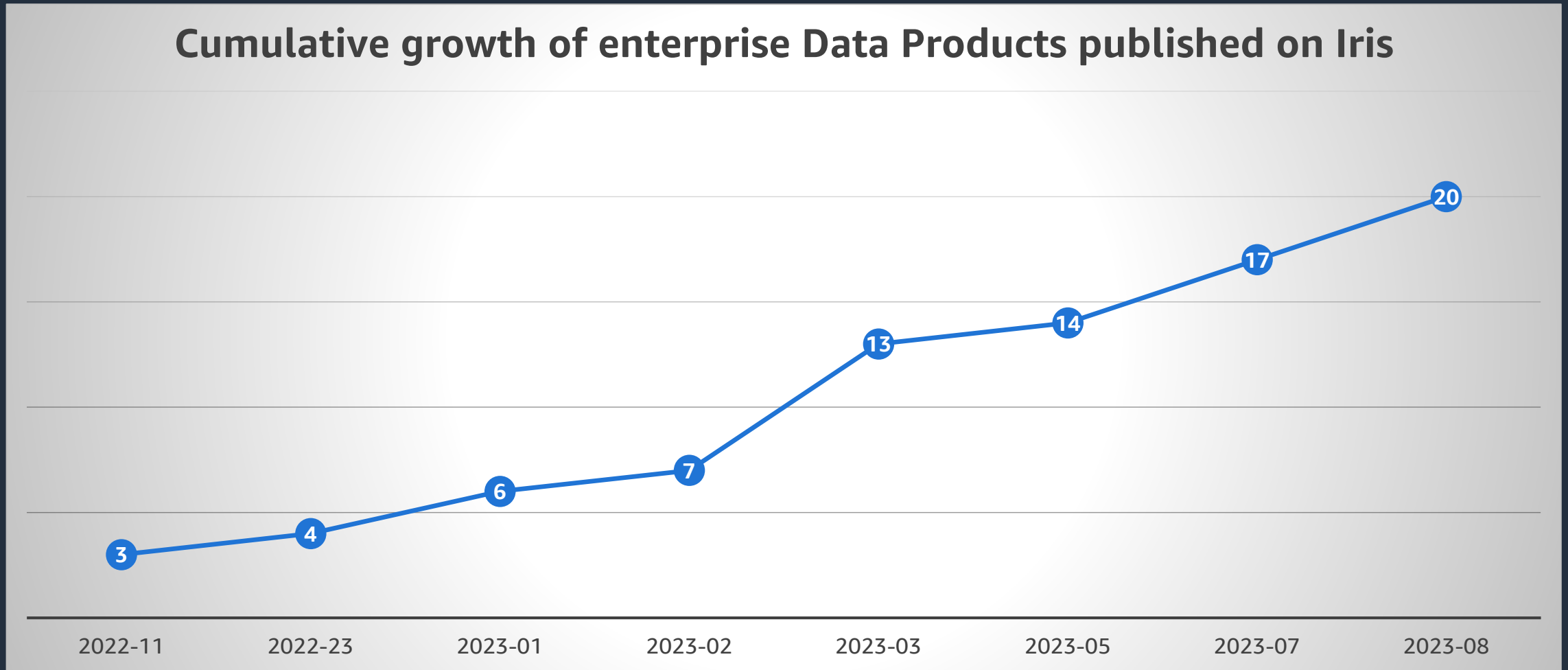
AWS data services

- Glue Data Catalog: Python based services can use the awswrangler library to interact with the catalog
- S3: Direct system to system access

Technical details

- Data Lifecycle and events
 - Data Producers are responsible for the life-cycle of data sets (included as an SLA)
- Permissions Model
 - Machine to machine access is governed by IAM permission policy
 - Data consumers via Jupyter notebooks / Dremio or Power B.I. through A.D. + SSO
- Best practices multi-account setup
 - Security OU: Logs & audit, Guard duty, Macie...
 - Services OU: Networking, Secrets manager...
 - Production OU: Domain accounts
 - Disaster Recovery (warm pilot)

Adoption



A Data Mesh Approach

- Perfect is the enemy of good. Start small and build
- Flywheel effect – as more data sets are published, the demand from consumers grows
- Governance is key - *Data Product Portfolio Management (DPPM)**
 - Aim for Bonsai not Kudzu or Desert
 - Protect value
 - Manage Risk
 - Guide Evolution

<https://aws.amazon.com/blogs/big-data/the-art-and-science-of-data-product-portfolio-management/>



Wrap Up

- At RenRe our Iris development platform is the architectural foundation of our modernization journey
- It facilitates our dev teams to more easily transition to a data mesh approach – a core part of our broader data strategy
- Data mesh enables us
 - replace legacy data patterns
 - democratize access to data across the enterprise
 - allow data producers to create a single governed version of their data for consumption
 - allow data consumers to consume the data products using their preferred tools