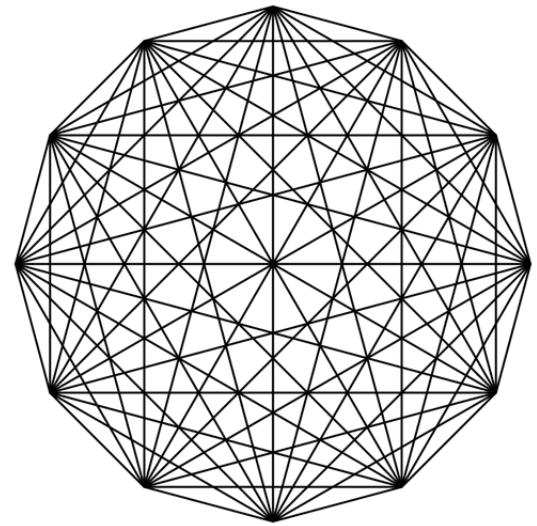




What is a Data Mesh for god sake !

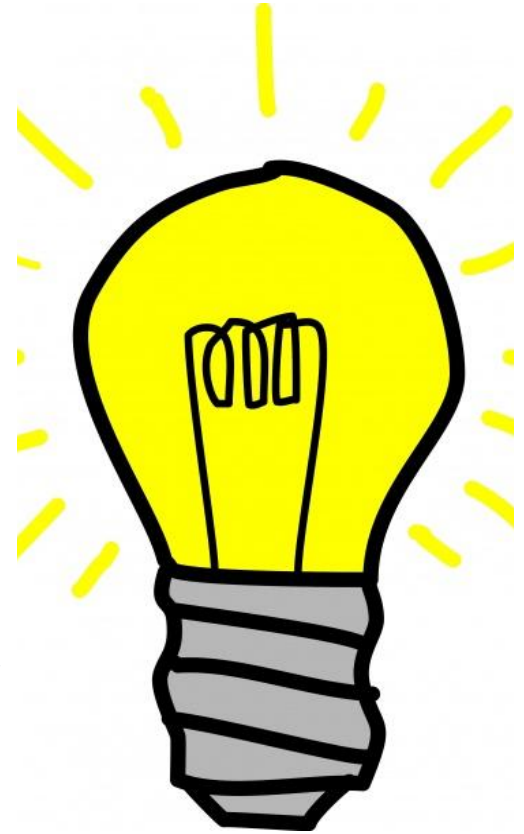
— Seriously...



00 A guideline — Reminder

"There is no sense in talking about the solution before we agree on the problem, and no sense talking about the implementation steps before we agree on the solution."

Efrat Goldratt-Ashlag



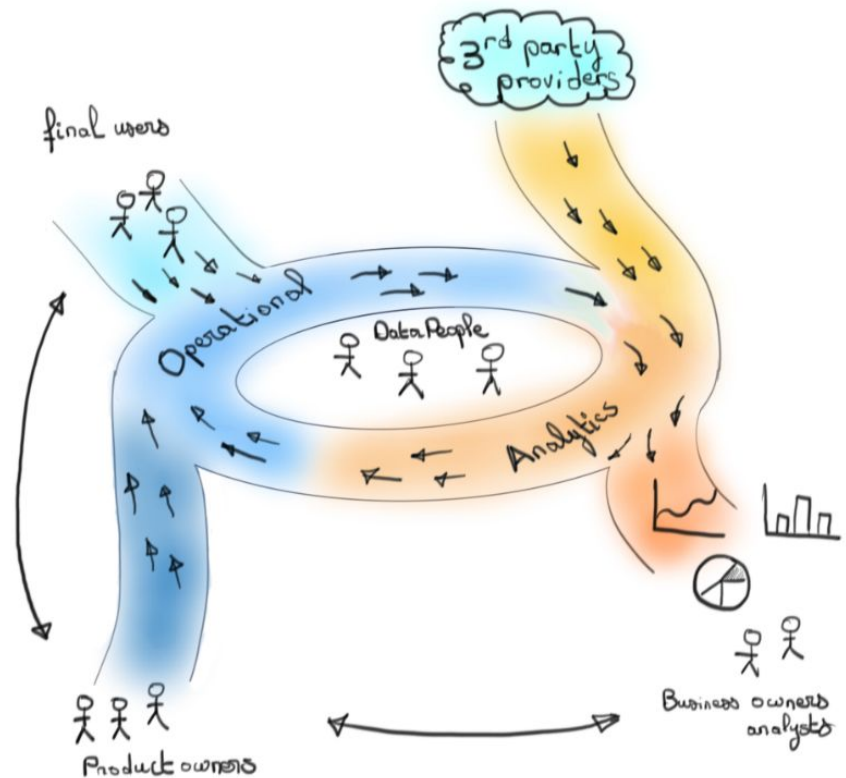
00

Prologue



Become and be data driven

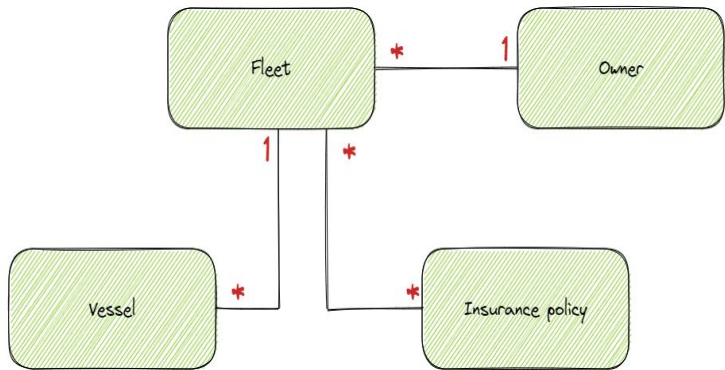
— Data supremacy



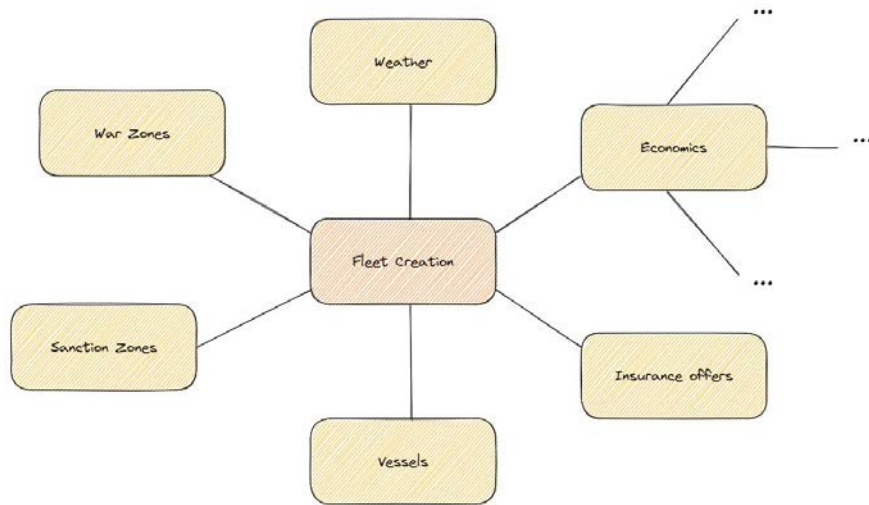


Reconcile to explorer

— Operational VS analytics



&



01

A problem

01

ETL centric

— Traditional approach

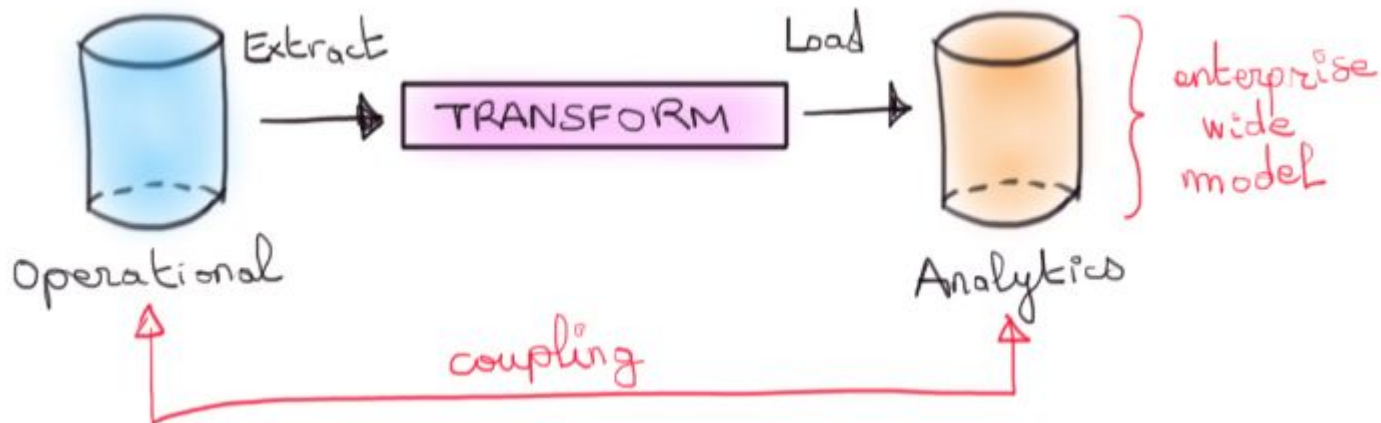


01 Transformation pipeline de transformation

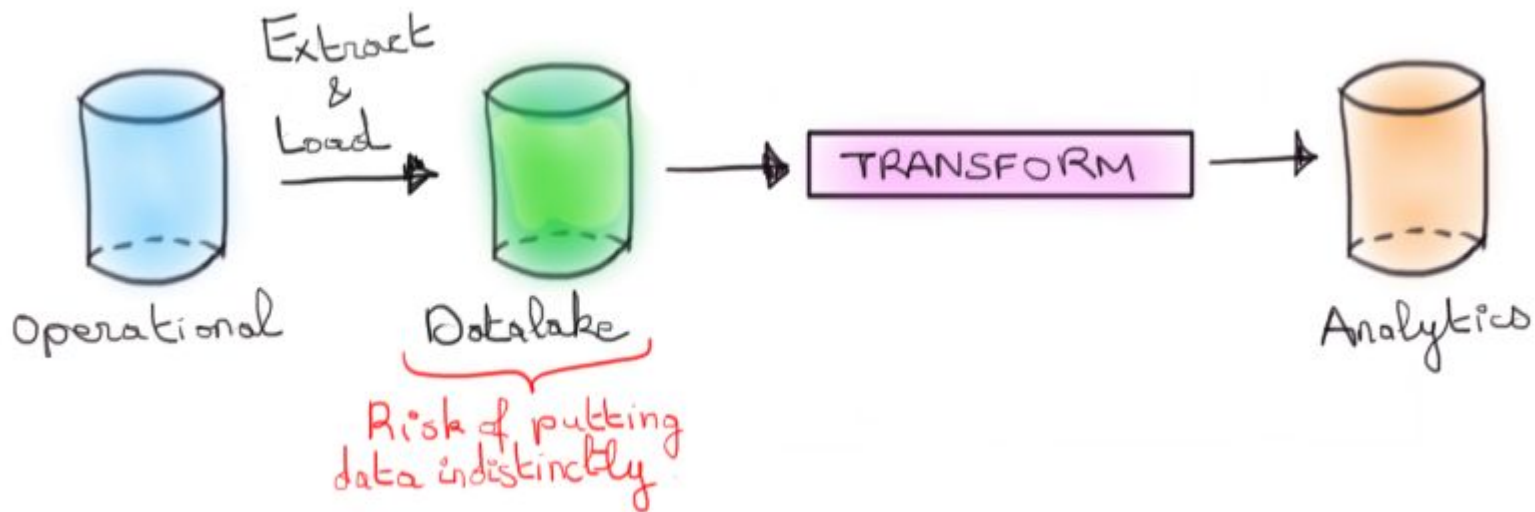
— ETL

- Extract data from all enterprise operational system
- Transform data into analytical model
- Load the resultant data into data-analysis oriented databases

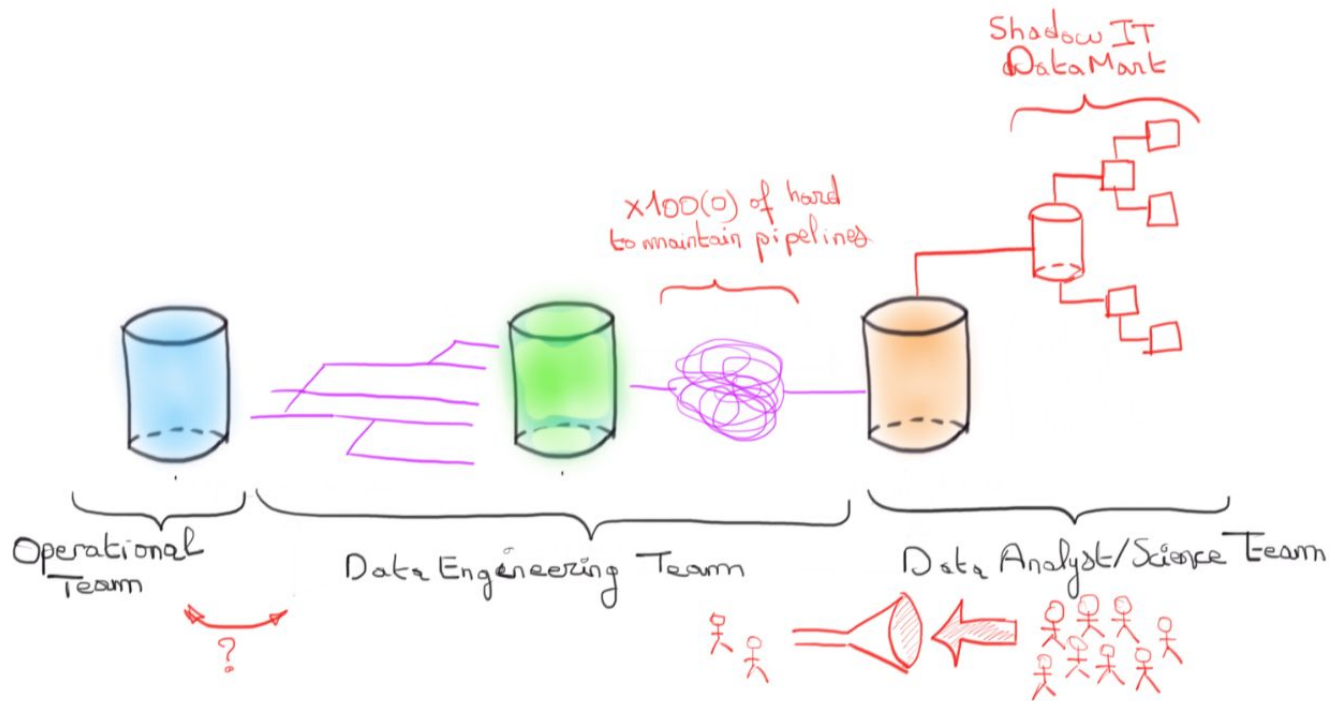
01 High volumetry analytics datastore — Data Warehouse



01 Une staging area pour le DWH — Data Lake



01 Conways'law illustration — Let's sum up



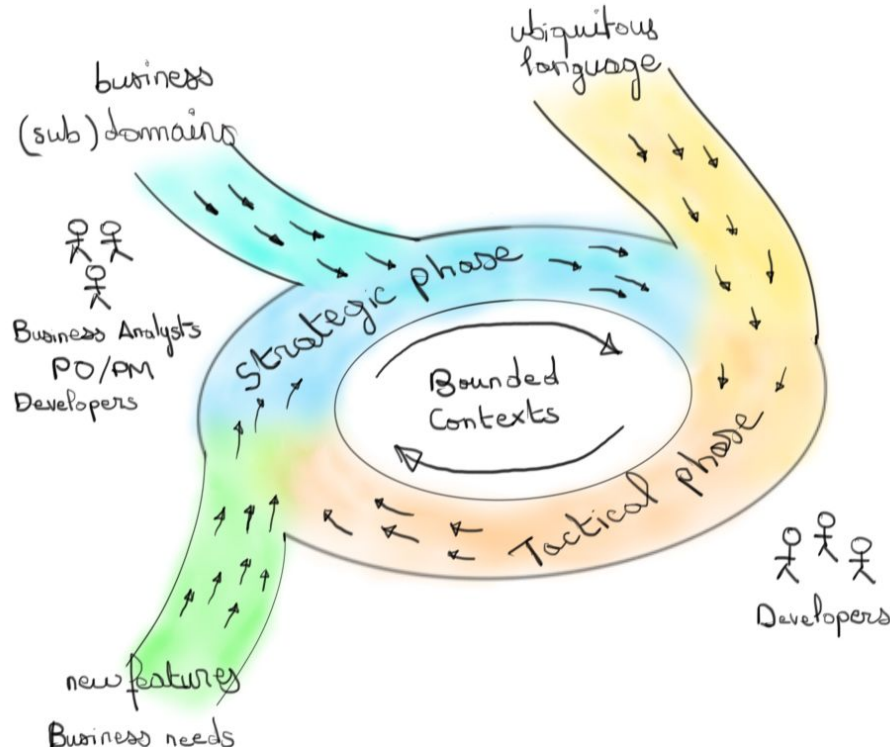
02

A solution

02

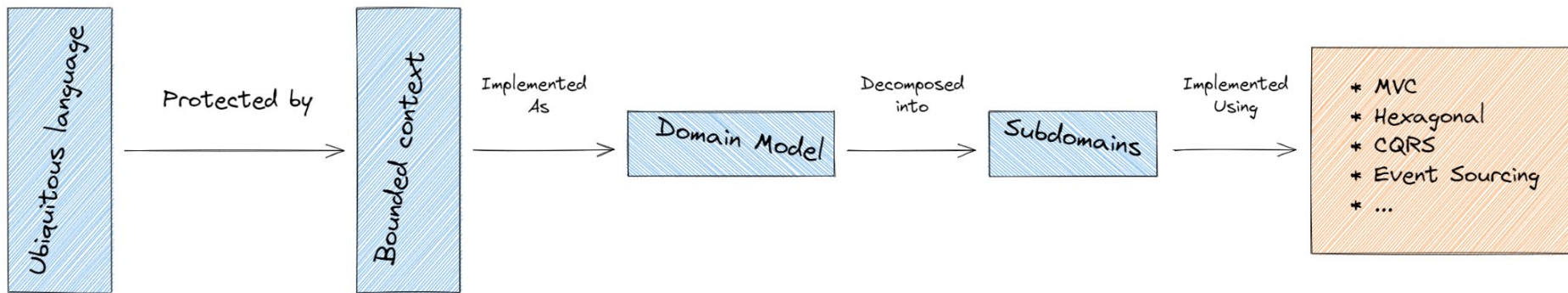
Let's focus a moment on software design

— Reminders about Domain Driven Design

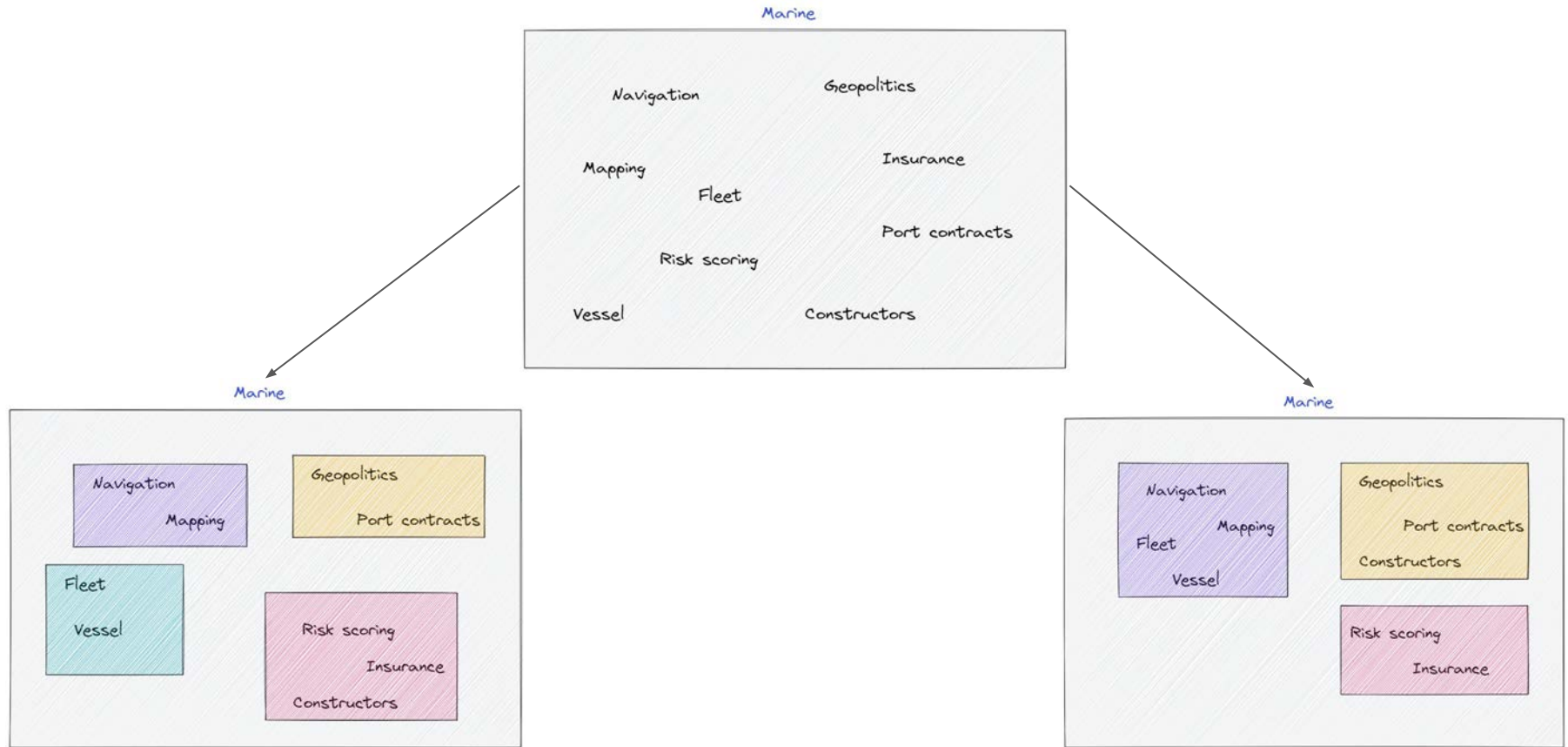


02 — A flow

A strategic phase followed by a tactical technical implementation

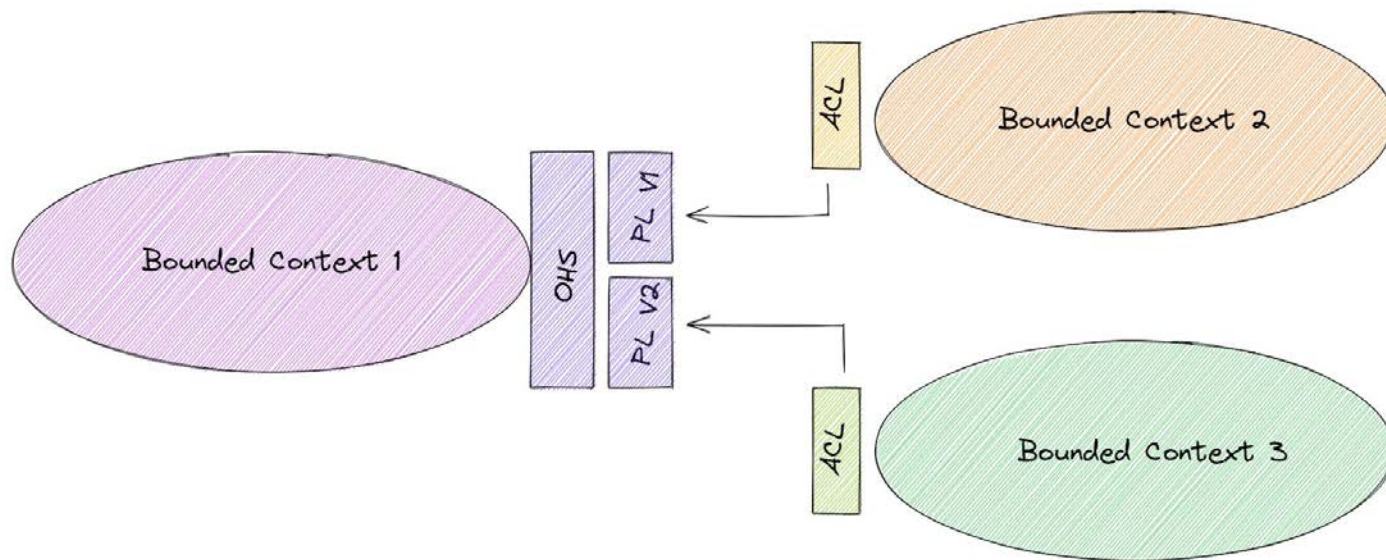


02 No unique gathering plan — Bounded contexts design example



02 Bounded contexts meant to communicate with one another

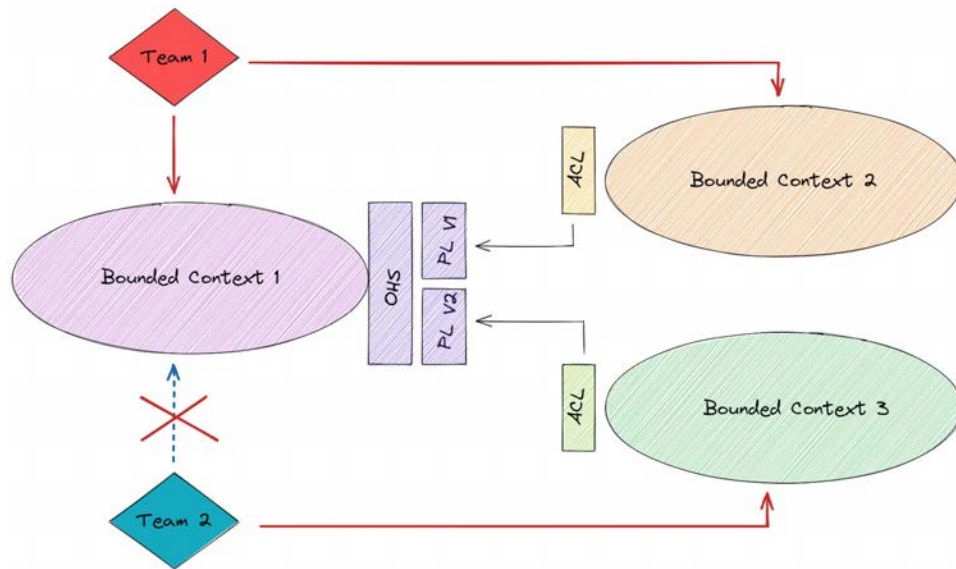
— Communications driven by contracts



- * **OHS**: Open Host Service
- ** **PL**: Published Language
- *** **ACL**: Anti Corruption Layer

02 Teams and bounded contexts, a one to many relationship

— Organisationally scalable



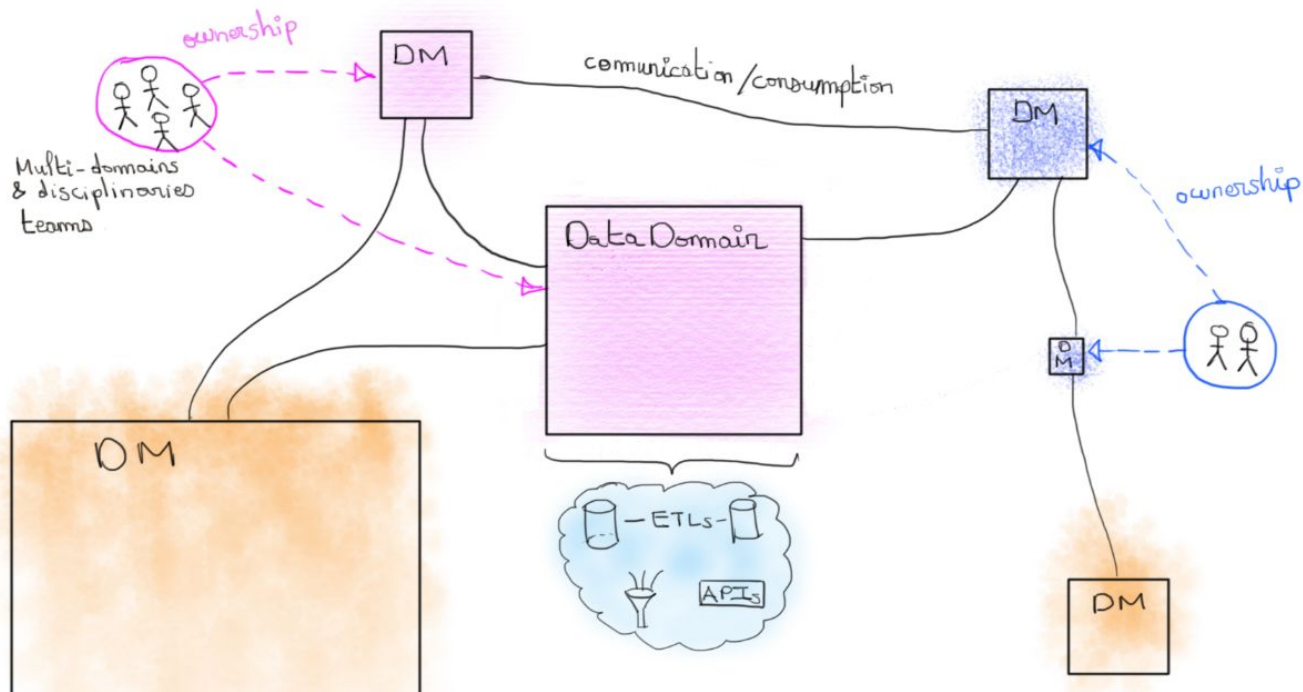
- Each team is responsible for its bounded contexts
- Is accountable for SLA and SLO
- Set metadata in a catalog to allow discoverability and consumption

03

Data Mesh

03 DDD applied to data analytics

— Main idea



03 Foundations of the Data Mesh

— The 4 pillars



03 Misconception to avoid

— Considerations

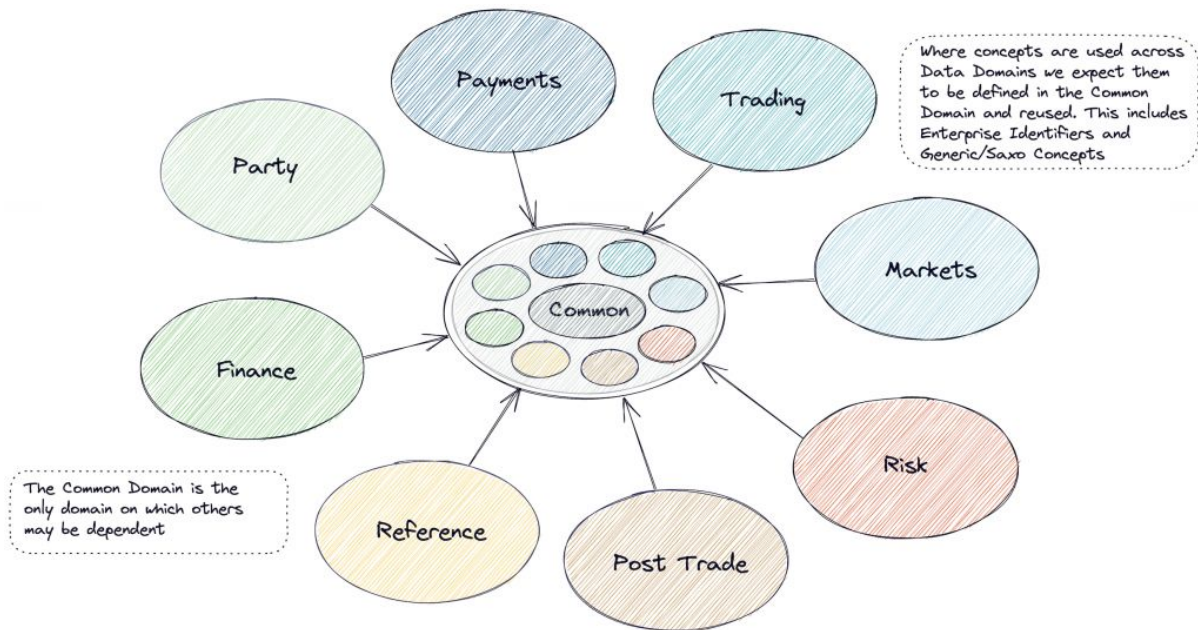
- Unlike Service Mesh, Data Mesh is not a purely technical concept
 - It's a socio-technical concept which solve an organisational scalability issue
- Beware of *Data Mesh ready* solutions
 - It however exists what we call *enablers*
- Data Mesh is a path for better collaboration between teams and data exploitations. It is not an end
 - A simple ETL centric approach if perfectly fine if your business domain is simple enough

04

Implementations

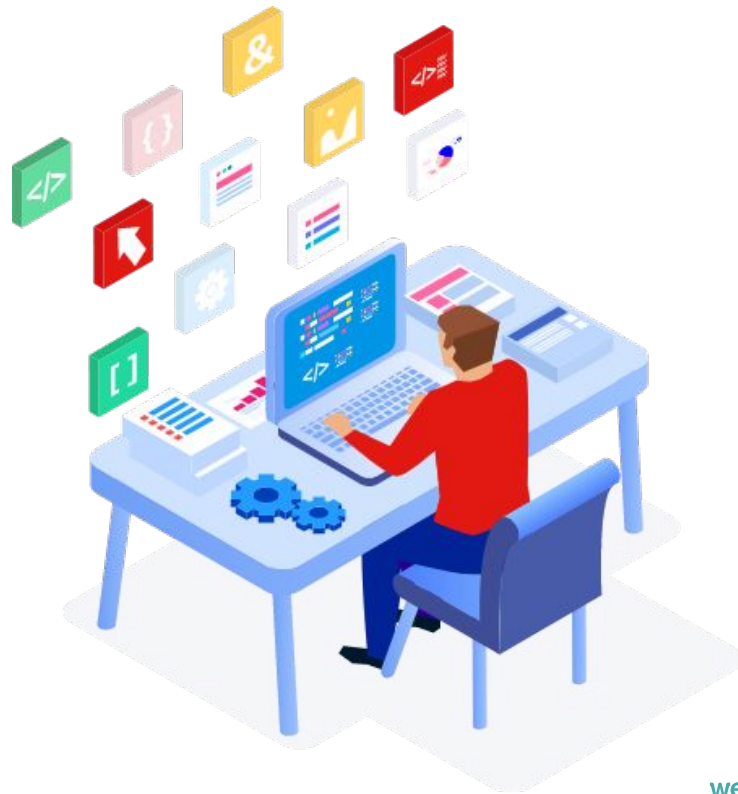
04 Implementations — The catalog

- Allow each domain to push their data as products into a common platform
- Subscribers can subscribe through this common platform, as a service
- This is where we define product definition like refresh frequency or transformation that should occur without asking the data team



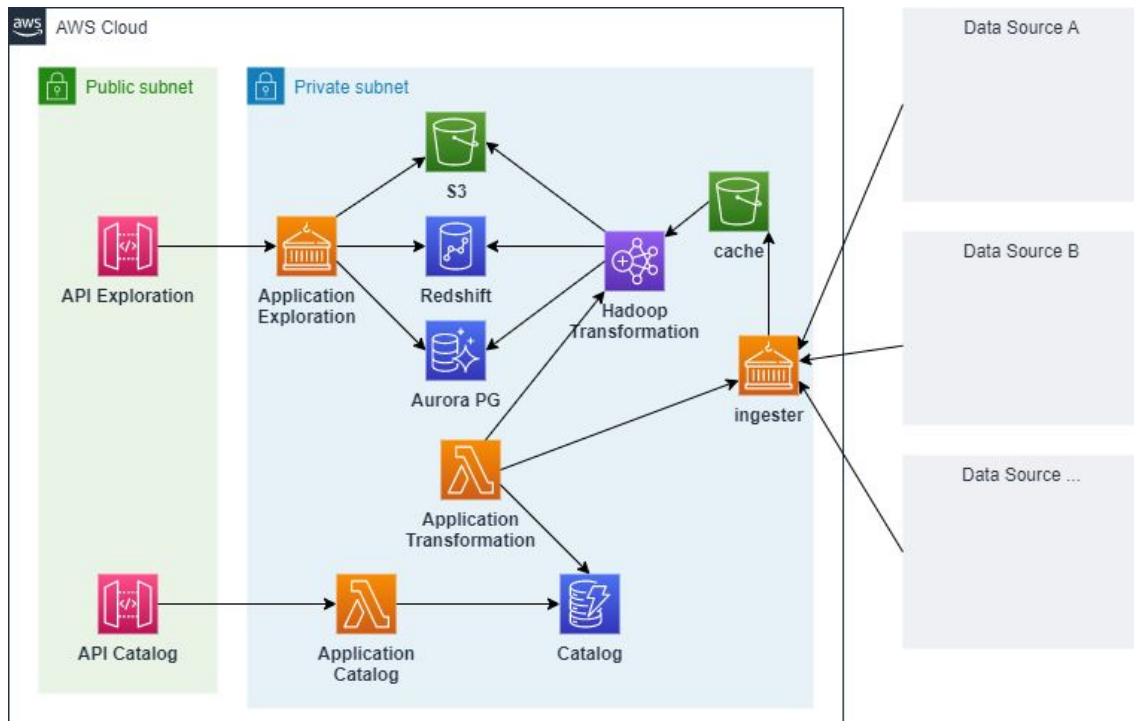
04 Implementations — The orchestrator

- Driving ETL operations
- Monitor all ETL stages
- Operate with the catalog
- provide an administration panel to users



04 Implementations — Data Domain - on AWS

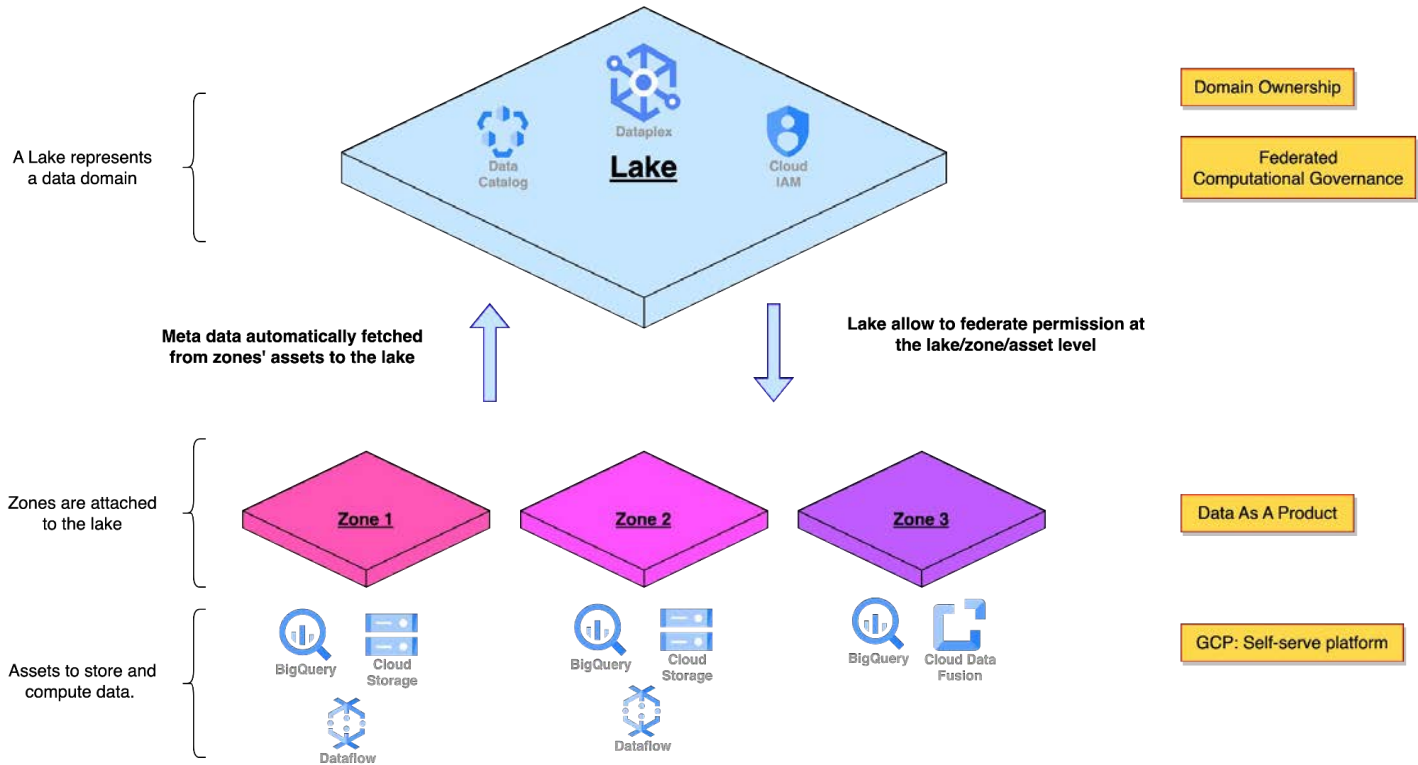
- Catalog and orchestrator based on serverless framework (lambda + DynamoDB)
- Exploration and pulling based containers and DynamoDB
- Transformation backed by EMR
- Storage on Redshift, aurora PG and S3



04 To conclude — GCP Dataplex

- A logical layer on already existing services : IAM, Data Catalog, BigQuery, Cloud Storage...
 - Is to the data what is the project for billing and API quota
 - Allow to quickly iterate on the Data Mesh pillars in a GCP environment

04 The vision — GCP Dataplex



**Thank you for your
attention.**

See you on the chat !

