

**BIG
DATA
DAYS
2023**



Silos to data-driven culture

A modern data strategy model

John Mousa - Sr. Solutions Architect – Amazon Web Services



/johnmousa



/in/johnmousa



**ENTERPRISE
BIG DATA
FRAMEWORK®**

Data strategy is crucial today

99%

Businesses
want to be data
driven



Successful Data Strategy

59.5%

Businesses
have some
success



8.1%

Technology
limitation



People or Technology

91.9%

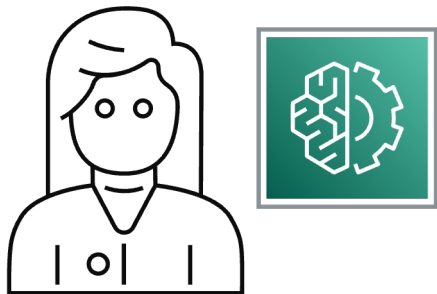
Cultural Factors,
People, or
Process



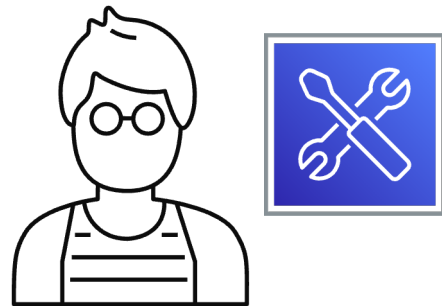
Meet the team



Buzz
from business

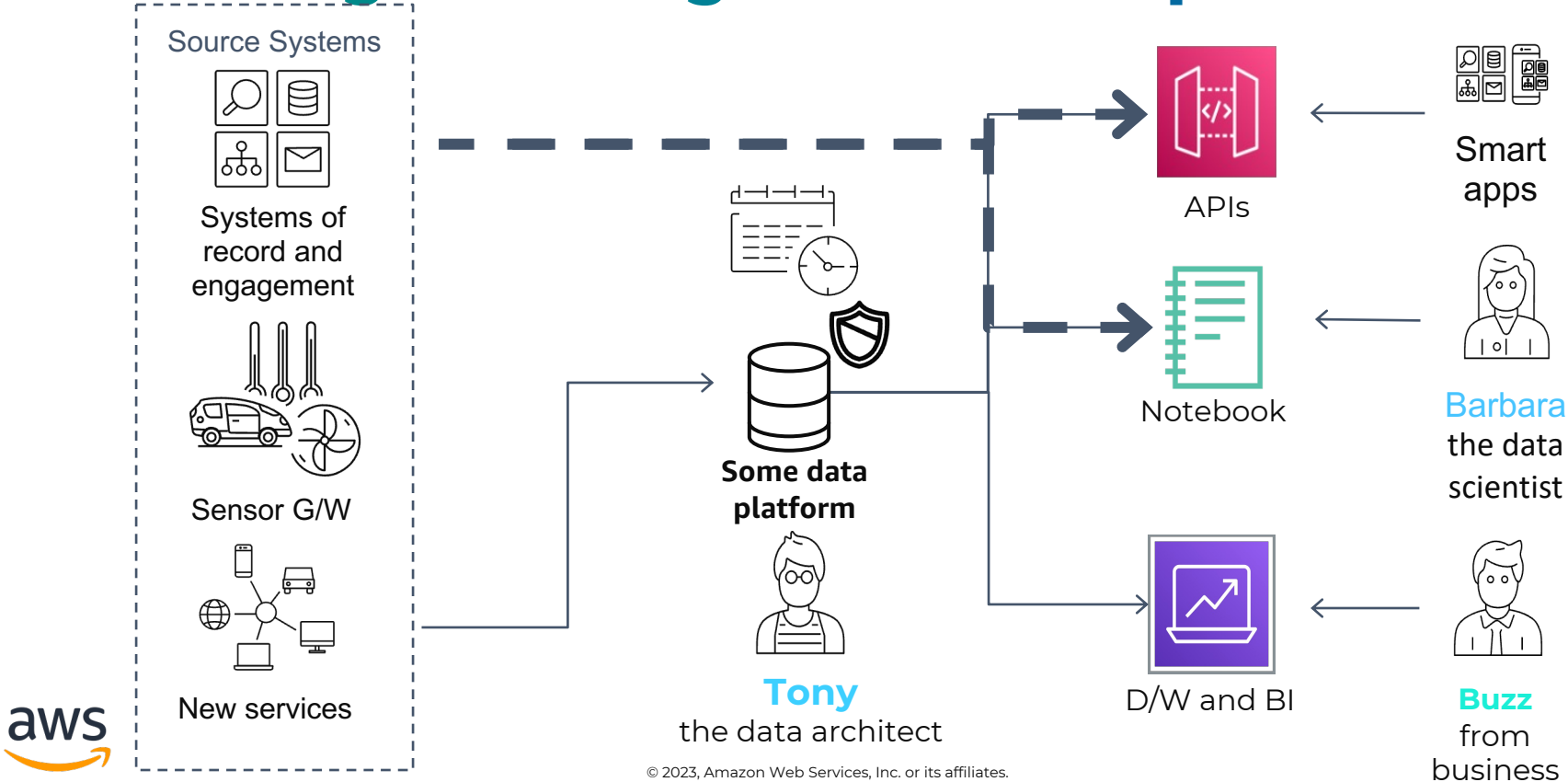


Barbara
the data scientist

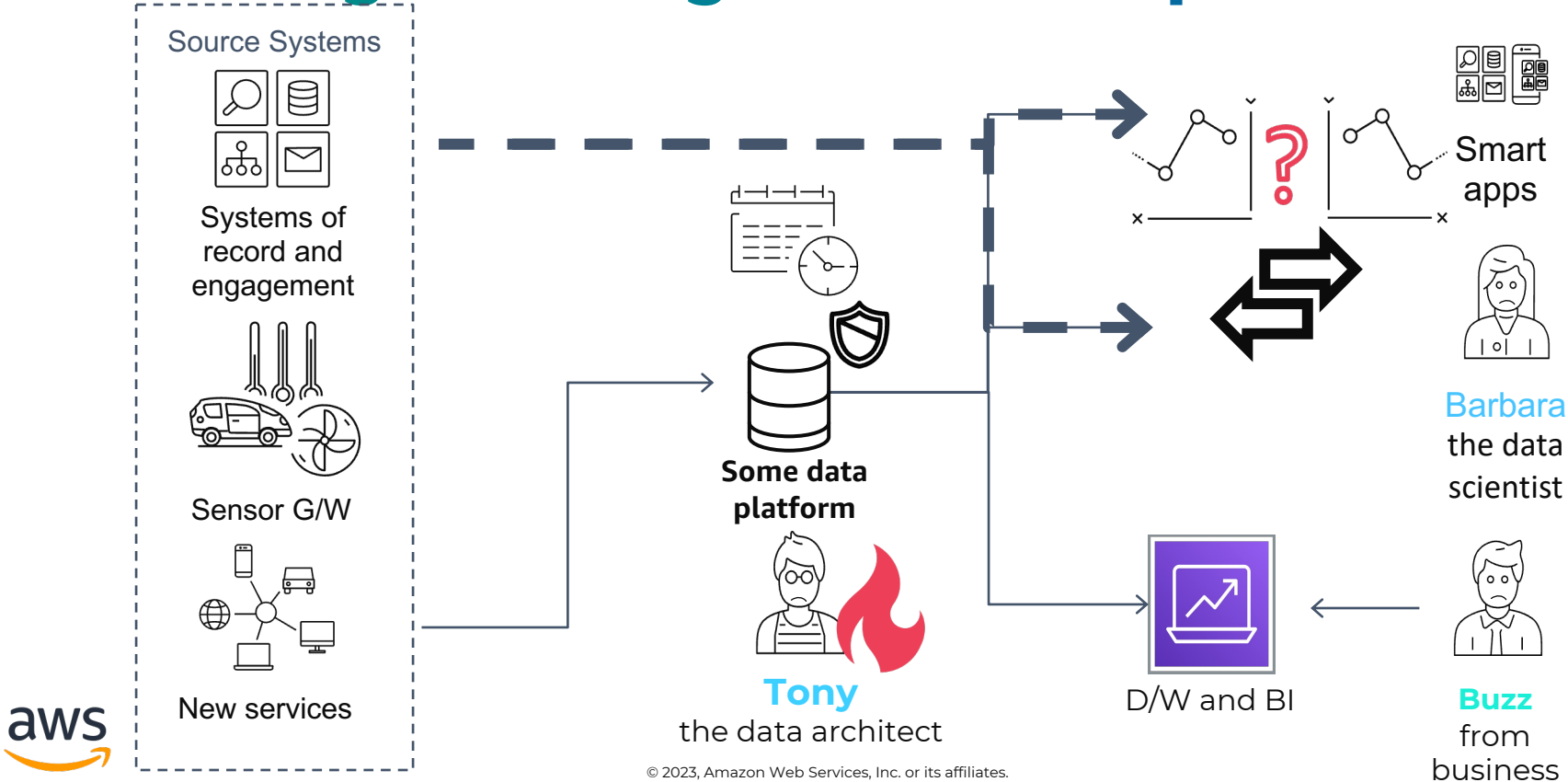


Tony
the data architect

Challenges working with the data platform



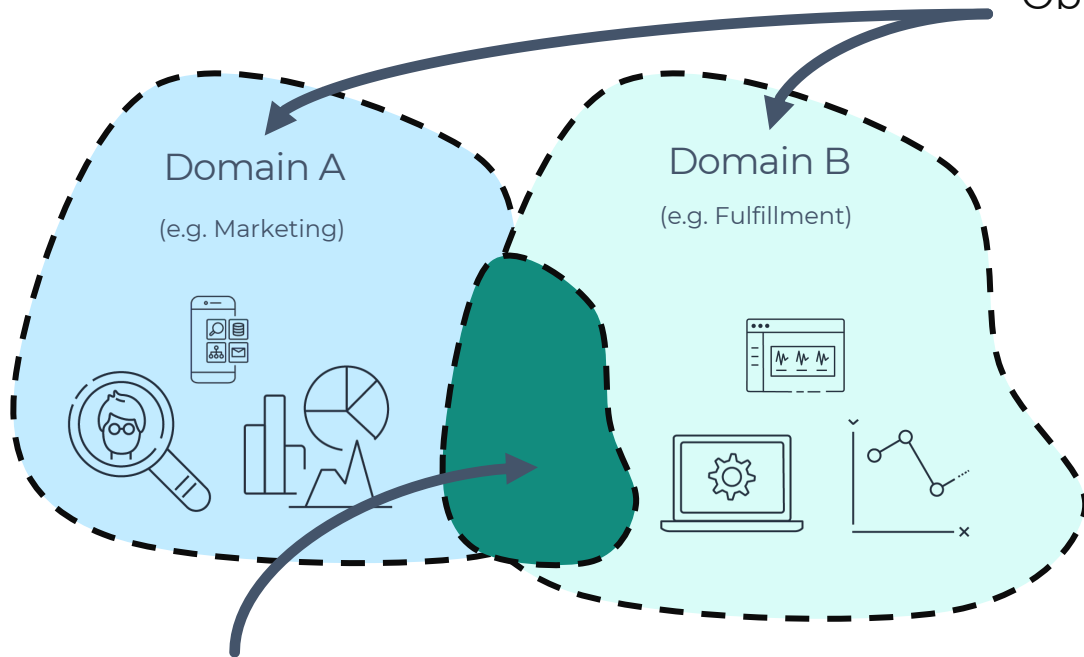
Challenges working with the data platform



Data integration value for business scalability

DATA-DRIVEN AND BEYOND

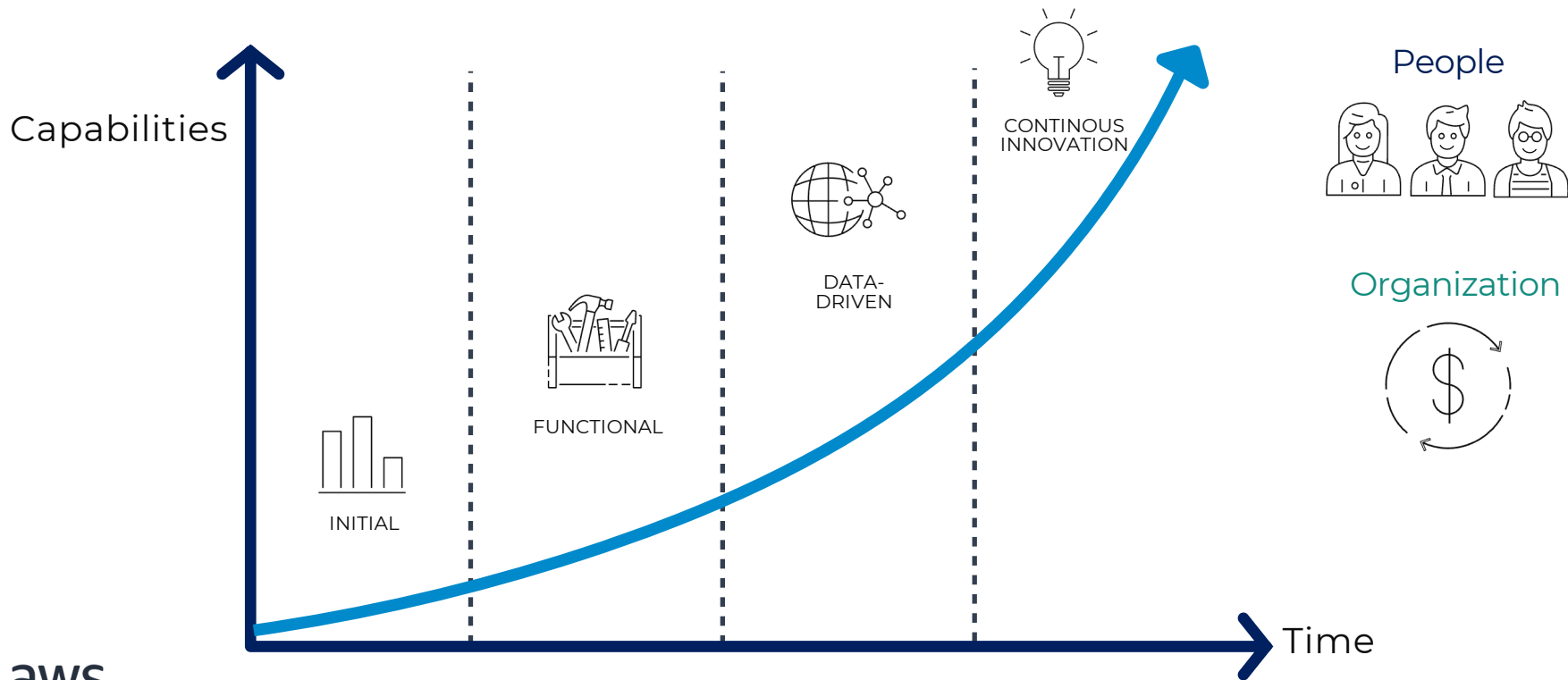
Obvious insights value



Differentiating integration value

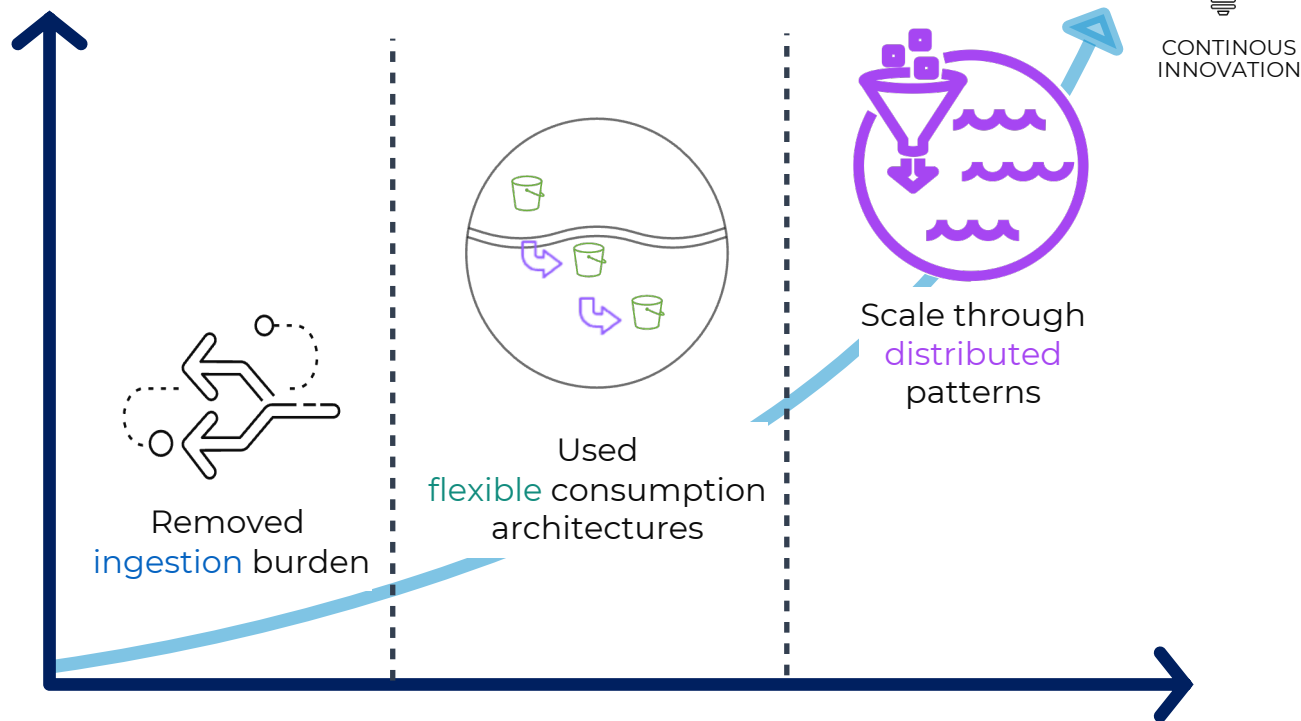


Advancing analytical capabilities model



Evolutionary architecture

MEET YOUR ORGANIZATION IN THE RIGHT PLACE IN ITS JOURNEY



Full recap can be found here

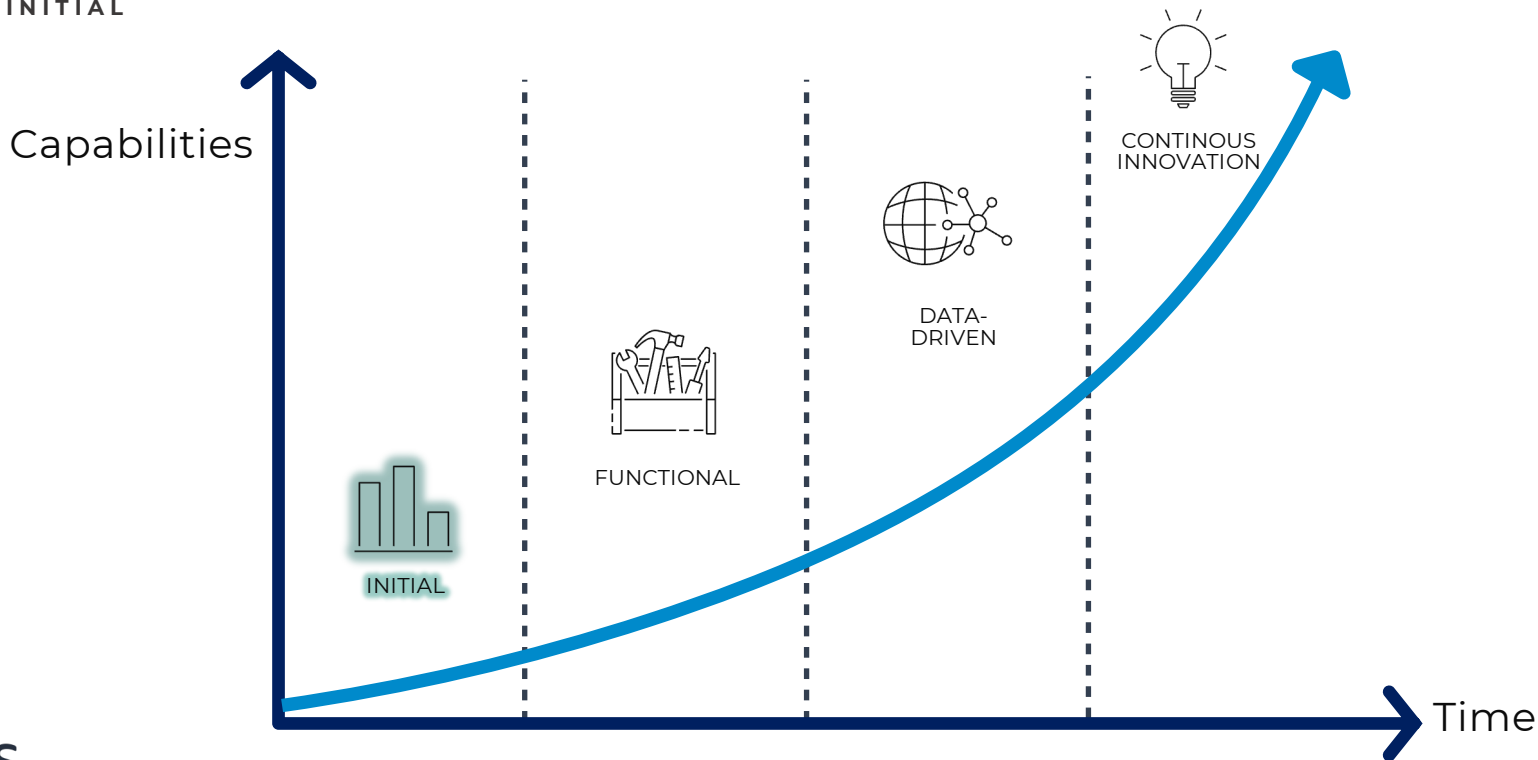


<https://youtu.be/P0qqBrNZ3AA>



Advancing analytical capabilities model

INITIAL



What is a safe data Landing Zone

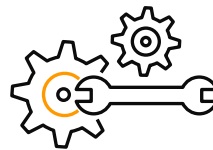
INITIAL



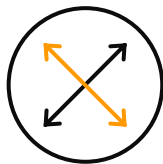
Protect your data
logically through
access controls



Protect your data
physically through
encryption controls



Automate tasks to
save time and reduce
risk



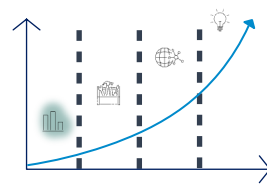
Scale with visibility and
control as your business
grows



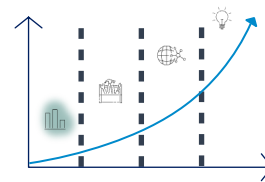
Use architecture
patterns to simplify
controls



Inherit global security
and compliance
controls



Safe Landing Zones in AWS



INITIAL



Identity & access management



Detection



Infrastructure protection



Data protection



Incident response



Compliance

AWS Identity & Access Management (IAM)

AWS Single Sign-On

AWS Organizations

AWS Directory Service

Amazon Cognito

AWS Resource Access Manager

AWS Security Hub
Amazon GuardDuty

Amazon Inspector

Amazon CloudWatch

AWS Config

AWS CloudTrail

VPC Flow Logs

AWS IoT Device Defender

AWS Firewall Manager

AWS Network Firewall

AWS Shield

AWS WAF – Web application firewall

Amazon Virtual Private Cloud (VPC)

AWS PrivateLink

AWS Systems Manager

Amazon Macie

AWS Key Management Service (KMS)

AWS CloudHSM

AWS Certificate Manager

AWS Secrets Manager

AWS VPN

Server-Side Encryption

Amazon Detective

CloudEndure DR

AWS Config Rules

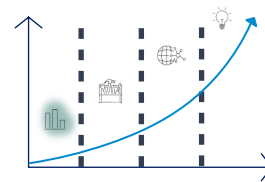
AWS Lambda

AWS Artifact
AWS Audit Manager



Data adapters and patterns

INITIAL

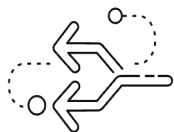
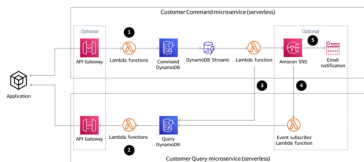


Build effort

DYI, CQRS

Aided

0 - ETL



Amazon DynamoDB
Streams



AWS
Database
Migration Service



Amazon
AppFlow



Amazon Aurora
0-ETL

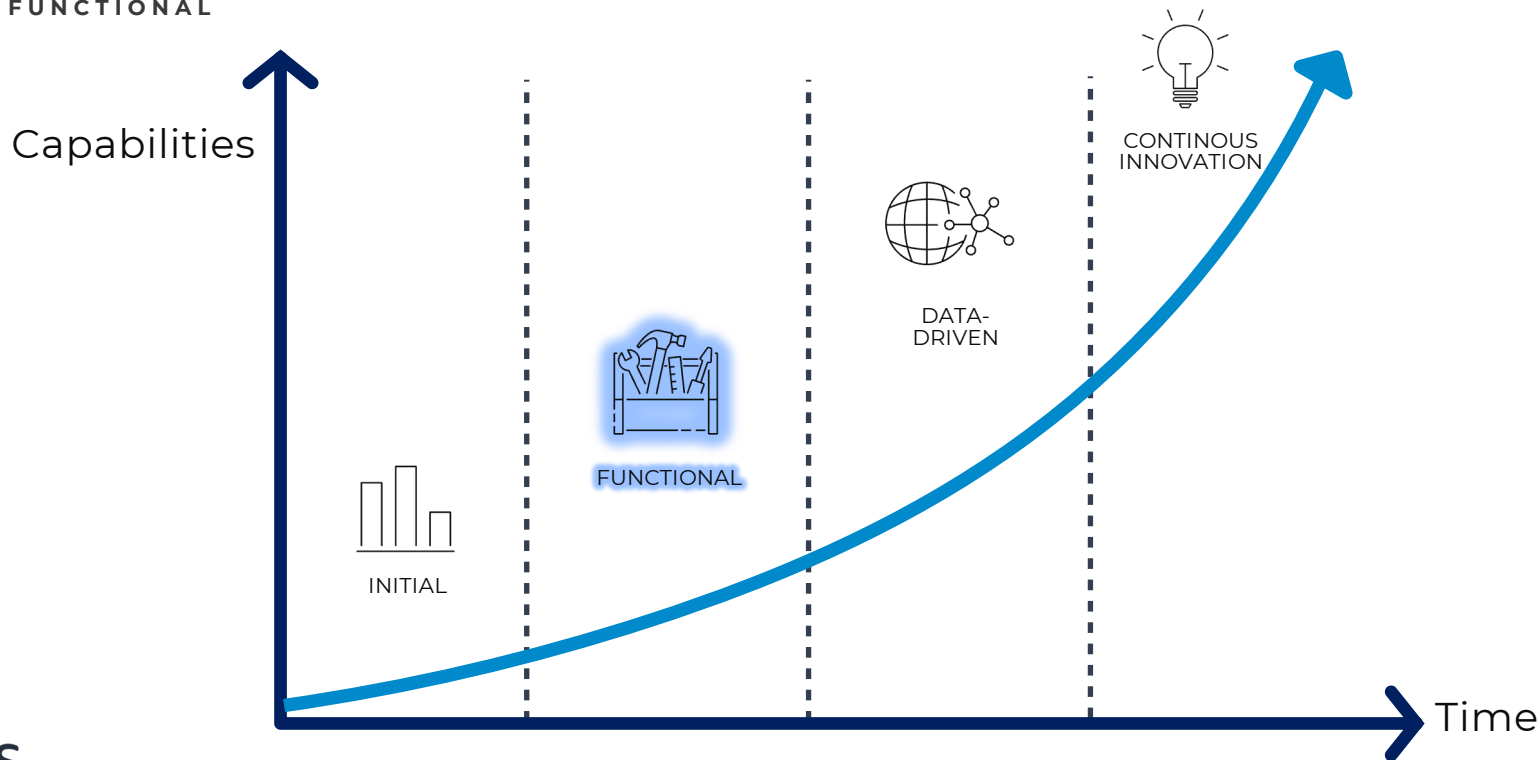


Amazon
Athena
Federated queries



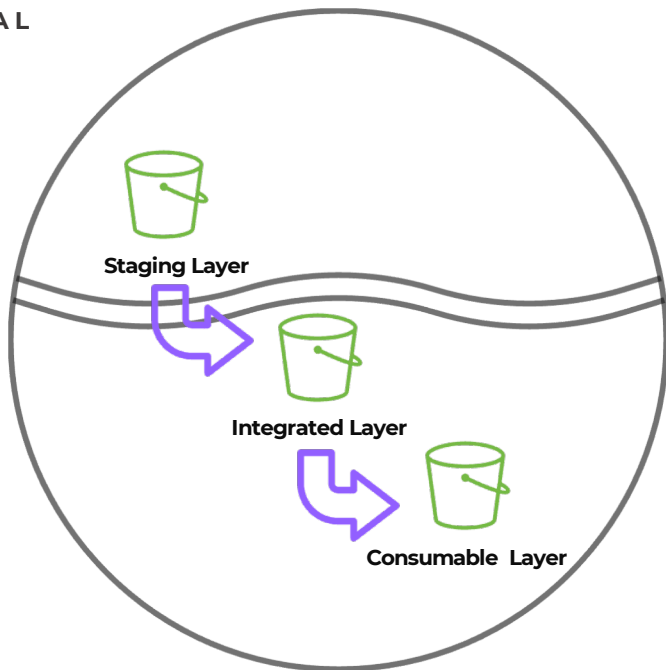
Advancing analytical capabilities model

FUNCTIONAL



Modern data strategy architecture

FUNCTIONAL



Staging Layer

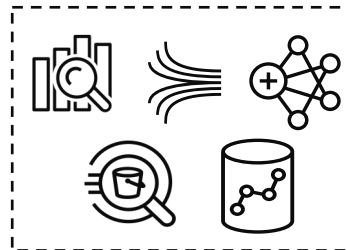
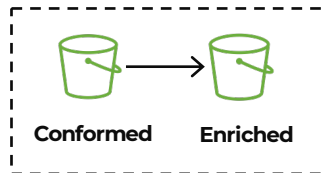
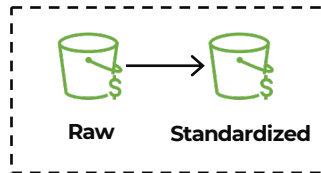
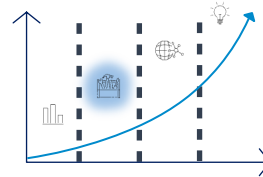
- Factual source data for all time
- Data treated as immutable
- **Cost effective**

Integrated Layer

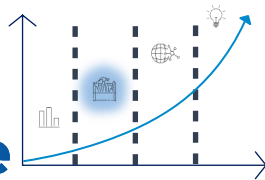
- Trustworthy
- Valuable
- Accessible, **Interoperable** and **Open**

Consumable Layer

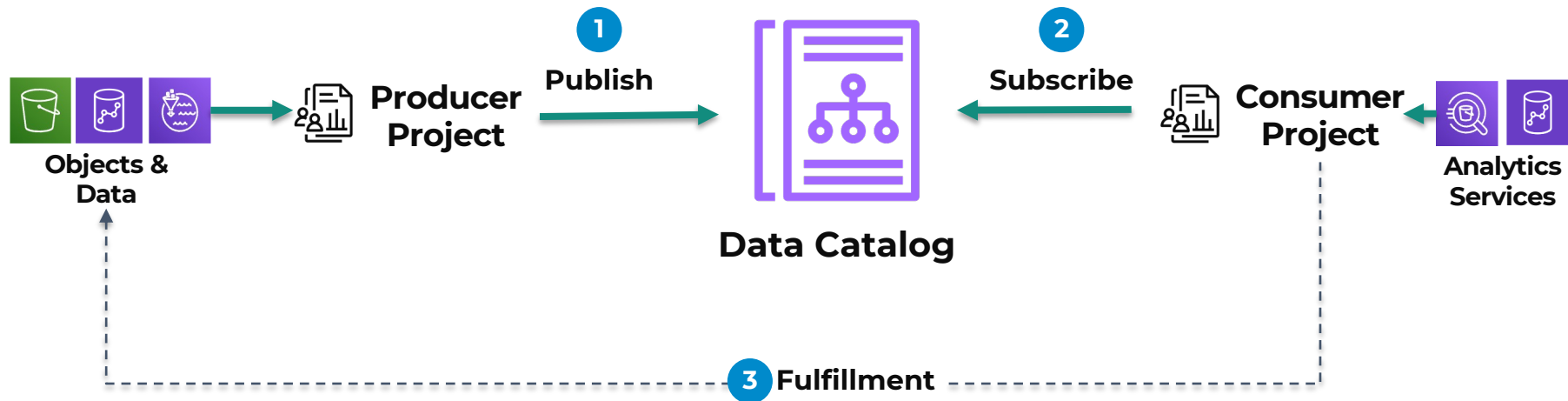
- Application-Specific format and granularity
- Fits consumer's technology
- Adds **Flexibility** to integrated layer



Data catalogs for discoverability enables self-service analytics

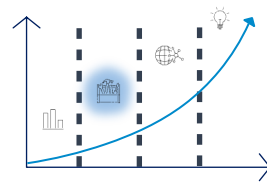


FUNCTIONAL



Data quality assurance

FUNCTIONAL



Accuracy

Data accurately represents the real world values

Example: incorrect spelling or product or person names or addresses.

Validity

Data conforms to the syntax of its definition (format, type, range).

Example: incorrect classification value for gender or customer type.

Timeliness

Data represents reality from the required point of time.

Example: customer address change which is effective on 1st of July, is entered into the system on the 15th.

Completeness

Data are complete in terms of required potential of data.

Example: customer address is missing zip code.

Uniqueness

Data are properly identified and recorded only once.

Example: Single customer is recorded twice in the database with different identifiers.

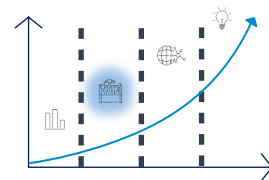
Consistency

Data are represented consistently across the data set.

Example: incorrect spelling or product or person names or addresses.

Data quality models

FUNCTIONAL



Data volumes

small - transactional

big - eventual

Proactive

Reactive



schema and model
consistency



AWS Glue
Data quality
jobs

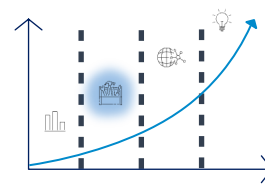


AWS Brew
Data sampling
and profiling

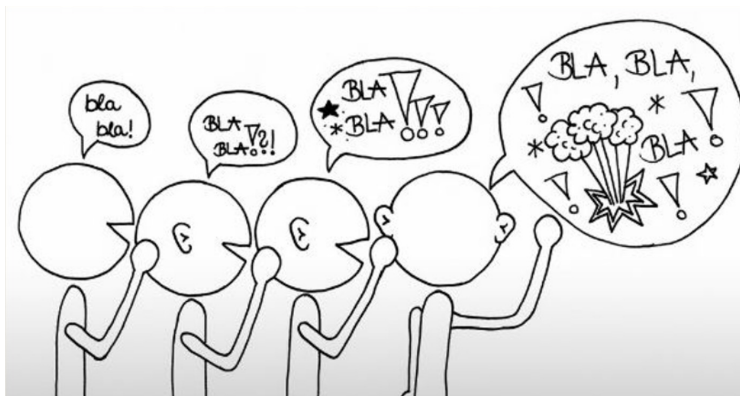


Data lineage

FUNCTIONAL



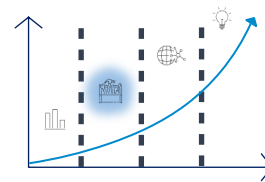
**“a guppy swims in
a shark tank.”**



**“The puppy that
spins and barks,
stank”**

„Game of
telephone“

Sample lineage solutions on AWS

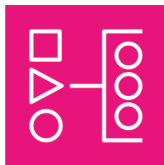


FUNCTIONAL

Data pipelines and jobs



AWS Glue



Amazon
MWAA



Amazon EMR



Lineage collection



Plugin and open lineage agent

Open Lineage



Amazon Neptune
Lineage backend

Visualization



MARQUEZ



AWS Fargate



Apache **Atlas**

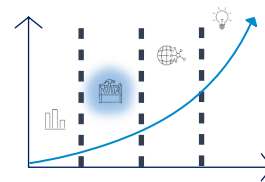


Amazon EMR



Data quality where you need it

FUNCTIONAL



Connect and
discover



AWS Glue
connectors



AWS Glue
crawler

Catalog and quality



AWS Glue Data Catalog

Open Lineage



AWS Glue
Data Quality

Transform



AWS Glue Data Brew

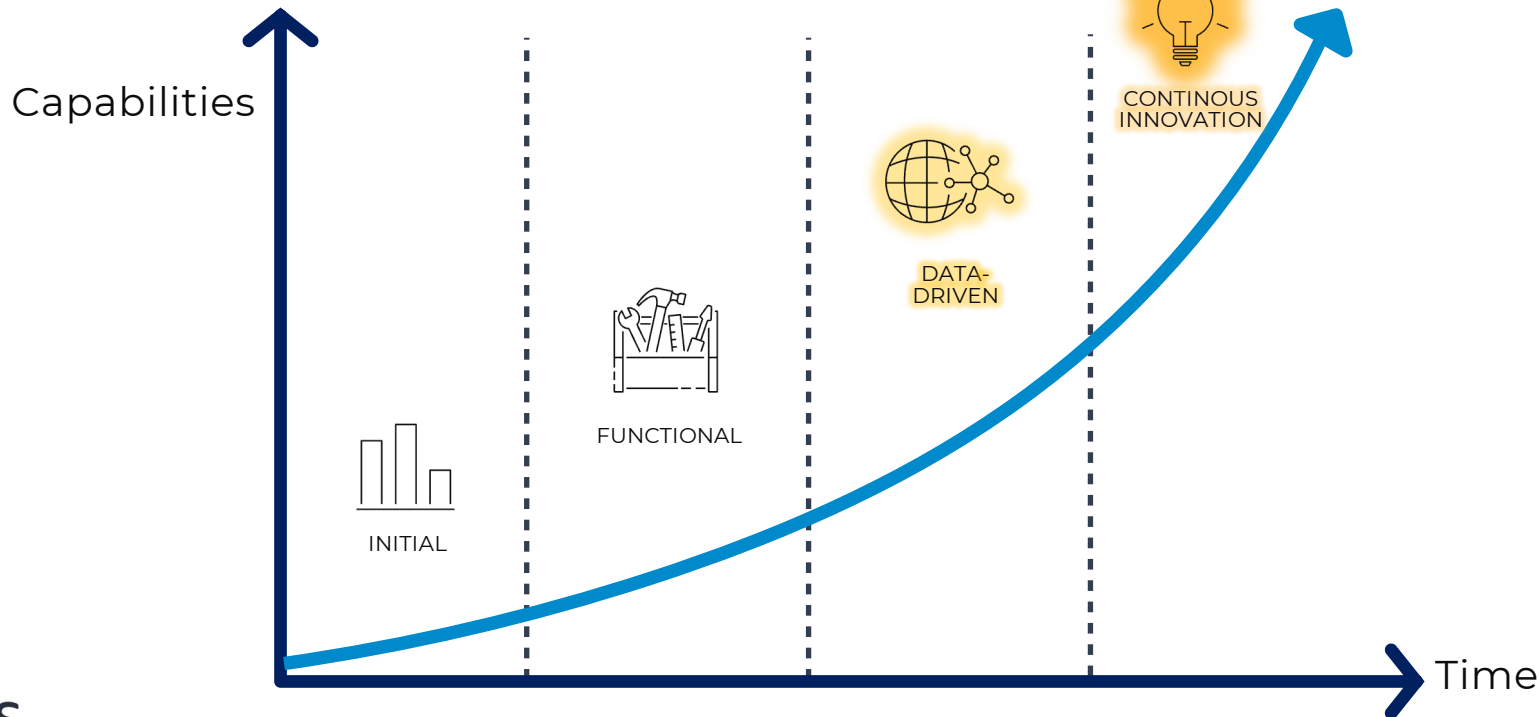


AWS Glue Studio
Visual transformations



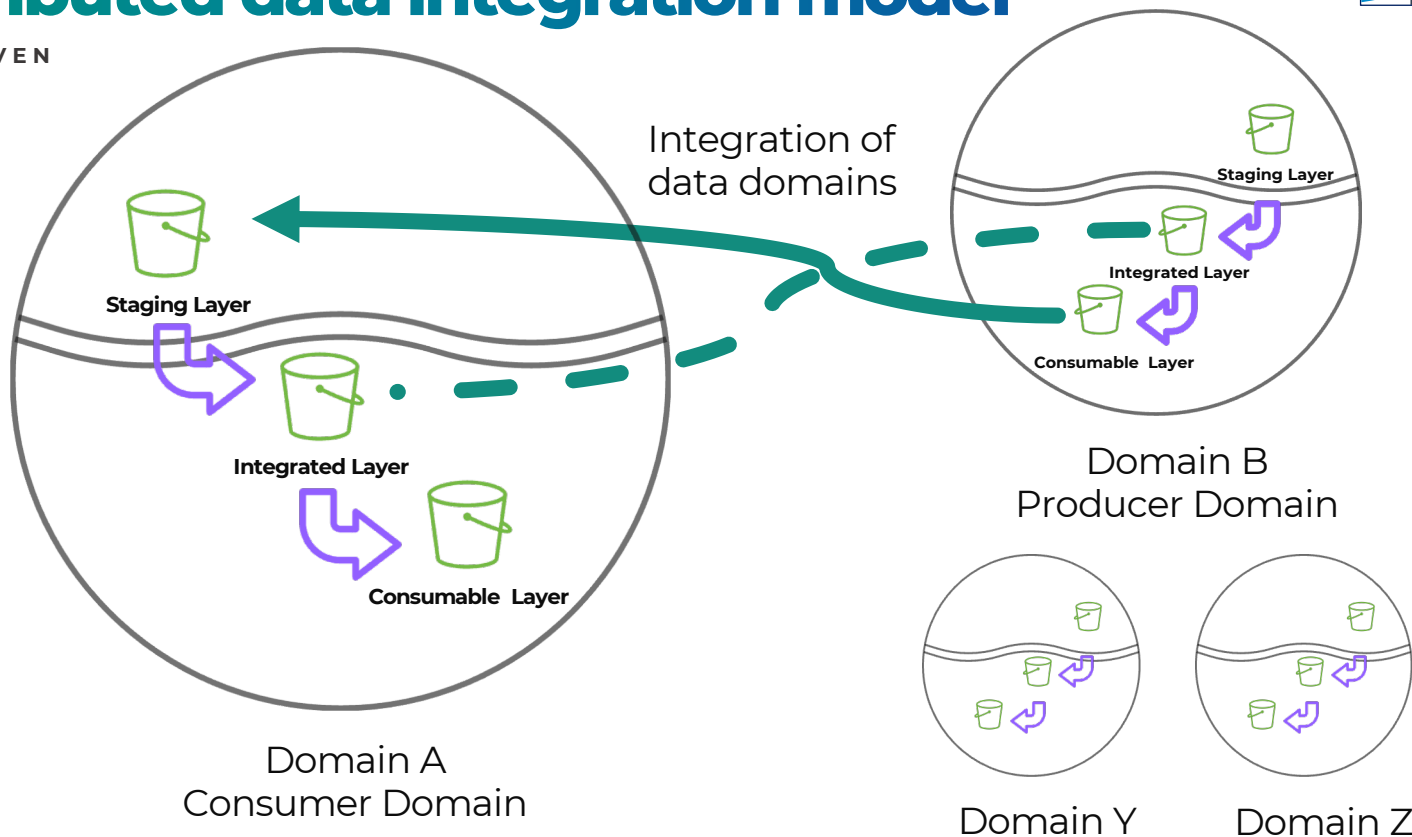
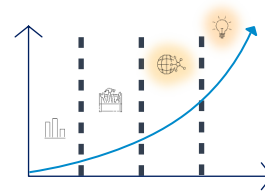
Advancing analytical capabilities model

DATA-DRIVEN



Distributed data integration model

DATA-DRIVEN



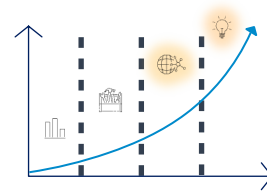
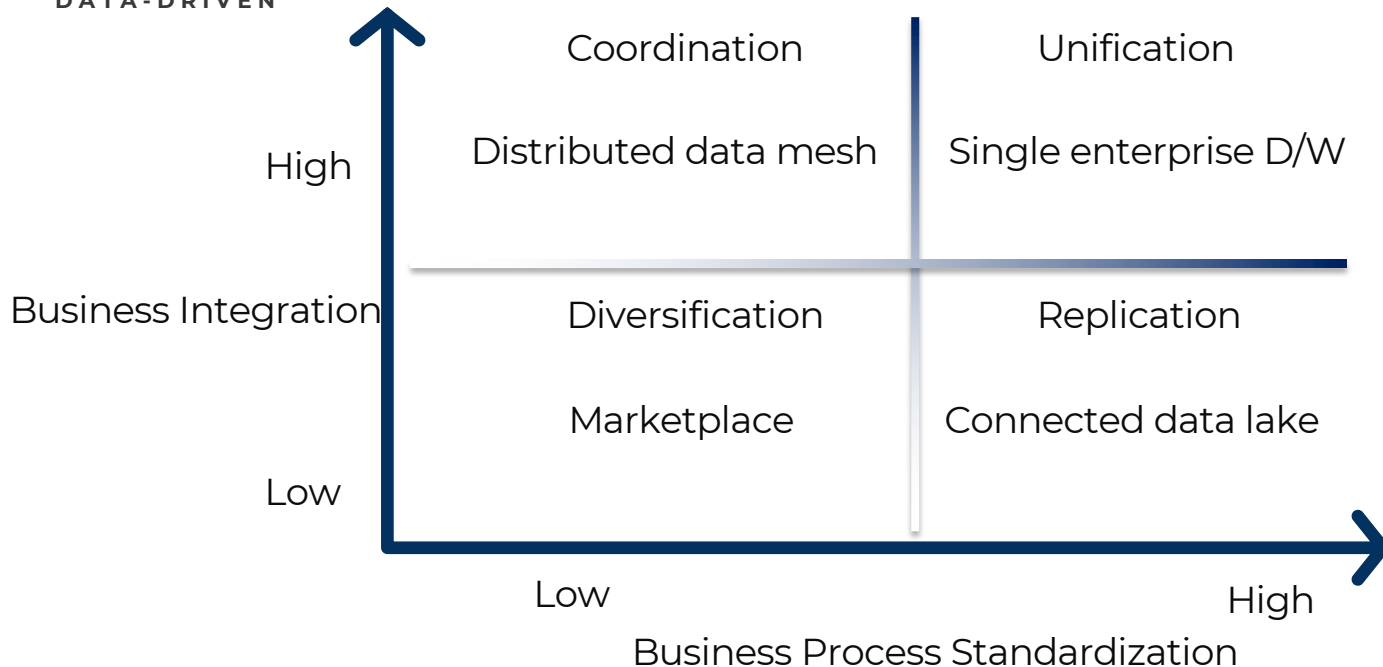
A Balancing Act

“The centralized data team was one of the last functionally divided teams, somewhat at odds with their current domain-oriented business and tech organizational design.”

Zhamak Dehghani, Data Mesh, Speaking about an example organization

Right-level of autonomy for scaling

DATA-DRIVEN

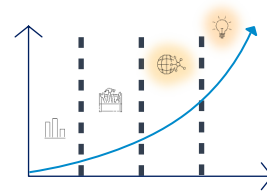
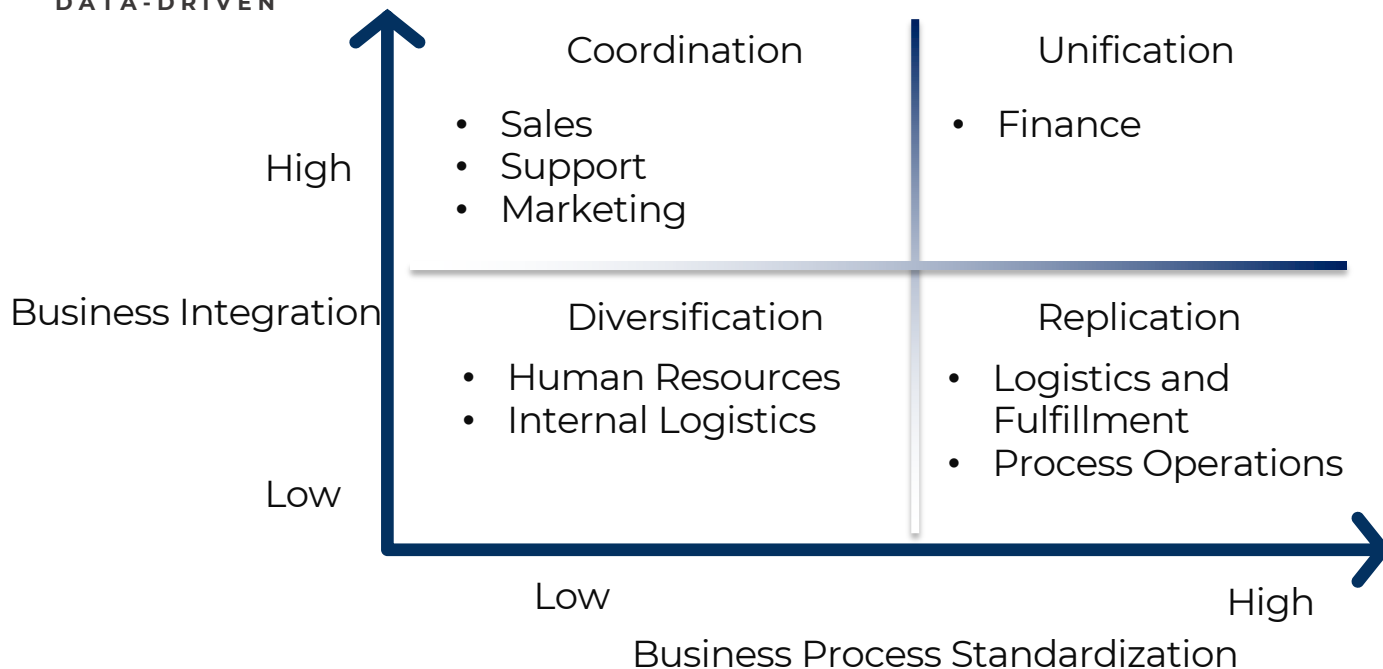


[The MIT Operating Models](#)



Right-level of autonomy for scaling

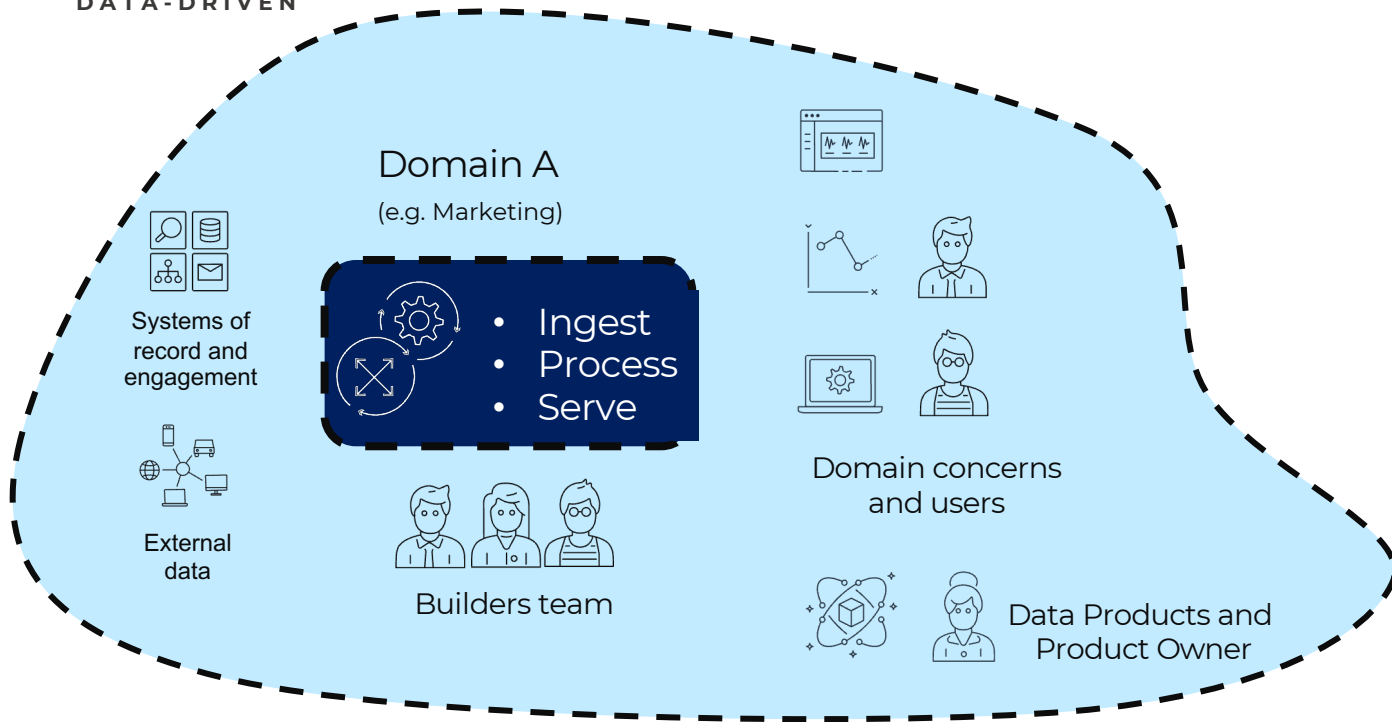
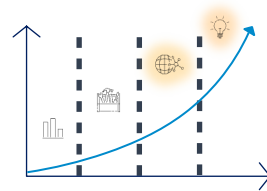
DATA-DRIVEN



[The MIT Operating Models](#)

Implications of encapsulated domains

DATA-DRIVEN



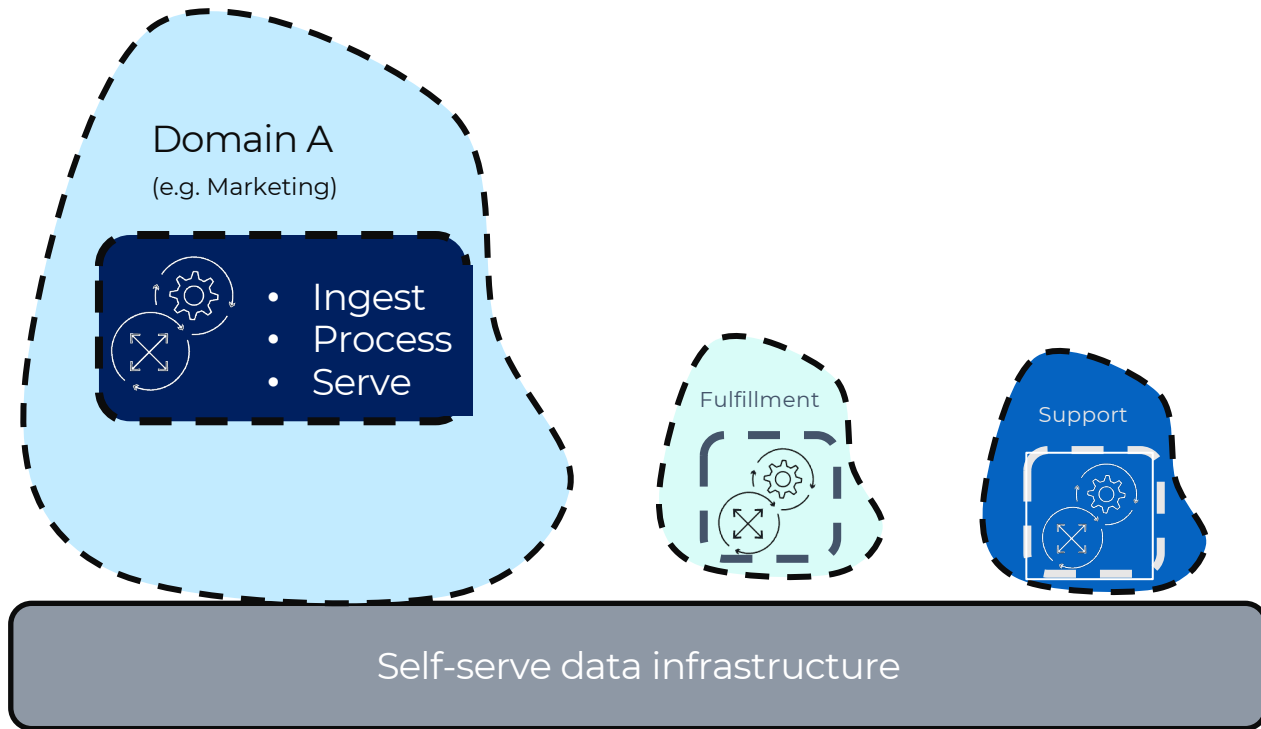
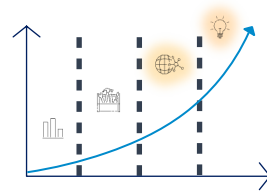
Other Domains

Data Product
Feature Request

- Priority
- Changes
- Perceived Value

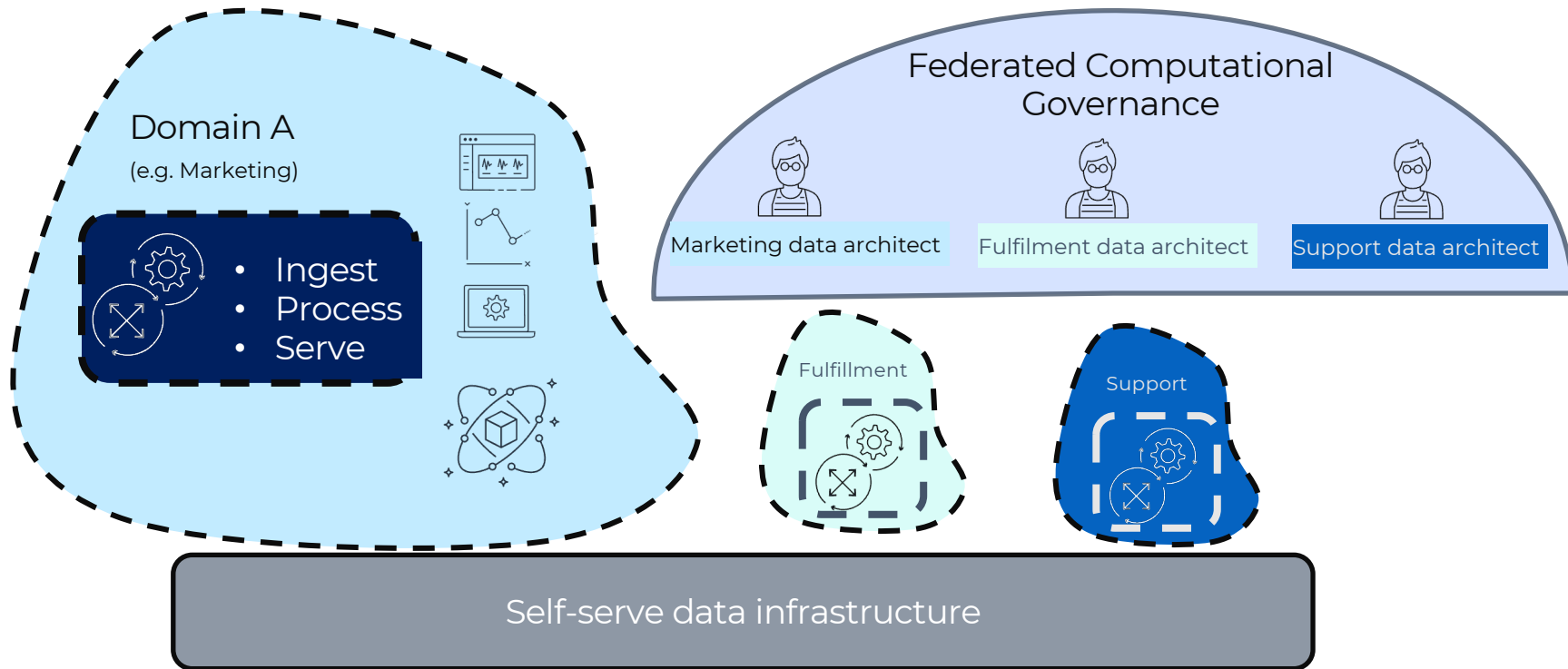
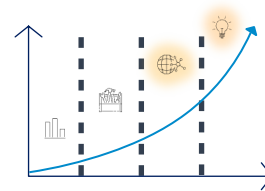
Self-serve data infrastructure for efficiency

DATA-DRIVEN



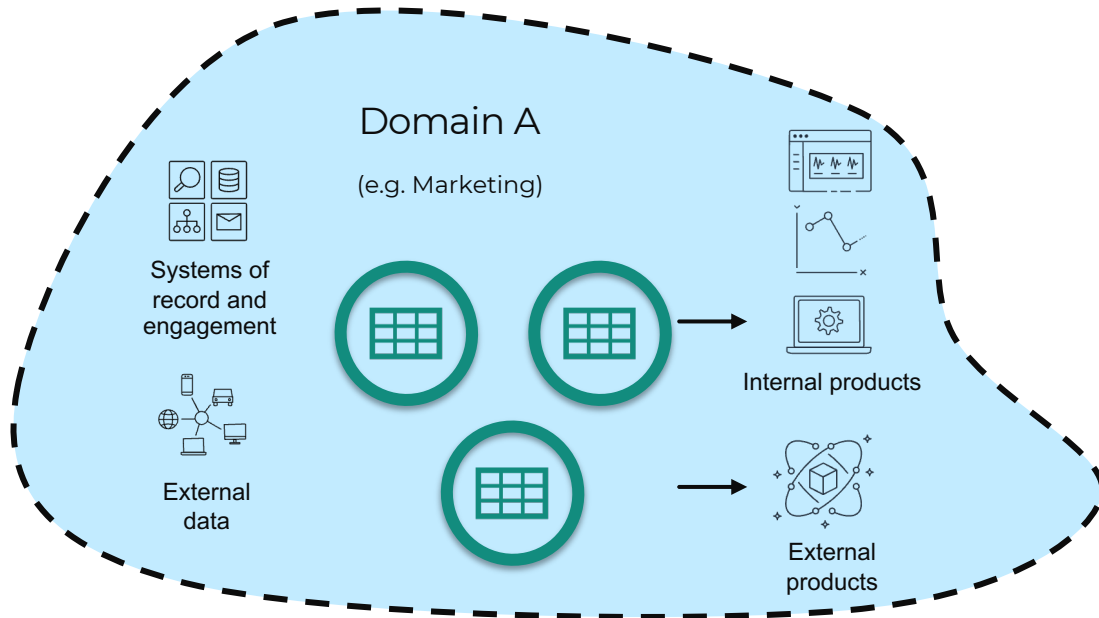
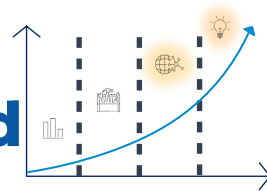
Federated governance balances interests

DATA-DRIVEN



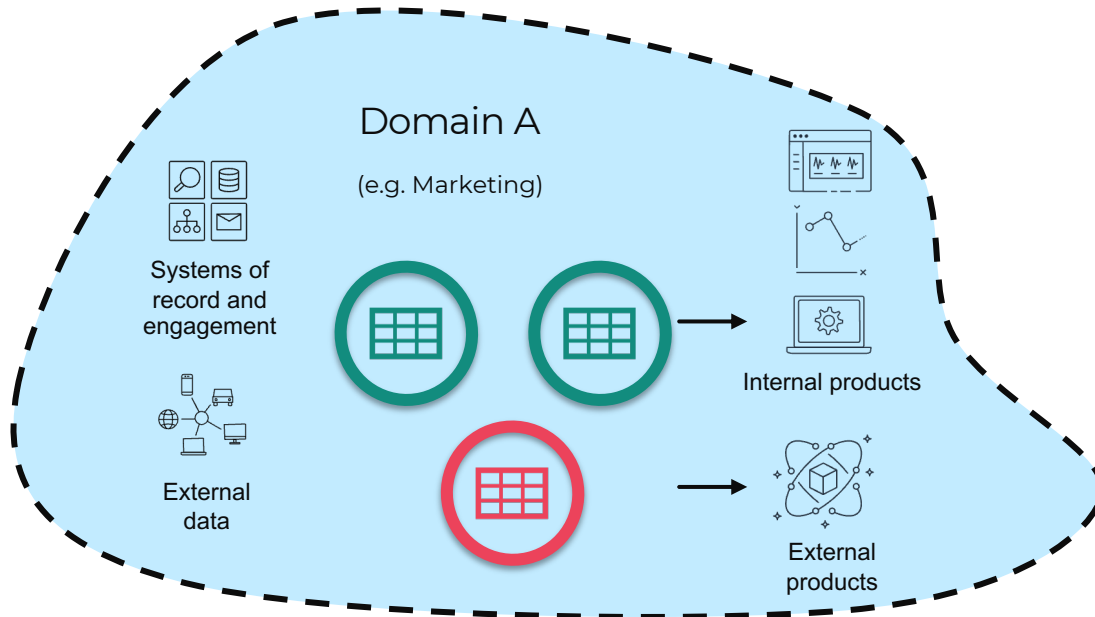
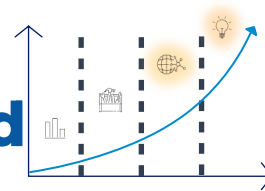
Domain models and architectures are encapsulated and exposed as products

DATA-DRIVEN



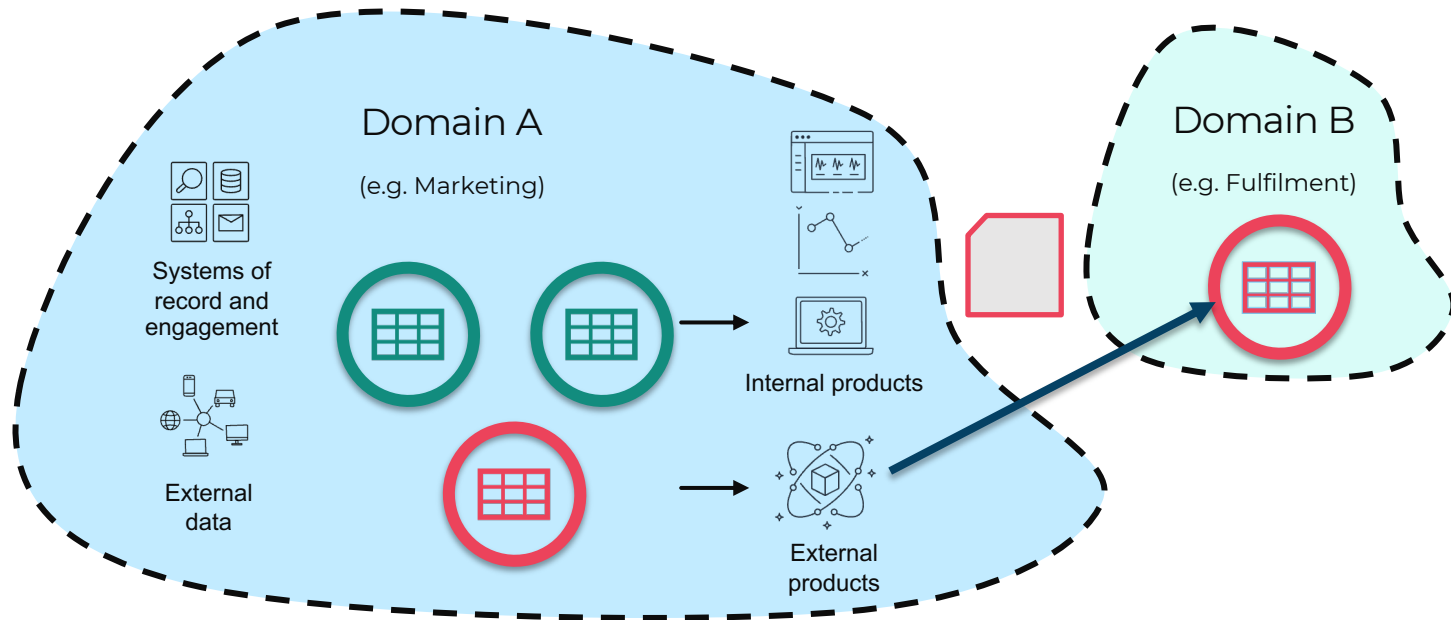
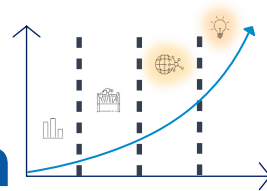
External product models are governed by federated governance

DATA-DRIVEN



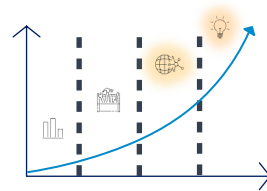
Federated governance prioritizes integration

DATA-DRIVEN

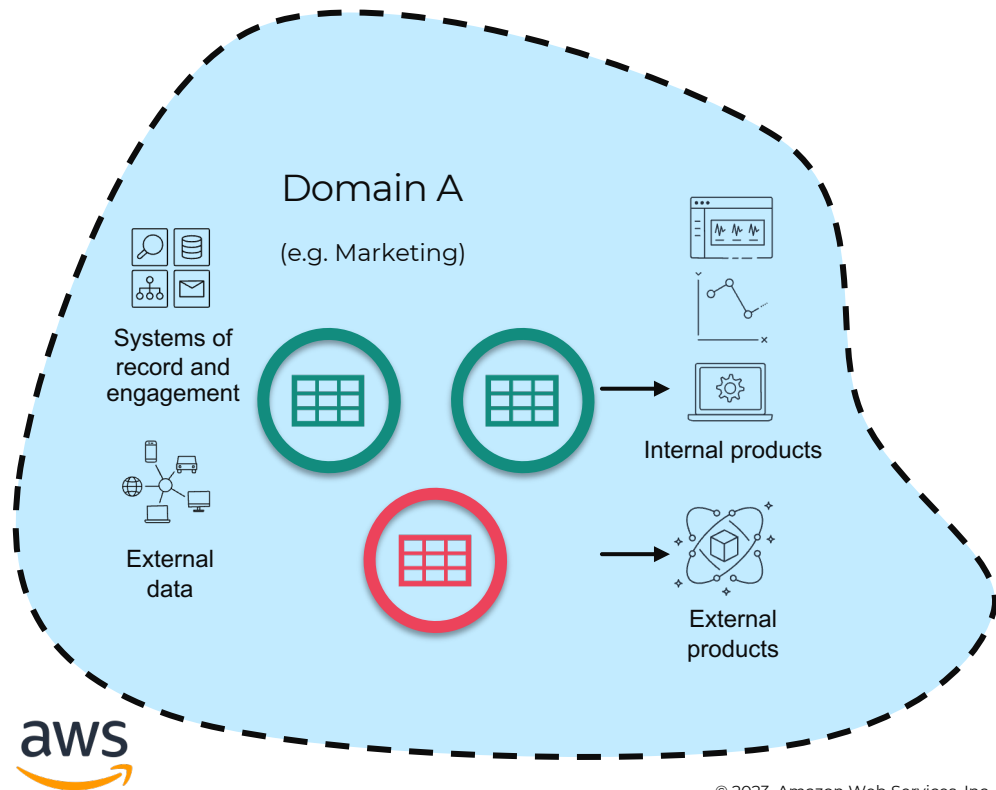


[Good] architectures and models should ...

DATA-DRIVEN

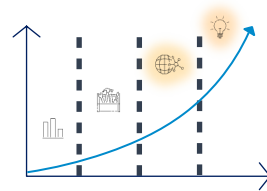


- Capture **domain** data effectively
- Working backwards from domain use cases
- Flexible enough to create **domain products for main domain stakeholders**
- Enables data products Adhering to governance standards
- Enables Integrated insights from multiple domains when needed

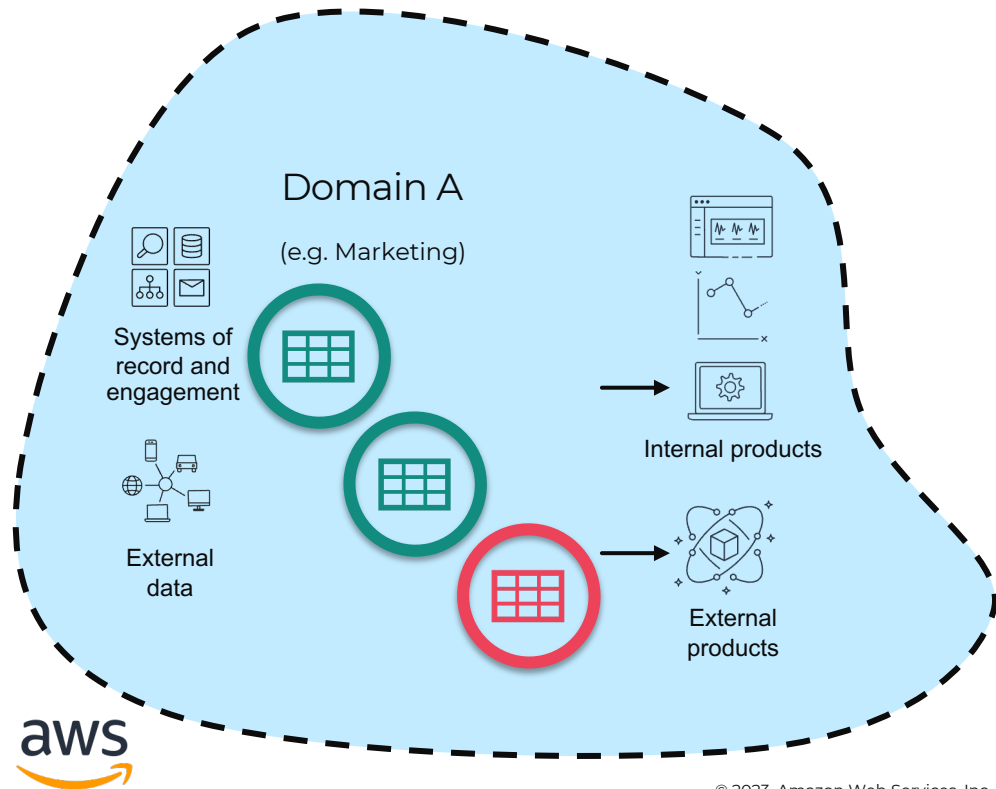


[Good] architectures and models should ...

DATA-DRIVEN



- Capture **domain** data effectively
- Working backwards from domain use cases
- Flexible enough to create **domain products for main domain stakeholders**
- Enables data products Adhering to governance standards
- Enables Integrated insights from multiple domains when needed

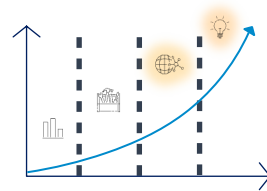


AWS Lake Formation

BUILD A SECURE
DATA LAKE IN DAYS

Amazon DataZone

SHARE, SEARCH, AND
DISCOVER DATA AT SCALE



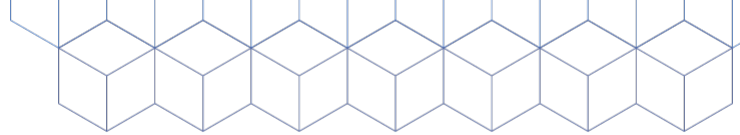
Build data lakes quickly



Simplify security management

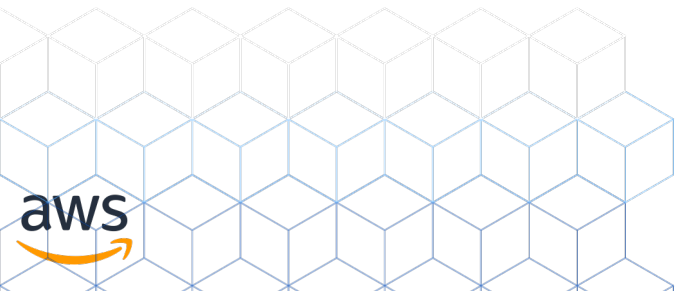


Provide self-service access to data

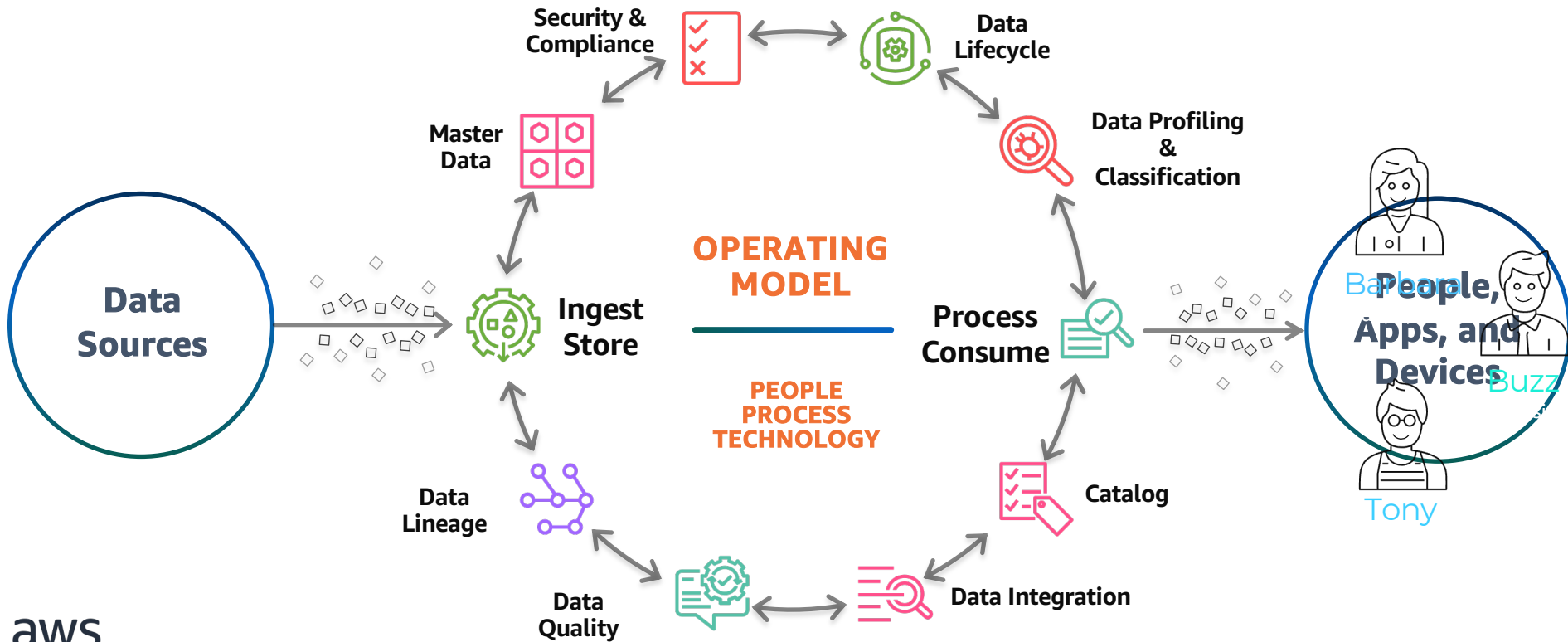


End of story?

.....



Data Governance enables you to scale



More Resources

Modern Data Architecture [Whitepaper](#)

Derive Insights from AWS Modern Data [Whitepaper](#)

Well Architected Framework – [Analytics Lens](#)

AWS serverless data analytics pipeline reference architecture [blog post](#)

AWS Prescriptive Guidance – [CQRS and enabling data persistence in microservices](#)

Data Lake Foundation on AWS – [Quick Start](#)

Build data mesh pattern at scale using AWS Lake Formation [blog post](#)



**BIG
DATA
DAYS
2023**



Thank You.

<https://pulse.buildon.aws/survey/HHQKTHOG>



John Mousa

jmousa@amazon.de



/in/johnmousa



/johnmousa



**ENTERPRISE
BIG DATA
FRAMEWORK®**



info@bigdataframework.org



<https://www.bigdataframework.org/>

