AWS re:Invent

NOV. 29 – DEC. 3, 2021 | LAS VEGAS, NV

GPS315

Building an internal data marketplace

Igor Alekseev (he/him)
Sr. Partner Solutions Architect
AWS

Subhasis Bhattacharya (he/him)
Data & Analytics Partner Segment Tech Lead
AWS

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Why an internal data marketplace?

Don’t we have concepts like data catalog, data warehouse, data mesh, data lakehouse . . . ?



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data consumption challenges

Top 3 frequently asked data questions:

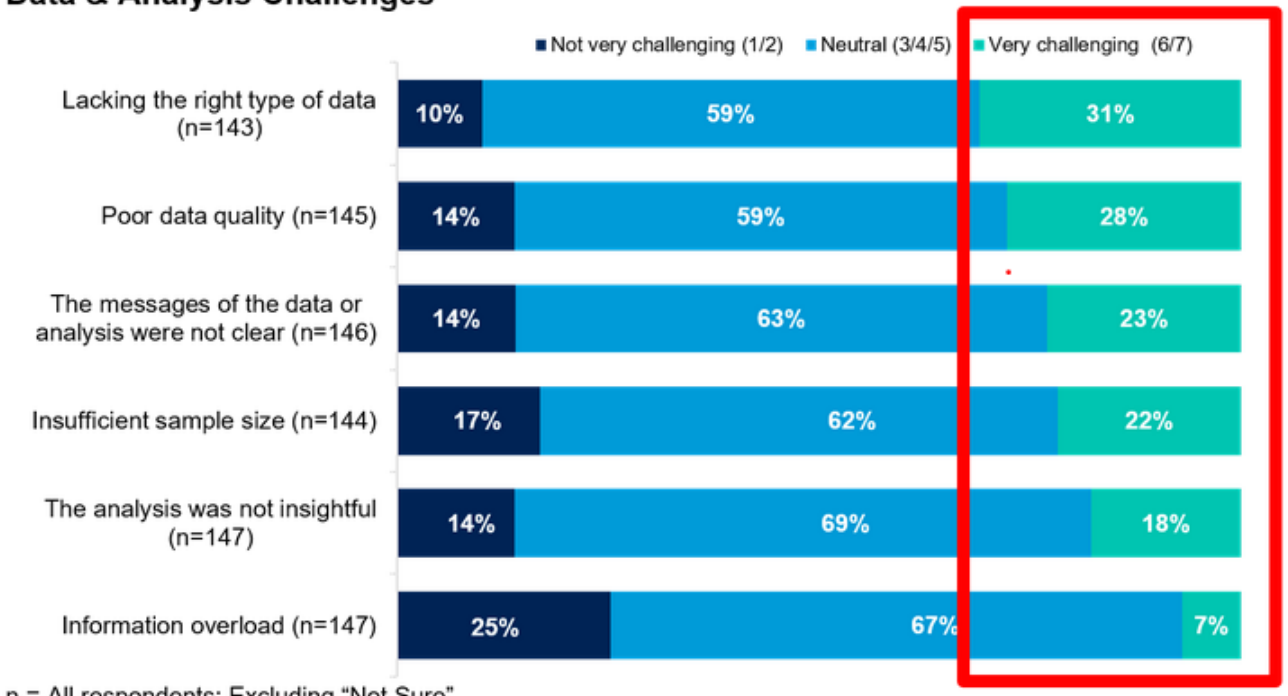
- Where can I find that data?
- How good is that data?
- How can I get access to that data?



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Top decision-making challenges

Most organizations' decisions are challenged to some degree by data and analysis issues



n = All respondents; Excluding "Not Sure"

Q05. How much of a challenge were the following data and analysis issues for the decision?

Source: Gartner 2020 Re-engineering the Decision

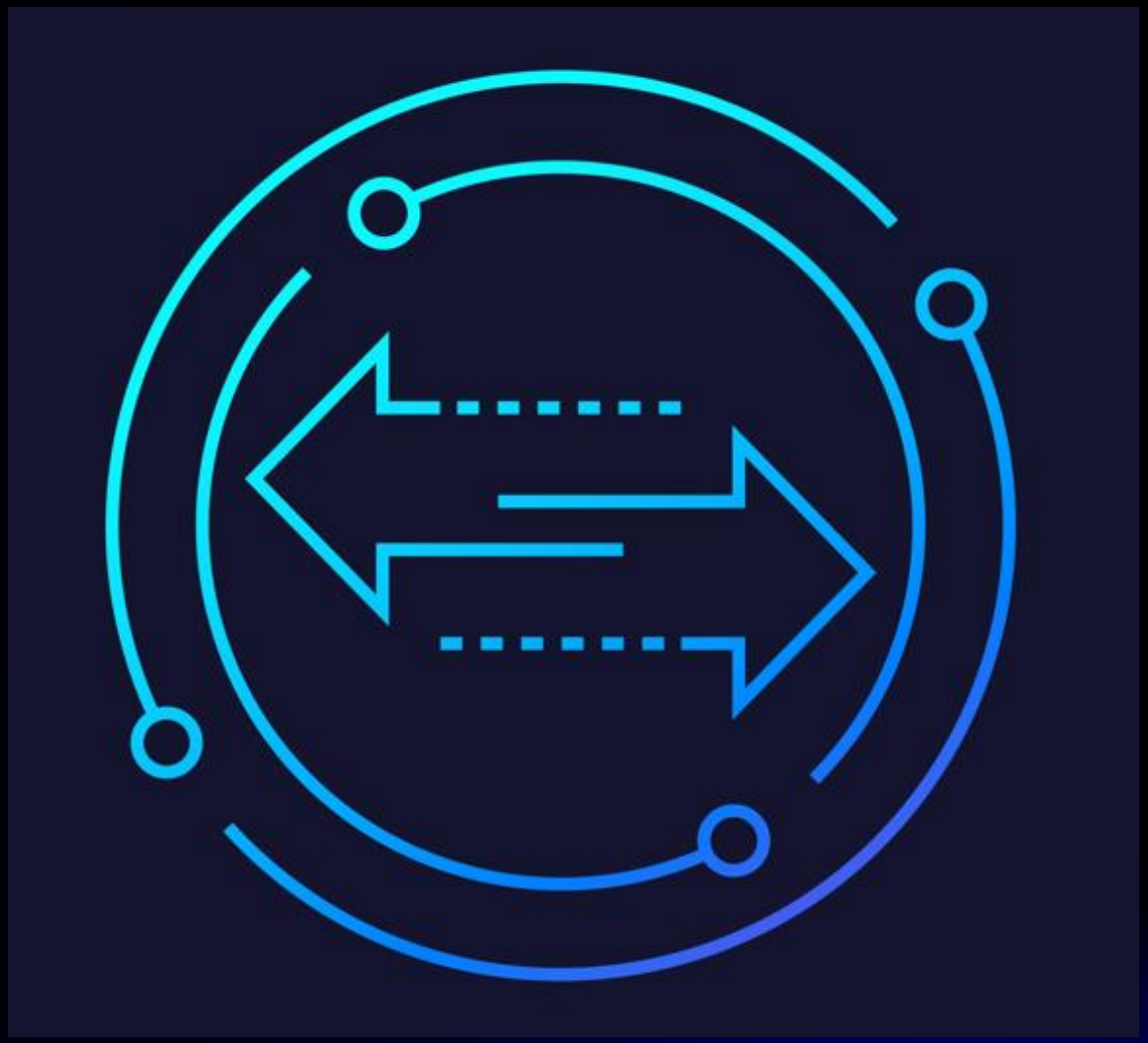
Gartner Research Circle members (One Circle)

Gartner

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

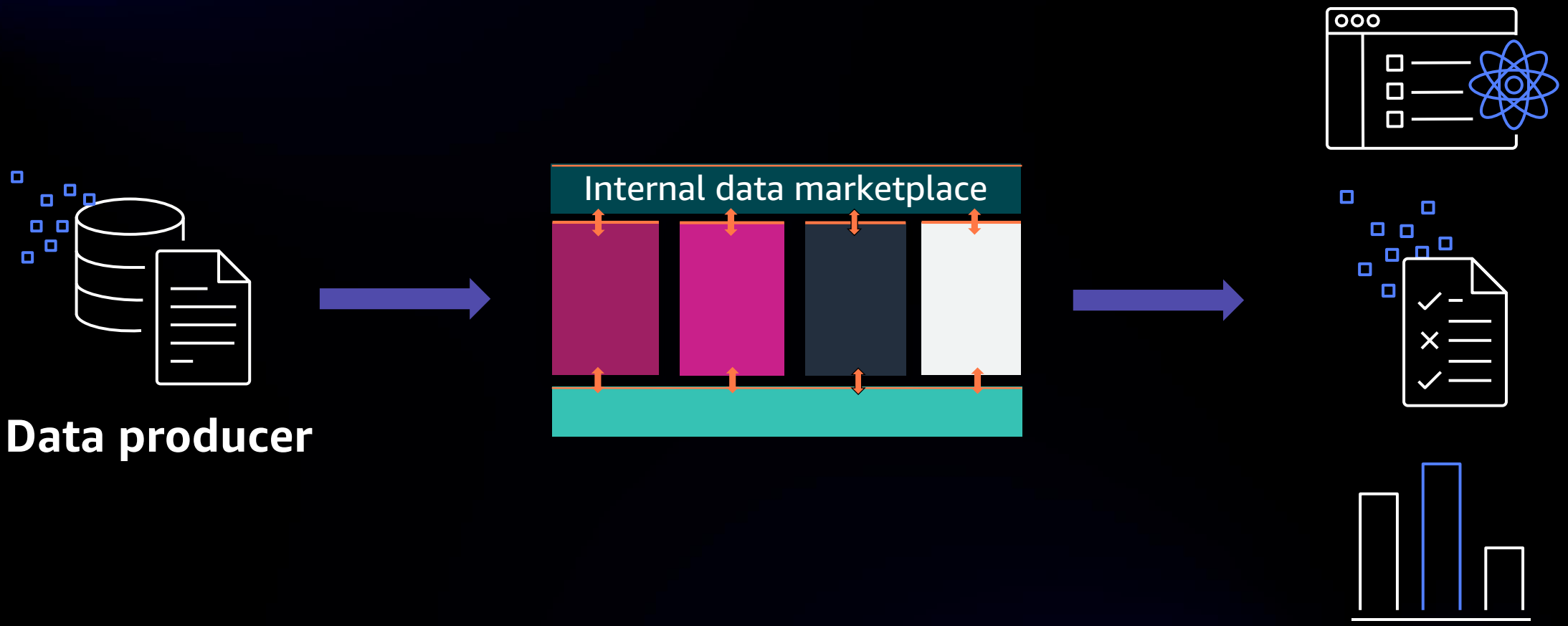
Public marketplaces and data brokers

- Public data marketplaces
- AWS Data Exchange
- Data brokers



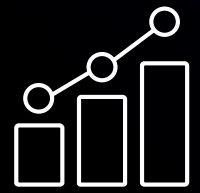
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Internal data marketplace

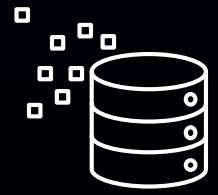


© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Customers want more value from their data



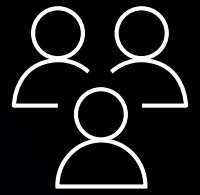
Growing exponentially



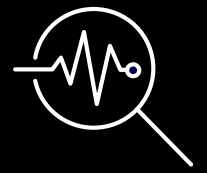
From new sources



Increasingly diverse



Used by many people



Analyzed by many applications

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Customer expectations are evolving too



Everything is easy to use



Everything is cloud-native



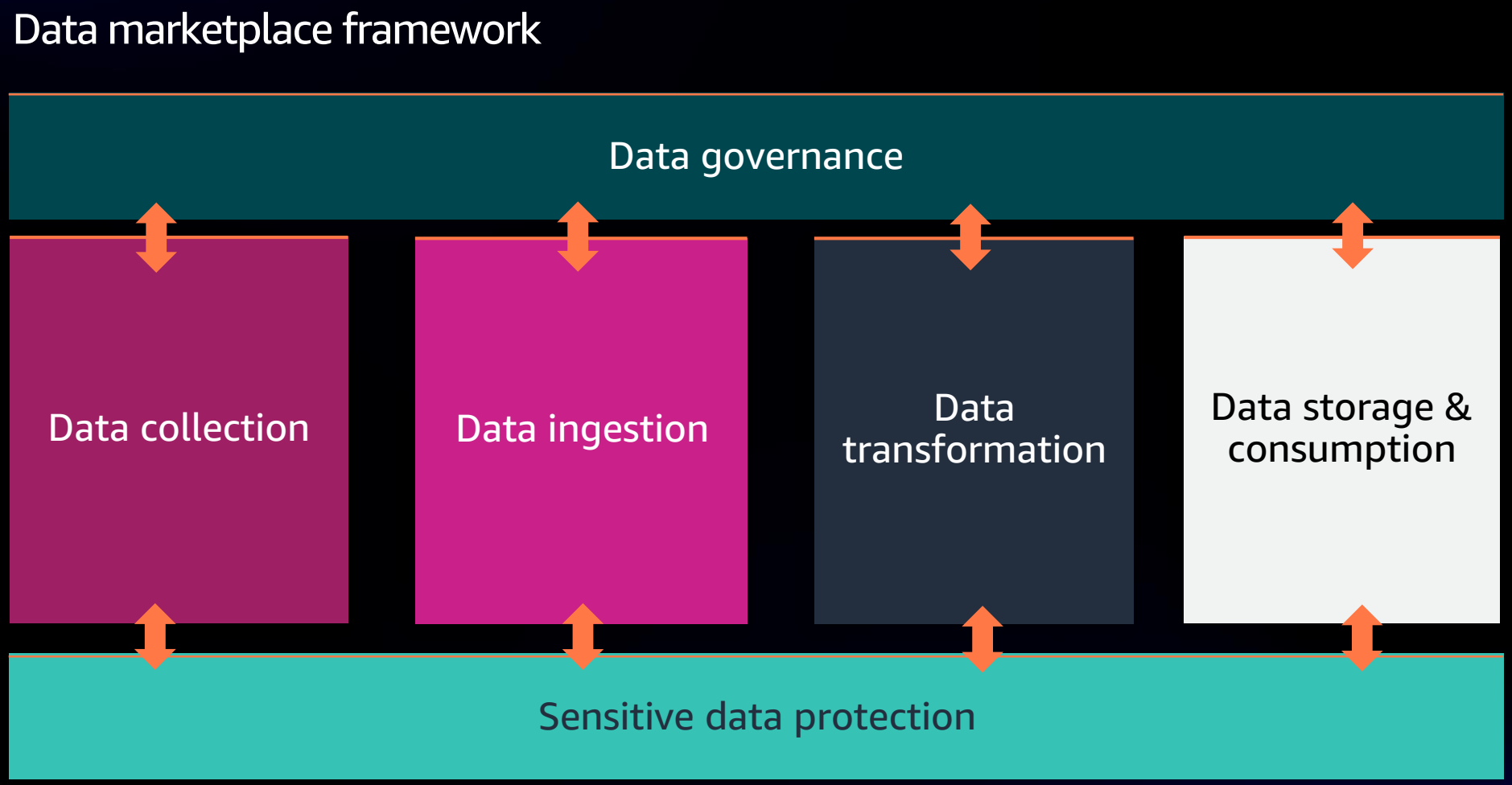
Seamless experience



Self-service preference

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Internal data marketplace framework



aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Internal data marketplace – Governance

Data governance

- Data discovery
- Data catalog
- Data lineage
- Data dictionary
- Data taxonomy
- Data classification
- Data stewardship
- KPI definition

Benefits

High trust on data, self-service, predictable model to consume and publish data effortlessly

Maintain high data quality throughout the data lifecycle

Growth

2019	$7.32 billion
2027	$35.45 billion

Harvard Business Review: CAGR of 21.9% from 2020 to 2027

Services



AWS Lake Formation



AWS Glue Data Catalog

Partner solution

Collibra

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Internal data marketplace – Data protection

PI data protection

- Role-based access control (RBAC)
- Data de-identification
  - Encryption
  - Masking
  - Tokenization
  - Hashing

Benefits

Protect data confidentiality, help comply with regulations such as PCI, CCPA, GDPR, and HIPAA

United Nations Conference on Trade and Development

- 82% countries with e-transaction laws
- 66% countries with privacy laws
- 56% counties with consumer protection laws

Services



AWS Key Management Service (AWS KMS)



AWS Identity and Access Management (IAM)

Partner solution

PROTEGRITY

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Internal data marketplace – Collection

Data collection

- Real-time, near-real-time and batch mode data collection from diverse sources

Benefit

Consume data from diverse sources with no setup time

Eliminate logical data corruption issues during data ingestion

Efficiency

Setting up a metadata-driven approach can save customers up to 90% development time and reduce time to market by 95%

Services



AWS Database Migration Service (AWS DMS)



Amazon Kinesis

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Internal data marketplace – Ingestion

Data ingestion

- Schema registration
- Schema validation
  - Data contract
- Data classification
  - Sensitive data tagging

Benefit

Eliminate logical data corruption issues, identify and secure sensitive data during ingestion

Purpose

Data ingestion acts as a backbone for ETL by efficiently handling large volumes of big data

Services



AWS Lambda



Amazon Macie

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Internal data marketplace – Transformation

Data transformation

- Data structuring
  - Aggregation
  - Normalize/ de-normalize
- Data cleansing/ filtering
- Data integration/ hydration

Benefit

Transform, clean, and enrich data for easy consumption

Governed transformation algorithm for quick development and deployment

Growth

2019	$49 billion
2026	$20 billion

Services



AWS Glue



Amazon EMR

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Internal data marketplace – Consumption

Data storage & consumption

- Business intelligence
- Machine learning
- Data warehousing
- Data lake
- Semantic layer

Benefit

Purpose build tools to store and consume data at scale and at ease

Integrated data visualization and ML features to gather KPIs

Growth

2019	$49 billion
2027	$103 billion

Forbes.com: Expected big data software market to reach $103 billion by 2027

Services



Amazon Redshift



Amazon Simple Storage Service (Amazon S3)



Amazon QuickSight

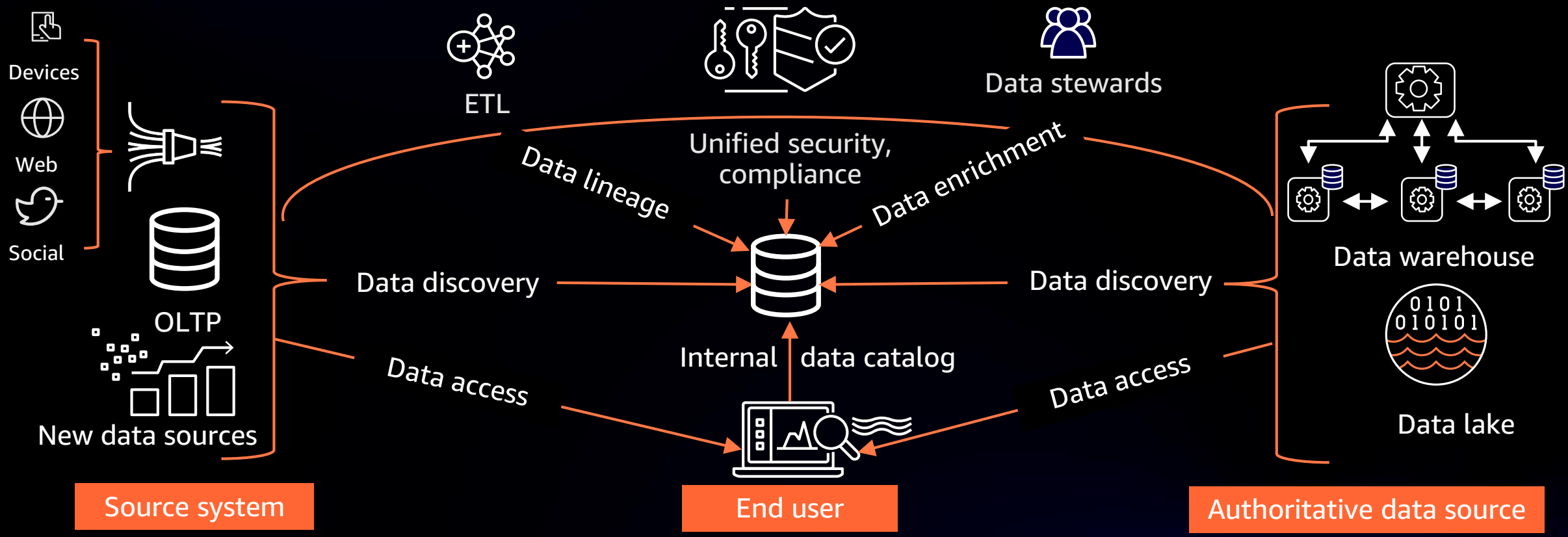


Amazon Athena

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

How does a data marketplace work?

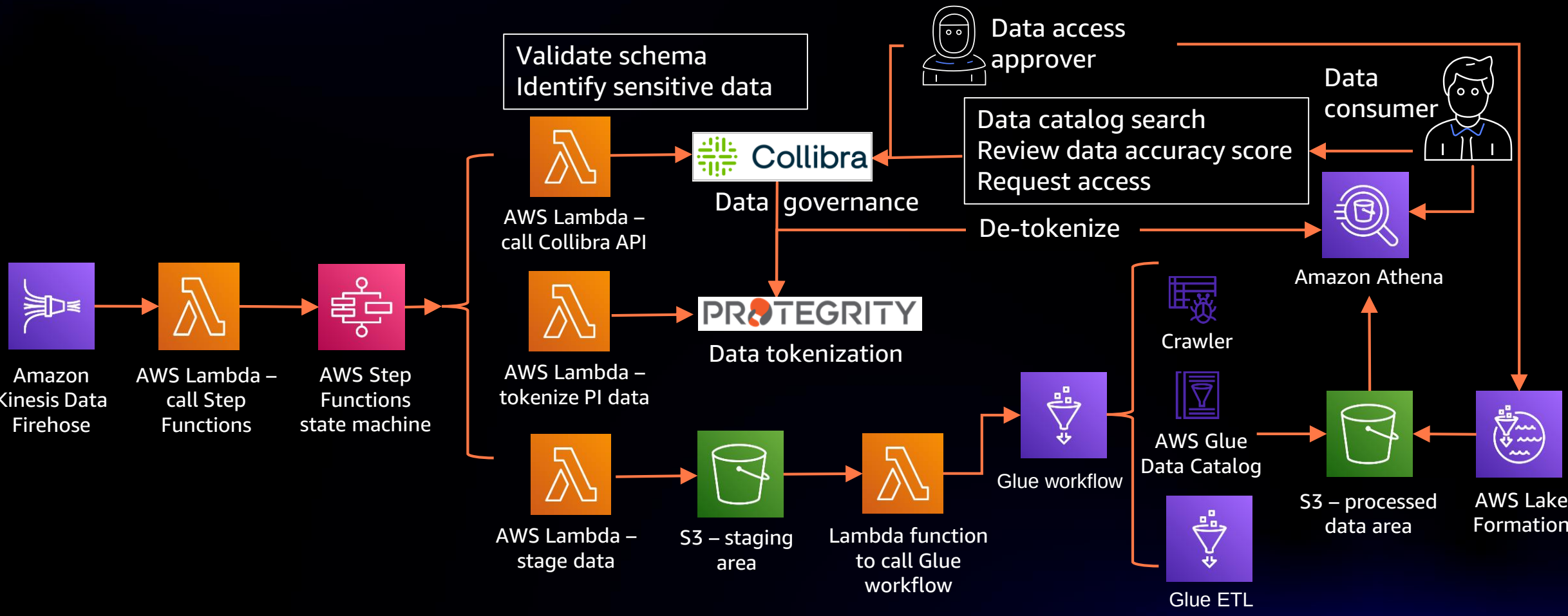
BUILDING AN INTERNAL DATA MARKETPLACE TO DEMOCRATIZE INSIGHTS AND ACCELERATE BUSINESS AGILITY



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Internal data marketplace – Implementation

REFERENCE ARCHITECTURE



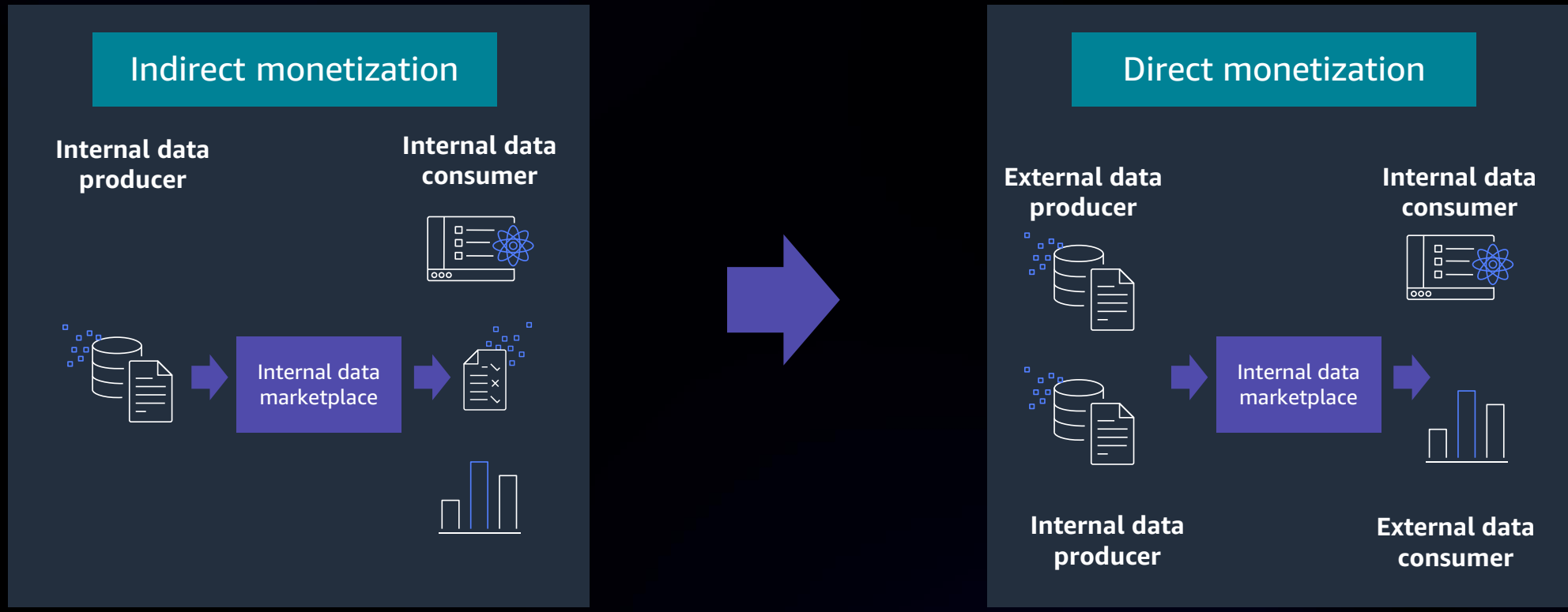
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Recap

- Data quality
- Data lineage
- Data catalog
- Data security
- Access control
- Shopping cart consumption experience
- AWS native services and partner integrations
- Data monetization

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data monetization journey



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Subhasis Bhattacharya
subhabh@amazon.com

Igor Alekseev
ialek@amazon.com

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.