

# **Implementare una architettura Data Lake serverless con AWS**

Francesco Diaz

EMEA Sr. GTM Spec, Analytics

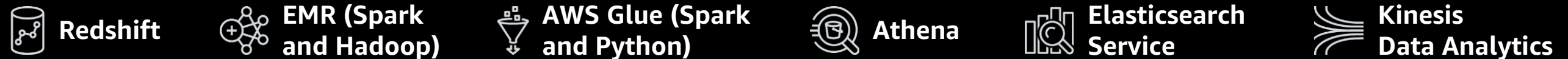
Amazon Web Services

# The AWS analytics portfolio

## DATA, VISUALIZATION, ENGAGEMENT, AND MACHINE LEARNING



## ANALYTICS



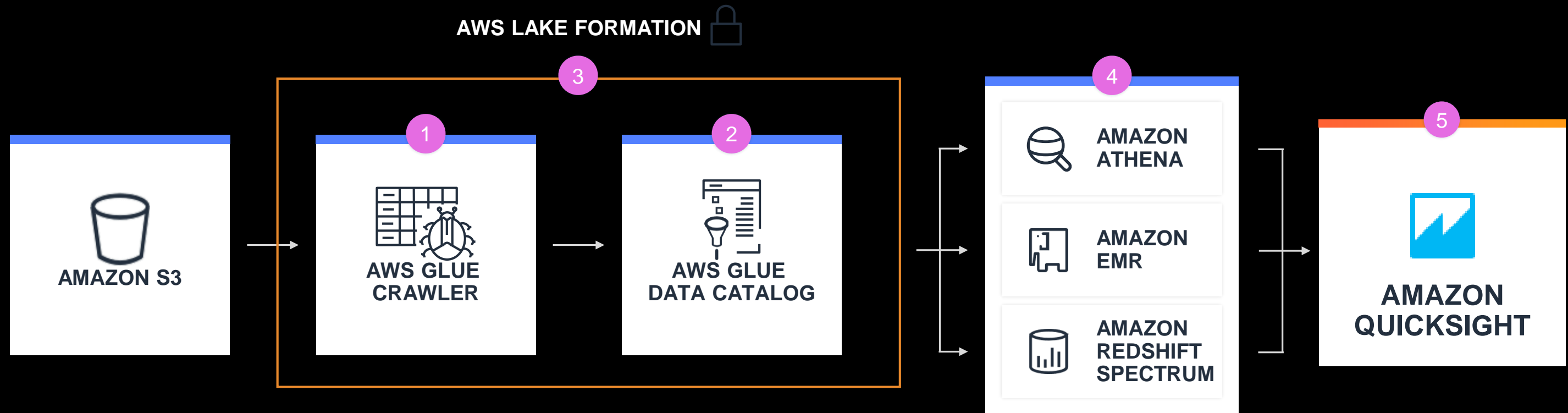
## DATA LAKE INFRASTRUCTURE AND MANAGEMENT



## DATA MOVEMENT

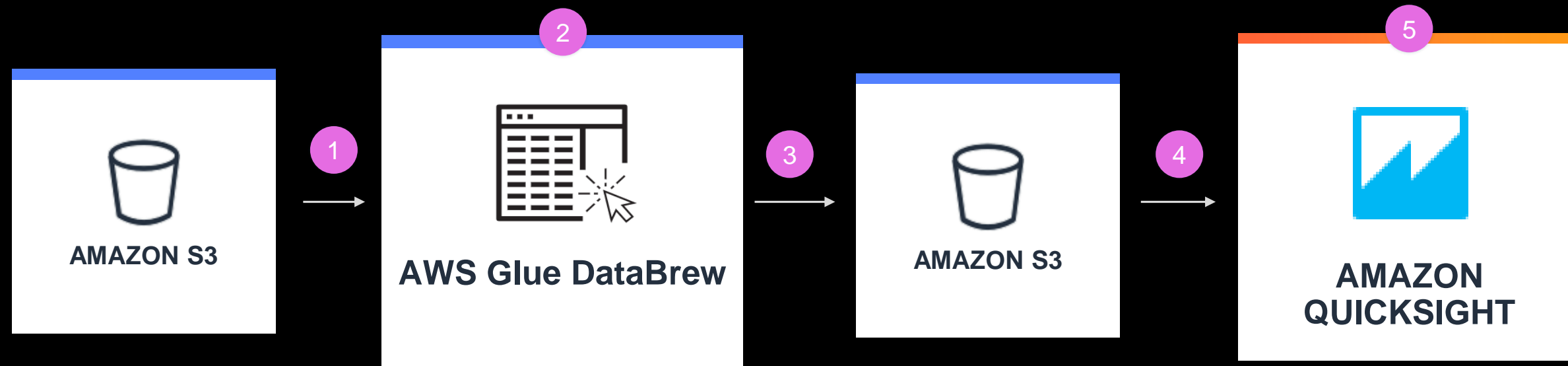
Database Migration Service | Snowball | Snowmobile | Kinesis Data Streams | Kinesis Data Firehose | Managed Streaming for Apache Kafka

# Serverless data lake & analytics with AWS



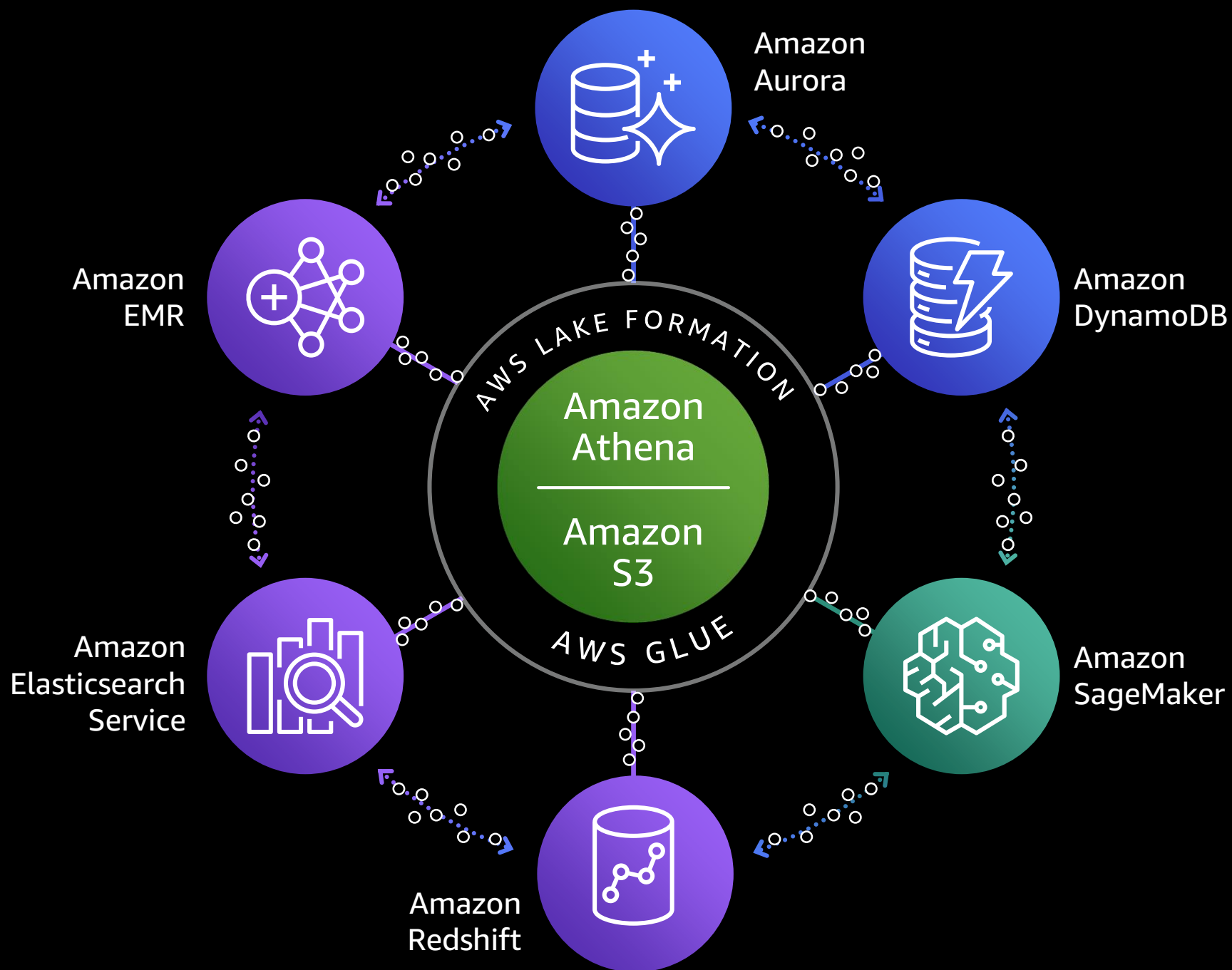
- 1 Crawlers scan your data sets and populate the Data Catalog
- 2 The Data Catalog serves as a central metadata repository
- 3 Once catalogued, apply security and permissions
- 4 Query your data immediately for analytics
- 5 Visualize with QuickSight

# Serverless data prep and visualization with AWS



- 1 Connect to data in S3 or upload files directly to DataBrew
- 2 Clean, prepare, transform data without writing code
- 3 Export prepared data to S3
- 4 Import data from S3 to SPICE
- 5 Visualize with QuickSight

# Lake House architecture on AWS



**SCALABLE DATA LAKES**

**PURPOSE-BUILT  
DATA SERVICES**

**SEAMLESS  
DATA MOVEMENT**

**UNIFIED GOVERNANCE**

**PERFORMANT AND  
COST-EFFECTIVE**

# References

- Athena engine v2: <https://docs.aws.amazon.com/athena/latest/ug/engine-versions-reference.html>
- Athena Bucketing vs Partitioning: <https://docs.aws.amazon.com/athena/latest/ug/bucketing-vs-partitioning.html>
- Using Athena with JDBC: <https://docs.aws.amazon.com/athena/latest/ug/connect-with-jdbc.html>
- Best practices to optimize Athena performances: <https://aws.amazon.com/blogs/big-data/top-10-performance-tuning-tips-for-amazon-athena/>
- Serverless tutorial and code: <https://github.com/aws-samples/serverless-data-analytics>
- Create a Glue crawler: <https://docs.aws.amazon.com/glue/latest/ug/tutorial-add-crawler.html>
- Getting started with Glue Studio: <https://docs.aws.amazon.com/glue/latest/ug/tutorial-create-job.html>
- Cloud Formation Template for Glue Studio tutorial: <https://aws.amazon.com/blogs/big-data/making-etl-easier-with-aws-glue-studio/>
- QuickSight – check permissions: <https://docs.aws.amazon.com/quicksight/latest/user/troubleshoot-connect-athena.html>
- QuickSight workshops: <https://learnquicksight.workshop.aws/en/>

# Thank you!

Francesco Diaz

@francedit

