The 4 Primary Stages

1. Data Ingestion & Staging (L1)

* Raw data from transactional, customer, and reference systems.
* Lands in Amazon S3 (L1 staging area).
* MWAA (Airflow) orchestrates pipelines, while Aurora manages orchestration metadata and checkpoints.

2. Data Transformation & Curation (L2 on EKS)

* Spark/Scala jobs running on Amazon EKS perform cleansing, enrichment, and transformations.
* Curated, analytics-ready datasets stored in S3 (L2 warehouse layer).
* Failure handling: Retry on failure; if persistent, alerts Ops.

3. Entity Resolution & Scoring (L3 on Quantexa + EKS)

* Quantexa services running on EKS resolve entities (accounts, identities, devices).
* Risk scoring and anomaly detection happens here.
* Resolved entities stored in S3 (L3 layer) and registered in AWS Glue Catalog for query/discovery.
* Failure handling: Quarantine problematic data; Ops notified.

4. Analytics, Reporting & AML Investigation

* Quantexa on EKS + additional Spark jobs on EKS run payment chain analysis & pattern detection.
* Outputs pushed into L3 Reporting store on S3.
* AML investigation dashboards/case management tools consume the data.
* Investigators see risk scores, suspicious transactions, and entity networks.

AML – Quantexa on Cloud Architecture

1. Data Ingestion (L1 Staging)

Raw data from upstream systems (transactions, customer info, reference data) lands in Amazon S3.  
Ensures secure, scalable, and cost-effective raw data storage.

2. Orchestration (MWAA + Aurora)

MWAA (Managed Airflow) schedules and triggers jobs for ingestion and processing.  
Aurora DB stores metadata, job status, and orchestration checkpoints.  
Failure Handling: If ingestion fails, an alert is sent to Ops.

3. ETL & Transformation (EKS)

Spark/Scala jobs running on Amazon EKS cleanse, enrich, and transform raw data.  
Generates structured datasets (Parquet/ORC format).  
Failure Handling: Failed jobs are retried; persistent failures trigger alerts.

4. Curated Data Warehouse (L2 Layer)

Transformed datasets are stored in an L2 curated layer on S3.  
Ready for downstream analytics and entity resolution.

5. Entity Resolution & Scoring (Quantexa on EKS)

Quantexa engines running on Amazon EKS resolve entities (linking customer accounts, identities, relationships).  
Suspicious behavior and risk scores are generated.  
Failure Handling: Problematic data quarantined, Ops notified.

6. Resolved Entities (L3 + Glue Catalog)

Resolved entities and networks are stored in L3 layer.  
AWS Glue Catalog registers schemas for easy querying and reporting.

7. Payment Chain Analysis (EKS Jobs)

Additional EKS Spark jobs analyze transaction flows and detect anomalies.  
Identifies circular transactions or unusual high-value transfers.  
Failure Handling: Failed analysis jobs trigger reprocessing.

8. AML Reporting (L3 Reporting Store)

Processed datasets are stored in the reporting layer (L3).  
Feeds regulatory reports, dashboards, and case management tools.

9. Investigation Platform

AML investigators consume dashboards and case management tools.  
They receive risk scores, entity networks, and suspicious transaction alerts.