



AWS Analytics Services Decision Guide

Table of Contents

1. List of AWS Analytics Services Covered

2. Streaming

3. Batch Processing & ETL

4. Data Warehousing & Lakes

5. Business Intelligence & Visualization

6. Data Governance & Collaboration

7. Search & Discovery

8. AWS Analytics Services Master Comparison Table

9. AWS Data Streaming Services Comparison

10. AWS Analytics Query Services Comparison

11. AWS Data Processing & ETL Services Comparison

12. AWS Orchestration Services Comparison

13. Data Governance & Cataloging Services Comparison

14. Business Intelligence & Visualization Services Comparison

15. Search & Log Analytics Services Comparison

16. Specialized Analytics Services

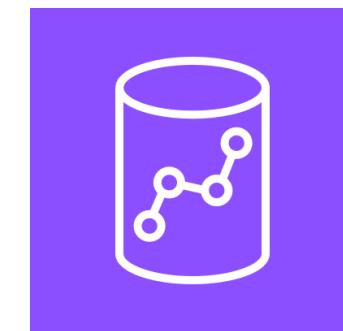
17. Data Storage & Format Optimization Comparison

18. Exam-Focused Decision Matrix

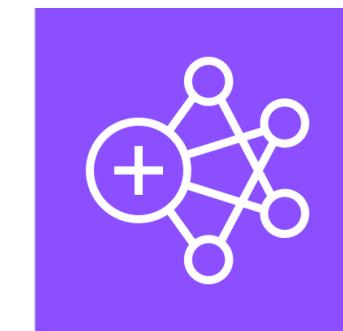
List of AWS Analytics Services Covered



1. Amazon Athena



2. Amazon Redshift



3. Amazon EMR



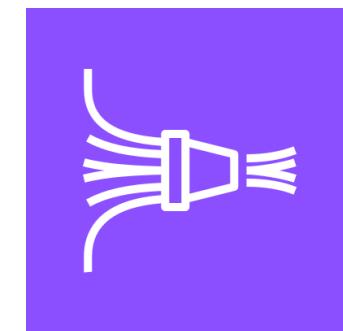
4. AWS Glue



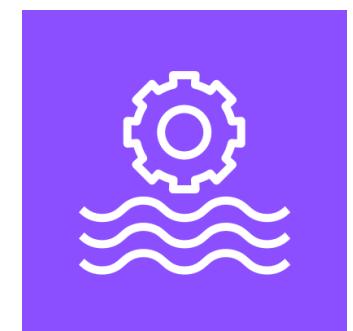
5. AWS Glue Data Catalog



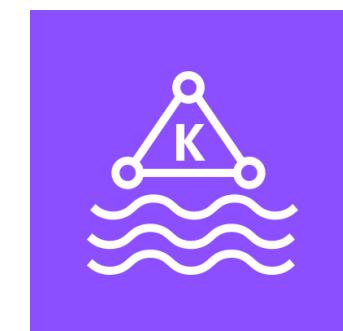
6. Kinesis Data Streams



7. Kinesis Data Firehose



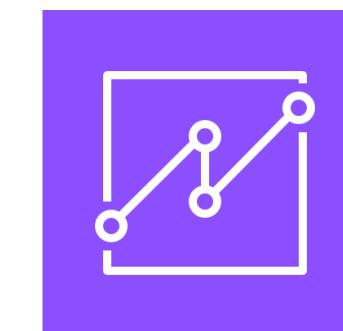
8. Managed Service for Apache Flink



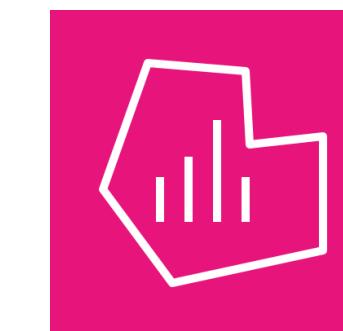
9. Amazon MSK



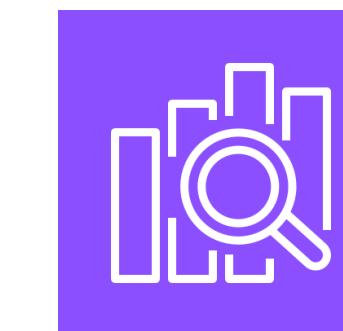
10. Amazon MWAA



11. Amazon QuickSight



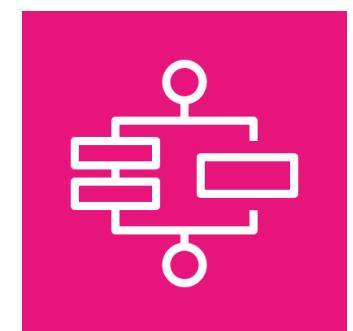
12. Amazon Managed Grafana



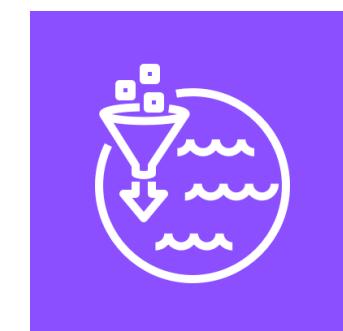
13. Amazon OpenSearch Service



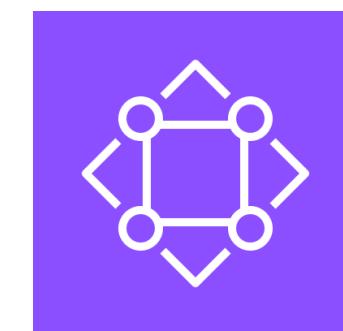
14. Amazon Timestream



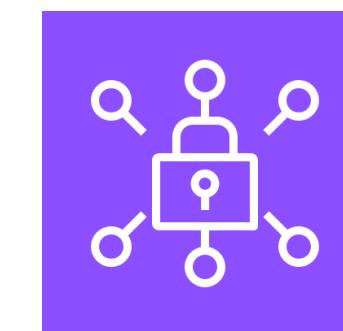
15. AWS Step Functions



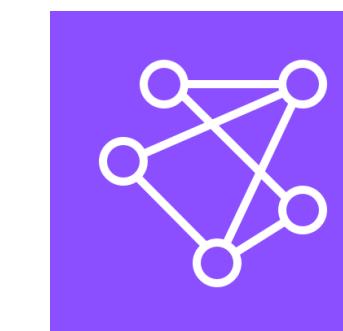
16. AWS Lake Formation



17. Amazon DataZone



18. AWS Clean Rooms



19. AWS Entity Resolution

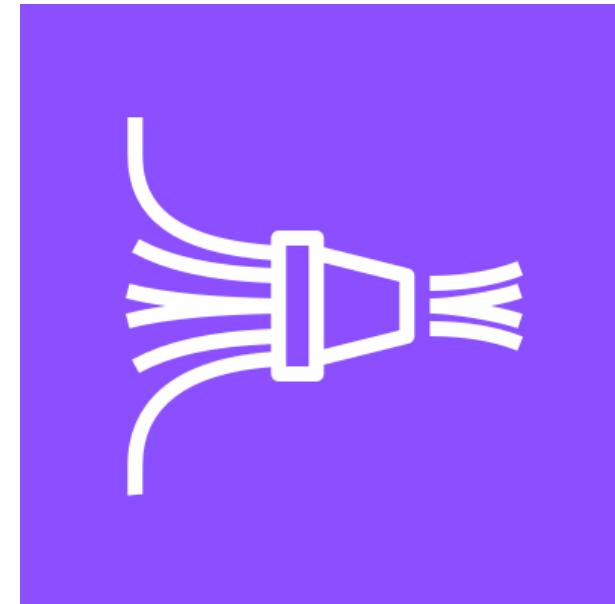


20. AWS Data Exchange

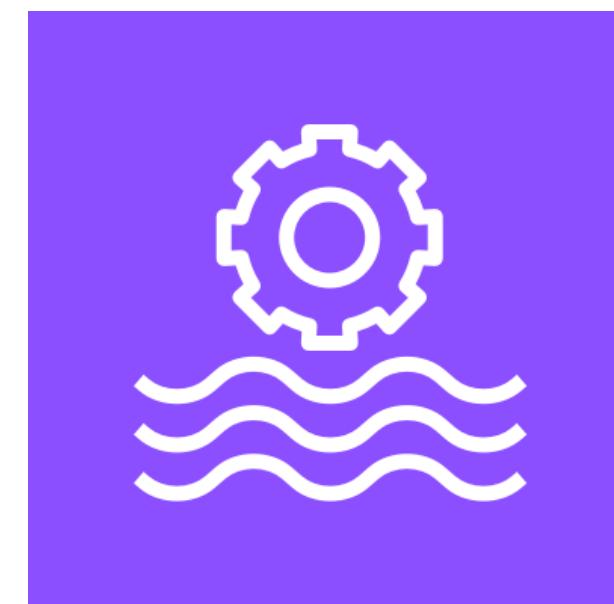
Streaming



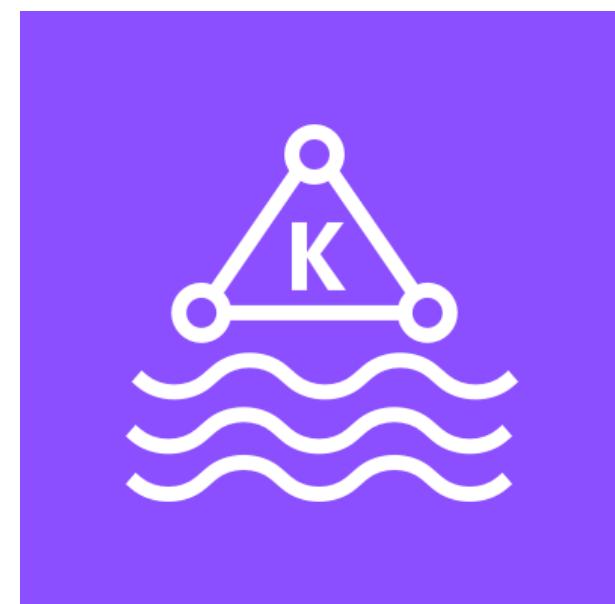
**1. Amazon Kinesis
Data Streams**



**2. Amazon Kinesis
Data Firehose**



**3. Amazon Managed Service
for Apache Flink**

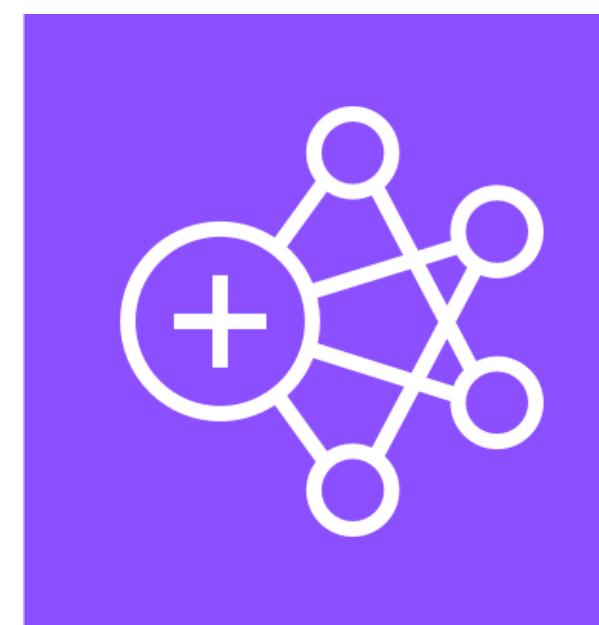


**4. Amazon Managed Streaming
for Apache Kafka (MSK)**

Batch Processing & ETL



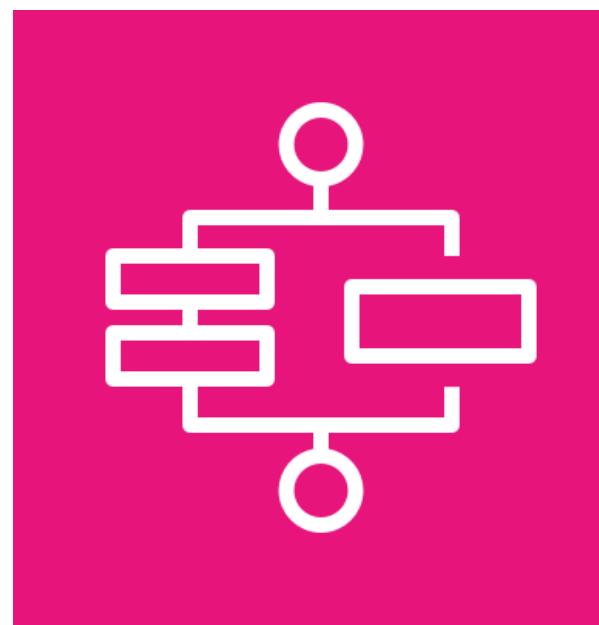
1. AWS Glue



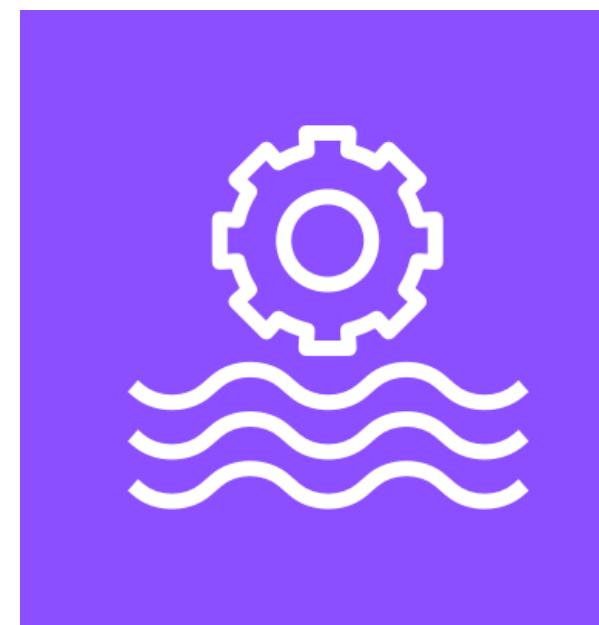
2. Amazon EMR



3. Amazon MWAA

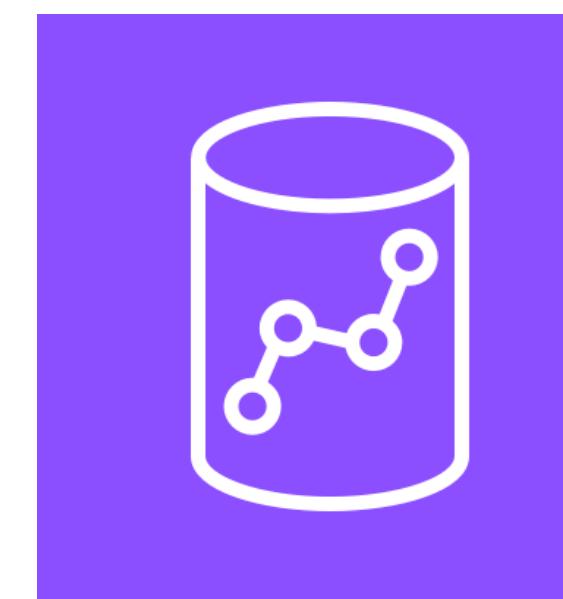


4. AWS Step Functions

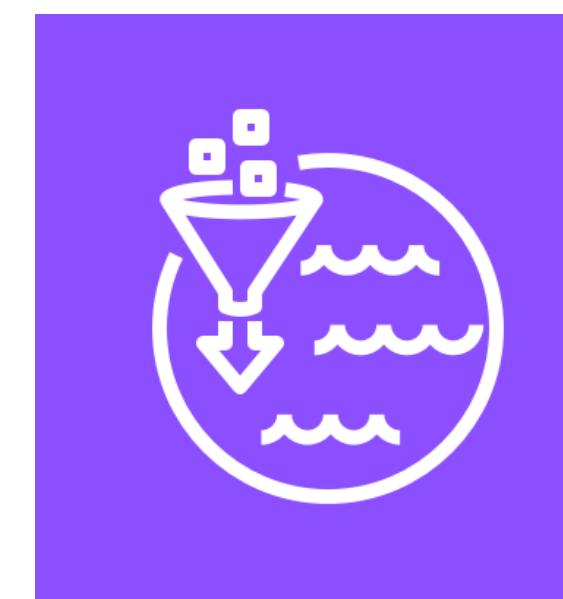


**5. Amazon Managed Service
for Apache Flink**

Data Warehousing & Lakes



1. Amazon Redshift



2. AWS Lake Formation

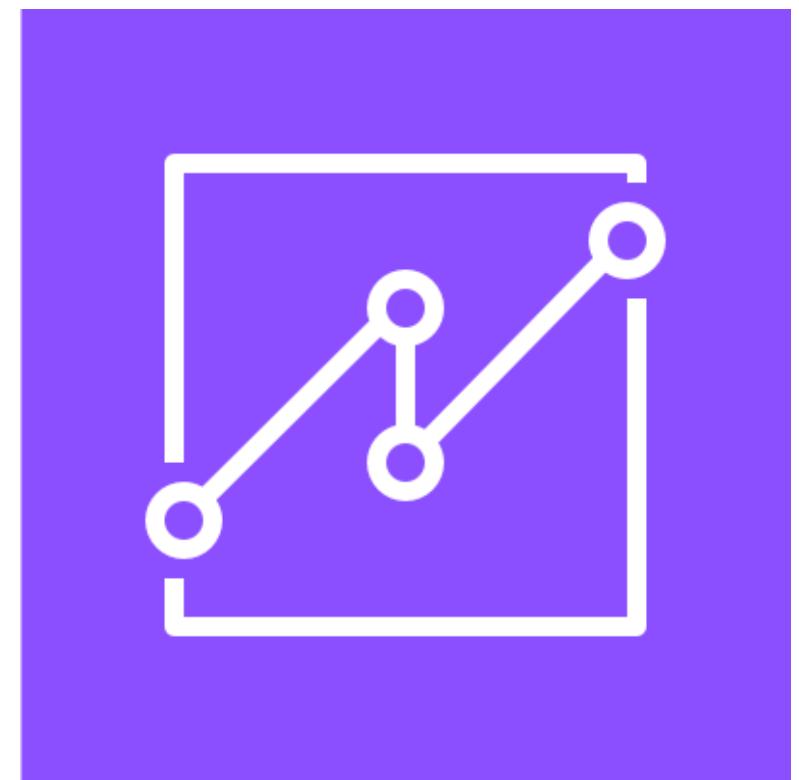


**3. AWS Glue
Data Catalog**



4. Amazon Timestream

Business Intelligence & Visualization



1. Amazon QuickSight

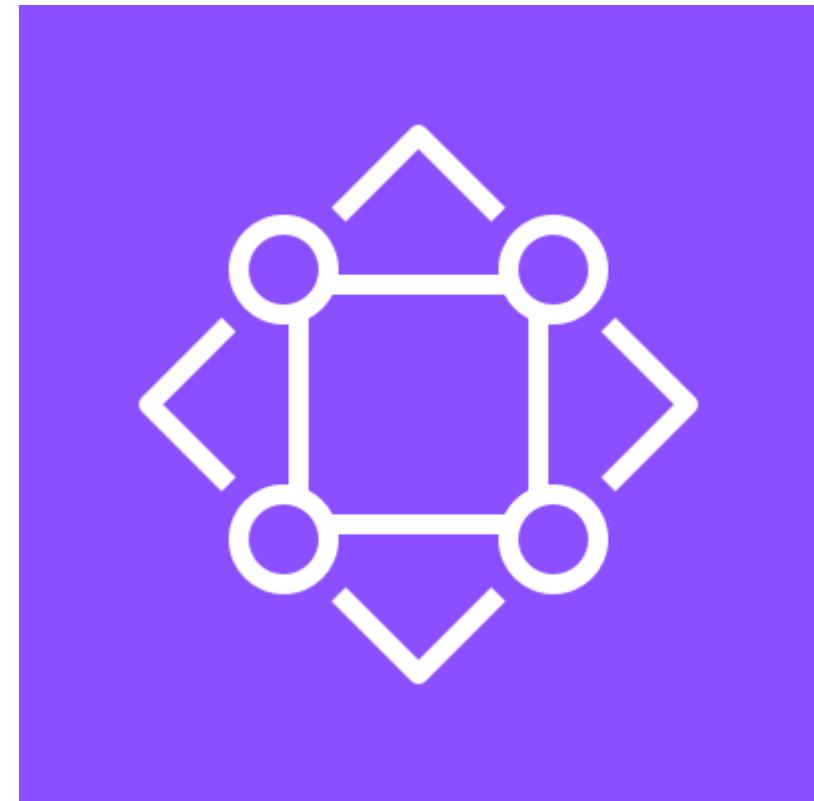


**2. Amazon Managed
Grafana**



3. Amazon Athena

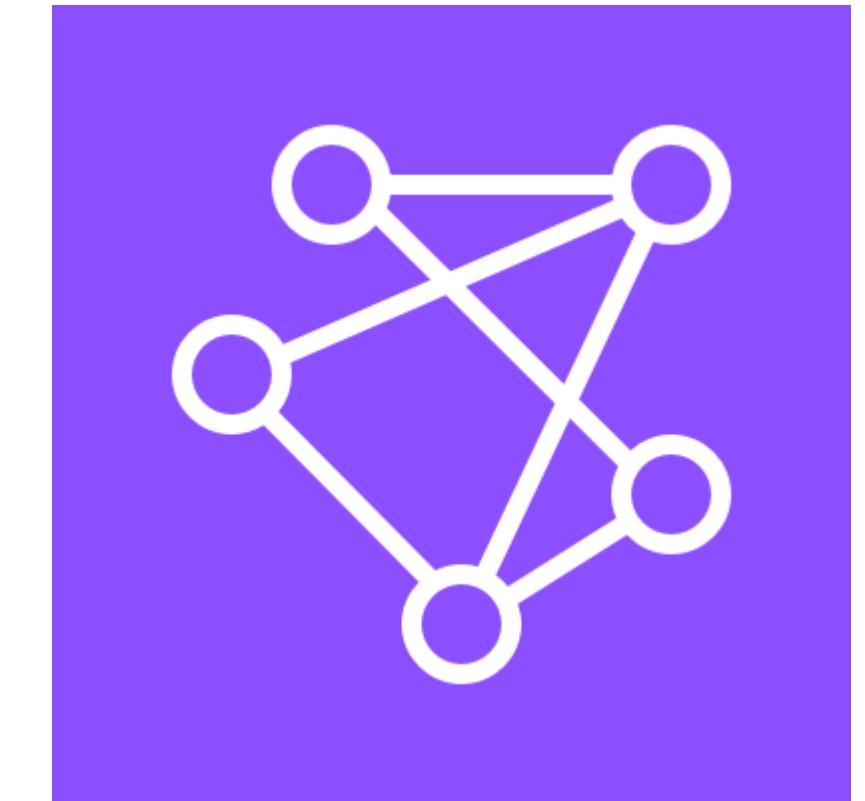
Data Governance & Collaboration



1. Amazon DataZone

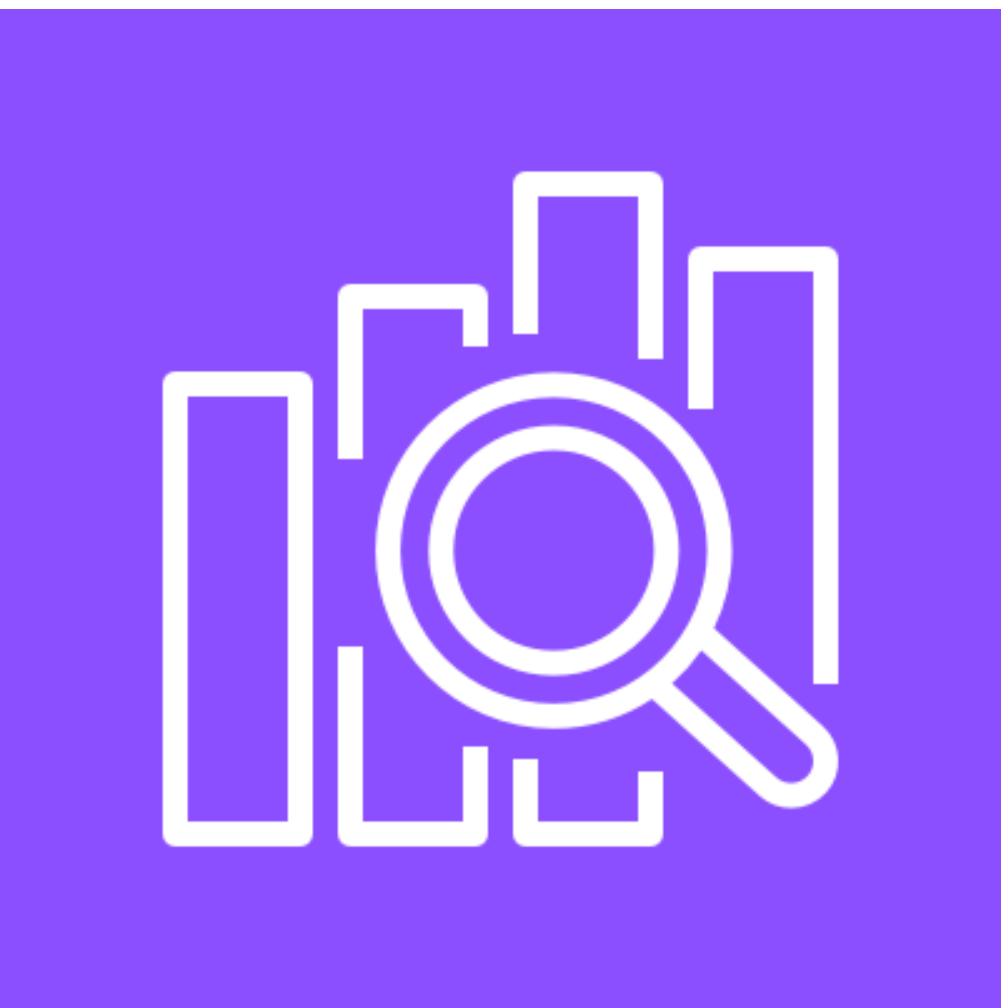


2. AWS Clean Rooms



**3. AWS Entity
Resolution**

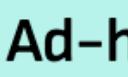
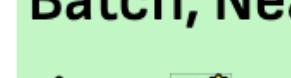
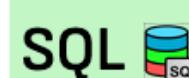
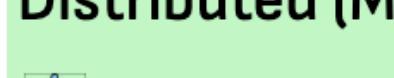
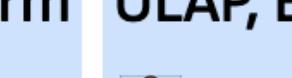
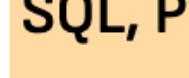
Search & Discovery



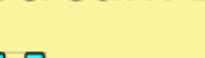
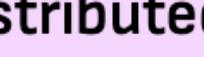
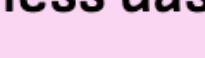
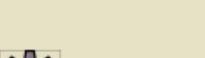
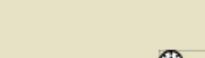
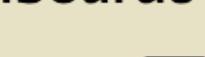
**1. Amazon OpenSearch
Service**

2. Amazon DataZone

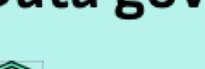
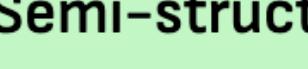
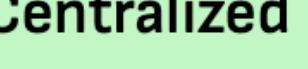
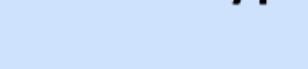
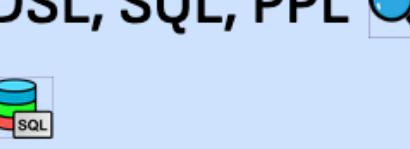
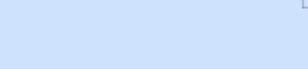
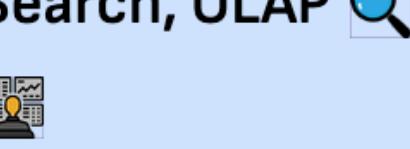
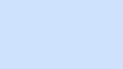
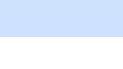
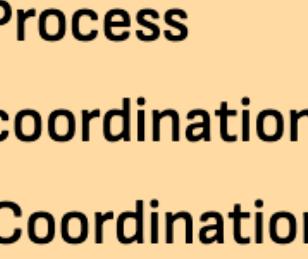
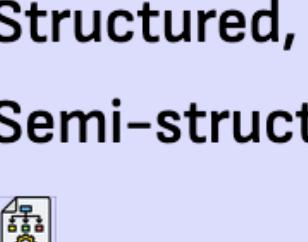
AWS Analytics Services Master Comparison Table

#	Service	Processing Pattern	Data Structure	Query/Data Model	Architecture	Data Storage	Processing Type	Primary Use Cases
1	Amazon Athena	Batch 	Structured, Semi-structured 	SQL 	Distributed 	Data Lake Query Engine 	OLAP 	Ad-hoc analytics  , Log analysis  , S3 data exploration 
2	Amazon Redshift	Batch, Near real-time 	Structured, Semi-structured 	SQL 	Distributed (MPP) 	Data Warehouse 	OLAP 	Enterprise BI  , Complex SQL analytics  , Data warehousing 
3	Amazon EMR	Batch, Streaming  	All data types 	Framework-dependent 	Distributed 	Big Data Platform 	OLAP, ETL, ML 	Big data processing  , ML at scale  , Custom framework needs 
4	AWS Glue	Batch, Streaming  	Structured, Semi-structured 	SQL, Python/Scala   	Distributed 	ETL Service 	ETL 	Serverless ETL  , Data cataloging  , Schema discovery 
5	Kinesis Data Streams	Streaming 	All data types 	N/A (storage layer) 	Distributed 	Stream Storage 	Real-time ingestion 	Real-time data ingestion  , Ingestion, Event streaming  , Multiple consumers 

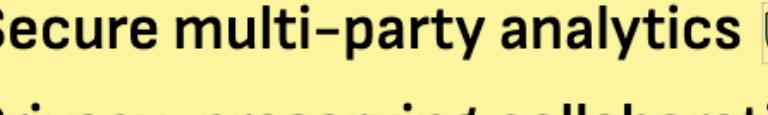
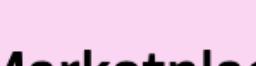
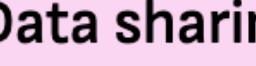
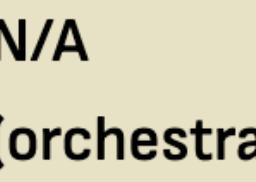
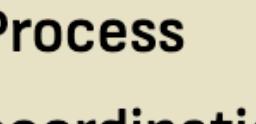
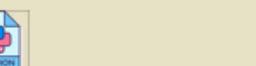
AWS Analytics Services Master Comparison Table

#	Service	Processing Pattern	Data Structure	Query/Data Model	Architecture	Data Storage	Processing Type	Primary Use Cases
⑥	Kinesis Data Firehose	Streaming 	All data types 	N/A (delivery) Delivery	Distributed 	Stream Delivery 	Near real-time ETL 	Stream delivery Delivery, Simple format conversion 
⑦	Managed Service for Apache Flink	Streaming, Batch 	Structured, Semi-structured 	SQL, Java, Python, Scala 	Distributed 	Stream Processing 	Real-time analytics 	Complex event processing  , Stateful analytics  , Time- windowed analysis 
⑧	Amazon MSK	Streaming 	All data types 	N/A (messaging) 	Distributed 	Kafka Service 	Event streaming 	Event backbone  , Stream processing foundation  , Messaging 
⑨	Amazon QuickSight	Batch, Near real-time 	Structured, Semi-structured 	SQL 	Centralized 	BI Service 	OLAP 	Business dashboards  , Self-service BI  , Embedded analytics 
⑩	Amazon Managed Grafana	Near real-time 	Structured, Semi-structured 	Time-series query 	Centralized 	Visualization 	Operational monitoring 	Operational dashboards  , Technical monitoring  , Observability 

AWS Analytics Services Master Comparison Table

#	Service	Processing Pattern	Data Structure	Query/Data Model	Architecture	Data Storage	Processing Type	Primary Use Cases
11	AWS Lake Formation	Batch 	All data types 	N/A (governance) 	Centralized 	Data Lake Governance 	Data governance 	Data lake security  , Fine-grained access control 
12	AWS Glue Data Catalog	Batch 	Structured, Semi-structured 	N/A (metadata) 	Centralized 	Metadata Repository 	Data cataloging 	Central metadata storage  , Schema discovery 
13	Amazon OpenSearch Service	Near real-time, Batch 	All data types 	DSL, SQL, PPL 	Distributed 	Search & Analytics 	Search, OLAP 	Full-text search  , Log analytics  , Observability  , Vector search  Vector
14	AWS Step Functions	Batch, Event-driven 	N/A (orchestration) 	N/A (state machine) 	Distributed 	Workflow Orchestration 	Process coordination 	Service coordination  , Multi-step workflows  , ETL orchestration 
15	Amazon DataZone	Batch 	Structured, Semi-structured 	N/A (discovery) 	Centralized Centralized	Data Discovery 	Data governance 	Business-oriented data discovery  , Self-service governance 

AWS Analytics Services Master Comparison Table

#	Service	Processing Pattern	Data Structure	Query/Data Model	Architecture	Data Storage	Processing Type	Primary Use Cases
16	AWS Clean Rooms	Batch 	Structured 	SQL 	Distributed 	Data Collaboration 	Privacy-preserving analytics 	Secure multi-party analytics  , Privacy-preserving collaboration 
17	AWS Entity Resolution	Batch 	Structured 	N/A (identity resolution) 	Distributed 	Identity Resolution 	Data matching 	Customer 360 views  , Record deduplication  , ID resolution 
18	Amazon Timestream	Near real-time, Batch  	Structured 	SQL 	Distributed 	Time Series Database 	Time-series analytics 	IoT data  , Monitoring metrics  , Time-series analytics 
19	AWS Data Exchange	Batch 	All data types 	N/A (marketplace) 	Centralized 	Data Marketplace 	Data sharing 	Third-party data discovery  , Data access 
20	Amazon MWAA	Batch 	N/A (orchestration) 	N/A (DAG) 	Distributed 	Workflow Orchestration 	Process coordination 	Complex data pipelines  , Python-defined workflows 

AWS Data Streaming Services Comparison

#	Service	Primary Purpose	Data Processing	Latency & Retention	Best For
1	Kinesis Data Streams	Real-time data ingestion  and enabling multiple consumers 	Real-time streaming  ingestion and consumption	Low latency (~70ms-200ms)  24h-365 days retention 	High-throughput data ingestion  with replay capability  and multiple consumers
2	Kinesis Data Firehose	Simple, managed delivery Delivery of streaming data to destinations 	Near real-time delivery  (buffered)	Near real-time (min 60s buffer)  No internal storage 	Simple, zero-code data delivery  to supported destinations   

AWS Data Streaming Services Comparison

#	Service	Primary Purpose	Data Processing	Latency & Retention	Best For
3	Managed Service for Apache Flink 	Complex real-time stream processing using Apache Flink 	Real-time stream processing  , stateful analytics 	Very low latency (milliseconds)  No built-in retention	Complex stateful stream processing  , event-time analytics  , windowing 
4	Amazon MSK 	Managed Apache Kafka service  for event streaming	Real-time streaming  , pub/sub messaging 	Low latency (sub-second)  Configurable retention 	Kafka-compatible applications  , event streaming backbone 

AWS Analytics Query Services Comparison

#	Service	Primary Purpose	Data Location	Query & Processing	Best For
1	 Amazon Athena	Serverless SQL querying  	Directly on S3  (schema-on-read)	Interactive SQL  ANSI Automatic scaling 	Ad-hoc exploration  , log analysis  , serverless data lake querying 
2	 Amazon Redshift	Petabyte-scale data warehouse 	Stored within cluster  or Redshift Managed Storage	Interactive SQL  Cluster nodes or RPUs 	Enterprise BI  , complex SQL analytics  , data warehousing 

AWS Analytics Query Services Comparison

#	Service	Primary Purpose	Data Location	Query & Processing	Best For
3	 Redshift Spectrum	Query S3 data from Redshift 	S3  (queries in-place)	Extension of Redshift  Uses Redshift cluster resources 	Extending data warehouse to lake Data Lake, hybrid queries 
4	 EMR with Presto/Spark SQL	Interactive analytics on big data platform 	S3  , HDFS, or various sources	SQL dialects  Manual cluster sizing 	Advanced customization  , specific frameworks  , large-scale processing 

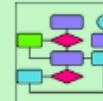
AWS Data Processing & ETL Services Comparison

#	Service	Primary Purpose	Processing Model	Key Components	Best For
1	AWS Glue	Serverless ETL and metadata management	Primarily batch ETL Supports streaming	Data Catalog Crawlers ETL Jobs Data Quality	Serverless ETL Schema discovery Metadata management
2	Amazon EMR	Managed big data processing platform	Batch Streaming Interactive	Clusters Applications (Spark, Hive) Steps	Big data processing ML at scale Custom framework needs

AWS Data Processing & ETL Services Comparison

#	Service	Primary Purpose	Processing Model	Key Components	Best For
3	Kinesis Data Analytics for SQL 	SQL analytics on streams 	Real-time SQL analytics (Legacy)  	SQL applications KPU-based scaling 	Simple stream analytics with SQL  (legacy) 
4	AWS Data Pipeline 	Batch workflow orchestration 	Batch processing (Legacy)  	Data Nodes Activities Resources Schedules    	Legacy batch ETL  (not recommended for new workloads) 

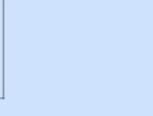
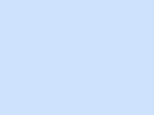
AWS Orchestration Services Comparison

#	Service	Primary Purpose	Definition Method	Workflow Types	Best For
1	 AWS Step Functions	General-purpose workflow orchestration 	Visual editor  JSON (ASL) 	Standard (exactly-once)  Express (at-least-once)  Up to 1 year duration 	  
2	 Amazon MWAA	Managed Apache Airflow service 	Python DAG files 	DAG-based workflows  Based on task configuration 	 Python-defined workflows 

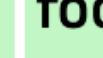
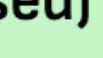
AWS Orchestration Services Comparison

#	Service	Primary Purpose	Definition Method	Workflow Types	Best For
3	 Glue Workflows	Glue-specific workflow orchestration 	Visual in console  API 	DAG of Glue jobs/ crawlers  Simple predecessor/ successor 	Simple Glue ETL orchestration 
4	 AWS Data Pipeline	Legacy batch data workflow service 	JSON pipeline definitions 	Batch scheduled workflows  Simple dependency chain 	Legacy ETL  (not recommended for new) 

Data Governance & Cataloging Services Comparison

#	Service	Primary Purpose	User Focus	Metadata Type	Best For
1	AWS Glue Data Catalog	Technical metadata storage 	Data engineers  Developers 	Technical schemas  Tables and partitions 	Central metadata repository Catalog for AWS analytics 
2	AWS Lake Formation	Data lake governance and security  	Data engineers  Security admins 	Security permissions  Fine-grained access control 	Data lake security  Fine-grained access control 
3	Amazon DataZone	Business data discovery  and governance 	Business users  Data stewards 	Business glossary  Data products and projects  	Business-oriented data discovery  Self-service governance 

Business Intelligence & Visualization Services Comparison

#	Service	Primary Purpose	Data Engine	User Types	Visualization Types	Data Sources	Embedding Capabilities	ML Features	Best For
1	Amazon QuickSight	Cloud-native business intelligence	SPICE (in-memory)  or Direct Query 	Authors and Readers (Standard/Pro) 	Standard business visualizations  , paginated reports 	40+ sources including AWS services  , SaaS, databases 	Full dashboard/visual embedding  , APIs 	Anomaly detection  , forecasting  , Q (natural language) 	AWS-integrated BI  , embedded analytics, pay-per-session model 
2	Amazon Managed Grafana	Operational dashboards and monitoring 	Time-series focused 	Technical users, DevOps 	Time-series  , logs  , traces  , visualization Trace	30+ including AWS services  , Prometheus  , CloudWatch 	Dashboard embedding (iframe) 	Limited (visualization focused) 	Operational monitoring  , technical dashboards 

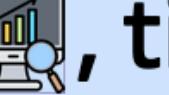
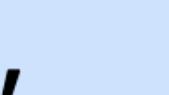
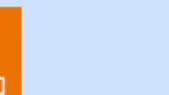
Search & Log Analytics Services Comparison

#	Service	Primary Purpose	Architecture	Query Capabilities	Scaling Model	Use Cases	Integration Points	Limitations
1	Amazon OpenSearch Service	Search, log analytics, observability	Managed clusters or serverless	Full-text, fuzzy, geo, vector search	Instance scaling or OCU-based	Full-text search, log analytics, SIEM, vector search	S3, Firehose, CloudWatch, Kinesis	Management overhead for provisioned clusters
2	CloudWatch Logs Insights	Interactive log analysis	Managed service	Log field extraction, pattern recognition	Automatic	Interactive analysis of CloudWatch logs	CloudWatch, Lambda	Limited to CloudWatch logs, not full-featured search

Specialized Analytics Services

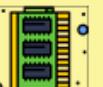
#	Service	Primary Purpose	Key Capabilities	Use Cases	Integration Points	Pricing Model
1	AWS Clean Rooms	Secure multi-party analytics	Cryptographic computing  , differential privacy 	Multi-party data collaboration  , privacy-preserving analytics 	S3  , Glue  , Redshift 	Collaboration compute  , queries  , storage 
2	AWS Entity Resolution	Match and link related records	Rule-based matching  , ML matching 	Customer 360 views  , record deduplication 	S3  , Glue  , identity providers 	Workflow-hour  , records processed 

Specialized Analytics Services

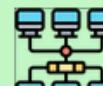
#	Service	Primary Purpose	Key Capabilities	Use Cases	Integration Points	Pricing Model
3	Amazon Timestream	Time series database 	Time-optimized storage tiers 	IoT data  , monitoring metrics  , time-series analytics 	IoT Core  , Kinesis  , Lambda 	Storage-hour  , processing  , query 
4	AWS Data Exchange	Data marketplace 	Discovery  , subscription  , delivery 	Finding and using third-party data 	S3  , Redshift  , Lake Formation 	Provider fees  , AWS fees 

Data Storage & Format Optimization Comparison

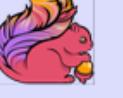
#	Storage/ Format	Best For	Query Performance	Storage Efficiency	AWS Service Support	Trade-offs
1	S3 (CSV/ JSON)	Simple compatibility, human readable	Poor (full scan) 	Low	All AWS analytics services	Easy to use but inefficient for analytics
2	S3 (Parquet)	Analytics workloads	Excellent (column pruning)	Very good	Athena , Redshift Spectrum , EMR Glue	Optimized for analytics but less human-readable
3	S3 (ORC)	Hive workloads	Excellent (column pruning)	Very good	Athena , Redshift Spectrum , EMR Glue	Similar benefits to Parquet, better for Hive/Hadoop

#	Storage/ Format	Best For	Query Performance	Storage Efficiency	AWS Service Support	Trade-offs
4	 S3 with partitioning	Optimizing query performance 	Significantly improved 	Neutral 	All query services 	Requires partition management 
5	 Redshift (internal)	Data warehouse queries 	Excellent for configured workloads 	Good (columnar, compression) Compression	Redshift 	Fast but requires loading 
6	 SPICE (QuickSight)	BI dashboards, visualization 	Very fast for prepared data 	In-memory model (capacity costs) 	QuickSight 	Fast visualization but limited capacity 
7	 Kinesis Streams	Temporary stream retention 	N/A - streaming 	Medium (24h-365d) 	KDS consumers 	Good for replay but time-limited 

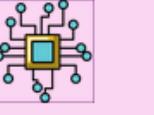
Exam-Focused Decision Matrix

#	Scenario Key Phrase	Primary Service	Secondary Option	Key Distinguishing Factors	Common Exam Distractors
1	"Ad-hoc SQL queries on S3 data" 	 Amazon Athena	 Redshift Spectrum	Athena: No infrastructure, pay per query  . Spectrum: Extension of existing Redshift  .	EMR Hive  (more complex than needed)
2	"Data warehouse for BI reporting" 	 Amazon Redshift	-	MPP architecture  , performance for complex analytics  , BI tool integration	RDS  (OLTP not OLAP), DynamoDB  (NoSQL not analytics)
3	"Real-time stream processing with state" 	 Managed Service for Apache Flink	 Kinesis Data Analytics for SQL	Flink: Complex state  , exactly-once. KDA SQL: Simpler SQL approach  .	Lambda  (lacks sophisticated state management)

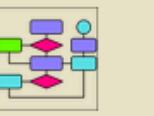
Exam-Focused Decision Matrix

#	Scenario Key Phrase	Primary Service	Secondary Option	Key Distinguishing Factors	Common Exam Distractors
4	"ETL jobs without managing infrastructure" 	 AWS Glue	-	Serverless  , integrated catalog  , visual builder option	EMR  (requires more management), Lambda  (limited for complex ETL)
5	"Apache Spark processing" 	 Amazon EMR	 AWS Glue	EMR: More control  , frameworks. Glue: Simpler, serverless  .	EC2 with Spark  (too much management)
6	"Stream data delivery to S3/Redshift" 	 Kinesis Data Firehose	-	Zero-code delivery  , managed scaling/buffering 	Kinesis Data Streams  (requires consumer development)

Exam-Focused Decision Matrix

#	Scenario Key Phrase	Primary Service	Secondary Option	Key Distinguishing Factors	Common Exam Distractors
7	"Business dashboards and reports" 	Amazon QuickSight 	-	Purpose-built for BI  , SPICE engine  , ML insights 	Managed Grafana  (more technical/operational)
8	"Data lake security and governance" 	AWS Lake Formation 	IAM policies only 	Fine-grained access  , central permissions model 	Direct S3 permissions  (too complex at scale)
9	"Apache Kafka workloads" 	Amazon MSK 	-	Native Kafka compatibility  , MSK Connect 	Kinesis  (different API, not Kafka-compatible)

Exam-Focused Decision Matrix

#	Scenario Key Phrase	Primary Service	Secondary Option	Key Distinguishing Factors	Common Exam Distractors
10	"Python-based workflow orchestration" 	Amazon MWAA 	-	Airflow compatibility  , DAG-based workflows 	Step Functions  (state machine, not Python-based)
11	"Coordinating AWS services" 	AWS Step Functions 	-	Visual workflow  , native AWS integration 	MWAA  (more complex for simple AWS coordination)
12	"Log analytics and search" 	Amazon OpenSearch Service 	CloudWatch Logs  Insights	Full-text search  , visualization  , Machine Learning 	Athena  (can query logs but not optimized for search)
13	"Self-service data discovery portal" 	Amazon DataZone 	Glue Data Catalog 	Business-oriented catalog  , user-friendly portal 	Glue Catalog  (technical metadata, not business-oriented)



**Thanks
for
Watching**