

Home Task 2

Part 1: Technology Selection

1. I choose AWS as cloud provider, S3 as source and destination storage, Amazon Kinesis streams for processing and Amazon Kinesis Data Analytics.
2. Choose justification:

Amazon S3 (Simple Storage Service):

- Scalable Storage: S3 provides virtually unlimited and highly scalable storage capacity, allowing to store large volumes of data, making it suitable for data lakes, backups, and archival storage.
- Data Durability and Redundancy: S3 offers high durability by storing data across multiple availability zones, ensuring data redundancy and availability, making it a reliable choice for critical data.
- Data Consistency: S3 provides strong read-after-write consistency, ensuring that when data is written or updated, subsequent read requests return the latest version.
- Security: S3 supports data encryption at rest and in transit, access controls through IAM policies and bucket policies, and integration with AWS Identity and Access Management (IAM), providing strong security measures.

Amazon Kinesis Streams:

- Real-time Data Streaming: Kinesis Streams enables the collection and processing of real-time data, making it suitable for applications that require low-latency data ingestion and analysis.
- Scalability: I can dynamically adjust the number of shards in a stream, providing automatic scalability to handle data volume variations.
- Data Retention: Kinesis Streams allows to configure data retention periods, ensuring data is available for processing over a specific time frame.
- Fault Tolerance: Kinesis Streams replicates data across multiple availability zones, offering fault tolerance and ensuring data durability.
- Integration: Kinesis Streams integrates with other AWS services like Lambda, Firehose, and Kinesis Data Analytics, making it a key component of many real-time data processing pipelines.

Amazon Kinesis Data Analytics:

- Real-time Analytics: Kinesis Data Analytics enables to process and analyze real-time data streams, making it suitable for applications that require immediate insights from streaming data.
- SQL-Based Processing: It allows to use standard SQL queries to analyze streaming data, which simplifies development for users familiar with SQL.
- Automatic Scaling: Kinesis Data Analytics automatically scales to handle the volume of incoming data, ensuring that your analytics workload can adapt to changes in data volume.

- Integration: It integrates with Kinesis Streams and various output destinations, allowing to easily store, visualize, or take action based on analyzed data.
- Low Latency: It provides low-latency processing capabilities, which is essential for applications where real-time decision-making is critical.
- Built-in Time Series Analytics: Kinesis Data Analytics offers built-in support for time series analytics, making it suitable for use cases like anomaly detection and trend analysis.

Part 2: Data Platform Design:

Design is in PowerPoint file.

Part 3 & 4: Infrastructure as Code IaC / Data Ingestion and Processing:

The script of IaC is in the text file.