Task 3

1. Amazon S3: Amazon Simple Storage Service (S3) is an object storage service known for its scalability, speed, and cost-effectiveness. S3 is like a data lake that can store any volume of data from any source. Data engineers use S3 to create web-based cloud solutions that expand automatically.

2. Amazon Kinesis: Kinesis is a collection of services designed to handle real-time streaming data. Data engineers use Kinesis to create new streams and begin streaming data. It allows for real-time analysis, without waiting for a complete data dump.

3. AWS Glue: Glue is a fully managed ETL service that allows for processing, improving, and migrating data between different data stores and streams. With Glue, data engineers can interactively analyze data, and with AWS Glue Studio, they can visually develop, run, and monitor ETL workflows.

4. AWS CloudWatch: CloudWatch is a monitoring service for AWS resources and applications. It consolidates system, application, and AWS service logs into a single service. CloudWatch Events allows scheduling services to launch during specific time frames.

5. Amazon Redshift: Redshift is a petabyte-scale data warehousing service that enables data engineers to discover insights about their clients and organization. Redshift Serverless allows importing and querying data in the data warehouse, while Query Editor v2 enables building schemas and tables and exploring database objects visually.

6. Amazon IAM: AWS Identity and Access Management (IAM) is a service that allows control over access to AWS resources. It provides features for managing authorizations for actions against services like S3 and SageMaker, and supports identity federation.

7. AWS Lambda: Lambda is a serverless computing service that executes code in response to events and manages the underlying computing resources. It's useful when collecting raw data, as Lambda functions can access API endpoints, process data, and save it to S3 or DynamoDB.

8. Amazon EMR: AWS Elastic MapReduce (EMR) is a service for large-scale data processing using Big Data technologies like Apache Hadoop, Apache Spark, Hive, and more. It allows data engineers to launch temporary clusters to run Spark, Hive, or Flink tasks, define dependencies, establish cluster setup, and identify the underlying EC2 instances.

9. Amazon DynamoDB: DynamoDB is a NoSQL database that supports several data types, including document, graph, key-value, memory, and search. It can store semi-structured data with a unique key and can be used to track the state of other services like Step Functions to prevent race conditions.

10. Amazon Athena: Athena is an interactive query service that allows easy analysis of data in S3 using SQL. Once metadata goes into the Data Catalog, data engineers can use Athena

to extract insights from the data. It's particularly efficient when accessing GBs of data in Parquet format with strong partitions.

Source: https://www.projectpro.io/article/aws-services-for-data-engineering/644