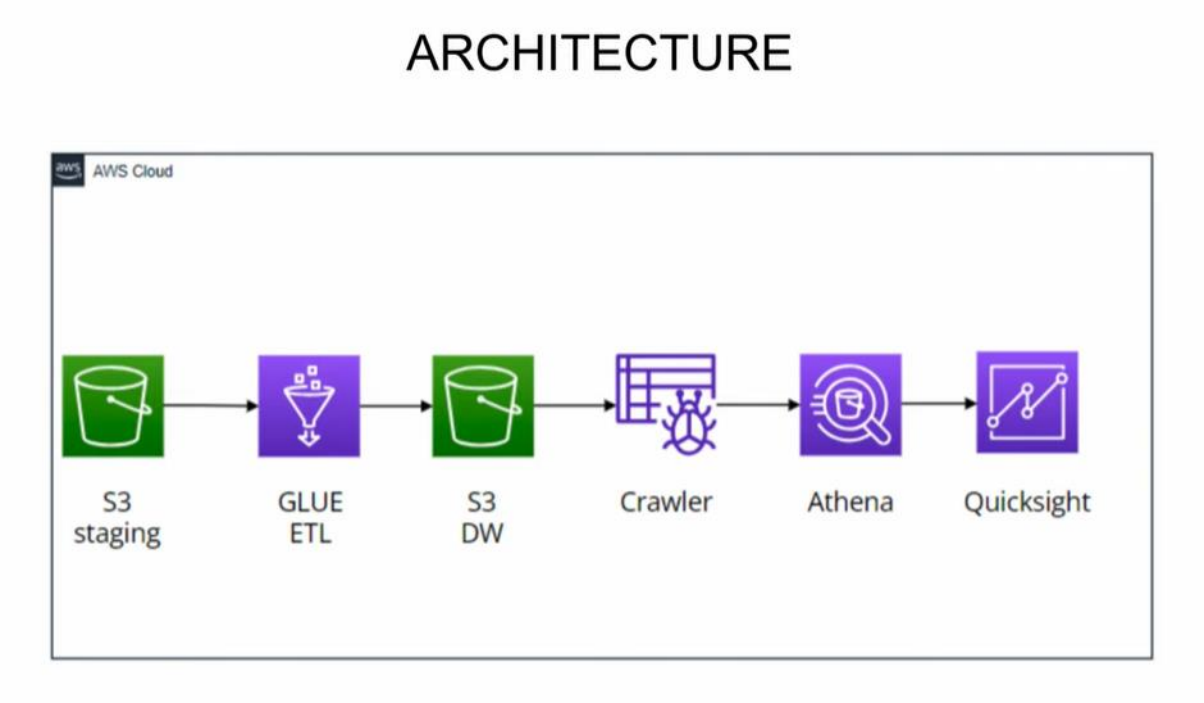
AWS – Spotify Analysis



This project harnesses the power of AWS services to analyze Spotify dataset comprising Albums, Artists, and Tracks. The dataset, stored in three CSV files, is first staged in an S3 bucket. Leveraging AWS Glue, a Visual ETL is created to efficiently transform the data. Jobs are then orchestrated to load the transformed data into another S3 bucket for further processing. A crawler is employed to scan the data warehouse S3 bucket, automatically creating a structured database and tables for seamless querying and analysis.

AWS Athena is utilized to execute queries on the data, with results stored in yet another S3 bucket for easy access. Finally, Amazon QuickSight is employed to visualize the analyzed data, enabling intuitive and insightful analysis. This end-to-end pipeline demonstrates the integration of various AWS services to derive valuable insights from Spotify's rich dataset.

1. **S3 Bucket**:
    - Storing the Spotify dataset in three CSV files.
    - Used as both staging area and data storage throughout the pipeline.
2. **AWS Glue**:
    - Visual ETL creation for transforming the dataset efficiently.
    - Facilitates data transformation and preparation for analysis.

3. **AWS Glue Jobs**:
    - Orchestrate the loading of transformed data into another S3 bucket.
    - Automate data movement and processing tasks.
4. **AWS Glue Crawler**:
    - Scans the data warehouse S3 bucket.
    - Automatically creates a structured database and tables for easy querying.
5. **Amazon Athena**:
    - Executes SQL-like queries on the structured data.
    - Enables ad-hoc querying for analysis purposes.
6. **Amazon Quicksight**:
    - Visualizes the analyzed data.
    - Provides intuitive and interactive dashboards for data exploration and insights.