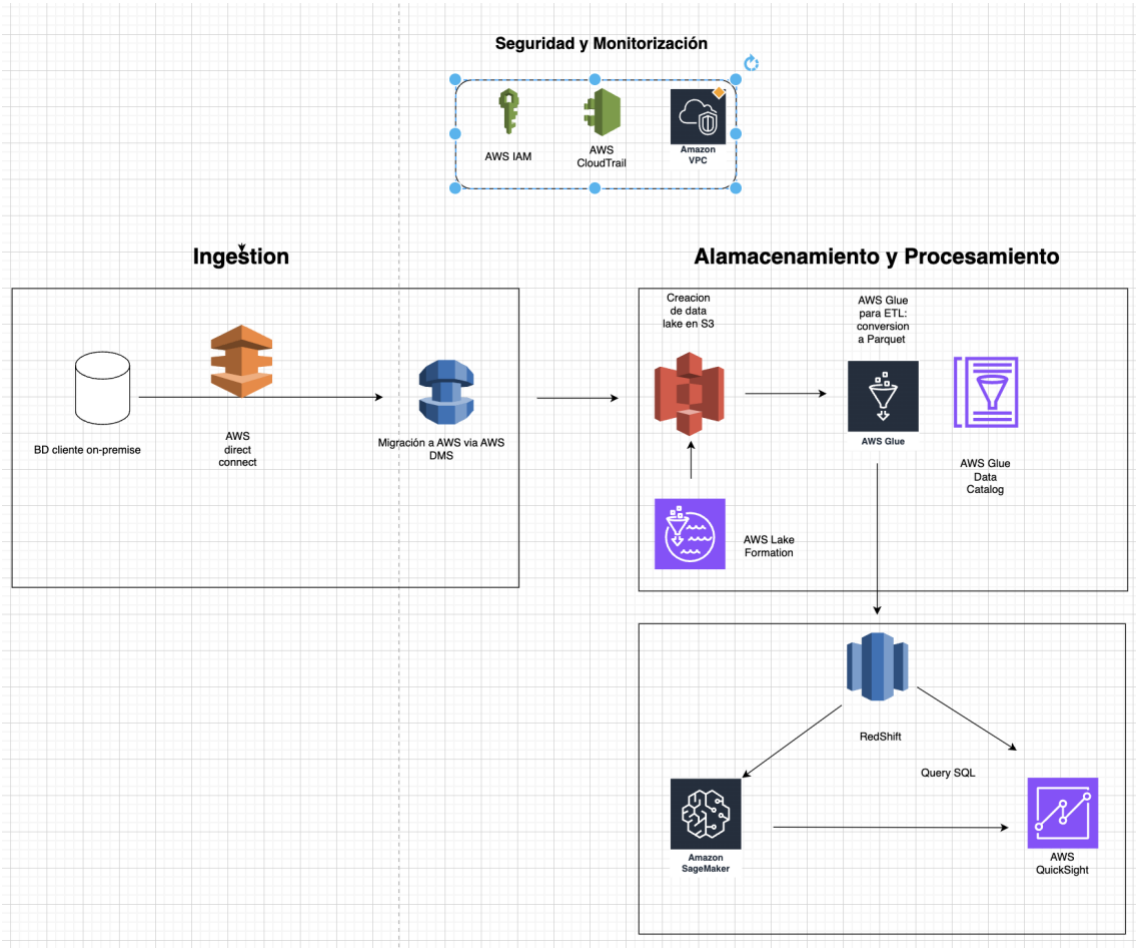
Solutions Architect problem

Problem statement

We are working for a company that wants to migrate its operational databases to the cloud and, in a later phase, make these data available for analytics in a Data Lake. The company aims to maintain as little code as possible. The Data Lake will be consumed by an internal reporting tool and Data Scientists. It is also crucial to consider data governance for regulatory compliance.

Design



Explanation of the Arquitecture:

1. **Secure Connection of the Client's Database to AWS Cloud via Direct Connect.**

2. **Ingestion Phase:**
   - The migration will be performed using **AWS Database Migration Service (DMS)** into a Data Lake created with **AWS Lake Formation** on Amazon S3. A

Data Lake is chosen instead of migrating to RDS or DynamoDB because the company utilizes the data for reporting and data science. This approach provides flexibility to store data in various formats (structured, semi-structured, and unstructured) at a low cost. Additionally, **AWS Lake Formation** enables centralized management of security and data access, ensuring compliance with regulatory requirements.

3. **Storage and Processing Phase:**
    - Since the primary purpose is reporting and data science, **AWS Glue** will be used to perform the ETL operations required to convert the data in the Data Lake into an analytics-ready format, such as Parquet. Glue can also partition or bucket data to improve query performance in **Amazon Redshift**. With **AWS Glue**, ETL pipelines can be created with minimal coding, as it leverages automatically generated PySpark scripts. Furthermore, the **AWS Glue Data Catalog** will serve as a unified source of metadata, facilitating data queries and management in Redshift during the consumption phase.

4. **Consumption Phase:**
    - The data transformed in the previous phase with AWS Glue will be loaded into **Amazon Redshift** for faster analysis via SQL queries. Finally, the company's analysts or data scientists can execute queries on Redshift to create dashboards or predictive models using tools such as **Amazon QuickSight** (reporting) or **AWS SageMaker** (machine learning).

5. **Security and Monitoring:**
    - **AWS IAM**: Defines the permissions and restrictions each employee will have within this AWS ecosystem.
    - **AWS CloudTrail**: Monitors and audits activities in the company's AWS account. This ensures detailed control over what each employee account is doing within the AWS environment developed for the company.
    - **AWS VPC**: Creates a secure and controlled environment in the AWS cloud for managing resources effectively and securely.