**1. PySpark Developer – ETL Modernization (2-3 Positions)**

**Role Summary:**  
This role focuses on building modern, scalable ETL pipelines using PySpark to replace legacy Informatica workflows. The developer will be responsible for transforming complex data logic into distributed PySpark jobs and ensuring high performance and maintainability on cloud platforms.

**Key Responsibilities:**

* Analyze legacy Informatica ETL logic and implement equivalent PySpark transformations.
* Develop scalable and modular PySpark-based ETL pipelines for data ingestion, transformation, and validation.
* Collaborate with data architects to align schema design with modernization goals.
* Perform unit and regression testing to ensure output matches existing Informatica workflows.
* Integrate PySpark jobs into orchestrators such as Airflow and cloud services like AWS Glue.

**Required Skills:**

* Strong hands-on experience with PySpark and distributed data transformation.
* Proficiency in SQL for complex joins, filters, and data validations.
* Exposure to AWS data stack including S3 and Glue is preferred.
* Familiarity with Informatica mappings and workflow patterns is an added advantage.
* Prior experience in ETL modernization/migration projects is desirable.

**2. AWS Data Engineer – Glue, Airflow, S3 (2 Positions)**

**Role Summary:**  
This role is responsible for building and orchestrating cloud-native ETL pipelines using AWS Glue, Airflow (MWAA), and S3 to support the modernization of legacy Informatica workloads. The engineer will help design the cloud execution layer for migrated PySpark-based pipelines.

**Key Responsibilities:**

* Build serverless ETL pipelines using AWS Glue integrated with S3 and PySpark.
* Design and implement workflow orchestration using Airflow (MWAA).
* Support cloud-native data ingestion, transformation, and data lake integration.
* Monitor, troubleshoot, and enhance performance of data pipelines in AWS.
* Collaborate with developers and architects to integrate migrated pipelines into production.

**Required Skills:**

* Strong experience with AWS Glue, Airflow (MWAA), and S3.
* Knowledge of PySpark and Python-based data engineering.
* Familiarity with legacy ETL modernization and migration best practices.
* Understanding of infrastructure-as-code and deployment automation is a plus.
* Experience integrating metadata-driven frameworks is preferred.

**3. Informatica ETL Developer – Migration & Support (2 Positions)**

**Role Summary:**  
This hybrid role involves supporting existing Informatica PowerCenter workflows and actively participating in their transformation to modern PySpark-based architectures. The developer will bridge legacy knowledge with future-state solutions.

**Key Responsibilities:**

* Analyze existing Informatica mappings and workflows to extract business and transformation logic.
* Collaborate with PySpark teams to convert legacy ETL logic into equivalent distributed processes.
* Maintain and troubleshoot current PowerCenter jobs during the migration period.
* Develop SQL-based logic for validations and comparative testing between legacy and modern pipelines.
* Contribute to the development of mapping specifications and transformation documentation.

**Required Skills:**

* Hands-on experience with Informatica PowerCenter (mapping, transformation, workflow design).
* Strong SQL skills for data manipulation, validation, and testing.
* Understanding of ETL migration strategies and patterns.
* Exposure to PySpark or AWS Glue is a significant advantage.
* Ability to work across both legacy systems and modern cloud-based platforms.