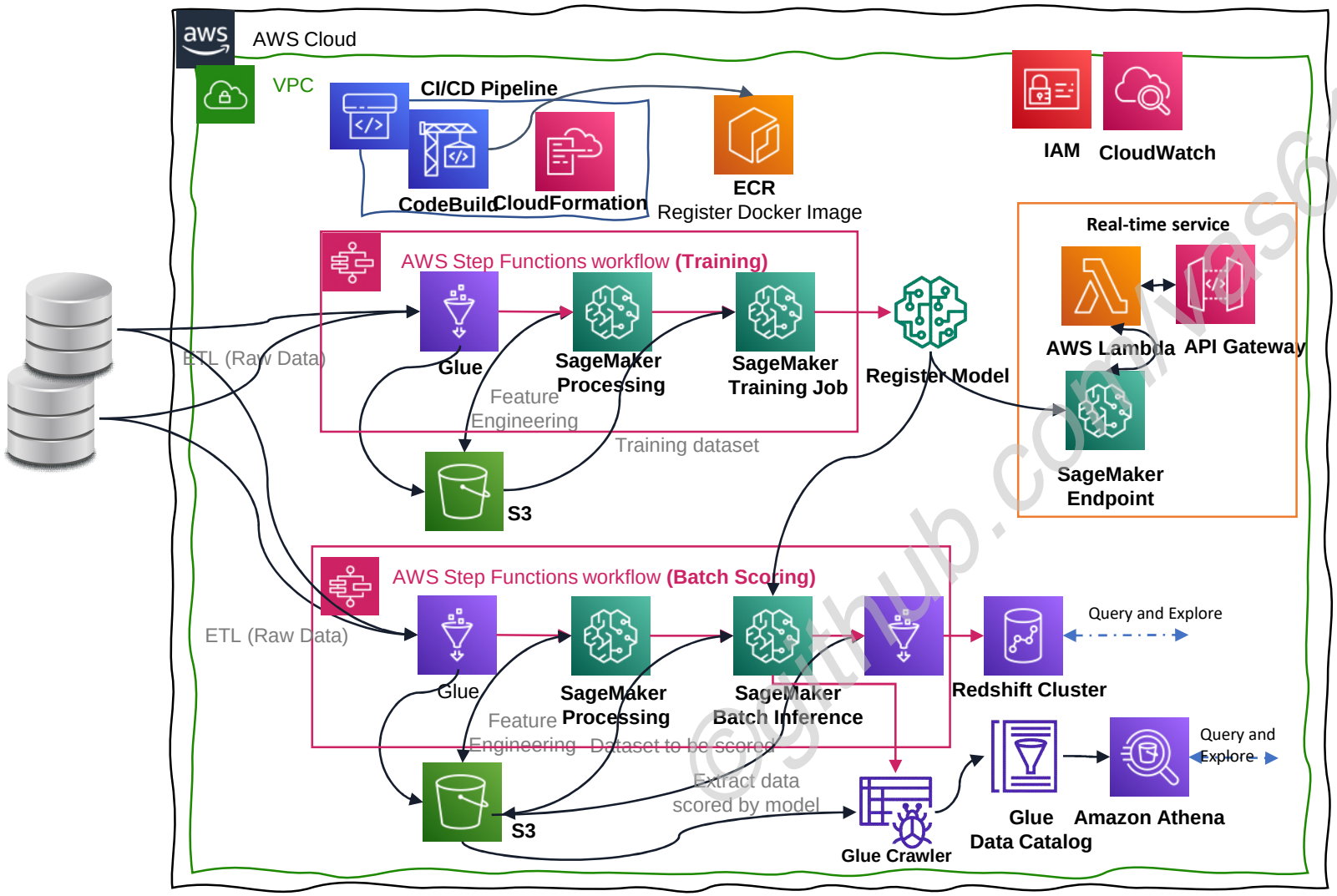
ML using AWS Ecosystem



1. Establish a **CI/CD pipeline** for both training and scoring pipeline. As part of this pipeline
   I. Register docker images required for SageMaker in **ECR**
   II. Use CloudFormation template to define all the resources required, **IAM** roles and policies
2. Use **AWS StepFunction** to orchestrate **training and batch scoring workflow**
3. Model **training pipeline** has the following step
   I. **Data extraction** from lake / warehouse using **Glue**
   II. **Feature engineering** using **SageMaker Processing job** (Spark / Python)
   III. **Model training** using **SageMaker training job** and **Register** the **model** with **SageMaker**
4. **Batch scoring pipeline** has the following step
   I. **Data extraction**
   II. **Feature Engineering**
   III. **Batch Inference** using **SageMaker Batch Transform**
   IV. **Load inference** to a **Redshift cluster** or update a Glue Catalog by running glue crawler
   V. Data Scientist / Engineers / Customers to **consume data** from **RedShift Cluster** Or **Athena query**
5. Use **CloudWatch** to log the **log messages** and **operational, model monitoring metrics**
6. The **Real-time Endpoint** is optional and only required if exposing model via REST API.