**Group 3 Project Synopsis**

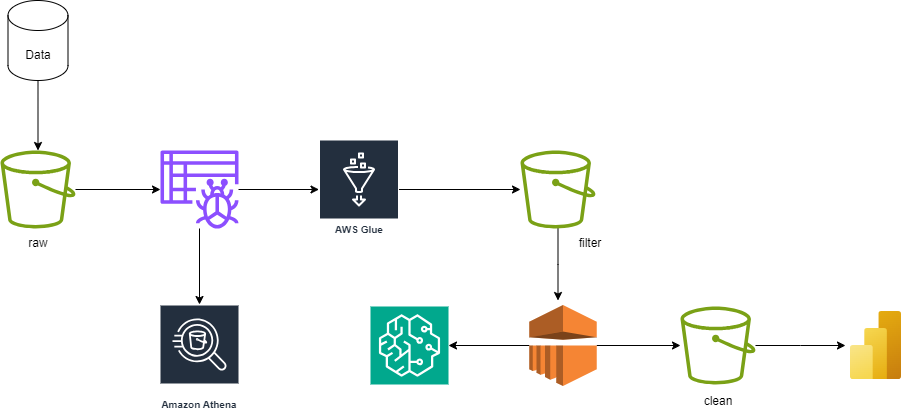
The project uses AI/ML technology to analyze marine biodiversity with the goal of improving conservation efforts along the Indian coastline. We will be forecasting crucial regions along the Indian coastline in this research, which calls for immediate conservation action. The forecast will concentrate on pinpointing areas of notable biodiversity that are in danger due to things like habitat loss, climate change, and human activity. The Ocean Biodiversity Information System (OBIS) dataset, which includes information on different marine species, habitats, and environmental conditions, will be utilized. The objective is to use patterns such as species distribution, small fish abundance, and ecological conditions to identify key areas that need immediate conservation.

To optimize the process, we'll leverage AWS services: Amazon SageMaker for model training and deployment, Amazon EMR/EC2 for data processing and exploratory data analysis, and Amazon S3 for data storage. In order to generate pertinent metrics, the project will also involve feature engineering, model experimentation, performance optimization, and Tableau visualization for insights that can be put to use. This methodology will furnish policymakers and environmentalists with focused suggestions for efficacious conservation tactics.

**Dataset** : [OBIS](https://obis.org/data/access/)

**Architectural Design**

* Data Ingestion and Storage: Amazon S3
* Data Processing and EDA: Amazon EMR/EC2 or Spark Engine
* Feature Engineering: Amazon EMR/EC2
* Model Building and Experimentation: Amazon SageMaker
* Model Deployment: Amazon SageMaker
* Visualization and Reporting: Tableau/Amazon QuickSight



**Future Possible Implementation** (*if time persists*):

1. **Automation and Monitoring:**

AWS Lambda, Amazon CloudWatch

MLOps/ Building CI/CD Pipeline

1. **Enhanced Data Integration and Management**

**Big Data Technologies**: Use advanced big data technologies and services (e.g., AWS Redshift, AWS Glue) to handle and process larger volumes of diverse data sources.

**Data Quality and Management**: Implement improved data quality control and management practices to ensure the accuracy and reliability of the data used for modeling and decision-making

Group Members

* Mandar Gajbhiye (050)
* Ayushree Kamde (009)
* Aryan Parashar (007)
* Gaurav Baviskar (012)
* Yashdeep Shrivastava (046)
* Vrushali Bankar (044)