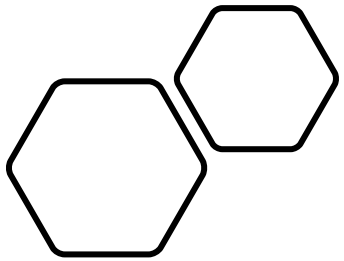
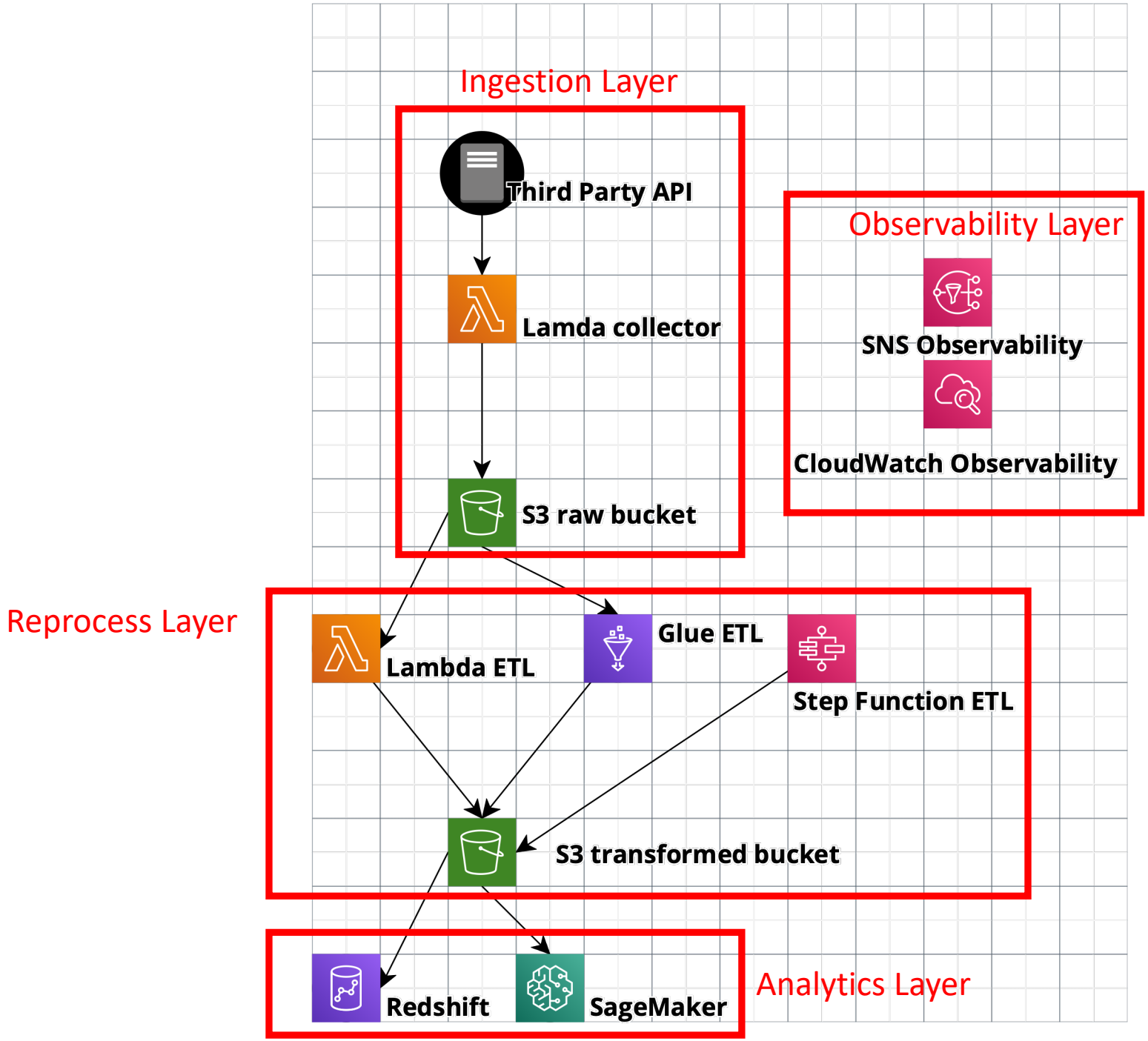
Architectural Proposal

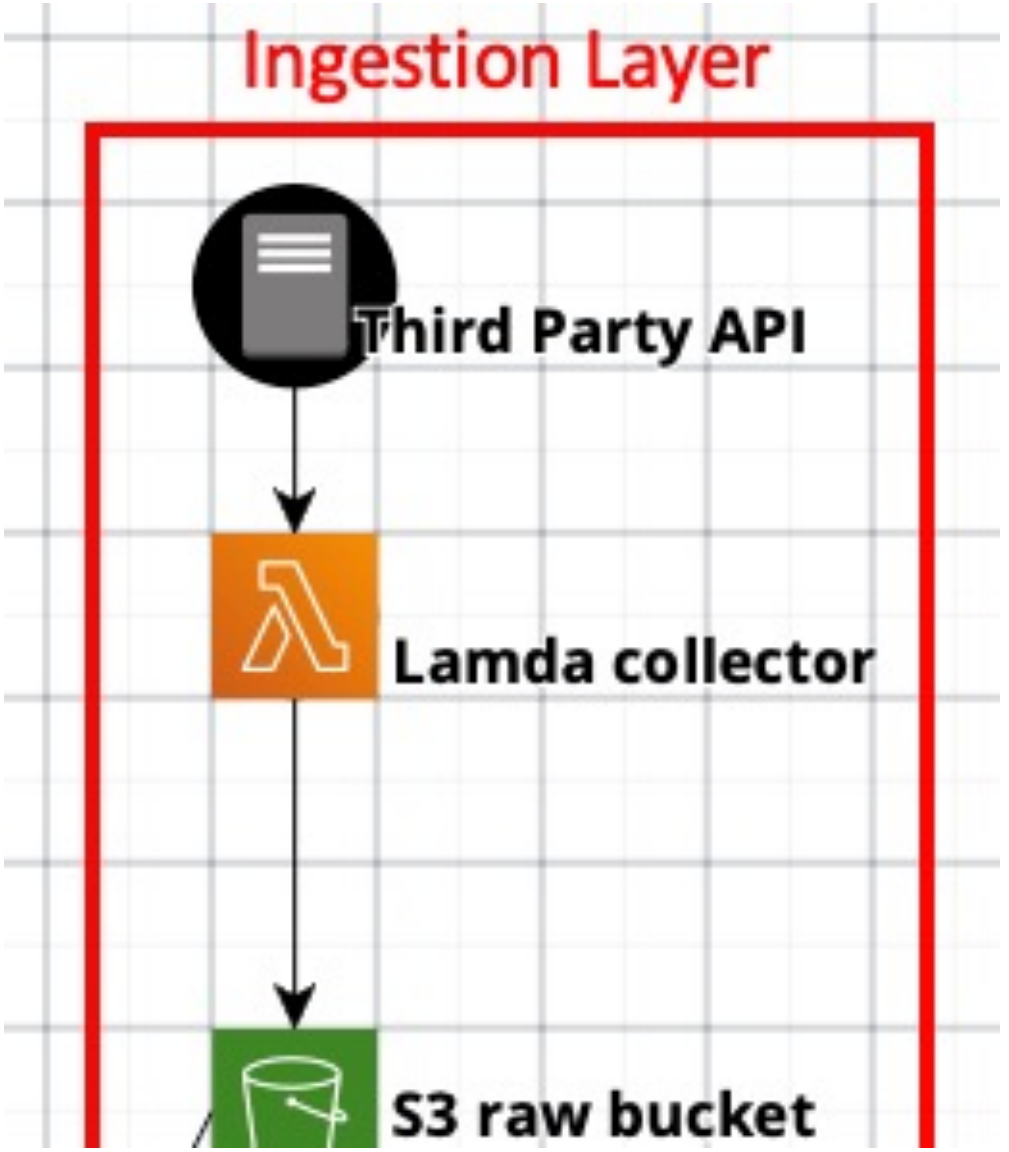


Tiago Melo (tiagotele)


Ingestion Layer
Third Party API
Lamda collector
S3 raw bucket
Observability Layer
SNS Observability
CloudWatch Observability
Reprocess Layer
Lambda ETL
Glue ETL
Step Function ETL
S3 transformed bucket
Redshift
SageMaker
Analytics Layer

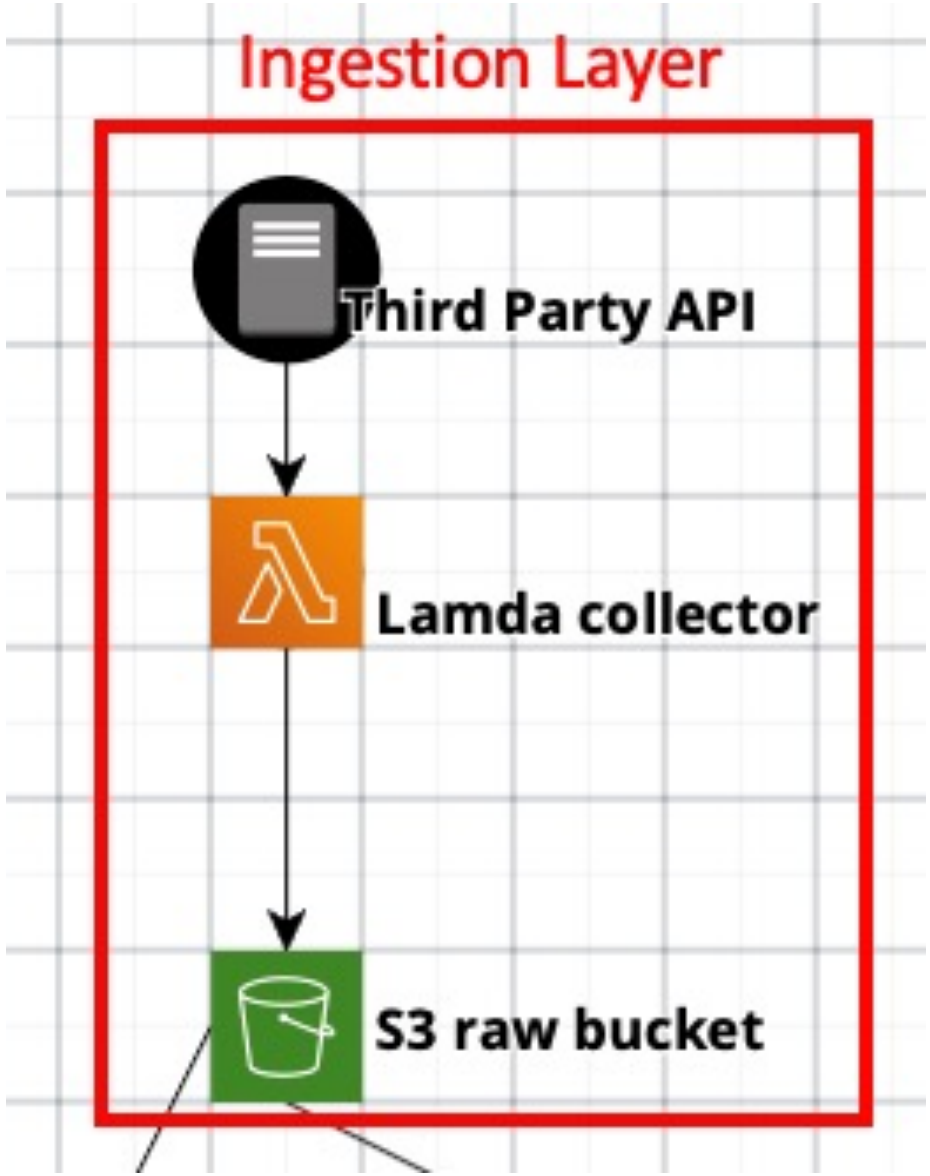
Ingestion Layer

- In this layer, a Lambda will be responsible to fetch data from external API.
- This lambda will triggered by cloudwatch scheduler.
- It will perform parallel requests and save the output on S3 as raw data.
- If something wrong happen it will perform some retries.



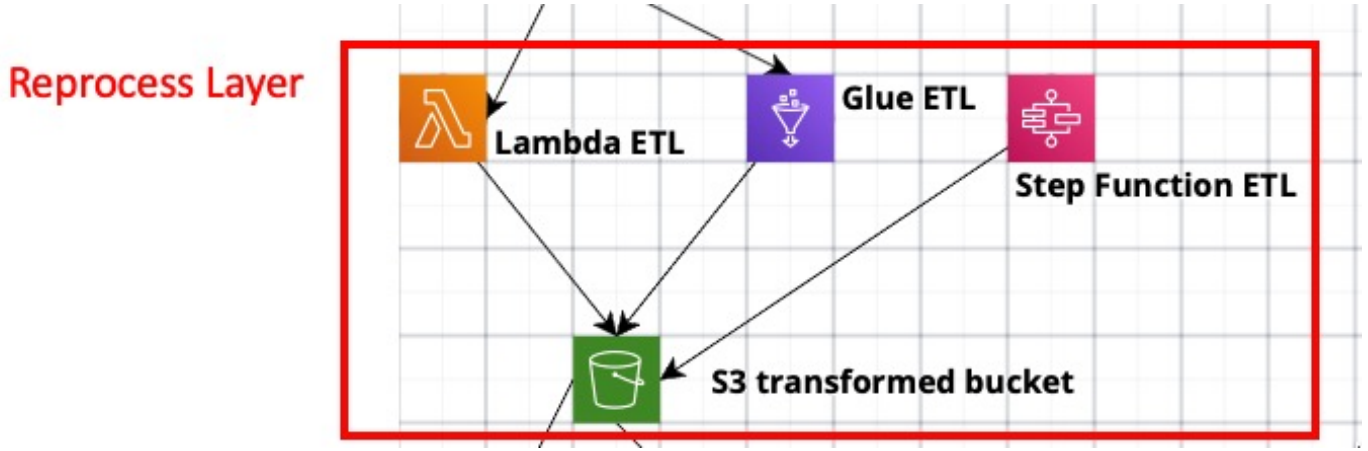
Ingestion Layer

In case of every try fail, it will save on bucket for later analysis.



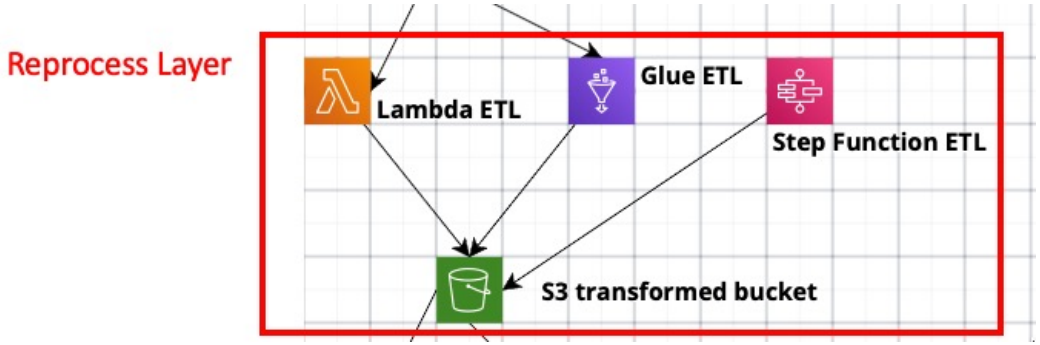
Reprocess Layer

- It will perform transformations on raw data and store it back on transformed bucket.
- There is 3 options depending of kind of data and needs for transform.
- For more basic transformations and simple storage on S3 a Lambda approach seems a better and cheaper solution.



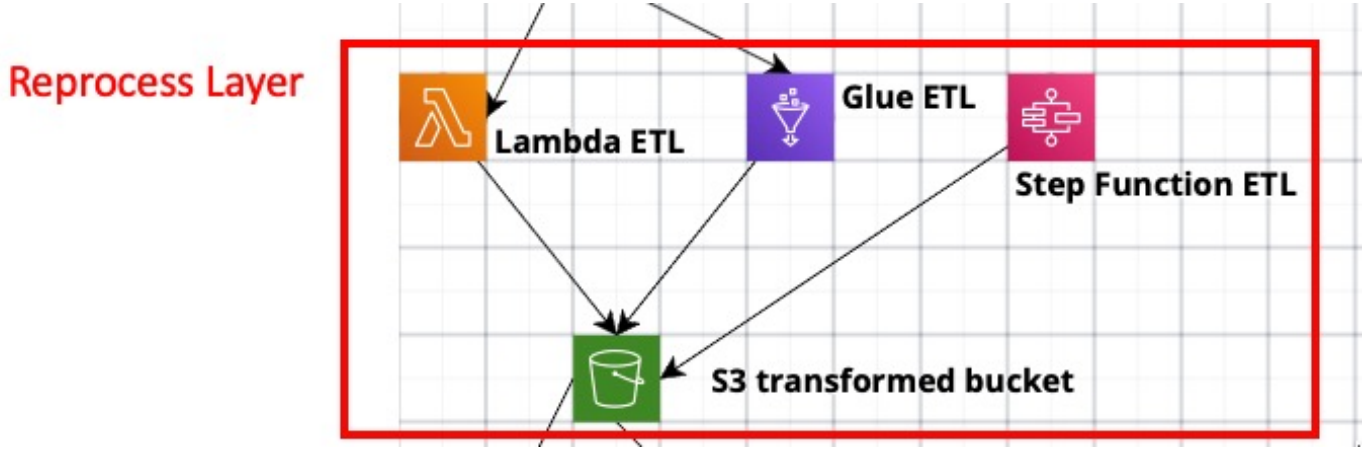
Reprocess Layer

- If there is more conditions/states during the transformations, the Step Functions looks like perfect solution.
- For a massive parallel computing using Spark jobs and Spark api the AWS Glue is gonna to solve most of the problems.



Reprocess Layer

- AWS Glue will take advantage of full set of tools as data catalog, data connectors, embedded Spark api, etc.



Analytics Layer

- On this layer data scientists will be able to explore and analysis data.
- For an initial analytics overview there is a Redshift cluster with transformed data.
- Data scientists might have first impressions of data using SQL queries.

Analytics Layer

- For a deeper data exploration, data visualization and finally Machine Learning modeling, data scientist would make intensive use of Sagemaker environment.

Observability Layer

- This layer will contain dashboard with performance, amount of data loaded/transferred, measurements of redshift and sagemaker clusters, and others.
- Also, will have some alert for undesired conditions, notifying people using SNS notifications.