Logical architecture with tools

A picture containing text, indoor

Description automatically generated

What are the preferred databases?

Datawarehouses like AWS Athena/Redshift can be used .No database required as the data resides on S3. External tables can be preferred

Elaborate the solution flows for both Real time and batch data/processing

Batch Data Processing

If the source is API or database based , AWS Lambda can be used if the data is not more than 1 or 2 gb else we can use AWS Glue if the data volume is high .(morethan 1gb)

Physical architecture/Hosting

The above architecture can be implemented for this requirement

* Scheduling can be done with in AWS lambda (application Specific) or GLUE using Cloudwatch
* Once it is scheduled lambda/Glue will process with the api/api token details or the db/port/table details from the AWS SSM parameter.
* Success or Failure Mail will be triggered to the respective stakeholders via AWS SNS which will be configured inside Glue or Lambda
* AWS Redshift/Athena external table can be created with a simple lambda function-Python and the data can be loaded to the table with transformation logics with another lambda function (python -pandas) or using AWS Glue

Package - **psycopg2,boto3,requests,pandas**

* A Single IAM role will be created for accessing all the services inside AWS and it is applied on the processing AWS service used (Lambda or Glue)

Note: The above can be processed via of EC2 and also via EMR but since there is only one functionality to be achieved ,there is no need of running an instance or a cluster

Which Reporting Tool would be most suitable?

Inside AWS the most recommended tool is AWS Insight , outside AWS we can go for PowerBI ,Tableau or Graphana

How would you design Recommendation Engine for new infrastructure?For recommendation Engine

We Can use AWS Personalize by connecting redshift tables