SYSTEM ARCHITECTURE DESIGN DECISIONS

The document seeks to clarify why certain design decisions were made.

1. Use of AWS Lambda- The lambda function is proposed in order to take advantage of the fact that it can be scheduled and can be initiated at any time with any time intervals. No need to provision a server to make API calls. The lambda executes the API call and delivers raw data to the S3 storage and then dies off. We benefit from cost savings too. No idle resources.

2. Use of scheduled tasks- Task scheduling allows process automation and with the processes separated, depending on load and complexity, one can schedule different tasks to run at different intervals. For example, dataset building will run at a much lower frequency compared to the API request in the lambda.

3. Separation of processes- Process decoupling allows the system to be flexible and individual components can scale as needed without affecting other services. For example, the API call service in the lambda does not need to scale at the same rate as the data set builder.

4. Separation of data- Data separation is a flexibility and future proof feature. It allows the system users to access data at each step in various states. In future, should another system or service need the raw data, it is available. Should another service need the cleaned data, it is also available.

5. Use of S3 storage- The AWS resource is easy to use and the technical expertise on how to use it is readily available. The resource is also cheap and can easily scale with system needs.