<https://d1.awsstatic.com/architecture-diagrams/ArchitectureDiagrams/meter-data-analytics-platform-ra.pdf?did=wp_card&trk=wp_card>

This meter data analytics architecture diagram for Utilities have the following AWS service components:

* Amazon SNS
* Amazon S3
* AWS Data Sync
* AWS Glue
* Amazon Redshift
* Amazon Athena
* Amazon SageMaker -Forecast (AI/ML Predictions)
* Amazon Quicksight, Amazon EMR- (Analytics Services)
* Amazon Pinpoint

**Amazon SNS**- This was used to ingest customer data from multiple sources. Amazon SNS provides high message durability. Every message stored in SNS are available throughout the geographical locations that makes the streaming easier and faster. We won’t be using SNS in our project. Our raw data containing tweet id’s will be fetched from decompressed files using python script and stored in AWS S3.

**AWS Data Sync** – It is used to store and transfer huge amount of data online between on-premises storage and S#. It handles tasks like scripting copy jobs, scheduling and monitoring transfers, validating data and network utilization.

**Amazon S3**- Amazon S3 provides scalability, availability, usability and performance. The Amazon S3 glacier is used for low cost storage for archival copies and retention and compliance. S3 provides easy to use management features to organize and fine tune the data to meet various needs. The diagram shows that the curated data is stored in S3 too. We will be using S3 to store both raw and processed data.

**AWS Glue** – It performs the ETL (Extract, Transform Load) process for this project. It is a fully managed ETL process and it prepares the data for analytics. It stores the metadata in the AWS catalog, which is hassle free and cost effective. The ETL process for our project will involve the following- python script, S3 and Dynamo DB.

**Amazon Redshift** – It is the data warehouse for this project for faster cloud storage capability and it has integrations with data lakes and other AWS services.

**Amazon Athena**- It is used to query huge semi-structured datasets with SQL. In our project we will use python script to query the database (tentatively).

**Amazon SageMaker**- for forecasting energy usage and predictive modelling. Our project will require to use AWS Comprehend for sentiment analysis of the tweets stored in the Dynamo DB. It will return the sentiments associated (positive, negative, mixed, neutral) with the tweets by BatchDetectSentiment.

**Amazon Quicksight, EMR** – This service is used for dashboards, reports building etc. For our project we will be using either Tableau/PowerBI for creating interactive dashboards.

**Amazon Pinpoint** – It is used to measure customer engagement and deliver customer centric experiences across different channels with the help of personalized messages. It is important for providing better customer experience and building stronger relationships.