**Solution Document**

Tech Stack used:

|  |  |
| --- | --- |
| AWS | Cloud platform |
| S3 bucket | Input data storage |
| Boto3 | Amazon library used to interact with AWS services |
| Dynamo DB | Output data storage |
| Python | Text processing |
| Yarn |  |
| Hadoop |  |
| Jupyter Notebook |  |
| EMR |  |