John Doe

+91-1234567879 | john.doe@gmail.com **|** Linkedin | Github

WORK EXPERIENCE

Some Company   **June 2021 – Present**
Data Engineer   *Delhi , India*

- Bank of America (Jan 2022 - May 2022)
  - Orchestrated a pipeline that would extract the DDL and table data of a table from Azure Synapse to ADLS Gen2 to help the client with the efficient backup of the DDL of the SQL tables.
  - Developed data ingestion pipeline using ctas to load terabytes of data more efficiently into the Synapse pool. The pipeline achieved a throughput rate of 10TB/hour which helped the client in loading data swiftly.
  - Developed dashboards and metrics to monitor the dwu used, memory used, and query execution time of the pipelines and SQL server along with different copy methods and settings. It helped the client with managing the cost and resources for the activities and arriving at an efficient and finest configuration for the services.
- AAR Aviation (August 2021 - December 2021)
  - Migration from dataframe/sparkSQL to dataset/spark (Functional Programming using PySpark).
  - Orchestrating data pipelines to ingest data in the Azure Synapse from a lot of varying sources and different formats as the client had an old dataset that needed to be freshened up to update standards.
  - Developed Azure Databricks notebooks which helped with data validation and transformation. This helped in transforming the data into a format that was efficient enough to be stored in Datamart.

Some Company   **May 2020 – June 2020**
Data Analyst Intern   *Delhi, India*

- Developed a web scraper that would scrape posts from different social networking platforms for specific hunt inquiries. The scraped posts were later used to feed into the ML model to generate sentiment reports from them.

PROJECTS

Covid19 Reporting | *Azure DataFactory, Azure Databricks, Azure SQL Server*   Github

- Orchestrated data pipelines to ingest, process, and load data to SQL server.
- Developed transformation solution using dataflow and databricks.
- Ingested data from multiple sources including URL, ADLS Gen2, and Azure Blob.
- Developed schedule and storage event trigger to run pipelines.
- Developed dashboard and alerts to monitor the pipelines.

Data Filtering | *Azure DataFactory, Azure Databricks, Azure SQL Server*   Github

- Orchestrated data pipelines to ingest, process, and load data to SQL server.
- Utilized Azure Key Vault to store credentials and use them in different activities.
- Interpreted the Schema from SQL server and validated if the data matches with the schema using azure databricks.
- Distributed files based on the schema validation and stored them in different containers.

EDUCATION

Some Universty   **2017-2021**
Bachelor of Technology in Computer Engineering CGPA : 8 / 10.00   *Faridabad, India*

SKILLS

Languages: Python, Javascript, SQL, PySpark
Technologies: Node.js, MySQL, MongoDB, PostgreSQL, SQLite, TSQL, Apache Spark
Developer Tools: Git, VS Code, Postman, SMSS, Azure Storage Explorer
Cloud Tools: Azure Synapse, Azure Data Factory, Azure Databricks, GCP, AWS

CERTIFICATIONS

AZ-900(Azure Fundamentals), DP-900(Azure Data Fundamentals), DP-203(Azure Data Engineer)