Varun Chitale
Data Engineer

SUMMARY:
With focused experience of 5 years on IBM and AMD big data projects as a Software Engineer, I excel in interacting with diverse technologies, completing tasks, and collaborating within teams. My background in software quality assurance and proficiency in Python, SQL, and PySpark enables me to tackle complex challenges with efficiency.

TECHNICAL SKILLS
- **Big Data Technologies:** Hadoop, Hive, Spark,
- **Cloud Technologies:** Azure Databricks,Azure Data Factory,Azure Synapse,ADLS
- **Programming Languages:** Python, PySpark
- **Version Control Tools:** Git
- **Database Management Systems:** PostgreSQL, MySQL
- **Project Management & Tracking:** Jira

TRAININGS AND CERTIFICATIONS
Big Data Spark And Hadoop Developer Skills**:** HDFS,Spark,Pyspark,Python,Sql

EXPERIENCE
Infobell It Solution Data Engineer- May 2019 to Present

RELEVANT PROJECTS:
Project Name : Data Migration from OnPrem to Azure Pipeline
Duration: May 2021- Ongoing
Description: The aim of the project is to build a scalable, real-time data pipeline on Azure, replacing NFS with Azure Storage, automating ingestion, transformation, and storage, and enabling efficient querying and analytics.
Technology :- Azure Data Factory, Databricks (Spark) Data Lake Storage, Postgres, Dremio, Data Explorer, APIs.
Contributions:
- Implement data ingestion from external sources via ingress APIs using **Azure Data Factory**.
- Migrate from NFS storage to **Azure Data Lake Storage** for scalable and secure data storage.
- Design Spark-based data transformations within **Azure Databricks** and store data in **format**.
- Configure **Dremio** for fast querying of data stored in **Azure Data Lake Storage**.
- Oversee and optimize Azure cloud infrastructure, including **Azure Data Factory**, **Azure Databricks**, and **Azure Data Lake Storage**, ensuring performance and cost-efficiency.
- Collaborate with cross-functional teams to ensure smooth deployment and operation of data pipelines across the **Azure ecosystem**.
- Document the data pipeline process and provide regular reports on system performance and data quality.
- Monitor scale, and optimize data pipelines for high availability and reliability using Azure monitoring tools.

Project Name: Carbon Accounting and Reduction Emission (CARE Project)
Duration: May 2019- 2021
Description: Led the development of a new feature on the IBM Cloud platform to calculate carbon emissions for all services, enabling enterprises to make more sustainable business decisions.
Technologies: Pyspark, IBM Cloud, Python, SQL, IBM DB2, Pytest, Power Query, Excel, Git
Role: Data Engineer
Contributions:

- Scheduled and led meetings with service owners to understand requirements, initiating and maintaining comprehensive project documentation.
- Assisted in the development of individual services and implemented **ETL** processes using IBM DataStage, optimizing data quality and accessibility.
- Performed manual data validation using **Power Query**, ensuring alignment with code outputs.
- Collaborated with data engineers to develop and maintain **robust data pipelines** capable of ingesting diverse data formats (**Parquet, JSON, CSV, Avro**).
- Utilized **IBM DataStage** activities to construct and optimize data pipelines.
- Executed SQL query optimization to enhance database performance and used Pandas for data cleansing, ensuring high data quality and integrity.
- Scheduled and monitored monthly jobs using **Code Engine**, ensuring timely and accurate data processing.
- Performed **unit testing** on codebases and provided production support, ensuring smooth operations.
- Developed a fully functional **PySpark** codebase for cloud-based data services
- Actively participated in agile development processes, including sprint planning, daily standups, and retrospectives, contributing to continuous improvement.

EDUCATION:

2019 | Pune University
Bachelor of Engineering B.E - Electronics Telecommunication