TEPPALA SATEESH

+91-9885501150 | sateeshteppala1@gmail.com | LinkedIn | GitHub |

SUMMARY

Experienced data engineer skilled in backend development, REST API, and crafting robust data pipelines for effective data processing and management.

TECHNICAL SKILLS

- **Big Data** - Apache Spark, Spark SQL, PySpark
- **Backend** - FastAPI, Flask
- **DevOps** - Docker, Kubernetes, IBM Cloud, Apache Airflow, AWS, TravisCI
- **Databases** - IBM DB2, PostgreSQL, RedShift, MySQL, MongoDB, Redis
- **Dashboard** - Apache Superset, Streamlit
- **Languages** - Python, Scala, SQL

EXPERIENCE

IBM India Pvt Ltd. **Bengaluru, India**

Data Engineer/ Backend Engineer January 2022 - Present

- Work as part of an agile team in the Finance and Operations domains.
- Find flaws and regularly occurring problems with the programme and fix them with an efficient problem-solving and change-management procedure.
- Interact with business and interfacing teams to address issues affecting the existing system and improve current application code.

Project:

Client- N/A (IBM Internal)

- Working in Marketing Platform. It serves as the source for all levels of marketing data: the individual, the company, and the linkage between them.
- Migrated existing DataStage jobs to Apache Spark using the ETL-Framework, a proprietary tool built with Spark 3.0.1 and Scala 2.12, to improve data processing efficiency and scalability.
- Developed and implemented automation solutions for development team using Python and DB2 to enhance productivity and reduce manual efforts.
- Created and deployed real-time monitoring dashboards in cloud for operations and development team using Python and DB2 to oversee job performance and identify issues.
- Worked on different file formats such as Avro and Parquet using Spark methods to optimize data storage and access.
- Implemented CI/CD pipelines using Travis CI, GitHub, and IBM Cloud to automate code testing, integration, and deployment.
- Deployed and managed Spark jobs on Kubernetes clusters using spark-operator.
- Developed and deployed REST API (Application Programming Interface) services using FastAPI for downstream teams to consume and integrate data.

PROJECTS

[Link] Data Pipeline ETL with PySpark and AWS -

Tech Stack: PySpark, AWS, AWS S3, SparkSQL, Apache Airflow, Python

- Build an end-to-end data pipeline that extracts data from different sources, transforms it using PySpark, and loads it into Amazon S3.
- Ingest data from various sources (e.g., CSV, JSON, or relational databases) into an S3 bucket.
- Use PySpark to perform transformations on the raw data (cleaning, filtering, aggregating). Load the processed data into AWS S3 bucket.

[Link] Dynamic Data Generation and RESTful API Delivery -

Tech Stack: Python3, FastAPI, PostgreSQL, Vercel

- Leverages FastAPI to dynamically create and serve data through a RESTful API.
- Retrieves a portion of the dynamic data from a PostgreSQL database table.
- Ensures seamless and efficient delivery of both dynamically generated and database-sourced data to clients through the API.

EDUCATION

Gayatri Vidya Parishad College of Engineering (Autonomous), Vishakhapatnam

Bachelor of Technology in Computer Science and Engineering (2018 - 2021)

GPA : 7.4