

Bhargav Teja J

571-355-3440 | jakku.bhargavteja@gmail.com | [Linkedln](#) | [Portfolio](#)

EXPERIENCE

Data Engineer

May 2019 – June 2021

Clorder Technology Inc.

Hyderabad, IND

- Deployed Google Analytics to collect terabytes of raw data and scheduled batch loading into Big Query by designing ETL data pipelines with Google Cloud Composer (Airflow) to transform JSON, XML files
- Collaborated cross-functionally with Data science teams and Marketing stakeholders, to develop 10 KPIs tailored to business objectives, integrated KPI queries into data transformation processes for better analytics
- Assisted in designing and optimizing relational databases, ensuring efficient data storage and retrieval
- Connected BigQuery and Tableau to build 24 dashboards, measuring performance metrics on marketing campaigns thereby improving global revenue and customer satisfaction
- Resolved data inconsistency by validating scripts in SQL Server, and saved over 30hrs/week
- Improved query efficiency by indices optimization, reduced costs spent by 1/3 on allocated resources
- Streamlined CI/CD and version control for Python code using Git and managed workloads using CRON jobs
- Developed on-demand Ad-hoc reports for Leadership on key outcomes, supporting data-driven growth strategies

PROJECTS

Covid-19 Data Integration | PostgreSQL, Talend, Big Query, GCP

Apr 2023

- Defined entities to create EER's and relational models to analyze EMR data (1M records) using PostgreSQL
- Performed ELT operations using Talend and modeled data using snowflake schema and OLAP data market
- Built robust pipeline by loading data to GCP buckets using shell, python and transformed data using Data Fusion
- Hosted data on CloudSQL, executed federated queries for rapidly changing data from Pub/Sub stream
- Ingested data to BigQuery using Data stream from local server to maintain data freshness, CDC and SCD's
- Diagnosed clinical dashboards in Looker for better healthcare support and managed access control using IAM

Graph Based Search Engine | AWS, PySpark, Neo4j, Web Dev

May 2023

- Engineered ETL pipeline using AWS Glue to extract raw faculty data from the university website to S3 buckets
- Transformed raw data using PySpark for optimal graph modeling and ingested into Neo4j Graph Database
- Modeled GraphDB with COSINE similarity and visualized research data in Neo4j Bloom, saving 50 hrs/week.
- Designed Web Application with 100+ faculty users to update information into the database using Restful APIs

Covid-19 Prediction Analysis | Python, Machine Learning

Dec 2022

- Conducted statistical analysis for critical patterns and visually interpreted data to reveal clinical assessment trends
- Forecasted virus outbreak transmission using Time Series Model(RNN) and recommended healthcare strategies

School Performance Analysis | SQL, Python

Dec 2021

- Collected and Performed EDA on School Card Reports using SQL and Python to improve Education quality
- Leveraged JOIN, CTEs, and window functions to measure the school's performance for better decision-making
- Visualized the key findings in Python and came up with strategies that led to a 40 percent increase in performance

EDUCATION

George Mason University

Fairfax, VA

Master of Science in Data Analytics

Aug 2021 – May 2023

Keshav Memorial Institute of Technology

India

Bachelor of Technology in Electronics and Communication Engineering

Aug 2016 – April 2020

TECHNICAL SKILLS

Languages: Scala, Python, SQL, R

Databases: NoSQL, T-SQL, SQL Server, PostgreSQL, Db2

Big Data Tools: Hadoop, Hive, Apache(Airflow, Spark, Kafka), Kubernetes,Docker

Cloud Services: GCP(BigQuery,Data Fusion,CloudSQL, Cloud Composer) AWS(Redshift,S3,EC2,Glue), Neo4j

Data Warehouse: Snowflake, Big query, Databricks,Redshift

Developer Tools: Git, Tableau, VS Code, Microsoft Suite, Agile(SAFE), Bash shell Scripting, UNIX

Libraries: Pandas, NumPy, MATPLOTLIB, Scikit-learn, TensorFlow, Seaborn, Pyflux