Rohith Kota

• San Francisco, CA • Kotarohit711@gmail.com • +1 (857)335-3738• LinkedIn • Open to Relocate

"Highly determined **Data Engineer** with **5+** years managing **petabytes of data**. Experienced in building **Scalable Data Lakehouses**, **200+ Batch** and **Streaming ETL pipelines** and **Big Data Infrastructure** to enhance data accessibility and reliability"

SKILLS

Programming Languages: Python (PySpark, Pandas, NumPy), SQL (T-SQL, PL/SQL, SparkSQL), Scala
Big Data Technologies: Spark, Airflow, Kafka, HDFS, Hive, Presto, Hadoop
ETL & Visualization: Airflow, Azure Data Factory, Dagster, DBT, Databricks, Fivetran, Power BI, Tableau, Looker
Databases & Storage: Snowflake, PostgreSQL, Redshift, ADLS, Azure Synapse, S3, MySQL, MSSQL, MongoDB (NoSQL)
Other Tools: Git, Agile, JIRA, Docker, VS code, Kubernetes, CI/CD Pipelines, MLflow

WORK EXPERIENCE

Product Data Engineer
META Platforms Inc
Aug 2024 – Present
Menlo Park, CA

- Engineered **Data foundations** by ingesting logging **scrapes** and building **Ads performance** metrics across multiple **0→1 product rollouts** for **Ads delivery system**; accelerated **feature** iterations and unlocking **~$300 M** in incremental revenue.
- Architected a **Product Growth framework** using **Airflow DAGs** to automate end-to-end **ETL** integration, **HIVE** tables and standardized **metrics**; adopted by **9+** product teams, slashing developer effort by **80%** and **infrastructure** overhead by **55%.**
- Designed a granular UTM-parameter matching framework in **Presto** to process **200 B+** records. standardizing **integration** of **Google Analytics** and **Adobe** logs via **API**, delivered a **5% conversion** uplift alongside a **12%** increase in **marketing** spend.
- Processed **2 PB/day Spark** + **Hive** pipeline to generate feature-rich **training datasets**, retrain **ML models** and **A/B experiments**, auto-flag high-CPA **Ad Campaigns** — delivering a **20%** lift in conversions and a **$150M** uptick in advertiser spend.
- **Frontend Products: multi rules, partner integration**

Data Engineer
T-Mobile, USA
Jun 2023 – Sep 2023
Seattle, WA

- Implemented a **Medallion-layer Data Mesh** on **Azure (ADF** + **Databricks** over **ADLS),** orchestrating **batch & streaming** pipelines that feed governed Gold views to **5 + Power BI** dashboards—cut **ETL** latency **20** % and time-to-insight **30 %.**
- Contributed development of comprehensive **Operational** and **Data Quality** monitoring infrastructure for T-Mobile, stabilizing **300+** tasks by creating **50+** data quality rules to ensure 2C data correctness.
- Analyzed T-Mobile device application performance in **Snowflake**. Leveraging **DBT** for efficient transformation, identified vendor charge discrepancies for unsuccessful installations. This led to proposed agreement modifications, cutting cost by **16%.**

Data Engineer
Drexel University
Sep 2022 –Mar 2024
Philadelphia, PA

- Processed **10+** million time series data records, building **batch** data pipelines using **BigQuery**, and **streaming pipelines** using **DataFlow (Apache Beam).** Used **Airflow** for orchestration and **Composer** for deployment.
- Designed and operationalized customized **ETL** pipelines by addressing **ad hoc** requests for different departments using **SSIS** for **MsSql**, Oracle Data Integrator (**ODI**) and **Teradata**, ensuring seamless **data flow** to cater to specific analytical demands.
- Enhanced Database efficiency, Memory utilization, and improved System Functionality by **15%**

Data Engineer/ Analyst
High Radius
Jun 2020 - Aug 2022
Hyderabad, TG

- Fabricated a **Data Warehouse** for an E-fitness client with **100+ tables** and composed **25+ ETL** Pipelines using **Azure Data Factory** from **MongoDB**, **MySQL**, App Store Connect **API** and WooCommerce to **MSSQL** Server.
- Played a role to create scalable, fault-tolerant data pipelines using **Kafka** and Amazon **Kinesis** for **real-time** data ingestion into **Redshift**, achieving a **25%** reduction in latency through parallel processing techniques.
- Revamped existing reporting and data management system utilizing **SQL** & **Tableau** reduced **data redundancy** by **38%**
- Collaborated with cross functions to implement Zoho **CRM** and automate Business process, increased growth visibility by **24%**

EDUCATION

Drexel University, Le Bow College of Business
Philadelphia, PA
Master of Science, Business Analytics
Chandigarh University, Apex Institute of Technology
Chandigarh, IN
Bachelor's in technology, Computer Science Engineering