

Rohit Kumar

Data Engineer

A Tech Enthusiast who loves to learn new things and helps Data reaching to its correct destination without any lag

✉ rohitroy37100@gmail.com

☎ 7488659743

📍 Bangalore

🌐 [linkedin.com/in/rohit-kumar-363820140](https://www.linkedin.com/in/rohit-kumar-363820140)

WORK EXPERIENCE

Data Engineer

Koantek

02/2022 - Present,

Koantek is a startup founded in 2021 and currently provides the services in Data Engineering, Data Science and Dev-ops Technologies. Currently all of our HR task is done through KIPi.Bi

- Migrated EMR spark pipelines to Databricks platform using notebook/Python Wheel methods that helped in increasing the efficiency of jobs and reduction in cost by performing cluster optimization.
- Created custom airflow DAG of Databricks operator using Databricks jobs 2.0 API to trigger and monitor the Databricksjobs/Notebook as an orchestration tool for legacy airflow version.
- Created and tested Virtual Data Connector by different methods for ADLS Gen 2 storage, snowflake and Big Query for Ascend Framework.
- Modified the existing Ascend framework to be compatible for GCP and AWS platform.
- Migrated data from Snowflake to Delta Lake tables in Databricks.
- Migrated Redshift views to Databricks SQL Platform.
- Created End to End to structured streaming pipeline using autoloader following medallion architecture (Bronze, Silver and Gold) from S3 bucket.
- Partitioned the table to have the data loaded into same gold table simultaneously to avoid concurrent update conflict from multiple sources in pipeline.
- Tuned the performance of existing cluster to increase the efficiency of data processing from S3 bucket to Delta Lake
- Exported Databricks logs to logz.io for easy visualization of logs and connected Databricks Delta Tables to Power BI for easy visualization of Data.
- Created a Python scripts to create and manage secret scope in Databricks using Secrets API available in Databricks.
- Connected the Delta Table data from Databricks to Power BI Desktop for easy visualization.
- Created handover documents with the concerned information and data.

Senior Project Engineer

Wipro

11/2021 – 02/2022

- Created Databricks notebook to load data in current pipeline.
- Responsible for Creating Timesheet web app which is used internally to estimate Demand VS Supply for Wipro Employee in RSA Project.

Project Engineer

Wipro

09/2019 – 11/2021

- Responsible for generating weekly, Monthly, Quarterly and Yearly timesheet report using Python which employee used to fill their timesheet based on the work request id which is unique for each development work.
- Created Python Automation script for existing batches to get progress update so that it can reduce manual efforts.
- Modified existing ETL jobs written in IBM DataStage tools to reflect new columns added in Data according to client requirement.
- Created High Level Design document and walkthrough it to the client for the development sprint.
- Worked on ServiceNow ticket to resolve the issue within SLA.
- Created ServiceNow Change Request to Push the code changes in production from Dev.
- Co-ordinated with other team during Deployment and successfully delivered the sprint from Discovery phase to Post Deployment Phase.

SKILLS

Pyspark

Spark SQL

Databricks

Airflow

EMR

Python

Spark

Selenium

DBMS

GitLab

Django

Pandas

ETL

AWS

Azure

GCP

EDUCATION

- **MTech (Data Science and Engineering)**
BITS, Pilani.
09/2019 - 10/2021.
- **BTech (E.C.E)**
Heritage Institute of Technology.
08/2015 - 06/2019

CERTIFICATIONS

- Solutions Architects Essentials - Databricks
- Databricks Developers Foundations-Databricks
- Databricks Developers Essentials-Databricks
- Using Databases with Python – Coursera.
- Using Python to Access Web Data – Coursera.
- Python Data Structures - Coursera.
- Python Data Structures –Coursera.