Rahul Mishra

Data Engineer

✉ rahul.mishra@gmail.com   📞 9876543210   in LinkedIn   Github   📍 Bengalore Karnataka

📇 PROFILE

4+ years of experience in various projects related to Data Modeling, Data Analysis, Design, and Development for Data warehousing environments. Practical understanding of Data modeling concepts like Start-Schema Modeling. Snowflake Schema Modeling, Fact and Dimension tables. Comprehensive knowledge and experience in process improvement, normalization/de-normalization, data extraction, data cleansing, and data manipulation. Exceptional troubleshooting skills with ETL technologies.

🧠 SKILLS

Programing Language:
Python, Java Scala, C++, and Linux Scripting

Databases
MySQL, Elastic Search, Cassandra, HBase, Hive, SQL Server

Dashboarding & Monitoring Tools
Grafana, Kibana, Power BI

AWS Services
S3 bucket, EMR, Lambda, Cloudwatch, DynamoDB, Redshift, MWAA, Glue, SNS, SQS

Code Management and Documentation
Git, BitBucket, Azure Develops, Circle CI, Travis CI

Web Development Framework
Django, Flask

Distributed Framework
Hadoop, Spark, Kafka, BigQuery

Workflow Management
Airflow, IBM Tivoli

Build Tools & Containers
Docker, Kubernetes

💼 PROFESSIONAL EXPERIENCE

Data Engineer, *Wipro*   May 2019 – present
                                                                                             Bangalore, India

Salesforce to Redshift Ingestion - Migration from Informatica to Native AWS

Tech Stack – Salesforce, Informatica, S3, Lambda, Glue, AppFlow, Redshift, SNS

- Crafted generic scalable Native AWS solution for Salesforce to Redshift ingestion
- It helped to move ingestion pipelines from third-party tool Informatica and saved cost for the heavy license fee
- This generic framework helped other business units for smooth ingestion of newly onboarded Salesforce objects into a Redshift Data lake

Dehleez - Report Scheduling Tool

Tech Stack – Python, JavaScript, Django, Azkaban, Docker, Hive, Ajax, Bootstrap, REST API, DataDog

- Enhanced Paytm's proprietary report scheduling tool which is used by business users working on data analysis where they can schedule their reports by writing HIVE/MySQL/Cassandra queries and report output in various formats
- Diff Checker - Admins can check the difference between queries before approving reports
- Time slot picker to schedule a report - The user can see scheduled reports for the next 4 hours from the intended schedule time and can pick the slot accordingly
- Dump report output into the S3 bucket - The user can take a dump of report output into the AWS S3 bucket
- Cassandra Connector - User can schedule reports having Cassandra query panels in addition to HIVE/MySQL

Data Engineer, *Capgemini*

Jan 2018 – May 2019
Mumbai, India

One checkout Platform – Real-time streaming application for Financial data

Tech Stack – Flink, Python, Kafka, Oracle, Kafka Connect, Docker, AWS, Airflow
- Building one checkout platform for different business domains of Expedia i.e; Vrbo, Hotels, Car Rentals, Lodging
- The real-time streaming platform captures & processes financial data for Accounting and helps suppliers to track the total amount to be paid via Auto Pay or Request Pay model
- Crafted a scalable streaming solution using Flink StateFun to handle ~800k booking streams each day

🎓 EDUCATION

Master of Computer Applications, *National Institute of Technology Allahabad (MNNIT), U.P*

Jun 2014 – Jun 2017
Prayagraj, India

Bachelor in Computer Science, *Lucknow University*

Jun 2011 – Jun 2014
Lucknow, India

📑 CERTIFICATES

- AWS certified Database specialty
- AWS certified Practitioner

- AWS certified Solution Architect
- Azure Fundamental