Koushik Dutta

Senior Data Engineer

📞 +91-9874658132   @ kdkoushikdutta282@gmail.com   🔗 https://github.com/koustreak   📍 Kolkata , INDIA

SUMMARY

Senior Data Engineer with over **8** years of experience developing data-intensive applications, building data pipelines, solving challenging architecture and scaling problems. Experience working for clients in the **healthcare, energy and banking** sectors. Currently looking for a Lead Data Engineer role in the B2B or B2C domain.

EXPERIENCE

Senior Data Engineer

Accenture   📅 2020 - Ongoing   📍 Kolkata , INDIA

- **Leading** a team of **10 data engineers**.
- **Migrated** legacy data warehouse to **AWS Redshift**, resulting in annual cost savings of **$475,000** and **20%** performance improvement.
- Designed **data quality** framework to remove impurities from datasets, resulting in a **35%** increase in data pipeline performance.
- Worked with SMEs to develop data marts and data pipelines for maritime business and contributed to **business continuity** that prevented a **$90 million** loss.
- **Reduced** data engineering workload by creating a custom **Spark** template, resulting in the delivery of over **50** data-pipeline on time.
- Ensured **99% data pipeline up-time** when ingesting streaming and batch data from 7 different data sources using Spark, Redshift, S3, Python.
- Developed tools that provide near real-time (15 minute latency) data on **international currency exchange rates**.
- Participated in the development of a solution for the calculation and visualisation (**Grafana** Dashboard) of the OEE (**Overall Equipment Effectiveness**) in near real time for the equipment in **2 power plants**, ensuring a plant **uptime** of over **90%**.

Data Engineer

Tata Consultancy Services   📅 2015 - 2020   📍 Kolkata , INDIA

- Worked with data modellers to create a data for a **sales analysis report** for a **healthcare** client and contributed to the development of the **price waterfall**.
- Automated manual data reconciliation between the different sources, parsing and reformatting data, and **resolving local language references** in data records increases data accuracy by **25%**.
- Created **preemption** logic to prioritise the ingestion of critical data and avoid errors caused by "**early arriving facts**".
- Produced 15-page **best practise** documentation with approaches to support continuous process automation for data ingestion and data pipeline workflow.
- Maintained **DDL Flyway** to control the change of metadata in Redshift and Glue Data Catalogue.
- **Led** a team of **5** in developing a **descriptive data lineage** to identify data sources from six different demographic regions and three different data warehouses.
- **Mentored** 9 junior data engineers in all aspects of data engineering.
- Contributed to and maintained the data architecture of event-based data ingestion and masking of **PII** (**personal Identifiable Information**) **data**.

PERSONAL PROJECTS

DDL Comparator

(open source contribution)

- Built using **Python 3.8**
- Compare two DDL statements, and highlight differences.
- Capable of handling 5 different SQL dialects.
- Provide optimization hint for DDL statement.
- Link_to_project

EDUCATION

B.Tech in Computer Science

Kalyani Govt. Engineering Collage   📅 2011 - 2015   📍 West Bengal, INDIA

SKILLS

Programming

`Python` `SQL` `Scala` `PL/SQL`

Cloud

`AWS` `Databricks` `Snowflake` `Confluent`

Data Processing and Orchestration

`Kafka` `Spark` `Hive` `Pig` `Airflow` `AWS Step Function`

SQL and NoSQL Databases

`Oracle` `MongoDB` `PostgreSQL` `AWS DynamoDB`

Additional Skills

`Git` `Docker` `Data Modeling` `Grafana`

FIND ME ONLINE

in LinkedIn
https://www.linkedin.com/in/koushik-dutta-9797a8209/

M Medium
https://medium.com/@koushik-dutta

CERTIFICATION

AWS Certified Developer Associate - 2021
Certification_Link

AWS Certified Data Analytics Specialty - 2022
Certification_Link

Databricks Lakehouse Fundamentals - 2023
Certification_Link