SOURABH JOSHI

Data Engineer

PERSONAL PROFILE

Passionate Data Engineer with over 8 years of experience diving into Data. Constantly working to improve my skills in any area of Data Engineering and Analytics. My expertise is dealing with automation of processes, creating data pipelines and data products which enable the teams I work with to take the right decisions as well as creating new methods of collecting, storing and analyzing data.

ACCOMPLISHMENTS

- Developed a sample data generator tool for ETL testing team, reducing the Testing time by 30%
- Facilitated the establishment and execution of the roadmap for Information delivery. Including Enterprise Data Warehouse, Data tools and Data Management
- Developed a Fuzzy match application improving efficiently by 60% to identify customer accounts
- Built an in-house collaborative platform for Project postings/Issue Hub for Tesco
- Active participant in the Open Source data tools Community
- Blog on Open Source Data Tools

SKILLS SUMMARY

- **Big Data** – Spark, Kafka, Hive
- **Cloud** – AWS (Lambda, Glue, Cloud Formation, Athena, S3)
- **Programming/Scripting** – Python, Shell, Perl, Base SAS 9.4, VBA, Go Lang (Newbie)
- **NoSQL** – Cassandra
- **Graph DB** - Neo4j
- **R Databases** – Oracle
- **Open Source Data Tools** – Apache Airflow, Great Expectations, Deequ, Soda SQL, Pandas Profiling, Apache Superset, Airbyte, dbt, Delta lake, Spline

GET IN TOUCH!

📞 +91 9886274050
✉️ soujoshi89@yahoo.com
in https://www.linkedin.com/in/dataengineersourabh
🖥️ https://dataexpertz-blr.github.io/

PROFESSIONAL CAREER

Senior Data Engineer

Tesco Bank | 2016 - present

- Architecture, design and continuous expansion of Spark Clusters within the Data team with focus on security, data ingestion and other data tools
- Currently leading the project (Insurance Product) to replicate ETL jobs from SAS to Spark platform.
- Leading an In house "Data Quality" product development program

Data Analyst

ANZ Operations and Technology | 2015 - 2016

- Involved in gathering the customer requirements, and preparing Design documents to be used by developers
- Worked in maintenance releases of the project after production implementation and enhancement of project as per requirements
- Led a team of four in a migration project from Datastage v8.1 to v11.5

ETL Developer

Accenture Services | 2013 - 2015

- Extensive hands on experience in ETL development on Datastage
- Involved in Unit Testing and Integration testing
- Involved in writing UNIX shell scripts for various purposes in the project to be included as part of the data Stage Jobs
- Involved in Design and Analysis of SQL queries and in performance tuning

Intern

Datagenic India | 2012 - 2013

- Created Python scripts that would extract data from the vendor specific site and load to staging area
- Built a python-based job monitoring tool
- SQL queries to generate vendor specific reports

SIDE PROJECTS

- Microscope Automation (IoT)
- NLP Resume Parser
- AWS IAM Graph

PASSION

- Photography
- Music
- Open Source
- IOT

EDUCATION

RNS Institute of Technology
B. Tech (Computer Science)