Shubham Mishra *Data Engineer*

✉ shubham.mishra@gmail.com ☎ 9876543210 **in** LinkedIn ⊙ Github ⦿ Bengalore Karnataka

PROFILE

Responsible Data Engineer, passionate about delivering outstanding quality and service. Offering 7 years of experience in the industry with a history of recognition for performance. Analytical thinker leverages problem-solving skills to drive continuous improvement. Works cross-functionally to adapt systems to changing business needs and formulate winning solutions. Able to work with clients to manipulate data to gain value. Strong foundation in technology and business applications. Well-versed in teaching, developing, and managing customers and team members to achieve self-sustainability. Deep knowledge of visual analysis design principles. Significant experience in large-scale and enterprise-level deployments.

PROFESSIONAL EXPERIENCE

Senior Data Engineer, *Mu Sigma*   10/2020 – present │ Bengalore, India

AVIVA Life Insurance

Migration, implementation, and scheduling of ETL job in Azure for customer and insurance-related data

Tech Stack: Azure Data Factory, Python, Azure Data Lake, Azure Steam Analytics, PySpark

- Investigated requests for maintenance and improved existing solutions
- Wrote maintenance procedures for software to help users and administrators get the most from designs.
- Developed standard procedures for testing and validating software.
- Collaborated across technical and design teams to produce innovative software applications.

Britannia Industries Ltd

Data Ingestion & Sync Process

Tech Stack: – Python, Hive, ElasticSearch, Scala Play Framework, SBT, EMR, Lambda, DynamoDB, Azkaban, Jenkins

Crafted data-sync logic by prioritizing datasets (High/Medium/Low tag) based upon criticality to meet SLO

- Built preemption logic to prioritize highly critical datasets when multiple low-priority sync processes are running
- Designed Rest API in data ingestion for retention of GA data in order to optimize cluster space
- Added exception handling scenarios in data sync logic to fix multiple bugs
- Fix for missing PG data from Kafka for UMP panel - Created a new pipeline to ingest missing data from HDFS to ElasticSearch in case of cluster failure

Data Engineer, *Amazon*   04/2017 – 09/2020 │ Hyderabad, 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

Near Real-Time Data Pipeline

Tech Stack – Java, Spark, Kafka, Datastax Cassandra, Datastax studio, Zookeeper, Maven

- Crafted a Cassandra based real time ingestion pipeline for marketplace data in order to help DWH team to reduce request.
load from production MySQL. The Objective was to shift business users from production, to overcome data leaks & security issues.
- Interacted with different business users to know about their use cases, ingestion tables, PII data and built data models accordingly for faster insertion/updation of data.
- Setup web interface Datastax Studio for users to query real-time data from Cassandra using LDAP authentication.

Expedia, *Data Engineer*   08/2015 – 03/2017
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
- 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

TECHNICAL SKILLS

Programing Language: *(Python, Java Scala, C++, and Linux Scripting)*,
Databases *(MySQL, Elastic Search, Cassandra, HBase, Hive, SQL Server)*,
Distributed Framework *(Hadoop, Spark, Kafka, BigQuery)*,
Dashboarding & Monitoring Tools *(Grafana, Kibana, Power BI)*, **Workflow Management** *(Airflow, IBM Tivoli)*,
AWS Services *(S3 bucket, EMR, Lambda, Cloudwatch, DynamoDB, Redshift, MWAA, Glue, SNS, SQS, Code Pipeline, ECS, ECR)*,
Build Tools & Containers *(Docker, Kubernetes)*,
Code Management and Documentation *(Git, BitBucket, Azure Develops, Circle CI, Travis CI)*,
Azure Services *(Azure Data Factory, Azure Analytics Tools, Data Lake Analytics, Azure Synapse Analytics, Azure Stream Analytics)*

EDUCATION

HSC, *Bhavan's College of Arts, Science and Commerce*   06/2009 – 06/2011 │ Mumbai, India

Bachelor of Engineering in Computer Science,   06/2011 – 06/2015 │ Prayagraj, India
New Horizon College Of Engineering

CERTIFICATES

- AWS certified Solution Architect
- Azure Fundamental
- Azure Data Scientist