




SHASHANK MISHRA

DATA ENGINEER

 noobdataengineer@gmail.com  +91 98XXXXXXX  Lucknow, Uttar Pradesh

EDUCATION

NIT ALLAHABAD

JUL 2014 - MAY 2017
MASTER OF COMPUTER APPLICATIONS
CGPA 8.5/10 (NIMCET Rank 43 of 20,000)

LUCKNOW UNIVERSITY

JUN 2011 - JUN 2014
B.SC IN COMPUTER SCIENCE

TECHNICAL SKILLS

PROGRAMMING LANGUAGES

Python, Java, Scala, C++, C,
Shell Scripting

BACKEND FRAMEWORKS

Django, Scala Play Framework,
Android Studio

WEB DEVELOPMENT

JavaScript, AJAX, JQuery, HTML,
CSS, BootStrap, NGINX

DATABASES

MySQL, Vektorwise, ElasticSearch,
Cassandra, HBase

BIG DATA FRAMEWORKS

Hadoop, Spark, Hive, Kafka,
BigQuery, DataRush

DASHBOARDING TOOL

Grafana, Kibana, DataStudio,
DataDog

SCHEDULERS & AWS

Azkaban, Airflow, S3, EMR, Glue,
Lambda, Cloudwatch DynamoDB

CONTROL SYSTEMS & DOCUMENTATIONS

Git, BitBucket, SVN, Jira, Confluence

WORK EXPERIENCE

PAYTM - DATA ENGINEER

JAN 2019 - PRESENT

Data Ingestion & Sync Process

- 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

GG VMN migration

- Migrated all **Facts/Olaps** written in Hive into **PySpark**

Near Real Time Data Pipeline – POC

- 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**
- Setup web interface **Datastax Studio** for users to query real time data from Cassandra using LDAP authentication

Hive Query Parser

- Query Validator and **Optimization** Engine - Created a Django web application to **parse** and **validate** user's hive queries..
- In case of a bad query (missing **partition** columns/unbalanced **joins**), it also provides **suggestions** to improve the query

OPERA SOLUTIONS - SOFTWARE ENGINEER

JAN 2017 - JAN 2019

Procurement Spend Optimization (Pharmaceutical)

- Developed CXO-level insights engine to manage **USD 60Bn**; engine enabled cost optimization using smart categorisation, benchmarking and **anomaly detection**
- Crafted a Big Data based solution using Hadoop Ecosystem (HDFS, YARN), Spark and Python for organised structured & unstructured data

Trip Narrative Platform (Aviation)

- Deployed an end to end solution for a **leading US airlines**; Aggregated a 360 view of customer's engagement throughout the life-cycle of the trip
- Developed **data pipelines** from scratch; optimised data aggregation from 10+ independent sources and automated the ETL process to roll out the solution

GEEKSFORGEES - TECHNICAL CONTENT WRITER

AUG 2016 - JAN 2017

- Individually contributed **150+** articles on Data Structures (Arrays, Stack, Queue, Linked list, Tree) and Algorithms (Dynamic Programming, Optimised Searching)
- Articles live on their website; received **30,000+** views and **2000+** upvotes
- Published 3 **YouTube** videos on Data structures and Python; received 4000+ views

ACHIEVEMENTS

- **Opera Ovation Award** : For exceptional work on a Trip Narrative Project, 2017
- **Runners Up** : Opera Python Hackathon; internal python coding competition, 2017
- Conducted a firm wide **Global Level session** for 200+ employees on SHM encompassing designing, executing and best practices on **workflows**
- Organised BigData session for India's Analytics team of 50+; demonstrated usage of **HIVE** platform; personally conducted 10+ workshops
- **Geek Of The Month** : GeeksForGeeks; for contribution in technical content, 2016
- **2 SILVER & 3 BRONZE MEDALS** : HourRank 18 and Week of Code 30; Hackerrank 2017, Week of Code 29, HourRank 17 and World CodeSprint 9; Hackerank 2017