

SHUBHAM JOSHI

DATA ENGINEER 📍 BENGALURU, INDIA ☎ 8447082238

◦ DETAILS ◦

Bengaluru

India

8447082238

joshi.shubham82@yahoo.com

◦ SKILLS ◦

Python

Programming

Software Logic

Software Troubleshooting

Java Spring Boot

Pyspark

◦ LANGUAGES ◦

English

👤 PROFILE

Experienced Data Engineer with expertise in design, installation, testing and maintenance of software systems. Able to effectively self-manage during independent projects, as well as collaborate as part of a productive team.

📁 EMPLOYMENT HISTORY

Data Engineer at ANZ, Bengaluru

June 2021 — Present

- Collaborated with the product and business team to understand requirements and business specifications about application, analytics and risk.
- Built Backend/Hadoop component for internal payment application.
- CI/CD pipelines using Bamboo.
- Used **Kubernetes** for deploying frontend/backend to openshift and **helm** charts for packaging.
- Used **pyspark** for writing complex batch processing jobs and **oozie** coordinators for orchestration.
- SAML and ADFS integration for SSO and authentication.

Data Engineer at Capgemini Invent, Gurgaon

April 2019 — June 2021

Client: SessionM - a customer engagement and loyalty platform

- Developed ETL pipelines for massive data transformation using **Apache NIFI**
- Developed highly scalable API on AWS lambda leveraging **Dynamo DB, Athena**
- Built schedulers to pull incremental data from various clients into our platform
- Python script to generate reports, analysis

Data/Machine learning Engineer at Hudson Data, Gurgaon

September 2017 — April 2019

Data Engineering

- Developed ETL pipeline for both streaming and batch processing using **Apache BEAM, Airflow, Spark-Streaming, PySpark**
- Developed and hosted APIs for credit report parsing on **Google's App Engine**.
- Created, scheduled, monitored data migration workflows using **Apache Airflow**

Machine Learning

- Feature Selection -> Used a variety of techniques such as forward selection, Genetic Algorithm, variable importance from a model on **H2O AI**
- Building Models -> Build Decision trees, GBM, GLM, Random Forest, XG Boost, using python API for H2O and H2O Flow
- Model Tuning -> Using Grid Search in GBM to tune
- Model Monitoring -> Various techniques like PSI, KS, IV, AUC to check model & population stability
- Visualization -> Used **DOMO** for creating various reports and model monitoring

DevOps Engineer at Overcart, Gurgaon

July 2016 — September 2017

- System Administration (AWS Cloud)
- Develop scripts to automate deployment, administration, and monitoring of this large scale Linux environment

- Ansible playbook for code deployment and Tower for managing inventory/resources
- Setup company-wide used reporting module on Metabase and Nginx/Application log visualization on Kibana
- Setting up auto scaling groups , alarms and alerts, AWS Lambda and APIGateway for micro services
- Significant Cost and Resource Optimization
- Datapipelines for migrating data from various sources to datawarehouse for reporting
- Developed various Reports/Dashboards on top it for different departments



EDUCATION



Bachelors in technology, Computer Science, IIIT JABALPUR, Jabalpur

August 2011 — August 2015



COURSES



GCP - Professional Data Engineer, Google

December 2018 — December 2020



INTERNSHIPS



Database Engineer at S&P Global Market Intelligence , Gurgaon

August 2015 — August 2016