

## Summary

- I bring total 3.3 years of experience working in one of the best IT Services provider Infosys Pvt. Ltd. with good knowledge in multiple roles – development and maintenance. My basic areas of expertise – Development.

## Key Domain and Technical Knowledge

- Google Cloud Platform
- Big Data (Hadoop) – Hive, HDFS, Pig, Spark
- Python, SQL, Shell

## Certifications

- Undergone internal training for GCP, Python, SQL, Business Intelligence Tools and Hadoop.
- Have done internal certification for Agile.

## Academic Qualification

- Bachelor of Engineering (B.E.) with specialization in Electronics & Communication Engineering, Vasavi College of Engineering, Hyderabad, India.

## Professional Experience

Have been trained in Infosys for a period of 5 months which was divided into 3 modules where the first module was on Python and SQL and the next was on Business Intelligence tool such as SSIS, SSMS, Cognos, PowerBI and then the rest was on Hadoop. There were assessments and project in each module.

### **Project 1: Financial Crime Risk - Data Reconciliation**

Company	: Infosys
Client	: HSBC
Project	: Development
Modules	: HQLs, Shell Scripts, Pyspark, Deployment
Duration	: July 2018 – Dec 2019

## Responsibilities:

- ✓ Have done pattern analysis of the data extraction from source.
- ✓ Set up the firefighter accounts to login to the production.
- ✓ Developed HQLs (Hive query language) for each country to match the transactional data from source and the target (monitoring system) with applying given set of filters.
- ✓ Written shell scripts for the automation of the HQLs written for each of the country and shell scripts for purging the old data which is not required for further use which can lead to storage issues.
- ✓ Generated reports for each reconciliation pair for all countries.
- ✓ Deployed the code developed to production server through GitHub and Jenkins.
- Technology: HIVE, HDFS, Shell, GIT, Jenkins

## ❑ Project 2: CMB-Marketing UDM

Company : Infosys  
Client : HSBC  
Project : Development  
Modules : Pyspark, GCP(Big query, Apache beam, GCS, Cloud SQL), Python, Unix  
Duration : Feb 2020 – Present

### **Responsibilities:**

- ✓ Developed Pyspark scripts to extract the data from CSVs and store in HDFS and Hive tables and the data stored is cleansed with help of various filters and converted to 1-normalised form and populated to Unified Data Model which is used to generate the reports on Qlick.
  - ✓ Written shell scripts to automate the above process.
  - ✓ Worked on GIT hub, ansible and Jenkins for deployment.
  - ✓ Workings on GCP as the data was later moved to GCS buckets.
  - ✓ As in GCP, have written python scripts in apache beam where I had to set up pipe lines to extract data from GCS buckets and cleanse the data and store it in Big query and then Normalize to store it into Unified data model(Cloud SQL).
  - ✓ Had setup Airflow DAGs for the Shell scripts to automate the above process.
- Technologies: Pyspark, GCP, Shell, Airflow, Ansible, GIT, Jenkins.

### **Personal Information**

- Candidate Name - Somani Ankith
- Job Role - Senior Systems Engineer
- Infosys date of joining - 01/01/2018
- Location of candidate - Hyderabad
- Mobile - 8125949569
- Email - somani.ankith@gmail.com

### **Declaration:**

I hereby declare that all the details mentioned above are in accordance with the truth and fact as per my knowledge and I hold the responsibility for the correctness of the above-mentioned particulars.

**Place:** Hyderabad

Somani Ankith