

Professional Synopsis

Data Engineer, Wipro Ltd.

2019 – Present

- Developed batch data pipelines for one of the leading automobiles industry using **Azure data Factory**, Azure Blob Storage, Snowflake, **Pyspark**, SQL
- Migrated 200+ tables from on-prem DB to Snowflake via AWS Services(EMR, EC2, S3), Pyspark, Sqoop, SQL
- Developed dynamic data pipelines using python dags in airflow
- Led, Designed, and implemented incremental data pipelines using azure data factory, azure blob storage, azure sql database for one of the critical project of Europe client
- Experienced with an Agile software development
- Experience in extracting data from multiple data sources and landing it to Snowflake.
- Won Excellence Award for outstanding performance in datalake migration project.
- Transformed terabytes of data using Pyspark and loaded into Snowflake via AWS S3 and Azure Blob Storage as stages.

Projects

1. World's Largest Automobile(North America), Data Engineer

2021 – Present

- Developed and deployed full load and **incremental** data pipelines using big data tech stack such as sqoop, **pyspark**, AWS **EMR**, **Airflow**, EC2, shell scripting and **Snowflake**.
- Expertise in working on Hortonworks Data Platform, Greenplum DB
- Developed **dynamic PySpark** scripts such as reading data from jdbc, AWS S3, HDFS, **Azure Blob Storage** and performing data transformations to enrich terabytes of data.
- Led, Designed and implemented azure data pipelines by integrating to Snowflake using storage integration, azure blob storage, azure data factory and airflow.
- **Migrated** around 200+ tables from on-prem Greenplum database to Snowflake using big data tech stack
- Led team of Manufacturing data engineers to develop end to end metadata driven incremental pipeline using ADF activities and respective azure services
- Led team of 5 GETs(Graduate Engineer Trainees) for Data Engineer role in a planned manner with individual task tracker and doubt solving sessions
- **Led Client calls** to understand business requirement and delivered the tasks in given time frame
- Developed dimensional data model by constructing fact and dimensional using snowflake sql scripts.
- Developed airflow dags by R & D for executing custom shell scripts and custom sql scripts in EMR
- **Reduced cost** by **optimizing** data pipelines such as conversion of spark action to snowflake sql to save cluster run time
- Performed PoC on **Snowpark** and received appreciation for the same

2. Multinational Energy Corporation(U.S.A), Data Engineer

September 2020 – Dec 2020

- Involved in understanding functional requirements with business and other team members.
- Completed Azure Fundamentals under given period of time that is 12 days as per client requirements.
- Completed DP-200 and DP-201 training and certifications under given period of time that is 1-2 months.
- Delivered Knowledge transfer sessions to new team members

3. Banking Subsidiary(United States), Data Engineer

May 2020 – August 2020

- To analyze the input csv data and map the schema using different functions and logics
- To deliver the resultant mapped data to apian developers

- Delivering knowledge transition to the newly joined team members.

4. Anaplan

[Dec 2019 – April 2020]

- Completed Anaplan L1, L2 and L3 certifications within given span of time
- Build Anaplan models on financial analysis and forecast using Anaplan modules and functions.
- Build a Data Hub model to create the central repository for data that can be more easily updated and used as a source for data in other models, such as the Supply Chain or FP&A models.
- Build real time dashboards according to the provided use cases

5. Telecommunication Client(Africa), Data Engineer

[August 2019 – Dec 2019]

- Developed sql scripts on various use cases for Airtel Africa Kenya Ltd.
- Prepared region wise real-time dynamic report showing no of inactive customers using graphana.
- Supported team members in installation and setting up the project environment.
- Performed Extract and Load using putty.
- Delivering KT sessions to the new joiners

Skills

- | | |
|---|-------------------|
| • Python | • SQL |
| • Hadoop(Sqoop, Oozie, Ambari) | • PySpark |
| • AWS(EC2, EMR, S3, Airflow) | • Snowflake |
| • Azure(Azure Data Factory, Azure Blob Storage, Azure DataBricks) | • Shell Scripting |

Education

Bachelor of Science , Maharshi Dayanand University	04/2016 – 05/2019
Awarded 1st Division with 70.5%	

Higher Secondary School , Govt. Sarvodaya Kanya Vidyalaya, New Delhi	04/2014 – 05/2015
Percentage : 82.6	

Secondary School Certificate , Sainik Public School, Haryana , Grade : 9.6 C.G.P.A	04/2012 – 05/2013
---	-------------------

ACHIEVEMENTS & CERTIFICATIONS

- Microsoft Certified AZ-900: Azure Fundamentals.
- Microsoft Data Ninja Series Participant : DP-203 Series
- Microsoft Certified Trainer(MCT)
- Microsoft Certified DP-200: Implementing an Azure Data Solution.
- Microsoft Certified DP-201: Designing an Azure Data Solution.
- Airflow Certification from Udemy
- Anaplan L1, L2, and L3 Certified Model Builder.
- Snowflake Fundamentals Certification from Udemy
- Certified in AI, ML, Data Science and NLP – Level 1.
- Participated and won **Bronze medal** in Spirit of Wipro(SOW) Run in 2019
- Udacity AWS Machine Learning Scholar
- **Delivered Talk in Google developer Groups Delhi Women's Day Celebration on Data Engineering**