

HARSH DUTT

Falls Church, Virginia

☎ [+1-5716099371](tel:+15716099371)

✉ harsh234102@gmail.com

🌐 [Linkedin](#)

🐙 [Github](#)

🔗 [LeetCode](#)

EDUCATION

Bachelors in Technology, DTU, DELHI COLLEGE OF ENGINEERING

2013 – 2017

B.Tech - Electronics/Computer Engineering - CGPA - 7.94

New Delhi

EXPERIENCE

Deutsche Telekom Digital Labs | Gurugram

December 2022 - August 2023

- Accomplished Data Modelling and Development of RDK(500K events per day) and HGW(40K events per day) Datawarehouse extracting telemetry data from S3 and building fact and dimension tables using Pyspark jobs and scheduling it using Airflow which impacted millions of users across Europe. Also, enhancing security features by enabling IAM authentication while reading/writing data to Redshift cluster using Pyspark scripts run on EMR.
- Developing Nifi pipelines to consume Kafka topic data, transforming it and, persisting in S3 data lake for further analysis and data warehousing purposes.
- Development of NiFi Custom processor in Java(Maven build) to automate conversion of Avro file to Json file/Parquet file and deploying it on Nifi cluster

Gspann Technologies | Gurugram

April 2022 - December 2022

- Built Python code to extract Tableau view data and pushed the processed data frame to MySQL in AWS.
- Creating pipelines involving Spark for transformations, Airflow to schedule the Spark code and using S3, Oracle DB, and Hive(HDFS) to store the data in the form of tables and flat files.
- Airflow DAGs to run spark jobs. HIVE vs Snowflake query validations and optimization across data layers for client i.e Nike

Bank of America | Gurugram

June 2017 - April 2022

- Enhanced Application (6 Deposit Apps) security by upgrading Teradata authentication using Kerberos Authentication. Recognized with Silver Award by LOB
- Improved 2 application's security features by enabling Kerberos authentication for Service IDs. At every login, a temporary password is generated using a Kerberos ticket, enhancing threat risk rating by 16 percent.
- Developing and tuning ETL pipelines using PySpark(Import) code from Teradata to HDFS and end-to-end testing for 6 applications in scope, importing data close to 56 million records impacting thousands of customers.

TECHNICAL SKILLS

Languages: Python, SQL, Java, Pyspark, Scala, JavaScript

Technologies/Frameworks: Teradata, Hive, Snowflake, Redshift, Spark(Python), Pandas, Numpy, SQLAlchemy, Autosys, DMX, Airflow, Web Scraping, Git, Hadoop, Sqoop, Ansible(Basic), AWS, S3, EMR, Athena, NiFi, Kafka, AWS CLI, IAM, Glue, Lambda, HTML, CSS, Shell, Unix, Agile, Grafana, ElasticSearch, Docker, Django.

Developer Tools: VS Code, PyCharm, IntelliJ

CERTIFICATIONS

- Oracle Database 12c: SQL Fundamentals, Introduction to Apache Hadoop, Introduction to PySpark Fundamentals, Machine Learning, Data Science and Deep Learning with Python, ISO 27001- Information Security Management Systems, Building RealTime REST APIs with SpringBoot, Apache NiFi Complete Master Course
 - **Leetcode:** Solved **240+** Problems [🔗](#)

AWARDS AND EXTRACURRICULAR

- I like to play Guitar and I've performed at a few events.
- Along with Data Engineering, I have an interest in Backend Development and Machine Learning, so in my spare time I try to learn something from these areas.
- Bronze award for successfully conducting DR exercise for 33 AITs.
- Bronze award to quickly learn and enhance data pipeline in Unix, Hadoop, Python, Spark ecosystem to successfully deploy an application.
- Recognized with Silver Award by LOB to successfully deploy 2 applications with enhancements within 6 months