SIDDHARTH NAHATA

Data Engineer ◇ (+91)7389629262 ◇ Bangalore,India

siddharthjain757@gmail.com ◇ [Linkedin](#) ◇ [GitHub](#)

SUMMARY

Data Engineer/Big Data engineer with around 4 Years of Experience working in Apache Spark,Pyspark,Python,SQL, Hive,Data,Hadoop,Azure Databricks,Data Management,Big Data,Data Engineering,Problem Solving,Data Analytics.

EDUCATION

Bachelor of Engineering(Computer Science),Csvtu University	May 2018

TECHNICAL SKILLS

Languages	Python, SQL, Scala
Big Data Frameworks	Spark, Pyspark,Spark Streaming, Spark Dataframes, Hadoop, Hive,Sqoop,Map Reduce
Databases	MySql, Hbase, MSSQL Server
AWS Services	S3,EMR,Athena,Glue,Redshift
Azure Services	Azure Blob,ADLS Gen2,Azure Data Lake,ADF,Databricks
Version Control	Git, Gitlab, Github
Methodologies Used	Agile,Scrum,CI-CD
Tools Used	Jira, Jenkins, Pycharm, Vscode
Operating System	Windows, Linux
Soft Skills	Communication, Teamwork, Leadership, Analytical Skills.

EXPERIENCE

1.Capgemini India — Oct 2021 - Till Date
(Data Engineer) — *Bangalore,India*

Client: Abu Dhabi National Insurance Company

Project: Abu Dhabi National Insurance CRM Entity Matching Solution:

Description: Entity Management is a tool that is developed for **deduplication** of Customer/policy records. It has a virtue to deduplicate multiple records across LOBs in the verge of bringing Golden Record result.

- Performed structured profiling on multiple data sources using python, Apache Spark,pyspark.
- Merging data from multiple sources and enriching the data using Join Optimization such as SMB join in Spark results in 10% improvement
- Worked on Matching and Merging rules for customers of ADNIC using **Apache Spark,Pyspark**
- Standardizing attributes which are required for project.
- Performed different threshold combinations with various rules to match and to get accurate results on large dataset using Apache spark, Azure databricks

2.Wipro Technologies — Feb 2019 - Sep 2021
(Data Engineer) — *Bangalore,India*

Client: National Australia Bank

- Ingested all tables from Oracle database to HDFS using Sqoop.
- Converted legacy SQL queries to Pyspark dataframes using Spark SQL.
- Later,Created data Pipeline using ADF to transfer our data from local to Azure Blob saving 30% time

AWARDS AND HONORS

- Earned appreciation from ADNIC CIO and Sr.Mgmt. stakeholders both from account side and practice side.
- Won Pragathi Award' from Wipro for developing a POC on Media Performance Characteristics using RTCP data in a Live Voice and Video call.