

NITESH KUMAR

Data Engineer

Mobile: +91 7004566137

Email: niteshkr2063@gmail.com

LinkedIn: <https://www.linkedin.com/in/contact2nitesh>

Professional Summary

- Highly committed, flexible and result oriented professional with total 5.8 years of experience on Hadoop, Hive, Impala, Sqoop, SQL, Python, PySpark, Bigquery, DataProc, Composer, Apache Airflow, Shell Script, GitHub and Jira.
- Experience in designing, developing, and managing data pipelines using Hadoop and GCP data Engineer services.
- Experience to work on large data warehousing across the different database like Bigquery, MySQL, Postgres, SQL Server and Oracle.
- Experience in processing data through Spark SQL, Python and PySpark programs.
- Experience on integrating the source code using Git and GitHub
- Experience in creating Airflow data pipelines to perform ETL and data validation using Python across different servers and databases.

Technical Skills

Programming Language: Python3, Scala, PySpark, SQL

Big Data Technologies: Hadoop, MapReduce, Hive, Sqoop, HBase

Databases: GCP Cloud SQL, GCP Cloud Spanner, Oracle, MySQL, Postgres, SQL Server

Cloud Computing: GCP Cloud Compute Engine, App Engine, Cloud Functions, Etc

Tools: Apache Airflow, Informatica, Control-M, Git, GitHub, Jira, Tableau

Operating System: Linux, Windows

Work Experience

Organisation : TEKsystems Global Services

Designation : Software Engineer

Duration : November 2021 - Current

Organisation : Tata Consultancy Services

Designation : Assistant System Engineer

Duration : January 2020 – November 2021

Organisation : REC Student Portal (UDAAN)

Designation : Data Analyst (Freelancing)

Duration : August 2017 - July 2019

Project Summary

Project Title: BI and Data Engineering

Client: Meta (Facebook)

Role: Data Engineer

Duration: November-2021 to Current

Skills: BigQuery, DataProc, Pub/Sub, Composer, Python, PySpark, Hadoop, Hive, Sqoop, SQL Server, SQL, GitHub, Control-M, DataSwarm(Airflow) and Tableau

Responsibilities:

- Designing and developing the PyScript code to create new tables and load the data from one layer to another layers of Hive (Data Lake, Base, Master, Data Mart, Analytics, Extract).
- Building Reusable Data ingestion and Data transformation frameworks using Python/PySpark.
- Performing testing on production data issue and optimizing the Hive queries using best practices and right parameters.
- Worked on processing of data from SQL Server to Hadoop and Hive using Sqoop framework.
- Worked on reading and writing multiple data file formats like JSON, Parquet, and HDFS using PySpark.
- Collaborating with cross-functional teams to gather requirements and understand data structures.
- Worked with stakeholders to identify data quality issues and implement solutions to address issues.

Project Title: Allianz Life Insurance North America**Client:** Allianz Life Insurance North America**Role:** Associate Data Engineer**Duration:** January 2020 – November 2021**Skills:** Python, PySpark, SQL, Airflow, Scala, GitHub, Shell Script, Bigquery, cloud functions an**Responsibilities:**

- Worked on Analysis, Design, Development of various business applications, different platforms using Informatica, Airflow, Oracle, and Bigquery.
- Worked on processing data through Python, Spark SQL, and PySpark programs.
- Used GCP services such as Cloud Storage, cloud function, and dataProc.
- Developed the Sqoop scripts in order to import or export the interaction between Hive and SQL.
- Monitor the running status of job and pipeline.

Project Title: Data Analytics**Client:** REC Student Portal**Role:** Data Analyst (Freelancing)**Duration:** August-2017 to July-2019**Skills:** Python, NumPy, Pandas, Matplotlib, SQL (MySQL, Postgres), Jupyter Notebook**Responsibilities:**

- Worked on designing data Model of REC Student Portal Web Application and Android App
- Worked on optimizing SQL queries, stored procedures, and views as per application team requirement.
- Worked on data processing, data extraction, data ingestion and data validation.
- Developed the reports using NumPy, Pandas and Matplotlib Python's libraries.

<i>Educational Qualification</i>				
EXAMINATION	INSTITUTE	UNIVERSITY/BOARD	YEAR	MARKS
B.Tech.	Ramgarh Engineering College, Ramgarh	VBU Hazaribag	2019	77.40%
12 th	B.I.S.S.S Secor- 2/C	CBSE	2014	65.16 %
10 th	B.I.V Sector- 12/E	CBSE	2012	77.90 %

Recognition, Certifications and Awards:

- Recognised by TGS for Rising Star Global Award and Spot Award.
- Certification of **GCP Associate Cloud Engineer**

Hobbies and Interests:

- Event Planning & Management

I hereby declare that the details provided hereby are true to the best of my knowledge.

Place: Bangalore

Nitesh Kumar