SANJAY KUMAR

Data Engineer

Result-oriented professional, targeting assignments in **Data Analytics** with a reputable organization preferably in India

+91 8082169024
J&K
sanjaymumbai90@gmail.com
https://www.linkedi
m/in/sanjay-
kumar-335854196/

CORE COMPETENCIES

Data Analytics

Hadoop Development

Data Mining & Insights

Data Visualization

Solution Deployment

Technical Design Documentation (TDD) & Architecture

Cross-functional Engagements

Stakeholder Management

SOFT SKILLS

Change Agent

Collaborator

Communicator

Innovator

TECHNICAL SKILLS

Languages: Python, SQL, Whistle language

Big Data: Hadoop Ecosystem (Sqoop, Hive, Spark, HDFS, HBase

Database: MySQL, SQL Server

Platform: Google Cloud Platform (GCP)

GCP Technologies: GCS, CloudDataFusion, BigQuery, HDE, Dataflow, DataProc
Tools: JIRA,Excel-vlookup,,pivot table,ETL

PROFILE SUMMARY

- Performance-driven professional with **nearly 4 years of experience** in Hadoop Development & Data Analytics, GCP environment includes GCS, Bigquery, Dataproc, Dataflow Healthcare API
- **Data Engineer** with experience in conceptualizing and generating infrastructure that allows big data to be accessed and analyzed; reformulating existing frameworks to optimize their functioning and testing such structures to ensure that they are fit for use
- **Playing a key role in** preparing raw data for manipulation by data scientists, detecting and correcting errors and ensuring that your work remains backed up and readily accessible to relevant coworkers
- **Hands-on experience in Microsoft AZURE and Amazon AWS, Hadoop Framework, and its Ecosystem** (HDFS, Hive, Sqoop, & Spark), and analyzing data through Hive QL programs and HBase tables as per assigned task requirements
- **Skilled in Technical Design Documentation (TDD) & Architecture** and implementation of a batch data pipeline using GCP Cloud Data Fusion (CDF) & Healthcare Data Engine (HDE)
- **Proficient in Software Development Life Cycle (SDLC),** having a thorough understanding of various phases like Requirements Analysis, Design, Development, and Testing
- **Collaborating with Stakeholders** while keeping them informed of progress and issues in order to manage expectations on all requirements & deliverables
- **Efficient organizer, motivator, and team player** with the capability to motivate teams to excel and win

WORK EXPERIENCE

Since Nov'20 with Quantiphi Inc., Mumbai as Data Engineer

Projects:

Adtalem and Walden University

Client: DeVry Education Group Inc.

Description: This project involves migrating processes and data from SQL Server to the GCP component Big Query

Role:

- Worked on tables and views are tested and retested, Similar-sounding tables exist in both SQL Server and BigQuery; data should match in both places; if not, problems should be reported to the table's owner, a developer
- Managed debugging of the particular table or view
- Created CTEs for a particular table and evening created DDLs and testing
- Performed re-testing, If the retest shows no problems, four tests will be included in a document that will be generated and shared with the table developer

Google Benchmarking and Analytics

Client: Google

Description: This project's goal was to use Sqoop, Hadoop, Spark, and other tools to transform various storage loads of data, ranging from a few MBs to TBs, with various CPU and SSD capabilities. Monitored the maximum, minimum, and average values in order to profile data processes in various ways and meet the needs of various internal Google Inc. clients. The project's scope comprises of

- Capability to work in GCP components like Datproc, big query, Monitoring, and debugging, SQL Servers, CDF, Spanner, and so on

EDUCATION

Post Graduate Program in Data Science from IIIT Bangalore with CGPA of 3.5 / 4.0

MCA (Computer Application) from IGNOU, Delhi with First Division

Bachelors in Science from Govt Gandhi Memorial Science College popularly Known as Prince of Wales College, Jammu, J&K affiliated to Jammu University with First Division

HSC from Sri Ranbir Senior Secondary School, Jammu, J & K Board with 72.0%

SSC from Jagriti Niketen Senior Secondary School, Jammu, J & K Board with 70%

- Working knowledge of Hadoop components such as Spark, Sqoop, BigTable, and so on
- The capacity to read data in a variety of forms, including tables, files, JSON, and so on

Role:

- Developed dataProc clusters in GCP through the command line with various configurations
- Designed the pipeline through Cloud data fusion and dataflow with various Hadoop and GCP components
- Monitored & performed debugging for the pipeline & completed the benchmarking process

Healthcare Data Engineering (HDE)

Client: Indiana University of Health (IUH), Google

Description: The goal of this project is to create a framework for converting old HL7 API (Health Level Seven International) formatted data into modern API FHIR (Fast Healthcare Interoperability Resources) prepared data in order to meet various criteria. The project's scope includes

- Managed both streaming and batch data sources
- Migrated files from Microsoft Teams to a properly structured folder in GCS,
- Implemented the Batch Ingestion pipeline to migrate data from GCS to the BigQuery tables
- Created a pipeline for harmonization to transfer data from Big Query to a healthcare API

Role:

- Designed JSON files for all .csv files to provide structure to data in BigQuery
- Managed Whistle language for data transformation in GCP dataflow
- Implemented Batch Pipeline through magic commands to migrate data from GCS to BigQuery and then from BigQuery to Healthcare Data Engine through the Harmonization pipeline

Aug'19-May'20 with Clairvoyant, Pune as Hadoop Developer

Projects:

Generic Data-Quality Framework (DQ)

Client: PayPal

Description: The goal of this project is to provide a general, multi-purpose framework that can be used to profile data in a variety of ways and to meet the needs of diverse PayPal internal clients. The project's range of activities includes:

- Aptitude for working with batch data sources.
- Capability to obtain data from several systems, including hive, oracle, hdfs, and so on
- Capacity to read data in a variety of forms, including tables, files, JSON, and so on

Role:

- Developed the architecture and planned the ways in which the framework can be modularized and collected the client requirements
- Designed different modules as per the requirements gathered

PayPal Polaris (I-Hub)

Role:

- Created Hadoop-Spark-based data pipelines to process, transform and load the data
- Developed HBase and Hive Partitioned tables to maintain historical data
- Performed migration of existing on-premise data pipeline into the GCP
- Managed performance tuning

PERSONAL DETAILS

Sanjay Kumar Sharma

Address: Jammu and Kashmir, Jammu city

Languages Known: Hindi, English, and Punjabi