

Drashti Mehta | +91 9974074251; drashti14395@gmail.com |

Data Engineer | <https://www.linkedin.com/in/drashti-mehta-b255b1104/> |

Summary

- ❑ Passionate data professional with rich working experience of over 4.1 years in building data systems.
- ❑ Knowledge of **Hadoop Core Components (HDFS, MapReduce), Hadoop Ecosystem (Hive, Sqoop, Pig) and Apache Spark.**
- ❑ Experience in ingesting real time/near real time data using **Apache Kafka.**
- ❑ Knowledge using build tools like Maven.
- ❑ Conversant with utilizing GCP cloud platform for creation of data ecosystems including ingestion, storage, transformations and data visualizations
- ❑ Well versed with **Cloud storage, Dataflow with apache beam, Pub-Sub, Datastore, Cloud function, Big Query, Big Table, Cloud SQL, Data Proc.**
- ❑ Experience in working with Python, Data Warehousing and SQL.
- ❑ Experience in Elastic Search and Kibana.
- ❑ Experience in working with source code repositories like **Git** and **Subversion (SVN).**
- ❑ Deep Knowledge of **Agile** methodology.
- ❑ Ability to learn and adapt quickly to the new emerging technologies as per the new requirement of the project.

Technical Skills

Big Data Technologies	: HDFS,MapReduce,Hive,Pig,Spark,PySpark,Spark-SQL,Spark- Streaming,Sqoop,Impala, HBase, Kafka
Cloud Data Engineering	: GCP(Dataflow, Big Query, GCS, gsutils, gcloud)
Programming & Scripting Skills	: Core Java, Python, Scala, Shell Scripting
Data Analysis	: Python, Numpy, Pandas, Matplotlib
Web Application Framework	: Flask
API tool	: Postman
Database	: MySQL
Platform	: Windows, Linux
Version Control	: Git, SVN
Process	: Agile

WORK EXPERIENCE

INFOSTRETCH CORPORATION PVT. LTD.

Data Engineer

July 2019- Present

Working as a data engineer for Infostretch in creating data pipelines ingestion, preprocessing, transformation, storage and optimization along with Vizualization for analytical purpose using Python, GCP, Scala and Spark.

TATA CONSULTANCY SERVICES

Big Data Developer

Dec 2016- July 2019

Worked as a big data developer for TCS to handle large amount of United Healthcare data using big data analytics.

PROJECTS

BioIntellisense

Working for a American pharmaceutical company i.e., Health Monitoring and Clinical Intelligence who develops remote patient monitoring devices, captures data and insights with clinical accuracy. Also, perform advanced analytics and data services for early detection patient's health.

Disease Ontology Datalake – GCP

This project is about creation of a self service analytics platform for client to enable him explore insights out of Disease Ontology data. Its a GCP based ecosystem spanning across ingestion, preprocessing, transformation, storage and optimization along with Vizualization for analytical purpose.

Real Time Data Streaming , Alerting and Anomaly Detection

Create an end-to-end pipeline for real-time data streaming using Kafka, Mysql, Elasticsearch and Kibana.

Twitter Sentiment Analysis

Twitter Sentiment Analysis is the process of determining whether a piece of writing (product/movie review, tweet, etc.) is positive, negative or neutral.

Data Warehouse

Implement a data pipeline to stream data from MySQL database to Matillion in the form of csv using Snowflake.

QUALIFICATIONS

- ☐ B.E , Gujarat Technological University in Year-2016, 7.57 CPI
- ☐ XII , 2012 GSHSEB– 87.43 %
- ☐ X , 2010 GSEB - 83.23 %

ACHIEVEMENTS

- ☐ Star Team Award from Tata Consultancy Services for outstanding performance
- ☐ Best Team Award from Infostretch Corp for performance.