Samana T

Data Engineer

# Contact

**Address**

Bengaluru, India 560085

**Phone**

+91-9066534301

**E-mail**

[samanat020@gmail.com](mailto:samanat020@gmail.com)

**LinkedIn**

* Seeking a responsible role that offers professional growth by being able to fully utilize and expand my knowledge and skills by making significant contributions to the organization
* Working as **Data Engineer** at HTC Global Services with overall experience of 5+ years in the industry and 3+ years of experience in learning and developing applications in big data, python, and Apache spark.
* Good knowledge in Python, Spark, PySpark & big data ecosystems. Passionate in learning and working in new technologies like Spark Streaming and Cloud Computing.

# Work History

[**www.linkedin.com/in/samana-t**](http://www.linkedin.com/in/samana-t)

# Technical Profile

* Python
* MySQL
* Linux
* Docker
* Pyspark
* ETL
* Unix
* AWS
* Azure

# Competencies

TECHNICAL SKILLS

Python, Pyspark

MySQL, Hive, Snowflake, Postgres

Big Data Technologies

Spark, Hadoop, and its Ecosystems, Debezium, kafka

Azure, AWS

**1**. **2022-01 -**

**Current**

**Data Engineer**

#### HTC Global Services, Bengaluru

* ***Project-1: HTC Data Platform*** *– An ETL pipeline product.*
* ***Tenure****: January 2022 – Current.*
* **Roles and Responsibilities**
* Designed and developed end to end pipeline to perform ETL operations on given large dataset.
* Designed the architectural flow of the CI/CD pipeline of the ETL tool which is developed from scratch.
* Collaborated and proactively worked with system architects, design analysts and others to understand business and industry requirements.
* Processed the data from various sources namely: different file formats like (csv, Json, xml, Avro, orc, parquet etc.), hive, different databases like (mongo DB, Cassandra, MySQL, Postgres, Snowflake) and different cloud services like (AWS S3, Azure Blob, Azure cosmos) etc.
* Developed 60+ transformations to handle various operations supported by using pyspark libraries.
* Documented and communicated database schemas using accepted notations.
* Worked on spark optimizations to handle large datasets by using some key features like Partitions, broadcast variables, accumulators, etc.
* Handled processing real time data (Change Data Capture) using debezium open-source connectors and kafka topics to hold the

streaming data and to perform UPSERT commands on various databases on real time basis like (MySQL, Postgres, snowflake) etc.

* + **Technologies & Tools**- Big Data, Datawarehouse, Spark, ETL, SQL, Unix, Cassandra, Kafka, AWS, Azure Blob, kafka, Debezium, Pyspark, Python.
  + **Project 2: Data Validation tool** – A end to end tool used to validate data between different sources.
  + ***Tenure****: June 2022 – Current*

### Roles and Responsibilities

* + Validating the data from various data source and sinks and storing the result.
  + End to end development and maintained the product.
  + **Technologies & Tools**- Big Data, Datawarehouse, Spark, ETL, SQL, Unix, Pyspark, Python
  + **Project – 3: Data Profiling** – a tool designed to get the profiling results and to calculate the richness of the data.
  + ***Tenure****: August 2022 – January 2023*

### Roles and Responsibilities

* + Profiling the data source and generating scorecards and visualizing the data using various libraries like matplotlib
  + **Technologies & Tools**- Pyspark, and various python libraries like pandas, NumPy, matplotlib etc.
  + **Project – 4: Test Data Management** – a tool designed to handle test data and to transform the data to validate further test analysis
  + **Tenure**: March 2022 – June 2022

### Roles and Responsibilities

* + Generated random test datasets according to the business requirement and validated with the original data.
  + Performed various transformations on the test data to match the results with the original dataset.
  + Handled various encryptions to maintain Data Integrity of the original datasets.
  + **Technologies & Tools**- Pyspark, Spark, Python, MySQL, AWS S3, AWS Glue Services, Azure Blob

### 2. 2018-12 -

**2021-12**

## Application Developer

#### Accenture, Bengaluru

* ***Project -1: QBE Insurance****.*
* ***Tenure****: December 2018 – December 2021*
* **Roles and Responsibilities:**
* Designed and developed application scripts for test scenarios.
* Collaborated with stakeholders regarding project capabilities and limitations to deliver optimal functionality.
* Worked on Screening the Customer data before carrying out business
* Deploying application to Dynatrace – A self- monitoring tool using AI technology which requires no manual intervention in monitoring the application
* Providing technical support to oracle application
* 6 months of experience in Migration and data handling in hive database and Hadoop ecosystems such as HDFS
* **Technologies & Tools** - Oracle Enterprise Data Quality, Dynatrace, Hive, Hadoop map-reduce, HDFS, SQL, ETL, Unix

# Education

**May-2018 Bachelor of Engineering: CSE**

*VTU (Global Academy of Technology)*

Percentage: 73.4

**April-2014 Intermediate**

*Sri Bhagwan Mahaveer Jain College*

Percentage: 86.00

**May-2012 Matriculation**

*Vijaya Bharathi Vidyalaya*

Percentage: 86.00

# Accomplishments

* + *Azure* cloud certified on AZ-900 from Microsoft Corporation
  + *Big Data Spark Certified from IBM (Big Data University*

*Questions :*

1. How conversant are you with Python/Scala on a scale 1-5

Ans:

Hands on with Python : Rated in Python : 3/5

1. Would you be interested to take a data structures and Algorithm test in any programming language of your choice

Ans: Yes , ready to take up the test .

1. What Development Techniques do you adopt for your data project?

Ans: Designing the pipeline, Development of each of model structure, handling the pipeline .

Technologies using: Python, Spark

1. Explain in brief what measures were taken to assure data quality.

Ans: Having data check-up . Will have some transformation to check out how good the data is . Will check up that is there any duplicate or irrelevant data.

1. What data Modelling approach did u adopt for enabling Business Intelligence If any

Ans: Not answered.

1. What technical gaps do you want to work on once you join our organisation?

Ans: Will contribute to the development area.

.