



# Sai Vamsi Dudigam

## DATA ENGINEER

Experienced Data Engineer with a strong foundation in Data Science and Data Analysis. Proven track record in building and optimizing scalable data pipelines and ETL processes across AWS, Azure, and GCP. Expertise in Snowflake, Databricks, and data warehousing, leveraging Python, Scala, and PowerShell for automation. Skilled in big data technologies like Spark, Hadoop, and Hive to design high-performance architectures that enable real-time reporting and data-driven decision-making.

 saivamsidudigam@gmail.com

 +1 (216) 333-7191

 Tampa, Florida, USA

 [LinkedIn](#)

 [Portfolio](#)

## TECHNICAL SKILLS

**Programming Languages:** C, C++, Java, R, SQL, PL/SQL, Python, Scala, Shell Scripting, MATLAB, Machine Learning.

**Big Data Technologies:** HDFS, MapReduce, Hive, Pig, Sqoop, Flume, Oozie, Kafka, Cassandra, Pyspark, Spark, Spark Streaming, HBase, Impala, Zookeeper.

**Cloud Platforms:** Azure, AWS, GCP, Snowflake.

**Databases:** Oracle, SQL Server, MySQL, MongoDB, Teradata, HBase.

**Data Warehousing & BI:** Snowflake, Amazon Redshift, Google BigQuery, Power BI, Tableau.

**Tools & Frameworks:** Apache Spark, Apache Hadoop, Airflow, Databricks, PowerShell, Git, Jenkins.

**Operating Systems:** UNIX, Linux (CentOS, Ubuntu), Windows.

## WORK EXPERIENCE

### Fifth Third Bank – Data Engineer | Remote | February 2024 – Present

- Designed and implemented ETL processes using Azure Data Factory and GCP tools, ensuring efficient data ingestion and transformation.
- Developed data pipelines in Azure Databricks and GCP Dataproc using PySpark, Spark SQL, Scala, and Python for data transformation and ingestion.
- Optimized performance using Spark's in-memory capabilities, partitions, and broadcasting.
- Integrated Azure, AWS, and GCP technologies (SSIS, YARN, Spark) for scalable data transformation and reporting.
- Created data models and schemas for analytics in Snowflake, Google BigQuery, and Amazon Redshift, optimizing query performance.
- Implemented data quality checks and cleansing using Azure Data Quality Services, GCP Dataflow, and automated workflows with Bash and PowerShell.

**Environment:** SQL, SSIS, SSAS, Azure, PowerShell, Power BI, MapReduce, HBase, JSON, Spark, Hive, Pig, Hadoop, YARN, Spark Core, Spark SQL, Scala, Python, Java, Hive, Sqoop, Impala, Oracle, Google BigQuery, Cloud Dataflow.

### Database Student Coordinator | Kent State University – India Centre (Part-time) | Jan 2023 – May 2023

- Managed the migration of the university website from Drupal to WordPress.
- Led database updates and ensured seamless integration of content during the transition.

### Perfexion – Data Associate | Hyderabad, India | May 2021 – June 2022

- Performed data transformations using Spark DataFrame API, Hive, and MapReduce, loading data into Hive.
- Led the migration of the application to AWS, utilizing services like EC2, S3, and AWS Data Pipeline.
- Optimized data processing using AWS Elastic MapReduce (EMR) and integrated data into Amazon Redshift and AWS Data Warehouse.
- Integrated data from NoSQL databases (e.g., HBase) into AWS Data Warehouse solutions.
- Utilized AWS Databricks for scalable data processing and machine learning.

**Environment:** AWS, AWS Databricks, AWS Data Warehouse, NoSQL, AWS Lambda, EC2, S3, EMR, MapReduce, Hive, Pig, Spark, Python, PySpark, Linux, Hadoop, Shell Scripting, PL/SQL, Agile Methodologies.

### HDLC Info Tech – Data Science Intern | Chennai, India | April 2020 – April 2021

- Implemented data transformation pipelines on Snowflake Cloud, optimizing data processing and utilizing native data-sharing features.
- Integrated Snowflake with Apache Spark for high-performance ETL processes and developed scalable data models.
- Automated job scheduling with Airflow for timely data processing and reporting.
- Designed frameworks to convert raw data into actionable insights and developed dynamic dashboards for business analytics.

**Environment:** Snowflake, Apache Spark, Airflow, Python, SQL, ETL.

### HDFC Bank – Data Science Apprentice | Chennai, Tamil Nadu, India | Remote | Feb 2020 – April 2020

- Analyzed and monitored cyber frauds to enhance security protocols for secure banking transactions.
- Generated Weekly MIS reports to track key performance indicators and provide actionable insights for business and security teams.
- Conducted bank account reconciliation, ensuring data integrity and accuracy across various financial systems.
- Contributed to data warehousing efforts to improve Customer Relationship Management (CRM) by aggregating and analyzing customer data for enhanced service delivery.

### Revolution Medicines – Data Entry Specialist (Freelance) | Kochi, Kerala, India | May 2018 – Jan 2020

- Specialized in hospital data entry, ensuring accurate and up-to-date patient records in electronic health systems (EHR).
- Managed confidential patient information with a focus on privacy and compliance with healthcare regulations.
- Supported healthcare teams by providing timely and accurate data updates, contributing to efficient medical operations.
- Utilized data management tools to maintain and streamline data entry processes, improving the accuracy and reliability of patient records.

## EDUCATION & CERTIFICATION

**MS in Computer Science | Kent State University – USA | August 2022 – December 2023**

**Bachelor's in Computer Science | Hindustan Institute of Technology and Science – India | June 2018 – May 2022**

Microsoft Certified: Azure Fundamentals (AZ-900)

Microsoft Certified: Azure Developer Associate (AZ-204)

AWS Certified: Data Analytics – Specialty

Google Professional Data Engineer