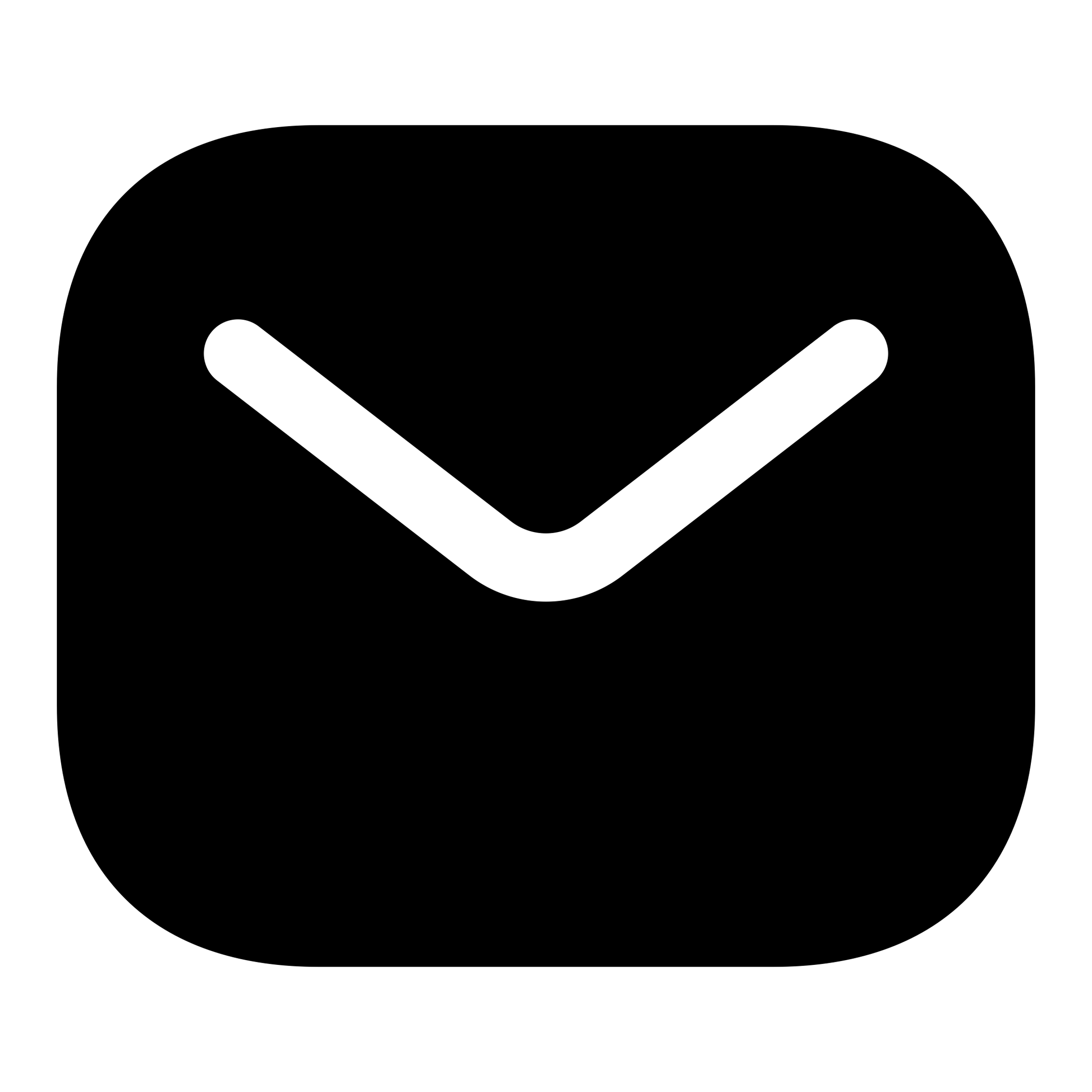
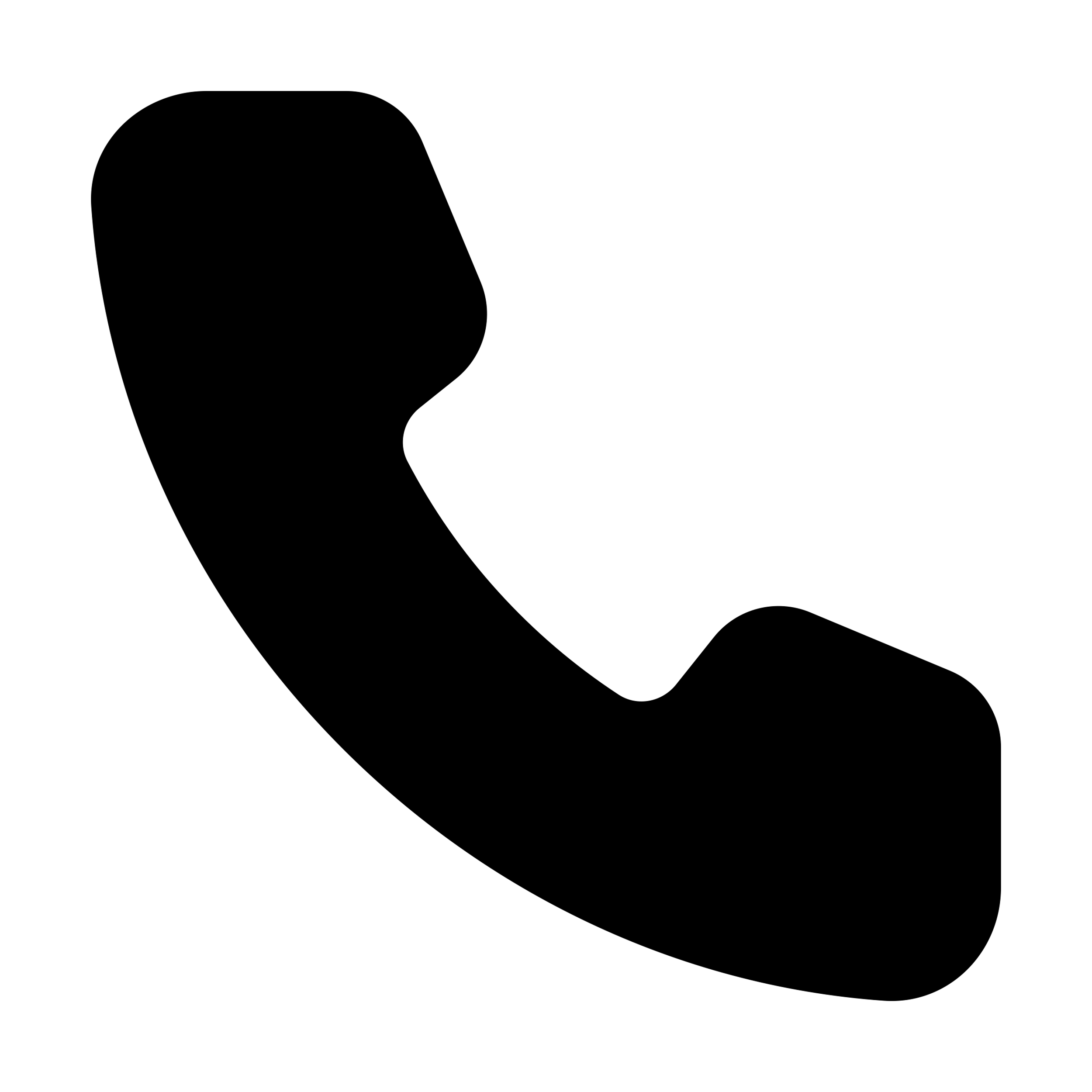
AKHIL REDDY BOKKALA

* Irving, TX |  [akhilreddy.r1127@gmail.com](mailto:akhilreddy.r1127@gmail.com) |  Contact: +1 (945) 232 9416

|  |
| --- |
| PPROFESSIONAL SUMMARY |

**GCP Data Engineer with 4+ years** of experience in designing and implementing data solutions on Google Cloud Platform. Achieved a 25% improvement in data processing efficiency through Cloud Dataflow and Apache Spark. Delivered actionable insights using BigQuery and created visualizations with Looker and Google Data Studio. Skilled in developing ETL pipelines, integrating diverse data sources, and managing real-time data streaming with Google Cloud Pub/Sub. **Google Cloud Certified - Professional Data Engineer.**

|  |
| --- |
| SKILLS |

**Databases:** BigQuery, Cloud SQL, Snowflake, PostgreSQL, MySQL, Oracle DB, MongoDB, HBase.

**Big Data:** Google Cloud Pub/Sub, Apache Spark, Cloud Dataflow, Cloud Dataproc, HDFS, MapReduce, Hive, Kafka, Sqoop, Spark SQL, Oozie.

**ETL Tools:** Cloud Data Fusion, Cloud Composer, Google Cloud Dataflow, Apache Airflow.

**Languages:** Python, Scala, SQL.

**Data Modeling:** Data Vault 2.0, Data Integration, Data Migration, BigQuery ML.

**Containerization and Orchestration:** Google Kubernetes Engine, Docker, Kubernetes.

**Monitoring and Logging:** Google Cloud Monitoring, Google Cloud Logging, Stack driver.

|  |
| --- |
| PROFESSIONAL EXPERIENCE |

# Fifth Third Bank Irving, TX

# Data Engineer 08/2023 – Present

* Built real-time data pipelines with GCP Pub/Sub, Dataflow, and Hadoop, processing 2 TB of daily IoT data.
* Optimized data ingestion workflows on Hadoop and GCP, reducing latency by 25% for large-scale data.
* Designed scalable storage solutions using GCP Cloud Storage and Hadoop HDFS, managing 8 TB+ of structured and unstructured data.
* Enhanced query performance using BigQuery and Hive, enabling analysis of 50M+ daily sensor events.
* Automated ETL processes with Cloud Dataflow and Spark, cutting manual data handling by 40%.
* Implemented predictive models in Looker with GCP and Hadoop, reducing system downtime by 15%.

# Addicor Technologies Private Limited Noida, India

# Junior Data Engineer 01/2020 – 04/2022

* Automated ETL workflows using Cloud Composer and Apache Oozie, orchestrating 50+ daily data pipelines from APIs and databases.
* Streamlined data ingestion with Cloud Dataflow and Hadoop MapReduce, reducing processing time by 30% for real-time business intelligence.
* Loaded over 10 TB of data into BigQuery from Hadoop HDFS, improving query performance by 20%.
* Optimized ETL processes with DataPrep and Apache Hive, cutting manual data wrangling time by 25%.
* Enhanced workflow efficiency by 20% using Cloud Composer and Airflow, automating and monitoring ETL tasks.

|  |
| --- |
| EDUCATION |

**The University of Texas at Arlington 12/2023**

Master of Science in Data Science

|  |
| --- |
| PROJECTS |

**Real-Time Big Data Streaming and Analytics Platform**

* Built a scalable Data Lakehouse with GCP Cloud Storage and BigQuery, managing 5+ PB of structured and unstructured data, integrated with Hadoop.
* Optimized query performance by 30% with BigQuery and Hadoop Hive for large-scale healthcare analytics.
* Streamlined data management, reducing storage costs by 20% with Cloud Storage lifecycle policies and Hadoop HDFS optimization, while improving processing efficiency by 25% with Dataproc.

**End-to-End ML Pipeline with Vertex AI**

* **Developed ML pipelines with Vertex AI and Kubernetes (GKE), handling 3+ TB of data and boosting prediction.**
* **Automated data processing using Dataflow, cutting preparation time by 40%, and optimized performance in BigQuery by 30%, reducing churn by 12%.**

|  |
| --- |
| PROJECTS |

**Google Cloud Certified - Professional Data Engineer**