|  |  |  |
| --- | --- | --- |
| **EDUCATION** | |  |
| **The University of Texas at Dallas** | | May 2021 |
| *M.S., Management Information System* | | **GPA 4** |
| **Visvesvaraya Technical University** | | June 2017 |
| *B.E., Electronics and Communication* | | **GPA 3.8** |
|  | |  |
| **TECHNICAL SKILLS** | | | |
| Languages: | SQL, Python, Scala, R, SAS, Shell, PostgreSQL, Hive | | |
| Hadoop Stack: | HDFS, Spark, Kafka, Airflow, Sqoop, Flink, Flume, YARN, Ambari, HUE | | |
| Software: | Tableau, Talend, Informatica, QlikView, Docker, Kubernetes, Power BI, Alteryx, GitHub | | |
| AWS services: | S3, EC2, EMR, DynamoDB, RDS, Redshift, Kinesis, Lambda, Athena, QuickSight | | |
| Database: | MySQL, PostgreSQL, MongoDB, Cassandra, Snowflake, HBase, Redis, Bigtable | | |
| Certification: | AWS Solution Architect, Cloudera Certified Associate, Tableau Desktop Specialist, Alteryx Designer | | |

**BUSINESS EXPERIENCE**

**Micron Technology, Boise, Idaho** May 2020 – August 2020

*Data Engineer*

* Developed the infrastructure required for optimal extraction, transformation, and loading of 100GB data from various sources like AWS S3, Dynamo DB, RDS to store in Redshift and resulting in increased efficiency by 7%
* Collaborated with various stakeholders including the Data scientists, Analysts and Business intelligence teams to support their data infrastructure needs and automated ETL process to reduce manual workload by 20 hours every month
* Designed and prepared interactive and intuitive dashboards and reports using Tableau from Redshift warehouse to contribute to Micron’s competitive intelligence effort
* Analyzed and determined the key strengths of our competitor’s devices in the area of performance, latency, power, cost

**Accenture, Bangalore, India** June 2017 – June 2019

*ETL Developer*

* Worked with 5+ clients to gather new requirements and prepare design and test plan documents for work parcel
* Migrated data from SQL Server and Oracle to Cloudera Hadoop Cluster-Impala using Sqoop which increased execution speed by more than 35%
* Utilized Big data technology tools to interact with HDFS for ETL of disparate data source like transactional data, web services data (RESTful), payment and marketing campaign data and load it into Snowflake warehouse
* Designed and developed various kinds of reports like scorecards & dashboards, KPI, operational reports, business reports and ad-hoc reports for business analytics in Apache superset
* Built SQL queries to improved CPU performance and server utilization by 60% by debugging bad queries

**RELEVANT PROJECT**

**Uber Real-time data mining using Spark ML**

* Created a data processing pipeline to detect and visualize user location practices from Uber event data using Spark
* Published driving location data to MapR-ES topic using Kafka API and transformed using Structured Streaming
* Deployed machine learning models (K-Means) on the transformed data to group the Uber customers
* Created Tableau dashboard for Geospatial analysis of the customer’s data stored in Cassandra

**Data Pipeline for COVID 19 analysis in AWS**

* Created a pipeline and orchestrated for continuous integration of the data processing using Airflow
* Real time streaming of data using NiFi and stored in data lake by parsing complex Json data into csv
* Parsed data was sent through Kafka and processed using PySpark
* Transformed data was stored in HDFS and visualized in tableau using Hive

**ORGANIZATIONS**

* Office of Student Volunteerism, The University of Texas at Dallas – VolunteerAugust 2019 – Present
* Data Science Club, The University of Texas at Dallas – Member August 2019 – Present