

VENKATA NELLURI

Dallas, TX | Contact: 1 510-953-2111 | email: nellurivenkata@gmail.com | [LinkedIn](#) | [HackerRank](#)

Technical Skills: Machine Learning, Artificial Intelligence, Pandas, NumPy, SciPy, Matplotlib, SQL, Python, Scala, Java, Tableau, PowerBI, GoogleDataStudio, MySQL, Oracle, Hadoop, HDFS, Hive, Pyspark, Hive, MongoDB, Sqoop, Kafka, SparkSQL, HBase, S3, EC2, EMR, CloudWatch, ADF, ADLS, Databricks, Event Hub, Snowflake, Git, JIRA, Confluence, Jenkins, CI/CD, Airflow.

EXPERIENCE

Uber – Platform and Capacity Engineering

Oct 2022 – Till Date

Data Engineer

Remote

- Leading the effort of migrating 50+ dashboards from Looker to Tableau for easy navigation and saving time by 30% in accessing the reports and allowing business to make budget and capacity forecast planning for compute and storage.
- Developing data pipelines in airflow to read 10M+ events every hour from Kafka and ingest into aggregation tables to monitor Data Infrastructure Insights, providing stakeholders with real-time access to key performance indicators and metrics.
- Enhancing existing data pipelines to fix the data anomalies and critical production issues ensuring that the data was well-organized and accessible to stakeholders, and scheduling them using Airflow to automate the process.
- Led the development effort for POC developing new Insights and demonstration to leadership and got project approval.
- Wrote python scripts to read 100K+ JSON files from terra blob having 900 attributes and load the load data to Hive.
- Developed a pipeline using Python and Airflow to load data and develop dashboards for 4 internal applications in Google Data studio to provide Application Insights like Active Users (Adoption), Latency metrics, Issue tracking and SLA metrics.

Apple – Ads Platform Engineering

Sep 2021 – Sep 2022

Big Data Engineer

Austin, TX

- Led the Ads Platform Prod Support Team in maintaining and improving performance of streaming applications with 600M+ events per day and improved pipeline efficiency and issue reduction by 65% and improve SLA by 50% by enhancing the application to track pipeline status, monitoring proactively and saving \$100K+ revenue loss per month.
- Enhanced applications using Airflow, EKS, S3, EMR, Time Series DB, and Cymatics to track processing delays and identify areas for improvement, ensuring that the applications met the highest standards of performance and reliability.
- Designed and implemented Data validation process ensuring accuracy and consistency across multiple databases.
- Collaborated with cross-functional teams identify business gaps and develop data solutions ensuring data consistency.

Albertsons – Data Governance and Enterprise Data Warehouse

Feb 2021 – Sep 2021

Data Engineer

Phoenix, AZ

- Designed and developed data pipelines using various technologies such as Kafka, Azure Data Factory, Azure Data Lake, and Snow pipe to ingest XML/JSON data into Snowflake data warehouse.
- Developed analytics data pipelines using ADF, Pyspark and Databricks to ETL data into Snowflake for business insights.
- Led the effort in enforcing data quality by applying business logics and implementing checks using iCEDQ tool.
- Worked closely with stakeholders to understand business requirements, perform gap analysis for Data Catalog and Data Lineage, and ensure metadata is curated and updated regularly.
- Developed python utilities to automate metadata curation and update the catalog with the latest information in Alation.
- Maintained documentation for data pipelines, processes, and best practices to ensure efficient knowledge transfer and onboarding for new team members.

Citi – Compliance, Digital and Enterprise Data Warehouse

Sep 2018 – Jan 2021

Data Engineer

Tampa, FL

- Designed and implemented automated data pipelines using Spark for daily batch ingestion from Hive to MongoDB, transforming and loading data for analysis and visualization purposes.
- Developed and maintained Python scripts for running queries in Hive and MongoDB, generating daily batch status reports, and performing data quality checks to ensure regulatory compliance.
- Migrated batch processing to real-time data ingestion using Spark Streaming for scalable data processing.
- Led the data acquisition and loading from various sources into MongoDB for building data visualizations and supporting machine learning applications to run marketing campaigns for Credit card customers with promotions.
- Managed ETL pipelines that processed 8K-10K batch jobs per day and 200K+ batch executions per month, handling over a million records each day from data warehouse platforms to AML target systems.
- Maintained business intelligence cubes and MicroStrategy dashboards, providing technical support for big data applications and automating repetitive data export tasks using SQL and SAS scripts.
- Improved process efficiency and saved 90% of time for manual processes by writing Python scripts for process automation.

Epathusa – HR Analytics

Jul 2018 – Aug 2018

Data Engineer

Des Moines, IA

- Designed and implemented data pipelines to extract, transform, and load data from various sources into a centralized repository using technologies such as SQL Server Integration Service.
- Developed and maintained real-time dashboards and reports using tools such as Power BI, Tableau, or other similar platforms, to provide stakeholders with actionable insights into key business metrics and trends.

Education

Master of Science| Computers & Information Science | Southern Arkansas University, Arkansas. (Graduation – 2018, GPA - 3.9)

Bachelor of Technology | Electronics and Communication Engineering | JNTU Kakinada, India. (Graduation – 2015, GPA - 3.3)