SAI SANTOSH EEDUPUGANTI

saisantosheedupuganti@gmail.com| +1 5184969581|LinkedIn | Github

EDUCATION:

Master's in Data Science

University at Albany, SUNY   May 2020

Bachelor of Mechanical Engineering

Jawaharlal Nehru Technological University, Hyderabad   June 2018

TECHNICAL SKILLS:

Programming Languages	: Python, R Programming, HTML, CSS
Frameworks/Libraries	: Hadoop, PySpark, Hive ,Pandas, NumPy, Scikit, BeautifulSoup, Selenium
DevOps/BI/Tools	: AWS S3, AWS EMR, AWS Batch, AWS Glue, Databricks, Docker, Jenkins, Tableau, Qlik sense, QuickSight, Apache Airflow
Databases and repositories	: AWS Redshift, Snowflake, MySQL, MongoDB, GIT, BitBucket, Jira

WORK EXPERIENCE:

Amazon, Business Intelligence Engineer   **January 2022 – Present**

ACES FMP Team -Develop Machine Learning and Analytical Solutions designed to optimize business processes and enhance operational efficiencies within warehouses

- Devised and executed complex SQL queries to establish ASN Compliance metrics, enabling the identification of high-defect vendors and driving a targeted coaching program that contributed to an annual savings of $36 million.
- Designed and developed ETL pipelines using Apache Airflow, automating data workflows that enhanced the monthly reporting of ASN compliance metrics, leading to a 20% increase in operational efficiency.
- Transitioned legacy ETL jobs to cloud technologies like AWS Glue and Airflow, achieving an 11% cost reduction through better resource utilization.
- Leveraged AWS QuickSight and Tableau to create dynamic, interactive dashboards that distilled complex datasets into clear, actionable reports for business stakeholders, improving data-driven decision-making.
- Built and deployed Python and Pyspark scripts for complex data transformations, boosting AWS Redshift's analytical capabilities by 21%.
- Crafted an AWS QuickSight dashboard, orchestrated via Apache Airflow, that aggregated key performance indicators (KPI's), providing a holistic view of business health for leadership's use in Monthly and Weekly Business Reviews.
- Designed and implemented an automated dashboard to track inventory reinstatement in non-sort warehouses, due to theft losses, and conducted Gross Merchandise Sales (GMS) analysis aligned with Amazon.com sales, deriving key trends and strategic insights for inventory management.
- Provided pivotal ad-hoc data analysis support to leadership during peak seasons, enabling agile decision-making and strategic operational adjustments.

Technologies: AWS Redshift, Python, Pyspark, AWS Quicksight, Tableau, Apache Airflow, Docker, ETL Pipelines, AWS S3, AWS CloudWatch, AWS Glue, Asana, Gitlab

Verizon, Data Analyst/ Data Engineer   **January 2021 – January 2022**

DENA Team -Collects all the Events related to Verizon FIOS media Service and provide Business analytics service to build visualizations provide insights from data.

- Developed end-to-end data pipelines with Python and Spark, integrating Kafka events into Snowflake for centralized data storage and management.
- Engineered continuous data ingestion into Snowflake utilizing Snow Pipes and Streams, with automated scripts for data loads from AWS S3 staging areas.
- Conducted comprehensive data analysis to identify peak channel viewership, churn rates, and webpage UI hotspots, enhancing ad placement strategies using PySpark on AWS EMR notebooks.
- Managed the creation of on-demand S3 tables, employing AWS services such as Lambda functions and AWS Glue, leveraging Python and PySpark for data manipulation.
- Crafted complex SQL queries for business intelligence reporting and constructed interactive dashboards for data visualization with Qlik Sense.
- Implemented Apache Airflow for robust ETL pipeline development, facilitating advanced analytical reporting and dashboard creation.
- Addressed data quality issues and monitored data flows, utilizing Splunk for effective data traffic oversight.

Technologies: Python, Snowflake, Qlik sense, PySpark, Apache Airflow, Docker, Splunk, AWS S3, AWS EMR, Bitbucket, JIRA

Capital One, Data Analyst/ Data Engineer   **August 2020 - January 2021**

Fraud analytics- Monitors Fraud's in Transactions, identify anomalies developing solutions to solve them

- Executed data migration processes from SQL Server to Snowflake, ensuring seamless data transition and integrity
- Developed an automation process to detect and log data discrepancies in transaction volumes, integrating with Jira for ticketing and team assignments.
- Devised a method to identify anomalous transactions by analyzing geo-referenced data and comparing against historical patterns
- Created dynamic Tableau dashboards to track and analyze transaction spikes and volume trends.
- Achieved over 10% cost savings in incident data processing by constructing a comprehensive ETL pipeline with Python and Airflow.
- **Technologies:** Python, Snowflake, Qlik sense, PySpark, Apache Airflow, Docker, Splunk, AWS S3, AWS EMR, Bitbucket, JIRA

Data Analyst, State university Of New York, Albany   **May 2019 - May 2020**

NYSDEC Team- Perform water quality analysis of New York state's water bodies laid throughout the state.

- Developed a data pipeline by extracting data through APIs using Python; analyzed and visualized over two decades of data with Tableau.
- Engineered multiple MySQL relational database schemas for efficient storage of JSON-formatted data retrieved from APIs.
- Conducted exploratory data analysis on historical datasets, deriving insights that contributed to resolving water quality issues.
- Constructed regression models, including Linear Regression and tree-based models, to forecast water chemical levels for consumption safety.
- Facilitated data management and accessibility by organizing and creating data frameworks for stakeholders and peers.