SALINI IVATURI
DATA ENGINEER
Phone: 6696009198 | Email: salini.i@careernb.com | Location: CA

SUMMARY

- 4+ years of experience as a Data Engineer with expertise in designing data-intensive applications using Cloud Data engineering, Data Warehouse, Hadoop Ecosystem, Big Data Analytical, Data Visualization, Reporting, and Data Quality solutions.
- Skilled in leveraging Big Data technologies, including the Hadoop and Spark frameworks, and proficient in associated tools such as MapReduce, Hive, Pig, BigQuery, HDFS, Spark, and HBase.
- Expertise in SQL, database design, and ETL pipeline implementation, experience in AWS (EC2, S3, Redshift, Lambda, IAM, Kinesis, DynamoDB, Glue), and Azure (Azure DevOps, Azure Data Lake, Azure Data Factory, Azure Databricks), showcasing expertise in leveraging diverse cloud services.
- Ability to implement continuous integration/continuous delivery (CI/CD) pipelines for automated data infrastructure deployments.

TECHNICAL SKILLS

Programming Language: Scala, Python, SQL
IDE's: PyCharm, Jupyter Notebook
Big Data Ecosystem: MapReduce, Hive, Pig, HDFS, Spark
Machine Learning: Linear Regression, Logistic Regression, Decision Tree, SVM, K mean, Naïve Bayes, Random Forest
Cloud Technologies: AWS (EC2, S3, Amazon Redshift, Glue, Lambda, IAM, Kinesis, AWS Pipeline), Azure (Azure DevOps, Azure Data Lake, Azure Data Factory, Azure Databricks)
Packages: NumPy, Pandas, Matplotlib, SciPy, Scikit-learn, Seaborn, TensorFlow, Kafka, PySpark
Reporting/Other Tools: Tableau, Power BI, SSRS
ETL/Database: MS SQL Server, PostgreSQL, MongoDB, MySQL, SSIS
Operating Systems: Windows, MacOS

EDUCATION

Masters of Science in Computer Science — **Dec 2023**
California State University East Bay, Hayward, CA — GPA: 3.8/4.0

Bachelor of Technology in Electronics and Communication Engineering — **April 2015**
Jawaharlal Nehru Technological University, Kakinada, India — GPA: 7.8/10

WORK EXPERIENCE

Allstate, TX — **Jan 2023 – Present**
Data Engineer

- Develop a real-time data processing pipeline using Kafka and Python, enabling faster anomaly detection and response.
- Automated 80% of data pipeline tasks using Python scripts and Airflow workflows, resulting in 15% reduced operational costs.
- Reduce ETL processing time by 30% through code optimization and efficient data transfer techniques.
- Implement 5+ MapReduce jobs to process 10TB of data daily, achieving 20% faster processing times than traditional ETL methods.
- Integrate Lambda with other AWS services (API Gateway, SNS, SQS) to build event-driven data architectures, enabling real-time notifications and actions.
- Establish complex HiveQL queries to join data from multiple sources, supporting diverse analytics requirements.
- Increase data pipeline throughput by 35% by optimizing Spark configurations and parallelization techniques.
- Improve data quality by 15% by implementing AWS Glue dynamic data filtering and deduplication capabilities, resulting in cleaner and more reliable data sets.
- Orchestrate complex data pipelines with 10+ stages for data ingestion, transformation, and analysis using AWS Pipeline's visual workflow builder.

Sigma Tech Solution, India — **Jan 2016 – Mar 2019**
Data Engineer

- Designed and implemented a star schema in Snowflake for efficient data warehousing, leading to a 15% improvement in query performance.
- Utilized NumPy and Pandas libraries in Python to clean and preprocess terabytes of data, improving data quality by 18%.
- Utilized Kafka offset management and replay capabilities to guarantee data reliability and consistency.
- Automated data extraction, transformation, and loading processes using SSIS, resulting in 20% improved data accuracy and reduced manual effort.
- Successfully integrated Azure Data Lake Storage and Data Lake Analytics to streamline data processing workflows, resulting in a 25% reduction in data processing times.
- Leveraged the autoscaling capabilities of Azure Databricks to optimize resource utilization and reduce data processing costs by 15%.
- Integrated code versioning and continuous integration/continuous delivery (CI/CD) practices with Azure DevOps, enabling faster deployments and reduced risk of regressions.
- Exploited Spark SQL to join complex datasets with a 25% improvement in query performance.

CERTIFICATION

AWS Solutions Architect - Associate (AWS SAA)