

NAGA SRIVALLI SOWDAMINI, NANDIGAMA

[in LinkedIn](#) | [515-771-7713](tel:515-771-7713) | nnsvsowdamini@gmail.com

Skills

- Spark | PySpark | Hadoop | Hive | Mapreduce | Pig | Unix | MSSQL | Python | Java | Sqoop | Impala | MongoDB | Git
- AWS | S3 | Redshift | DynamoDB | Cloud Formation | Cloud Watch | Lambda | IAM | Azure | Azure Data Factory | Azure Databricks
- Azure SQL | Tableau | Informatica | Machine Learning Algorithms | R | Visual Studio | Win SCP | R Studio

Experience

AWS Data Engineer

Wellmark Inc., Contract

Des Moines, IA, USA 03/2023 - 08/2023

- Designed and implemented ETL framework using **AWS S3, IAM, Glue, Lambda, Cloud watch, Cloud Formation** technologies which extracts CSV data from S3, using py-spark code performing multiple transformations, aggregations on columns.
- Ensured Data Quality checks on the data through regular data validation and cleansing procedures. Improved data storage and retrieval efficiency by optimizing SQL queries.
- Designed and implemented scalable ETL processes, resulting in a 30% reduction in data processing time, maintained data pipelines using Python and Apache Spark, handling 1TB+ of data daily.
- Automated routine data tasks, reducing manual efforts by 40%. Handled special characters in data through User-built functions in existing Stored Procedures, deployed the changes through service now CRQ to multiple environments.
- Continuous Integration/Deployment Pipeline Integration, pull requests, code reviews, unit/regression/integration/e2e testing.

Senior Analyst

Accenture

Bangalore, India 05/2021 - 12/2021

- As a Senior Analyst, worked with data scientists to provide insights on the pharmaceutical data in structured format, for the Directors accessing a portal for drug analytics from previous months using the latest technologies of **Databricks, Py-spark, AWS S3**.
- Designed and implemented py-spark scripts to include the parameters like variance, correlation, RMSE, MSE, R2 and adjusted R2 to understand the raw data and the relation of each column and their dependency on the other columns.
- Used Linear regression, Classification models on various columns to deduce the data patterns on particular columns mentioned in the business requirement.
- Led the development of several components of E2E, from identifying business requirements and data dependencies..
- Continuous Integration/Deployment Pipeline Integration, pull requests, code reviews, unit/integration/e2e testing

Azure Data Engineer

Infosys

Hyderabad, India 05/2020 - 04/2021

- Developed complex ETL processes ensuring efficient data extraction, transformation, and loading using technologies like **Azure Functions, Azure Data Factory and Azure Databricks**..
- Designed and developed the Py-spark scripts to load the historical data and delta data from various on premises data like SAP HANA etc.
- Implemented CI/CD of changes through **Azure DevOps**, raising CRQ's in Service now for the approval of the changes to deploy changes.
- Played a key role in migrating on-premises data to Azure, resulting in cost savings of 10% and improved data availability.

Spark Developer

Infosys

Bangalore, India 08/2017 - 04/2020

- Developed Spark applications in **Scala** and **Python**, optimizing performance for critical data processing tasks, Created complex ETL pipelines using Spark for data transformation, cleansing, and aggregation.
- Conducted performance tuning and optimization of **Spark** jobs like broadcasting, salting, resulting in a 25% improvement in processing.
- Developed **Mapreduce** UDF in **hadoop** which could be used for optimizing data load through **Hive** scripts, before migrating to Spark..
- Worked on Grafana Dashboard to populate the data on the jobs, summary of daily patterns used by the board of directors in the decision-making process for the new products being launched.

Education

Master of Sciences

Northwest Missouri State University

Maryville, Missouri, USA 02/2022 - 03/2023

- Major in Data Analytics

Bachelor of Technology

Koneru Lakshmaiah College of Engineering

Vijayawada, India 07/2013 - 04/2017

- Major in Computer Science

Projects

- **Machine Learning For Health Insurance Premium**: Built machine Learning models for health insurance premium prediction (03/2023)
- **Comparison of BigData Algorithms**: Published Paper on Health care dataset on clustering, classification, regression algorithms (02/2017)

Others

- **AWS Cloud Practitioner Certification** : Cloud Practitioner Certification (30/12/2022)