ARAVIND KOMMINENI | DATA ENGINEER

PROFESSIONAL SUMMARY

As a versatile professional Data Engineer 3.5+ Years with a proven track record combining technical expertise with strategic thinking to deliver innovative and scalable data solutions. Adept at leveraging a diverse skill set in Python scripting, data acquisition, manipulation, and modelling to design and optimize robust ETL pipelines. Proficient in database design, CI/CD implementation, and cloud services such as AWS. Demonstrated success in collaborative candidate evaluation for data engineering positions, assessing proficiency in big data technologies and cloud platforms.

WORK HISTORY

Data Engineer | Sep 2022 - Current
EPIC - USA

- Actively writing Python scripts for Data Acquisition, Data Manipulation, and Data Modelling, creating robust and scalable Data Integration (ETL) pipelines in Agile development process.
- Applying knowledge of data management fundamentals, data storage principles, ETL, Data Modelling, and Data Architecture to build and optimize pipelines.
- Designed and created RDBMS tables, views, user-created data types, indexes, stored procedures, cursors, and triggers. Implemented optimization techniques for existing pipelines and queries.
- Instantiated, created, and maintained CI/CD pipelines, applying automation to environments and applications. Utilized automation tools like GIT for deploying packages in test and production environments, and conducting end-to-end testing.
- Created proof of concepts and working models for technical feasibility studies using the latest features in technologies like AWS, Google Cloud Platform, Python, REST API, and various databases.
- Played a crucial role in collaborative candidate evaluation for data engineer positions, assessing proficiency in big data technologies, Python, SQL, and AWS services to build a skilled and capable team.
- and deployed data solutions using AWS services, focusing on enhancing pipeline efficiency and scalability through the application of AWS-specific tools and frameworks.
- Utilized AWS's machine learning and AI services to enhance data models and predictive analytics capabilities, contributing to more advanced data-driven decision-making.
- Demonstrated expertise in managing databases using AWS's managed services, ensuring optimal performance and reliability of data storage and processing environments.

Data Engineer | Jun 2019 - Jun 2021
Magna Info Tech - India

- Designed and implemented scalable big data architectures tailored to business needs, balancing performance requirements with considerations for data volume, velocity, and variety.
- Applied in-depth Extract, Transform, Load (ETL) skills to orchestrate the smooth movement of data, optimizing processes for efficiency and maintaining data integrity throughout.
- Applied expertise in data modelling to design structures that optimize storage, retrieval, and processing, catering to t unique challenges posed by varying data volumes and complexities.

CONTACT

Address: Fairfax, VA, USA 22030
Phone: +1 (540)753-1598
Email: aravind.k99@aogjob.com

SKILLS

Data Components: HDFS, Hue, MapReduce, PIG, Hive, HCatalog, HBase, Sqoop, Impala, Zookeeper, Flume, Kafka, Yarn, Cloudera Manager, Kerberos, Pyspark Airflow, Kafka Snowflake
Spark Components: Apache Spark, Data Frames, Spark SQL, Spark, YARN, Pair RDDS
Cloud tools: AWS (Redshift, Athena, Glue, PageMaker, S3), Data pipelines, Airflow, IAM, CloudFormation, EC2, ELB/CLB, Lamda
Methodologies: SDLC, Agile/ Scrum, Waterfall
Language: Python, SQL, R
Packages: Pandas, NumPy, MatplotLib, SciPy, Scikit-Learn, SeaBorn, PyTorch. TensorFlow, ggplot2, Plotly, Keras, LangChain
IDES: Visual Studio Code, PyCharm, Jupyter Notebook
Data Analytics Skills: Data Manipulation, Predictive Analysis, Data Cleaning, Data Mining, Data Visualization, Statistical Modelling, Exploratory Data Analysis
Machine Learning: Predictive Modelling, Neural Networks, Regression, Clustering, Statistics, LSTM
Databases: MySQL, PostgreSQL, MongoDB, MSSQL, Oracle SQL
Tools: Tableau, Power BI, ArcGIS, MongoDB, Alteryx, Neo4j, Django, OpenAI, AI Advanced Excel (Pivot Tables, VLOOKUP), Talend, Microsoft office Suite (Word, PowerPoint, Excel, Outlook), Jira, Looker
ML Algorithms: Linear Regression, Logistic Regression, Decision Trees, Supervised Learning, Unsupervised Learning, Classification, SVM, Random Forests, Naive Bayes, KNN, K Means, Time Series Forecasting
Other Technical Skills: Natural Language Processing, A/B Testing, Hypothesis testing, ETL, Databricks, Hadoop, Spark, Snowflake, Big Query, Apache Airflow, Critical Thinking, Communication Skills, Presentation Skills, Problem-Solving.
Version Control Tools: Git, GitHub
Operating Systems: Windows, Linux and mac OS

clusters and Azure Databricks. Utilized AWS (IAM) for identity and access management.

- Conducted performance tuning for ETL processes, resulting in a resource utilization optimization of 20%. Utilized AWS Redshift, Athena, and Glue for optimized data processing.
- Implemented caching strategies and query optimizations in Hive, leading to a 30% improvement in analytical query performance by utilizing AWS.
- Established data quality checks and validation processes, reducing data errors by 15%.
- Implemented monitoring solutions, resulting in a 40% reduction in data quality issues through proactive identification and resolution. Used AWS CloudWatch.

EDUCATION

Master of Science, **Information Studies | Dec 2023**
Trine University - Angola, USA

Master of Science, **Computer Science | Dec 2022**
George Mason University - Virginia, USA

Bachelor of Technology, **Computer Science | Aug 2021**
Gokaraju Rangaraju College of Engineering - India

PROJECTS

Credit Risk Prediction:

- Combined the training and validation set and performed data preprocessing on both numerical and categorical columns.
- Conducted up and down sampling techniques for the highly imbalanced target variable.
- Performed various classification algorithms Gaussian Naive Bayes baseline model, knn, Random Forest Classifier achieving the best accuracy for the last model.

Color Channel Perturbation Attack for Fooling CNNs:

- Investigated the performance of CNNs like ResNets and VGG in case of Color Channel Perturbation (CCP) attack and found that the performance of CNNs dropped drastically.
- Introduced a defense mechanism with Data Augmentation to overcome such attack.
- Implemented object detection using YOLO on the CCP attacked images to investigate the effect of CCP attack over object detection.