HARSHITHA ILAVALA

Denton, TX | Tel: +1 (945)-274-0849 ||harshithailavala4@gmail.com

SUMMARY:

- Data Specialist with over 3+ years of experience, including 10 months in machine learning and AI and more than 2+ years as a Big Data Engineer focused on advanced data ecosystems, cloud computing, and developing efficient data workflows.
- Skilled in Python libraries (Pandas, NumPy, Scikit-learn, TensorFlow, PyTorch) and R programming for creating predictive models, data transformation, and performing analytical computations.
- Proficient in designing and optimizing both batch and real-time data pipelines using cutting-edge technologies like Apache Spark, Kafka, and Flink.
- Extensive experience with distributed data systems, including Hadoop, Apache Spark (streaming and batch), and Kafka, to manage high-volume data processing tasks.
- Committed to ensuring data governance, integrity, and regulatory compliance while building pipelines that enable machine learning workflows.
- Skilled in non-relational databases (MongoDB, Cassandra, Cosmos DB) and SQL-based systems (MySQL, PostgreSQL, Azure Synapse Analytics) for efficient data structuring, querying, and analysis.
- Advanced knowledge of cloud services (Azure Blob Storage, Azure Functions, Azure Machine Learning) for scalable data processing pipelines and deployment of ML solutions.
- Proficient in building automated ETL workflows using tools such as Apache Airflow, Azure Data Factory, and NiFi for seamless integration of multiple data sources.
- Experienced in containerized environments and orchestration tools like Docker, Kubernetes, and Infrastructure-as-Code frameworks (Azure Resource Manager, Terraform) to streamline application scaling and deployment.
- Expertise in system observability tools such as Prometheus, Grafana, and Azure Monitor for proactive system performance tracking and anomaly detection.

TECHNICAL SKILLS:

- **Programming & Scripting:** Python (Pandas, NumPy, SciPy, Matplotlib, Seaborn, Scikit-learn, TensorFlow, Keras, PyTorch), R, SQL (MySQL, PostgreSQL), Java, Scala, Shell Scripting
- **Machine Learning & Deep Learning:** Supervised Learning (Logistic Regression, Decision Trees, Random Forest, Unsupervised Learning, Neural Networks, Time Series Forecasting (ARIMA, Prophet, LSTM), NLP (Tokenization, Sentiment Analysis)
- **Big Data Technologies:** Apache Hadoop (HDFS, YARN), Apache Spark (batch & streaming), Apache Kafka, Apache Flink, Apache Hive, Pig, Sqoop, AWS Redshift, Google BigQuery, NoSQL Databases (MongoDB, Cassandra, HBase, DynamoDB)
- **Cloud Platforms & Services:** AWS (S3, Lambda, SageMaker, Redshift), Azure (Azure ML, Databricks, Synapse Analytics, Data Factory), Google Cloud (BigQuery, Cloud Dataflow, AI Platform)
- **Data Pipelines & Orchestration:** Apache Airflow, Talend, Apache NiFi, AWS Glue, Apache Kafka, RabbitMQ, AWS Kinesis, Data Lakes (AWS S3, Azure Data Lake, Google Cloud Storage)
- **ETL Tools:** Informatica PowerCenter, Apache Nifi, Apache Kafka, SSIS, AWS Glue, Google Cloud Dataflow, Apache Airflow, Data Integration
- **Containers & Automation:** Docker, Kubernetes, Terraform, AWS CloudFormation, Jenkins, GitLab CI, CircleCI, Ansible
- **Data Governance & Quality:** Data Lineage, Metadata Management (Apache Atlas), Data Privacy, Data Validation, Profiling
- **Version Control & Collaboration Tools:** Git, GitHub, Bitbucket, Jira, Confluence, ServiceNow, Databricks
- **Data Visualization & Reporting:** Matplotlib, Seaborn, Plotly, Tableau, Power BI, ggplot2, Jupyter Notebooks
- **Monitoring & Logging:** Prometheus, Grafana

WORK EXPERIENCE:

Hartford Financial Service Group **Feb 2024 - Current**

Data Scientist/ Data Engineer

- Architect and deploy comprehensive machine learning workflows using Python, TensorFlow, and Keras on Azure Machine Learning (AML) to model customer churn, achieving a 15% improvement in retention rates.
- Engineered ETL/ELT pipelines using DataStage to extract, transform, and load data from sources like AWS S3, Oracle, and SQL Server into Snowflake, applying business rules for data cleansing and transformation.
- Conduct in-depth data analysis and exploratory visualization using Pandas, NumPy, and Seaborn on Azure Databricks, uncovering actionable insights that enhance campaign personalization.
- Develop and operationalize deep learning architectures (RNNs, LSTMs) for NLP applications, such as sentiment analysis and automated support chat systems, using Azure ML and Azure Cognitive Services.
- Utilize Azure Databricks and Azure Synapse Analytics to efficiently handle and analyze massive datasets, enhancing processing throughput and boosting model precision.
- Direct the transformation of legacy ETL infrastructures into modern, cloud-native data ecosystems leveraging Azure Data Factory and Azure Synapse, improving scalability and uptime.

- Design and optimize data ingestion and streaming frameworks with Azure Stream Analytics and Azure Databricks, leading to a 30% reduction in processing latency.
- Build and manage machine learning deployments in Azure Machine Learning, ensuring smooth integration and reliability in live production environments through Azure Kubernetes Service (AKS).
- Automate ETL workflows using Azure Data Factory and Apache Airflow on Azure, streamlining complex data pipelines for scalable machine learning systems.
- Spearhead the creation of a real-time recommendation engine using collaborative filtering and deep learning models on Azure ML, driving higher user engagement metrics.
- Collaborate with diverse stakeholders to align business needs with tailored data science methodologies and robust machine learning frameworks using Azure ML.
- Leverage version control tools (Git, GitHub) to ensure efficient collaboration, traceability, and deployment of machine learning solutions on Azure DevOps.

Coforges **Jan 2020 - Dec 2022**

Data Engineer

- Designed and optimized high-performance data pipelines using Apache Kafka and Google Cloud Dataflow, enhancing data ingestion and processing workflows.
- Administered and optimized enterprise data warehouses using Google BigQuery to support robust business intelligence initiatives.
- Crafted and fine-tuned complex SQL queries for high-efficiency data retrieval and analysis across relational databases like MySQL and PostgreSQL.
- Orchestrated the integration of real-time streaming data with Apache Kafka and Google Cloud Pub/Sub, enabling instant processing and actionable insights.
- Enhanced data accessibility and traceability by deploying metadata frameworks with tools like Apache Atlas and Alation.
- Automated data quality assurance through validation, profiling, and anomaly detection workflows, ensuring compliance with GDPR and HIPAA.
- Directed the development of unified data lake architectures on Hadoop, consolidating data from diverse sources into a cohesive structure.
- Built and maintained RESTful APIs to seamlessly integrate external data sources, improving interoperability within the data ecosystem.
- Partnered with data scientists to deliver well-structured, clean datasets, boosting the reliability and performance of predictive models.
- Employed Git-based version control to manage and monitor changes in complex data engineering workflows, fostering team collaboration.
- Configured advanced networking protocols and managed distributed systems, ensuring seamless and rapid data transfers.
- Developed interactive dashboards and analytics reports with Tableau and Power BI, providing senior management with critical business insights.

EDUCATION:

- **Master's In Information Systems & Technology**
  University of North Texas

- **Bachelor's in Mechanical Engineering**
  Sreenidhi Institute of Science & Technology

CERTIFICATION:

- Azure Data Engineer Associate DP 203: Microsoft
- Azure Data Fundamentals DP 900: Microsoft
- Azure Fundamentals AZ 900: Microsoft
- Cloud Digital Leader: Google
- Machine Learning: Verzeo
- Python Programming: Hackerrank
- SQL Programming: Hackerrank