Sandhya Raghavan
www.linkedin.com/in/raghavansandhya
https://github.com/raghavansandhya/
Email:raghavansandhya@gmail.com
Mobile: +91 9008732338
Location: Bengaluru, Karnataka

A self-driven and enthusiastic engineering professional with 15+ years of experience designing and developing solutions for complex business problems involving large scale data warehousing, real-time analytics and reporting solutions. Managing large and small teams providing mentorship and technical leadership. Have experience recommending the right AI and Big Data tools for the business need and creating a stable architecture that helps organizations effectively analyze and process large structured and unstructured data. Have hands-on experience in performing analytics and data visualizations, applying ML/Statistical algorithms to real-world problems, and building end-to-end machine learning platforms.

TECHNICAL SKILLS

Coding Languages	Python, R, UNIX Shell Scripting, SQL, Scala, Go(learning)
Big Data Tools	Spark, Kafka, NIFI, Hadoop, Hive, Spark-Streaming, Yarn, Presto
Database	Oracle, MySQL, PostgreSQL, SQLite, AWS Redshift, Cassandra, Teradata, MongoDB, HBASE
Viz Tools	Tableau, DASH, Plotly, Streamlit, D3, Seaborn, ggplot
Data Platforms	AWS, Databricks, Hortonworks, Cloudera
Version Control	GIT, Ab Initio EME, Clear case, Tortoise SVN
DevOps Tools	Airflow**,** Jenkins, Ab Initio–CC, Docker, Autosys, Control-M, Kubernetes(learning)
ML Libraries	Koalas**,** Numpy, Pandas, scikit-learn, Tensor flow, Keras, GeoPanda, Horovod, MLFlow
ETL Tools	Ab Initio 3.3.2.18, Express >IT, Conduct >IT
Other Tools	JIRA, Confluence, ActiveMQ, MQTT, QGIS, Collibra, MHUB, HP Service Manager

EXPERIENCE

Virgin Hyperloop (previously Virgin Hyperloop One) **Apr 2019 - Current**
Role: Data Engineering Manager/ Specialist
Location: Los Angeles (CA, USA)
Working for the Machine Intelligence and Analytics Team which is responsible for creating automated solutions including Demand Prediction, Real time Trip Management, Command and Control Software, Optimized Alignment Generation and Hyperloop Simulations. Wearing many hats as Software Architect, Team Lead, Data Engineer, ML Engineer. Video showcasing our work [link].
Job Duties:

- Designed, built, and maintained a cloud-based AI simulation execution and analysis infrastructure: NIFI Data Pipeline, python(scripts), PostgreSQL, EC2, S3, ActiveMQ, MLflow, Spark, Athena, Delta and Tableau. Purpose: mass scaled simulations & batch analysis for generative transport designs. Result: 50x faster time-to-insight for executives' reports - from 6months to 3days
- Leading a team of 3 engineers. Guided the team to use the state-of-the-art tools like Spark/koalas which sped-up python analytics code 150x, reducing processing time from 9 hours to 7 minutes
- Created a GIS data pipeline in Spark with Databricks using packages like Geopandas, GeoTrellis, Rasterframes and Geospark which would take millions of GIS vector data and convert into raster grids for generating optimized alignments
- Designed the architecture for Data Governance, created data dictionary and business catalogs, collected the provenance and created flow-level lineage representations of the processes

Consultant from ITIDATA **Nov 2011 – Apr 2019**
Role: Data Architect/Lead Data Engineer
Client: Citi Group, Wells Fargo
Location: Warren (NJ, USA), Charlotte (NC, USA)
Created several data extraction, warehousing and reporting frameworks with logging and support automation in Ab Initio with reusable templates that reduced development efforts and cost. Built highly scalable,

distributed systems using different open-source tools as well as designing and optimizing large, multi-terabyte data warehouses. Integrated new tools and developed technology frameworks/prototypes to accelerate the data integration process and empower the deployment of predictive analytics. Exposed the data quality metrics and lineage on Critical Data Elements as part of Data governance and Basel requirements.

Job Duties:

- Leading a team of 8 member both onsite and offshore in an Agile environment, holding daily standup calls and managing daily tasks
- Integrated Hadoop into traditional ETL, accelerating the extraction, transformation, and loading of massive structured and unstructured data
- Designed and developed ETL framework to get the data lineage metadata from the report applications and loaded them to MHUB for visualization
- Performed data analysis using R and Python to understand the source data and underlying business rules written in Talend. Designed and developed a migration framework based on the analysis to load the legacy data to the new KYC portal on aggressive timelines
- Created python scripts for rule parsing and extracting the rule metadata for performing ETL operations in Ab Initio
- Developed frameworks for data ingestion and extraction from and to several HDFS file formats like Parquet, AVRO and expose them as Hive tables for performing SQL analysis
- Installed and configured Apache Hadoop, Hive and Pig environment on the prototype server
- Created and enhanced templates for the generic ETL components based on the user requirements and provide application and environment support from installation to production in the user application
- Design, develop, enhance and support feeds to transform and load data from multiple vendors like S&P, Reuters, Bloomberg based on the business requirements provided

Consultant from WIPRO **Sep 2006 – Nov 2011**

Clients: Capital One Finance and Capital One Auto Finance

Role: Tech Lead/Senior Data Engineer

Location: Richmond (VA, USA), Dallas (TX, USA), Chennai (TN, India)

Created ETL process in Ab Initio for several data and process migration projects. Worked on different lines of business from Cards, Auto Finance to Mortgages. Helped modelling the ODS and ADS for Mortgages and Cards. Created ETL pipeline with data streaming process to load the data warehouse for Digital Analytics initiative.

Full list of my employment (>5) on linked.in/sraghavan.

For my personal projects please check out my GitHub.

EDUCATION

Anna University, Chennai, India

Bachelor of Engineering in Computer Science 2002 – 2006

COURSEWORK

Statistics with R Specialization - Coursera 2017 – 2018

Deep Learning Specialization – Coursera 2017 – 2018

Machine Learning Nanodegree – Udacity 2016 – 2017

PATENTS

Patent US 63/107249: Optimization, Infrastructure Digitization and Simulation System and Method – J. Wei, S. Raghavan, P. Oleniuk, Published in US Patent Office, October 2020 - This patent protects the simulation orchestration software I developed for Virgin Hyperloop. Published in US Patent Office 2020

CONFERENCES/ARTICLES

Spark + AI Summit 2020 - Generative Hyperloop Design: Managing Massively Scaled Simulations Focused on Demand Modelling

Tech Post - Koalas: A field guide on seamlessly switching your pandas code to Apache Spark