AVINASH RAMESH

San Jose, CA | (408) 338-5862 | avinash.ramesh@sjsu.edu | www.linkedin.com/in/avinash94/ | Github: github.com/rameshavinash94/

SUMMARY

A Software Engineer & problem solver with hands-on experience in the big data stack for 5 years, across functions such as data ingestion, business understanding, development of data pipelines, and automation of workflows.

EDUCATION

San José State University *San Jose, CA, USA*
Master of Science in Computer Software Engineering Specializing in Data science ***August 2021 - April 2023***
Courses: *Enterprise Software Systems, Data Mining, Advanced Data Mining*

Anna University, Meenakshi Sundararajan Engineering College *Chennai, TN, INDIA*
Bachelor of Technology in Information Technology ***June 2012 - April 2016***
Courses: *Data Structures and Algorithms, Web Technologies, Object-Oriented Programming, Database Management Systems, Computer Networks*

EXPERIENCE

ITIDATA, *Chennai, India*
Analyst ***August 2016 - June 2019***
Associate Data Engineer ***July 2019 - July 2021***
Client: Citibank (Projects: Security Master Central, Price Master Central)

Roles/Responsibilities:

- Built infrastructure to process internal & external vendor data feeds for global use within Citigroup and apply business rules & data selection hierarchy on individual data attribute levels to create a golden data source.
- Deployed Sqoop scripts to import, export, and update data between HDFS, Hive, and relational databases.
- Created multiple Hive tables with partitioning and bucketing for efficient data access.
- Implemented pipelines to extract data from various vendors, wrangle data, apply business transformation rules, and load into desired formats.
- Migrated Ab Initio graphs/plans to Spark distributed framework (40% faster), based on business needs.
- Automated ETL pipelines using Autosys scheduler saving manual workloads.
- Applied various data sources and formats, including structured, semi-structured, and complex file formats.
- Identified, designed, and implemented internal process improvements: automating manual processes, optimizing data delivery, re-designing infrastructure for greater scalability, etc.
- Collaborated with business analysts/stakeholders to resolve data-related technical issues.
- Setup audit triggers in Oracle Database tables to track DML changes.
- Coordinated with team and Developed framework to generate daily Adhoc reports for ETL feeds.

SKILLS

Programming languages: *Python, Unix Shell Scripting, SQL, Java*
ETL tools: *Ab Initio, Google OpenRefine*
Big Data Frameworks/Tools: *Spark, Hive, Hadoop, HDFS, YARN*
Data Formats: *CSV, JSON, Parquet, Avro, RDBMS tables*
Scheduler: *Autosys, Crontab*
Visualization: *Tableau, Microsoft Excel, Streamlit*
CI/CD tools: *IBM UrbanCode Deploy, AWS code pipeline, GCP Build, Jenkins*
Version Control: *Github, Bitbucket*
Web Frameworks: *HTML, CSS, Bootstrap, Flask*
Container Technology & Management: *Docker, Kubernetes*
Cloud: Amazon Web Services (AWS) – *EC2, S3, EKS, ELB, IAM, LAMBDA. GCP - Cloud Run, Cloud Build, GKS, Vertex AI*
Databases: *Oracle 11g & 12c, SQLite, HBase*
ML Concepts: *Classification, Regression, Clustering, Dimensional Reduction, NLP, Ensemble Techniques*
Data Manipulation/ ML Libraries: *NumPy, Pandas, Matplotlib, Pyspark, Spacy, Scikit-learn, Keras, Pycaret*

PROJECTS

Status : *Completed* ***September 2021 - December 2021***

- **Wikipedia Based Question & Answering (QA) Application**:
  Repo: *https://github.com/rameshavinash94/Wiki_QA_System*
  Deployment URL: *https://cmpe256-q4uake3apq-uc.a.run.app*
  Techniques: *Information Retrieval, Cosine Similarity, Word/Sentence Embedding, QA systems, BERT*
  Libraries: *Python - Spacy, Transformers, Wikipedia-API, Streamlit, Pandas, Numpy*

- **Web App for Detection & Classification Of ECG Images**:
  Repo: *https://github.com/rameshavinash94/Cardiovascular-Detection-using-ECG-images*
  Deployment URL: *https://cmpe255-project-q4uake3apq-uc.a.run.app*
  Techniques: *RGB2Gray, Resize, Gaussian filter, De-noising, Thresholding, Contour, GridSearchCv, Ensemble*
  Libraries: *Python - Scikit-Learn, Scikit-Image, Matplotlib, Pandas, Numpy, Joblib, Streamlit*

- **Patient Management System (Book/Cancel Appointments with Doctors):**
  Repo: *https://github.com/rameshavinash94/CMPE272_PMS*
  Libraries: *Python - Flask, Flask-oidc, okta, Dialogflow,sqlite Web - HTML, CSS, JS, Bootstrap*