

Avinash Ramesh

San Jose, CA | (408) 338-5862 | [avinashramesh2312@gmail.com](mailto:avinashramesh2312@gmail.com) | [www.linkedin.com/in/avinash94/](https://www.linkedin.com/in/avinash94/) | Github: [github.com/rameshavinash94/](https://github.com/rameshavinash94/)

SUMMARY

I am a skilled software professional with nearly 6 years of experience in the backend engineering, specializing in data-driven design, OOAD patterns (*strategy, singleton, factory, builder, chain of responsibility, observer, dependency injection,*), REST APIs, relational databases, data pipeline design, warehouse modeling(*Dimensional/ Datavault*), and workflow automation.

EDUCATION

**San José State University** *San Jose, CA, USA*  
**Master of Science in Computer Software Engineering (Specializing in Data science)** *August 2021 - May 2023*  
**Courses:** *Enterprise Software Systems, Data Mining, Advanced Data Mining, Machine Learning, Deep Learning, Software Systems Engineering, Special topics in ML and Reinforcement Learning. GPA:3.9/4*

**Anna University** *Chennai, TN, INDIA*  
**Bachelor of Technology in Information Technology** *June 2012 - April 2016*

EXPERIENCE

**Carl Zeiss Meditec**, *Dublin, CA* *May 2022 - December 2022*  
**Software Engineer Internship - Building fullstack big data product for clinical analysis**

- Spearheaded the entire product development lifecycle for **three** key functionalities of CRUD-based DICOM application: uploading series of scans, annotating studies, comprehensive metadata search and exporting dicom metadata & scans.
- Designed frontend using React, built event-driven data pipelines using Azure services (*ADF, Azure Batch, and Azure Databricks*), developed scalable REST APIs (*Fastapi, ORM-SQLAlchemy*) handling nearly **hundreds of concurrent users**, performed API pagination, deployed database models (*PostgreSQL*), and created an internal cache store.
- Follow object oriented design patterns(*strategy, singleton, factory*) & dependency ingestion techniques to develop loosely coupled applications.
- Collaborated with managers and downstream users to plan, design and develop an internal data product to fasten clinical research and analytics for AI teams.

**ITI DATA**, *Chennai, India*  
**Associate - Software Engineering / Data Engineering (Big Data Platforms)** *July 2019 - July 2021*

- Planned 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.
- Write automation scripts using python/unix shell script for data validation and processing tasks to reduce man hours.
- Created Hive tables with partitioning and bucketing for efficient data access; deployed spark pipelines; optimized existing SQL queries by **30%**; and wrote Sqoop jobs to import, export, and update data between HDFS, Hive, and relational databases.
- Migrated existing data pipelines(*Abinitio*) to Spark distributed framework (POC & productionize pipelines) nearly **40%** faster thereby reducing licensing costs spent on ETL tools.
- Automated ETL pipelines across **millions of rows** using Autosys scheduler reducing manual workloads.
- Identified, reviewed, and implemented internal process improvements: automating manual processes, optimizing data delivery by **50%**, re-designing infrastructure for greater scalability, etc.

**Analyst - Software Engineering / Data Engineering** *August 2016 - June 2019*

- **Designed data marts** and written **REST API** endpoints for handling various data extraction and querying tasks for BI engineering to interact with databases to design dashboards, and perform real-time data analysis.
- Architected ETL pipelines to extract more than **2 million** records on a weekly basis from various vendors(Reuters, Bloomberg, etc.), wrangle ingested data, apply business transformation rules, and load into desired formats.
- Created a framework to recognize **data anomalies** and produce daily automated reports to reduce turnaround time to nearly **half** for addressing data problems.
- Designed db models and applied **Slowly Changing Dimensions(SCD)** in Data Warehouse to track historic changes.
- Remodeled functional logic of python similarity script using Numba jit, numpy arrays and map operations resulting in decrease in time complexity by nearly **3/4th**.

SKILLS

**Programming languages:** Python, Unix Shell Scripting, SQL, Java, Javascript  
**Big Data Frameworks/Tools:** Spark, Hive, Hadoop, HDFS, YARN, Sqoop  
**Web Frameworks/API:** HTML, CSS, React, Flask, Fastapi  
**Data Formats:** CSV, JSON, Parquet, Avro, DICOM  
**ML Libraries:** Numpy, Pandas, Matplotlib, Pyspark, Spacy, Scikit-learn, Keras, Tensorflow, Pytorch  
**Orchestration & ETL:** Autosys, Crontab, Airflow, Ab Initio, OpenRefine, ADF  
**Databases:** Oracle 11g & 12c, SQLite, HBase, Postgresql, Redis  
**Visualization:** Tableau, Microsoft Excel, Streamlit  
**CI/CD tools:** IBM UrbanCode Deploy, AWS code pipeline, Azure Devops  
**Container Technology & Management, Configuration:** Docker, Kubernetes, Ansible