SRIDHAR RAMASAMY

Greater Seattle Area | rsridhar086@gmail.com | +1-608-695-6561

OBJECTIVE

- Software Engineer with 8 years of experience in data engineering, API design/REST and graph systems.
- Specialties: Large-scale streaming, event-driven, highly available data systems, API design, REST services.
- Looking for a role to be a proactive team member, lead initiatives, take responsibilities & be accountable.

TECHNICAL SKILLS

Languages: Java, Scala, Python, Bash
Frameworks: Spring-Boot, Liquibase, AWS/OCI, Hadoop, Apache Spark, Luigi/Airflow
Database: Cassandra, Hive SQL, Postgres, Oracle DB
Dev. Tools: Kubernetes, Grafana, Argo, Gitlab-CI/CD, Docker

WORK EXPERIENCE

Sr. Software Engineer **Oracle Inc. Seattle, USA** **Sep '18 - Current**

- **Project: Advertising Data Pipelines** (March 2023 - Current)**:**
  - Handle voluminous Kafka stream data with variety of records and make it available for the org. to consume.
  - The data pipelines receive 1000's of records every minute and are processed in OCI spark dataflow.
  - The data is made available in both AWS and OCI, while processing happens only in OCI.
  - Migrating our workflow from AWS to OCI, setting up new orchestration of pipeline using airflow/argo.
- **Project: Data API/Data Collaboration Platform** (July 2021 – March 2023)**:**
  - Designed model to facilitate data storage, access, and enforce data lease contracts.
  - Data API defines schema, transformation and joins across schema, source and destination location.
  - Designed API for data model, entities, and its relationships & persistence.
  - Usage of IAM roles/policy to allow limited access to a leased data location and enforce contracts and TTL.
  - Designed and built **security model** for the Data Lease Service with RSA based auth and auto JWT generation.
  - Integrated with other systems to effectively manage source data and lease data according to contracts.
- **Project: User Consent Management System** (July 2020 – July 2021)**:**
  - Opt-out data auditor will be capable of keeping a register of dataset, export opt-outs and trigger alerts to dataset owners and generate periodic reports and keep counts.
  - Load testing the service using Apache JMeter for an endpoint that needs to capture few million records.
  - Productionizing **Apache Pulsar** cluster with TLS certs and setting it up on OCI Kubernetes.
- **Project: Dynamic Data Packaging and Delivery Project** (Sep 2018 – July 2020)**:**
  - Designed and implemented a new data packaging & delivery project to replace an existing batch processing.
  - The project saves about $2M/year in computing cost, brought down compute time from 96 hours to 14 hours.
  - The different supporting pieces are - Data selection, Data Matching, Data Transformation and Packaging, Data Encryption, Data Compression and Data Delivery.
  - Automatically deliver reports and notify the stake holders, clients, and dev team about the deliveries.
  - Interacted with the Product team gathering requirement for the system & drafting functional specification.
  - Train the operations team with the new system and oversee on-time resolutions of support tickets.
- Tech Stack: **Java/Spring Boot/Kubernetes**. Data Processing: **Apache Spark/Qubole/AWS**.
- Job Orchestration: **Luigi/Apache Airflow**.
- Migration of all our CI/CD pipelines from **Jenkins** to **Gitlab.** Replicate all the functionalities of existing pipeline.
- Perform unit testing and integration testing to seamlessly integrate with other existing systems.
- Coordinate with other teams/pods to enhance design/development and support cross-team projects.

Software Developer **Red Hen Systems, Fort Collins, CO** **Nov '16 – Sep '18**

- Led a Field Report creation project for Anadarko on C#/.Net to view and publish survey data.
- The paper-based forms were converted to digital application based on a .Net Windows platform capable of data entry, sketching with stylus or touch pad and attaching photos.
- The tool is capable of formatting the data in the pdf, resize images/customer sketch, publish data to a database and send an email with the pdf as an attachment.

- Responsible for development of IsWhere, Media Geo-Tagger softwares - **C#/.Net/WPF/Forms**.
- Introduced online licensing & management system for all our software products reducing manual effort.
- Improved the software by adding tools to clip videos, modified UI, provided multiple file compatibility.
- Batch automation process for ETL workflow – Data extraction, processing, analytics, and delivery.
- Worked on GIS projects for various oil & natural gas companies, universities, and government agencies.

Grad. Research Asst. **Computer Networking Research Lab, Fort Collins, CO** **Feb '15 – May '17**

Reconstruction of the Topology of Undirected Graphs from Partial Information

- Used extended Robust PCA, an accurate method for matrix completion by low rank approximation.
- Applied on 2D & 3D WSN's and real-world internet & social networks with 500 - 5000 nodes.
- Facebook, E-mail, Collaboration, CAIDA etc. are some of the real world online/social networks.
- Reconstructed the topology of network/graphs with less than 20% information achieving a mean error less than 6% for hop distances and 0.5 hop average error.

EDUCATION

Master of Science Electrical & Computer Engineering	**Colorado State University, USA**	GPA 3.5/4.0	May '17
Bachelor of Engineering Electrical & Electronics Engineering	**Anna University, India**	GPA 9.02/10	Apr '13

PERSONAL/ACADEMIC PROJECTS

Building a Distributed, Replicated, and Fault Tolerant File System - Java

- Implemented a Distributed File System in Java. It consists of three components – Controller Node, Chunk Server, and Client Node.
- The files are stored with in chunks of 64KB and replication level of 3. The chunk severs also stores message digest (SHA-1) to check for integrity and if it fails it contacts other chunk servers to fix errors.

Scalable Server design using Thread pools and load balancing on the server - Java

- Designed and implemented Thread pools to handle load on the server without using 3rd libraries.
- Observed that mean per client throughput X No of active connections equals server throughput. Std. Deviation of per client throughput is low.

Content Searching in a Distributed Application Layer Network – Structured & Unstructured - Java

- Implemented Unstructured & Structured (Chord Algorithm) P2P network on an 80-node test bed.

PUBLICATIONS

- Capture and Reconstruction of the Topology of Undirected Graphs from Partial Coordinates: A Matrix Completion Based Approach, **Thesis/Dissertation – Colorado State University 2017**.
- Network Topology Mapping from Graph Geodesics and Partial Virtual Coordinates, **IEEE/ACM Transactions on Networking 2019**.
- Topology Maps and Distance Free Localization from Partial Virtual Coordinates for IoT, **IEEE ICC 2016**.