

ETHIRAJ SRINIVASAN

CONTACT

- ethirajsrinivasan@gmail.com
- +91 9841153360
- ethigEEK.com
github.com/ethirajsrinivasan

JOB OBJECTIVE

Aspire to lead data-driven innovation by architecting scalable platforms, fostering engineering excellence, and aligning technology strategy with organizational growth and long-term vision.

CORE COMPETENCIES

- Strategic Data Platform Planning
- Big Data Analytics & Processing
- Data Pipeline & ETL Architecture
- Real-Time & Batch Processing
- Data Quality & Governance
- Cloud Data Solutions
- Performance Optimization & Resource Management
- Analytics & Reporting
- Team Leadership & Mentoring
- Cross-Functional Collaboration
- Ingestion Application Development
- Technical Innovation & Strategy

AWARDS

- Shopee, Singapore - Value Star Award
- DNA Capitals, Singapore - Winner - Hackathon
- Tata Consultancy Services - Star of the Learners group, On the Spot Award

EDUCATION

- 2018: National University of Singapore, Singapore - M.Tech Software Engineering, CGPA - 4.36/5
- 2013: Anna University, India - B.E Electronics and Communication, CGPA - 8.56/10

PROFILE SUMMARY

- Strategic Data Engineering Leader with a decade of experience driving enterprise-scale data platforms, architecture, and analytics modernization across diverse industries. Possesses deep expertise in data pipeline design, distributed processing, ETL, and real-time and batch frameworks, with a proven ability to deliver scalable, high-performance, and reliable data ecosystems.
- Leads and mentors cross-functional teams in building and managing data ingestion, ETL, and analytics platforms, ensuring strong data governance, data quality, observability, and operational excellence. Skilled in designing and maintaining data lakes, data warehouses, ML-ready datasets, and real-time analytics pipelines to support enterprise decision-making.
- Expert in implementing data quality frameworks, monitoring systems, and cost-optimized solutions to improve reliability, reduce latency, and enhance efficiency. Proficient in Spark, Hadoop, Kafka, Hudi, Hive, HBase, AWS (EMR, S3, Athena), Python, Spring Boot, Redis, and other modern data technologies. Recognized for combining technical depth with strategic vision to drive data-driven innovation and organizational growth.

WORK EXPERIENCE

Aug'25 – Present | Zgrow Solutions, Chennai | Head of Data Engineering

Stack - Spark, S3, Athena, EMR

Key Result Areas:

- Led Data Engineering, mentoring a high-performing team and implementing governance, best practices, and process optimizations to deliver scalable, reliable, and high-quality data platforms.
- Architected an end-to-end Spark-based ML pipeline on AWS EMR for logistics anomaly and weight prediction, integrating a distributed ML model and automated feature engineering, achieving 99.9% uptime.

Jun'20 – Feb'25 | Shopee, Singapore | Expert Data Engineer

Stack - Spark, Hadoop, Hudi, Hive, Kafka, HBase, Python, Redis, Docker, Spring Boot, Ruby on Rails

Key Result Areas:

Batch Ingestion Team Leadership:

- Led the batch ingestion team managing 23k MySQL-to-Hive pipelines, 6k Hive-to-Hive pipelines, and 2k Hive-to-HBase and HBase-to-Hive pipelines with daily and hourly frequencies.
- Ensured seamless operations on a 14PB data lake, achieving an outstanding 99.95% job success rate for 31k ingestion pipelines.

Performance Optimization & Resource Utilization:

- Maintained 80% memory utilization, processing 1,059,221 GB/hour and 245,414 Vcore/hour, with 97% of jobs completing within 30 minutes.
- Optimized cross-IDC Hive-to-Hive pipelines, saving approximately 94,325 GB/hour, 13,487 Vcore/hour, and 500 TB of storage.

Platform Upgrade & Cost Optimization:

- Upgraded the ingestion platform to Spark 3, resulting in a 9% reduction in processing time and a 14% reduction in resource usage.

Pipeline Development & Monitoring:

- Designed and monitored batch, real-time, and cross-country ingestion pipelines, including hourly Hive-to-Hive jobs with filters, incremental dumps, and Data Quality Checks (DQC).

TALKS & ENGAGEMENTS

- Jury - HackElite' 25 - SRM Institute of Science and Technology
- Guest of honor - International Conference on Computer Communication & Power Systems - ICCCPS 2020
- Guest of honor - Outcome Based Industry 4.0 at Sri Venkateswaraa College of Technology
- Resource Person - Faculty Development Program On Artificial Intelligence For Robotics (AICTE Training and Learning (ATAL) Academy) At Sai Ram Engineering College

OPEN SOURCE CONTRIBUTIONS

- Ruby Gems
 - o Collaborator - Sunspot (Apache Solr) gem
 - o Owner - Collection2csv, devise-foundation-views, rails-fort, rails-protip

PERSONAL DETAILS

- DOB: 3rd Oct 1991
- Languages: English, Telugu & Tamil
- Address: Chennai

Data Pipeline & Metrics Management:

- Designed metrics for job performance and resource utilization, and managed ingestion pipelines handling diverse sources (**MySQL, CSV, Kafka**) and sinks (**HDFS, Hive, Hudi, Kafka**) using **Spark, DeltaStreamer, Spring Boot, and Maxwell**.

Ingestion Applications Development:

- Built and maintained key applications: **Core Engine** (job automation), **AutoAdaptation Framework** (dynamic optimization), **Operations Platform, Data Quality Module, Configuration Service**, and **Notification Service**.

Feb'19 – May'20 | Lomotif, Singapore | Data Engineer

Stack - Spark, Kinesis, Lambda, EMR, Athena, Quicksight, S3, Python, Redis, Elasticsearch, Postgres, Flask

Key Result Areas:

- **Data Infrastructure Development:** Led the design of a scalable **data infrastructure platform**, including a **data lake** using **S3, Kinesis, Lambda, and EMR (Spark)**, and implemented an **automated recommendation engine** to enhance feeds and follower engagement.
- **Big Data Processing & Analytics:** Built and monitored **PySpark jobs on EMR** for diverse business use cases, developed advanced **metrics for ranking videos**, and conducted **graph analysis** to identify communities and influencers.
- **Real-Time Visualization:** Created interactive **QuickSight dashboards** with **Athena** as the query engine, providing actionable insights for stakeholders and enabling data-driven decisions.

May'18 – Jan'19 | Knorex, Singapore - Big Data and Analytics Engineer

Stack - Spark, Kinesis, Lambda, S3, Scala, Vertx

Key Result Areas:

- Migrated an ad events tracker to **Vert.x 3**, achieving **50K QPS** with optimized throughput using **Kinesis and Lambda architecture**; built a **data lake** on **S3** integrated with real-time sources, developed **PySpark jobs on EMR** for analytics.
- Created a **reporting tool** with **Spark (Scala)** for attributions, and designed **QuickSight dashboards** using **Athena** to deliver actionable insights through visualization and advanced metrics.

Aug'15 – Dec'16 | Pramati Technologies, India | Senior Software Engineer

Stack - Ruby on Rails, Memcached, Redis, Solr, Postgres

Key Result Areas:

- Developed a web application for managing patents with a robust role-based content management system for content manipulation, authorization, and access; additionally, organized and facilitated Open Source Meetups to foster community engagement and collaboration.

Jan'14 – Jul'15 | Tata Consultancy Services, India | Assistant System Engineer

Stack - Ruby on Rails, Memcached, Redis, Solr, Postgres

Key Result Areas:

- Headed a team of two members to develop a collaborative platform for sharing content, ideas, and processes across the Tata group of companies