

# SHAMEERA RATHNAYAKA YODAGE

## Senior Data Infrastructure Engineer

24 Hoitsma Ct, Fair Lawn, NJ 07410 | (812) 558 4824 | [shameerayodage@gmail.com](mailto:shameerayodage@gmail.com)

LinkedIn: <https://www.linkedin.com/in/shameerarathnayaka/> | Github: <https://github.com/syodage>

### PROFILE

Senior Engineer with excellent big data processing skills with different infrastructure including Google Dataflow, Google Dataproc, Apache Hadoop, Apache Kafka etc. A senior member of Spotify's elite data infrastructure team who designs, implements, and manages the whole Spotify data processing, storing, monitoring and scheduling echo system for both real-time and batch processing. An expert Java and Scala developer and open source enthusiast.

### SKILLS

Java, Scala, Functional Programming, Apache Beam, Google Dataflow, BigQuery, GCS, Hadoop, Map Reduce, HDFS, Streaming, Batch, Pipeline, ETL, Workflows, Spark, Flink, Kafka, BigTable, Datastore, Cloud SQL, Spanner, Cloud Pub/Sub, Docker, Kubernetes, Ansible, Google Cloud Platform (GCP), Apache Avro, Parquet, Protobuf, Python, JavaScript, RabbitMQ, BigData, Cloud, Vue Js, React, AWS (S3, Kinesis), MySQL, API, Data retention, Access Control, Agile, Scrum

### EXPERIENCE

#### **Spotify USA Inc, New York City** - *Senior Data Infrastructure Engineer*

Sep 2019 - Present

- Implemented Spotify's new A/B testing platform's data processing, scheduling, monitoring and validation features and led the rest of the team in development efforts.
- A core developer of the Spotify Scio and its echo system, which is used by data engineers to write data pipelines and run them on Google Dataflow.
- Improved the runtime performance of data pipelines by reducing data shuffling before SMB (Sort, Merge, Bucket) Joins.
- Improved the user experience of PII data access in the data pipelines by implementing a scala library which auto decrypts encrypted personal data.
- Designed and built A/B testing data pipelines(ETL) to produce high-quality statistics which are used to make business decisions.
- Develop ETL pipelines to collect user emitted events and produce high-quality data sets.
- Integrated optimized distributed distinct count algorithm HyperLogLog++ support with sketches to Spotify Scio; a scala wrapper to Apache Beam and exposed it as an API to provide accurate distinct count to end users.
- Implemented a real-time notification system for dataset anomaly detection.
- Conducted 'Data Processing Best Practices with Internal Tools', an advanced data engineering course to train other engineers in the company.

- Implemented Jupiter notebooks to validate the input data quality of A/B testing user events.

### **Spotify USA Inc, New York City - Data Infrastructure Engineer II**

Jan 2017 - Aug 2019

- Developed Spotify data retention platform supporting all common data storage formats, Apache Avro, Apache Parquet, Google Protobuf, BigQuery etc.
- Developed Spotify's GDPR compliant access control platform to control access to all the datasets (PII and non-PII), both opt-in based and inheritance-based from streaming to batch level.
- Implemented real-time data schema service for BigQuery, Apache Avro, Google Protobuf.
- Migrated in-house Hadoop jobs to Google Cloud Platform by forklifted to run on data proc and re-wrote in Apache Beam(Scio) to run with Google Dataflow.
- Developed a metadata service to collect all the metadata of data pipelines, a.k.a workflows to build the dataset lineage.
- Have written the 'Access Control and Retention Golden Path' documents which were widely adopted by engineers in the company to learn the echo system.
- Integrated Apache Beam JdbcIO support with Spotify Scio.

### **Indiana University - SciGap Group, Bloomington, IN - Graduate Research Assistant**

Aug 2014 - Dec 2016

- Designed and Implemented component-based microservice architecture to execute and manage computational jobs and workflows on distributed, grid, and cloud compute resources.
- Deployed and managed Apache Mesos and Aurora to run Scientific jobs.
- Profile, cluster, and standalone streaming performances of Apache Kafka and RabbitMQ on IoT domain.

### **WSO2 Inc, Sri Lanka - Distributed Software Engineer**

Apr 2013 - Aug 2014

- Designed and Implemented a lightweight JAXRS engine for WSO2 admin services.
- Automated user-friendly patch tool for WSO2 core module.
- Case-study to implement OSGi level multi-tenancy with bundle resolve framework hooks.

## **EDUCATION**

### **Indiana University, Bloomington - M.Sc in Computer Science**

Aug 2014 - Dec 2016 | GPA: 3.9

### **University of Moratuwa, Sri Lanka - B.Sc in Computer Science**

Jun 2008 - Apr 2013 | GPA: 3.5

## **AWARDS**

Speaker, Scale by the Bay 2020 Tech Conference - 'Type class materialization with Magnolia'.

Apache PMC member ( Axis2, Airavata) and Committer.

Google Summer of Code Student (2007, 2008) and Mentor (2010).