

KARTHIK RAMAKRISHNAN

Senior Data Engineer

karthik.ramakrishnan0522@gmail.com | +1 (469) 951-8854 | linkedin.com/in/karthikrk1 | github.com/karthikrk1

PROFESSIONAL SUMMARY

Senior Data Engineer with 10+ years of experience designing and operating large-scale data platforms across retail, financial services and workplace analytics. Builds petabyte-scale streaming and batch pipelines on GCP and AWS (Kafka, BigQuery, Airflow, dbt, Spark, Glue), with a focus on data modelling, data quality and query performance. Has led engineering teams and delivered production systems spanning ETL, search, recommendations and text analytics. M.S. in Computer Science.

CORE SKILLS

Languages: Python, SQL, Java, Scala, R, Unix Shell

Data Engineering: Apache Kafka, Apache Airflow, dbt, Apache Spark, Hadoop, NiFi, Storm, ETL/ELT, data modelling, data warehousing, data quality

Cloud: GCP (BigQuery, Cloud Functions, App Engine, Cloud SQL); AWS (Glue, Lambda, SageMaker, Comprehend, Translate)

Databases & Search: MySQL, Oracle, Cassandra, HBase, OpenSearch, Solr

Other: Recommendation systems, NLP / sentiment analysis, A/B testing platforms, Kerberos security

PROFESSIONAL EXPERIENCE

Walmart Global Tech

Jun 2022 – Present

Senior Data Engineer

- Designed and built ETL pipelines with Apache Kafka and Google BigQuery processing petabytes of transaction data to detect shrink (inventory loss) at self-checkout lanes.
- Redesigned the shrink-identification data model, reducing table footprint and improving query performance by up to 40%.
- Built a data-quality framework that detects and alerts on erroneous upstream transaction data and excludes it from daily reporting.
- Migrated the production analytics dashboard to dbt models, improving resilience and data governance.
- Developed Apache Airflow workflows that deliver daily business reports to stakeholders.

Gallup, Inc.

May 2017 – Jun 2022

Software Engineer

- Led a team of five developers building an action-planning product for managers that helped teams improve engagement scores by 30%.
- Re-architected the survey aggregation engine from a relational store to AWS Glue (Spark), cutting survey compute time from about a day to minutes.
- Designed context-aware, natural-language search for Gallup's content platform using OpenSearch.
- Designed and built a hybrid recommendation system for the content platform.
- Built a text analytics engine on AWS Translate and AWS Comprehend to generate sentiment scores from open-ended survey responses.
- Designed the data pipeline for an A/B/n testing framework used as an experimentation platform for ad hoc research.

Wipro Technologies (client: Mastercard)

Dec 2013 – Jul 2015

Project Engineer – Big Data Administrator / Developer

- Installed, configured and tuned enterprise production Hadoop clusters for Mastercard's analytics team.
- Implemented Kerberos security (cross-realm authentication), high availability, and backup and recovery.
- Built a fraud-detection proof of concept integrating Spark, Kafka and a Java API.

INTERNSHIPS

pSolv LLC

2016 – 2017

Technology Intern

- Built streaming analysis of fraudulent transaction data with Apache Spark, and an ETL pipeline using Kafka, Storm and Oracle with Python/Shell automation and Drools business rules.
- Designed an Apache NiFi data flow to consolidate varied sources into Oracle with consistency transformations.

Copart, Inc.

Aug 2016 – Dec 2016

Data Services Intern

- Built a Kafka pipeline streaming database records to Apache SolrCloud, and Pentaho BI report dashboards for executives and customers.

EDUCATION

Master of Science, Computer Science

May 2017

The University of Texas at Dallas – Richardson, Texas, USA

Bachelor of Technology, Information Technology

Apr 2013

Anna University – Chennai, Tamil Nadu, India