

# Gunasuriya Balasubramani

SENIOR DATA ENGINEER · ARCHITECT

hello@gunasuriya.com · gunasuriya.com · /in/gunasuriya · @gunasuriya ·  
Bangalore, IN · Remote-first

● V.2026.05  
OPEN · REMOTE / BLR / UAE

## PROFILE

Senior Data Engineer with **7+ years** designing *cross-cloud streaming architectures* and unified data platforms. Currently the technical authority for a 6-engineer pod at **SoundHound AI** (acq. Interactions LLC, Sep 2025), consolidating GCP, OCI, AWS and on-prem sources into a single source of truth for Voice & Conversational AI workloads. Reduced infrastructure spend by **~\$1M annually** while eliminating hardware failures and improving streaming latency 5x. Open to **remote Staff / Principal roles** in Bangalore or UAE; available for client meetings and conferences in Dubai.

ANNUAL SAVINGS	COST CUT	THROUGHPUT	CLOUDS UNIFIED
~\$1 <sup>M</sup>	-87%	5.0 <sup>x</sup>	4 <sup>x</sup>

## EXPERIENCE

### Interactions LLC → SoundHound AI

2022 – Present  
3.5+ yrs

SENIOR DATA ENGINEER · VOICE & CONVERSATIONAL AI

- Architected the cross-cloud **Unified Data Platform (UDP)** consolidating streaming, analytical and transactional sources from GCP, OCI, AWS and on-prem into a single source of truth.
- Modernized a 26-node Cloudera CDH on-prem cluster (~\$1.14M/yr) to a GCP-native UDP (~\$144K/yr) — saving **~\$992K annually**, eliminating hardware failures and maintenance overhead.
- Designed per-cloud regional Kafka pattern with cloud-native processing (OCI Dataflow / AWS EMR / Spark on Dataproc) writing into BigQuery via the Storage Write API; secured cross-cloud auth with Vault, WIF and private interconnect.
- Standardized zero-downtime migration playbook (*parallel build* → *dual-write* → *validate* → *cutover*); re-engineered the ClickHouse hot-path consumer onto GCP Dataflow without losing real-time semantics.
- Acted as **technical authority for a 6-engineer pod in Bangalore** coordinating daily with US, UK and Ukraine engineers — architecture design, code review, design approval, mentorship.

Stack: GCP · OCI · AWS · Apache Spark / PySpark · Apache Kafka · Apache NiFi · Dataproc · BigQuery · Dataflow · CloudSQL · GCS · GKE · Vault · WIF · Terraform · Airflow · ArgoCD · Jenkins

### eDirectSys

2021 – 2022  
~2 yrs

FULL-STACK & DATA ENGINEER · DIGITAL MARKETING

- Built data pipelines and the analytics surface for a high-volume email-marketing platform — bulk-send infrastructure, instrumentation, and dashboards.
- Designed instrumentation for opens, clicks, conversions and downstream attribution; powered client-facing campaign-performance dashboards.

Stack: Python · SQL · ETL pipeline design · Campaign instrumentation · Visualization & dashboards · Full-stack web

- Worked on *SolarPulse* — a real-time telemetry product for offshore solar plants. Streamed live IoT data from inverter, grid and environmental sensors via gateway devices into a central time-series store; powered live operational dashboards and forecasting models.

Stack: IoT ingestion · Real-time streaming · Time-series storage · Live dashboards  
· Forecasting

SIGNATURE PATTERNS

MODERNIZATION · CASE STUDY

Cloudera CDH → *GCP-native UDP*.

26-node on-prem Hadoop replaced with autoscaling Dataproc + BigQuery. ~\$1M/yr saved, 5x faster, hardware failures eliminated.

CROSS-CLOUD STREAMING · CASE STUDY

Per-cloud Kafka → *BigQuery, securely*.

Regional Kafka per source cloud, processed locally with OCI Dataflow / AWS EMR / Spark, written to BQ via the Storage Write API.

HOT-PATH MIGRATION · CASE STUDY

ClickHouse MV → *Dataflow*.

Re-engineered Kafka → ClickHouse Materialized-View consumer onto GCP Dataflow without losing real-time semantics.

ZERO-DOWNTIME CUTOVER · METHODOLOGY

Parallel build → dual-write → *validate* → *cutover*.

Standardized migration playbook with reconciliation checks. Zero-incident cutover, every time.

TECHNICAL STACK

Apache Spark / PySpark	core
Apache Kafka · NiFi	streaming
BigQuery · Dataflow · Dataproc	compute
OCI Dataflow · AWS EMR	multi-cloud
CloudSQL · GCS · ClickHouse	storage
Terraform · Airflow · ArgoCD	ops
GKE · Vault · WIF	security

CORE COMPETENCIES

- Cross-cloud streaming architecture
- Cost optimization (~87% / ~\$1M/yr)
- Unified Data Platform design
- Migration playbooks (zero-downtime)
- Security-first patterns (Vault, WIF)
- Petabyte-scale data systems
- Distributed-pod technical authority

CERTIFICATIONS



Google Cloud Associate Cloud Engineer  
▲ SITTING · MAY 2026



Google Professional Data Engineer  
▲ SITTING · JUNE 2026