ANKIT KHARE

San Jose, CA | ankit@ankitkhare.dev | ankitkhare.dev | linkedin.com/in/ankitkhare97 | github.com/ankitkhare-joyent

SUMMARY

Data platform technical lead, 5 years building data and AI infrastructure on Kubernetes for Samsung Private Cloud, now leading two engineers. I helped build an AWS EMR on EKS compatible Spark service from its first month, then the Iceberg lakehouse on it and a Text2SQL API that answers infrastructure questions in Slack.

TECHNICAL SKILLS

Data: Apache Spark, Apache Iceberg, Trino, Apache Airflow 3, AWS Glue, DuckDB, PyIceberg, Apache Superset, Cube, Kafka, Flink, Parquet
AI and LLM: Text2SQL, LLM evaluation (gold and hold-out sets), Model Context Protocol (MCP), RAG, OpenRouter, Claude, prompt design, sqlglot query guards, OPA, Claude Code
Platform: Kubernetes, Helm, Terraform, Temporal, Ray Serve, FastAPI, Docker, spark-operator, ingress-nginx, cert-manager, HashiCorp Vault, Consul
AWS: EKS, EMR on EKS, S3, Glue, IAM, EC2, ECR, Route 53, RDS
Languages: Python, Go, SQL, TypeScript (React), Bash
Operations: GitHub Actions, Jenkins, Grafana, Telegraf, InfluxDB, Graylog, PagerDuty, Prometheus metrics, PostgreSQL, MariaDB, Redis

EXPERIENCE

JOYENT, INC. (Samsung Cloud) Sunnyvale, CA
Data Platform Technical Lead *Nov 2026 - Present*
Senior Data Platform Engineer *Feb 2024 - Nov 2026*
Generalist Engineer *Aug 2021 - Feb 2024*

Internal Analytics: lakehouse and AI query API **2025 - Present**

- Lead the Internal Analytics lakehouse, 188 Apache Iceberg tables and 51 views on S3 and AWS Glue that teams query through Trino and Superset. Wrote 40+ of its Spark ETL jobs and Airflow 3 DAGs, which pull from EC2, VPC and IAM admin APIs, DCIM, Nagios, Ansible, ServiceNow, Jira and PagerDuty.
- On my own initiative, folded per-table Spark jobs and DAGs into a few parameterized ones and scheduled each ETL right after its data lands. Full refresh dropped from 2-3 hours to minutes and now runs hourly, with DAG files down from 185 to 97 and about 41K fewer lines of code.
- Moved about 59 GB of Snowflake history (2024 to present), along with the ingestion and ETL that fed it, onto our self-hosted Iceberg lakehouse and Airflow. That saves about $3,700 a month in compute and storage.
- Cut a 43M-row Iceberg MERGE from 19.3 s to 2.1 s by pruning partitions at planning time, so hourly loads no longer join against all 929M rows of the tags table.
- Wrote most of the IA REST API (FastAPI on Ray Serve, DuckDB, PyIceberg, Redis), the data source behind the team's MCP server for AI agents. Its live tables read straight from the admin APIs and return 20K classified EC2 instances in under a second.
- Built a Text2SQL endpoint that answers plain-English infrastructure questions with Trino SQL, in the API and in a Slack bot, using Claude Haiku through OpenRouter. Every generated query passes a sqlglot guard, and OPA policies hide PII columns (nested fields too) and fail closed.
- Wrote its eval harness, 83 gold questions with 19 held out on tables the prompts never saw: 92.8% overall and 84% on the hold-out. Compared Claude, Gemini, GLM and GPT-4.1 mini on it and cut prompts from 22.3K to 14.7K tokens.

Managed Data Analytics Service (MDAS) **2022 - Present**

- Joined MDAS in its first month (May 2022). It is an EMR on EKS compatible Spark service in Go on Temporal and MariaDB, now in production in Korea, Singapore, Europe and the US. Built ManagedEndpoints (Jupyter Enterprise Gateway kernels), JupyterLab Workspaces, autoscaling, auto-termination, quota checks and the metering APIs.
- Wrote the Terraform for EKS clusters, VPCs, IAM roles and ECR policies, plus Helm charts for spark-operator, ingress-nginx, cert-manager, cluster-autoscaler and the Spark UI proxy. Automated EKS Pod Identity setup at job submission and added mixed instance types for node groups.
- Deployed the Go gateway and Temporal workers with Vault and Consul, and set up metrics in Grafana (via Telegraf and InfluxDB), logs in Graylog and PagerDuty alerts with an SOP for each. I run the team's on-call.
- Added hourly canary jobs in every production region with Slack alerts, and moved CI/CD from Jenkins to GitHub Actions with an image promotion pipeline. Also upgraded to aws-sdk-go-v2 and fixed critical CVEs.

Storage platform **2021 - 2022**

- Built lifecycle, health-monitoring and config-validation tools for Joyent's distributed storage cluster.

Open-source labs (github.com/ankitkhare-joyent): text-to-SQL agents compared three ways (Cube semantic layer, raw SQL, hybrid); Kafka, Flink SQL and Debezium streaming into Iceberg; agent evals with OpenTelemetry and Langfuse.

EDUCATION AND AWARDS

Syracuse University, M.S., Computer Science Syracuse, NY | May 2021
SRM Institute of Science and Technology, B.Tech., Computer Science Chennai, India | Apr 2019
Best Paper Award, International Conference on Internet of Things (ICIOT), 2019.