KRISHNA GOJE

Staff Analytics Engineer | Data Platform Architecture | AI-Augmented Engineering

Phone: +91 8977572730 | Email: krishna19.gk@gmail.com | LinkedIn | Portfolio | GitHub

---

PROFESSIONAL SUMMARY

Staff Analytics Engineer with 14 years building data platforms at American Express, Grubhub, Booking.com, and Opendoor. Pioneered an AI engineering ecosystem connecting 15+ enterprise platforms with parallel agent orchestration and self-learning feedback loops – compressing week-long workflows into under one hour while improving quality and consistency. At Opendoor, architected the operational data layer powering 70%+ of company dashboards, consolidated 100+ datasets into 3 unified models ($300K/year saved), and built the data pipeline behind an AI VoiceBot projected to save $5M annually. Seeking Principal-level roles at the intersection of data platform architecture and AI-augmented engineering.

IMPACT BEYOND ROLE

- **Awarded 2 formal recognition awards** at Opendoor within first year from Sr. Director and Director of Operations for platform impact and delivery velocity
- **Credited by name in company-wide announcement** for architecting the data pipeline behind Opendoor's AI VoiceBot launch, projected to save $5M annually
- **Granted autonomous authority** by Opendoor's Data Director to deprecate and redefine company-wide operational metrics, a responsibility typically reserved for senior leadership
- **Solved Spark's small file problem** on American Express's Open-Source platform, publishing optimization patterns adopted by engineering teams across the 60,000-person enterprise
- **Organized technical and cultural events** engaging 1,000+ colleagues at American Express; published internal technical blogs on Spark and Hive optimization
- **Recognized as Hackathon Finalist (2017) and Data Geek award winner (2018)** for innovation in data engineering approaches at American Express

AI-AUGMENTED ENGINEERING PLATFORM | Built at Opendoor | 2024–Present

Pioneered a first-of-its-kind AI engineering ecosystem that connects 15+ enterprise platforms into a unified, self-improving system – turning one engineer into a high-velocity data team. Reduced week-long analytics workflows to under one hour while simultaneously improving data quality, consistency, and stakeholder response times.

- **Ecosystem Architecture**: Designed and connected 15+ Model Context Protocol (MCP) integrations – Snowflake, Databricks, QuickSight, dbt, Airflow, Slack, Glean, GitHub, Linear, Select Star, Google Workspace, Gmail, Google Calendar, and a custom semantic memory server – into a single AI orchestration layer. Built custom workarounds and API bridges where native MCP connectors did not exist, establishing end-to-end handshake mechanisms across the entire enterprise data stack

- **Parallel Agent Orchestration**: Engineered a multi-agent execution framework where 2-3 AI sub-agents work simultaneously, each running up to 8 parallel data warehouse queries. A 24-query cross-market investigation completes in under 60 seconds. Built 76 domain-specific automation skills that auto-detect task type (data investigation, dashboard update, model creation, stakeholder response) and load specialized execution strategies without manual invocation

- **Self-Learning Feedback System**: Designed 9 automation hooks that capture error patterns, winning approaches, anti-patterns, and domain knowledge at session end, then automatically update system configuration. The platform compounds in speed and accuracy with every use – mistakes are never repeated, optimizations are permanent, and institutional knowledge accumulates infinitely via a custom semantic memory layer

- **Production Automation Library → Open Source (quicksight-mcp)**: Built a 4,874-line Python library (146 functions) for programmatic BI dashboard management – then extracted and open-sourced as an MCP server with 31 tools, 59 tests, and a self-learning engine that tracks error patterns and SPICE timing. Eliminated manual AWS console operations across 6+ production executive dashboards

- **Autonomous Stakeholder Support → Open Source (slack-data-bot)**: Deployed an L1 data support bot that monitors team communication channels, auto-investigates data questions using parallel warehouse queries, drafts responses through a Writer/Reviewer quality loop, and delivers polished answers. Open-sourced with configurable data sources and human-in-the-loop safety

---

EXPERIENCE

Staff Analytics Engineer | Opendoor (NYSE: OPEN) | Aug 2024 – Present

Data Platform & Executive Analytics

- Architected end-to-end operations data platform spanning home assessment through resale, serving as the single source of truth for 70%+ of operational decision-making across the company
- Consolidated 100+ fragmented datasets into 3 unified analytical models, cutting BI compute and storage costs by 60% (~$300K/year in measured savings)
- Built 6 executive dashboards powering Weekly Business Reviews, Sales Performance, and Operations Planning used by VP and C-level stakeholders daily
- Overhauled 8+ core dbt data models, resolving 800+ data quality issues in assessment classification that had silently corrupted downstream reporting
- Eliminated 7+ manual spreadsheet dependencies from production dashboards, removing human-error risk from executive reporting workflows
- Partnered with 15+ cross-functional stakeholders across Operations, Finance, Sales, and Engineering to define metric definitions and reporting standards
- Earned autonomous authority from Opendoor's Data Director to deprecate and redefine company-wide operational metrics, reflecting deep trust in platform judgment and domain expertise

AI Products & Innovation

- Architected data pipeline for Ernie AI VoiceBot, targeting 80% of seller leads that human agents never reach – projected $5M annual savings from automated lead qualification; personally credited by name in company-wide launch announcement
- Built AI-generated executive summary system using Snowflake Cortex, delivering COO-level operational narratives from raw metrics without manual analyst intervention

Lead Data Engineer | Booking.com (via Tavant Technologies) | Jan 2023 – Aug 2024

- Migrated 15+ financial reporting pipelines from legacy MySQL and Hadoop-Hive to cloud-native architecture, reducing report generation latency by 40% and improving data freshness from daily to near-real-time
- Built automated data quality monitoring framework across 50+ financial KPIs, catching reconciliation discrepancies before they reached downstream audit surfaces
- Partnered with Finance and Accounting teams to translate multi-currency reconciliation logic into scalable SQL models, replacing fragile Excel-based workflows
- Implemented incremental loading patterns and partition pruning strategies that reduced pipeline compute costs by 35% while maintaining sub-hour data freshness

Lead Data Engineer | Grubhub | Nov 2021 – Dec 2022

- Led development of Universal Listing Directory, aggregating and normalizing restaurant data from Google, Uber Eats, DoorDash, and internal sources into a single canonical dataset serving 300K+ restaurant partners
- Built ETL pipelines processing multi-source restaurant metadata with deduplication and entity resolution, powering operations dashboards used by 200+ internal stakeholders
- Provided sales teams with unified competitive intelligence on listing coverage, enabling data-driven restaurant onboarding and incentive targeting at scale

---

Lead Data Engineer | American Express | Jan 2017 – Nov 2021

- Designed and built the feature engineering pipeline for Next-Gen Membership Rewards fraud detection, computing 2,000+ behavioral features from 40+ data sources, contributing to a system that prevented an estimated $100M+ in annual fraudulent charges
- Tripled Spark ETL processing speed through pipeline parallelization, partition optimization, and memory tuning – reducing critical batch windows from 6+ hours to under 2
- Solved Spark's small file problem and published memory optimization patterns on American Express's internal Open-Source platform, addressing industry-wide performance bottlenecks adopted by engineering teams across the 60,000-person enterprise
- Designed real-time campaign tracking pipeline connecting 50+ marketing channels to customer conversion events, enabling attribution modeling that informed acquisition spend decisions
- Built Customer Engagement analytics layer measuring cardholder lifecycle metrics (activation, spend velocity, attrition risk) across 60M+ card members
- Published technical blogs on Spark and Hive optimization for internal knowledge sharing, establishing reusable best practices referenced by data engineering teams enterprise-wide
- Organized technical and cultural events engaging 1,000+ colleagues across the organization, building cross-functional community beyond core engineering responsibilities
- Earned Hackathon Finalist recognition (2017) and Data Geek award (2018) for innovation in data engineering approaches

Data Engineer | American Express (via Scepter Technologies) | Apr 2015 – Jan 2017

- Built Real-Time Federated Query Engine enabling analysts to query across NoSQL databases (HBase, Cassandra, Elasticsearch) through a unified SQL interface, reducing cross-platform data access time from days to minutes
- Developed query optimization and caching layers that improved federated query performance by 10x for repeated analytical workloads

OPEN SOURCE

- **quicksight-mcp** – The most comprehensive AWS QuickSight MCP server. 31 tools, 59 tests, self-learning engine. Born from managing 6+ executive dashboards in production. Python, MIT License.
- **slack-data-bot** – Autonomous data Q&A bot with 8-strategy Slack monitoring, parallel AI investigation agents, and Writer/Reviewer quality loop. Human-in-the-loop safety. Python, MIT License.

TECHNICAL SKILLS

Data Platforms: Snowflake, Databricks, AWS (QuickSight, S3, Glue, Lambda, EMR), GCP, Apache Spark, Hadoop, Hive, Kafka

Data Engineering: dbt, Apache Airflow, SQL (advanced), Python, Java, Scala, Shell Scripting, ETL/ELT

BI & Analytics: Amazon QuickSight, Snowflake Cortex (LLM), Data Modeling, Dimensional Modeling, Data Warehousing

AI/ML & Automation: LLM Integration, Prompt Engineering, AI Agent Development, MCP (Model Context Protocol)

Infrastructure: Terraform, Kubernetes, Docker, CI/CD, Git, GitHub Actions

EDUCATION

Master of Science, Computer Science | East Texas A&M University | GPA: 3.52 | 2014

Bachelor of Technology, Information Technology | JNTU Hyderabad | 2012

CERTIFICATIONS

- Databricks Certified Associate Developer for Apache Spark
- AWS Certified Solutions Architect – Associate
- Oracle Certified Professional, Java SE 6 Programmer