# VIBHOR SINGH

contact@vibhorsingh.com | vibhorsingh.com | linkedin.com/in/iamvibhorsingh | github.com/iamvibhorsingh

## PROFESSIONAL SUMMARY

Senior Data & Platform Engineer with 5+ years of expertise architecting enterprise-scale distributed systems and event-driven architectures. Proven track record managing mission-critical Apache Kafka infrastructure across 10+ production clusters serving >100 million daily events with 99.95% uptime. Builds AI-agent tooling on the Model Context Protocol (MCP), including an internal Oracle GoldenGate operations MCP server at Fastenal for LLM-driven root-cause analysis and a geospatial MCP server published to npm and the official MCP registry. Developer tools adopted by government platforms — BoundingBox used in PDOK's official documentation; LookStuffUp Chrome extension featured on Chrome Web Store (5.0 rating). Published researcher. Specialized in real-time streaming, Infrastructure-as-Code, AI-assisted operations, and zero-downtime migrations across hybrid cloud.

## PROFESSIONAL EXPERIENCE

### Data Integration Engineer — Fastenal | *July 2021 – Present*

***Kafka Infrastructure & Platform Engineering***

- Architected and led enterprise-wide **Kafka-as-a-Managed-Service** platform modernization across **10+ production clusters** (on-prem and cloud-native) using Terraform, Puppet, and Azure Cloud Suite, achieving **60% reduction in incident response time** and **99.95% uptime SLA**
- Led critical **zero-downtime migration of 4,000+ Kafka topics** across data centers, coordinating with 15+ cross-functional teams while eliminating **40+ hours/week** of manual operations
- Managed resource allocation and capacity planning for **50+ RHEL VMs** and cloud resources, optimizing annual infrastructure spend and reducing cluster load by **70%** through Puppet automation

***AI & Agentic Operations***

- Built an internal **Model Context Protocol (MCP) server** that gives AI agents structured access to **Oracle GoldenGate** replication state (lag, throughput, discard rates, recent errors), enabling **LLM-driven root-cause analysis and monitoring** in place of manual log and dashboard triage
- Designed the agent harness around deterministic, pre-parsed telemetry rather than raw terminal access, so the model reasons over facts instead of guessing in production database operations; prototyped first as an open-source project (Oracle-Golden-Gate-Mcp)

***Stream Processing & Integration Architecture***

- Engineered **100+ custom SMT Kafka connectors** and Oracle GoldenGate integrations, reducing integration time by **90%** and enabling real-time CDC pipelines for critical financial and enterprise systems
- Designed Kafka Streams applications for real-time data transformation, aggregation, and enrichment, processing **50M+ records daily** with sub-100ms latency
- Established self-service deployment framework enabling **150+ developers** to independently deploy topics and connectors while maintaining governance

***DevOps, Security & Automation***

- Built comprehensive CI/CD automation in Azure DevOps (YAML pipelines) reducing deployment cycle time from **2 weeks to 4 hours** and achieving **95% reduction** in configuration errors
- Developed monitoring and observability stack using **Prometheus, Grafana**, and custom alerting (PagerDuty) reducing **MTTR by 70%**
- Led security remediation for **50+ critical vulnerabilities** across the Data Platform ecosystem
- Mentored **7 junior engineers** on Kafka administration, stream processing patterns, Oracle GoldenGate/SQL suite, and DevOps practices

## SELECTED PROJECTS & PUBLICATIONS

**BoundingBox — Advanced GIS Coordinate Tool & MCP Server** | 2025

Built a full-stack GIS platform adopted by **PDOK (Netherlands' National Geo Platform)** as the recommended coordinate tool in their official OGC API learning modules. Published the **first-of-its-kind dedicated geospatial MCP server** to npm and the official Model Context Protocol registry: 6 tools giving AI agents **EPSG coordinate projection (3,900+ codes), Uber H3 spatial indexing, OpenStreetMap POI discovery, geocoding, and multi-format export** — capabilities previously requiring manual GIS tooling by users. Every response carries a shareable map link so humans can verify agent output; agents find it via llms.txt, an MCP server card, and an agent-skills index.

*Tech: TypeScript, Node.js, Leaflet, Uber H3, Proj4js, MCP SDK, Overpass API, Mapbox Geocoding*

**Oracle GoldenGate MCP Server — AI Root Cause Analysis** | Open Source | 2026

Open-source, lighter counterpart to the internal Fastenal tool. Lets LLM agents diagnose replication lag through the GoldenGate Microservices REST API using structured facts (lag, throughput, discard rates, errors) instead of guesses. **Safety by design:** read-only by default and enforced twice (write tools never advertised, plus a hard block in the routing layer), stateless in-memory operation with opt-in SQLite baselining, and a redacted, **SHA-256 hash-chained audit log** that makes tampering evident.

*Tech: Python, MCP, GoldenGate Microservices REST API, SQLite, LLM Agents*

**LookStuffUp — Chrome Extension** | **Featured** on Chrome Web Store | ★ 5.0 Rating | Production

Published privacy-first browser extension providing instant context for selected text via intelligent multi-source search. One of, if not the first, chromium extension that integrated on-device AI for summarization and translation (20+ languages) using Chrome's built-in ML capabilities. Zero tracking, ads, or data collection. Demonstrates full product lifecycle from design to Chrome Web Store deployment and user acquisition.

*Tech: JavaScript, Chrome APIs, On-Device AI, Privacy Engineering*

**Self-Healing Distributed System POC** | Prototyping | 2024

Engineered a fault-tolerant system demonstrating resilience and automation in hyperscale environments. Containerized microservices communicating via Redis Streams with ML-based anomaly detection (scikit-learn Isolation Forest) to automate system recovery.

*Tech: Python, Docker, Redis Streams, scikit-learn, Microservices*

**Quasi-Euclidean Distance Metric** | Published (AICAI 2020) | arXiv | Google Scholar Profile

Developed a novel quasi-Euclidean distance metric for image similarity evaluation, benchmarked against SSIM and Euclidean distance. Peer-reviewed and published. Full implementation open-sourced.

*Tech: Python, Computer Vision, Image Processing, Research*

## TECHNICAL SKILLS

**Streaming & Messaging:** Apache Kafka (Connect, Streams, ksqlDB), Confluent Platform & Cloud, Apache Flink, Redis Streams, Oracle GoldenGate, Debezium
**AI & Agents:** Model Context Protocol (MCP) servers, LLM tool & agent-harness design, AI-assisted operations (RCA), on-device AI (Chrome built-in AI), adversarial model evaluation, scikit-learn, PyTorch, TensorFlow
**Infrastructure & Cloud:** Terraform, Puppet, Ansible | Azure (Container Apps, Event Hubs, DevOps, AAD, Key Vault, Monitor), AWS, GCP | Cloudflare (Workers, D1) | Kubernetes, Docker, AKS | Linux/RHEL, Bash, PowerShell
**Data Platforms:** MongoDB Atlas, MS SQL Server, MySQL, Oracle, PostgreSQL, Elasticsearch, Couchbase Capella
**DevOps & Observability:** Azure DevOps, Jenkins, GitHub Actions, CircleCI | Prometheus, Grafana, Splunk, Elastic, LogicMonitor, PagerDuty | Git, Gradle
**Programming:** Python (pandas, PySpark, scikit-learn, PyTorch), Java, Go, JavaScript, TypeScript, SQL, C++, R, Groovy
**Frameworks & Tools:** Node.js, Express.js, FastAPI, Flask, GraphQL, Leaflet, Chrome Extensions API, OpenCV, SparkML

## EDUCATION

**Master of Science in Computer Science** *2019 – 2021*
New York University, New York