

HITARTH BHARAD

San Francisco, CA, 94105

☎ +1 (520)-342-9637 ✉ hitarth.bharad@gmail.com 🔗 linkedin.com/in/hitarthbharad 🌐 [Website](#)

Summary

Software Engineer with experience in designing and deploying scalable data pipelines, ETL systems, and cloud-based analytics infrastructure. Expert in building real-time data processing systems handling 10M+ data points daily using Python, SQL, and AWS services (S3, Lambda, EC2, CloudWatch). Proven track record in web scraping, data extraction, transformation, and loading processes across enterprise systems. Strong background in ML-driven analytics platforms and AI-powered decision systems.

Education

University of Arizona

MS Information Science - Machine Learning

Aug 2023 - May 2025

Tucson, Arizona

Dhirubhai Ambani Institute of Information and Communication Technology

Bachelors Engineering Information and Communication Technology

Aug 2017 - May 2021

India

Work Experience

Waterlily

Senior Software Engineer – Data & Machine Learning

Aug 2025 – Present

San Francisco, CA

- Designed and maintained multi-stage ETL pipelines for insurance document processing, including OCR-based data extraction from diverse PDF templates across multiple insurance carriers, achieving 97%+ accuracy in data point extraction.
- Built scalable serverless data processing engine using AWS Lambda, S3, EC2, and CloudWatch, handling real-time financial calculations for long-term care planning with sub-second response times and 99.9% uptime.
- Developed and maintained production web scraping systems for 5 insurance carriers, extracting premium, benefit, and coverage data points; automated CSV-to-PostgreSQL ingestion pipelines ensuring data freshness and integrity.
- Architected PostgreSQL schemas for storing structured insurance data, personal health information, and financial projections; optimized queries for real-time retrieval reducing latency by 40%

FERO.AI

Software Engineer

Sep 2021 – Jul 2024

Dubai, UAE

- Designed and deployed end-to-end ETL pipelines processing 10M+ data points daily across 20+ enterprise logistics clients, implementing real-time data ingestion, transformation, and loading workflows with 99.9% uptime
- Built scalable data infrastructure using AWS (S3, EC2, SQS, CloudWatch) for time-series forecasting systems, optimizing data storage and retrieval patterns that reduced query latency by 35% and storage costs by 25%.
- Implemented comprehensive observability infrastructure with Prometheus, Grafana, and distributed tracing for performance profiling and debugging production systems, maintaining 15-minute incident response SLA and reducing system downtime by 40% through automated monitoring
- Developed dimensional data models for supply chain analytics, implementing feature engineering pipelines for time-series forecasting (XGBoost) that achieved 25% improvement in prediction accuracy.

Verse Innovation

Associate Software Engineer

Jun 2021 – Aug 2021

Bangalore, India

- Designed real-time data ingestion pipelines for streaming analytics processing 5M+ user interactions, ensuring 99.9% uptime and low-latency data availability for recommendation systems.
- Analyzed and optimized API performance metrics (latency, throughput) for high-traffic systems, implementing caching and query optimization reducing response time by 40%.

Projects

Aegion – Real-time Agent Orchestration Platform

aegion.app

FastAPI, Next.js, WebSockets, Vector Search (Pinecone), PostgreSQL, AWS

- Architected low-latency real-time communication system using WebSockets and async Python, implementing efficient message serialization protocols and connection pooling to support concurrent multi-agent workflows with sub-50ms response times
- Built high-performance decision engine with optimized graph traversal algorithms and vector similarity search (Pinecone), processing complex workflow trees with minimal computational overhead through efficient data structures and caching

- Implemented performance monitoring infrastructure with real-time metrics tracking, profiling CPU and memory usage, and optimizing async I/O patterns for scalable multi-user concurrent operations

Hyperspectral Image Classification, VSI Lab – University of Arizona

Python, PyTorch, Transformers, Pandas, NumPy, Data Preprocessing & Annotation

- Developed vision transformer-based architecture achieving low latency and high accuracy for hyperspectral image classification with significant impact on remote sensing and agriculture applications
- Performed extensive data cleaning, annotation, and standardization across multiple hyperspectral datasets with varying dimensional structures, ensuring consistent data quality for model training and reducing preprocessing time by 40%
- Implemented parameter-efficient training techniques reducing computational requirements by 30% while maintaining model performance for resource-constrained deployment scenarios

Tucson Crime Pattern Analysis Dashboard

R (Shiny App, Quarto), PostgreSQL, Leaflet.js, Time-Series Analysis, REST APIs

- Built end-to-end data pipeline fetching and processing crime data from Local Police Department API, implementing automated ETL workflows for real-time data ingestion into PostgreSQL with 99.9% uptime
- Developed interactive analytics dashboard using R Shiny for real-time visualization, integrating Leaflet for geospatial mapping allowing users to filter crime hotspots by time, type, and location
- Applied time-series and geospatial statistical models (ARIMA, spatial clustering) for crime pattern analysis, enabling authorities to allocate resources efficiently based on predictive insights

Anomaly Detection in Time-Series Patterns

Python, Pandas, NumPy, Scikit-learn, Prophet, PostgreSQL, ARIMA, EMA

- Developed time-series anomaly detection pipeline processing and cleaning real-time datasets from financial and sensor sources, implementing automated data quality checks and transformation logic handling 1M+ data points daily
- Applied statistical methods (ARIMA, EMA, Prophet) for detecting anomalies in time-series data, achieving 92% precision in identifying outliers and abnormal patterns with minimal false positives
- Created visualization dashboards using Matplotlib and Seaborn for anomaly detection results, enabling stakeholders to interpret insights and take data-driven actions reducing incident response time by 35%

Technical Skills

Languages: Python, SQL, Shell/Bash, Java, Node.js, JavaScript, TypeScript, R, C++, Rust

Data Engineering & ETL: Data Pipelines, ETL/ELT, Real-time Processing, Batch Processing, Data Ingestion & Web Scraping (Scrapy, BeautifulSoup), Data Cleaning, Data Annotation, OCR Processing

Cloud & Infrastructure (AWS): S3, EC2, Lambda, CloudWatch, SQS, SageMaker, DLQ, Serverless Architecture & CI/CD, Jenkins, Kubernetes, Docker, Microservices

Databases: PostgreSQL, MySQL, MongoDB, Redis, Vector DBs (Pinecone), Query Optimization

Big Data & Analytics: Dimensional Modeling, Data Warehousing, Feature Engineering, Schema Design & Power BI, Statistical Analysis, Time-Series Analysis (ARIMA, Prophet)

Python Ecosystem: Pandas, NumPy, SciPy, Scikit-learn, PyTorch, TensorFlow, FastAPI, Django

Real-time & Messaging: WebSockets, SQS, Kafka, Redis, Async Processing, Event-driven Architecture

Monitoring & Logging: CloudWatch, Elastic Search, Application Performance Monitoring, Observability

ML/AI: LLMs (OpenAI), NLP, Computer Vision, XGBoost, Hugging Face Transformers

& Neural Networks, Model Deployment, ML Pipelines

Backend Development: Python (Django, FastAPI, DRF), Node.js (Express.js, Next.js), Java (Spring Boot) & RESTful APIs, Microservices, API Optimization

Frontend: React, Next.js, TypeScript, Tailwind CSS

Automation & Scripting: Bash/Shell Scripting, Cron Jobs, Deployment Automation, CI/CD Pipelines