

Tejpreet Singh

✉ singhtejpreet2004@gmail.com ☎ (+91) 84272-11756 🌐 singhtejpreet.com in tejpreet-singh-6593a6217

Professional Summary

Final-year Computer Science engineer specializing in Data Engineering, with proven expertise in architecting real-time streaming pipelines, orchestrating ETL workflows, and producing detailed technical documentation. Strengthened by research projects and hackathon success, I combine technical depth with adaptability and collaborative drive to deliver scalable, resilient data systems.

Skills

Languages

Python, PySpark, SQL, Bash, C/C++

Tools & Technologies

Data Engineering & Pipeline Development - Apache Kafka, Apache Spark

Databases: PostgreSQL, InfluxDB, MinIO, XAMPP Server

Infrastructure: Docker(YAML, Compose), Linux Administration, Git(Github, Gitea)

Monitoring & Observability: Grafana, Prometheus

Methodologies

Data Engineering: ETL/ELT Pipelines, Real - Time Streaming Analytics, Event - Driven Architecture

System Administration: Networking Basics, Service Management (systemd, systemctl, launchd), Structured Logging & Telemetry

Soft Skills

Technical Communication & Documentation, Root Cause Analysis & Troubleshooting, Project Leadership & Team Coordination, Agile Development & Sprint Planning

Experience

Research Intern - Data Pipelines | Dixon IoT Lab, Plaksha University

May 2025 – Present

- Engineered a real-time streaming analytics pipeline (Kafka → ML inference → Flask), scaling to 10+ live streams with $\leq 0.1\%$ data loss and sub-200ms latency.
- Deployed a self-hosted Gitea instance on NAS with PostgreSQL backend and RBAC, enabling secure, offline version control and team collaboration across projects.
- Prototyped a Raspberry Pi-based k3s edge cluster (master-slave nodes) as a testbed for distributed data pipelines, supporting a pilot study for an upcoming research paper.

Business Analyst Intern | Roton Consultancies Pvt. Ltd.

July 2024 - Jan 2025

- Collaborated with cross-functional teams across business and technical units, streamlining project communication and reducing feedback cycles by 15%.
- Developed time management and workflow practices under client deadlines, supporting a 10% faster delivery of market studies.

Education

Bachelor of Engineering – Computer Science and Engineering

July 2026

Chandigarh University - Morinda, Punjab

- CGPA: 7.19/10.00

Higher Secondary

April 2022

Stepping Stones Sr. Sec. School - Chandigarh

- CGPA: 8.20/10.00

Matriculation

Sahibzada Ajit Singh Academy - Ropar, Punjab

April 2020

- CGPA: 8.90/10.00

Projects

Enterprise-Scale Real-Time Video Stream Analytics Pipeline

June 2025

Technologies: Apache Kafka, Docker, Flask, Grafana/Prometheus, InfluxDB, Redpanda Console, Git

- Engineered a fault-tolerant Kafka producer with multithreading, systemd automation, and buffered pools, sustaining 24×7 ingestion of 100 events/sec per stream with $\leq 1\%$ data loss..
- Orchestrated a modular consumer pipeline decoding timestamped frames, running TensorFlow Lite classification, and streaming results via Flask, achieving sub-500ms per-frame stream latency.
- Implemented millisecond-granular telemetry capturing frame IDs, CPU/memory/bandwidth metrics, and structured error events, accelerating root-cause analysis and recovery by 60%.
- Dockerized producer/consumer microservices and optimized a secure Kafka cluster topology, scaling reliably to 10+ concurrent live video streams (≥ 20 GB/day).

Orchestrated ETL Pipeline with Airflow, Spark & Object Storage

February 2024

Technologies: Apache Airflow, PySpark, MinIO, PostgreSQL, Docker

- Designed an Airflow DAG orchestrating raw data ingestion \rightarrow Spark transformation \rightarrow curated Parquet loads, automating 100% of ETL workflows..
- Automated data processing, reducing manual entry workload by 70% and enabling operations team to focus on strategic initiatives.
- Deployed PySpark jobs reading/writing via S3A connector to MinIO buckets, processing 50 GB datasets with 2× faster throughput than baseline.
- Integrated validation and schema checks as Airflow PythonOperators, catching 95% of data-quality anomalies pre-load.

System Resource Monitoring Dashboard on macOS

November 2024

Technologies: Python (psutil), MariaDB, Grafana, Launchd, Docker

- Developed a Python service using psutil, capturing CPU, memory, and disk metrics every 10 seconds.
- Automated ingestion into MariaDB via launchd scheduling, ensuring uninterrupted long-term telemetry collection.
- Integrated Grafana with MariaDB, visualizing real-time trends and cutting manual monitoring effort by 30%.

Research Work

Patents

Sept 2024, Published

- An Integrated Automobile Helmet to Promote User Ease And Protection

Papers

Aug 2024 - Jan 2025

- Comparative Analysis of EHRs for SML - aided Thyroid Prognosis

Extra Curriculars

Competitions

- Winner | Hack'Ndore - Municipal Cooperation Indore (1L Cash Prize)
- 2nd Position | Innospark Ideathon - Chandigarh University
- Top 5 Teams | Bharat Tech Experience - Google Developers Group

Volunteering

- PR Coordinator, IEEE CUSB
- Editorial Coordinator, CIS CUSB