

Sumit Kumar

Lead Software Engineer / Backend Specialist

Jamshedpur, JH 831015 • rsumit123@gmail.com • +917077856293 • <https://profile.skdev.one>

SUMMARY

5+ years experienced engineer with experience building scalable cloud systems, AI-driven pipelines, and web data platforms

PROFESSIONAL EXPERIENCE

MINDFIRE SOLUTIONS

Associate Team Lead

Remote

2021-Present

Scalable GCP Infrastructure for Web Crawling

- Led the transition to a scalable cloud infrastructure on GCP for a high-volume web crawling application, improving system efficiency by over 50%.
- Authored Terraform and Ansible scripts for automating Kubernetes and Dataproc deployments, boosting deployment speed and system reliability.
- Designed and deployed infrastructure to crawl millions of pages daily with enhanced observability, cost efficiency, and failover support.
- **Technologies:** GCP, Kubernetes, Terraform, Ansible, MongoDB, Dataproc

High-Volume Crawl Data Processing & Analytics

- Engineered data pipelines to process millions of crawled pages per customer per month, enabling weekly trend reporting on SEO health.
- Designed a multi-stage architecture using Parquet, GCS, and BigQuery for batch processing, followed by streaming analytics via ClickHouse.
- Optimized ClickHouse queries with S3 functions, caching, Bloom filters, and schema design for efficient reporting on SEO issues and anomalies.
- Monitoring at each steps to capture anomalies and issues early using Grafana dashboards
- **Technologies:** GCS, BigQuery, ClickHouse, Parquet, PyArrow, Python, Kubernetes, Grafana

AI-Driven Keyword Generation via LLMs

- Developed an AI prompting system to generate SEO-optimized keywords from crawl data using LLMs, improving automation of content analysis.

- Integrated page sampling and prompt tuning to extract high-quality keywords contextually relevant to customer webpages.
- **Technologies:** Python, LLM APIs, Prompt Engineering

GREENDECK
ML/Data Engineer

Indore, MP
2019-2021

Monolithic to Microservice migration

- Led a team project to migrate a monolithic application to multiple microservices
- Used job control queues like Redis / Kafka to couple different microservices through data pipelines.
- **Technologies Used:** Python, Redis, Kafka, MongoDB, Elastic Search, GCP

Exact product matching using Deep-Learning

- Developed a deep learning solution for exact product matching on an e-commerce platform, using both image and textual descriptions.
- Conducted experiments with multimodal architectures (Corrnet, VSE++, DeepRank) and various text embeddings (BERT, SpaCy, Flair, FastText) to optimize matching accuracy.
- **Technologies Used:** TensorFlow, Python, GCP

EDUCATION

SILICON INSTITUTE OF TECHNOLOGY
Bachelor of Technology, C.S.E

Bhubaneswar, Odisha
2015-2019

CENTRAL BOARD OF SECONDARY EDUCATION
Secondary and Higher Secondary

Jamshedpur, JH
2001-2014

ADDITIONAL INFORMATION

- Program Management: JIRA, Teams, Slack
- Technical Skills:
 - o - Languages: Python
 - o - Backend: Django, Flask, FastAPI
 - o - Cloud & Infra: GCP, AWS, Kubernetes, Terraform, Ansible, Docker
 - o - Data & Storage: BigQuery, ClickHouse, SQL, MongoDB, Parquet, PyArrow

- o - Monitoring: Grafana
 - o - Frontend: ReactJS
 - o - DevOps & Tools: Git, CI/CD, Web Scraping
- Certifications: Deep Learning Specialization (Coursera), Machine Learning (Coursera), Kubernetes Developer (Udemy), Ultimate AWS certified SAA (udemy)