Shubham Tiwari

PROFESSIONAL SUMMARY

A highly motivated Data Engineer with expertise in Python, Scala, cloud technologies, and big data pipelines, seeking to leverage my experience in the design, development, and optimization of data systems. Delivering cost-effective solutions by implementing i.e. scheduler to different compute instances to on and off, regularly monitoring billing for different resources, while ensuring scalability, security (Allocating min. access to resources), and performance (optimizing query size, API request, global variable). Passionate about solving complex data challenges with a focus on automation and cloud infrastructure optimization. Proven track record of delivering results, including saving significant costs for organizations through efficient data engineering strategies and always ready to master new tech-stack.

TECHNICAL SKILLS

- **Programming Languages**: Python, C, C++, PHP, Scala, R (basic), Dart, JavaScript, C#
- **Databases**: BigQuery, PostgreSQL, MySQL, SQLite3, PL/SQL
- **Libraries/Frameworks**: Pandas, NumPy, Matplotlib, Seaborn, TensorFlow, OpenCV, Scikit-learn, MediaPipe, Apache Beam
- **Cloud Platforms**: Google Cloud Platform (GCP)
- **Web Frameworks**: Flask, Django, Bootstrap, HTML, CSS
- **Other Tools**: Docker, Container, Nginx, Jenkins, Ansible, JIRA, Linux, Git, GitLab, Gerrit, DBeaver, Azure, Maya 2020
- **Automation Tools**: CI/CD pipelines, automation scripts deployment, testing, analysis

---

PROFESSIONAL EXPERIENCE

Data Engineer
Sep. 2023 – Present
Ericsson Global India Pvt. Ltd., INDIA
Technologies Used: Python, GCP, Apache Beam, SQL, BigFrame, Pandas, PostgreSQL

- Optimized data infrastructure by implementing a more cost-efficient solution, removing unnecessary cloud Dataproc & Cloud Composer components, reducing costs by ~~30% (~~$1M per year).
- Managed high-volume data ingestion pipelines (6.4 million files daily) and transformed them for ingestion into BigQuery tables, SQL table(for operation purpose)
- Conducted multiple proof-of-concepts (POCs) with Cloud Run, Cloud Functions, Dataproc, and Dataflow's batch & streaming to evaluate the most efficient tools for processing data pipelines.

- Performed comprehensive production load testing to ensure reliability and performance of large-scale data processing workflows.
- Regularly monitoring billing and optimizing cost for different labs- UAT, DEV, DATA, PIPE, PROD
- Collaborated with cross-functional teams to identify and resolve bottlenecks in data processing, improving system efficiency.

Internship at Ericsson India Global Services
Jan. 2023 – Sep. 2023
Technologies Used: Python, GCP, BigQuery, SQL

- Addressed issues related to cloud overspending, lack of automation, and security/compliance concerns within the company's infrastructure.
- Developed a proof-of-concept for a lightweight data ingestion and aggregation pipeline, optimizing data flow and storage.
- Proposed and implemented an automation solution that reduces costs, time, and resource usage, while proactively predicting service degradation.
- Gained valuable experience in a multinational work environment, collaborating with cross-disciplinary teams to solve real-time issues and improve service delivery.

Project Experience / Open Source

DURI – Digital Universal Remote Interface (Research Paper + Patent)
March 2022 – July 2022
Technologies Used: Python, Computer Vision, Nvidia Jetson Nano, MediaPipe, PyQt5, Autopy

- Developed an AI-based software solution allowing users to control public computers (e.g., ATMs) with hand gestures, achieving 98% accuracy.
- The solution aimed to reduce the spread of infectious diseases by eliminating the need to physically touch shared public computers.
- Designed an interface that can be adapted for controlling robots and significantly reduced the manufacturing cost of input devices (e.g., keyboards, mice) by 90%.

Lakhan Mitra: Book to eBook
Nov 2021 – July 2022
Technologies Used: Python, OpenCV, Tesseract OCR, PyQt5, Django, HTML, CSS

- Developed a tool that converts physical books into eBooks in multiple languages (e.g., Tamil, Telugu, Sanskrit), saving users 90% of the time required for manual typing.
- Integrated OCR technology (Tesseract) and image processing (OpenCV) for seamless scanning and text extraction.
- Supported over 100 languages to preserve the literary value of books while facilitating global accessibility.

EDUCATION

Bachelor of Technology in Computer Science Engineering
University Name – Thapar Institute of Engineering & Technology, Patiala

---

ADDITIONAL INFORMATION

- **Competitive Programming**: 5-star rating on HackerRank, 3-star rating on Codechef
- **Certifications**: [Include relevant certifications such as GCP, AWS, or specific data engineering courses]
- **Languages**: English (fluent), [Add any other languages you speak]