YASH RAI

📞 902-633-2517 ✉ Yashrai1224@gmail.com 🔗 linkedin.com/in/yashrai1224 ⭘ github.com/Yash-rai-29 🌐 yash-rai-29.github.io/portfolio/

SUMMARY — Data Engineer specializing in scalable data pipeline development and big data technologies. Proven ability to optimize data workflows, enhance system performance, and reduce operational costs. Adept at leveraging cutting-edge tools to efficiently manage and process large datasets. Seeking to contribute technical expertise and innovative problem-solving skills to a forward-thinking organization.

TECHNICAL PROFICIENCIES

Programming Languages:	Python, SQL	Cloud Services:	Google Cloud Platform (GCP)
Big Data Tools:	Apache Beam, Pub/Sub, Dataform, Dataflow, Databricks, PySpark	Database Systems:	BigQuery, Bigtable, Firebase, NoSQL
		Version Control:	Git, GitHub, Jira

PROFESSIONAL EXPERIENCE

Clarity   **Feb 2024 – Jun 2024**
Software Development Engineer Intern
- Developed a **Customer Data Platform (CDP)** web application that improved user data management, enabling better detection of events, payments, and fraud.
- Engineered a **real-time big data pipeline** using **Python** and **Apache Beam**, processing over **5 million user events** daily from **Pub/Sub** to **BigQuery** and **Bigtable**.
- Implemented an **ETL pipeline** to manage real-time data from Pub/Sub, process with Apache Beam, and store in BigQuery, reducing data handling costs by **10%**.

BinPlus Technologies   **Nov 2023 – Jan 2024**
UI/UX Developer Intern
- Worked on a **casino game project**, adding **Socket.IO** components to enable real-time gameplay connectivity.
- Developed **React components** to facilitate online gameplay interactions.
- Built a **betting website** from scratch for a client, integrating various APIs to enable dynamic user management.
- Implemented **authentication and authorization** mechanisms using Node.js, securing APIs with token-based authentication.

Abhyaz   **Dec 2022 – May 2023**
Web Developer Intern
- Worked on **three different websites** on the Zoho platform, maintaining and developing dynamic and responsive web applications using Zoho sites.
- Implemented various features and functionalities to enhance user experience, including **forms**, **calendar event markers**, and other interactive elements.

EDUCATION

SR Group of Institutions, Jhansi   **2020 – 2024**
B. Tech in Computer Science & Engineering   *AKTU University*

PROJECTS

Real-Time Streaming Data Pipeline
Python, Apache Beam, Google Cloud Pub/Sub, Apache Kafka, BigQuery, Bigtable, Dataflow
- Developed a **real-time streaming data pipeline**: Engineered a pipeline to read data from **Google Pub/Sub** topics that are sourced from **Apache Kafka**. This project dynamically handles 20-25 topics.
- Utilized **Apache Beam** (Python) to create a data pipeline that **transforms and cleans** the incoming data.
- Dynamically writes the cleaned data into **BigQuery** tables according to the topic, ensuring appropriate **data partitioning** and **schema management**.
- For specific topics, data is additionally written into **Bigtable** for **high-throughput** and **low-latency** access.
- Deployed and managed the pipeline on **Google Dataflow**, ensuring **real-time processing** and **monitoring**.

Raw Data to Incremental Table
Google Dataform, BigQuery
- Designed and implemented a **raw data to incremental views** pipeline: Used **Google Dataform** to read raw data from various sources.
- Transformed the data according to business requirements and created different **views** for different tables.
- Joined data from multiple views into a single **incremental table** for further **data analytics** purposes.
- Ensured **data integrity** and **accuracy** by applying necessary transformations and validations.
- Automated the pipeline to run at scheduled intervals, enabling **real-time data availability** for analytics.