Prathmesh Pethkar

San Jose, CA | +1 (551) 727 8705 | prathmesh.pethkar2200@gmail.com | LinkedIn

EDUCATION

Master of Science, Computer Science Aug 2023 - May 2025
San Jose State University, CA, USA | (GPA: 3.7)

Bachelor of Engineering, Computer Science Aug 2018 - May 2022
Pimpri Chinchwad College of Engineering, Pune, India | (GPA: 4.0)

TECHNICAL SKILLS

Languages & Frameworks: Python, Java, JavaScript, C, C++, Shell Script, Flask, Django
Databases: Postgres, MySQL, SQLite, Oracle
Technologies: Google Cloud Platform, Azure, Hadoop, Apache Airflow, Apache beam, PySpark, Pig, Snowflake, Kafka
Tools: Git, Jira, Confluence, Visual Studio Code, Google Cloud CLI
Others: Strong debugging skills, logical thinking, problem-solving skills

PROFESSIONAL BACKGROUND

Data Engineer Intern, Onix, San Jose, CA, USA May 2024 - Present

- Developed a robust dataflow streaming pipeline to efficiently acquire data from Kafka and Google Cloud sources, utilizing Python for data processing, and published the required information to Kafka, ensuring real-time insights.
- Managed the production migration of a live system to a new environment, ensuring minimal downtime for over 10,000 users through meticulous planning and real-time monitoring.
- Built an Apache Beam pipeline to migrate data from Azure to Google Cloud Platform, ensuring data integrity and increasing data processing efficiency by 35% post-migration.
- Integrated client's framework with existing cloud-based system, ensuring full compatibility by conducting thorough testing and fine-tuning configurations, resulting in a seamless transition and reduction in integration issues.

Data Engineer, Datametica Solutions Pvt. Ltd, Pune, India Sep 2022 - Jul 2023

- Constructed end-to-end workflows integrating diverse technologies such as Apache Airflow, SQL, JavaScript, and Spark, reducing operational costs by 25% through optimized data processing.
- Collaborated with industry experts on a solution to process large volumes of complex unstructured data within tight deadlines by implementing Python-based dynamic parallel processing of the data, achieving an 80% boost in operational efficiency.
- Led the reverse engineering process of on-prem Java and Spark workflows, transitioning to cloud-based equivalents, enhancing system scalability by 50% and providing greater flexibility for future developments.
- Resolved critical workflow discrepancies between client's system and cloud platform, restoring downstream application stability within 2 weeks through targeted working sessions and detailed system analysis.

Intern, Datametica Solutions Pvt. Ltd, Pune, India Sep 2021 - Aug 2022

- Spearheaded the development of a Java framework enabling seamless data ingestion into BigQuery from diverse sources, integration with 10+ data sources, and optimized processes leading to a 40% reduction in latency
- Managed Git version control for the framework, overseeing documentation for 5+ releases, reducing conflicts by 30%, and boosting development efficiency by 20%.
- Utilized a variety of GCP services such as Google Cloud Storage, BigQuery, Dataflow, Cloud Function, and Dataproc, harnessing functionalities to optimize data processing, analysis, and workflow automation.
- Conducted client presentation on a custom GCP data ingestion framework through code walkthrough. Clarified complexities, and highlighted increased data processing efficiency, fostering deeper client engagement and understanding.

PROJECT EXPERIENCE

Heart Disease Prediction, Pune University Aug 2021 - Mar 2022

- Analyzed and compared 6+ machine learning techniques like Random Forest, Naive Bayes, Support Vector Machine, and k-Nearest Neighbor for heart disease prediction based on various heart-related medical parameters.
- Integrated model having 90% accuracy with a web application with features for mass prediction, individual reports, and exploratory data analysis for lab personnel.

Blood Bank Management System, Pune University Oct 2020 - Feb 2021

- Built a web application Blood Bank Management System' using flask framework integrating with PostgreSQL and PostGIS to better serve the blood requirements in emergencies.
- Empowered users to conveniently locate nearby blood banks using Google Maps API, assess availability, and trigger notifications to the five closest blood banks regarding urgent requirements.