AYANTIKA NANDI

Bloomington, Indiana | ayanandi@iu.edu | (930)-333-4913 |
linkedin.com/in/ayantika-nandi | github.com/ayantikanandi18

Education

Indiana University Bloomington **Aug 2023 – May 2025**
Master of Science, Data Science Bloomington, Indiana
Coursework: Machine Learning, Big Data, Large Language Model, Advanced Database Technologies, Database Concepts, Cloud Computing, Algorithms, R for Statistical Analysis, Data Visualization, Data Mining, Social Media Mining

Sikkim Manipal University **July 2017 – Aug 2021**
Bachelor of Technology, Computer Science and Engineering Sikkim, India

Skills

Languages: Python, Java, C/C++, JavaScript, HTML/CSS, R, BigQuery SQL, NodeJS, React
Python Libraries: NumPy, pandas, PySpark, Keras, scikit-learn, TensorFlow, PyTorch, Matplotlib, Seaborn, SciPy, PyMuPDF, cv2, os, re, BeautifulSoup, NLTK, Flask, FastAPI, Streamlit, LangChain, Scrapy, Plotly
Database: PostgreSQL, Ansi SQL, SQL (RDMS) , Oracle, Snowflake, sqlite3, NoSQL: MongoDB, Neo4j,
Tools: Apache Kafka, Grafana, Tableau,Git, Docker, Apache Airflow, PySpark, Scala, Hadoop, AWS , Google BigQuery

Experience

Maurer School of Law, Indiana University Bloomington ***Bloomington, Indiana***
Research Data Engineer (Office of the Vice President for Research funded) *May 2024 – Present*

- Used spaCy for NER and Gensim for topic modeling, accurately classifying judicial documents with 90% prediction accuracy.
- Analyzed judicial datasets with Apache Spark and identifying a 40% rise in redacted content and refining workflows.
- Developed data pipelines using PySpark and PostgreSQL, ensuring a high data quality and optimizing query performance.
- Built and maintained automated data workflows to integrate structured and unstructured data and streamlining analytics.

Cognizant Technology Solutions **Hyderabad, India**
Data Engineer | Cognizant Digital Business (CDB-AIA) *Sept 2021 – April 2023*

- Integrated Hadoop with GCP and BigQuery using Dataflow, boosting efficiency by 20% and improving data insights by 40%.
- Designed scalable pipelines with Google Cloud tools to seamlessly migrate 10TB+ of data, reducing transfer time by 30%.
- Optimized BigQuery queries and storage solutions, reducing retrieval times by 30% and cutting storage costs by 15%.
- Automated workflows using Cloud Functions, Cloud Composer, and Pub/Sub, reducing manual effort by 25%.
- Developed validation workflows with BigQuery ML to ensure data consistency and accuracy across large datasets.
- Utilized Google Cloud Monitoring and Cloud Logging to track pipeline performance and resolve anomalies in real time.
- Implemented data archival strategies using Google Cloud Storage, ensuring efficient data retrieval and cost management.

Google Cloud Data Engineer intern | Cognizant Digital Business (CDB-AIA) *April 2021 – Sept 2021*

- Built ETL pipelines using Hadoop, Hive, Pig, and Spark to process 2TB+ raw data, improving reporting accuracy by 25%.
- Designed automated workflows integrating Oracle, HDFS, and BigQuery for seamless data ingestion and real-time analytics.
- Deployed optimized Spark tasks and Python scripts to improve pipeline scalability and processing speed for operations.

Projects

Technical Layoff Trend Analysis **Feb 2025- Mar 2025**

- Developed a system that tracks layoffs using Google Trends data, capturing spikes in searches like "layoffs" and "unemployment" in real time by integrating Apache Kafka in Docker for continuous data streaming and MySQL for structured
- Used Grafana to create interactive dashboards that showcase trends and sudden shifts in search interest, enabling real-time insights into layoff patterns.

Geospatial and Temporal Analysis of Rental Property Trends **Oct 2024- Dec 2024**

- Processed and analyzed geospatial, temporal, and sociological datasets for 10,975 rental properties in Bloomington to track market trends using data mining and statistical modeling.
- Identified a 25% rise in rental permits post-2020 through time-series analysis and clustering techniques, revealing proximity-driven demand shifts and gender disparities, providing actionable insights for urban planning.

Professional Certifications

- Neo4j Professional Certificate
- LLM Engineering: Master AI, Large Language Models & Agents (Udemy)
- AWS Certified Data Engineer Associate 2025- Hands On (Udemy)