AMEY SADANAND BHILEGAONKAR

📞 480-616-3980 ✉ ameybhilegaonkar3@gmail.com ⌗ github.com/ameygoes in linkedin.com/in/amey-bhilegaonkar

With over 3+ years of experience as a **Data Engineer** specializing in ETL data pipelines for **large-scale distributed database** systems, I bring both expertise and enthusiasm for handling **complex data** challenges to your team.

WORK EXPERIENCE

BigCommerce
Austin, Texas
Data Science Intern — June 2023 - August 2023

· Designed and managed a **large-scale Snowflake** data retrieval pipeline for efficient data warehousing.
· Implemented **logistic regression**, and predictive models, improving **customer retention** prediction accuracy by 12%.
· Collaborated with data infrastructure teams to ensure data availability and resolve data-related issues.
· Leveraged **advanced data-mining** techniques to process and analyze millions of data points, extracting critical features for **search indexing and ranking**.

Publicis Sapient
Bangalore, India
Data Engineer - II — June 2019 - July 2022

· Engineered complex **ETL** pipelines using **Apache Spark**, optimizing data **extraction, transformation,** and **loading** from diverse sources, such as **Redshift**, **S3**, **Kinesis Streams**, and **Kafka**.
· Utilized distributed computing frameworks such as **Apache Spark** to manage **large-scale data processing** tasks, enhancing performance and resource utilization by 15%.
· Independently designed, and built database tools and scripts to simplify and automate reduced operation toil by 15%.
· Improved database performance by **optimizing queries** and **tuning indexes**, resulting in 20% reduction in query execution.
· Revamped and maintained **real-time data streaming** solutions with **Apache Spark,** and **GCP Cloud Run** in large-scale infrastructure markets with over 30 million daily customer transactions, resulting in a 15% revenue increase.
· Utilized **AWS Lambda, EC2, and SNS** for scheduled data transformations, cutting pipeline run-time by 2 hours weekly.

PROJECTS

Search Engine for All file types - Opportunity Hackathon - Meta Sponsored

· Spearheaded **Elasticsearch** implementation for blazing-fast search responses, with **millisecond response times**.
· Converted and stored every file type data as **vector embeddings**, ensuring **low-latency search** capabilities.
· Led **Python FAST API development**, providing efficient data access and retrieval.
· Used Machine Learning techniques such as BERT, OCR, ResNet50, and Image Captioning to parse Image features.
· **Collaborated** effectively with team members, optimizing task distribution for streamlined project completion.

Scalable Data Processing Pipeline - Neo4J, Docker, Kafka and Minikube

· Designed and implemented a highly scalable and available data processing pipeline using **Kubernetes**, Kafka, Docker, **Neo4j**.
· Orchestrated the setup of Kafka and Apache **Zookeeper** using **Minikube**, a lightweight Kubernetes implementation.
· Streamlined data ingestion, and processing into Neo4j, applying **PageRank** and **BFS** for graph-based data exploration.

Email Automation Marketing Tool

· Initiated and completed the development of a robust email automation project, streamlining job application outreach and networking efforts with individuals of similar interests. employed **RESTFul** API integration to acquire and manage a comprehensive contact database from CRM tools, demonstrating prowess in **full-stack development** and **API design**.

Speech Emotion Detection

· Researched and optimized existing emotion detection approaches by combining CNN and LSTM networks.
· Discovered emotion-affecting attributes in voice by analyzing audio signal features (MFCC, ZCR, Pitch, Chroma).

EDUCATION

Arizona State University, Tempe, USA — August 2022 - May 2024
Masters of Science in Computer Science (GPA: 4/4)

Pune Institute of Computer Technology, Pune, India — July 2015 - May 2019
Bachelors of Engineering in Electronics and Telecommunications (GPA: 8.78/10)

TECHNICAL SKILLS

Programming Languages	Python, Unix / Linux Scripting
Cloud Platforms & Databases	SQL, GCP, AWS, Big Query, Cassandra
Data Engineering	SnowFlake, Airflow, Spark, Kafka, Pandas, Tableau, D3.js
DevOps / SRE	CI/CD, Git, Jenkins, Docker, Kubernetes
Certified Google Cloud Platform	Associate Cloud Engineer.