

# AMEY BHILEGAONKAR

 [+1-480-616-3980](tel:+14806163980)  [Portfolio](#)  [amey.bhilegaonkar@asu.edu](mailto:amey.bhilegaonkar@asu.edu)  [LinkedIn](#)  [GitHub](#)

## Education

### Master of Computer Science

Arizona State University, AZ, USA

August 2022 - May 2024

3.89 / 4.00

### Bachelor of Engineering in Electronics and Telecommunications

Pune Institute of Computer Technology, MH, India

June 2015 - July 2019

8.78 / 10.00

## Relevant Coursework

- Data Structures and Algorithms
- Object Oriented Programming (OOP)
- Database Management
- Spatial Data Science
- Statistical Machine Learning
- Data Processing at Scale
- Knowledge Representation
- Data Mining

## Technical Skills

**GCP Associate Cloud Engineer** certified.

**Programming Languages:** Python, Java, SQL, Unix / Linux Scripting

**Cloud Platforms and Databases:** GCP, AWS, DBT, MySQL, DynamoDB, Cassandra, PostgreSQL

**Data Engineering:** Airflow, Tableau, ETL, BigQuery, Redshift, Kafka, Spark, Big data, Data warehousing, Data Analytics, Data Visualisation, Pandas, Numpy, Scipy

**DevOps / SRE:** CICD, Git, Version control, Jenkins, Terraform, Automation, Docker, Kubernetes, S3, Data Lake, GCS, SQS, EC2, ECR, SNS, VPC

## Experience

### ASDE 2 Data Engineer - McDonald's Customer Real-Time Data Services

January 2021 - July 2022

*Publicis Sapient, Bangalore, India*

- Designed and implemented efficient data pipelines to extract, transform, and load data from various sources into data warehouse using Cassandra and Java - Spark reducing manual efforts by 60%
- Developed and maintained documentation for data pipelines and data models to understand usage for the team
- Imported data from AWS S3 into spark RDD and performed transformations and actions on RDDs
- Maintained data pipeline up-time of 99.4% while ingesting streaming and transnational data across 8 different data sources
- Improved data processing performance by 20% by performing data partitioning and optimization techniques
- Deployed solutions via automated CI/CD processes to execute data pipelines running in batch.
- **Capabilities:** Java, EC2, ECR, ECS, S3, SNS, VPC, DynamoDB, Scala, Cassandra, DevOps, Terraform, Apache Spark, Data Cleaning, Data Modeling, Data Visualization, Distributed Systems

### ASDE 1 Data Engineer - McDonald's Japan Analytics

June 2019 - December 2020

*Publicis Sapient, Bangalore, India*

- Developed an Automated Data Validation Module to verify Data flow in **Python**, boosting efficiency by 30%
- Created Translation module using google translate API to translate business Data Dashboards into different languages, reducing manual efforts to 10%
- Configured Airflow jobs to orchestrate data processing tasks in GCP cloud environment.
- Constructed Data Ingestion pipelines to process 50 TBs of file data every day to GCP's BigQuery
- Developed and maintained Data Models and enhanced the efficiency of data querying and storage for the purpose of data analysis
- Utilized test-driven development principles to design data transformations running on Apache Spark or DBT that are orchestrated via Airflow.
- **Capabilities:** Python, DBT, Airflow, Kubernetes, GCP, BigQuery, DataFlow, Cloud Storage, AI platform, Data Collection, Data Validation, Data Cleaning, Data Integration

## Projects

### Java Data Engineering using Apache Spark | Java, Apache Spark, AWS, CSV, JSON

December 2022

- Implemented efficient data pipelines to extract, transform, and load data from various sources into the data lake using Apache Spark, Spark Streaming, and Kafka
- applied joins of data frames to various filter transformations, aggregation transformations, etc.
- Integrated and ran Spark Redditi Application on Amazon EMR cluster
- Designed and developed a data processing platform using Java and Apache Spark to analyze large volumes of data in real time

### Data Engineering: YouTube Data Analytics | Python, AWS, Data Ingestion, ETL

September 2022

- Developed and maintained ETL processes using SQL and Python to load data into the data warehouse
- Created data lakes and data warehouses using AWS technologies such as S3 and Redshift
- Partitioned different tables and written ETL jobs using AWS spark, used AWS glue for cataloging
- Created dashboards and reports operating with BI tools such as AWS Quicksight to provide data-driven insights
- Worked with programming languages used for data manipulation, including: PySpark, Spark SQL

### Python Django API based web App | Python, Django, SQL, HTML, CSS, DevOps, AWS, Git

January 2022

- Designed and devised a **REST API** based web application using Python and Django framework to streamline the data-driven processes
- Executed efficient database models and optimized querying and data storage for efficient data analysis
- Improved user experience and engagement by accomplishing responsive design and user-centered features