

# Mars Bansiti

Data Engineer

-  Hà Nội, Việt Nam
-  Zalo +84 565799694
-  Marsbansiti@gmail.com
-  Portfolio: Marsbansiti



Data Engineer with 5+ years of experience specializing in building scalable data pipelines and ETL processes. Proficient in big data technologies like Apache Spark, Kafka, and Airflow. Expert in designing and implementing cloud-based data solutions on AWS, GCP, and Azure. Experienced in data warehouse design and implementation across SQL and NoSQL technologies. Passionate about building robust, efficient data infrastructure that enables data-driven decision making.

## Education

2017 – 2019  
**Master of Computer Science (MSc)**  
National University of Hanoi, Vietnam

2013 – 2017  
**Bachelor of Software Engineering (B.Tech)**  
Delhi University, India

## Technical Skills

### Data Engineering

- Apache Airflow, Kafka, Spark
- ETL/ELT with Python & PySpark
- Data pipeline optimization
- CI/CD for data infrastructure

### Database Systems

- SQL (MySQL, PostgreSQL, Oracle)
- NoSQL (MongoDB, Cassandra)
- Data warehousing (Snowflake, BigQuery)
- Data lakes (S3, Azure Data Lake)

### Programming

- Python (Pandas, NumPy, PySpark)
- SQL & Shell scripting
- Docker & Kubernetes
- Git version control

### Data Analysis

- Power BI & Tableau
- Statistical analysis
- Data visualization
- Dashboard development

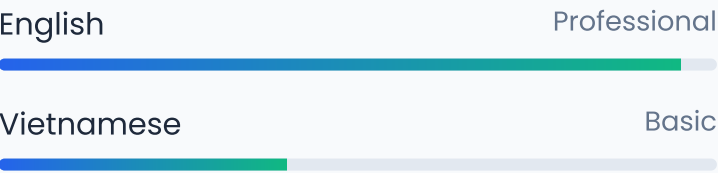
## Certifications

**AWS Certified Data Analytics**  
Amazon Web Services (2023)

**Google Cloud Professional Data Engineer**  
Google Cloud (2022)

**Apache Spark Certification**  
Databricks (2021)

## Languages



## Soft Skills

- Problem Solving
- Agile Methodology
- Attention to Detail
- Teamwork
- Self-motivated
- Mentoring

## Interests

- Tech Blogging
- Photography
- Open Source
- Hiking

## Professional Experience

03/2022 – 02/2025  
Hà Nội, Việt Nam  
**Senior Data Engineer**  
TechVision Vietnam

- Designed and implemented scalable data pipelines using Apache Spark and Airflow
- Built real-time data processing systems with Kafka and Flink
- Developed data warehouse solutions using Snowflake and BigQuery
- Optimized ETL processes reducing processing time by 60%
- Implemented CI/CD pipelines for data infrastructure using GitHub Actions
- Mentored junior data engineers and conducted code reviews

06/2019 – 02/2022  
Đà Nẵng, Việt Nam  
**Data Engineer**  
Digital Solutions Co.

- Built batch and streaming data pipelines using Python and PySpark
- Developed data models and schemas for analytics databases
- Integrated various data sources including APIs, databases, and files
- Implemented monitoring and alerting for data pipelines
- Created comprehensive documentation for data infrastructure
- Collaborated with data scientists on feature engineering

02/2019 – 05/2019  
Hồ Chí Minh, Việt Nam  
**Data Analyst**  
WebCraft Solutions

- Developed SQL queries and reports for business intelligence
- Created data visualizations and dashboards
- Cleaned and prepared datasets for analysis
- Identified data quality issues and implemented improvements
- Automated reporting processes using Python scripts

05/2018 – 01/2019  
Bangalore, Ấn Độ  
**Data Intern**  
CodeCraft Technologies

- Assisted in data cleaning and preparation tasks
- Participated in database design and implementation
- Learned and applied data analysis techniques

## Key Projects

### Data Engineering Solutions

Collection of production-grade data engineering implementations including ETL pipelines, real-time processing systems, and data warehouse solutions. Demonstrated expertise in Apache ecosystem and cloud data technologies.

- Apache Spark
- Airflow
- Kafka
- Python
- Data Pipelines

[View Project Repository](#)

### Bank Customer Analytics

End-to-end analytics platform for banking customer data with segmentation, behavior analysis, and predictive modeling. Delivered actionable insights through interactive dashboards.

- Power BI
- Python
- Machine Learning
- SQL
- Customer Analytics

[View Project Repository](#)

### Vietnam Jobs Market Analysis

Comprehensive analysis of Vietnam's job market trends with salary distributions, in-demand skills, and regional variations. Included automated data collection and interactive visualizations.

- Web Scrapping
- Data Visualization
- Market Analysis
- Python
- Tableau

[View Project Repository](#)

## Technical Expertise

### Data Architecture

Expert in designing scalable data pipelines, stream processing systems, and data lakes. Proficient with Apache ecosystem (Spark, Airflow, Kafka), cloud data solutions, and containerization technologies.

### Data Processing

Strong proficiency in ETL/ELT processes, data integration patterns, fault-tolerant processing, data cleansing techniques, schema design, and building resilient data pipelines for high-volume processing.

### Data Analysis

Skilled in descriptive and diagnostic analytics, statistical methods, exploratory data analysis, and supporting data scientists with feature engineering and dataset preparation for machine learning models.

### Database Systems

Comprehensive knowledge of database architecture, indexing strategies, and optimization techniques across SQL and NoSQL platforms. Experienced in distributed database systems and cloud-native solutions.