

# Sai Ram Bade

[sairam.bade@utdallas.edu](mailto:sairam.bade@utdallas.edu) | (214) 940-9313 | <https://www.linkedin.com/in/sairambade> | <https://github.com/sairam-bade>

Richardson, TX

## EDUCATION

### The University of Texas at Dallas

Master of Science, Business Analytics – Data Science

December 2023

GPA: 3.9

### Rajiv Gandhi University of Knowledge Technologies | India

April 2016

Bachelor of Technology, Mechanical Engineering

GPA: 3.16

## EXPERIENCE

### Data Analytics Intern | ASML, United States

May 2023 – Till Date

- Generating reports and dashboards using Spotfire, Tableau and analyzing them to provide actionable business conclusions.
- Delivered a forecasting algorithm utilizing Python, Scikit-learn, Pandas, and NumPy to predict the refill required in machines.
- Developed a robust data migration framework within Azure Databricks to copy data from network drives to Azure blob storage.
- Built an automation tool utilizing SMTP to address data processing delays, ensuring seamless and uninterrupted operations.
- Engaging with stakeholders to identify opportunities for reporting improvements, enhancing data-driven decision-making.

### Senior Data Engineer | KPMG, India

June 2021 - June 2022

- Led data engineering teams in driving end-to-end data pipeline development, covering integration from raw to curation layers, and employed Databricks notebooks for streamlined data collection, cleansing, and ingestion into Azure Data Lake.
- Managed entire Business Intelligence solutions, overseeing data analysis, technical design, and ETL pipeline development for internal stakeholders. Additionally, provided ongoing support and optimization for existing pipeline infrastructure.
- Spearheaded a significant big data project that involved requirement gathering, design, data collection, data modeling, data reporting, and unit testing, successfully meeting the designated timeline.
- Designed and executed Databricks notebooks utilizing Spark and SQL to extract data from sources including Data Lake, SQL Server, and Oracle SQL. Performed data cleaning, aggregation, manipulation, and tailored ingestion into Azure Data Lake or SQL Server, aligning with specific project requirements.
- Deployed ADF pipelines with scheduled triggers, cutting manual intervention by 40% in automating the ETL process.
- Worked with structured and unstructured data, built pipelines, and collaborated with data scientists to build statistical models.
- Engaged and collaborated with internal stakeholders, maintained analytical reports and interactive dashboards using Power BI and Tableau (Customer Retention, Sales Analysis, and Financial Analysis) to effectively track evolving business initiatives.
- Trained 20+ business users and new hires on Data Modeling, and ETL and worked on new functionality POCs.

### Data Engineer | JPMorgan Chase & Co., India

September 2019 - May 2021

- Developed ETL applications using Python, Scala Spark, and Kafka to transform and load data into Hive tables and HDFS.
- Transitioned the full Java ETL framework to Scala, elevating code readability, maintenance, and performance by 20%.
- Delivered an interactive and visually impactful dashboards using Tableau for the business team to make informed decisions.
- Implemented Python framework to monitor long running and resource intensive jobs in cluster and integrated the live data to Tableau to obtain insightful dashboards for real-time performance monitoring.
- Collaboratively partnered with business analysts to translate business insights into actionable decisions and initiatives.
- Honed expertise in performance tuning techniques for Spark and Hive, contributing to a substantial reduction in cluster resource usage and costs, approximately by 30%.
- Proactively launched a pilot machine learning project to conduct customer risk analytics. Employed decision trees, logistic regression, k-means, and random forest methods, subsequently presenting key insights to the management.

### Big Data Engineer | Tata Consultancy Services, India

January 2017 - August 2019

- Developed Spark applications to clean, transform, and load petabyte scale data into Hive tables in Parquet and Avro formats.
- Created Shell scripts to automate server health checks and disk utilities, scheduled them as CRON and Control-M tasks, significantly diminished the need for manual intervention.
- Collected data from diverse sources, aggregating, manipulating, and utilizing it for analysis, then crafting Power BI dashboards.
- Proficient in data modeling, the development of complex SQL logic, and handling extensive encrypted archived data.
- Worked closely with Big Data administration team to monitor the resource usage to better understand the code performance.
- Regularly engaged with the stakeholders to identify and prioritize key tasks, and to understand and align with the expectations.

## SKILLS

<b>Certifications</b>	Microsoft Azure Data Engineer, Alteryx core designer, AWS Cloud Fundamentals, PG Data Science
<b>Programming</b>	R, Scala, Python - Pandas, NumPy, Scikit-learn, Matplotlib, and NLTK, Python Scripting, Linux, VBA, SAS
<b>Tools/Framework</b>	Hadoop, Spark, Databricks, HDFS, Hive, Kafka, Sqoop, GitHub, Azure Data Factory, Alteryx, Agile
<b>Analytical Tools</b>	Tableau, Microsoft Power BI, Tibco Spotfire, Advanced Microsoft Excel, Minitab
<b>Databases</b>	MySQL, Oracle SQL, PostgreSQL, Microsoft SQL Server, MongoDB