Varun Chitale

LinkedIn +918668616458 varunchitale2015@gmail.com

Summary

Proactive Software Engineer with over 4 years of experience in Big Data engineering, specializing in data pipeline optimization, ETL architecture, and data quality assurance. Expert in transforming large, complex datasets into meaningful insights through advanced data transformation techniques. Skilled in designing and implementing scalable, high-performance solutions that enhance data accessibility, integrity, and cross-platform reliability. A strong team player with a collaborative approach, consistently driving data-informed decisions through meticulous engineering and innovative problem-solving in dynamic, fast-paced environments.

Skills

- **Programming Languages**: Python, PySpark, SQL, PostgreSQL, MySQL, Azure SQL
- **Cloud Technologies**: Azure, Azure Data Factory, Azure Databricks, Azure Synapse, Azure ADLS, AWS S3, AWS RDS, AWS Glue
- **Big Data Technologies**: Hadoop, Hive, HBase, Spark, Kafka, Airflow, Apache Nifi, Flink
- **Data Integration & Validation**: Docker, Git, Jira, Excel (Advanced), Power Query, Jenkins, CI/CD, ETL, Data Warehousing

Work Experience

Software Engineer (Big Data) — **Infobell It Solution** — *Bangalore, KA, IND* — **01/2023 – Current**

Project : Carbon Emission And Reduction Emission

- Led the strategic optimization of ETL pipeline architecture, accelerating data transformation processes by 50% to ensure prompt, high-integrity insights across expansive data ecosystems
- Optimized data pipelines alongside engineering teams, implementing automated quality checks and data validation in cloud-based data integration platforms to ensure consistent, reliable, and efficient data access across systems.
- Facilitated the ingestion and processing of diverse formats (Parquet, JSON, CSV, Avro), ensuring smooth data flow, seamless integration,and compatibility across systems, supporting project requirements and improving data efficiency.
- Optimized complex SQL queries and database structures, reducing query run times by 30% and significantly enhancing analytics efficiency for faster, real-time data processing and insights.
- Utilized Python Pandas to cleanse and transform data, ensuring data integrity, enhancing accuracy, optimizing performance.
- Built scalable PySpark applications, enabling comprehensive carbon emission analyses, real-time insights into enviromental impact and driving data-driven sustainability initiatives for improved decision-making

Junior Data Analyst — **Ubisoft** — *Pune, MH, IND* — **03/2020 - 03/2022**

Project : Xdefiant

- Efficiently managed and optimized large CSV datasets from various sources, ensuring secure cloud storage and efficient retrieval.
- Streamlined data extraction and integration into cloud-based databases, leveraging PostgreSQL for advanced querying, metric analysis, and seamless data synchronization for optimized performance and scalability.
- Engineered and fine-tuned Stored Procedures in PostgreSQL, enhancing query efficiency and optimizing complex data retrieval processes
- Performed thorough validation and cross-referencing of query results using Excel, ensuring data accuracy and consistency, and alignment
- Collaborated in CI/CD deployments using Jenkins, streamlining data updates, automating processes, and ensuring seamless continuous integration within development pipelines for faster delivery and improved quality

Education

Post Graduate Diploma — **Cdac,Kharghar** — *Mumbai, MH, IND* — **02/2022 - 09/2022**

- Major in Big Data Analytics

Bachelor of Engineering — **Pune University** — *Pune, MH, IND* — **01/2014 - 03/2019**

- Major in Electrical and Telecommunication Engineering

Certificates

- Big Data Hadoop&Spark Developer
- HackerRank SQL
- HackeRank Python