

# Kaushik Varma N

Chicago,IL | +1 872-222-6796 | [LinkedIn](#) | [Github](#) | [Hackerrank](#) | [Email](#) | [Website](#)

## **Summary:**

- Master's in Computer Science with over 3+ Years of overall IT experience in Data Engineering.
- Working experience in Hadoop ecosystem (Gen-1 and Gen-2) and its various components such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node, Resource Manager (YARN).
- Experience with components such as Cloudera distribution encompassing components like MapReduce, Spark, SQL, Hive, HBase, Sqoop, Pyspark.
- Good skills on NoSQL Database- Cassandra.
- Proficient in developing Hive scripts for various business requirements.
- Knowledge in Data Warehousing Concepts in OLTP/OLAP System Analysis and developing Database Schemas like Star Schema and Snowflake Schema for Relational and Dimensional Modeling.
- Good hands on in creating custom UDF in Snowflake Data Warehouse.
- Load and transform large sets of structured, semi-structured and unstructured data from Relational Database Systems to HDFS and vice-versa using Sqoop tool.
- Good Experience on architecture and components of Spark, and efficient in working with Spark Core, Data Frames/Datasets/RDD API/Spark SQL, Spark streaming and expertise in building PySpark and Spark-Scala applications for interactive analysis, batch processing and stream processing.
- Hands-on experience in Spark, Scala, SparkSQL, Hive Context for Data Processing.
- Knowledge on GCP tools like Cloud Function, Dataproc, Big Query.
- Experience on Azure cloud i.e., ADF, ADLS, Blob Storage, Databricks, Synapse etc.
- Extensive working experience in an Agile development Methodology & Working knowledge on Linux.
- Expertise in working with big data distributions like Cloudera and Hortonworks. .
- Experience with MSSQL SSIS or other ETL Tools and having advanced hands-on experience with SQL.
- Experienced with reading data models and have worked in an agile software development lifecycle (SDLC).
- Expertise in developing batch data processing applications using Spark, Hive and Sqoop.

## **Experience:**

### **ExxonMobil (contract)**

**Data Engineer** (jan 2023- Nov 2023)

- Engineered SQL scripts to automate query processes, successfully eliminating the need for manual intervention and boosting query volume and accuracy by an impressive 40%
- Designed and implemented an automated system using PowerShell and Azure Cloud Shell, effectively streamlining the deployment of Azure data solutions.
- Crafted advanced data models that enhanced data processing efficiency in Azure, resulting in a substantial 30% increase in overall productivity
- Utilized Azure services such as Data Factory to construct robust data pipelines, enabling the seamless migration of data from legacy SQL servers to Azure Database. Accomplished this through the integration of Data Factories and Python scripting
- Collaborated closely with Analytics and BI teams to develop globally utilized metrics and reports, significantly reducing the need for manual data analysis by more than 15%

### **Tech Mahindra**

**Associate Software Engineer** (Dec 2020 - Dec 2021)

- Designed and implemented a highly scalable data model and data warehouse using Snowflake, leading to a remarkable 10% enhancement in data processing speed and a 15% reduction in storage costs
- Engineered and deployed data pipelines to elevate data quality, resulting in a notable 30% increase in data accuracy.

- Streamlined and fine-tuned ETL processes for seamless data loading into Snowflake, achieving a remarkable 50% reduction in data loading time and a 15% enhancement in overall data quality.
- Pioneered the development of highly optimized Spark code using PySpark and Spark-SQL, resulting in an exceptional 20% improvement in data processing speed and unparalleled data accuracy.
- Building and architecting multiple data pipeline, end to end ETL and ELT process for Data ingestion and transformation.

### **Adiroha**

**Data Engineer** (Nov 2019 - Dec 2020)

- Optimized an existing data pipeline to improve its performance and scalability.
- Created and maintained documentation of data models, ETL processes, and data security policies, resulting in a 30% reduction in onboarding time for new team members and ensuring consistent data governance practices.
- Built a data warehouse in snowflake to capture historical data.
- Conducted data analysis to identify patterns and trends in customer behavior

**Intern** (June 2019 - Nov 2019)

- Developed a new data pipeline to collect and load data from a new data source into the company's data warehouse.
- Implemented complex SQL queries to extract the data from the data warehouse.
- Optimized SQL Queries using indexes as per the company requirements.

### **Projects:**

#### **Winkart | django,sql,javascript,docker([link](#))**

- Built a complete end to end ecommerce website where a user can buy apparels. Used session keys to implement add to cart function to increment/decrement/delete items in the cart. Integrated PayPal payment system so a user can purchase products. Implemented token based logging in for enhanced security

#### **Credit Card Spends | Sql([link](#))**

- Retrieved transaction details for each card type when it reaches a cumulative of 100000. Retrieved which card and expense type combination saw highest month over month growth in Jan-201. Found which city took least number of days to reach its 500th transaction after the first transaction in that city

#### **Snowflake ETL Pipeline | ([link](#))**

- Created 3 layers to store the data and to capture the CDC data overtime. For each layer created 3 pipes, 3 tables and 3 Streams to build the continuous data flow from amazon s3. Used snowflake tasks to automate copying data into tables. Transformed the data in the 2nd layer(curating zone) and sent it to the consumption layer for performing data analysis.

### **More Side Projects**

### **Technical Skills:**

<b>Languages</b>	:	Java, Python, C/C++, SQL, JavaScript
<b>Data Warehousing</b>	:	Snowflake,Pentaho, AWS RedShift
<b>Azure Cloud Tools</b>	:	Data Lake, Blob Storage, VM, Synapse, Data Factory, cosmos
<b>Big Data Tools</b>	:	Hadoop, Hive, Spark, Metastore, Presto, Flume, Kafka
<b>Developer Tools</b>	:	Git, VS Code, Visual Studio, PyCharm, IntelliJ, Eclipse
<b>Libraries</b>	:	Pandas, NumPy, Matplotlib
<b>ML Frameworks</b>	:	ScikitLearn, TensorFlow
<b>No SQL</b>	:	HBase, Cassandra, MongoDB

**Education:**

MS in Computer Science, **Western Illinois University** (Completed-May 2023)

BS in Computer Science, **KL University**(Completed-May 2020)

**Certifications:**

1. AWS Certified Cloud Practitioner
2. Hackerrank SQL intermediate