

KARAN KUMAR RAVELLA

Phone Number: +1 704-675-3003

Email Address: karankumarravella17@gmail.com

LinkedIn Profile: <https://www.linkedin.com/in/ravella-karan-kumar-461312233/>

Portfolio: karan300305.github.io

Professional Summary:

- Accomplished Big Data and Cloud Engineer with 7+ years of hands-on experience, specializing in designing data-intensive applications using the **Hadoop Ecosystem and Cloud Data Engineering**.
- Demonstrated proficiency in implementing end-to-end solutions on Big Data, designing solutions on **Cloudera (CDH3 & CDH4), and Hortonworks** distribution systems.
- Expert in Big Data processing using the entire **Hadoop Ecosystem (MapReduce, Pig, Spark, Scala, Hive, Sqoop, Flume, HBase, Cassandra, MongoDB, Kafka Framework, Zookeeper, Oozie, Storm)** for implementation, maintenance, ETL, and Big Data analysis operations.
- Adept in programming languages like **Scala and Apache Spark**, with extensive experience in **Hadoop, Hive, HBase, Sqoop, Oozie, and Zookeeper**.
- Skilled in analyzing data using **SQL, Hive, and Spark SQL** for **Data Mining and Data Cleansing operations**.
- Extensive use of **Python libraries**, including **PySpark, Pytest, Pymongo, Oracle, PyExcel, Boto3, Psycopg, embedPy, NumPy, Pandas, Starbase, Tabulae, matplotlib, and Beautiful Soup**.
- Proficient in writing Python code for **manipulating data, including data loads, extracts, statistical analysis, modelling, and data munging**.
- Developed data pipelines using **AWS services (EC2, S3, Redshift, Glue, Lambda functions, Step functions, CloudWatch, SNS, DynamoDB, and SQS)** and implemented **ETL processes using IICS Data Integration**.
- Hands-on experience collecting real-time streaming data, creating pipelines for raw data using **Kafka, and storing data into HDFS and NoSQL using Spark**.
- Extensive experience migrating SQL databases to **Azure Data Lake, Azure Data Lake Analytics, Azure SQL Database, Databricks, and Azure SQL Data Warehouse**. Proficient in controlling and granting database access, along with migrating On-premises databases to **Azure Data Lake store using Azure Data Factory**.

- Experienced in **Erwin data modeling iterations (Schematic/Logical/Physical)**, and forward and reverse-engineering processes. Proficient in writing shell scripts in Linux for various data-related operations.
- Utilized **Tableau** for data visualization and analysis on large datasets, drawing insightful conclusions for business teams.
- Collaborated with cross-functional teams, translating intricate data requirements into scalable and efficient solutions using **Cross-functional Collaboration**.
- Collaborated seamlessly with DevOps for Docker-based containerization, elevating deployment efficiency, and ensuring consistent environments in **DevOps Collaboration**.

Skills:

Technical Skills	Tools & Technologies	Database Management
Hadoop Ecosystem	Cloudera (CDH3 & CDH4), Hortonworks	
Big Data Processing	MapReduce, Pig, Spark, Scala, Hive, Sqoop, Flume, HBase, Cassandra, MongoDB, Kafka, Zookeeper, Oozie, Storm	
Programming Languages	Scala, Apache Spark, Python	
Data Analysis and Mining	SQL, Hive, Spark SQL	
Python Libraries	PySpark, Pytest, Pymongo, Oracle, PyExcel, Boto3, Psycopg, embedPy, NumPy, Pandas, Starbase, Tabulae, matplotlib, Beautiful Soup	
AWS Services	EC2, S3, Redshift, Glue, Lambda, Step Functions, CloudWatch, SNS, DynamoDB, SQS	
ETL Processes	IICS Data Integration	
Real-time Streaming	Kafka, Spark.	HDFS, NoSQL
Cloud Platforms	Google Cloud Platform (GCP), Azure	
Data Modeling	Erwin, Scripting and Automation	
Data Visualization Tableau	Shell scripting (Linux)	
Database Migration	Azure Data Lake, Azure Data Lake Analytics, Azure SQL Database, Databricks, Azure SQL Data Warehouse	
ETL Tools	Informatica, DataStage, Snowflake	
Database Design	Normalization, De-Normalization	
Schema Design	Snowflake Schemas	
Collaboration and Leadership	Cross-functional Collaboration, Data Quality Assurance Authority, DevOps Collaboration Expertise, Mentoring and Team Collaboration.	

Professional Experience:

Role: Data Engineer

Client: Dollar General

Location: Goodlettsville, TN

February 2022 - Present

- Designed and implemented robust data infrastructure solutions utilizing **Python, Apache Airflow, and Docker**, ensuring seamless data extraction, transformation, and loading processes architected scalable data solutions.
- Led the development and maintenance of sophisticated data models for optimal storage efficiency and enhanced retrieval speed, incorporating advanced techniques for complex data structures for **advanced data modelling**.
- **Cross-functional Collaboration** with cross-functional teams to define and implement data requirements, translating them into scalable solutions while ensuring alignment with business objectives.
- **Real-time Data Processing Expertise** spearheaded initiatives for real-time data processing, integrating technologies like **Apache Kafka and Spark** to enable instantaneous data insights.
- **Optimization Guru** conducted performance tuning and optimization of **SQL queries**, achieving a significant improvement in database query response time, and enhancing overall system efficiency.
- **Big Data Implementation** successfully integrated **Hadoop** into the data ecosystem, enabling efficient processing and analysis of large datasets, and paving the way for **advanced analytics and machine learning applications**.
- Implemented CI/CD pipelines using **Kubernetes**, streamlining deployment processes and ensuring consistent environments across development and production.
- **Data Quality Assurance Leadership** established and enforced robust data quality checks and monitoring systems, reducing errors and fortifying overall data integrity.
- Spearheaded initiatives for **database optimization**, leveraging techniques such as **partitioning and indexing** to enhance query performance and reduce processing time.
- Provided **technical guidance and mentorship** to junior team members, fostering knowledge sharing and contributing to skill development within the team.

Role: Data Engineer

Client: Emigrant Bank

Location: New York, NY

March 2020 - January 2022

- Develop and execute strategic data architecture plans, incorporating advanced data modelling techniques and overseeing the integration of emerging technologies such as **Apache Kafka and Spark** for real-time data processing.
- Led the integration of Big Data technologies, including **Hadoop and Spark**, to support the scaling of data processing capabilities, enabling the analysis of large datasets for business intelligence and machine learning applications.
- Drive performance improvements through comprehensive **SQL optimization** strategies, resulting in a reduction in database query response time and enhanced system efficiency.
- Architect and implement scalable data solutions on cloud platforms, specializing in **Amazon Redshift, optimizing data storage, and ensuring seamless ETL processes** for diverse and complex datasets.
- Spearheaded the implementation of robust **CI/CD pipelines using Kubernetes, automating deployment processes, and ensuring consistent environments across development, testing, and production.**
- Establish and enforce comprehensive **data quality governance policies and procedures, reducing errors and enhancing overall data integrity** throughout the data lifecycle.
- **Advanced Analytics Enable** advanced analytics by integrating machine learning models into data pipelines, facilitating data-driven decision-making and predictive analytics for business insights.
- **Expert Database Optimization** utilizes advanced database optimization techniques, including partitioning and indexing, to enhance query performance and reduce processing time, ensuring optimal system functionality.
- **Cross-functional Leadership** collaborate closely with cross-functional teams to understand complex data requirements, translating them into scalable and efficient solutions aligned with organizational objectives.
- **Mentorship and Talent Development** provide leadership and mentorship to the data engineering team, fostering a culture of continuous learning, knowledge sharing, and skill development.

Role: Data Engineer

Client: Cipher Health

Location: New York City, NY

August 2019 – March 2020

- **Strategic Data Planning will** Lead the development and execution of strategic data plans, specializing in crafting comprehensive data architectures to support the organization's evolving business needs.
- **Advanced Data Modeling Techniques** employ advanced data modelling techniques to design and maintain intricate data models, optimizing storage efficiency and enhancing retrieval speed for diverse datasets.
- Utilize **Python and Apache Airflow** to engineer cutting-edge **ETL pipelines, ensuring seamless data extraction, transformation, and loading processes for large-scale and complex datasets.**
- Architect and implement scalable data solutions on cloud platforms, with a focus on **Amazon Redshift, facilitating efficient data storage and optimizing performance** for analytical workloads.
- Spearheaded initiatives for **real-time data processing**, integrating technologies like Apache Kafka and Spark, enabling instantaneous data insights and supporting time-sensitive analytics.
- Implement advanced **database optimization strategies**, including partitioning and indexing, to enhance query performance and reduce processing time, ensuring optimal system functionality.
- Collaborate closely with **cross-functional teams** to understand intricate data requirements, translating them into scalable solutions aligned with organizational objectives.
- Drive **continuous improvement** by implementing best practices in data engineering, optimizing processes, and ensuring the organization stays at the forefront of technological advancements.
- Establish and enforce comprehensive **data governance policies**, ensuring data quality, integrity, and security throughout the data lifecycle.

Role: Data Engineer

Client: Tvisha Technologies

Location: Hyderabad, Telangana, India.

May 2017 – July 2019

- Integrate advanced analytics into **data pipelines, collaborating with data scientists** to deploy machine learning models for predictive analytics and business intelligence applications.
- Implement and optimize streaming data processing solutions using technologies such as **Apache Kafka and Spark**, ensuring real-time insights and timely decision-making.
- The automation of routine data infrastructure tasks using **Docker and Kubernetes**, ensuring efficient deployment processes and consistent environments across different stages.
- Architect and implement cloud-native data solutions, leveraging platforms like **AWS and Azure to enhance scalability, flexibility, and cost-effectiveness**.
- Design and implement robust data quality checks and monitoring systems, reducing errors and enhancing overall data integrity throughout **the data processing lifecycle**.
- Conduct performance tuning and optimization of **SQL queries, achieving significant improvements in database query response time** and overall system efficiency.
- Collaborate seamlessly with cross-functional teams to gather and understand data requirements, translating them into scalable and efficient data solutions.
- Manage and optimize databases, specializing in both relational databases like **MySQL and large-scale data warehousing solutions such as Amazon Redshift**.
- **Problem-solving and Troubleshooting will** play a key role in on-call rotations, utilizing problem-solving skills to troubleshoot and resolve data-related issues promptly and maintain 24/7 data availability.

References:

Available upon request.