Meha Dave

✉ meha@jobhuntmails.com 📞 669-204-8139 📍 San Jose, USA 💼 Linkedin

Profile

Accomplished Data Engineer with 3+ years of experience in designing, developing, and implementing data engineering systems. Proven expertise in software development life cycle development (SDLC), QA testing and software engineering. Provided data architecture concepts and data integration support, resulting in improved data quality, performance, and scalability, leading to a 20% increase in system efficiency, data access and a 15% reduction in data processing time.

Skills

Languages: Python, R, SQL, Unix, Scala, MATLAB

Databases: MySQL, MongoDB, PostgreSQL, SQL Server, Oracle, Data Warehouse

Data Processing & Streaming: PySpark, Apache Kafka, Pandas, NumPy, Airflow, Snowflake, T-SQL, PL-SQL, ETL, Apache Airflow, Seaborn, Informatica, SnapLogic

Big Data Tools: Google Data Flow, Hadoop, HDFS, Spark, Hive, Pig, MapRaduce, EMR, Kafka, Sqoop, Oozie

Others: Git, GitHub, Docker, CI/CD pipelines, Jenkins, Power BI, HTML, CSS, Agile, JIRA, AWS, Microsoft Azure, Machine Learning, Deep Learning, NLP, Tensorflow, Keras, Pytorch, KNN, RNN

Professional Experience

Data Engineer, *Lowe's*   01/2023 – Present │ Remote, USA
- Developed and implemented a comprehensive data integration framework utilizing Azure Data Factory, proficiently extracting JSON data from REST APIs, performing transformative processes, and seamlessly loading it into Azure SQL Server, resulting in a notable 20% enhancement in data accuracy.
- Proficiently engineered efficient PySpark and SparkSQL notebooks within Azure Databricks and Synapse environments, skillfully orchestrated through Apache Airflow, ensuring precise ingestion of financial data into Azure Data Lake and Data Warehouse systems.
- Migrated 100 TB of on-premises data from SQL Server and MongoDB to Snowflake Data Warehousing tool via Azure Data Factory, significantly bolstering data management efficiency and optimizing overall data infrastructure.
- Utilized a systematic approach by integrating DBT with Airflow DAGs to define, document, and execute data transformations within Snowflake, resulting in enhanced pipeline efficiency and sustainability.
- Implemented streamlined data streaming processes via Apache Kafka, handling CSV, JSON, and Parquet formats, seamlessly integrating them into Databricks Delta Lake, resulting in a notable 50% reduction in data latency and enhancing overall processing efficacy.
- Pioneered the development of comprehensive CI/CD pipelines utilizing GitHub and Jenkins, overseeing the management, tracking, and deployment of over 100 changes in ELT code, resulting in significant improvements in both code quality and deployment reliability.
- Deployed scalable machine learning pipelines in Python, leveraging Scikit-learn and TensorFlow, to refine dataset performance for real-time analysis and predictive model optimization, thus enabling informed decision-making processes.
- Implemented robust data governance and security measures, ensuring compliance with industry standards such as GDPR and HIPAA, thereby safeguarding sensitive data throughout the entire data lifecycle.
- Developed comprehensive data quality monitoring and anomaly detection systems using tools like Apache NiFi and Prometheus, enabling proactive identification and resolution of data inconsistencies, thus enhancing overall data reliability and integrity.
- Led cross-functional teams in conducting thorough data exploratory analysis and insightful data visualization using tools such as Tableau and Power BI, facilitating data-driven decision-making at strategic levels within the organization.

Data Engineer, *Amazon (Internship)*   05/2022 – 08/2022 │ Seattle, USA
- Developed and optimized trials data pipelines using AWS Redshift Serverless, integrating services like Athena, CloudWatch, S3, VPC, IAM, and EC2 for enhanced data management.
- Overhauled Healthtech trials data architecture, implementing network, coding, and security enhancements, leading to a 20% cost reduction and automated data access.
- Pioneered the shift from manual to automated data access systems, establishing a robust data analytics framework to streamline data accessibility and enhance operational efficiency.

Data Engineer, *Larsen & Toubro Smart World & Communication*   05/2019 – 08/2021 | Ahmedabad, India
- Developed and optimized Python-based ETL pipelines using NumPy and Pandas, streamlining data extraction, transformation, and loading processes for large-scale datasets, resulting in a 20% reduction in processing time.
- Improved SQL query performance within Data Warehouses by implementing indexing strategies and query optimization techniques, achieving a 30% enhancement in data retrieval speed and efficiency.
- Automated intricate data manipulation tasks in Excel using VBA scripting, reducing manual effort by 50% and enabling swift analysis and reporting for business stakeholders.
- Designed and implemented an AWS-based Data Lake architecture utilizing S3, Glue, Redshift, Athena, and EMR (Elastic MapReduce), facilitating seamless integration of diverse data sources and reducing storage costs by 25%.
- Developed interactive reports and dashboards using Power BI, integrating multiple data sources to provide real-time insights and empower stakeholders to make data-driven decisions with enhanced visualization capabilities.
- Implemented advanced statistical analysis techniques, such as regression modeling and clustering, to uncover hidden patterns and trends within datasets, providing deeper insights into customer behavior and market dynamics.
- Conducted A/B testing and hypothesis testing to evaluate the effectiveness of marketing campaigns and product features, leading to data-driven recommendations for optimization and improvement.
- Collaborated with cross-functional teams to define key performance indicators (KPIs) and develop automated reporting solutions, ensuring stakeholders have access to timely and accurate performance metrics for informed decision-making.

Education

Master of Science, *San Jose State University*   08/2021 – 05/2023 | San Jose, USA

Bachelor of Technology, *Birla Vishvakarma Mahavidyalaya*   08/2017 – 05/2021 | Gujarat, India

Projects

Crashless Versatile Multipurpose Drone
- Design and development of a collision-tolerant quadcopter from inception to completion, demonstrating proficiency in project management and technical expertise.
- Implemented Machine Learning algorithms to analyze data acquired during confined flight operations, enabling the real-time detection of structural defects such as cracks, contributing to enhanced safety measures.
- Employed LoRa (Long Range) technology for seamless network transmission of critical data, facilitating efficient communication and control during remote operations.
- Successfully secured a competitive grant of 182,000 INR from the Student Start-up and Innovation Cell, Gujarat, in recognition of the project's innovative approach and potential impact.

Enterprise WAN Network – Network Slicing
- Engineered the network architecture to interconnect multiple geographically dispersed locations, ensuring seamless communication and data transfer across the enterprise.
- Configured and optimized network devices and servers with meticulous attention to detail, employing advanced protocols including HTTPS, DNS, and Email Server configurations via Cisco Packet Tracer's Command Line Interface (CLI).
- Implemented Network Slicing techniques to enhance data transmission efficiency for devices situated in the UK location, effectively minimizing packet loss and optimizing network performance.
- Utilized Modular QoS Command Line Interface (MQC) to define and enforce QoS policies, ensuring prioritized data delivery and optimal resource allocation across the network infrastructure.