Ishan Khanna

Gurgaon, India 122003

+919758599025 –

Professional summary

Knowledgeable Data Engineer with a robust background in data architecture and pipeline development. Proven ability to streamline data processes and enhance data integrity through innovative solutions. Demonstrates advanced proficiency in SQL and Python, leveraging these skills to support cross-functional teams and drive data- driven decision-making.

Skills

Python & Scripting Languages
Big Data Technologies & Hadoop Ecosystem
Data Engineering & ETL Processes
Cloud & Data Platforms (Azure Data Factory & MLStudio)
Containerization & Orchestration (Docker & Kubernetes)

SQL & Data Analysis
Version Control & Collaboration Tools (Git, Jenkins, JIRA)
Data Cleaning & Pre- Processing
Infrastructure Planning & System Analysis
Communication & Problem-Solving

Work history

01/2025 to Current Senior Data Engineer

IBM – Gurgaon, India

Enhanced Month End Processing automation to next level by migrating the scripts to Spark-SQL

Strengthened data reliability by enhancing the validation steps before and after the processing.
Managed and Monitored automated tasks running on LINUX/UNIX Servers.
Worked closely with business to understand the unique needs and proposed tailored solutions.
Currently working on a development and testing of a critical process transition where a Major component of the process is being decommissioned.

11/2023 to 01/2025 Associate

Cognizant Technology Solutions – Gurgaon, India

Automated the month-end processing (ME Process) by developing Python scripts to execute HQL queries in the required sequence, reducing manual intervention and ensuring timely completion.
Implemented threading for parallel execution in Python to optimize processing time and increase efficiency.
Integrated pre- and post-process validations and quality checks to ensure data accuracy and process reliability.
Scheduled and monitored automated processes on Linux/UNIX environments, addressing failures promptly to prevent business disruptions.
Ensured correct execution order of queries and performed necessary validations and quality checks to maintain operational integrity.
Developed a search engine tool to streamline information retrieval from team-specific database tables, enhancing productivity and facilitating the identification of relevant column descriptions.
Designed and implemented machine learning models (TF-IDF Vectorizer and Word2Vec) in Python to optimize search accuracy and relevance.
Enabled keyword-based searches and added functionality for suggesting similar terms and synonyms to improve search results.
Collaborated with clients to gather requirements and define tailored use cases to meet business needs.
Worked closely with the front-end development team to explain functionality and support backend development as needed.

05/2022 to 09/2023 Associate Software Engineer – II

Optum Global Solutions – Noida, India

Automated processes for monitoring and training call centre executives by deploying machine learning models for analysing and predicting key performance metrics, saving over $1.5M by replacing manual tasks.
Designed and optimized pipelines for extracting, transforming, and validating large datasets using Spark-SQL and Python for cloud and server environments.
Deployed machine learning models on cloud infrastructure using Azure Data Factory (ADF), significantly improving process efficiency.
Implemented parallel processing using Celery and Redis on Linux/UNIX environments, accelerating execution times.
Automated workflows with UNIX/Shell scripting, ensuring seamless operations and robust error handling.
Optimized Hive queries for efficient data extraction from the data warehouse.
Conducted comprehensive data validation and facilitated smooth project handoff to the production team.
Collaborated with clients to gather requirements and enhance automation workflows.

08/2020 to 05/2022 Business Analyst

EXL Services – Noida, India

Delivered reports for the AIS Data Delivery project, enabling clients to manage and track medication usage, eligibility, and treatments availed by consumers.
Processed reports using Hive, Python, PySpark, and SAS, archiving data in HDFS for analytical purposes.
Worked with diverse data sources, including Mainframe DSN, Data Warehouse, and input files in formats such as JSON, EXCEL, CSV, TXT, TSV, and XML, within Linux/UNIX environments.
Wrote and optimized complex Hive-based queries to extract and transform data from the Warehouse, performing joins across multiple tables to create reports.
Collaborated closely with business teams to understand requirements and provided solutions for operational challenges.
Led and managed production runs, ensuring efficient job execution and troubleshooting issues to maintain smooth operations.

EDUCATION