Optional: Overview of the DataOps Methodology

Gartner defines DataOps as a collaborative data management practice focused on improving the communication, integration, and automation of data flows between data managers and consumers across an organization. DataOps aims to create predictable delivery and change management of data, data models, and related artifacts. DataOps uses technology to automate data delivery with the appropriate levels of security, quality, and metadata to improve the use and value of data in a dynamic environment."

(Source: https://blogs.gartner.com/nick-heudecker/hyping-dataops/)

A small team working on a simpler or limited number of use cases can meet business requirements efficiently. As data pipelines and data infrastructures get more complex, and data teams and consumers grow in size, you need development processes and efficient collaboration between teams to govern the data and analytics lifecycle. From data ingestion and data processing to analytics and reporting, you need to reduce data defects, ensure shorter cycle times, and ensure 360-degree access to quality data for all stakeholders.

DataOps helps you achieve this through metadata management, workflow and test automation, code repositories, collaboration tools, and orchestration to help manage complex tasks and workflows. Using the DataOps methodology ensures all activities occur in the right order the right security permissions. It helps set in a continual process that allows you to cut wastages, streamline steps, automate processes, increase throughput, and improve continually.

Several DataOps Platforms are available in the market, some of the popular ones being IBM DataOps, Nexla, Switchboard, Streamsets, and Infoworks.

DataOps Methodology:

The purpose of the DataOps Methodology is to enable an organization to utilize a repeatable process to build and deploy analytics and data pipelines. Successful implementation of this methodology allows an organization to know, trust, and use data to drive value.

It ensures that the data used in problem-solving and decision making is relevant, reliable, and traceable and improves the probability of achieving desired business outcomes. And it does so by tackling the