

DataOps Methodology

Establish DataOps - Prepare for operation

You will learn about the people who are involved in defining data, curating it for use by a wide variety of data consumers, and how they can work together to deliver data for a specific purpose.

Establish DataOps – Optimize for operation

You will learn about how the DataOps team works together in defining the business value of the work they undertake to be able to clearly articulate the value they bring to the wider organization.

Iterate DataOps - Know your data

You will learn about the capabilities that you will need to use to understand the data in repositories across an organization. Data discovery is most appropriately employed when the scale of available data is too vast to devise a manual approach or where there has been institutional loss of data cataloging. It utilizes various techniques to programmatically recognize semantics and patterns in data. It is a key aspect of identifying and locating sensitive or regulated data to adequately protect it, although in general, knowing what stored data means unlocks its potential for use in analytics. Data Classification provides a higher level of semantic enrichment, enabling the organization to raise data understanding from technical metadata to a business understanding, further helping to discover the overlap between multiple sources of data according to the information that they contain.

Iterate DataOps – Trust your data

You will learn that understanding data semantics helps data consumers to know what is available for consumption, but it does not provide any guidance on how good that data is. This module is all about trust, how reliable a data source can be in providing high fidelity data that can be used to drive key strategic decisions, and whether that data should be accessible to those who want to use it; whether the data consumer is permitted to see and use it. This module will address the common dimensions of data quality, how to both detect and remediate poor data quality. And it will look at enforcing the many policies that are needed around data quality, not least the need to respect an individual's wishes and rights around how their data is used:

Iterate DataOps – Use your data

You will learn that providing useful data in a catalog can often necessitate some transformation of that data. Modifying original data can optimize data ingestion in various use-cases, such as combining multiple data sets, consolidating multiple transaction summaries, or manipulating non-standard data to conform to international standards. This module will examine the choices for data preparation, how visualization can be used to facilitate the human understanding of the data and what needs to be changed, and the various options for single use, optimization of data workflows and ensuring the regular production of transformations for operational use. Furthermore, this module will show you how to plan and implement the data movement and integration tasks that are required to support a business use case. The module is based on a real-world data movement and integration project required to support implementation of an AI-based SaaS analytical system for supply chain management running in the Google cloud. The module will cover the major topics that need to be addressed to complete a data movement and integration project successfully.

Improve DataOps

In this lesson you will learn about evaluating the last data sprint, observe what worked and what did not, and make recommendations on how the next iteration could be improved.