

ENSURING DATA QUALITY STANDARD WITH DEVOPS

1. Ensuring Data Quality Standards with DevOps

00:00 - 00:08

High-quality data is one of the essential assets for any organization. Let's look closer at how DevOps help maintains it.

2. DevOps ensures good code

00:08 - 00:24

The goal of most software is to handle data, and good software manages the data well and maintains high quality. DevOps is about developing, testing, and deploying high-quality software, therefore helping maintain high data quality.

3. Data quality

00:24 - 00:56

Data quality is about how trusted the information in a dataset is. It is crucial to capture the most recent and accurate information in our data sets. Imagine conducting data analysis on a dataset that has false information. The results of the data analysis will also be incorrect. High-quality data can be tricky and costly, and all data may require a different quality. Therefore it is essential to define data quality factors and look at them closely.

4. Elements of data quality

00:56 - 01:15

Elements of data quality are parameters that help us break down and define the quality. These elements are characteristics of a data set that should be traced and standardized. Code quality elements are Accuracy, Completeness, Consistency, Relevance, and Timeliness.

5. Accuracy

01:15 - 01:37

Accuracy is about the correctness of data in every detail. Imagine the courses we didn't take would be listed on our DataCamp accounts. It would be inaccurate because we did not take those courses. We want only the courses we took to be listed in our accounts. The accuracy of the data should be closely monitored and tracked.

6. Completeness

01:37 - 02:00

Completeness is about the comprehensiveness of data. It is important to check no data is lost while it is stored in a database or when it is moved between databases. Imagine if our DataCamp account does not list the courses we completed. How frustrating it would be! We want all the finished courses to be listed entirely in our account.

7. Consistency

02:00 - 02:29

Consistency of the data is about the reliability of the information. Data should not contradict other information in the same system. Imagine if we click on a course and see we have finished it on the course page, but it is not listed on our account. This would mean the data is not consistent within the same system. Consistency or integrity of data is vital and should be tested regularly.

8. Relevance

02:29 - 02:41

Relevance is about holding and storing only the necessary information. If a dataset has too much irrelevant information, handling that data will waste time and resources.

9. Timeliness

02:41 - 03:11

Timeliness is about how up-to-date the information is. Timeliness is most of the time extremely important because old data might be replaced with more recent information. In data analysis, raw data is not used. Raw data could be the data stored within microservices. Microservices usually communicate with each other via APIs to get the most recent information from other systems and not rely on stored data.

10. Let's practice!

03:11 - 03:26

Data Quality can only be achieved with high-quality code and software, and DevOps is our best chance to accomplish that. Let's go to the exercises and practice the Data Quality standards.