Preventing data quality issues with data contracts: a proactive solution

Vanya Petrova Kostova
Global TECH, HelloFresh SE, Germany.

Abstract

Data quality is a critical aspect of data product management and a major challenge in the field of data engineering. It refers to the availability, accuracy, completeness, and consistency of data, which are essential factors for reliable and informed data-driven business decisions. In data flows, data quality is often compromised by errors, missing values and inconsistencies that occur already in the initial source systems. In practice, such issues are getting addressed with filtering and post processing techniques to clean and format the data accordingly but delays from the time of detection until resolution may cause unacceptable risks in data-driven decision-making processes and thus, may harm the overall business.

We propose data contracts as a mechanism to address data quality issues at the root cause. Data contracts between data producers and data consumers enable defining and tracking data lineage and ensure that a data consumer can analyse and model the data in time-critical business situations. Compared to the more traditional approaches of data quality monitoring and alerting, which are designed to identify and raise an issue, data contracts can help organisation to avoid data quality issues before they affect data flows and business operations.

Keywords: data quality; data contracts; data products; data-driven decision-making.