Google

Reading: Schema-validation checklist

In this course, you have been learning about the tools business intelligence professionals use to ensure conformity from source to destination: schema validation, data dictionaries, and data lineages. In another reading, you already had the opportunity to explore data dictionaries and lineages. In this reading, you are going to get a schema validation checklist you can use to guide your own validation process.

Schema validation is a process used to ensure that the source system data schema matches the target database data schema. This is important because if the schemas don't align, it can cause system failures that are hard to fix. Building schema validation into your workflow is important to prevent these issues.

Common issues for schema validation

- **The keys are still valid:** Primary and foreign keys build relationships between tables in relational databases. These keys should continue to function after you have moved data from one system into another.
- **The table relationships have been preserved:** The keys help preserve the relationships used to connect the tables so that keys can still be used to connect tables. It's important to make sure that these relationships are preserved or that they are transformed to match the target schema.
- **The conventions are consistent:** The conventions for incoming data must be consistent with the target database's schema. Data from outside sources might use different conventions for naming columns in tables– it's important to align these before they're added to the target system.

Using data dictionaries and lineages

You've already learned quite a bit about data dictionaries and lineages. As a refresher, a data dictionary is a collection of information that describes the content, format, and structure of data objects within a database, as well as their relationships. And a data lineage is the process of identifying the origin of data, where it has moved throughout the system, and how it has transformed over time. These tools are useful because they can help you identify what standards incoming data should adhere to and track down any errors to the source.

The data dictionaries and lineages reading provided some additional information if more review is needed.

Key takeaways

Schema validation is a useful check for ensuring that the data moving from source systems to your target database is consistent and won't cause any errors. Building in checks to make sure that the keys are still valid, the table relationships have been preserved, and the conventions are consistent before data is delivered will save you time and energy trying to fix these errors later on.