

 100 XP

Determine operational needs

3 minutes

Once you've identified the kind of data you're dealing with (structured, semi-structured, or unstructured), the next step is to determine how you'll use the data. For example, as an online retailer you know customers need quick access to product data, and business users need to run complex analytical queries. As you work through these requirements, taking your data classification into account, you can start to plan your data storage solution.

Here, you'll go through some of the questions to ask when determining what to do with your data.

Operations and latency

What are the main operations you'll be completing on each data type, and what are the performance requirements?

Ask yourself these questions:

- Will you be doing simple lookups using an ID?
- Do you need to query the database for one or more fields?
- How many create, update, and delete operations do you expect?
- Do you need to run complex analytical queries?
- How quickly do these operations need to complete?

Let's walk through each of the data sets with these questions in mind and discuss the requirements.

Product catalog data

For product catalog data in the online retail scenario, customer needs are the highest priority. Customers will want to query the product catalog to find, for example, all men's shoes, then men's shoes on sale, and then men's shoes on sale in a particular size. Customer needs may require lots of read operations, with the ability to query on certain fields.

In addition, when customers place orders, the application must update product quantities. The update operations need to happen just as quickly as the read operations so that users don't put an item in their shopping carts when that item has just sold out. This will not only result in a large number of read operations, but will also require increased write operations for product

catalog data. Be sure to determine the priorities for all the users of the database, not just the primary ones.

Photos and videos

The photos and videos that are displayed on product pages have different requirements. They need fast retrieval times so that they are displayed on the site at the same time as product catalog data, but they don't need to be queried independently. Instead, you can rely on the results of the product query, and include the video ID or URL as a property on the product data. So, photos and videos need only be retrieved by their ID.

In addition, users will not make updates to existing photos or videos. They may, however, add new photos for product reviews. For example, a user might upload an image of themselves showing off their new shoes. As an employee, you also upload and delete product photos from your vendor, but that update doesn't need to happen as fast as your other product data updates.

In summary, photos and videos can be queried by ID to return the whole file, but creates and updates will be less frequent and are less of a priority.

Business data

For business data, all the analysis is happening on historical data. No original data is updated based on the analysis, so business data is read-only. Users don't expect their complex analytics to run instantly, so having some latency in the results is okay. In addition, business data will be stored in multiple data sets, as users will have different access to write to each data set. However, business analysts will be able to read from all databases.

Summary

When deciding what storage solution to use, think about how your data will be used. How often will your data be accessed? Is your data read-only? Does query time matter? The answers to these questions will help you decide on the best storage solution for your data.

Next unit: Group multiple operations in a transaction

Continue >