✓ 100 XP

Choose a storage solution on Azure

7 minutes

Choosing the correct storage solution can lead to better performance, cost savings, and improved manageability. Here, you'll apply what you've learned about the data in your online retail scenario, and find the best Azure service for each data set.

Product catalog data

Data classification: Semi-structured because of the need to extend or modify the schema for new products

Operations:

- Customers require a high number of read operations, with the ability to query on many fields within the database.
- The business requires a high number of write operations to track the constantly changing inventory.

Latency & throughput: High throughput and low latency

Transactional support: Required

Recommended service: Azure Cosmos DB

Azure Cosmos DB supports semi-structured data, or NoSQL data, by design. So, supporting new fields, such as the "Bluetooth-enabled" field or any new fields you need in the future, is a given with Azure Cosmos DB.

Azure Cosmos DB supports SQL for queries and every property is indexed by default. You can create queries so that your customers can filter on any property in the catalog.

Azure Cosmos DB is also ACID-compliant, so you can be assured that your transactions are completed according to those strict requirements.

As an added plus, Azure Cosmos DB also enables you to replicate your data anywhere in the world with the click of a button. So, if your e-commerce site has users concentrated in the US, France, and England, you can replicate your data to those data centers to reduce latency, as you've physically moved the data closer to your users.

Even with data replicated around the world, you can choose from one of five consistency levels. By choosing the right consistency level, you can determine the tradeoffs to make between consistency, availability, latency, and throughput. You can scale up to handle higher customer demand during peak shopping times, or scale down during slower times to conserve cost.

Why not other Azure services?

Azure SQL Database would be an excellent choice for this data set if you could identify the subset of properties that are common for most of the products and the variable properties that might not exist in some products. Azure SQL Database enables you to combine structured data in the columns, and semi-structured data stored as JSON columns that can be easily extended. Azure SQL Database can provide many of the same benefits of Azure Cosmos DB, but it provides little benefit if the structure of your data is changing in different entities, and you cannot pre-define a set of common properties that are repeated in most of the entities. Unlike Azure CosmosDB that index every property in the documents, in Azure SQL Database you need to explicitly define what properties from semi-structured documents should be indexed. Azure Cosmos DB is better choice for highly unstructured and variable data where you cannot predict what are the properties that should be indexed.

Other Azure services, such as Azure Table storage, Azure HBase as a part of HDInsight, and Azure Cache for Redis, can also store NoSQL data. In this scenario, users will want to query on multiple fields, so Azure Cosmos DB is a better fit. Azure Cosmos DB indexes every field by default, whereas the other services are limited in the data they index, and querying on non-indexed fields results in reduced performance.

Photos and videos

Data classification: Unstructured

Operations:

- Only need to be retrieved by ID.
- Customers require a high number of read operations with low latency.
- Creates and updates will be somewhat infrequent and can have higher latency than read operations.

Latency & throughput: Retrievals by ID need to support low latency and high throughput. Creates and updates can have higher latency than read operations.

Transactional support: Not required

Recommended service: Azure Blob storage

Azure Blob storage supports storing files such as photos and videos. It also works with Azure Content Delivery Network (CDN) by caching the most frequently used content and storing it on edge servers. Azure CDN reduces latency in serving up those images to your users.

By using Azure Blob storage, you can also move images from the hot storage tier to the cool or archive storage tier, to reduce costs and focus throughput on the most frequently viewed images and videos.

Why not other Azure services?

You could upload your images to Azure App Service, so that the same server that is running your app is serving up your images. This solution would work if you didn't have many files. But if you have lots of files, and a global audience, you'll get more performant results by using Azure Blob storage with Azure CDN.

Business data

Data classification: Structured

Operations: Read-only, complex analytical queries across multiple databases

Latency & throughput: Some latency in the results is expected based on the complex nature of the queries.

Transactional support: Required

Recommended service: Azure SQL Database

Business data will most likely be queried by business analysts, who are more likely to know SQL than any other query language. Azure SQL Database could be used as the solution by itself, but pairing it with Azure Analysis Services enables data analysts to create a semantic model over the data in SQL Database. The data analysts can then share it with business users, so that they only need to connect to the model from any business intelligence (BI) tool to immediately explore the data and gain insights.

Why not other Azure services?

Azure SQL Data Warehouse supports OLAP solutions and SQL queries. But your business analysts will need to perform cross-database queries, which Azure SQL Data Warehouse does not support.

Azure Analysis Services could be used in addition to Azure SQL Database. But your business analysts are more well-versed in SQL than in working with Power BI. So they'd like a database

that supports SQL queries, which Azure Analysis Services does not. In addition, the financial data you're storing in your business data set is relational and multidimensional in nature. Azure Analysis Services supports tabular data stored on the service itself, but not multidimensional data. To analyze multidimensional data with Azure Analysis Services, you can use a direct query to the SQL Database.

Azure Stream Analytics is a great way to analyze data and transform it into actionable insights, but its focus is on real-time data that is streaming in. In this scenario, the business analysts are looking at historical data only.

Summary

Each type of data has different storage requirements, and it's your job to figure out which solution is best. Always consider the type of data, the operations required, expected latency, and the need for transactional support.

Next unit: Summary

Continue >