

- Inventory management
- Order management
- Reporting database
- Accounting

Examples

- Organize data into series of two-dimensional tables with rows and columns
- Structured Query Language (SQL) for data retrieval and management
- Conforms to ACID (Atomic, Consistent, Isolated and Durable) for data updates
- Schema on write model: data structure defined and read/write operations use schema
- Strong consistency guarantees
- Does not shard
- Must be normalized

Concepts

Relational database management system

Azure Services

- Azure SQL Database
- Azure Database for
 - MySQL
 - PostgreSQL
 - MariaDB

Workload

- Records are frequently created and updated.
- Multiple operations have to be completed in a single transaction (atomic)
- Relationships are enforced using database constraints.
- Indexes are used to optimize query performance.

Data types

- Data is highly normalized.
- Database schemas are required and enforced.
- Many-to-many relationships between data entities in the database.
- Constraints are defined in the schema and imposed on any data in the database.
- Data requires high integrity.
- Indexes and relationships need to be maintained accurately.
- Data requires strong consistency. Transactions operate in a way that ensures all data are 100% consistent for all users and processes.
- Size of individual data entries is small to medium-sized.



Compiled by Dr T Oberholster
Data store models

- [Website & Blog](#)
- [TeePublic Shirts](#)
- [RedBubble Merch](#)
- [SpreadShirt Apparel](#)
- [Contrado Fashion](#)

- Data caching
- Session management
- User preference and profile management
- Product recommendation and ad serving



Compiled by Dr T Oberholster
Data store models

- [Website & Blog](#)
- [TeePublic Shirts](#)
- [RedBubble Merch](#)
- [SpreadShirt Apparel](#)
- [Contrado Fashion](#)

Key/value stores

Examples

Concepts

Azure Services

Data types

Workload

- A key/value store associates each data value with a unique key
- Support simple query, insert, and delete operations
- Must overwrite the existing data for the entire value
- Read/write of values is atomic
- Schema information must be provided by the application
- Retrieves or stores value by key
- Optimized for simple lookups
- Extremely scalable & distributable

- Azure Cosmos DB Table API
- Azure Cache for Redis
- Azure Table Storage

- Each key is associated with a single value.
- There is no schema enforcement.
- No relationships between entities.

- Data is accessed using a single key, like a dictionary.
- No joins, lock, or unions are required.
- No aggregation mechanisms are used.
- Secondary indexes are generally not used.

- Product catalog
- Content management
- Inventory management



Compiled by Dr T Oberholster
Data store models

- [Website & Blog](#)
- [TeePublic Shirts](#)
- [RedBubble Merch](#)
- [SpreadShirt Apparel](#)
- [Contrado Fashion](#)

- Data can be managed in de-normalized way.
- Size of individual document data is relatively small.
- Each document type can use its own schema.
- Documents can include optional fields.
- Document data is semi-structured, meaning that data types of each field are not strictly defined.

Concepts

- A document database stores a collection of documents, where each document consists of named fields and data
- Simple values or complex elements (lists and child collections)
- Retrieved by unique keys
- Contains the data for single entity (customer/order)
- May contain information that would be spread across several relational tables in an RDBMS
- Documents don't need to have the same structure
- Can store different data in documents

Azure Services

- Azure Cosmos DB SQL API

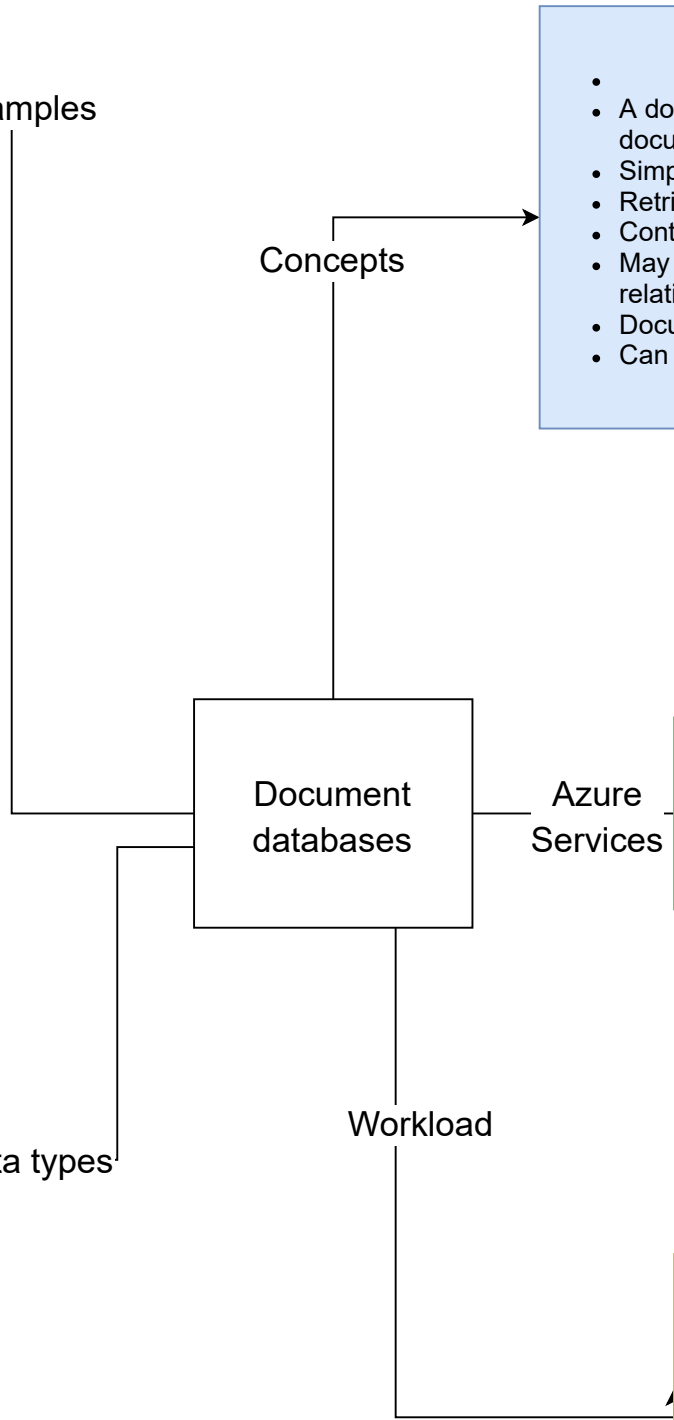
Workload

- Insert and update operations are common.
- No object-relational impedance mismatch. Documents can better match the object structures used in application code.
- Individual documents are retrieved and written as a single block.
- Data requires index on multiple fields.

Examples

Document databases

Data types



- Organization charts
- Social graphs
- Fraud detection
- Recommendation engines



Compiled by Dr T Oberholster
Data store models

- [Website & Blog](#)
- [TeePublic Shirts](#)
- [RedBubble Merch](#)
- [Spreadshirt Apparel](#)
- [Contrado Fashion](#)

- Nodes and relationships.
- Nodes are similar to table rows or JSON documents.
- Relationships are just as important as nodes, and are exposed directly in the query language.
- Composite objects, such as a person with multiple phone numbers, tend to be broken into separate, smaller nodes, combined with traversable relationships

Concepts

- Stores two types of information, nodes and edges
- Edges specify relationships between nodes.
- Nodes and edges can have properties that provide information about that node or edge, similar to columns in a table.
- Edges can also have a direction indicating the nature of the relationship.
- Perform queries across the network of nodes and edges and analyze the relationships between entities
- Custom query language to traverse a network of relationships efficiently

Graph databases

Azure Services

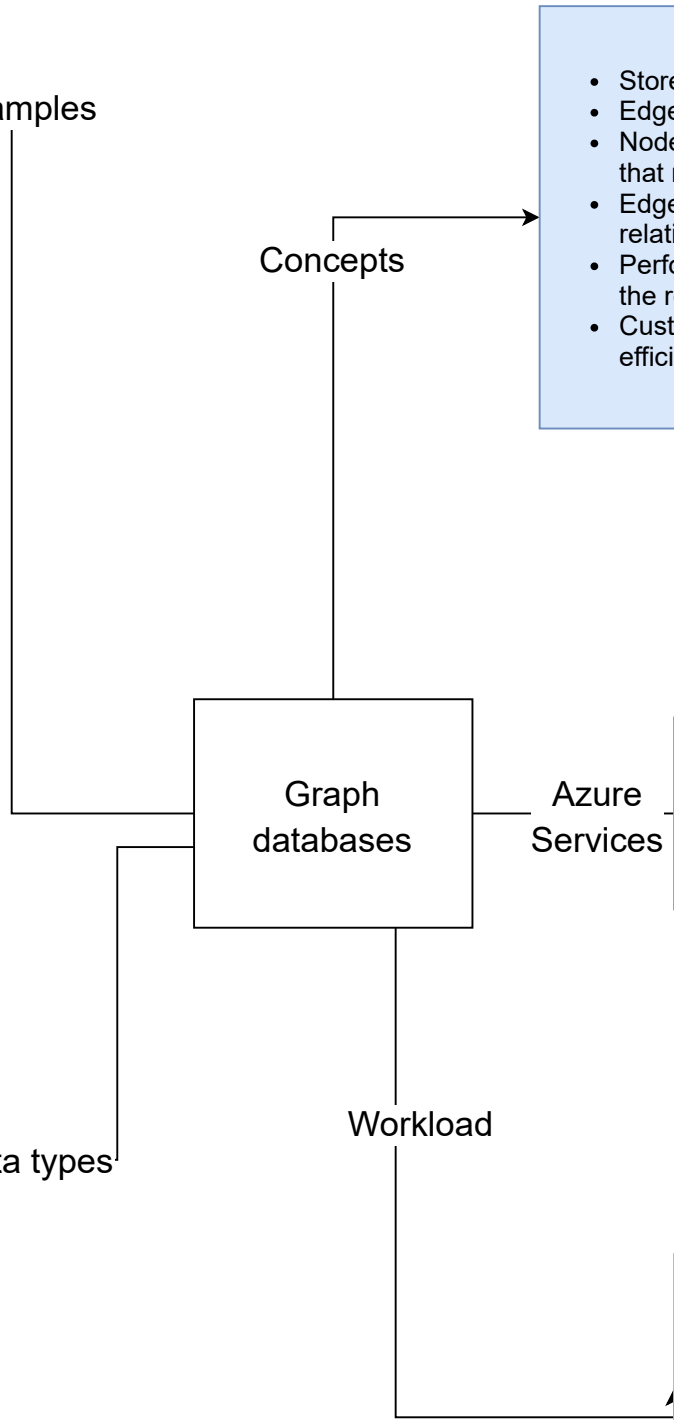
- Azure Cosmos DB Gremlin API
- SQL Server

Workload

- Complex relationships between data items involving many hops between related data items.
- The relationship between data items are dynamic and change over time.
- Relationships between objects are first-class citizens, without requiring foreign-keys and joins to traverse.

Examples

Data types



- Enterprise data warehouse

Examples

Concepts

- Data analytics stores provide massively parallel solutions for ingesting, storing, and analyzing data
- Data is distributed across multiple servers to maximize scalability
- Large data file formats such as delimiter files (CSV), parquet, and ORC
- Historical data stored in data stores (blob storage or Azure Data Lake Storage Gen2) and accessed by Azure Synapse, Databricks, or HDInsight as external tables



Compiled by Dr T Oberholster
Data store models

- [Website & Blog](#)
- [TeePublic Shirts](#)
- [RedBubble Merch](#)
- [Spreadshirt Apparel](#)
- [Contrado Fashion](#)

Data analytics

Azure Services

- Azure Synapse Analytics
- Azure Data Lake
- Azure Data Explorer
- Azure Analysis Services
- HDInsight
- Azure Databricks

- Historical data from multiple sources.
- Usually denormalized in a "star" or "snowflake" schema, consisting of fact and dimension tables.
- Usually loaded with new data on a scheduled basis.
- Dimension tables often include multiple historic versions of an entity, referred to as a *slowly changing dimension*.

Data types

Workload

- Data analytics
- Enterprise BI

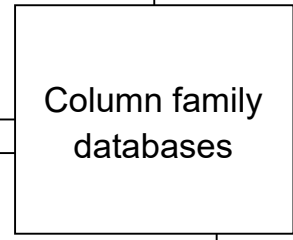
- Recommendations
- Personalization
- Sensor data
- Telemetry
- Messaging
- Social media analytics
- Web analytics
- Activity monitoring
- Weather and other time-series data



Compiled by Dr T Oberholster
Data store models

- [Website & Blog](#)
- [TeePublic Shirts](#)
- [RedBubble Merch](#)
- [Spreadshirt Apparel](#)
- [Contrado Fashion](#)

- Data is stored in tables consisting of a key column and one or more column families.
- Specific columns can vary by individual rows.
- Individual cells are accessed via get and put commands
- Multiple rows are returned using a scan command.



← Examples

Concepts →

Azure Services

Workload

← Data types

- Organizes data into rows and columns
- Appears similar to relational database
- Denormalized approach to structuring sparse data
- Holds tabular data with rows and columns
- Column families divide columns into groups
- Column family holds a set of columns that are logically related together and are retrieved/manipulated as a unit
- Within a column family, new columns can be added dynamically, and rows can be sparse (rows don't need a value for every column).
- Store structured, volatile data since rows for any given object in a column family can vary dynamically
- Store data in key order
- Indexes can be implemented over columns in a column family to retrieve data by columns value, rather than row key.
- Row read/write operations are atomic with a single column-family (some may have row atomicity)

- Azure Cosmos DB Cassandra API
- HBase in HDInsight

- Most column-family databases perform write operations extremely quickly.
- Update and delete operations are rare.
- Designed to provide high throughput and low-latency access.
- Supports easy query access to a particular set of fields within a much larger record.
- Massively scalable.

- Product catalogs
- Site search
- Logging



Compiled by Dr T Oberholster
Data store models

- [Website & Blog](#)
- [TeePublic Shirts](#)
- [RedBubble Merch](#)
- [Spreadshirt Apparel](#)
- [Contrado Fashion](#)

- Semi-structured or unstructured text
- Text with reference to structured data

Examples

Concepts

Search Engine Databases

Azure Services

- Azure Search

Workload

Data types

- Data indexes from multiple sources and services.
- Queries are ad-hoc and can be complex.
- Full text search is required.
- Ad hoc self-service query is required.

- Allows applications to search for information held in external data stores
- Index massive volumes of data
- Provide near real-time access to these indexes
- Indexes can be multi-dimensional
- Support free-text searches across text data
- Exact or fuzzy search (fuzzy search finds documents that match a set of terms and calculates how closely they match)
- May have linguistic support (synonyms, genre expansion and stemming)

- Monitoring and event telemetry.
- Sensor or other IoT data.

Examples

Concepts

- Values organized by time
- Large amounts of data in real time from a large number of sources
- Updates are rare
- Deletes done as bulk operations
- Total data size grows rapidly with large number of small records



Compiled by Dr T Oberholster
Data store models

- [Website & Blog](#)
- [TeePublic Shirts](#)
- [RedBubble Merch](#)
- [SpreadShirt Apparel](#)
- [Contrado Fashion](#)

Time Series Databases

Azure Services

- Azure Time Series

Workload

- A timestamp is used as the primary key and sorting mechanism.
- Tags may define additional information about the type, origin, and other information about the entry.

Data types

- Records are generally appended sequentially in time order.
- An overwhelming proportion of operations (95-99%) are writes.
- Updates are rare.
- Deletes occur in bulk, and are made to contiguous blocks or records.
- Data is read sequentially in either ascending or descending time order, often in parallel.

- Images, videos, office documents, PDFs
- Static HTML, JSON, CSS
- Log and audit files
- Database backups

Examples

Concepts

- Optimized for storing and retrieving large binary objects (images, files, video and audio streams, large application data objects and documents, virtual machine disk images, cav, parquet, orc).
- Manage extremely large amounts of unstructured data



Compiled by Dr T Oberholster
Data store models

- [Website & Blog](#)
- [TeePublic Shirts](#)
- [RedBubble Merch](#)
- [Spreadshirt Apparel](#)
- [Contrado Fashion](#)

Object storage

Azure Services

- Azure Blob Storage
- Azure Data Lake Storage Gen2

Workload

Data types

- Data size is large.
- Value is opaque.

- Identified by key.
- Content is typically an asset such as a delimiter, image, or video file.
- Content must be durable and external to any application tier.

- Legacy files
- Shared content accessible among a number of VMs or app instances



Compiled by Dr T Oberholster
Data store models

- [Website & Blog](#)
- [TeePublic Shirts](#)
- [RedBubble Merch](#)
- [SpreadShirt Apparel](#)
- [Contrado Fashion](#)

- Files in a hierarchical set of folders.
- Accessible with standard I/O libraries.

Concepts

- Optimized for storing and retrieving large binary objects (images, files, video and audio streams, large application data objects and documents, virtual machine disk images, cav, parquet, orc).
- Manage extremely
- Stores simple flat files
- Access across a network
- Highly scalable distributed services given appropriate security and concurrent access control
- Perform basic, low-level operations such as simple read and write requests.
amounts of unstructured data

Shared files

Azure Services

- Azure Files

Workload

- Migration from existing apps that interact with the file system.
- Requires SMB (server message block) interface.

Examples

Data types

