

0.1 Apache Ignite

Apache Ignite is an in-memory distributed database, caching, and processing platform for transactional, analytical, and streaming workloads, delivering in-memory speeds at petabyte scale [hid-sp18-501-ApacheIgnite].

Unlike in-memory databases, Apache Ignite works on top of existing databases and requires no rip-and-replace or any changes to an existing RDBMS. Users can keep their existing RDBMSs in place and deploy Apache Ignite as a layer above it. Apache Ignite can even automatically integrate with different RDBMS systems, such as Oracle, MySQL, Postgres, DB2, Microsoft SQL and others. This feature automatically generates the application domain model based on the schema definition of the underlying database and then loads the data. Moreover, IMDBs typically only provide a SQL interface while Apache Ignite provides a much wider ecosystem of supported access and processing paradigms in addition to ANSI SQL. Apache Ignite supports key/value stores, SQL access, MapReduce, HPC/MPP processing, streaming/CEP processing and Hadoop acceleration, all in one well-integrated in-memory data fabric [hid-sp18-501-GridGain].

0.2 Azure Blob Storage

Microsoft Azure BLOB storage service can be used to store and retrieve Binary Large Objects (BLOBs), or what are more commonly known as files [hid-sp18-501-RedGate]

This service can be used to share files with clients and to off-load some of the static content from web servers to reduce the load on them. Azure BLOB storage also provides persistent storage. By using this service, developers get dedicated virtual machines to run code without having to worry about managing those virtual machines. Azure BLOB Storage can store any type of file, such as Image files, database files, text files, or virtual hard drive files. However, when they are uploaded to the service they are stored as either a Page BLOB or a Block BLOB depending on how one plans on using the file or the size of the file. Page BLOBs are optimized for random reads and writes so they are most commonly used when storing virtual hard drive files for virtual machines. Each Page BLOB is made up of one or more 512-byte pages of data, up to a total size limit of 1 TB per file. The majority of files would benefit from being stored as Block BLOBs, which are written to the storage account as a series of blocks and then committed into a single file. One can create a large file by breaking it into blocks, which can be uploaded concurrently and then then committed together into a single file in one operation. This allows faster upload times and better throughput. The client storage libraries manage this process by uploading files of less than 64 MB in size in a single operation, and uploading larger files across multiple operations by breaking down the files and running the concurrent uploads. A Block BLOB has a maximum size of 200 GB [hid-sp18-501-RedGate]

0.3 Azure Cosmos DB

Azure Cosmos DB is a globally-distributed data service that allows elastically scaling throughput and storage across any number of geographical regions while guaranteeing low latency, high availability and consistency [hid-sp18-501-CosmosDB]. It can support multiple data models using one backend. This means that it can be used for document, key value, relational, and graph models. It is more or less a NoSQL database because it does not rely on any schemas. However, because it uses query language similar to SQL and can easily support ACID transactions, some people have been classifying it as a NewSQL type of database. What differentiates it from other NewSQL databases, however, is that it does not have a relational data model [hid-sp18-501-Stackify]

0.4 Google BigQuery

Google BigQuery is a cloud-based big data analytics web service for processing very large read-only data sets. BigQuery can analyze data on the order of billions of rows, using a SQL-like syntax. It runs on the Google Cloud Storage infrastructure and can be accessed with a REST-oriented application program interface (API) [**hid-sp18-501-TechTarget**].

BigQuery enables the creation of a logical data warehouse over managed, columnar storage as well as data from object storage, and spreadsheets. It also allows the capture and analysis of data in real-time using its powerful streaming ingestion [**hid-sp18-501-GoogleCP_BQ**].

0.5 Oracle Big Data Cloud Service

Oracle Big Data Cloud Service is an automated service that provides a high-powered environment tailor-made for advancing businesses' analytical capabilities. With automated lifecycle management and one-click security, Big Data Cloud Service is designed to optimally and securely run a wide variety of big data workloads and technologies while simplifying operations [**hid-sp18-501-OracleCloud**] Oracle Big Data Cloud Machine use connectors to seamlessly integrate with other Oracle services such as Oracle R Advanced Analytics for Hadoop (this enables the development of models using R that run in parallel on Hadoop, and accelerated using Spark), Oracle Data Integrator (this enables the transformation and enrichment of data within an enterprise big data clusters), Oracle SQL Connector for Hadoop and Oracle Loader for Hadoop (this allows the integration with data in online or offline modes), Big Data Spatial and Graph (this enables processing and enriching geospatial data), and Oracle R Advanced Analytics for Hadoop (this enables building models using R that run in parallel on Hadoop, and accelerated using Spark) [**hid-sp18-501-OracleCloud**]

References