A picture containing tableware, dishware

Description automatically generated



Mongo DB & GemFire

Brandon Rocha Díaz.

Software developer.

Shape, rectangle

Description automatically generated

**MONGODB**

MongoDB is document based NoSQL Database. Most widely used NoSQL Database in the world.

* Stores Data in JSON like documents. (BSON)
* Has very flexible structure to store the data
* Large amount of data can be easily managed

*SQL VS NOSQL*

|  |  |
| --- | --- |
| SQL | NoSQL |
| Relational database management system (rdbms) | Non-relational or distributed database system. |
| These databases have fixed or static or predefined schema | They have dynamic schema |
| These databases are not suited for hierarchical data storage. | These databases are best suited for hierarchical data storage. |
| These databases are best suited for complex queries | These databases are not so good for complex queries |
| Vertically Scalable | Horizontally scalable |
| Follows ACID property | Follows CAP(consistency, availability, partition tolerance) |
| Examples: MySQL, PostgreSQL, Oracle, MS-SQL Server etc | **Examples:**MongoDB, GraphQL, HBase, Neo4j, Cassandra etc |

*MONGODB TERMS*

|  |  |
| --- | --- |
| SQL | MONGO DB |
| Database | Database |
| Table | Collection |
| Column | Field |
| Row | Document |
| Foreign key | Sub Document (Reference) |

*WHEN TO USE SQL*

SQL is a good choice when working with related data. Relational databases are efficient, flexible, and easily accessed by any application. A benefit of a relational database is that when one user updates a specific record, every instance of the database automatically refreshes, and that information is provided in real-time.

SQL and a relational database make it easy to handle a great deal of information, scale as necessary and allow flexible access to data — only needing to update data once instead of changing multiple files, for instance. It’s also best for assessing data integrity. Since each piece of information is stored in a single place, there’s no problem with former versions confusing the picture.

Most of the big tech companies use SQL, which build their own database systems, use SQL to query and analyze data.

*WHEN TO USE NOSQL*

While SQL is valued for ensuring data validity, NoSQL is good when it’s more important that the availability of big data is fast. It’s also a good choice when a company will need to scale because of changing requirements.

NoSQL is also a good choice when there are large amounts of (or ever-changing) data sets or when working with flexible data models or needs that don't fit into a relational model. When working with large amounts of unstructured data, document databases are a good fit. For quick access to a key-value store without strong integrity guarantees, Redis may be the best choice. When a complex or flexible search across a lot of data is needed, Elastic Search is a good choice.

Scalability is a significant benefit of NoSQL databases. Unlike with SQL, their built-in sharding and high availability requirements allow horizontal scaling. Furthermore, NoSQL databases like Cassandra, developed by Facebook, handle massive amounts of data spread across many servers, having no single points of failure and providing maximum availability.

Other big companies that use NoSQL systems because they are dependent on large volumes of data not suited to a relational. In general, the more extensive the dataset, the more likely that NoSQL is a better choice.

**The *MongoTemplate***follows the standard template pattern in Spring and provides a ready-to-go, basic API to the underlying persistence engine.

**The repository** follows the Spring Data-centric approach and comes with more flexible and complex API operations, based on the well-known access patterns in all Spring Data projects.

For both, we need to start by defining the dependency — for example, in the *pom.xml*, with Maven:

<dependency>

<groupId>org.springframework.data</groupId>

<artifactId>spring-data-mongodb</artifactId>

<version>3.0.3.RELEASE</version>

</dependency>

*XML CONFIGURATION FOR MONGO TEMPLATE*

A simple configuration for mongo template:

<mongo:mongo-client id="mongoClient" host="localhost" />

<mongo:db-factory id="mongoDbFactory" dbname="test" mongo-client-ref="mongoClient" />

We first need to define the factory bean responsible for creating Mongo instances.

Then is needed to actually define (and configure) the template bean:

<bean id="mongoTemplate" class="org.springframework.data.mongodb.core.MongoTemplate"> <constructor-arg ref="mongoDbFactory"/> </bean>

And finally, we need to define a post processor to translate any MongoExceptions thrown in @Repository annotated classes:

<bean class= "org.springframework.dao.annotation.PersistenceExceptionTranslationPostProcessor"/>

*JAVA CONFIGURATION*

Let’s create a similar configuration using Java config by extending the base class for MongoDB configuration AbstractMongoConfiguration

@Configuration

public class MongoConfig extends AbstractMongoClientConfiguration {

@Override

protected String getDatabaseName() {

return "test"; }

@Override

public MongoClient mongoClient() {

ConnectionString connectionString = new ConnectionString("mongodb://localhost:27017/test"); MongoClientSettings mongoClientSettings = MongoClientSettings.builder() .applyConnectionString(connectionString) .build(); return MongoClients.create(mongoClientSettings);

}

@Override

public Collection getMappingBasePackages() { return Collections.singleton("com.baeldung"); } }

Note that we didn’t need to define MongoTemplate bean in the previous configuration since it’s already defined in AbstractMongoClientCondfiguration.

We also use our configuration from scratch without extending AbstractMongoClientConfiguration:

@Configuration

public class SimpleMongoConfig {

@Bean

public MongoClient mongo() {

ConnectionString connectionString = newConnectionString("mongodb://localhost:27017/test"); MongoClientSettings mongoClientSettings = MongoClientSettings.builder() .applyConnectionString(connectionString) .build();

return MongoClients.create(mongoClientSettings);

}

@Bean

public MongoTemplate mongoTemplate() throws Exception { return new MongoTemplate(mongo(), "test");

} }

*XML CONFIGURATION FOR MONGO REPOSITORY*

To make use of custom repositories (extending the MongoRepository), we need to continue the configuration from section and set up the repositories:

<mongo:repositories base-package="com.baeldung.repository" mongo-template-ref="mongoTemplate"/>

*JAVA CONFIGURATION*

Similarly, we’ll build on the configuration we already created, and add a new annotation into the mix:

@EnableMongoRepositories(basePackages = “com.baeldung.repository”)

*CREATE THE REPOSITORY*

After the configuration, we need to create a repository – extending the existing MongoRepository interface:

public interface UserRepository extends MongoRepository<User, String> { // }

Now we can auto-wire this UserRepository and use operations from MongoRepository or add custom operations.

*ANNOTATIONS*

Finally, let’s also go over the simple annotations that spring data uses to drive these API operations.

The field level @Id annotation can decorate any type, including long and string:

@Id

private String id;

If the value of the *@Id* field is not null, it's stored in the database as-is; otherwise, the converter will assume we want to store an *ObjectId* in the database (either*ObjectId*,*String*or*BigInteger*work).

We'll next look at *@Document*:

@Document

public class User { // }

This annotation simply marks a class as being a domain object that needs to be persisted to the database, along with allowing us to choose the name of the collection to be used.

**GEMFIRE**

GemFire is a high-performance distributed data management infrastructure that sits between application cluster and back end data sources.

With GemFire, data can be managed in-memory, which makes the access faster. Spring Data provides an easy configuration and access to GemFire from Spring application.

To make use of the Spring Data GemFire support, we first need to add the following dependency in the pom.xml:

<dependency>

<groupId>org.springframework.data</groupId>

<artifactId>spring-data-gemfire</artifactId>

<version>1.9.1.RELEASE</version>

</dependency>

*GEMFIRE BASICS FEATURES*

The cache GemFire provides the essential data management services as well as manages the connectivity to other peers.

The cache configuration (cache.xml) describes how data will be distributed among different nodes:

<cache>

<region name="region"><region-attributes>

<cache-listener>

<class-name> ... </class-name>

</cache-listener> </region-attributes>

</region> ...

</cache>

*PIVOTAL IS A GEMFIRE*

Pivotal GemFire is a data management platform that provides real-time, consistent access to data intensive applications through widely distributed cloud architectures.

GemFire pools memory, CPU, network resources, and optionally local disk across multiple processes to manage application objects and behavior. It uses dynamic replication and data partitioning techniques to implement high availability, improved performance, scalability, and fault tolerance. In addition to being a distributed data container.

*MAIN CONCEPTS AND COMPONENTS*

Caches are an abstraction that describe a node in a GemFire distributed system. Application architects can arrange these nodes in peer-to-peer or client/server topologies.

Within each cache, you define data regions.

GemFire locators provide both member discovery and load-balancing services. You configure clients with a list of locator services and the locators maintain a dynamic list of member servers.

GemFire uses continuous querying to enable event-driven architectures. GemFire ties events and data together so that when an event is processed, the data required to process the event is available without additional queries to a disk-based database. Clients can subscribe to change notifications so that they can execute tasks when a specific piece of data changes.

In addition to peer-to-peer and client/server topologies, GemFire supports multi-site configurations that allow you to scale horizontally between disparate, loosely-coupled distributed systems over a wide geographically separated network. A wide-area network (WAN) is the main use case for the multi-site topology.

*REGIONS*

Data regions are a logical grouping within a cache for a single data set.

Simply put, a region lets us store data in multiple VMs in the system without consideration to which node the data is stored within the cluster.

Data regions are analogous to tables in a relational database and manage data in a distributed fashion as name/value pairs. A replicated region stores identical copies of the data on each cache member of a distributed system. A partitioned region spreads the data among cache members. After the system is configured, client applications can access the distributed data in regions without knowledge of the underlying system architecture. You can define listeners to create notifications about when data has changed, and you can define expiration criteria to delete obsolete data from a region.

Regions are classified into three broad categories:

* **Replicated region** holds the complete set of data on each node. It gives a high read performance. Write operations are slower as the data update need to be propagated to each node:

<region name="myRegion" refid="REPLICATE"/>

* **Partitioned region** distributes the data so that each node only stores a part of region contents. A copy of the data is stored on one of the other nodes. It provides a good write performance.

<region name="myRegion" refid="PARTITION"/>

* **Local region** resides on the defining member node. There is no connectivity with other nodes within the cluster.

<region name="myRegion" refid="LOCAL"/>

GemFire provides a query language called OQL (Object Query Language) that allows us to refer the objects stored in GemFire data regions. This is very similar to SQL in syntax.

A basic query can be such as: SELECT DISTINCT \* FROM exampleRegion

GemFire’s QueryService provides methods to create the query object.

To manage the data serialization-deserialization, GemFire provides options other than Java serialization that gives a higher performance, provides greater flexibility for data storage and data transfer, also support for different languages.

With that in mind, GemFire has defined Portable Data eXchange(PDX) data format. PDX is a cross-language data format that provides a faster serialization and deserialization, by storing the data in the named field which can be accessed directly without the need of fully deserialization the object.

In GemFire, a function can reside on a server and can be invoked from a client application or another server without the need to send the code itself.

The caller can direct data-dependent function to operate on a particular data set or can lead an independent data function to work on a particular server, member or member group.

With continuous querying, the clients subscribe to server-side events by using SQL-type query filtering. The server sends all the events that modify the query results. The continuous querying event delivery uses the client/server subscription framework.

The syntax for a continuous query is similar to basic queries written in OQL. For example, a query which provides the latest stock data from Stock region can be written as:

SELECT \* from StockRegion s where s.stockStatus = ‘active’;

To get the status update from this query, an implementation of C QListener need to be attached with the StockRegion:

<cache>

<region name="StockRegion>

<region-attributes refid="REPLICATE"> ... <cache-listener>

<class-name>...</class-name> </cache-listener> ... </region-attributes>

</region>

</cache>

*SPRING DATA GEMFIRE SUPPORT*

To simplify configuration, spring data GemFire provides various annotations for configuring core GemFire components:

@Configuration

public class GemfireConfiguration {

@Bean

Properties gemfireProperties() {

Properties gemfireProperties = new Properties(); gemfireProperties.setProperty("name","SpringDataGemFireApplication"); gemfireProperties.setProperty("mcast-port", "0");

gemfireProperties.setProperty("log-level", "config");

return gemfireProperties; }

@Bean

CacheFactoryBean gemfireCache() {

CacheFactoryBean gemfireCache = new CacheFactoryBean(); gemfireCache.setClose(true);

gemfireCache.setProperties(gemfireProperties()); return gemfireCache;

}

@Bean(name="employee")

LocalRegionFactoryBean<String, Employee> getEmployee(final GemFireCache cache) { LocalRegionFactoryBean<String, Employee> employeeRegion = new LocalRegionFactoryBean();

employeeRegion.setCache(cache); employeeRegion.setName("employee");

// ... return employeeRegion;

}

}

To set up the GemFire cache and region, we have to first setup few specific properties. Here mcast-port is set to zero, which indicates that this GemFire node is disabled for multicast discovery and distribution. These properties are then passed to CacheFactoryBean to create a GemFireCache instance.

Using GemFireCache bean, an instance of LocalRegionFatcoryBean is created which represents the region within the Cache for the Employee instances.

The library provides support to map objects to be stored in GemFire grid. The mapping metadata is defined by using annotations at the domain classes. We can use:

* @Region, to specify the region instance of the x class
* @Id, to annotate the properly that shall be utilized as a cache key
* @PersistenceConstructor, which helps to mark the one that will be used to create entities, in case multiple constructors available

The repositories allow the definition of query methods to efficiently run the OQL queries against the region the managed entity is mapped to.

It also have annotation support available – to simplify working with GemFire function execution.

There are two concerns to address when we make use of functions, the implementation, and the execution.

A POJO can be exposed as a GemFire function using Spring Data annotations like:

@Component

public class FunctionImpl {

@GemfireFunction

public void greeting(String message){

// some logic }

// ... }

But it’s needed to be activated annotation processing explicitly for @GemfireFunction to work as:

@Configuration

@EnableGemfireFunctions

public class GemfireConfiguration { // ... }

For function execution, a process invoking a remote function need to provide calling arguments, a function id, the execution target (onServer, onRegion, onMember, etc.):

@OnRegion(region="employee")

public interface FunctionExecution {

@FunctionId("greeting

public void execute(String message); // ... }

*ADVANTAGES OF USING SPRING OVER PIVOTAL GEMFIRE cache.xml*

Spring Data for Pivotal GemFire’s XML namespace supports full configuration of the Pivotal GemFire In-Memory Data Grid (IMDG). The XML namespace is one of two ways to configure Pivotal GemFire in a Spring context in order to properly manage Pivotal GemFire’s lifecycle inside the Spring container. The other way to configure Pivotal GemFire in a Spring context is by using [annotation-based configuration](https://docs.spring.io/spring-gemfire/docs/current/reference/html/#bootstrap-annotation-config).

While support for Pivotal GemFire’s native cache.xml persists for legacy reasons, Pivotal GemFire application developers who use XML configuration are encouraged to do everything in Spring XML to take advantage of the many wonderful things Spring has to offer, such as modular XML configuration, property placeholders and overrides, SpEL ([Spring Expression Language](https://docs.spring.io/spring/docs/current/spring-framework-reference/core.html#expressions)), and environment profiles. Behind the XML namespace, Spring Data for Pivotal GemFire makes extensive use of Spring’s FactoryBean pattern to simplify the creation, configuration, and initialization of Pivotal GemFire components.

Pivotal GemFire provides several callback interfaces, such as CacheListener, CacheLoader, and CacheWriter, that let developers add custom event handlers. Using Spring’s IoC container, you can configure these callbacks as normal Spring beans and inject them into Pivotal GemFire components. This is a significant improvement over native cache.xml, which provides relatively limited configuration options and requires callbacks to implement Pivotal GemFire’s Declarable interface.

In addition, IDEs, such as the Spring Tool Suite (STS), provide excellent support for Spring XML namespaces, including code completion, pop-up annotations, and real time validation.

To simplify configuration, Spring Data for Pivotal GemFire provides a dedicated XML namespace for configuring core Pivotal GemFire components. It is possible to configure beans directly by using Spring’s standard <bean> definition. However, all bean properties are exposed through the XML namespace, so there is little benefit to using raw bean definitions.

To use the Spring Data for Pivotal GemFire XML namespace, declare it in your Spring XML configuration meta-data, as the following example shows:

**<?xml version="1.0" encoding="UTF-8"?>**

<beans xmlns="http://www.springframework.org/schema/beans"

xmlns:gfe="https://www.springframework.org/schema/gemfire" (1)(2)

xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"

xsi:schemaLocation="

http://www.springframework.org/schema/beans https://www.springframework.org/schema/beans/spring-beans.xsd

https://www.springframework.org/schema/gemfire https://www.springframework.org/schema/gemfire/spring-gemfire.xsd (3)

">

<bean id ... >

<gfe:cache ...> (4)

</beans>

In this example:

1. Spring Data for Pivotal GemFire XML namespace prefix. Any name works, but, through this reference documentation, gfe is used.
2. The XML namespace prefix is mapped to the URI.
3. The XML namespace URI location. Note that, even though the location points to an external address (which does exist and is valid), spring resolves the schema locally, as it is included in the Spring Data for Pivotal GemFire library.
4. Example declaration using the XML namespace with the gfe prefix.

To see more information about GemFire details visit:

[Spring Data for Pivotal GemFire Reference Guide](https://docs.spring.io/spring-gemfire/docs/current/reference/html/#data-access)