**Micro Services**

1. **What is Micro Services Architecture?**

Micro services architecture allows to avoid monolith application for large system. It provide loose coupling between collaborating processes which running independently in different environments with tight cohesion.

For example imagine an online shop with separate microservices for user-accounts, product-catalog order-processing and shopping carts.

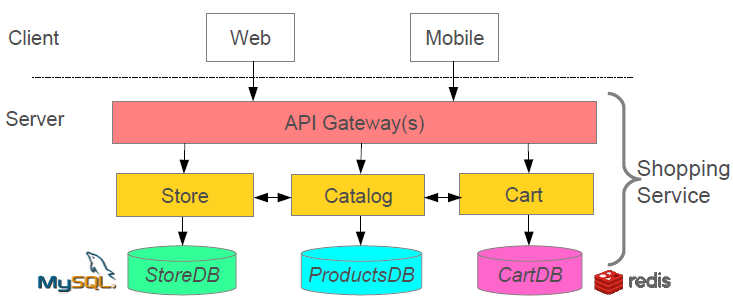
* **Shopping system without Micro services (Monolith architecture)**

In this architecture we are using Monolith architecture i.e. all collaborating components combine all in one application.



* **Shopping system with Micro services**

In this architecture style the main application divided in a set of sub applications called micro services. One large Application divided into multiple collaborating processes as below.



1. **Micro Services Benefits and Challenges?**

**Micro Services Benefits**

* Smaller code base is easy to maintain.
* Easy to scale as individual component.
* Technology diversity i.e. we can mix libraries, databases, frameworks etc.
* Fault isolation i.e. a process failure should not bring whole system down.
* Better support for smaller and parallel team.
* Independent deployment.
* Deployment reduce time.

**Micro Services Challenges**

* Difficult to achieve strong consistency across services
* ACID transactions do not span multiple processes.
* Distributed System so hard to debug and trace the issues
* Greater need for end to end testing
* Required cultural changes in across teams like Dev and Ops working together even in same team.

1. **Service Discovery / Registry**

**Implementing Service Discovery**

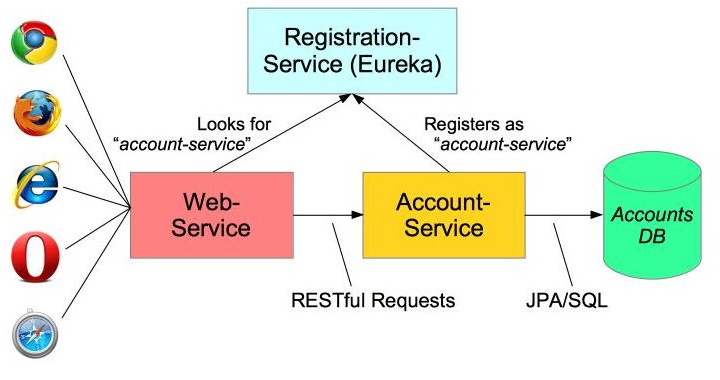
Use Eureka created by Netflix

**Client-side Load Balancing**

Each service typically deployed as multiple instances for fault tolerance and load sharing.

**Implementing Client-Side Load Balancing**

We will use Netflix Ribbon, it provide several algorithm for Client-Side Load Balancing. Spring provide smart **RestTemplate**for service discovery and load balancing by using ***@LoadBalanced***annotation with **RestTemplate**instance.

****

1. **Developing Simple Micro services Example**

**For build a simple micro services system following steps required**

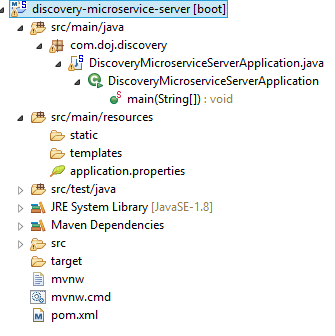
1. Creating Discovery Service (Creating Eureka Discovery Service)
2. Creating Micro Service (the Producer)
   1. Register itself with Discovery Service with logical service.
3. Create Micro service Consumers find Service registered with Discovery Service
   1. Discovery client using a smart **RestTemplate**to find micro service.

**Maven Dependencies**

1. **<dependencies>**
2. **<dependency>**
3. **<groupId>**org.springframework.cloud**</groupId>**
4. **<artifactId>**spring-cloud-starter**</artifactId>**
5. **</dependency>**
6. **<dependency>**
7. **<groupId>**org.springframework.cloud**</groupId>**
8. **<artifactId>**spring-cloud-starter-eureka**</artifactId>**
9. **</dependency>**
10. **<dependency>**
11. **<groupId>**org.springframework.boot**</groupId>**
12. **<artifactId>**spring-boot-starter-web**</artifactId>**
13. **</dependency>**
14. **<dependency>**
15. **<groupId>**org.hsqldb**</groupId>**
16. **<artifactId>**hsqldb**</artifactId>**
17. **<scope>**runtime**</scope>**
18. **</dependency>**
19. **<dependency>**
20. **<groupId>**org.springframework.boot**</groupId>**
21. **<artifactId>**spring-boot-starter-test**</artifactId>**
22. **<scope>**test**</scope>**
23. **</dependency>**
24. **</dependencies>**

**Step 1: Creating Discovery Service (Creating Eureka Discovery Service)**

* Eureka Server using Spring Cloud
* We need to implement our own registry service as below.



**application.yml**

1. # Configure this Discovery Server
2. eureka:
3. instance:
4. hostname: localhost
5. client: #Not a client
6. registerWithEureka: **false**
7. fetchRegistry: **false**
8. # HTTP (Tomcat) port
9. server:
10. port: 8761

**or application.properties**

**eureka.client.register-with-eureka=false** [By Default , the registry will also attempt to register itself , so we’ll need to disable that.]

**eureka.client.fetch-registry=false**

**server.port=8761**

**DiscoveryMicroserviceServerApplication.java**

1. **package** com.doj.discovery;
2. **import** org.springframework.boot.SpringApplication;
3. **import** org.springframework.boot.autoconfigure.SpringBootApplication;
4. **import** org.springframework.cloud.netflix.eureka.server.EnableEurekaServer;
5. @SpringBootApplication
6. @EnableEurekaServer
7. **public** **class** DiscoveryMicroserviceServerApplication {
8. **public** **static** **void** main(String[] args) {
9. SpringApplication.run(DiscoveryMicroserviceServerApplication.**class**, args);
10. }
11. }

Spring Cloud’s @EnableEurekaServer to standup a registry that other applications can talk to. This is a regular Spring Boot application with one annotation added to enable the service registry.

**pom.xml**

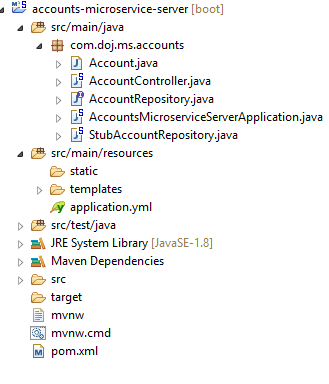
1. <!-- Eureka registration server -->
2. **<dependency>**
3. **<groupId>**org.springframework.cloud**</groupId>**
4. **<artifactId>**spring-cloud-starter-eureka-server**</artifactId>**
5. **</dependency>**

**Step 2: Creating Account Producer Micro Service**

Micro service declares itself as an available service and register to Discovery Server created in **Step 1**.

* Using ***@EnableDiscoveryClient***
* Registers using its application name

Let’s see the service producer application structure as below.



**application.yml**

1. ### Spring properties
2. # Service registers under this name
3. spring:
4. application:
5. name: accounts-microservice
6. # Discovery Server Access
7. eureka:
8. client:
9. serviceUrl:
10. defaultZone: http://localhost:8761/eureka/
11. # HTTP Server (Tomcat) Port
12. server:
13. port: 7777
14. # Disable Spring Boot's "Whitelabel" default error page, so we can use our own
15. error:
16. whitelabel:
17. enabled: **false**

or **application.properties**

server.port=7777

spring.application.name=accounts-microservice

#Eurkeka

eureka.client.serviceUrl.defaultZone= http://localhost:8761/eureka/

eureka.instance.preferIpAddress: false

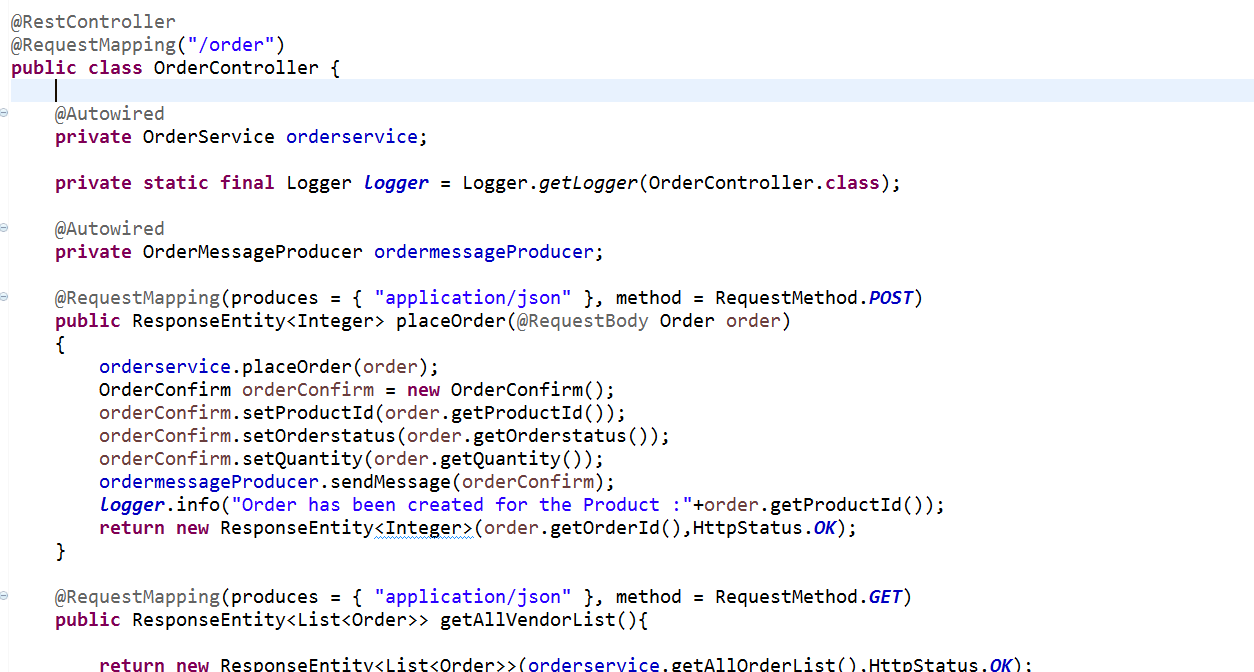
**AccountsMicroserviceServerApplication.java**

1. **package** com.doj.ms.accounts;
2. **import** org.springframework.boot.SpringApplication;
3. **import** org.springframework.boot.autoconfigure.SpringBootApplication;
4. **import** org.springframework.cloud.client.discovery.EnableDiscoveryClient;
5. @SpringBootApplication
6. @EnableDiscoveryClient
7. **public** **class** AccountsMicroserviceServerApplication {
8. **public** **static** **void** main(String[] args) {
9. SpringApplication.run(AccountsMicroserviceServerApplication.**class**, args);
10. }
12. }

The @EnableDiscoveryClient activates the Netflix Eureka DiscoveryClient implementation.

**pom.xml**

1. **<dependencies>**
2. **<dependency>**
3. **<groupId>**org.springframework.cloud**</groupId>**
4. **<artifactId>**spring-cloud-starter**</artifactId>**
5. **</dependency>**
6. **<dependency>**
7. **<groupId>**org.springframework.cloud**</groupId>**
8. **<artifactId>**spring-cloud-starter-eureka**</artifactId>**
9. **</dependency>**
10. **<dependency>**
11. **<groupId>**org.springframework.boot**</groupId>**
12. **<artifactId>**spring-boot-starter-web**</artifactId>**
13. **</dependency>**
14. **</dependencies>**

****

**Using the @RestController Annotation**

Spring 4.0 introduced **@RestController**, a specialized version of the controller which is a convenience annotation that does nothing more than add the **@Controller** and **@ResponseBody** annotations. By annotating the controller class with **@RestController** annotation, you no longer need to add **@ResponseBody** to all the request mapping methods. The **@ResponseBody** annotation is active by default.

1. **What is Spring Boot?**

First of all Spring Boot is not a framework, it is a way to ease to create stand-alone application with minimal or zero configurations. It is approach to develop spring based application with very less configuration. It provides defaults for code and annotation configuration to quick start new spring projects within no time.

1. **How does it work? How does it know what to configure?**

Auto-configuration works by analyzing the classpath  
– If you forget a dependency, Spring Boot can't configure it  
– A dependency management tool is recommended  
– Spring Boot parent and starters make it much easier  
• Spring Boot works with Maven, Gradle, Ant/Ivy

1. **How are properties defined? Where?**

In spring boot, we have to define properties in the application.properties file exists in classpath of application as follow.  
Example: configure default DataSource bean

1. database.host=localhost
2. database.user=admin
3. **What embedded containers does Spring Boot support?**

Spring Boot includes support for embedded Tomcat, Jetty, and Undertow servers.  
By default the embedded server will listen for HTTP requests on port 8080.

1. **What does @EnableAutoConfiguration do? What about @SpringBootApplication?**

**@EnableAutoConfiguration annotation** on a Spring Java configuration class

– Causes Spring Boot to automatically create beans it thinks you need  
– Usually based on classpath contents, can easily override

1. @Configuration
2. @EnableAutoConfiguration
3. **public** **class** MyAppConfig {
4. **public** **static** **void** main(String[] args) {
5. SpringApplication.run(MyAppConfig.**class**, args);
6. }
7. }

**@SpringBootApplication**was available from Spring Boot 1.2  
It is very common to use @EnableAutoConfiguration, @Configuration, and @ComponentScan together.

1. @Configuration
2. @ComponentScan
3. @EnableAutoConfiguration
4. **public** **class** MyAppConfig {
5. ...
6. }

**With @SpringBootApplication annotation**

1. @SpringBootApplication
2. **public** **class** MyAppConfig {
3. ...
4. }
5. **Can you control logging with Spring Boot? How?**

Yes, we can control logging with spring boot.  
 **Customizing default Configuration for Logging:**  
By adding **logback.xml** file to the application we can override the default logging configuration providing by the Spring Boot. This file place in the classpath (src/main/resources) of the application for Spring Boot to pick the custom configuration.  
  
**Spring Boot can control the logging level**  
– Just set it in application.properties  
• Works with most logging frameworks  
– Java Util Logging, Logback, Log4J, Log4J2

1. logging.level.org.springframework=DEBUG
2. logging.level.com.acme.your.code=INFO
3. **How to reload my changes on Spring Boot without having to restart server?**

Include following maven dependency in the application.

1. **<dependency>**
2. **<groupId>**org.springframework**</groupId>**
3. **<artifactId>**springloaded**</artifactId>**
4. **<version>**1.2.6.RELEASE**</version>**
5. **</dependency>**

**Automatic restart**

Applications that use spring-boot-devtools will automatically restart whenever files on the classpath change. This can be a useful feature when working in an IDE as it gives a very fast feedback loop for code changes. By default, any entry on the classpath that points to a folder will be monitored for changes.

1. **<dependency>**
2. **<groupId>**org.springframework.boot**</groupId>**
3. **<artifactId>**spring-boot-devtools**</artifactId>**
4. **<optional>**true**</optional>**
5. **</dependency>**

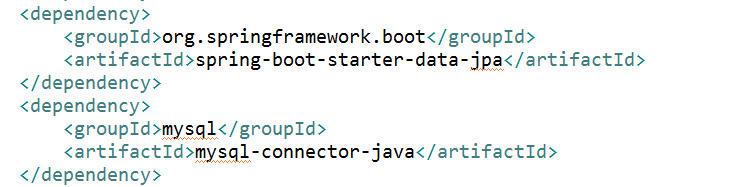
This can be achieved using DEV Tools. With this dependency any changes you save, the embedded tomcat will restart. Spring Boot has a Developer tools (DevTools) module which helps to improve the productivity of developers. One of the key challenge for the Java developers is to auto deploy the file changes to server and auto restart the server. Developers can reload changes on Spring Boot without having to restart my server. This will eliminates the need for manually deploying the changes every time. Spring Boot doesn’t have this feature when it has released it’s first version. This was a most requested features for the developers. The module DevTools does exactly what is needed for the developers. This module will be disabled in the production environment.

1. **How to implement security for spring boot application?**

Add spring security starter to the boot application

1. **<dependency>**
2. **<groupId>**org.springframework.boot**</groupId>**
3. **<artifactId>**spring-boot-starter-security**</artifactId>**
4. **</dependency>**
5. **How to configure datasource using Spring boot?**

* Use either spring-boot-starter-jdbc or **spring-boot-starterdata-jpa** and include a JDBC driver on classpath



• Declare properties in the application.properties

1. spring.datasource.url=jdbc:mysql://localhost/test
2. spring.datasource.username=dbuser
3. spring.datasource.password=dbpass
4. spring.datasource.driver-**class**-name=com.mysql.jdbc.Driver

**Spring Cache**

**Enable Caching**

1. **@EnableCaching** to enable caching because caching in spring is not enabled by default. The caching feature can be declaratively enabled by simply adding the **@EnableCaching** annotation to any of the configuration classes.
2. @SpringBootApplication
3. @EnableCaching
4. **public** **class** SpringBootCacheApplication {
5. **public** **static** **void** main(String[] args) {
6. SpringApplication.run(SpringBootCacheApplication.**class**, args);
7. }
8. }

In the XML we can you use following tag for enable caching.

1. **<cache:annotation-driven** **/>**

**Use Caching With Annotations**

After enabling the cache we can use following list of declarative annotations.

* @Cacheable
* @CacheEvict
* @CachePut
* @Caching
* @CacheConfig

**@Cacheable:**It is one of the most important and common annotation for caching the requests. If you annotate a method with @Cacheable, if multiple requests are received by the application, then this annotation will not execute the method multiple times, instead it will send the result from the cached storage.

The simplest way to enable caching behavior for a method is to demarcate it with @Cacheable and parameterize it with the name of the cache where the results would be stored:

1. @Cacheable(value="cities")
2. **public** List<City> findAllCity(){
3. **return** (List<City>) cityRepository.findAll();
4. }

The findAllCity() call will first checks the cache cities before actually invoking the method and then caching the result.

Spring framework also supports multiple caches to be passed as parameters:

1. @Cacheable(value={"cities","city-list"})
2. **public** List<City> findAllCity(){
3. **return** (List<City>) cityRepository.findAll();
4. }

**@CacheEvict:**

If we annotate all methods with**@Cacheable** then the size of cache may be some problem. **We don’t want to populate the cache with values that we don’t need often.** Caches can grow quite large, quite fast, and we could be holding on to a lot of stale or unused data. **@CacheEvict**annotation is used for removing a single cache or clearing the entire cache from the cache storage so that fresh values can be loaded into the cache again:

1. @CacheEvict(value="cities", allEntries=**true**)
2. **public** List<City> findAllCity(){
3. **return** (List<City>) cityRepository.findAll();
4. }

Here **allEntries** indicated whether all the data in the cache has to be removed.

**@CachePut:**

@CachePut annotation helps for updating the cache with the latest execution without stopping the method execution. The difference between @Cacheable and @CachePut is that @Cacheable will skip running the method, whereas @CachePut will actually run the method and then put its results in the cache.

1. @CachePut(value="cities")
2. **public** List<City> findAllCity(){
3. **return** (List<City>) cityRepository.findAll();
4. }

**@Caching:**

What if you want to use multiple annotations of the same type for caching a method? @Caching annotation used for grouping multiple annotations of the same type together when one annotation is not sufficient for the specifying the suitable condition. For example, you can put multiple @CacheEvict or @CachePut annotation inside @Caching to narrow down your conditions as you need.

1. @Caching(evict = {
2. @CacheEvict("cities"),
3. @CacheEvict(value="city-list", key="#city.name") })
4. **public** List<City> findAllCity(){
5. **return** (List<City>) cityRepository.findAll();
6. }

**@CacheConfig:**

You can annotate @CacheConfig at the class level to avoid repeated mentioning in each method. For example, in the class level you can provide the cache name and in the method you just annotate with @Cacheable annotation.

1. @CacheConfig(cacheNames={"cities"})
2. **public** **class** CityMasterService {
3. @Cacheable
4. **public** List<City>) findAllCity() {
5. **return** (List<City>)) cityRepository.findAll();
6. }
7. }

**Conditional Caching**

Sometimes we don’t want to cache some result there are no need to cache. For example – reusing our example from the @CachePut annotation – this will both execute the method as well as cache the results each and every time:

1. @CachePut(value="cities")
2. **public** List<City>) findAllCity(State state){
3. **return** (List<City>)) cityRepository.findAll(state.getStateCode());
4. }

**Condition Parameter:**

1. @CachePut(value="cities", condition="#state.stateName=='UP'")
2. **public** ListList<City>) findAllCity(State state){
3. **return** (List<City>)) cityRepository.findAll(state.getStateCode());
4. }

@CachePut can be parameterized with a condition parameter that takes a SpEL expression to ensure that the results are cached based on evaluating that expression.

**Unless Parameter:**

We can also control the caching based on the output of the method rather than the input – via the unless parameter

1. @CachePut(value="cities", **unless**="#result.length()>25")
2. **public** List<City>) findAllCity(String state){
3. **return** (List<City>)) cityRepository.findAll(state.getStateCode());
4. }

The above annotation would cache only those cities of state that have minimum 25 cities in the state.

**Agile**

* 1. **What are the differences between Agile and Traditional Project management?**

Agile encourages that a little of everything, including design, development and testing, is done at the same time, as opposed to the traditional approach to projects, where one phase is closed and completed before the next begins. Agile encourages short, frequent feedback loops and embraces changes to requirements. In [Waterfall](https://en.wikipedia.org/wiki/Waterfall_model), feedback is usually not collected until the very end of the project and changes are discouraged.

* 1. **What are the roles in Scrum?**

Scrum only prescribes 3 roles - **the Product Owner, Scrum Master and the Delivery team**. These roles should, ideally, be cross-functional and not shared among other projects.

Most Scrum Masters have not had the opportunity to work with a team that was cross-functional or dedicated due to the organization’s resistance or inability to allow for what some refer to as a “luxury”. This question sometimes leads them to ask how you would handle working with a team that did not have a designer or tester within the team or how you would handle a team that was not dedicated.  Be ready!

* 1. **What is the Daily Stand-Up?**

Every day, preferably in the morning, the team meets for no more than 15 minutes to answer three questions:

* + 1. What did you do yesterday?
    2. What do you plan on doing today?
    3. Are there any blocks or impediments that keep you from doing your work?

This Scrum ceremony is not meant to be a status meeting for stakeholders, but a way to energize the team and get them to set focus for the day.

* 1. **Describe what happens in the Sprint planning meeting.**

In Sprint planning, the Product Owner presents the goal of the sprint and discusses the high priority product backlog items. The Delivery team then chooses the amount of work for the next sprint.

* 1. **What is the role of the Scrum Master?**

A Scrum Master serves the team and shields them from any distractions that would prevent them from completing a sprint goal. They also remove blocks, teach the team to become self-organized, and serve as coach who teaches Agile and Scrum values and principles.

* 1. **Is there a difference between Agile and Scrum?**

Yes! Agile is the broader umbrella which Scrum falls under. [Agile has four main values and twelve principles](https://www.smartsheet.com/comprehensive-guide-values-principles-agile-manifesto). Scrum has its own set of values and principles and provides a lightweight “framework” to help teams becomeAgile.

* 1. **What are the four Main values and twelve principles of Agile?**

**The Four values of the Agile Manifesto**

* + 1. **Individuals and Interactions Over Processes and Tools**
    2. **Working Software Over Comprehensive Documentation**
       - Agile documents requirements as user stories, which are sufficient for a software developer to begin the task of building a new function.  
         The Agile Manifesto values documentation, but it values working software more.
    3. **Customer Collaboration Over Contract Negotiation**
       - Negotiation is the period when the customer and the product manager work out the details of a delivery, with points along the way where the details may be renegotiated. Collaboration is a different creature entirely. With development models such as Waterfall, customers negotiate the requirements for the product, often in great detail, prior to any work starting. This meant the customer was involved in the process of development before development began and after it was completed, but not during the process. The Agile Manifesto describes a customer who is engaged and collaborates throughout the development process, making. This makes it far easier for development to meet their needs of the customer. Agile methods may include the customer at intervals for periodic demos, but a project could just as easily have an end-user as a daily part of the team and attending all meetings, ensuring the product meets the business needs of the customer.
    4. **Responding to Change Over Following a Plan**
       - Traditional software development regarded change as an expense, so it was to be avoided. The intention was to develop detailed, elaborate plans, with a defined set of features and with everything, generally, having as high a priority as everything else, and with a large number of many dependencies on delivering in a certain order so that the team can work on the next piece of the puzzle.
       - With Agile, the shortness of an iteration means priorities can be shifted from iteration to iteration and new features can be added into the next iteration. Agile’s view is that changes always improve a project; changes provide additional value.

**The Twelve Agile Manifesto Principles**

1. **Customer satisfaction through early and continuous software delivery.**
2. **Accommodate changing requirements throughout the development process.**
3. **Frequent delivery of working software.**
4. **Collaboration between the business stakeholders and developers throughout the project.**
5. **Support, trust, and motivate the people involved.**
6. **Enable face-to-face interactions.**
7. **Working software is the primary measure of progress.**
8. **Agile processes to support a consistent development pace.**
9. **Attention to technical detail and design enhances agility.**
10. **Simplicity –**Develop just enough to get the job done for right now.
11. **Self-organizing teams encourage great architectures, requirements, and designs.**
12. **Regular reflections on how to become more effective.**
    1. **What is velocity?**

Velocity is the average number of points from that past 3-4 sprints and is used to help predict when backlog items will be delivered.

* 1. **What type of metrics or reports have you used?**

Sprint and release burn-down and burn-up charts are standard reports.

* 1. **What is a burn-down chart?**

A burn down chart displays the amount of work a team has burned through such as hours during the sprint.

A Scrum burn down chart should consist of

* + - * X-axis that displays working days
      * Y-axis that displays remaining effort
      * Ideal effort as guideline
      * Real Progress of effort
  1. **How many Scrum teams have you managed at one time?**

This is a popular question! Don’t offer that Scrum guidelines states only one Scrum Master per team as your answer! In this new role, you may be required to lead more than one team. Notice the use of the word “managed” versus “led” as Scrum Masters do not manage, they lead teams so be sure to use this word in your response.

* 1. **What type of requirements did you use for your teams?**

Requirements in Scrum are written as user stories using a standard, “As a , I want… so that I .”  As a Scrum Master, you don’t necessarily write user stories, but you would assist the Product Owner to ensure that user stories are written, prioritized and ready for the sprint.

* 1. **How much time should a person expect to spend on ScrumMaster activities?**

A ScrumMaster should make this role their top priority to focus on benefits of the overall team. Their load will vary from sprint to sprint depending on what impediments and issues the team is dealing with. Newly formed teams typically take more ScrumMaster time; 50%-100%, while experienced ScrumMasters with established well functioning teams might spend 50% or less time on the ScrumMaster role.

* 1. **What is difference between Epic, User stories & Tasks?**

**Epic** is a group of related user stories.  
**User Stories** define the actual business requirement. Generally created by the business owner.  
**Task:** To accomplish the business requirements, development team create tasks.

* 1. **How study Board can be defined in agile?**

A Story Board is a visual representation of a software project’s progress. There are generally four columns ‘To do’, In Progress’, ‘Test’, and ‘Done’. Different colored post, its notes are placed in each column indicating the progress of individual development items. A story board is typically used in agile development.

* 1. **How much time should a person expect to spend on ScrumMaster activities?**

A ScrumMaster should make this role their top priority to focus on benefits of the overall team. Their load will vary from sprint to sprint depending on what impediments and issues the team is dealing with. Newly formed teams typically take more ScrumMaster time; 50%-100%, while experienced ScrumMasters with established well functioning teams might spend 50% or less time on the ScrumMaster role.

* 1. **How the velocity of sprint is measured?**

If capacity is measured as a percentage of 40 hours weeks then completed  
= story points \* team capacity  
If capacity is measured in man hours then completed story points / team capacity.

* 1. **How to measure velocity if our iteration lengths change?**

You can’t measure it easily. Velocity’s value comes from its inherent consistency. A fixed iteration length helps drive the reliable rhythm of a project. Without this rhythm, you are constantly revising, re-estimating, and reconciling, and the ability to predict out in the future is minimized due to inconsistent results.  
If, on the other hand, almost everyone is going to be out a week for the holidays or a couple days for company-wide meetings then by all means adapt iteration dates or velocity accordingly. Like most agile practices, these are guidelines, not strict rules.

* 1. **List out what are the artifacts of Scrum process?**

Scrum process artifacts include

* + - * Sprint backlog
      * Product backlog
      * Velocity chart
      * Burn-down chart
  1. **Explain what is scrum of scrum?**

Scrum of scrum is used to refer the meeting after the daily scrum.  The responsible person from each team attends the meeting and discuss their work and answer the questions like

* + - * Since the last meeting, what is the progress of the team?
      * What your team is expected to do or should accomplish, before the next meeting?
      * What are the obstacles your team faced while completing the task?
      * Were you going to allot any of your work to the following team?
  1. **Explain when Scrum cannot be useful?**

Ideally scrum is useful to monitor work with 5 to 10 people, who are committed to achieving the sprint goal.  It does not go well with huge groups or team having more responsibilities.  For larger team, scrum can be applied by splitting the team into small groups and practice scrum.

* 1. **What is Spike?**

There may be some technical issues or design problem in the project which needs to be resolved first. To provide the solution of these problem “Spikes” are created. Spikes are of two types- Functional and Technical.

**Map Reduce and HDFS**

* 1. **How to compress mapper output but not the reducer output?**

With MR2, now we should set

conf.set("**mapreduce.map.output.compress**", true)

conf.set("**mapreduce.output.fileoutputformat.compress**", false)

By Default the property(mapreduce.map.output.compress) has been set as false.

* 1. **Explain what is distributed Cache in MapReduce Framework ?**

Distributed Cache is an important feature provided by map reduce framework. When you want to share some files across all nodes in Hadoop Cluster, DistributedCache  is used.  The files could be an executable jar files or simple properties file.

* 1. **Explain what is JobTracker in Hadoop? What are the actions followed by Hadoop?**

In Hadoop for submitting and tracking MapReduce jobs,  JobTracker is used. Job tracker run on its own JVM process .

Hadoop performs following actions in Hadoop

* + - * Client application submit jobs to the job tracker
      * JobTracker communicates to the Namemode to determine data location
      * Near the data or with available slots JobTracker locates TaskTracker nodes
      * On chosen TaskTracker Nodes, it submits the work
      * When a task fails, Job tracker notify and decides what to do then
      * The TaskTracker nodes are monitored by JobTracker
  1. **Explain what is Speculative Execution?**

It is the option for Hadoop to specify backup tasks if it detects that there are some slow tasks on a few of the cluster nodes.  The backup tasks will be preferentially scheduled on the faster nodes.  Whichever of the duplicate tasks finishes first becomes the one that is used in further operations.  
  
This is good to have on if you might have some slow nodes for whatever reason and there are available hardware resources to support additional tasks on the other nodes**.  Indeed it is on by default.**  However, you will want to turn it off if you are resource-limited.  You will also want to make sure it is off if you are copying any information into a distributed database to avoid duplicate entries.    
  
**To turn off speculative execution entirely, just set the values of the following to false:**

* **mapred.map.tasks.speculative.execution** (mapreduce.map.speculative in Hadoop 2.x)
* **mapred.reduce.tasks.speculative.execution** (mapreduce.reduce.speculative in Hadoop 2.x)

I additionally have a special case where I use mappers to write data to files outside of the standard output collection code.  If I need to append data to those files, speculative execution must be off or else I will get duplicate data in those files as well.

* 1. **Explain what is the function of MapReducer partitioner?**

The function of MapReducer partitioner is to make sure that all the value of a single key goes to the same reducer, eventually which helps evenly distribution of the map output over the reducers.

* 1. **Explain what is difference between an Input Split and HDFS Block?**

Logical division of data is known as Split while physical division of data is known as HDFS Block.

* 1. **Mention what are the main configuration parameters that user need to specify to run Mapreduce Job ?**

The user of Mapreduce framework needs to specify

* + - * Job’s input locations in the distributed file system
      * Job’s output location in the distributed file system
      * Input format
      * Output format
      * Class containing the map function
      * Class containing the reduce function
      * JAR file containing the mapper, reducer and driver classes
  1. **Explain what is sqoop in Hadoop ?**

To transfer the data between Relational database management (RDBMS) and Hadoop HDFS a tool is used known as Sqoop. Using Sqoop data can be transferred from RDMS like MySQL or Oracle into HDFS as well as exporting data from HDFS file to RDBMS.

* 1. **Explain how can you debug Hadoop code?**

The popular methods for debugging Hadoop code are:

* + - * By using web interface provided by Hadoop framework
      * By using Counters

# [**How to do I copy data from one HDFS to another HDFS?**](https://stackoverflow.com/questions/31862904/how-to-do-i-copy-data-from-one-hdfs-to-another-hdfs)

DistCp (distributed copy) is a tool used for copying data between clusters. It uses MapReduce to effect its distribution, error handling and recovery, and reporting. It expands a list of files and directories into input to map tasks, each of which will copy a partition of the files specified in the source list.

Usage: $ hadoop distcp <src> <dst>

example: $ hadoop distcp hdfs://nn1:8020/file1 hdfs://nn2:8020/file2

file1 from nn1 is copied to nn2 with filename file2

Distcp is the best tool as of now. Sqoop is used to copy data from relational database to HDFS and vice versa, but not between HDFS to HDFS.

If the clusters are running identical versions of hadoop, then the hdfs scheme is appropriate to use.

$ hadoop distcp hdfs://namenode1/foo hdfs://namenode2/bar

The data in /foo directory of namenode1 will be copied to /bar directory of namenode2. If the /bar directory does not exist, it will create it. Also we can mention multiple source paths.

Similar to rsync command, distcp command by default will skip the files that already exist. We can also use -overwrite option to overwrite the existing files in destination directory. The option -update will only update the files that have changed.

$ hadoop distcp -update hdfs://namenode1/foo hdfs://namenode2/bar/foo

distcp can also be implemented as a MapReduce job where the work of copying is done by the maps that run in parallel across the cluster. There will be no reducers.

If trying to copy data between two HDFS clusters that are running different versions, the copy will process will fail, since the RPC systems are incompatible. In that case we need to use the read-only HTTP based HFTP filesystems to read from the source. Here the job has to run on destination cluster.

$ hadoop distcp hftp://namenode1:50070/foo hdfs://namenode2/bar

* 1. **What do you mean by data locality?**

[***Data locality***](http://www.edureka.co/blog/mapreduce-tutorial/?utm_source=blog&utm_medium=content-link&utm_campaign=hadoop-interview-questions-mapreduce/#data_locality) talks about moving computation unit to data rather data to the computation unit. MapReduce framework achieves data locality by processing data locally i.e. processing of the data happens in the very node by Node Manager where data blocks are present.

* 1. **What happens when a DataNode fails during the write process?**

When a DataNode fails during the write process, a new replication pipeline that contains the other DataNodes opens up and the write process resumes from there until the file is closed. NameNode observes that one of the blocks is under-replicated and creates a new replica asynchronously.

* 1. **Explain the differences between a combiner and reducer. When is it suggested to use a combiner in a MapReduce job?**

Combiner can be considered as a mini reducer that performs local reduce task. It runs on the Map output and produces the output to reducers input. It is usually used for network optimization when the map generates greater number of outputs.

* + - * Unlike a reducer, the combiner has a constraint that the input or output key and value types must match the output types of the Mapper.
      * Combiners can operate only on a subset of keys and values i.e. combiners can be executed on functions that are commutative.
      * Combiner functions get their input from a single mapper whereas reducers can get data from multiple mappers as a result of partitioning.

Combiners are generally used to enhance the efficiency of a MapReduce program by aggregating the intermediate map output locally on specific mapper outputs. This helps reduce the volume of data that needs to be transferred to reducers. Reducer code can be used as a combiner, only if the operation performed is commutative. However, the execution of a combiner is not assured.

* 1. **Explain the process of spilling in MapReduce? Why the output of map tasks are stored (spilled) into local disc and not in HDFS?**

The output of a map task is written into a circular memory buffer (RAM). The default size of buffer is set to 100 MB which can be tuned by using **mapreduce.task.io.sort.mb** property. Now, spilling is a process of copying the data from memory buffer to disc when the content of the buffer reaches a certain threshold size. By default, a background thread starts spilling the contents from memory to disc after 80% of the buffer size is filled. Therefore, for a 100 MB size buffer the spilling will start after the content of the buffer reach a size of 80 MB.

**Note:** One can change this spilling threshold using **mapreduce.map.sort.spill.percent** which is set to 0.8 or 80% by default.

The outputs of map task are the intermediate key-value pairs which is then processed by reducer to produce the final aggregated result. Once a MapReduce job is completed, there is no need of the intermediate output produced by map tasks. Therefore, storing these intermediate output into HDFS and replicate it will create unnecessary overhead.

#################### PRINTED TILL HERE ###############################

**HDFS**

* 1. **How the client communicates with Name node and Data node in HDFS?**

The communication mode for clients with name node and data node in HDFS is **SSH**.

**Secure Shell** (**SSH**) is a [cryptographic](https://en.wikipedia.org/wiki/Cryptography) [network protocol](https://en.wikipedia.org/wiki/Network_protocol) for operating network services securely over an unsecured network. SSH provides a [secure channel](https://en.wikipedia.org/wiki/Secure_channel) over an unsecured network in a [client-server](https://en.wikipedia.org/wiki/Client-server) architecture, connecting an [SSH client](https://en.wikipedia.org/wiki/SSH_client) application with an [SSH server](https://en.wikipedia.org/wiki/SSH_server).

## ****Explain the HDFS Architecture and list the various HDFS daemons in HDFS cluster?****

While listing various HDFS daemons, you should also talk about their roles in brief. Here is how you should answer this question:

Apache Hadoop [***HDFS Architecture***](http://www.edureka.co/blog/apache-hadoop-hdfs-architecture/?utm_source=blog&utm_medium=content-link&utm_campaign=hadoop-interview-questions-hdfs-2) follows a Master/Slave topology where a cluster comprises a single NameNode (Master node or daemon) and all the other nodes are DataNodes (Slave nodes or daemons).  Following daemon runs in HDFS cluster:

* + - * **NameNode:**It is the master daemon that maintains and manages the data block present in the DataNodes.
      * **DataNode:** DataNodes are the slave nodes in HDFS. Unlike NameNode, DataNode is a commodity hardware, that is responsible of storing the data as blocks.
      * **Secondary NameNode:**The Secondary NameNode works concurrently with the primary NameNode as a *helper daemon*. It performs checkpointing.

## ****Apache Hadoop HDFS**** ****Architecture****

## HDFS Topology - Apache Hadoop HDFS Architecture - Edureka

## **Apache HDFS** or **Hadoop Distributed File System** is a block-structured file system where each file is divided into blocks of a pre-determined size. These blocks are stored across a cluster of one or several machines. Apache Hadoop HDFS Architecture follows a Master/Slave Architecture, where a cluster comprises of a single NameNode (Master node) and all the other nodes are DataNodes (Slave nodes). HDFS can be deployed on a broad spectrum of machines that support Java. Though one can run several DataNodes on a single machine, but in the practical world, these DataNodes are spread across various machines.

## ****NameNode:****

## Apache Hadoop HDFS Architecture - Edureka

NameNode is the master node in the Apache Hadoop HDFS Architecture that maintains and manages the blocks present on the DataNodes (slave nodes). NameNode is a very highly available server that manages the File System Namespace and controls access to files by clients. The HDFS architecture is built in such a way that the user data never resides on the NameNode. The data resides on DataNodes only.

### **Functions of NameNode:**

* 1. It is the master daemon that maintains and manages the DataNodes (slave nodes)
  2. It records the metadata of all the files stored in the cluster, e.g. the location of blocks stored, the size of the files, permissions, hierarchy, etc. There are two files associated with the metadata:
     + - **FsImage:** It contains the complete state of the file system namespace since the start of the NameNode.
       - **EditLogs:** It contains all the recent modifications made to the file system with respect to the most recent FsImage.
  3. It records each change that takes place to the file system metadata. For example, if a file is deleted in HDFS, the NameNode will immediately record this in the EditLog.
  4. It regularly receives a Heartbeat and a block report from all the DataNodes in the cluster to ensure that the DataNodes are live.
  5. It keeps a record of all the blocks in HDFS and in which nodes these blocks are located.
  6. The NameNode is also responsible to take care of the **replication factor**of all the blocks.
  7. In **case of the DataNode failure**, the NameNode chooses new DataNodes for new replicas,balance disk usage and manages the communication traffic to the DataNodes.

## ****DataNode:****

The DataNode is a block server that stores the data in the local file ext3 or ext4.

### **Functions of DataNode:**

* These are slave daemons or process which runs on each slave machine.
* The actual data is stored on DataNodes.
* The DataNodes perform the low-level read and write requests from the file system’s clients.
* They send heartbeats to the NameNode periodically to report the overall health of HDFS, by default, this frequency is set to 3 seconds.

## ****Secondary NameNode:****

Apart from these two daemons, there is a third daemon or a process called Secondary NameNode. The Secondary NameNode works concurrently with the primary NameNode as a **helper daemon.**And don’t be confused about the Secondary NameNode being a**backup NameNode because it is not.**



### **Functions of Secondary NameNode:**

* The Secondary NameNode is one which constantly reads all the file systems and metadata from the RAM of the NameNode and writes it into the hard disk or the file system.
* It is responsible for combining the EditLogs with FsImage from the NameNode.
* It downloads the EditLogs from the NameNode at regular intervals and applies to FsImage. The new FsImage is copied back to the NameNode, which is used whenever the NameNode is started the next time.

Hence, Secondary NameNode performs regular checkpoints in HDFS. Therefore, it is also called CheckpointNode.

## ****Blocks:****

Now, as we know that the data in HDFS is scattered across the DataNodes as blocks. **Let’s have a look at what is a block and how is it formed?**

Blocks are the nothing but the smallest continuous location on your hard drive where data is stored. In general, in any of the File System, you store the data as a collection of blocks. Similarly, HDFS stores each file as blocks which are scattered throughout the Apache Hadoop cluster. The default size of each block is 128 MB in Apache Hadoop 2.x (64 MB in Apache Hadoop 1.x) which you can configure as per your requirement.

HDFS ARCHITECTURE – HDFS TUTORIAL Introduction In this blog, we are going to talk about HDFS Architecture. From my previous blog, we already know that HDFS is a distributed file system which is deployed on low cost commodity hardware. I discussed many of its features too. So, its high time that we take a deep dive into Apache Hadoop HDFS Architecture and unlock its beauty. The topics that will be covered in this blog are as follows: • HDFS Master/Slave Topology • NameNode and DataNode • What is a block • Replication Management • Rack Awareness • HDFS Read/Write – Behind the scenes HDFS Architecture HDFS or Hadoop Distributed File System is a block-structured file system where each file is divided into blocks of a pre-determined size. These blocks are stored across a cluster of one or several machines. Apache Hadoop HDFS Architecture follows a Master/Slave Architecture, where a cluster comprises of a single NameNode (Master node) and all the other nodes are DataNodes (Slave nodes). HDFS is based on Java programming language, due to which HDFS can be deployed on broad spectrum of machines that support Java. Though one can run several DataNodes on a single machine, but in practical world, these DataNodes are spread across various machines. NameNode and DataNode NameNode: NameNode is the master of HDFS that maintains and manages the blocks present on the DataNodes (slave nodes). Think of the NameNode as a Lamborghini in midst of various other cars. Thus, like a Lamborghini, NameNode is a very highly available server that manages the File System Namespace and controls access to files by clients. I will be discussing this High Availability feature of Apache Hadoop HDFS in my next blog. The HDFS architecture is built in such a way that the user data is never stored in the NameNode. The data resides on DataNodes only. Functions of NameNode: • It is the master daemon that maintains and manages the DataNodes (slave nodes) • It records the metadata of all the files stored in the cluster, e.g. location of blocks stored, size of the files, permissions, hierarchy, etc. There are two files associated with metadata: o FsImage: An image of the file system on starting the NameNode. o EditLogs: A series of modifications done to the file system after starting the NameNode. • It records each change that takes place to the file system metadata. For example, if a file is deleted in HDFS, the NameNode will immediately record this in the EditLog. • It regularly receives a Heartbeat and a block report from all the DataNodes in the cluster to ensure that the DataNodes are live. • It keeps a record of all the blocks in HDFS and in which nodes these blocks are located. • The NameNode is also responsible to take care of the replication factor of all the blocks which we will discuss in detail later in this HDFS tutorial blog. • In case of a DataNode failure, the NameNode chooses new DataNodes for new replicas, balances disk usage and manages the communication traffic to the DataNodes. DataNode: DataNodes are the slave nodes in HDFS, just like any average car in front of a Lamborghini! Unlike NameNode, DataNode is a commodity hardware, that is, a non-expensive system which is not of high quality or high-availability. DataNode is a block server that stores the data in the local file ext3 or ext4. Functions of DataNode: • These are slave daemons or process which runs on each slave machine. • The actual data is stored on DataNodes. • DataNodes perform the low-level read and write requests from the file system’s clients. • They send heartbeats to the NameNode periodically to report the overall health of HDFS, by default, this frequency is set to 3 seconds. So, till now, you folks must have realized that the NameNode is pretty much important to us. If it fails, we are doomed. But don’t worry, we will be talking about how Hadoop solved this single point of failure problem in the next HDFS tutorial blog. So, just relax for now and let’s take one step at a time. Secondary NameNode Apart from these two daemons there is a third daemon or process called Secondary NameNode. The Secondary NameNode works concurrently with the primary NameNode as a helper daemon. And don’t confuse Secondary NameNode as a backup NameNode because it is not. Functions of Secondary NameNode: • The Secondary NameNode is one which constantly reads all the file systems and metadata from the RAM of the NameNode and writes it into the hard disk or the file system. • It is responsible for combining the EditLogs with FsImage from the NameNode. • It downloads the EditLogs from the NameNode at regular intervals and applies to FsImage. The new FsImage is copied back to the NameNode, which is used whenever the NameNode is started the next time. Hence, Secondary NameNode just performs regular checkpoints in HDFS. Therefore, it is also called CheckpointNode. Blocks Now as we know that the data in HDFS is scattered across the DataNodes as blocks. Let’s have a look on what is a block and how is it formed? Blocks are the nothing but smallest continuous location in your hard drive where data is stored. In general, in any of the File System the data are stored as collection of blocks. Similarly, HDFS stores each file as blocks which is scattered throughout the Apache Hadoop cluster. The default size of each block is 128MB in Apache Hadoop 2.x (64 MB in Apache Hadoop 1.x) which you can configure as per your requirement. It is not necessary that in HDFS, each file is stored in exact multiple of the configured block size (128MB, 256MB etc.). Let’s take an example where I have a file “example.txt” of size 514MB as shown in above figure. Suppose, we are using the default block size configuration which is 128Mb. So, how many blocks will be created? 5, Right. First four blocks will be of 128 MB. But, the last block will be of 2 MB size only. Now you must be thinking why we need to have such a huge blocks size i.e. 128MB? Well, whenever we talk about HDFS, we talk about huge data sets i.e. terabytes and petabytes of data. So, if we had a block size of let’s say 4KB as in Linux file system, we would be having too many of blocks and therefore too much of metadata. So, managing these no. of blocks and metadata will create huge overhead which is something, we don’t want. As you understood what a block is, lets understand how these blocks are places in the next section. Replication Management and Rack Awareness Replication Management: HDFS provides a reliable way to store huge data in a distributed environment as data blocks. The blocks are also replicated to provide fault tolerance. The default replication factor is 3 which is again configurable. So, if you want to store a file of 1GB in your HDFS, you will be consuming a space of 3GB (replication factor =3) and 24 (1GB/128MB=8 data blocks) data blocks in total, considering the default configuration. Don’t worry guys, if you didn’t get the math in one go. Take your time then move ahead. The NameNode collects block report from DataNode periodically to maintain the replication factor. Therefore, whenever a block is over-replicated or under-replicated the NameNode deletes or add replicas as needed. Rack Awareness: Anyways, moving ahead, let’s talk more about how replica are placed and what is rack awareness? Again, the NameNode also ensures that all the replicas are not stored on the same rack or a single rack. It follows an in-built Rack Awareness Algorithm to reduce latency. Considering the replication factor is 3, the Rack Awareness Algorithm says that the first replica of a block will be stored on a local rack and the next two replicas will be stored on a different (remote) rack but on a different DataNode within that (remote) rack. If you have more replicas, the rest of the replicas will be placed on random DataNodes provided not more than two replicas reside on the same rack, if possible. This is how an actual Hadoop production cluster looks like. Here, you have multiple racks populated with many DataNodes. Advantages of Rack Awareness: So now you will be thinking why do we need Rack Awareness algorithm? The reasons are: • To improve the network performance: The communication between nodes residing on different racks is directed via switch. In general, you will find greater network bandwidth between machines in the same rack than the machines residing in different rack. So, the Rack Awareness helps you to have reduce write traffic in between different racks and thus providing a better write performance. Also, you will be gaining increased read performance because you are using the bandwidth of multiple racks. • To prevent loss of data: We don’t have to worry about the data even if an entire rack fails because of the switch failure or power failure. And if you think about it, it will make sense, as it is said that never put all your eggs in the same basket. HDFS Read/ Write Architecture Now let’s talk about how the data read/write operations are performed on HDFS. HDFS follows Write Once - Read Many philosophy. So, you can’t edit files already stored in HDFS. But, you can append new data in a file. HDFS Write Architecture: Imagine a situation where a HDFS client, want to write a file named “example.txt” of size 248MB. Assume that the system block size is configured to 128 MB (default). So, the client will be dividing the file “example.txt” into 2 blocks – one of 128MB (Block A) and the other of 120 MB (block B). Now, the following protocol will be followed whenever the data is written into HDFS: • At first, the HDFS client will reach out to the NameNode for a Write Request against the two blocks, say, Block A & Block B. • The NameNode will then grant client the write permission and will provide the IP addresses of the DataNodes where the file blocks will be copied eventually. • The selection of IP addresses of DataNodes is purely randomized based on availability, replication factor and rack awareness that we have discussed earlier. • Let’s say the replication factor is set to default i.e. 3. Therefore, for each block the NameNode will be providing the client a list of (3) IP addresses of DataNodes. The list will be unique for each block. • Suppose, the NameNode provided following lists of IP addresses to the client: o For Block A, list A = {IP of DataNode 1, IP of DataNode 4, IP of DataNode 5} o For Block B, set B = {IP of DataNode 3, IP of DataNode 7, IP of DataNode 9} • Each block will be copied in three different DataNodes to maintain the replication factor consistent. • Now the whole data copy process will happen in three stages: 1. Set up of Pipeline 2. Data streaming and replication 3. Shutdown of Pipeline (Acknowledgement stage) 1. Set up of Pipeline: Before writing the blocks, client confirms whether the DataNodes present in each of the list of IPs are ready to receive the data or not. For doing so, the client creates a pipeline for each of the blocks by connecting the individual DataNodes in the respective list for that block. Let us consider Block A. The list of DataNodes provided by the NameNode is : For Block A, list A = {IP of DataNode 1, IP of DataNode 4, IP of DataNode 5}. So, for block A, the client will be performing following steps to create a pipeline:  Client will choose first DataNode in the list (DataNode IPs for Block A) which is DataNode 1 and will open a TCP/IP connection.  Client will inform DataNode 1 to be ready to receive the block. It will also provide the IPs of next two DataNodes (4 and 5) where the block will be replicated.  DataNode 1 will connect to DataNode 4. DataNode 1 will inform DataNode 4 to be ready to receive the block and will give it the IP of DataNode 5. Then, DataNode 4 will tell DataNode 5 to be ready for receiving the data.  Next, the acknowledgement of readiness will follow the reverse sequence i.e. From DataNode 5 to 4 and then to 1.  At last DataNode 1 will inform the client that all the DataNodes are ready and a pipeline is formed between client, DataNode 1, 4 and 5.  Now pipeline set up is complete and the client will finally begin the data copy or streaming process. 2. Data Streaming: As the pipeline has been created, the client will push the data into the pipeline. Now, don’t forget that in HDFS, data is replicated based on replication factor. So, here Block A will be stored to three DataNodes as the assumed replication factor is 3. Moving ahead, the client will copy the block (A) to DataNode 1 only. The replication is always done by DataNodes sequentially. So, following steps will take place during replication:  Once the block has been written to DataNode 1, it will connect to DataNode 4.  Then, it will push the block in the pipeline and data will be copied to DataNode 4.  Again, DataNode 4 will connect to DataNode 5 and will copy the block. 3. Shutdown of Pipeline or Acknowledgement stage: Once the block has been copied into all the three DataNodes, a series of acknowledgements will take place to ensure the client and NameNode that the data has been written successfully. And client will finally close the pipeline to end the TCP session. As shown in the figure, the acknowledgement happens in the reverse sequence i.e. from DataNode 5 to 4 and then to 1. Finally, the DataNode 1 will push three acknowledgements (including its own) into the pipeline and send it to client. Client will inform NameNode that data has been written successfully. NameNode will update its metadata and the client will shut down the pipeline. Similarly, Block B will also be copied into the DataNodes in parallel with Block A. But, following things are to be noticed here:  Client will copy Block A and Block B to the first DataNode simultaneously.  Therefore, in our case two pipelines will be formed for each of the block and all the process discussed above will happen in parallel in these two pipelines.  Client writes the block into first DataNode and then the DataNodes will be replicating the block sequentially. Folks! Its time for a quiz now: In HDFS, blocks of a file are written in parallel, however the replication of the blocks are done sequentially: a. True b. False True. Right guys. So, if you are still confused, go through the HDFS write architecture again. I am sure you will understand it this time. HDFS Read Architecture: HDFS Read architecture is comparatively easy to understand. Let’s the above example again where a HDFS clients want to read the file “example.txt” which is already there in HDFS. Now, following steps will be taking place while reading the file: • Client will reach out to NameNode asking for the block metadata for the file example.txt. • The NameNode will return the list of DataNodes where each block (Block A and B) are stored. • After that client, will connect to the DataNodes where the blocks are stored. • Client starts reading data parallel from the Data nodes (Block A from DataNode 1 and Block B from DataNode 5). • Once it gets all the required file blocks, it will combine these blocks to form a file. While serving read request of the client, HDFS selects the replica which is closest to the client. This reduces the read latency and the bandwidth consumption. Therefore, that replica is selected which resides on the same rack as the reader node, if possible. 

It is not necessary that in HDFS, each file is stored in exact multiple of the configured block size (128 MB, 256 MB etc.). Let’s take an example where I have a file “example.txt” of size 514 MB as shown in above figure.  Suppose that we are using the default configuration of block size, which is 128 MB. Then, how many blocks will be created? 5, Right. The first four blocks will be of 128 MB. But, the last block will be of 2 MB size only.

**Now, you must be thinking why we need to have such a huge blocks size i.e. 128 MB?**

Well, whenever we talk about HDFS, we talk about huge data sets, i.e. Terabytes and Petabytes of data. So, if we had a block size of let’s say of 4 KB, as in Linux file system, we would be having too many blocks and therefore too much of the metadata. So, managing these no. of blocks and metadata will create huge overhead, which is something, we don’t want.

**Setting up Hadoop Cluster**

* 1. **Which are the modes in which Hadoop can run ?**

We have three modes in which Hadoop can run and that are:

* ***Standalone (local) mode***: Default mode of Hadoop, it uses the local file system for input and output operations. This mode is mainly used for debugging purpose, and it does not support the use of HDFS. Further, in this mode, there is no custom configuration required for mapred-site.xml, core-site.xml, hdfs-site.xml files. In stand-alone mode, there are no daemons, everything runs on a single JVM. It has no DFS and it utilizes the local file system. Stand-alone mode is suitable only for running MapReduce programs during development for testing. It is one of the most least used environments.
* ***Pseudo-distributed mode:*** In this case, you need configuration for all the three files mentioned above. In this case, all daemons are running on one node and thus, both Master and Slave nodes are on the same machine.
* ***Fully distributed mode:*** This is the production phase of Hadoop where data is distributed across several nodes on a Hadoop cluster. Separate nodes are allotted as Master and Slaves.
  + 1. **What is configured in /etc/hosts and what is its role in setting Hadoop cluster?**

/etc/hosts file contains the hostname and their IP address of that host. It maps the IP address to the hostname. In Hadoop cluster, we store all the hostnames (master and slaves) with their IP addresses in /etc/hosts so, that we can use hostnames easily instead of IP addresses.

## ****What are the default port numbers of NameNode, ResourceManager & MapReduce JobHistory Server?****

* + - * Namenode – ’50070’
      * ResourceManager – ’8088’
      * MapReduce JobHistory Server – ’19888’.

## ****What are the main Hadoop configuration files?****

4 main configuration files in Hadoop.

* **core-site.xml:**core-site.xml informs Hadoop daemon where NameNode runs on the cluster. It contains configuration settings of Hadoop core such as I/O settings that are common to HDFS & MapReduce.
* **hdfs-site.xml:** hdfs-site.xml contains configuration settings of HDFS daemons (i.e. NameNode, DataNode, Secondary NameNode). It also includes the replication factor and block size of HDFS.
* **mapred-site.xml**: mapred-site.xml contains configuration settings of the MapReduce framework like number of JVM that can run in parallel, the size of the mapper and the reducer, CPU cores available for a process, etc.
* **yarn-site.xml:**yarn-site.xml contains configuration settings of ResourceManager and NodeManager like application memory management size, the operation needed on program & algorithm, etc.

These files are in the conf/hadoop/directory inside Hadoop directory.

## ****What is command to extract the compressed file in tar.gz format?****

## tar -xvf /file\_location/filename.tar.gzcommand will extract the tar.gz compressed file.

## 6. ****What is the default replication factor and how will you change it?****

The default replication factor is 3.

♣ Tip: Default Replication Factor could be changed in three ways. Answering all the three ways will show your expertise.

1. By adding this property to **hdfs-site.xml**:

|  |  |
| --- | --- |
| 1  2  3  4  5 | <property>  <name>dfs.replication</name>  <value>5</value>  <description>Block Replication</description>  </property> |

1. Or you can change the replication factor on per file basis using following command:

**hadoop fs –setrep –w 3 /file\_location**

1. Or you can change replication factor for all the files in a directory using the following command:

**hadoop fs –setrep –w 3 -R /directory\_location**

## 7. What happens if you get a ‘connection refused java exception’ when you type hadoop fsck /?

If you get a ‘connection refused java exception’ when you type hadoop fsck, it could mean that the NameNode is not working on your VM.

## 8. How can we view the compressed files via HDFS command?

We can view compressed files in HDFS using **hadoop fs -text /filename**command.

## 9. ****What is the command to move into safe mode and exit safe mode?****

Safe Mode in Hadoop is a maintenance state of the NameNode during which NameNode doesn’t allow any changes to the file system. During Safe Mode, HDFS cluster is read-only and doesn’t replicate or delete blocks.

* To know the status of safe mode, you can use the command: ***hdfs dfsadmin -safemode get***
* To enter safe mode: ***hdfs dfsadmin -safemode enter***
* To exit safe mode: ***hdfs dfsadmin -safemode leave***

## 10. ****How can I restart Namenode?****

This question has two answers, answering both will give you a plus point. We can restart NameNode by following methods:

1. You can stop the NameNode individually using***.*** ***/sbin /hadoop-daemon.sh stop namenode*** command and then start the NameNode using***.*** ***/sbin/hadoop-daemon.sh start namenode***
2. Use .***/sbin/stop-all.sh*** and and then use .***/sbin/start-all.sh*** command which will stop all the daemons first and then start all the daemons.

## 11) What are the different commands used to startup and shutdown Hadoop daemons?

1. ***./sbin/start-all.sh*** to start all the Hadoop daemons and .***/sbin/stop-all.sh*** to stop all the Hadoop daemons.
2. Then you can start all the dfs daemons together using***.*** ***/sbin/start-dfs.sh***, yarn daemons together using***.*** ***/sbin/start-yarn.sh***and MR job history server using***.*** ***/sbin/mr-jobhistory-daemon.sh start historyserver***. To stop these daemons similarly we can use***.*** ***/sbin/start-yarn.sh***, ***./sbin/start-yarn.sh****&****.*** ***/sbin/mr-jobhistory-daemon.sh stop historyserver***.
3. The last way is to start all the daemons individually and stop them individually:

***./sbin/hadoop-daemon.sh start namenode***

***./sbin/hadoop-daemon.sh start datanode***

***./sbin/yarn-daemon.sh start resourcemanager***

***./sbin/yarn-daemon.sh start nodemanager***

***./sbin/mr-jobhistory-daemon.sh start historyserver***

and stop them similarly.

## 12) What does hadoop-metrics.properties file do?

## hadoop-metrics.properties is used for ‘Performance Reporting‘ purposes. It controls the reporting for Hadoop. The API is abstract so that it can be implemented on top of a variety of metrics client libraries. The choice of client library is a configuration option, and different modules within the same application can use different metrics implementation libraries. This file is stored inside /etc/hadoop.

## 13) What happens to a NameNode, when ResourceManager is down?

## When a ResourceManager is down, it will not be functional (for submitting jobs) but NameNode will be present. So, the cluster is accessible if NameNode is working, even if the ResourceManager is not working.

## 14) How can we format HDFS?

Hadoop distributed file system(HDFS) can be formatted using ***bin/hadoop namenode -format***command. This command formats the HDFS via NameNode. This command is only executed for the first time. Formatting the file system means initializing the directory specified by the **dfs.name.dir** variable. If you run this command on existing filesystem, you will lose all your data stored on your NameNode. Formatting a Namenode will not format the DataNode. It will format the **FsImage** and **edit logs** data stored on the NameNode and will lose the data about the location of blocks stored in HDFS.

Never format, up and running Hadoop filesystem. You will lose all your data stored in the HDFS.

### **15) What happens when two clients try to access the same file in the HDFS?**

HDFS supports exclusive writes only.

When the first client contacts the “NameNode” to open the file for writing, the “NameNode” grants a lease to the client to create this file. When the second client tries to open the same file for writing, the “NameNode” will notice that the lease for the file is already granted to another client, and will reject the open request for the second client.

### **16) State the reason why we can’t perform “aggregation” (addition) in mapper? Why do we need the “reducer” for this?**

This answer includes many points, so we will go through them sequentially.

* We cannot perform “aggregation” (addition) in mapper because sorting does not occur in the “mapper” function. Sorting occurs only on the reducer side and without sorting aggregation cannot be done.
* During “aggregation”, we need output of all the mapper functions which may not be possible to collect in the map phase as mappers may be running on different machine where the data blocks are stored.
* And lastly, if we try to aggregate data at mapper, it requires communication between all mapper functions which may be running on different machines. So, it will consume high network bandwidth and can cause network bottlenecking.

### 17) **How do “reducers” communicate with each other?**

The “MapReduce” programming model does not allow “reducers” to communicate with each other. “Reducers” run in isolation.

**Apache Hive**

## Apache Hive is a data warehouse system built on top of Hadoop and is used for analyzing structured and semi-structured data. Internally, these queries or HQL gets converted to map reduce jobs by the Hive compiler.

## ****Where does the data of a Hive table gets stored?****

## By default, the Hive table is stored in an HDFS directory – /user/hive/warehouse. One can change it by specifying the desired directory in hive.metastore.warehouse.dir configuration parameter present in the hive-site.xml.

## ****2) What is a metastore in Hive?****

## Metastore stores the meta data information using RDBMS and an open source ORM (Object Relational Model) layer called Data Nucleus which converts the object representation into relational schema and vice versa. The reason for choosing RDBMS instead of HDFS is to achieve low latency. We can implement metastore in following three configurations:

### **Embedded Metastore:**



## Both the metastore service and the Hive service runs in the same JVM by default using an embedded Derby Database instance where metadata is stored in the local disk. This is called embedded metastore configuration. In this case, only one user can connect to metastore database at a time. If you start a second instance of Hive driver, you will get an error. This is good for unit testing, but not for the practical solutions.

### **Local Metastore:**



## This configuration allows us to have multiple Hive sessions i.e. Multiple users can use the metastore database at the same time. This is achieved by using any JDBC compliant database like MySQL which runs in a separate JVM or a different machine than that of the Hive service and metastore service which are running in the same JVM as shown above. In general, the most popular choice is to implement a MySQL server as the metastore database.

### **3. Remote Metastore:**

### Remote Mode Metastore - Hive Tutorial - Edureka

## In the remote metastore configuration, the metastore service runs on its own separate JVM and not in the Hive service JVM. Other processes communicate with the metastore server using Thrift Network APIs. You can have one or more metastore servers in this case to provide more availability. The main advantage of using remote metastore is you do not need to share JDBC login credential with each Hive user to access the metastore database.

## 3) ****What is the default database provided by Apache Hive for metastore?****

## By default, Hive provides an embedded Derby database instance backed by the local disk for the metastore. This is called the embedded metastore configuration.

## Suppose I have installed Apache Hive on top of my Hadoop cluster using default metastore configuration. Then, what will happen if we have multiple clients trying to access Hive at the same time?

The default metastore configuration allows only one Hive session to be opened at a time for accessing the metastore. Therefore, if multiple clients try to access the metastore at the same time, they will get an error. One has to use a standalone metastore, i.e. Local or remote metastore configuration in Apache Hive for allowing access to multiple clients concurrently.

Following are the steps to configure MySQL database as the local metastore in Apache Hive:

* One should make the following changes in **hive-site.xml**:
  + **javax.jdo.option.ConnectionURL** property should be set to jdbc:mysql://host/dbname?createDataba  
    seIfNotExist=true.
  + **javax.jdo.option.ConnectionDriverName** property should be set to com.mysql.jdbc.Driver.
  + One should also set the username and password as:
    - **javax.jdo.option.ConnectionUserName** is set to desired username.
    - **javax.jdo.option.ConnectionPassword** is set to the desired password.
* The JDBC driver JAR file for MySQL must be on the Hive’s classpath, i.e. The jar file should be copied into the Hive’s lib directory.
* Now, after restarting the Hive shell, it will automatically connect to the MySQL database which is running as a standalone metastore.

## ****Is it possible to change the default location of a managed table?****

## Yes, it is possible to change the default location of a managed table. It can be achieved by using the clause – LOCATION ‘<hdfs\_path>’.

## ****5)When should we use SORT BY instead of ORDER BY ?****

## We should use SORT BY instead of ORDER BY when we have to sort huge datasets because SORT BY clause sorts the data using multiple reducers whereas ORDER BY sorts all of the data together using a single reducer. Therefore, using ORDER BY against a large number of inputs will take a lot of time to execute.

## What is Tables, Partitions and Buckets?

## Hive’s data model provides a high-level, table-like structure on top of HDFS. It supports three data structures: tables, partitions, and buckets, where tables correspond to HDFS directories and can be divided into partitions, which in turn can be divided into buckets.

## What is the Temporary data file path?

## hive.exec.scratchdir: is the temporary data file path . By Deafult it is /tmp/hive-${user.name}.

## What are the commands from the Beeline to Hive.

## - Prepare the data as follows:

## -bash-4.1$ vi employee.txt

**Michael|Montreal,Toronto|Male,30|DB:80|Product:Developer^DLead**

**Will|Montreal|Male,35|Perl:85|Product:Lead,Test:Lead**

**Shelley|New York|Female,27|Python:80|Test:Lead,COE:Architect**

**Lucy|Vancouver|Female,57|Sales:89,HR:94|Sales:Lead**

- Log in to Beeline with the proper HiveServer2 hostname, port number, database name, username, and password:

**-bash-4.1$ beeline**

**beeline> !connect jdbc:hive2://localhost:10000/default**

**scan complete in 20ms Connecting to**

**jdbc:hive2://localhost:10000/default**

- Create a table using ARRAY, MAP, and STRUCT composite

data types ( It will create the Managed table/Internal Table)

**jdbc:hive2://> CREATE TABLE employee**

**. . . . . . .> (**

**. . . . . . .> name string,**

**. . . . . . .> work\_place ARRAY<string>,**

**. . . . . . .> sex\_age STRUCT<sex:string,age:int>,**

**. . . . . . .> skills\_score MAP<string,int>,**

**. . . . . . .> depart\_title MAP<string,ARRAY<string>>**

**. . . . . . .> )**

**. . . . . . .> ROW FORMAT DELIMITED**

**. . . . . . .> FIELDS TERMINATED BY '|'**

**. . . . . . .> COLLECTION ITEMS TERMINATED BY ','**

**. . . . . . .> MAP KEYS TERMINATED BY ':';**

**No rows affected (0.149 seconds)**

-Verify the table’s creation:

**jdbc:hive2://>!table employee**

**+---------+------------+------------+--------------+---------+**

**|TABLE\_CAT|TABLE\_SCHEMA| TABLE\_NAME | TABLE\_TYPE | REMARKS |**

**+---------+------------+------------+--------------+---------+**

**| |default | employee | MANAGED\_TABLE| |**

**+---------+------------+------------+--------------+---------+**

**jdbc:hive2://>!column employee**

**+--------------+-------------+---------------+---------------+**

**| TABLE\_SCHEM | TABLE\_NAME | COLUMN\_NAME | TYPE\_NAME |**

**+--------------+-------------+---------------+---------------+**

**| default | employee | name | STRING |**

**| default | employee | work\_place | array<string> |**

**| default | employee | sex\_age |**

**struct<sex:string,age:int>|**

**| default | employee | skills\_score | map<string,int>|**

**| default | employee | depart\_title | map<string,array<string>>**

**| +--------------+-------------+---------------+---------------+**

-Load data into the table: ( Load from the localpath)

**jdbc:hive2://>LOAD DATA LOCAL INPATH '/home/hadoop/employee.txt'**

**. . . . . . .>OVERWRITE INTO TABLE employee;**

**Hive Database**

Whenever a new database is created, Hive creates a directory for each database at **/user/hive/warehouse**, defined in **hive.metastore.warehouse.dir**. For example, the myhivebook database is located at **/user/hive/datawarehouse/myhivebook.db**.

**However, the default database doesn’t have its own directory.**

**Core DDL(Data Definition Language) for Hive Database.**

* Create the database without checking whether the database already exists:

**jdbc:hive2://> CREATE DATABASE myhivebook;**

* Create the database and check whether the database already exists:

**jdbc:hive2://> CREATE DATABASE IF NOT EXISTS myhivebook;**

* Create the database with location, comments, and metadata information:

**jdbc:hive2://> CREATE DATABASE IF NOT EXISTS myhivebook**

**. . . . . . .> COMMENT 'hive database demo'**

**. . . . . . .> LOCATION '/hdfs/directory'**

**. . . . . . .> WITH DBPROPERTIES ('creator'='dayongd','date'='2015-01-**

**01');**

* Drop the empty database:

**jdbc:hive2://> DROP DATABASE IF EXISTS myhivebook;**

Note that Hive keeps the database and the table in directory mode. In order to remove the parent directory, we need to remove the subdirectories first. By default, the database cannot be dropped if it is not empty, unless CASCADE is specified. CASCADE drops the tables in the database automatically before dropping the database.

Drop the database with CASCADE:

**jdbc:hive2://> DROP DATABASE IF EXISTS myhivebook CASCADE;**

* Alter the database properties.

The ALTER DATABASE statement can only apply to the table properties and roles (Hive 0.13.0 and later) on the table. The other metadata about the database cannot be changed:

**jdbc:hive2://> ALTER DATABASE myhivebook**

**. . . . . . .> SET DBPROPERTIES ('edited-by' = 'Dayong');**

**jdbc:hive2://> ALTER DATABASE myhivebook**

**SET OWNER user dayongd;**

**What is Hive Internal and External Tables?**

Each table associates with a directory configured in **${HIVE\_HOME}/conf/hive-site.xml** in HDFS. By default, it is **/user/hive/warehouse** in HDFS.

All the data in the table will be kept in the directory. The Hive table is also referred to as **internal or managed** tables.

When there is data already in HDFS, an external Hive table can be created to describe the data. It is called **EXTERNAL** because the data in the external table is specified in the **LOCATION** properties instead of the default warehouse directory. When keeping data in the internal tables, Hive fully manages the life cycle of the table and data. This means the data is removed once the internal table is dropped. If the external table is dropped, the table

metadata is deleted but the data is kept. Most of the time, an external table is preferred to avoid deleting data along with tables by mistake.

**DDLs for Hive internal and External table**

Create the internal table and load the data:

jdbc:hive2://> CREATE TABLE IF NOT EXISTS employee\_internal

. . . . . . .> (

. . . . . . .> name string,

. . . . . . .> work\_place ARRAY<string>,

. . . . . . .> sex\_age STRUCT<sex:string,age:int>,

. . . . . . .> skills\_score MAP<string,int>,

. . . . . . .> depart\_title MAP<STRING,ARRAY<STRING>>

. . . . . . .> )

. . . . . . .> COMMENT 'This is an internal table'

. . . . . . .> ROW FORMAT DELIMITED

. . . . . . .> FIELDS TERMINATED BY '|'

. . . . . . .> COLLECTION ITEMS TERMINATED BY ','

. . . . . . .> MAP KEYS TERMINATED BY ':'

. . . . . . .> STORED AS TEXTFILE;

No rows affected (0.149 seconds)

jdbc:hive2://> LOAD DATA LOCAL INPATH '/home/hadoop/employee.txt'

. . . . . . .> OVERWRITE INTO TABLE employee\_internal;

Create the external table and load the data:

jdbc:hive2://> CREATE EXTERNAL TABLE employee\_external

. . . . . . .> (

. . . . . . .> name string,

. . . . . . .> work\_place ARRAY<string>,

. . . . . . .> sex\_age STRUCT<sex:string,age:int>,

. . . . . . .> skills\_score MAP<string,int>,

. . . . . . .> depart\_title MAP<STRING,ARRAY<STRING>>

. . . . . . .> )

. . . . . . .> COMMENT 'This is an external table'

. . . . . . .> ROW FORMAT DELIMITED

. . . . . . .> FIELDS TERMINATED BY '|'

. . . . . . .> COLLECTION ITEMS TERMINATED BY ','

. . . . . . .> MAP KEYS TERMINATED BY ':'

. . . . . . .> STORED AS TEXTFILE

. . . . . . .> **LOCATION '/user/dayongd/employee';**

No rows affected (1.332 seconds)

jdbc:hive2://> LOAD DATA LOCAL INPATH '/home/hadoop/employee.txt'. . .

. . . .> OVERWRITE

INTO TABLE employee\_external;

**What will happen if folder does not exist in the Location Property.?**

If the folder in the path does not exist in the LOCATION property, Hive will create that folder. If there is another folder inside the folder specified in the LOCATION property, Hive will NOT report errors when creating the table, but will report an error when querying the table.

**What are the other format supported for STORE AS?**

For the STORE AS property, it is set to **AS TEXTFILE** by default. Other file format values, such as **SEQUENCEFILE**, **RCFILE**, **ORC**, **AVRO** (since Hive 0.14.0), and **PARQUET** (since Hive 0.13.0) can also be specified.

**What is the Create the table as Select (CTAS) and its restrictions.?**

jdbc:hive2://> CREATE TABLE ctas\_employee

. . . . . . .> AS SELECT \* FROM employee\_external;

No rows affected (1.562 seconds)

CTAS copies the data as well as table definitions. The table created by CTAS is atomic; this means that other users do not see the table until all the query results are populated.

CTAS has the following restrictions:

The table created cannot be a partitioned table

The table created cannot be an external table

The table created cannot be a list bucketing table

A CTAS statement will trigger a map job for populating the data;

**What will the Alter the table’s statement to rename the table.?**

jdbc:hive2://> ALTER TABLE cte\_employee RENAME TO c\_employee;

No rows affected (0.237 seconds)

**What will be the Alter the table’s file format statement?**

jdbc:hive2://> ALTER TABLE c\_employee SET FILEFORMAT RCFILE;

No rows affected (0.235 seconds)

**What will the Alter the table’s location statement?**

jdbc:hive2://> ALTER TABLE c\_employee

. . . . . . .> SET LOCATION

. . . . . . .> 'hdfs://localhost:8020/user/dayongd/employee';

No rows affected (0.169 seconds)

**What will be Alter the table’s enable/disable protection to NO\_DROP, which prevents a table from being dropped, or OFFLINE, which prevents data (not metadata) in a table from being queried?**

jdbc:hive2://> ALTER TABLE c\_employee ENABLE NO\_DROP;

jdbc:hive2://> ALTER TABLE c\_employee DISABLE NO\_DROP;

jdbc:hive2://> ALTER TABLE c\_employee ENABLE OFFLINE;

jdbc:hive2://> ALTER TABLE c\_employee DISABLE OFFLINE;

**What will be the Alter the table’s concatenation to merge small files into larger files.?**

**--Convert to the file format supported**

jdbc:hive2://> ALTER TABLE c\_employee SET FILEFORMAT ORC;

No rows affected (0.160 seconds)

**--Concatenate files**

jdbc:hive2://> ALTER TABLE c\_employee CONCATENATE;

No rows affected (0.165 seconds)

**--Convert to the regular file format**

jdbc:hive2://> ALTER TABLE c\_employee SET FILEFORMAT TEXTFILE;

No rows affected (0.143 seconds)

In Hive release 0.8.0, RCFile is added to support fast block-level merging of small RCFiles using the CONCATENATE command. In Hive release 0.14.0 ORC, the files that are added support fast stripe-level merging of small ORC files using the CONCATENATE command. Other file formats are not supported yet. In case of RCFiles, the merge

happens at block level and ORC files merge at stripe level thereby avoiding the overhead of decompressing and decoding the data. MapReduce is triggered when performing concatenation.

**What is Hive Partitions?**

By default, a simple query in Hive scans the whole Hive table. This slows down the performance when querying a large-size table. The issue could be resolved by creating Hive partitions, which is very similar to what’s in the RDBMS. In Hive, each partition corresponds to a predefined partition column(s) and stores it as a subdirectory in the table’s directory in HDFS. When the table gets queried, only the required partitions (directory) of data in the table are queried, so the I/O and time of query is greatly reduced.

**What is Hive Buckets?**

Besides partition, bucket is another technique to cluster datasets into more manageable parts to optimize query performance. Different from partition, the bucket corresponds to segments of files in HDFS. For example, the employee\_partitioned table from the previous section uses the year and month as the top-level partition. If there is a further request to use the employee\_id as the third level of partition, it leads to many deep and small partitions and directories. For instance, we can bucket the employee\_partitioned table using employee\_id as the bucket column. The value of this column will be hashed by a user-defined number into buckets. The records with the same employee\_id will always be stored in the same bucket (segment of files). By using buckets, Hive can easily and efficiently do sampling and map side joins .

**--Create the buckets table**

**jdbc:hive2://> CREATE TABLE employee\_id\_buckets**

**. . . . . . .> (**

**. . . . . . .> name string,**

**. . . . . . .> employee\_id int,**

**. . . . . . .> work\_place ARRAY<string>,**

**. . . . . . .> sex\_age STRUCT<sex:string,age:int>,**

**. . . . . . .> skills\_score MAP<string,int>,**

**. . . . . . .> depart\_title MAP<string,ARRAY<string >>**

**. . . . . . .> )**

**. . . . . . .> CLUSTERED BY (employee\_id) INTO 2 BUCKETS**

**. . . . . . .> ROW FORMAT DELIMITED**

**. . . . . . .> FIELDS TERMINATED BY '|'**

**. . . . . . .> COLLECTION ITEMS TERMINATED BY ','**

**. . . . . . .> MAP KEYS TERMINATED BY ':';**

**No rows affected (0.104 seconds)**

Bucketing has close dependency on the underlying data loaded. To properly load data to a bucket table, we need to either set the maximum number of reducers to the same number of buckets specified in the table creation (for example, 2) or enable enforce bucketing as follows:

**jdbc:hive2://> set map.reduce.tasks = 2;**

**No rows affected (0.026 seconds)**

**jdbc:hive2://> set hive.enforce.bucketing = true;**

**No rows affected (0.002 seconds)**

To populate the data to the bucket table, we cannot use LOAD keywords such as what was done in the regular tables since LOAD does not verify the data against the metadata. Instead, INSERT should be used to populate the bucket table as follows:

**jdbc:hive2://> INSERT OVERWRITE TABLE employee\_id\_buckets**

**. . . . . . .> SELECT \* FROM employee\_id;**

**No rows affected (75.468 seconds)**

**--Verify the buckets in the HDFS**

**-bash-4.1$ hdfs dfs -ls /user/hive/warehouse/employee\_id\_buckets**

**Found 2 items**

**-rwxrwxrwx 1 hive hive 900 2014-11-02 10:54**

**/user/hive/warehouse/employee\_id\_buckets/000000\_0**

**-rwxrwxrwx 1 hive hive 582 2014-11-02 10:54**

**/user/hive/warehouse/employee\_id\_buckets/000001\_0**

**What is hive.fetch.task.conversion = more property ?**

SELECT \* scans the whole table/file without triggering MapReduce jobs, so it runs faster than SELECT <column\_name>. Since Hive 0.10.0, the simple SELECT statements, such as SELECT <column\_name> FROM

<table\_name> LIMIT n, can also avoid triggering the MapReduce job if the Hive fetch task conversion is enabled by setting hive.fetch.task.conversion = more.

**What is “*Correlating expression cannot contain unqualified column references”* error in Hive?**

If the alias (see the following example for the employee table) is not specified before columns (name) in the WHERE condition, Hive will report the error Correlating expression cannot contain unqualified column references. This is a limitation of the Hive subquery.

**jdbc:hive2://> SELECT name, sex\_age.sex AS sex**

**. . . . . . .> FROM employee a**

**. . . . . . .> WHERE a.name IN**

**. . . . . . .> (SELECT name FROM employee**

**. . . . . . .> WHERE sex\_age.sex = 'Male'**

**. . . . . . .> );**

**How to skip header while loading the csv file in the Hive table ?**

CREATE TABLE `warehouse`(`w\_warehouse\_sk` int,

`w\_warehouse\_id` string,

`w\_warehouse\_name` string,

`w\_warehouse\_sq\_ft` int,

`w\_street\_number` string,

`w\_street\_name` string,

`w\_street\_type` string,

`w\_suite\_number` string,

`w\_city` string,

`w\_county` string,

`w\_state` string,

`w\_zip` string,

`w\_country` string,

`w\_gmt\_offset` decimal(5,2)) row format delimited fields terminated by ',' stored as textfile **tblproperties("skip.header.line.count"="1")** ;

LOAD DATA Local INPATH '/home/cloud/data.csv' OVERWRITE INTO TABLE warehouse;

**Why hive only supports Equal Join instead of unequal Join?**

Hive only supports equal JOIN instead of unequal JOIN, because unequal JOIN is difficult to be

converted to MapReduce jobs.

**Can JOIN operation be performed among more than two tables ?**

The JOIN operation can be performed among more tables (three tables in this case),as follows:

**jdbc:hive2://> SELECT emp.name, empi.employee\_id, emph.sin\_number**

**. . . . . . .> FROM employee emp**

**. . . . . . .> JOIN employee\_hr emph ON emp.name = emph.name**

**. . . . . . .> JOIN employee\_id empi ON emp.name = empi.name;**

**What is Self-Join?**

Self-join is a special JOIN where one table joins itself. When doing such joins, a different alias should be given to distinguish the same table:

**jdbc:hive2://> SELECT emp.name**

**. . . . . . .> FROM employee emp**

**. . . . . . .> JOIN employee emp\_b**

**. . . . . . .> ON emp.name = emp\_b.name;**

**Why it is suggested for JOIN statements to put the big table at the end.?**

It is suggested for JOIN statements to put the big table right at the end for better performance as well as

avoiding **Out Of Memory** (**OOM**) exceptions, because the last table in the sequence is streamed through the reducers where the others are buffered in the reducer by default.

Also, a hint, such as /\*+STREAMTABLE (table\_name)\*/, can be specified to tell which table is streamed as follows:

**jdbc:hive2://> SELECT /\*+ STREAMTABLE(employee\_hr) \*/**

**. . . . . . .> emp.name, empi.employee\_id, emph.sin\_number**

**. . . . . . .> FROM employee emp**

**. . . . . . .> JOIN employee\_hr emph ON emp.name = emph.name**

**. . . . . . .> JOIN employee\_id empi ON emph.employee\_id =**

**empi.employee\_id;**

**What is MAPJOIN ?**

The MAPJOIN statement means doing the JOIN operation only by map without the reduce job. The MAPJOIN statement reads all the data from the small table to memory and broadcasts to all maps. During the map phase, the JOIN operation is performed by comparing each row of data in the big table with small tables against the join conditions. Because there is no reduce needed, the JOIN performance is improved. When the

**hive.auto.convert.join** setting is set to true, Hive automatically converts the JOIN to MAPJOIN at runtime if possible instead of checking the map join hint.

**jdbc:hive2://> SELECT /\*+ MAPJOIN(employee) \*/ emp.name, emph.sin\_number**

**. . . . . . .> FROM employee emp**

**. . . . . . .> CROSS JOIN employee\_hr emph WHERE emp.name <> emph.name;**

**What is LOAD command in Hive?**

To *move* data in Hive, it uses the LOAD keyword. Move here means the original data is moved to the target table/partition and does not exist in the original place anymore.

The LOCAL keyword specifies where the files are located in the host. If the LOCAL keyword is not specified, the files are loaded from the full **Uniform Resource Identifier** (**URI**) specified after INPATH or the value from the **fs.default.name** Hive property by default. The OVERWRITE keyword is used to decide whether to append or

replace the existing data in the target table/partition.

**The examples to load files into Hive tables:**

**Load local data to the Hive table:**

**jdbc:hive2://> LOAD DATA LOCAL INPATH**

**. . . . . . .> '/home/dayongd/Downloads/employee\_hr.txt'**

**. . . . . . .> OVERWRITE INTO TABLE employee\_hr;**

**No rows affected (0.436 seconds)**

**Load local data to the Hive partition table:**

**jdbc:hive2://> LOAD DATA LOCAL INPATH**

**. . . . . . .> '/home/dayongd/Downloads/employee.txt'**

**. . . . . . .> OVERWRITE INTO TABLE employee\_partitioned**

**. . . . . . .> PARTITION (year=2014, month=12);**

**No rows affected (0.772 seconds)**

**Load HDFS data to the Hive table using the default system path:**

**jdbc:hive2://> LOAD DATA INPATH**

**. . . . . . .> '/user/dayongd/employee/employee.txt'**

**. . . . . . .> OVERWRITE INTO TABLE employee;**

**No rows affected (0.453 seconds)**

**Load HDFS data to the Hive table with full URI:**

**jdbc:hive2://> LOAD DATA INPATH**

**. . . . . . .>**

**'hdfs://[dfs\_host]:8020/user/dayongd/employee/employee.txt'**

**. . . . . . .> OVERWRITE INTO TABLE employee;**

**No rows affected (0.297 seconds)**

**What is dynamic Partition?**

**What is IMPORT and EXPORT in HIVE?**

When working with Hive, sometimes we need to migrate data among different environments. Or we may need to back up some data. Since Hive 0.8.0, EXPORT and IMPORT statements are available to support the import and export of data in HDFS for data migration or backup/restore purposes.

The EXPORT statement will export both data and metadata from a table or partition.

Metadata is exported in a file called **\_metadata**. Data is exported in a subdirectory called **data**:

**jdbc:hive2://> EXPORT TABLE employee TO '/user/dayongd/output3';**

**No rows affected (0.19 seconds)**

After EXPORT, we can manually copy the exported files to other Hive instances or use Hadoop **distcp** commands to copy to other HDFS clusters. Then, we can import the data in the following manner:

**Import data to a table with the same name. It throws an error if the table exists:**

jdbc:hive2://> IMPORT FROM '/user/dayongd/output3';

Error: Error while compiling statement: FAILED: SemanticException

[Error 10119]: Table exists and contains data files

(state=42000,code=10119)

**Import data to a new table:**

jdbc:hive2://> IMPORT TABLE empolyee\_imported FROM

. . . . . . .> '/user/dayongd/output3';

No rows affected (0.788 seconds)

**Import data to an external table, where the LOCATION property is optional:**

jdbc:hive2://> IMPORT EXTERNAL TABLE empolyee\_imported\_external

. . . . . . .> FROM '/user/dayongd/output3'

. . . . . . .> LOCATION '/user/dayongd/output4' ;

No rows affected (0.256 seconds)

**Export and import partitions:**

jdbc:hive2://> EXPORT TABLE employee\_partitioned partition

. . . . . . .> (year=2014, month=11) TO '/user/dayongd/output5';

No rows affected (0.247 seconds)

jdbc:hive2://> IMPORT TABLE employee\_partitioned\_imported

. . . . . . .> FROM '/user/dayongd/output5';

No rows affected (0.14 seconds)

**What is ORDER BY AND SORT BY in HIVE? Why Sort by should be used instead of Order by?**

**ORDER BY (ASC|DESC):** A sorted order is maintained across all of the output from every reducer. It performs the global sort using only one reducer, so it takes a longer time to return the result.

jdbc:hive2://> SELECT name FROM employee ORDER BY NAME DESC;

**SORT BY (ASC|DESC):** This indicates which columns to sort when ordering the reducer input records. This means it completes sorting before sending data to the reducer. The SORT BY statement does not perform a global sort and only makes sure data is locally sorted in each reducer unless we set **mapred.reduce.tasks=1**. In this

case, it is equal to the result of ORDER BY. It can be used as follows:

**--Use more than 1 reducer**

**jdbc:hive2://> SET mapred.reduce.tasks = 2;**

**No rows affected (0.001 seconds)**

**jdbc:hive2://> SELECT name FROM employee SORT BY NAME DESC;**

**What is DISTRIBUTE BY in HIVE?**

Rows with matching column values will be partitioned to the same reducer. When used alone, it does not guarantee sorted input to the reducer. The DISTRIBUTE BY statement is similar to GROUP BY in RDBMS in terms of deciding which reducer to distribute the mapper output to. **When using with SORT BY, DISTRIBUTE BY must be specified before the SORT BY statement**. And, **the column used to distribute must appear in the select column list**. It can be used as follows:

**jdbc:hive2://> SELECT name**

**. . . . . . .> FROM employee\_hr DISTRIBUTE BY employee\_id;**

**Error: Error while compiling statement: FAILED: SemanticException**

**[Error 10004]: Line 1:44 Invalid table alias or column reference**

**'employee\_id': (possible column names are: name)**

**(state=42000,code=10004)**

**jdbc:hive2://> SELECT name, employee\_id**

**. . . . . . .> FROM employee\_hr DISTRIBUTE BY employee\_id;**

**+----------+--------------+**

**| name | employee\_id |**

**+----------+--------------+**

**| Lucy | 103 |**

**| Steven | 102 |**

**| Will | 101 |**

**| Michael | 100 |**

**+----------+--------------+**

**4 rows selected (38.92 seconds)**

**--Used with SORT BY**

**jdbc:hive2://> SELECT name, employee\_id**

**. . . . . . .> FROM employee\_hr**

**. . . . . . .> DISTRIBUTE BY employee\_id SORT BY name;**

**+----------+--------------+**

**| name | employee\_id |**

**+----------+--------------+**

**| Lucy | 103 |**

**| Michael | 100 |**

**| Steven | 102 |**

**| Will | 101 |**

**+----------+--------------+**

**4 rows selected (38.01 seconds)**

**What is ClUSTER BY in Hive?**

This is a shorthand operator to perform DISTRIBUTE BY and SORT BY operations on the same group of columns. And, it is sorted locally in each reducer.

The CLUSTER BY statement does not support ASC or DESC yet. **Compared to ORDER BY, which is globally sorted, the CLUSTER BY operation is sorted in each distributed group.** To fully utilize all the available reducers when doing a global sort, we can do CLUSTER BY first and then ORDER BY. This can be used as follows:

**jdbc:hive2://> SELECT name, employee\_id**

**. . . . . . .> FROM employee\_hr CLUSTER BY name;**

**What are the configuration parameters need to set to set the transaction support in Hive?**

SET hive.support.concurrency = true;

SET hive.enforce.bucketing = true;

SET hive.exec.dynamic.partition.mode = nonstrict;

SET hive.txn.manager = org.apache.hadoop.hive.ql.lockmgr.DbTxnManager;

SET hive.compactor.initiator.on = true;

SET hive.compactor.worker.threads = 1;

For now, all the transactions are autocommuted and only support data in the **Optimized Row Columnar**

**(ORC)** file format and in bucketed tables.

# **What is the Hive SQL COALESCE function, what does it do, and why on earth is it useful?**

There is a function called COALESCE that exists in Hive SQL. It has the following syntax.

**COALESCE(field1, field2, … , fieldn)**

**What does it do?**

Of the fields above (field1, field2, … , fieldn), for each record returned it returns the value of the first field of the ones listed in the COALESCE function that is not NULL. If all of the fields are NULL, it returns NULL.

You use it to create proxy values for NULLs in a field.

It is very useful when you have dirty data and have to use another field as an approximation of the dirty data’s actual value. If you have a field that is full of NULLs, you can use another field to put values in for those NULLs that you think provide a good approximate value of what should be there.

**What is Aggregation Condition – HAVING ?**

**HAVING is added to support the conditional filtering of GROUP BY results. By using HAVING, we can avoid using a subquery after GROUP BY. The following is an example:**

**jdbc:hive2://> SELECT sex\_age.age FROM employee**

**. . . . . . .> GROUP BY sex\_age.age HAVING count(\*)<=1;**

**If we do not use HAVING, we can use a subquery for instance as follows:**

**jdbc:hive2://> SELECT a.age**

**. . . . . . .> FROM**

**. . . . . . .> (SELECT count(\*) as cnt, sex\_age.age**

**. . . . . . .> FROM employee GROUP BY sex\_age.age**

**. . . . . . .> ) a WHERE a.cnt<=1;**

**IMPORTANT:- What is Windowing Functions in Hive?**

**IMPORTANT:- What is SAMPLING in Hive?**

**What is EXPLAIN and ANALYZE statements in Hive?**

The EXPLAIN and ANALYZE statements that can be used as utilities to check and identify the performance of queries.

**The EXPLAIN Statement**

The EXPLAIN command return a query execution plan without running the query. We can use an EXPLAIN command for queries if we have a doubt or a concern about performance.

**The syntax for EXPLAIN is as follows:**

**EXPLAIN [EXTENDED|DEPENDENCY|AUTHORIZATION] hive\_query**

* **EXTENDED**: This provides additional information for the operators in the plan, such as

file pathname and abstract syntax tree.

* **DEPENDENCY**: This provides a JSON format output that contains a list of tables and

partitions that the query depends on. It is available since HIVE 0.10.0.

* **AUTHORIZATION**: This lists all entities needed to be authorized including input and

output to run the Hive query and authorization failures, if any. It is

available since HIVE 0.14.0.

A typical query plan contains the following three sections.

**Abstract syntax tree (AST):** Hive uses a pacer generator called ANTLR to automatically generate a tree of syntax for HQL.

**Stage dependencies:** This lists all dependencies and number of stages used to run the query.

**Stage plans:** It contains important information, such as operators and sort orders, for running the job.

**The ANALYZE Statement**

Hive statistics are a collection of data that describe more details, such as the number of rows, number of files, and raw data size, on the objects in the Hive database. Statistics is a metadata of Hive data. Hive supports statistics at the table, partition, and column level.

These statistics serve as an input to the **Hive Cost-Based Optimizer (CBO)**, which is an optimizer to pick the query plan with the lowest cost in terms of system resources required to complete the query.

The statistics are gathered through the ANALYZE statement since Hive 0.10.0 on tables, partitions, and columns as given in the following examples:

**jdbc:hive2://> ANALYZE TABLE employee COMPUTE STATISTICS;**

No rows affected (27.979 seconds)

jdbc:hive2://> **ANALYZE TABLE employee\_partitioned**

. . . . . . .> PARTITION(year=2014, month=12) COMPUTE STATISTICS;

No rows affected (45.054 seconds)

jdbc:hive2://> **ANALYZE TABLE employee\_id COMPUTE STATISTICS**

**. . . . . . .> FOR COLUMNS employee\_id;**

No rows affected (41.074 seconds)

Once the statistics are built, we can check the statistics by the DESCRIBE EXTENDED/FORMATTED statement. From the table/partition output, we can find the statistics information inside the parameters, such as parameters:{numFiles=1,

COLUMN\_STATS\_ACCURATE=true, transient\_lastDdlTime=1417726247, numRows=4,

totalSize=227, rawDataSize=223}). The following is an example:

**jdbc:hive2://> DESCRIBE EXTENDED employee\_partitioned**

**. . . . . . .> PARTITION(year=2014, month=12);**

jdbc:hive2://> **DESCRIBE EXTENDED employee;**

… parameters:{numFiles=1, COLUMN\_STATS\_ACCURATE=true,

transient\_lastDdlTime=1417726247, numRows=4, totalSize=227,

rawDataSize=223}).

**What is the property to persist the Hive Statistics in the Metastore to avoid computing them every time.?**

jdbc:hive2://> SET hive.stats.autogather=ture;

**How many types of logs available in Hive?**

There are two types of logs available in Hive:

1) **System log** :- The system log contains the Hive running status and issues. It is configured in

**{HIVE\_HOME}/conf/hive-log4j.properties**. The following three lines for Hive log can be found:

hive.root.logger=WARN,DRFA

hive.log.dir=/tmp/${user.name}

hive.log.file=hive.log

To modify the status, we can either modify the preceding lines in **hive-log4j.properties** (applies to all users) or set from the Hive CLI (only applies to the current user and current session) as follows:

**hive --hiveconf hive.root.logger=DEBUG,console**

2) **Job log :-** The job log contains Hive query information and is saved at the same place, **/tmp/${user.name}**, by default as one file for each Hive user session. We can override it in **hive-site.xml** with the **hive.querylog.location** property. If a Hive query generates MapReduce jobs, those logs can also be viewed through the Hadoop JobTracker Web UI.

**What are the features available for the Performance Optimizations in HIVE?**

Since Hive 0.13.0, Hive includes the following new features for performance optimizations:

* **TEZ** :- Tez is an application framework built on Yarn that can execute complex directed acyclic graphs (DAGs) for general data-processing tasks. Tez further splits map and reduce jobs into smaller tasks and combines them in a flexible and efficient way for execution. Tez is considered a flexible and powerful

successor to the MapReduce framework. To configure Hive to use Tez, we need to overwrite the following settings from the default MapReduce:

**SET hive.execution.engine=tez;**

* **Vectorization :-** Vectorization optimization processes a larger batch of data at the same time rather than one row at a time, thus significantly reducing computing overhead. Each batch consists of a column vector that is usually an array of primitive types. Operations are performed on the entire column vector, which improves the instruction pipelines and cache usage. Files must be stored in the **Optimized Row**

**Columnar** (**ORC**) format in order to use vectorization.

**To enable vectorization, we need to do the following setting:**

**SET hive.vectorized.execution.enabled=true;**

**What is the property to set the Default File Format? What is the Default File Format?**

SET hive.default.fileformat=<File\_Format> --default file format for table

TEXTFILE is the default file format.

**How we can load the data in other file format other than TEXTFILE?**

We can load a text file directly to a table with the TEXTFILE format. To load data to the table with other file formats, we need to load the data to a TEXTFILE format table first.

Then, use **INSERT OVERWRITE TABLE <target\_file\_format\_table> SELECT \* FROM**

**<text\_format\_source\_table>** to convert and insert the data to the file format as expected.

**What is Hadoop Archive File (HAR)?**

**Hadoop Archive File** (**HAR**) is another type of file format to pack HDFS files into archives. This is an option (not a good option) for storing a large number of small-sized files in HDFS, as storing a large number of small-sized files directly in HDFS is not very efficient.

**How we can compress the intermediate files produced by Hive between multiple MapReduce Jobs?**

we need to set the following property (false by default) in the Hive CLI or the hive-site.xml file:

**jdbc:hive2://> SET hive.exec.compress.intermediate=true**

**what are the solutions to deal with too many small-file issues?**

* **Hadoop Archive and HAR :-** These are toolkits to pack small files.
* **SequenceFile Format :-** This is a format to compress small files to bigger files.
* **CombineFileInputFormat:-** A type of InputFormat to combine small files before

map and reduce processing. It is the default InputFormat for Hive.

* **HDFS federation** :- It makes namenodes extensible and powerful to manage more

files.

**What are the configurations required for merging files of query results to avoid recreating small files.?**

* **hive.merge.mapfiles**: This merges small files at the end of a map-only job. By default, it is true.
* **hive.merge.mapredfiles**: This merges small files at the end of a MapReduce job. Set it to true since its default is false.
* **hive.merge.size.per.task**: This defines the size of merged files at the end of the job. The default value is 256,000,000.
* **hive.merge.smallfiles.avgsize:** This is the threshold for triggering file merge. The default value is 16,000,000.

When the average output file size of a job is less than the value specified by **hive.merge.smallfiles.avgsize**, and both **hive.merge.mapfiles (for map-only jobs) and hive.merge.mapredfiles (for MapReduce jobs)** are set to true, Hive will start an additional MapReduce job to merge the output files into big files.

**What is JVM reuse?**

By default, Hadoop launches a new JVM for each map or reduce job and runs the map or reduce task in parallel. When the map or reduce job is a lightweight job running only for a few seconds, the JVM startup process could be a significant overhead. The MapReduce framework (version 1 only, not Yarn) has an option to reuse JVM by sharing the JVM to run mapper/reducer serially instead of parallel. JVM reuse applies to map or reduce tasks in the same job. Tasks from different jobs will always run in a separate JVM. To enable

the reuse, we can set the maximum number of tasks for a single job for JVM reuse using

the **mapred.job.reuse.jvm.num.tasks** property. **Its default value is 1**:

**jdbc:hive2://> SET mapred.job.reuse.jvm.num.tasks=5;**

**We can also set the value to –1 to indicate that all the tasks for a job will run in the same JVM.**