

# **Elastic MapReduce**

## **Basic EMR Issues**

### **Product Documentation**



## Copyright Notice

©2013-2024 Tencent Cloud. All rights reserved.

Copyright in this document is exclusively owned by Tencent Cloud. You must not reproduce, modify, copy or distribute in any way, in whole or in part, the contents of this document without Tencent Cloud's the prior written consent.

## Trademark Notice



All trademarks associated with Tencent Cloud and its services are owned by Tencent Cloud Computing (Beijing) Company Limited and its affiliated companies. Trademarks of third parties referred to in this document are owned by their respective proprietors.

## Service Statement

This document is intended to provide users with general information about Tencent Cloud's products and services only and does not form part of Tencent Cloud's terms and conditions. Tencent Cloud's products or services are subject to change. Specific products and services and the standards applicable to them are exclusively provided for in Tencent Cloud's applicable terms and conditions.

# Basic EMR Issues

Last updated : 2022-07-04 11:23:37

## How do I view task logs?

You can log in to any EMR server and run the following command to view task logs:

```
yarn logs -applicationId application_1507732460084_0057
```

Note :

You need to run this command with the `Hadoop` username. If it is a task of another user, you can add the `appOwner username` parameter.

To view the cause of a task exception, run the following command:

```
yarn logs -applicationId application_1507732460084_0057|grep -A20 Exception
```

## How do I adjust the computing resources of a cluster?

Cluster computing resources are determined by the following two configuration items in `yarn-site.xml` :

```
<property>
<name>yarn.nodemanager.resource.cpu-vcores</name>
<value>4</value>
</property>
<property>
<name>yarn.nodemanager.resource.memory-mb</name>
<value>14745</value>
</property>
```

By default, `cpu-vcores` is equal to number of CPU cores of the server, and `memory-mb` is equal to 91% of the memory size of the server. You can adjust them based on your actual needs, but if they are too large, there may be a risk of server failure.

## How do I fix an out of memory error?

If an out of memory error occurs when you are submitting a MapReduce task or running a SQL script through Hive, fix it by setting the following parameters:

```
set mapreduce.map.java.opts=-Xmx4096m;  
set mapreduce.reduce.java.opts=-Xmx4096m;
```

The memory parameter can be adjusted based on your actual computation needs. It can also be written in the `~/.hiverc` file in Hive and will be executed automatically when submitted.

## How do I estimate the cluster size?

Suppose that you need to run a SQL query. If 64 vcores and 128 GB memory are needed for getting the query result in the specified time period, and the business requires 10 concurrencies, then the required resources would be 640 vcores and 1,280 GB memory. If the server specification you are using is 24 cores and 48 GB memory, then you need around  $1280 / 48 = 27$  servers.

## How do I set up a fetch query in Hive?

The default query in Hive is as follows:

```
select * from tablename where a='1' limit 10;
```

The default query does not start a computation task. You can start a distributed query by adding the `set hive.fetch.task.conversion=none` parameter.

## How do I choose the cluster storage media?

An EMR cluster supports the following storage media: HDD local disk, SSD local disk, HDD cloud disk, SSD cloud disk, and COS. You can choose the most appropriate one based on your actual needs:

- If your application scenario is large-scale data warehouse analysis which is not very latency-sensitive, we recommend you use COS as the underlying storage.
- If you are very familiar with HDFS and the cost of migration to COS would be too high, you can use HDD cloud disks.
- If your application is a massive columnar database (HBase) which requires efficient writing and querying, we recommend you use SSD local or cloud disks.

## Can I log in to an EMR cluster over the public network?

You can log in to the cluster after enabling the public IP as instructed in [Logging in to Clusters](#).

## Does an EMR cluster support LDAP authentication?

LDAP authentication is subject to the product version. On v2.3.0 and later, it is supported and enabled by default and cannot be disabled. It is not supported on earlier versions.

## Can I add more Common nodes?

No.

### **Does EMR support ClickHouse version upgrade?**

At present, there are three releases in an EMR ClickHouse cluster, and ClickHouse has been upgraded in each release. For more information, see [Version Overview](#).

### **Will the big data components selected in the EMR console be automatically installed?**

Components can be directly installed in the console.

### **Does EMR HDFS enable WebHDFS by default?**

Yes.

### **Can I modify the project of an EMR cluster?**

No. If this affects your business, [submit a ticket](#) for assistance.

### **Can I run backend JAR packages on MapReduce nodes?**

When creating a cluster, you can use bootstrap actions to set a custom bootstrap script to achieve this (after node initialization, before cluster start, or after cluster start). For more information, see [Bootstrap Action](#).

### **How do I select cluster specifications based on business use cases?**

EMR provides six types of clusters for you to choose from based on your business needs. For more information, see [Business Evaluation](#).