# Elastic MapReduce

# Getting Started

# Product Documentation


©2013-2019 Tencent Cloud. All rights reserved.

Copyright Notice

©2013-2019 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

Tencent Cloud

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.

# Contents



©2013-2019 Tencent Cloud. All rights reserved.

# Getting Started
# Business Assessment

Last updated：2019-07-26 17:38:28

Please evaluate your business situations before using EMR:

## Selecting Model Specifications

EMR offers a wide variety of CVM instance models, including EMR Standard, EMR Compute, EMR High IO, EMR MEM-optimized, and EMR Big Data. If you need CPM models, please contact us by submitting a ticket.

You can choose the most appropriate model based on your business needs and budget.

- If you require low latency for offline computation, we recommend you to choose the model with a local disk.
- If you need to use the real-time database HBase, we recommend you to choose the EMR High IO model with a local SSD disk for optimal performance.

### Node Specification Recommendations

- **Master node**: A master node mainly performs cluster scheduling and task submission. It does not require high computing power; however, depending on the actual situation, it may require a larger memory. EMR Standard 4-core 8 GB, 4-core 16 GB, or higher-spec models are usually recommended.

- **Core node**: As a core node is used for computing and storage tasks, it has high requirements for CPU, memory, and disk. However, if your data is completely stored in COS, then the role of the core node is basically the same as that of a task node. In this case, a local disk is recommended to improve the I/O capability and get the computation results faster.

- **Task node**: As a task node is responsible for computation only and the data for computation comes from a core node or COS, its disk size does not need to be too large; however, it is recommended to select at least 500 GB for it to ensure the computation efficiency.

- **Common node**: Currently, the specification of a common node is EMR Standard 2-core 4 GB model with a local disk of 100 GB by default.

©2013-2019 Tencent Cloud. All rights reserved.

- **Router node**: A router node is mainly used to relieve the load of the master node and as a task submitter; therefore, it is recommended to select a model with larger memory, preferably not lower than the master node specification.

## Network and Security

To ensure the network security of the EMR cluster, it is placed in a VPC, and a security group policy is added to the VPC. In addition, to ensure that the web UI of Hadoop can be easily accessed, one of the master nodes is configured with a public IP which is billed by traffic. A router node has no public IPs by default. If you need one for it, you can bind it to an EIP in the CVM Console.

- A public IP is enabled for a master node when a cluster is created, but you can disable it based on the actual conditions.
- Enabling the public IP for the master node is mainly for SSH login and component viewing in the web UI.
- A master node with a public IP enabled is billed by traffic with a bandwidth of up to 5 Mbps. Once the cluster is created, you can make adjustments to its network in the console.

©2013-2019 Tencent Cloud. All rights reserved.

# Creating EMR Clusters

Last updated：2019-07-26 18:05:32

## Application Scenario

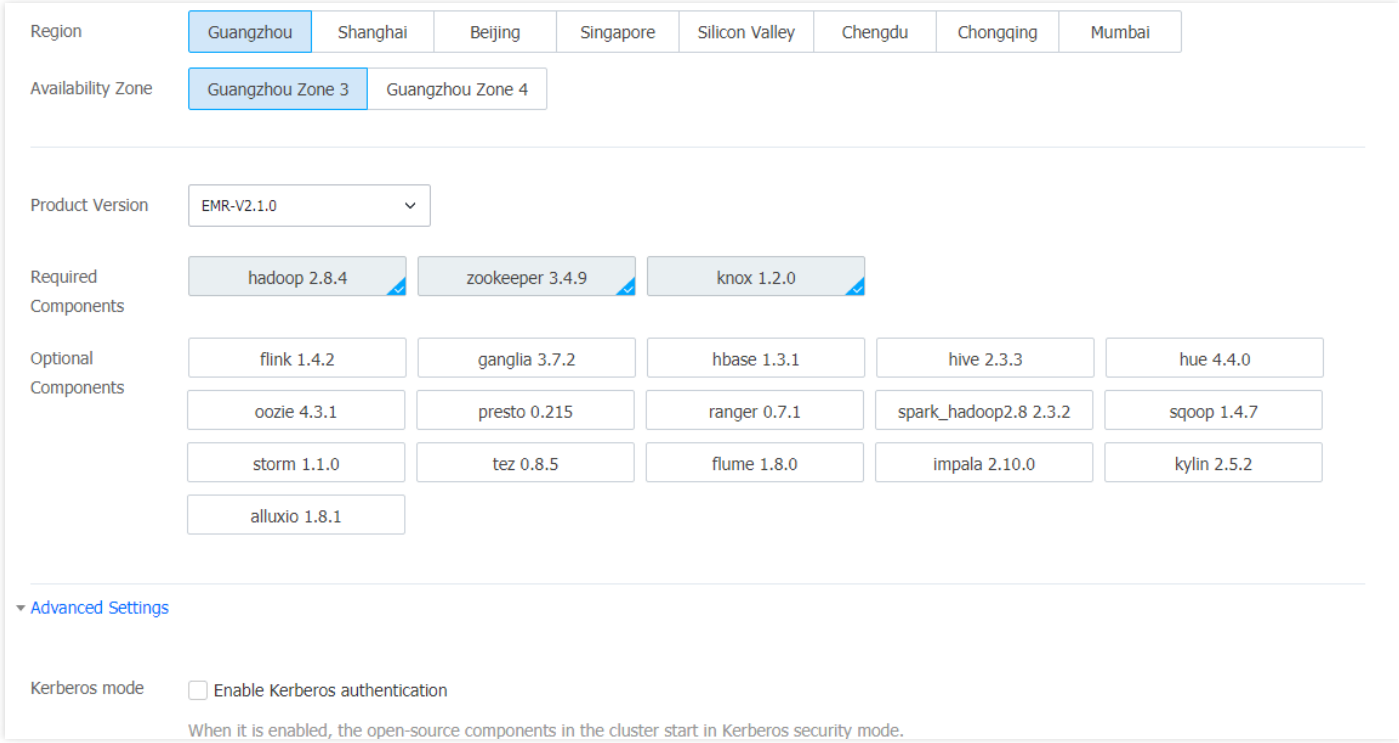
This document describes how to create an EMR cluster in the EMR Console.

## Directions

Log in to the EMR Console and click **Create Cluster** on the cluster list page.

### 1. AZ and Software Configuration

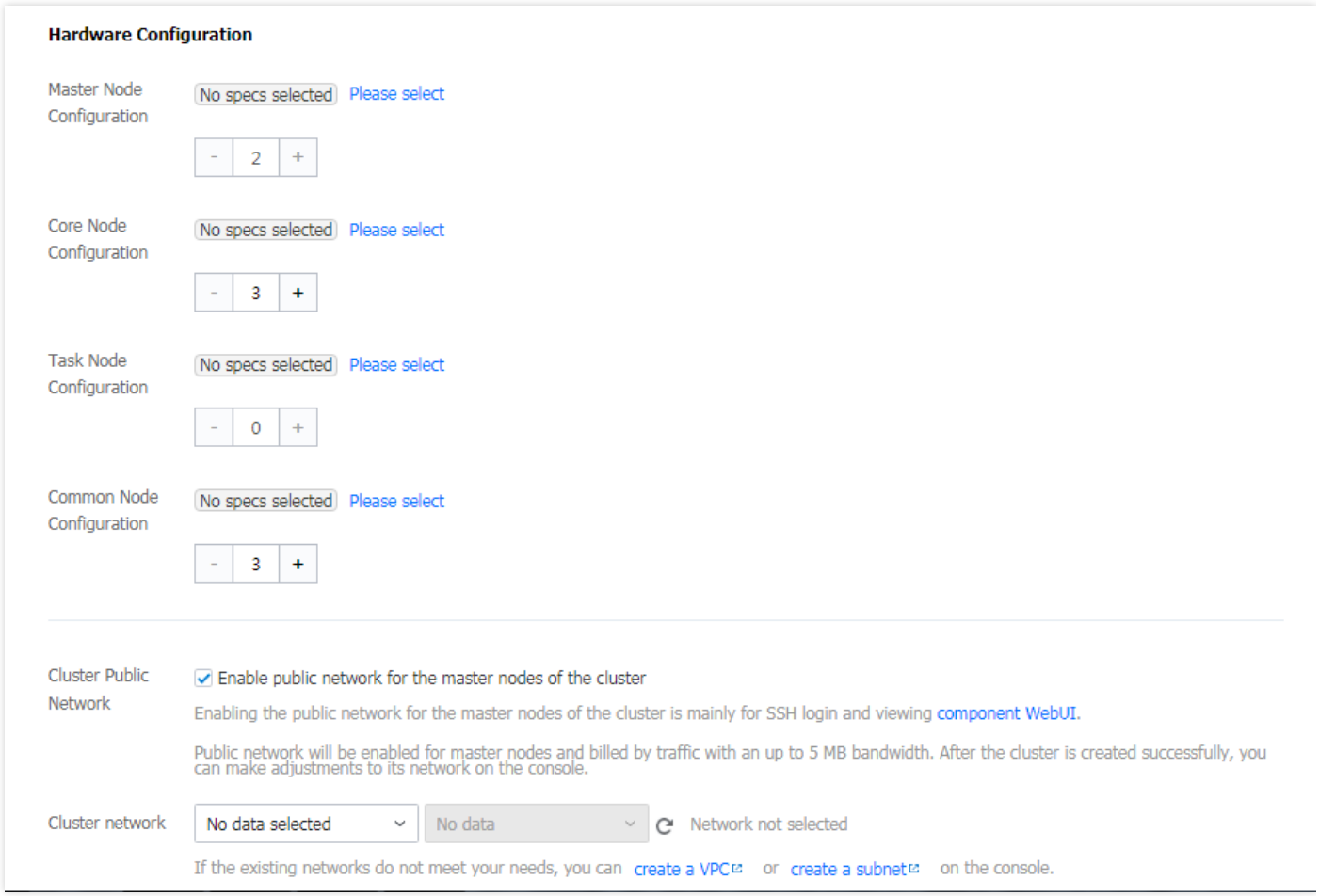
- **Region and AZ**: Currently supported regions include Guangzhou, Shanghai, Beijing, and Silicon Valley. Tencent Cloud products in different regions cannot communicate with each another over a private network.
- **Version and Components**: EMR recommends some commonly used combinations of components for Hadoop. You can also combine the components based on your needs.
- **Kerberos Secure Cluster**: This specifies whether to enable Kerberos authentication for the cluster. This feature is not required for individual users and disabled by default.




©2013-2019 Tencent Cloud. All rights reserved.

## 2. Hardware Configuration

- **High-availability (HA) cluster**: By default, once you've selected **Enable high availability**, 2 master nodes, at least 3 core nodes, and 3 common nodes will be enabled. For more information about node types, see Node Type Descriptions.
- **Hardware Configuration**: Node specification. EMR provides a variety of node specifications. You can choose your node type, core quantity, memory size, disk type, and disk size to meet your business needs.
- **Cluster Network**: To ensure the security of the EMR cluster, all nodes of the cluster are placed in a VPC; therefore, you need to set up a VPC before your can successfully create the EMR cluster.



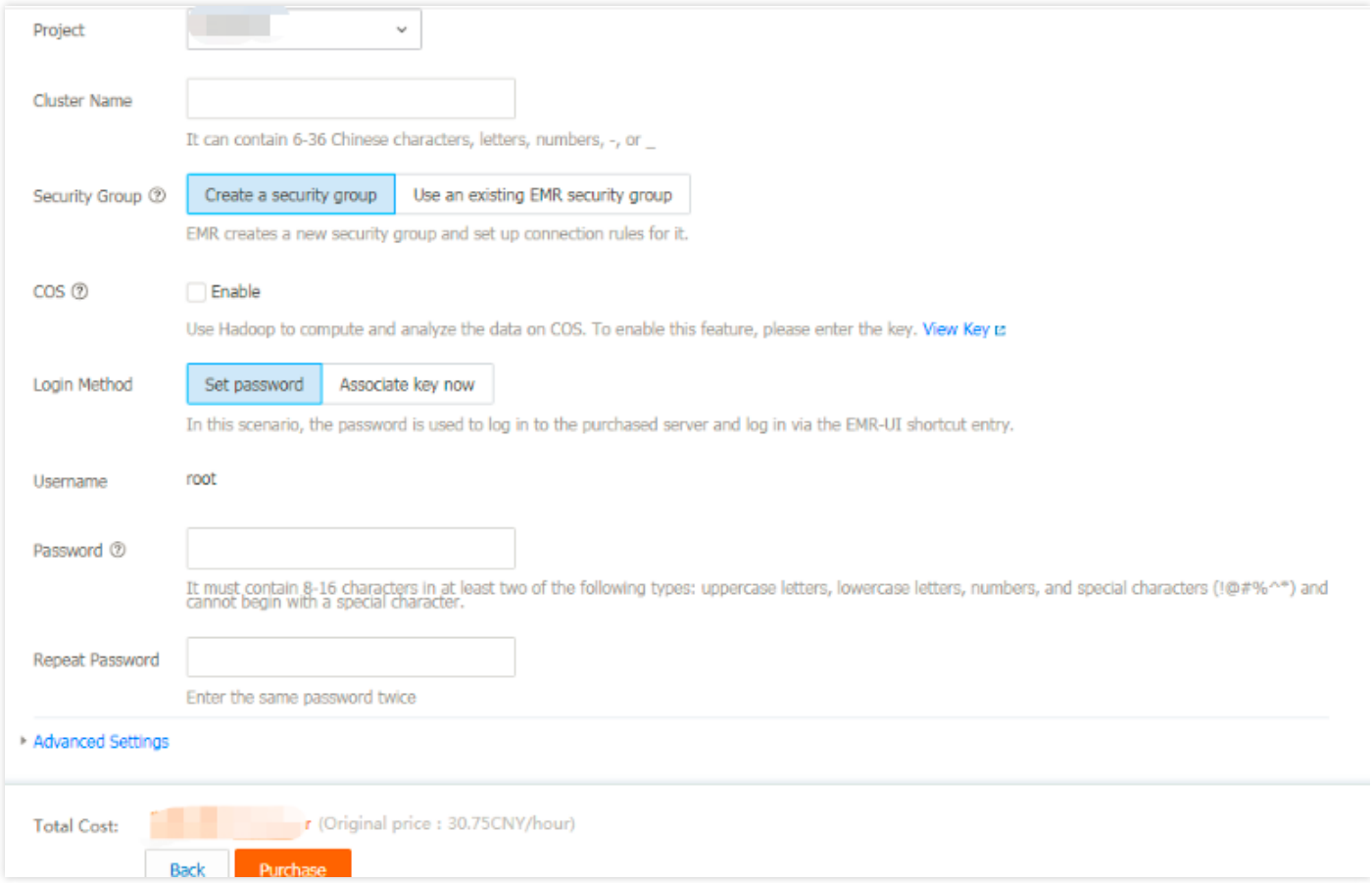

## 3. Basic Configuration

- **Cluster Name**: EMR clusters are differentiated by cluster name.
- **Security Group**: Security groups act as a virtual firewall and control inbound and outbound traffic for CVM instances. Create a new security group if you don't have one. Otherwise, choose an existing one.
  - Create a security group: EMR will create a security group that allows traffic going through ports 22 and 30001 as well as all traffic from the necessary private network IP range.

©2013-2019 Tencent Cloud. All rights reserved.

- Use an existing EMR security group: Select an existing EMR security group as the security group for your instance. Open ports 22 and 30001 as well as the required IP range for communication over the private network.

- **COS**: After COS is enabled, the EMR cluster can directly compute the data stored in COS so that the compute and storage can be separated, thus reducing the costs of big data processing. To ensure that EMR can access to your data stored in COS, please enter your COS API key.




## 4. Completing the Creation

After completing the configurations above, click **Purchase** to make the payment. Your EMR cluster will be automatically created once the payment is received. Wait for about 10 minutes, then you will find the cluster you just created in the EMR Console.

> To ensure your EMR cluster working properly, you can view, but please **do not** modify the configuration information of each instance in the CVM Console.

©2013-2019 Tencent Cloud. All rights reserved.

# Collaborator/Sub-account

Last updated：2019-07-26 17:38:21

You can grant permissions to each collaborator/sub-account in the console. For the authorization method, see CAM.

EMR currently provides three permission roles. You can search for the roles by the keyword “EMR” on the policy page in the CAM Console.

| Role | Permissions |
| --- | --- |
| QcloudEMRObserverAccess (observer access) | API for getting component configuration information.<br>API for getting monitoring information.<br>API for getting node information.<br>Get cluster list. |
| QcloudEMROperationAccess (OPS access) | Inherit the observer access.<br>Modify parameters and distribute configurations.<br>Restart services.<br>Change passwords. |
| QcloudEMRAdministratorAccess (admin access) | Inherit OPS access.<br>Restart the specified service.<br>Scale out.<br>Scale in.<br>Create a cluster.<br>Terminate the specified cluster.<br>Change passwords for the specified cluster. |

©2013-2019 Tencent Cloud. All rights reserved.

# Logging In to a Cluster

Last updated：2019-07-26 17:38:47

## Logging in Using a Remote Login Tool (on Windows)

This section uses Xshell as an example to describe how to log in to an EMR cluster with a password using a remote login tool on Windows.

### Applicable OS

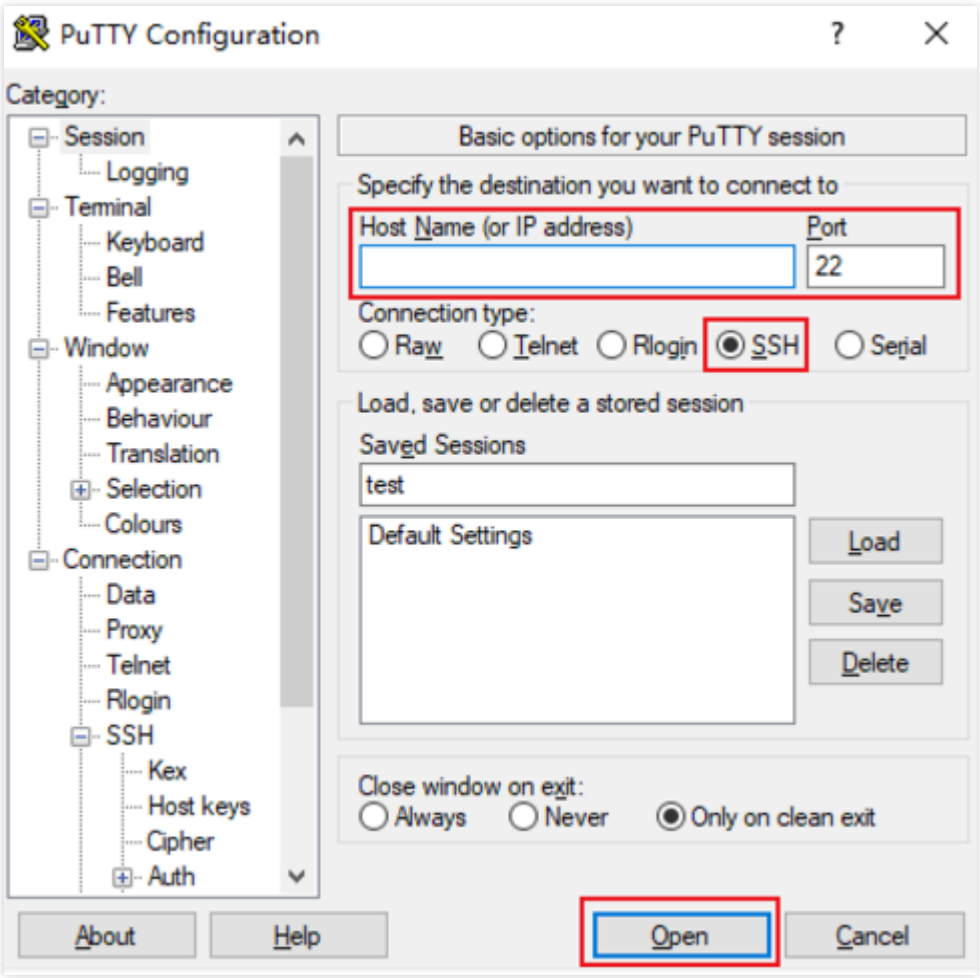
Windows

### Logging in with a Password

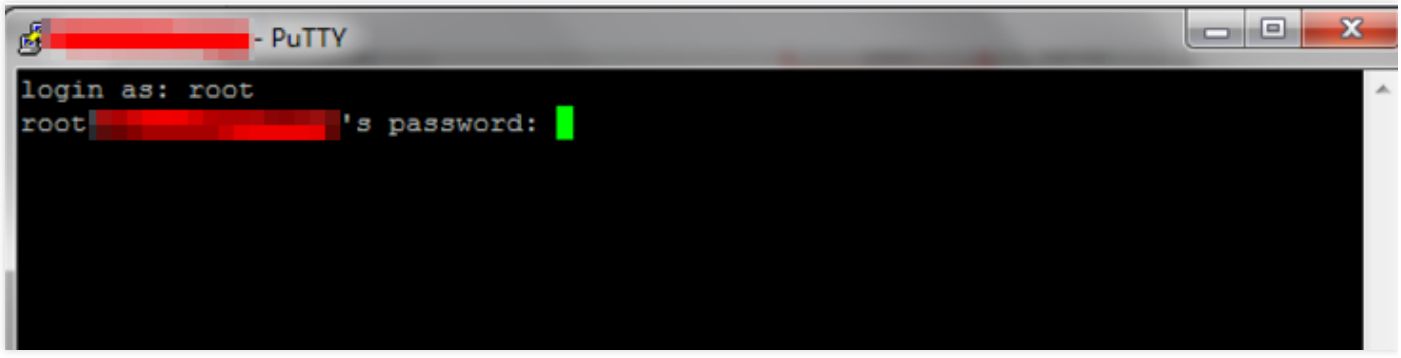
1. Download PuTTY (a remote login tool on Windows) here and install it.
2. Launch the PuTTY client, enter the following in the PuTTY Configuration window, and click **Open** to create a session as shown below:
   - **Host Name**: The public IP address of the EMR cluster, which can be viewed on the list page or details page in the EMR Console.
   - **Port**: The port number of the CVM instance, which has to be 22. (Please make sure that the port 22 in the CVM instance is opened. For more information, see Security Group and Network ACL).

©2013-2019 Tencent Cloud. All rights reserved.

- **Connect type**: Select **SSH**.




3. In the PuTTY session window, enter the obtained admin account and press Enter.
4. Enter the obtained login password and press Enter to log in as shown below:




# Log in Using SSH (on Linux or macOS)

This section describes how to log in to an EMR cluster using SSH on Linux or macOS.

©2013-2019 Tencent Cloud. All rights reserved.

### Applicable OS

Linux or macOS

### Logging in with a Password

1. On macOS, launch Terminal and run the following command. On Linux, run the following commands directly:

   ```
   ssh <username>@<hostname or IP address>
   ```

   - username: The admin account, such as root.
   - hostname or IP address: The public IP address or custom domain name of your EMR cluster.
2. Enter the obtained password (only the input but not the output is displayed here) and press Enter to log in.

©2013-2019 Tencent Cloud. All rights reserved.