

云数据库 Cassandra

用户指南

天翼云科技有限公司

目 录

目录

1. 产品概述	4
1.1 产品定义	4
1.2 基本概念	5
1.3 产品架构	5
1.4 应用场景	6
2. 快速入门	9
2.1 产品规格	9
2.2 计费方式	9
2.3 创建实例	9
2.4 连接实例	12
2.4.1 多语言 SDK 访问	13
2.4.2 Cqlsh 接入	14
2.5 高可用配置	14
2.6 可调一致性	14
2.7 数据迁入迁出	14
3. 操作指导	15
3.1 实例生命周期管理	15
3.1.1 变更实例	15

3.2	命令行	17
3.3	用户管理	18
3.3.1	新增数据库用户	18
3.3.2	密码重置	19
3.4	监控告警	19
3.5	数据库配置	21
3.6	节点管理	21
4.	最佳实践	21
5.	常见问题	22
5.1	包年包月和按需计费切换	22
5.2	实例开通后 VPC 切换	22
5.3	备份	22
5.4	实例可用性	23
5.5	可用区域	23

修订历史

版本号	日期	修订人	修订内容	修订原因

1. 产品概述

1.1 产品定义

天翼云数据库 Cassandra 版是一款高可靠的 NoSQL 分布式数据库服务，Always Online，单点故障不影响服务可用性；拥有优异的写入性能，能够支持很高的多客户线程并发写和突发的峰值。支持类 SQL 语法 CQL，开发者容易上手；一致性可配置，支持多种应用场景。100%兼容开源 Apache Cassandra。

优势	概况词	说明
易扩展	按需线性扩展	- 支持线性扩展节点，实现水平扩容，集群处理能力线性上升 - 支持磁盘扩容，实现按需扩容
高性能	优异写性能	- 拥有优异的写入性能，能够支持很高的多客户线程并发写性能和突发的峰值 - 线性扩容节点，集群处理能力可实现线性上升
高可用	Always Online	- 去中心化部署，无需担心关键节点宕机 - Always Online，单点故障完全不影响服务可用性 - 多副本，可调一致性级别，满足多种写入场景
易使用	容易上手、接入	- 支持类 SQL 语法的 CQL 语言，开发者容易上手 - 支持 Java、Python、Go、C++等十几种语言等原生客户端接入

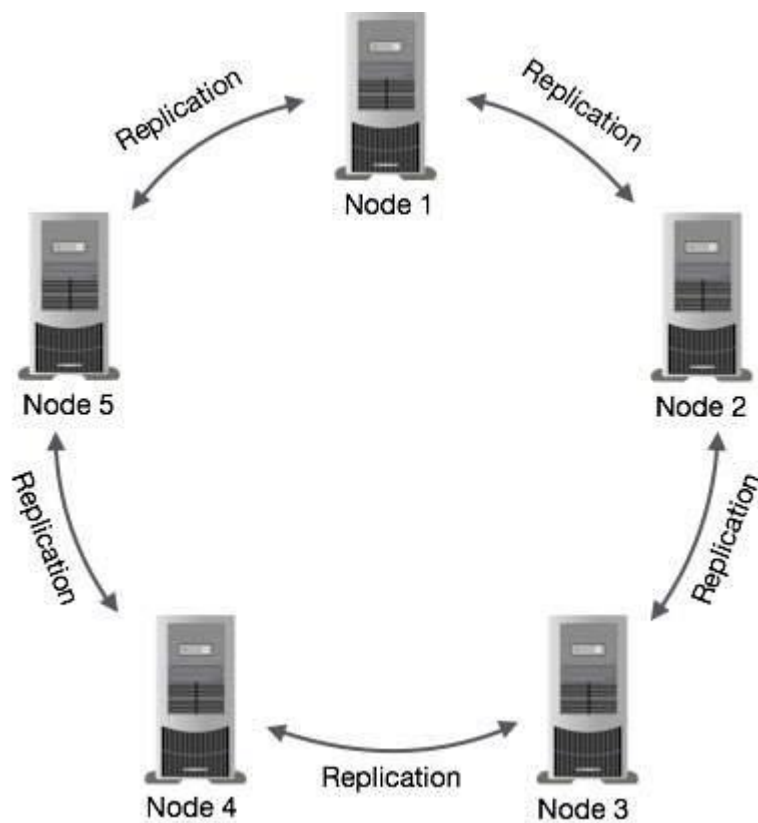
--	--	--

1.2 基本概念

集群	Cassandra 数据库集群包含多个 cassandra 节点，每个节点都包含一个或多个数据副本节点。当 cassandra 节点故障后，数据副本节点会接替故障节点持续性工作。Cassandra 以环形方式将节点分配在集群中，并为其分配数据。每台 Cassandra 节点都必须指定同一个 Cluster 名称。
KeySpace	数据最外层的容器，可以把它想象成 RDBMS 中的一个数据库，户可以在 keyspace 这个级别指定副本策略。
列	列” 由一个 name，一个相关的 value 和一个时间戳组成。name 和 value 可以是任何类型，name 不需要是字符串。一个列就是一个 Name-Value-Timestamp 集合
列族	可以想象成在关系数据库中一行所包含的一个 key 和一些列
超列族	就是一个列族，这个列族中每个列的值就是一个列的集合
Partitioner	定义了数据如何在集群中的节点分布，哪个节点应该存放数据的第一份副本。基本上就是个计算分区键 token 的哈希函数
副本策略	副本策略用来确定将副本存放在哪个节点上面，由用户在创建 keyspace 时指定
副本因子	副本数，表示数据在集群中存了几“份”（副本），由用户在创建 keyspace 时指定
CQL	cassandra 中提供的一种类 SQL 查询语言

1.3 产品架构

- 1、Cassandra 实例节点在同一 VPC 内（暂不支持跨 IDC 部署），每个节点部署一个 Cassandra 进程
- 2、所有节点都是对等的
- 3、客户端可与服务端每个节点连接



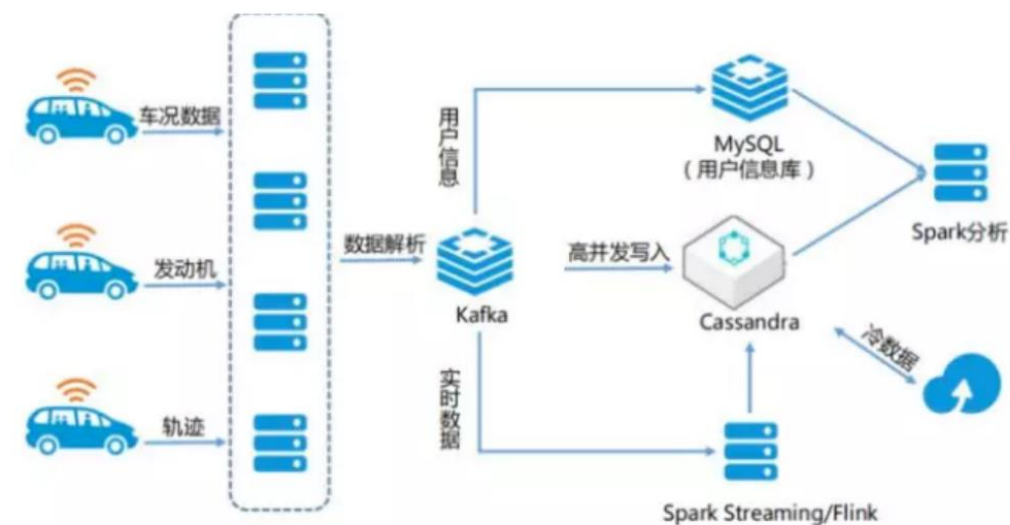
Cassandra 使用 Gossip 协议，以允许节点彼此通信并检测集群中的任何故障节点

1.4 应用场景

Cassandra 其特征是支持大并发低延时的访问需求、具备高可用和弹性扩容能力、灵活的数据模型以及优异的写吞吐非常适合大数据量的互联网在线应用、存储分析以及写密集场景。以下是具体应用：

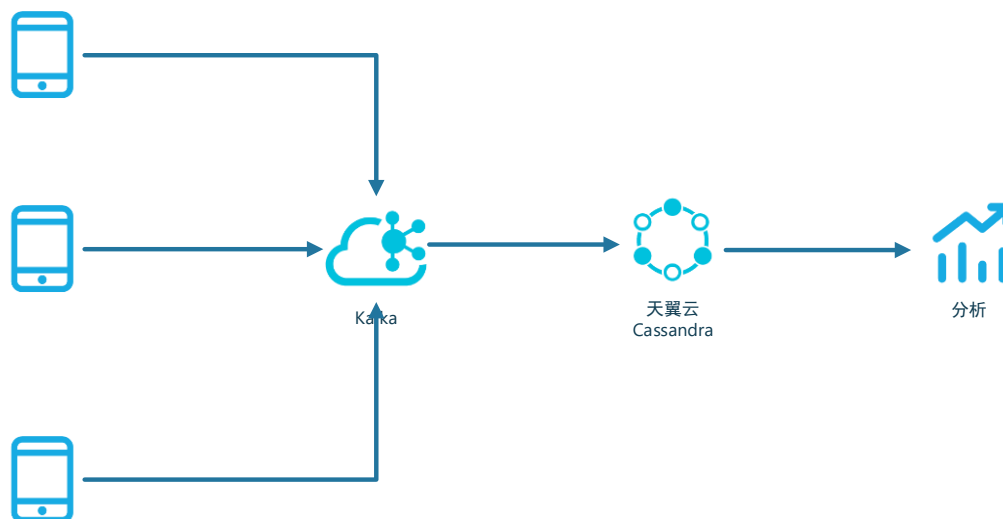
应用场景一

写密集、统计和分析型工作。Cassandra 是为优异的写吞吐量而特别优化的，能够支持很高的多客户线程并发写性能和突发的峰值，这些特性使得 Cassandra 能够很好支持写多于读的场景，例如 IOT 设备状态相应、用户状态更新、社交网络、建议/评价以及应用统计等。



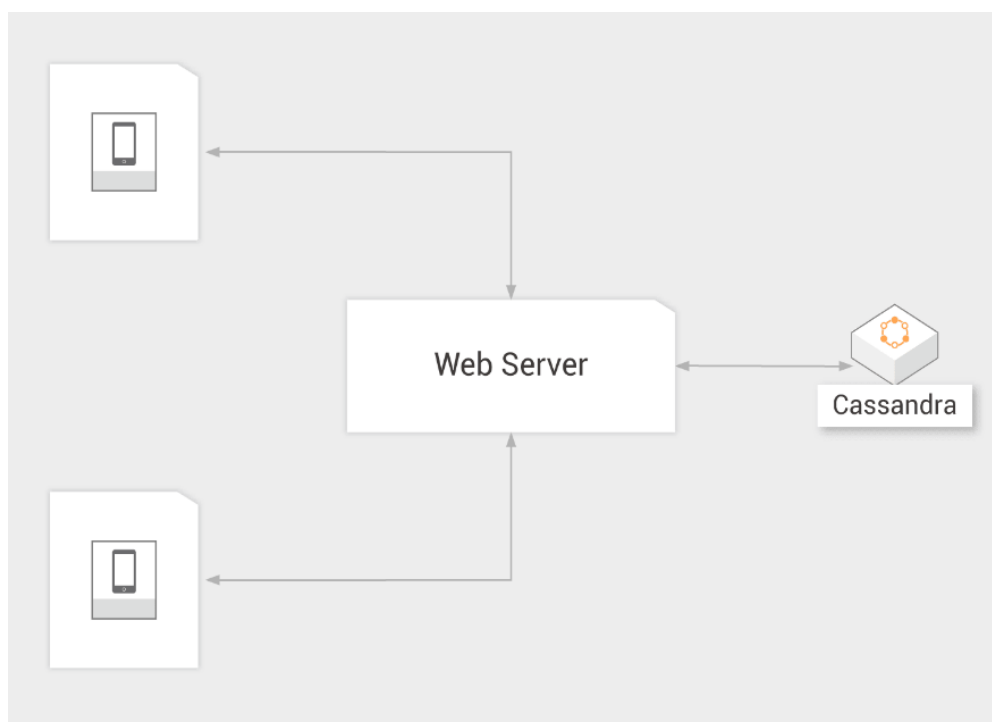
应用场景二

海量数据分析场景。天翼云 Cassandra 可以支持数十个节点的集群规模，适合大数据量的存储。在一些需要应用大量数据对用户行为进行分析的场景中，可以通过整合多种数据来源，存储用户行为数据，构建用户画像，实时存储在天翼云 Cassandra 中，提供大数据风控、推荐等服务。



应用场景三

在线应用场景。 Cassandra 能够支持大并发低延时的访问需求，具备高可用和弹性扩容能力，适合消息、feed 流、订单查询、网站等各种大数据量的互联网在线应用场景。



2. 快速入门

2.1 产品规格

产品 CPU/内存规格支持：2C8G、4C16G、8C32G、16C64G、32C128G

存储类型：普通 IO（SATA）

节点存储容量大小：不超过 6000G。

2.2 计费方式

计费方式	说明	注意事项
预付费 (包年包月)	预付费方式，在新建实例时就需要支付费用。 对于包月的实例，购买 1 年即可享受价格 85 折优惠，且购买的年数越长，折扣更低。	无法变更为按量付费。
按量付费	计费周期为 1 个小时。不足 1 小时，也按 1 小时收费。 账单生成后会自动从您的账户余额中扣除费用以结算账单。	实例可以随时释放，释放后不再收费。不可变更成包年包月

2.3 创建实例

- 1、登入天翼云门户
- 2、选择相应的资源池
- 3、进入服务列表，选择云数据库 Cassandra，进入控制台

实例列表页点击“订购实例”进入开通页进行实例创建。

* 计费模式:

包年/包月

按需计费

版本:

3.11

* 实例名称:

AnalyticPG-FZbk

购买完成后进行创建的实例名称。4~64位之间，必须以字母开头，区分大小写，可以包含字母、数字、中划线或下划线，不能包含其他特殊字符。

存储类型:

普通IO

规格:

2核8G

节点数量:

- 3 +

节点可选择范围3~12

节点存储空间 (GB):

100GB

1000GB

2000GB

3000GB

4000GB

5000GB

6000GB

-

100

+

数据库配置

密码设置:

立即设置

稍后设置

数据库用户名:

请输入

4~64位之间，必须以字母开头，区分大小写，可以包含字母、数字、中划线或下划线，不能包含其他特殊字符。

密码:

请输入密码

请再次确认密码

大写、小写、数字、特殊字符占三种，长度为8 - 26位；特殊字符为!@#\$\$%^&*()_+~=

网络

* 虚拟私有云: [刷新](#) [管理VPC](#)

* 子网:

* 安全组:

基础配置:

参数	描述
计费模式	“包年/包月”或“按需计费”。 - 包年包月用户选购完服务配置后，可以根据需要设置购买时长，系统会一次性按照购买价格对账户余额进行扣费。 - 按需付费,用户选购完服务配置后，无需设置购买时长，系统会根据消费时长对账户余额进行扣费。
版本	大版本 3.11，参考社区 3.11.5 版本
实例名称	对开通的实例命名，方便标识
存储类型	普通 IO，SATA 类型
节点数	实例包含的 Cassandra 节点数

参数	描述
节点存储空间	每个 Cassandra 节点的存储空间大小

数据库配置：

参数	描述
计费模式	“包年/包月”或“按需计费”。 - 包年包月用户选购完服务配置后，可以根据需要设置购买时长，系统会一次性按照购买价格对账户余额进行扣费。 - 按需付费,用户选购完服务配置后，无需设置购买时长，系统会根据消费时长对账户余额进行扣费。
管理员用户名	4-64 位之间，必须以字母开头，区分大小写，可以包含字母、数字、中划线或下划线，不能包含其他特殊字符
管理员用户名密码	大写、小写、数字、特殊字符占三种，长度为 8-26 位；特殊字符为!@#%&*()_+-=

2.4 连接实例

- 1、创建数据库用户名和密码；
- 2、获取访问地址，进入 Cassandra 控制台，实例管理-实例详情，查看连接地址
- 3、通过客户端进行访问

2.4.1 多语言 SDK 访问

支持 JAVA, python, Ruby, Node.js, PHP, C++ 和 C#/.NET 等多种客户端接入方式, 具体参考 github 社区文档。

以下 JAVA 连接示例：

Maven 依赖

```
<dependency>
  <groupId>com.datastax.cassandra</groupId>
  <artifactId>cassandra-driver-core</artifactId>
  <version>3.7.2</version>
</dependency>
```

```
public class Demo {

    public static void main(String[] args) {

        // 此处填写数据库连接点地址，控制台有几个就填几个。
        // 实际上 SDK 最终只会连上第一个可连接的连接点并建立控制连接，填写多个是为了防止单个节点
        // 挂掉导致无法连接数据库。
        // 此处无需关心连接点的顺序，因为 SDK 内部会先打乱连接点顺序避免不同客户端的控制连接总是
        // 连一个点。
        String[] contactPoints = new String[]{
            "ipaddress"
            // 填写 ip 地址
        };
        Cluster cluster = Cluster.builder()
            .addContactPoints(contactPoints)
            // 填写端口
            .withPort(port)
            // 填写账户名密码（如果忘记可以在用户管理处重置）
            .withAuthProvider(new PlainTextAuthProvider("cassandra", "cassandra"))
            .build();
        // 初始化集群，此时会建立控制连接（这步可忽略，建立 Session 时候会自动调用）
        cluster.init();
        // 连接集群，会对每个 Cassandra 节点建立长连接池。
        // 所以这个操作非常重，不能每个请求创建一个 Session。合理的应该是每个进程预先创建若干
        // 个。
        // 通常来说一个够用了，你也可以根据自己业务测试情况适当调整，比如把读写的 Session 分开管
        // 理等。
        Session session = cluster.connect();
        // 查询，此处我们查一下权限相关的表，看看我们创建了几个角色。
        // 这是系统表，默认只有 cassandra 这个 superuser 账户有 SELECT 权限。
        // 如果你使用其他帐号测试，可以换一个表或者授权一下。
        ResultSet res = session.execute("SELECT * FROM system_auth.roles");
        // ResultSet 实现了 Iterable 接口，我们直接将每行信息打印到控制台
        res.forEach(System.out::println);
    }
}
```

```
// 关闭 Session  
session.close();  
// 关闭 Cluster  
cluster.close();  
}  
}
```

2.4.2 Cqlsh 接入

可以在通 VPC ECS 主机上安装 Cassandra，通过 cqlsh 工具访问数据库 Cassandra。

进入 Cassandra 安装目录，bin/cqlsh \$host \$port -u \$username -p \$password

2.5 高可用配置

- 1、数据副本数大于等于 2，建议配置 3 副本。
- 2、Cassandra 集群节点数需大于等于副本数。
- 3、读写一致性配置：在副本数配置大于等于 3 时，建议配置 Quorum 读写一致性。不可配置要求所有数据副本所在节点均读/写成功才成功的策略，例如配置了三副本，一致性配置了 ALL，即读或写操作要求三个数据副本所在节点均读或写成功才成功，此情况下，若数据副本所在的任何一个节点故障，则数据库就无法读/写，无高可用保障。

2.6 可调一致性

可用的一致性级别包括 ONE、TWO、THREE，分别指定必须响应请求的复制节点的绝对数量。一致性级别 QUORUM 要求大多数复制节点响应(复制因子/2+1)。一致性级别 ALL 要求所有副本都响应。对于读写操作，ANY、ONE、TWO、THREE 都是弱一致性，QUORUM 和 ALL 是强一致性。

经常会用下面的公式来计算强一致性节点数：

$R+W>N$ = 强一致性，在公式中，R、W、N 分别代表读副本数、写副本数、复制因子

2.7 数据迁入迁出

cqlsh 提供 COPY 命令，包括 COPY TO 和 COPY FROM 命令支持数据导入导出，可以使用 COPY TO 命令将已有的 Cassandra 实例数据导出，再通过 COPY FROM 命令导入到 RDBMS 或者新 Cassandra 实例。

3. 操作指导

3.1 实例生命周期管理

进入数据库 Cassandra 控制台实例列表页，可进行实例的开通、续费、退订；

包周期实例到期不可访问，续费可恢复，退订后实例默认保留半个月后进行销毁。

按需计费实例退订立即释放，数据不保留。

3.1.1 变更实例

3.1.1.1 存储空间扩容

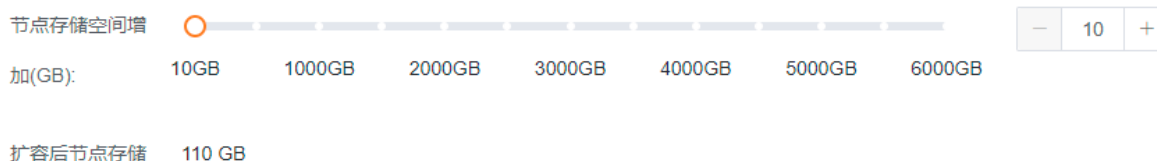
随着业务的发展，数据库存储容量不能满足数据的增加时，可以进行磁盘扩容。

需要注意的是：磁盘容量只允许扩容，不能缩容。无需重启实例，在此期间，服务不中断，不影响您正常使用数据库。

- 1、登录云数据库 Cassandra 控制台
- 2、在实例列表，选择目标实例，更多-存储空间扩容

当前配置

实例名称:	AnalyticPG-eold
实例ID:	0e51a829614a4794a069059ae2ef26dc
当前规格:	2核8G
当前节点数量:	3
存储类型:	普通IO
当前节点存储空间:	100GB



磁盘扩容是针对每个 Cassandra 节点，扩容的总容量=节点数*节点扩容的容量。

3.1.1.2 升 CPU/内存规格

当实例节点 CPU 和内存规格无法满足业务需求时，可以进行规格变更；变更只能扩容规格不能降低规格。变更 CPU/内存规格需要重启，建议业务空闲时进行；变更过程不建议进行 DDL 操作。

- 1、登录云数据库 Cassandra 控制台
- 2、在实例列表，选择目标实例，更多-升 CPU/内存。

3.1.1.3 增加节点

随着业务发展，当前实例节点不能满足需求时，可以为实例添加节点，不支持节点缩容。

- 1、登录云数据库 Cassandra 控制台
- 2、实例列表，选择目标实例，更多-增加节点

当前配置	
实例名称:	AnalyticPG-eold
实例ID:	0e51a829614a4794a069059ae2ef26dc
当前规格:	2核8G
当前节点数量:	3
当前节点存储空间:	100GB

增加节点数量:

-	1	+
---	---	---

您最多可以持有12个节点数。
添加节点会有短暂的OPS下降，建议业务空闲时添加。

总节点数量:

4

增加的节点个数受最大节点数 12 的限制。

3.2 命令行

命令窗口页面提供执行 CQL 语句能力，方便快速接入使用数据库。具体可查看示例操作。

● 示例

运行

命令运行结果

1.通过 cqlsh 创建 keyspace

Cassandra中的keyspace和关系型数据库中的database概念比较类似，一个 keyspace 可以包含一个或多个 tables 或 column families。当您启动 cqlsh 时没有指定 keyspace，那么命令提示符为 cqlsh>，您可以使用 CREATE KEYSPACE 命令来创建 keyspace，具体如下：实现语句：

```
CREATE KEYSPACE test_keyspace WITH replication = {'class': 'SimpleStrategy', 'replication_factor': 1};
```

上面命令创建了名为 test_keyspace 的 keyspace；并且采用 SimpleStrategy 进行副本复制。由于当前的测试集群只有单个节点，所以设置副本因子（replication factor）为 1。如果是生产环境，请勿将副本因子设置为 1，建议将副本因子设置为 3。

2.创建表

您可以通过下述命令在 test_keyspace 里面建表：

```
CREATE TABLE test_keyspace.test_user (first_name text , last_name text, PRIMARY KEY (first_name)) ;
```

3.读写数据

您可以通过下述命令在插入数据：

```
INSERT INTO test_keyspace.test_user (first_name, last_name) VALUES ('test', 'Hadoop');  
INSERT INTO test_keyspace.test_user (first_name, last_name) VALUES ('Zhang', 'San');  
INSERT INTO test_keyspace.test_user (first_name) VALUES ('Wu');
```

3.3 用户管理

用户管理模块可进行数据库用户的创建、删除和密码重置。

3.3.1 新增数据库用户

可以创建普通用户和超级用户，创建完用户后可通过命令行功能进行授权操作。

新增用户

* 用户名

密码

确认密码

是否超级用户 ☐

取消

确定

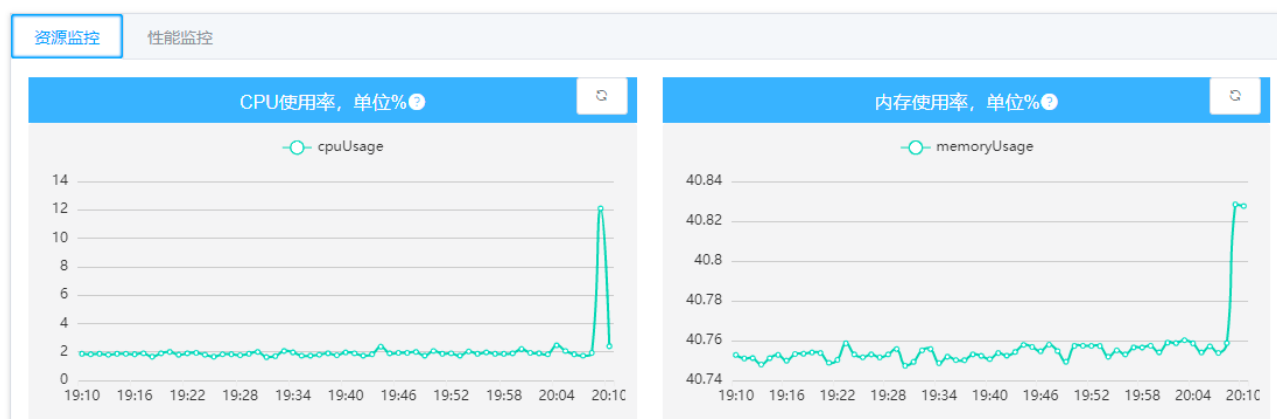
3.3.2 密码重置

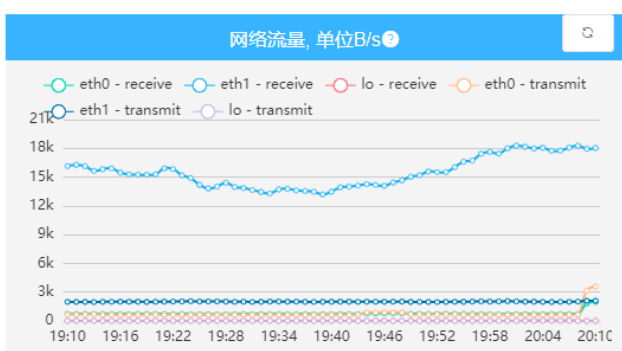
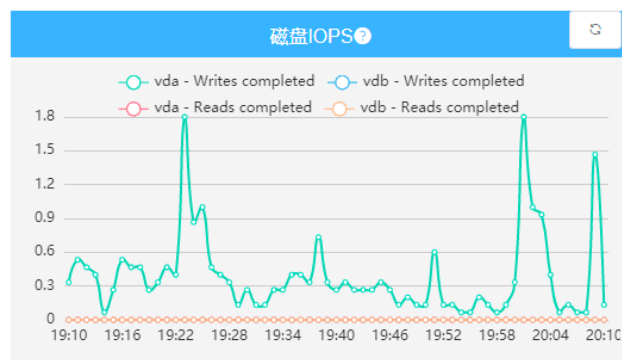
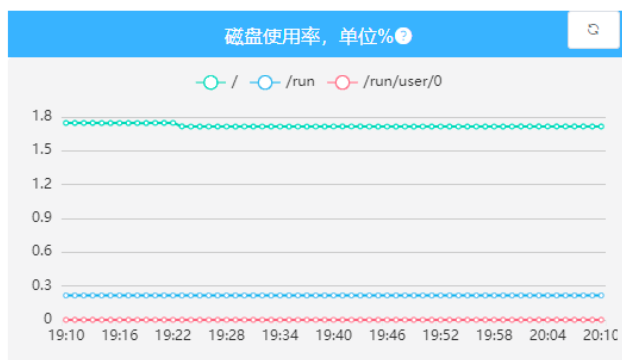
提供用户密码的重置修改能力。

3.4 监控告警

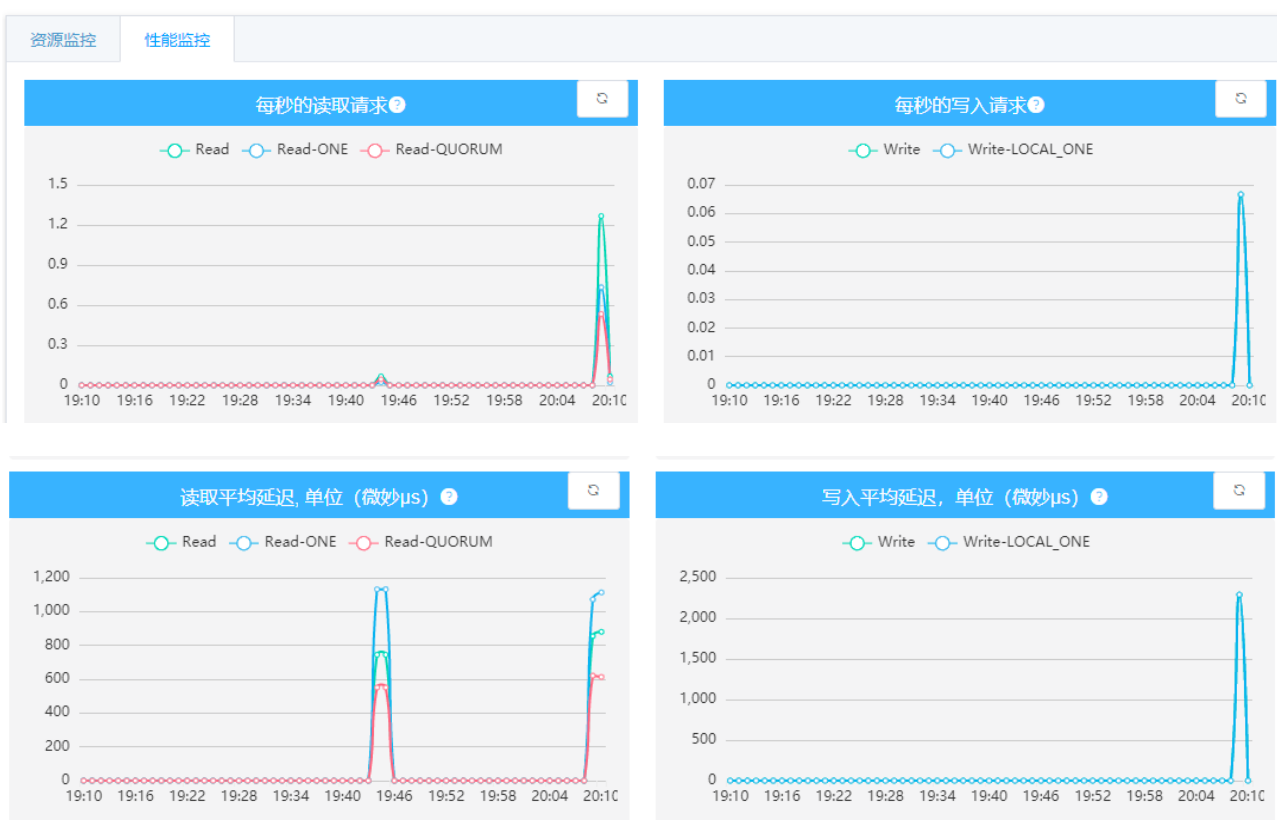
监控告警模块提供 Cassandra 各节点包括物理资源和实例节点指标监控

物理资源监控包括：cpu、内存、磁盘使用率、磁盘 IO 和 IOPS、网络流量。





节点性能监控包括读写 TPS、读写延时。



3.5 数据库配置

数据库配置提供实例节点参数的常用参数的调整，各参数功能可详见参数列表备注。

3.6 节点管理

4. 最佳实践

- 1、数据建模方面，不要把 Cassandra 数据模型类比成关系型数据库 Table，想象成嵌套的、排序的 Map 数据结构。当设计 Cassandra column family 的时候考虑它是一个 Map 中嵌入另一个 Map；外部 Map 的 key 为 row key，内部 Map 的 key 为 column key，两个 Map 的 key 都是有序的如 `SortedMap<RowKey, SortedMap<ColumnKey, ColumnValue>>`
- 2、围绕查询模式进行 Column Family 建模，Cassandra 通过创建二级索引或者编写复杂 SQL（使用 joins, order by, group by）来新建或修改查询不是件容易的事情，牢记前面提到的嵌入排序 map 数据结构，在考虑如何组织你的数据到 map，以满足快速查询 / 排序 / 分组 / 过滤 / 聚合的要求
- 3、必要时通过反范式和冗余提升性能。Cassandra 数据是分布式存储的并不支持 join 操作，可以通过反范式化来提升查询性能。
- 4、使用宽行进行排序、分组和过滤，但不宜太长。Column name 物理有序存储，这有利于宽行排序和过滤，对于一次查多个数据，可以将数据分组保存到一个宽行中如此可以快速读取数据
- 5、选择合适的 rowkey
- 6、读多数据和写多数据分离，充分利用 off-heap 行缓存特性
- 7、使用合适的 comparator 和 validator，尽量不要使用默认的 BytesType comparator 和 validator。一个不合适的 comparator 将不能按照你预期所想的顺序将 column name 排序并存放到磁盘上。它将很难（或者不能）根据 column name 进行范围查询。可以通过 Cassandra 文档看看 comparators 和 validators 支持的数据类型。
- 8、将数据模型设计为支持幂等。

- 9、 优先使用内置的 composite type 而不是自定义类型。如果 sub-columns 是不同的数据类型，自定义类型将不能正常工作。比如，composite key<state|zip|timeuuid> 将不能适用于根据类型感知排序（state 是字符串，zip 是整型，timeuuid 是时间类型）。
- 10、 优先使用静态 composite type。如果一个 column family 中的 column keys 都是相同的 composite type，则一定要使用静态 composite type。创建动态 composite type 最初的目的是为了在一个 column family 中保存多用户自定义索引。如果可能，请不要在一行中使用不同的 composite type，除非真的有必要这么做
- 11、 对于 super columns 优先使用 composite columns。
使用 super columns 存在以下问题：
 - Super column 的子 column 不能被索引，读取一个字 column 意味着要反序列化整个 super column
 - 子 column 无法使用内建二级索引
 - Super column 无法实现超过两层的结构容易产生性能瓶颈，类似的功能，可以通过 Composite column 实现，它是包含子 column 的标准的 column。
- 12、 尽量让数据均匀分布在集群的每个节点中
- 13、 最小化每次读取使用的分区数量

5. 常见问题

5.1 包年包月和按需计费切换

目前暂不支持包年包月和按需计费切换

5.2 实例开通后 VPC 切换

实例创建完后暂不支持更换 VPC。

5.3 备份

备份功能暂缺，已纳入版本计划中

5.4 实例可用性

实例可用性= (1 - 故障时间/服务总时间) × 100%

故障时间是指数据库实例创建成功后，运行期间累计发生故障的总时长。服务总时间指数据库实例创建成功后运行的总时长。

5.5 可用区域

区域是一个地理区域的概念。我国地域面积广大，由于带宽的原因，不可能只建设一个数据中心为全国客户提供服务。因此，我们根据地理区域的不同将全国划分成不同的区域。

服务可选区域以订购页提供为准，根据就近原则选择。