



零基础学习 **hadoop** 生态系统 到上手工作线路指导（**HBASE** 篇） **hbase** 从入门到编程

问题导读：

- 1.**hbase** 是什么？
- 2.**hbase** 原理是什么？
- 3.**hbase** 使用中会遇到什么问题？
- 4.如何通过 **eclipse**，操作 **hbase**？
- 5.**hbase** 经常和哪些软件一起使用？

。 。 。 。 。 。 。 。 。 。



本文链接：<http://www.aboutyun.com/thread-8391-1-1.html>

hbase 从入门到编程

认识 **hbase**

1.hbase 简介

hbase 我们或许已经知道了它是 nosql，但是什么是 nosql，我们不太清楚，nosql 是一种基于列的数据库，而我们的传统数据库则是基于行的数据库。想对 nosql 进一步了解，参考 [Nosql 数据库入门分享](#)

HBase— Hadoop Database，是一个高可靠性、高性能、面向列、可伸缩的分布式存储系统，利用 HBase 技术可在廉价 PC Server 上搭建起大规模结构化存储集群。



HBase 是 GoogleBigtable 的开源实现，类似 Google Bigtable 利用 GFS 作为其文件存储系统，HBase 利用 HadoopHDFS 作为其文件存储系统；Google 运行 MapReduce 来处理 Bigtable 中的海量数据，HBase 同样利用 Hadoop MapReduce 来处理 HBase 中的海量数据；Google Bigtable 利用 Chubby 作为协同服务，HBase 利用 Zookeeper 作为对应。

更多内容参考：

Hbase 简介

<http://www.aboutyun.com/thread-6138-1-1.html>



2.hbase 应用场景

我们知道了 hbase 是什么，那么我们什么时候使用。

当我们对于数据结构字段不够确定或杂乱无章很难按一个概念去进行抽取的数据适合使用什么数据库？答案是什么，如果我们使用的传统数据库，肯定留有多余的字段，10 个不行，20 个，但是这个严重影响了质量。并且如果面对大数据库，pt 级别的数据，这种浪费更是严重的，那么我们该使用是什么数据库？hbase 数个不错的选择。更多内容参考：

hbase 常识及 habse 适合什么场景

<http://www.aboutyun.com/thread-7073-1-1.html>



3.hbase 与传统数据的区别

我们整体了解了 hbase, 那么 hbase 与传统数据库到底有哪些区别, 我们的增删改查, nosql 有没有增删改查, 答案是有的。下面我们以两个图来表示:

图 1, 是我们经常见到的传统数据, 图 2 则是 nosql 数据库, 从这个图中, 或许我们知道了, nosql 为什么被称之为基于列的数据库了。

想了解更多参考下面内容

图解 Nosql (hbase) 与传统数据库的区别

<http://www.aboutyun.com/thread-7804-1-1.html>

hbase 与传统数据的区别

<http://www.aboutyun.com/thread-6720-1-1.html>

图 1 传统数据库

	列1	列2
记录1		
记录2		
记录3		
记录4		

图 2: hbase 数据库

	列族	列族
记录1	列1..列N	列1..列N
记录2	列1、列2	列1..列3
记录3	列1..列4	列1
记录4	列1.列5	列1..列10



4.hbase 与 hadoop 的关系

我们了解的越多，有时候问题就越多，hbase 与 hadoop 是什么关系，有了 hadoop，我们为什么使用 hbase.这个有点像我们的磁盘与数据库，比如我们的把 mysql 或则 sqlserver 放到 D 盘一样。hadoop 提供了介质，hbase 存储在 hdfs 上，同理 hive 也是如此，对于 hive 这里不在详述，可以参考[零基础学习 hadoop 到上手工作线路指导（中级篇）](#)。hbase 与 hive 及 hadoop 的关系，更详细参考下面帖子：

hive 与 hbase 的十大区别与联系

<http://www.aboutyun.com/thread-7870-1-1.html>



5.hbase 术语及原理

hbase 中出现了 Region，RegionServer，ROOT- 和.META 表，Region 是 HBase 数据存储和管理的基本单位。一个表中可以包含一个或多个 Region。每个 Region 只能被一个 RS（RegionServer）提供服务，RS 可以同时服务多个 Region，来自不同 RS 上的 Region 组合成表格的整体逻辑视图。

HBase 中有两张特殊的 Table，-ROOT-和.META.

META.: 记录了用户表的 Region 信息，.META.可以有多个 region

ROOT-: 记录了.META.表的 Region 信息，-ROOT-只有一个 region

当我们执行添加、删除数据的时候，相应的 hbase 的 META，ROOT 都会有相应的改变。

更多内容参考：

HBase 各部件的作用

<http://www.aboutyun.com/thread-5862-1-1.html>

hbase 中什么是 Region，什么是 RegionServer？

<http://www.aboutyun.com/thread-7159-1-1.html>

HBASE 原理简述

<http://www.aboutyun.com/thread-7199-1-1.html>



hbase 部署

我们掌握理论知识，并不是真正的认识，因为计算机是一门实践学科，下面我们开始部署 hbase，什么是部署，这里是安装的意思，因为 Linux 的安装软件与 window 有所区别的，所以对于刚接触 Linux 的同学，在部署方面还需要花不少功夫的。

hbase 的部署分为单机部署及集群部署，建议采用集群部署，在部署 hbase 之前，需要安装 hadoop。部署文档可以参考：

hbase 0.96 整合到 hadoop2.2 三个节点全分布式安装高可靠文档

<http://www.aboutyun.com/thread-7746-1-1.html>

在部署的时候，需要注意 hadoop 版本与 hbase 的版本匹配。

更多内容，可参考下面：

Hbase 单节点部署及其基础使用

<http://www.aboutyun.com/thread-7677-1-1.html>

HBASE 分布式安装视频下载分享

<http://www.aboutyun.com/thread-6612-1-1.html>

hadoop、hbase、hive 版本对应关系

<http://www.aboutyun.com/thread-7295-1-1.html>

hbase 与 zookeeper 可以说是不可分离的，因此 hbase 有自带的 zookeeper，如果不使用自带 zookeeper，可以禁用 zookeeper，使用外部 zookeeper。

Hbase 禁用自带 ZooKeeper，使用已经安装的 ZooKeeper

<http://www.aboutyun.com/thread-7451-1-1.html>

上面我们完成基本的配置，在后面的使用过程中，我们对集群的要求增高，所以我们就需要了解更多的配置项的作用，下面帖子可以参考。

hbase 配置项说明

HBase 默认配置说明（收藏版）

<http://www.aboutyun.com/thread-7914-1-1.html>

Hbase 配置项粗解（1）

<http://www.aboutyun.com/thread-8274-1-1.html>

Hbase 配置项粗解（2）

<http://www.aboutyun.com/thread-8275-1-1.html>

Hbase 配置项粗解（3）

<http://www.aboutyun.com/thread-8276-1-1.html>

其他：

hbase-0.94.10.tar.gz 包下载

<http://www.aboutyun.com/thread-6114-1-1.html>

hadoop_HBase 安装，数据分析与建模，实战案例剖析

<http://www.aboutyun.com/thread-5146-1-1.html>

Ganglia 监控 Hadoop 及 Hbase 集群性能(安装配置)

<http://www.aboutyun.com/thread-8129-1-1.html>

更多的 **jar** 包、安装包

[hadoop 家族、strom、spark、Linux、flume 等 jar 包、安装包汇总下载\(持续更新\)](#)



hbase 整合

hbase 不是单独存在的，hbase 可以与 hive、Impala、Flume 整合。我们为什么要整合？比如我们 hbase、hive 整合之后，当我们将数据插入 hbase 之后，hive 随之同步，我们无须操作 hive，这就是整合的好处。对于 flume 整合，当 flume 整合搜集到数据之后，会自动发送到 hbase，这样省去了我们很多的功夫去编程。我们只需要关心我们的业务逻辑即可。

hive 与 hbase 整合原理介绍

<http://www.aboutyun.com/thread-7824-1-1.html>

阐述：

- 1.Hive 与 hbase 整合的原理？
- 2.Hive 与 hbase 整合后的使用场景？
- 3.Hive 与 hbase 整合后的使用方法？

hive 为什么与 hbase 整合

<http://www.aboutyun.com/thread-7317-1-1.html>

阐述：

- 1.hive 为什么与 hbase 整合？
- 2.hive 整合 hbase 的优缺点是什么？

hbase 0.96 整合到 hadoop2.2 三个节点全分布式安装高可靠文档（推荐）

<http://www.aboutyun.com/thread-7746-1-1.html>

此文档详细介绍了 **hbase 0.96** 整合到 **hadoop2.2** 整合的步骤及遇到的相关问题

Impala 与 HBase 整合实践

<http://www.aboutyun.com/thread-7856-1-1.html>

编译和使用 hive 与 HBase 通信包--hive-hbase-handler.jar 及下载

<http://www.aboutyun.com/thread-7817-1-1.html>

阐述:

- 1.hive 与 hbase 对应版本
- 2.hive-hbase-handler.jar 在于 hbase、hive 中起的作用是什么?
- 3.hive-hbase-handler.jar 是否有版本之分, 不同版本是否都能使用这个包来整合 hbase 与 hive?
- 4.整合过程中 hive-hbase-handler.jar 应该放在 hive 的哪个文件夹中?

hbase0.96 与 hive0.12 整合高可靠文档及问题总结

<http://www.aboutyun.com/thread-7881-1-1.html>

阐述下面问题:

- 1.hive 安装是否需要安装 mysql?
- 2.hive 是否分为客户端和服务端?
- 3.hive 的元数据库有哪两种?
- 4.hive 与 hbase 整合的关键是什么?
- 5.hive 的安装是否必须安装 hadoop?
- 6.hive 与 hbase 整合需要做哪些准备工作? 7.hive 元数据库启动卡住代表的含义是什么?

Flume-ng 将数据插入 hdfs 与 HBase-0.96.0

<http://www.aboutyun.com/thread-7912-1-1.html>

阐述下面问题:

- 1.如何配置分布式 flume
- 2.master 与 node 之间该如何配置, 有什么异同?
- 3.启动命令是什么?
- 4.flume 把数据插入 hbase, 该如何配置?

hive0.13 调整 hbase 0.96.2 hadoop2.2.0 问题总结

<http://www.aboutyun.com/thread-7893-1-1.html>

阐述下面问题:

- 1.hive.aux.jars.path 参数的作用是什么?
- 2.Job Submission failed with exception 'java.io.FileNotFoundException'这个问题该如何解决?

Flume-1.4.0 和 Hbase-0.96.0 整合实践

<http://www.aboutyun.com/thread-7418-1-1.html>

阐述下面问题:

- 1.需要修改那些文件?
- 2.如何测试整合成功?
- 3.兼容问题该如何解决?

Flume-0.9.4 和 Hbase-0.96 整合实践

<http://www.aboutyun.com/thread-7417-1-1.html>

阐述下面问题：

- 1.都需要修改那些文件？
- 2.为什么修改这些文件？
- 3.代码有的地方需要改动，猜测原因什么？



hbase 使用

上面讲了很多，从理论到实践安装，我们终于可以使用了，我们很想尝试增删改查到底是什么样子的？

我们可以安装下面格式，来操作：

名称	命令表达式
创建表	<code>create '表名称', '列名称 1','列名称 2','列名称 N'</code>
添加记录	<code>put '表名称', '行名称', '列名称:', '值'</code>
查看记录	<code>get '表名称', '行名称'</code>
查看表中的记录总数	<code>count '表名称'</code>
删除记录	<code>delete '表名', '行名称', '列名称'</code>
删除一张表	先要屏蔽该表，才能对该表进行删除，第一步 <code>disable '表名称'</code> 第二步 <code>drop '表名称'</code>
查看所有记录	<code>scan "表名称"</code>
查看某个表某个列中所有数据	<code>scan "表名称", ['列名称:']</code>
更新记录	就是重写一遍进行覆盖

这里举个具体例子：

1.创建一个表

```
hbase(main):011:0>create 'member','member_id','address','info'  
0 row(s) in 1.2210seconds
```

2.drop 一个表

```
hbase(main):029:0>disable 'temp_table'  
0 row(s) in 2.0590seconds
```

更详细参考下面帖子：

HBase Shell 常用命令

<http://www.aboutyun.com/thread-6151-1-1.html>

上面是基本的操作，使用的时间越长，我们的需求就会越多，比如删除数据会遇到 region 不释放，我们想备份数据、定期删除数据，压缩数据、查询优化等更详细内容查看下面帖子：

HBase 如何实现多条件查询

<http://www.aboutyun.com/thread-6685-1-1.html>

开启 hadoop 和 Hbase 集群的 lzo 压缩功能

<http://www.aboutyun.com/thread-8349-1-1.html>

阐述问题：

- 1.如何启动 **hadoop**、**hbase** 集群的压缩功能？
- 2.**lzo** 的作用是什么？
- 3.**hadoop** 配置文件需要做哪些修改？

HBase 实现记录定期定量删除

<http://www.aboutyun.com/thread-8307-1-1.html>

阐述问题：

- 1、如何定期删除数据？
- 2、如何在数据超过阈值时删除数据？

hbase 数据删除不释放 region 解决办法

<http://www.aboutyun.com/thread-8306-1-1.html>

阐述问题：

- 1.删除 **hbase** 数据有几种方法？
- 2.删除数据，**region** 不释放，你认为该如何解决？

HBase 利用 Coprocessor 实现聚合函数

<http://www.aboutyun.com/thread-7840-1-1.html>

阐述问题：

- 1、**HBase** 默认不支持聚合函数，那我们该用什么来实现呢 ？
- 2、怎么用编程的方式去实现呢 ？

大数据应用之 HBase 数据插入性能优化之多线程并行插入测试案例

<http://www.aboutyun.com/thread-8011-1-1.html>

阐述问题：

- 1、单线程下 HBase 的插入性能如何？
- 2、如何在多线程下了解 HBase 的性能？

hbase 与 hadoop2.X 在 CentOS6.4 下源码编译

<http://www.aboutyun.com/thread-7150-1-1.html>

阐述问题：

- hadoop2.2.0 编译需要注意什么问题？
- HBase 部署的时候需要注意什么问题？
- HBase0.96.0 编译前需要使用什么为相应的版本生成 pom 文件？

hbase meta 表修复方式总结

<http://www.aboutyun.com/thread-7998-1-1.html>

阐述问题：

- 1.hbase hbck -fixMeta 的作用是什么？
- 2.如何重新将 hbase meta 表分给 regionserver？
- 3.出现 region 的 hole 该如何修复？

优化 hbase 的查询优化-大幅提升读写速率

<http://www.aboutyun.com/thread-7657-1-1.html>

阐述问题：

- 1.本文通过什么方法优化查询效率的？
- 2.如何增大 RPC 数量？
- 3.如何调整 hbase 内存？

hadoop 及 hbase 的超时设置

阐述问题：

<http://www.aboutyun.com/thread-7552-1-1.html>

- 1.hadoop 超时，该如何设置超时时间？
- 2.hbase 超时时间限制，该如何设置？

优化 hbase 的查询提升读写速率优化案例及性能提升的几种方法

<http://www.aboutyun.com/thread-7468-1-1.html>

阐述问题：

- 1.完全发挥不出 hbase 的效率的原因是什么？

2.使用 bloomfilter 和 mapfile_index_interval 如何提升性能?

3.如何设置 hbase 的内存?

4.如何增大 RPC 的数量?

扩展:

5.为什么 HBase 是基于列模式的存储?

Hbase 数据备份和恢复

<http://www.aboutyun.com/thread-7296-1-1.html>



hbase 工具

所谓工具就是帮助我们的, hbase 工具, 则是帮助我们使用 hbase, phoenix 的作用是什么? phoenix 的操作 sql 是通过 jdbc 发送到 HBase 的。phoenix 的查询语句会转化为 hbase 的 scan 操作和服务端端的过滤器。如果我们手工使用 HBase 的 api 去写这些代码, 也会得到相同的运行结果和执行速度。但是, 使用 phoenix 的效果却会带来更快的开发效率。更多详细内容, 见下面帖子:

hbase 的 sql 操作的框架-phoenix

<http://www.aboutyun.com/thread-6688-1-1.html>

hbase 有哪些可视化工具,图形界面管理工具

<http://www.aboutyun.com/thread-6257-1-1.html>

phoenix 实战 (hadoop2、hbase0.96)

<http://www.aboutyun.com/thread-8208-1-1.html>

Phoenix 介绍: 实现向 HBase 发送标准 SQL 语句

<http://www.aboutyun.com/thread-8153-1-1.html>



hbase 设计

hbase 设计这个是一个比较大的话题，很多都是根据自己项目的具体情况来设计，这里只是简单说一下设计中需要注意的问题 1.防止数据倾斜

2.性能提升

防止数据倾斜，rowkey 的设计还是比较关键的。

性能提升，在设计之初，将要查询的字段组合到 rowkey 中，否则后期可能查询是一个问题。更多详细内容参考下面帖子。

Hbase 初步入门-- 表该如何构造和设计

<http://www.aboutyun.com/thread-7828-1-1.html>

HBase 设计

<http://www.aboutyun.com/thread-5903-1-1.html>

hbase 之 rowkey 的设计讨论

<http://www.aboutyun.com/thread-8171-1-1.html>

HBase 的 rowkey 设计

<http://www.aboutyun.com/thread-7119-1-1.html>

hbase 数据快速备份：HBase snapshot 分析

<http://www.aboutyun.com/thread-8038-1-1.html>

修复 hbase 元数据.meta.表空洞的问题

<http://www.aboutyun.com/thread-7894-1-1.html>

hbase Balancer 源码分析-负载均衡

<http://www.aboutyun.com/thread-7643-1-1.html>

HBase Bug 知多少

<http://www.aboutyun.com/thread-7135-1-1.html>



hbase 编程

hbase 分为两种是一种开发 API，就是所谓的二次开发，一种是客户端 API，也就是我们使用 hbase 来完成一些事情。对于 hbase 源码获取，我们需要具备 maven 知识，同 hadoop 获取源码的方式是一样的，具体可以参考：[从零教你如何获取 hadoop2.4 源码并使用 eclipse 关联 hadoop2.4 源码](#)。

通过 eclipse 获取源码之后，我们会看到一些 maven 语法错误，如果遇到 execution 错误，可以参考下面帖子。

hbase、hadoop 通过 eclipse m2e maven 插件获取源码产生 execution 错误解决方案
<http://www.aboutyun.com/thread-8353-1-1.html>

我们获取源码是为了查看更好的使用 hbase，hbase 的基本操作，在 shell 中能够完成，我们通过 Java API eclipse 环境同样能完成。除了增删改查、分页之外，我们还可以开发插件。

hbase-0.90.2 中创建表、插入数据，更新数据，删除数据实例
<http://www.aboutyun.com/thread-7496-1-1.html>

Java 操作 hbase 编程
<http://www.aboutyun.com/thread-7075-1-1.html>

spark 使用 java 读取 hbase 数据做分布式计算
<http://www.aboutyun.com/thread-8242-1-1.html>

hbase 编程：通过 Java api 操作 hbase
<http://www.aboutyun.com/thread-7151-1-1.html>

hbase HTable 之 Put、delete、get 等源码分析
<http://www.aboutyun.com/thread-7644-1-1.html>

Hbase Java 编程实现增删改查
<http://www.aboutyun.com/thread-6901-1-1.html>

总结 Eclipse 远程连接 HBase 问题及解决方案大全
<http://www.aboutyun.com/thread-5866-1-1.html>

HBase 中如何开发 LoadBalance 插件
<http://www.aboutyun.com/thread-8350-1-1.html>

Hbase 与 eclipse 集成的第一个例子
<http://www.aboutyun.com/thread-7837-1-1.html>

hbase 分页应用场景及分页思路与代码实现

<http://www.aboutyun.com/thread-7030-1-1.html>

HBase MapReduce 排序 Secondary Sort

<http://www.aboutyun.com/thread-7304-1-1.html>

CDH4 源码搭建 hbase 开发环境

<http://www.aboutyun.com/thread-7259-1-1.html>

Thrift 了解 4: C#通过 Thrift 操作 HBase 实战

<http://www.aboutyun.com/thread-7142-1-1.html>

hbase API

hadoop2.2.0 帮助手册下载 API 及 HBase 0.98.1-hadoop2 API

<http://www.aboutyun.com/thread-6113-1-1.html>

HBase 数据迁移（1）-使用 HBase 的 API 中的 Put 方法

<http://www.aboutyun.com/thread-8336-1-1.html>

hbase 编程：Java API 连接 Hbase 进行增删改查讲解实例

<http://www.aboutyun.com/thread-8290-1-1.html>



hbase 问题

hbase 使用和开发过程中会遇到各种问题，比如插入数据越来越慢，master 启动之后又挂掉，hbase 如何存储图片，线上 regionserver 无缘无故下线，这里整理了一下。

hive-hbase 整合后查询缓慢

<http://www.aboutyun.com/thread-7935-1-1.html>

hbase 插入数据，为什么速度越来越慢

<http://www.aboutyun.com/thread-6564-1-1.html>

hbase 排错:be reached after 1 tries, giving up.

<http://www.aboutyun.com/thread-6579-1-1.html>

hbase master 启动了又挂了

<http://www.aboutyun.com/thread-5882-1-1.html>

hbase 配置、运行错误总结

<http://www.aboutyun.com/thread-8319-1-1.html>

hbase 删除数据的问题

<http://www.aboutyun.com/thread-8304-1-1.html>

hbase 和 hive 整合问题

<http://www.aboutyun.com/thread-8227-1-1.html>

HBase 如何把图片存进去呢？

<http://www.aboutyun.com/thread-8219-1-1.html>

hbase 编写自定义 count 功能的问题

<http://www.aboutyun.com/thread-8191-1-1.html>

请问 hbase 如何设置 region 大小啊

<http://www.aboutyun.com/thread-7974-1-1.html>

hbase regionserver 下线

<http://www.aboutyun.com/thread-8138-1-1.html>

hbase 伪分布式是不是没有 regionserver 啊

<http://www.aboutyun.com/thread-8158-1-1.html>

hbase 的“-ROOT-”表所在的 block 丢失,该如何恢复

<http://www.aboutyun.com/thread-7988-1-1.html>

hbase 插入数据, 出现 java.lang.OutOfMemoryError

<http://www.aboutyun.com/thread-7896-1-1.html>

hive0.13 与 hbase0.98.2 通过向 Hbase 表中导出数据出现的错

<http://www.aboutyun.com/thread-7892-1-1.html>

hbase 编程如何获取动态列

<http://www.aboutyun.com/thread-7763-1-1.html>

获取源码遇到问题解决方案:

hbase、hadoop 通过 eclipse m2e maven 插件获取源码产生 execution 错误解决方案

<http://www.aboutyun.com/thread-8353-1-1.html>

解决问题的根本

hbase 日志

Hadoop 和 **Hbase** 重要日志位置

<http://www.aboutyun.com/thread-6106-1-1.html>



hbase 在大企业应用

我们在使用 hbase 的过程中, 我们所遇到的问题, 走过的路, 或许别人已经遇到并且解决, 对于阿里、360、Facebook 等大公司是在我们前面的, 下面的一些经验可以借鉴。

hbase 在 360 的应用及使用过程中遇到的问题及解决方案

<http://www.aboutyun.com/thread-8298-1-1.html> 阐述问题:

1.360 为什么使用 hbase?

2.调用 Put 接口写入数据, 写入性能开高效的原因是什么? 该如何解决?

3.bulkImport 的数据导入阶段较慢原因是什么?

4.bulkImport 后, compaction 操作会产生大量 IO 原因是什么? 该如何解决?

HBase 在淘宝主搜索的 Dump 中的性能调优

<http://www.aboutyun.com/thread-8285-1-1.html>

阐述问题:

- 1.hbase 在运用中，如何有效降低延时？
- 2.如何对 Dump 进行性能调优？

淘宝搜索分析系统 Pora2 的应用之 HBase 高并发读写性能优化

<http://www.aboutyun.com/thread-8025-1-1.html>

阐述问题:

- 1、如何理解淘宝的搜索分析系统 Pora？
- 2、redis 超时的原因有哪些？

Facebook 使用 HBase 构建实时信息系统：能每月存储 1350 亿条信息

<http://www.aboutyun.com/thread-7499-1-1.html>

阐述问题:

- 1.Facebook 为什么选择 hbase？

扩展:

- 2.hbase 为什么被称之为面向列的数据库？
- 3.hbase 为什么能能承受如此多的数据？

淘宝之 HBase MapReduce 实例分析

<http://www.aboutyun.com/thread-7072-1-1.html>

阐述问题:

- 1.hbase MapReduce 那么它和 hadoop 的 MapReduce 有什么异同？
- 2.hbase MapReduce 基本模型是什么？
- 3.对于 InputFormat 和 OutputFormat 二者的作用是什么？
- 4.HBase 通过对哪些类的扩展（继承）来方便 MapReduce 任务来读写 HTable 中的数据？
- 5.HBase 中 Mapper 类继承哪个类？
- 6.HBase 中 Reducer 类继承哪个类？
- 7.HBase 在提交作业时设置 inputFormat 成什么？outputFormat 设置成什么？
- 8.HBase 中 TableMapReduceUtil 类的作用是什么？

淘宝 hbase 业务实践

<http://www.aboutyun.com/thread-7031-1-1.html>

阐述问题:

- 1.HTablePool 与传统数据块连接池那个对应？
- 2.通过那个函数可以释放连接池？
- 3.habse 在什么情况下会比较慢？
- 4.RowKey 该如何设计？

HBase 在内容推荐引擎系统中的一些问题

<http://www.aboutyun.com/thread-7147-1-1.html>

阐述问题:

- 1.随机读取性能成倍下降的原因是什么?
- 2.Snappy 压缩是为了解决 hbase 什么问题?
- 3.原生 HBase 最大的问题之一就是数据随机读写速度太慢, 该如何解决这个问题?
- 4.一个服务器宕机, 其服务器的数据, 做如何处理?
- 5.HBase 需要通过 Compaction 解决什么问题?
- 6.Region Server 假死或则退出, 会造成什么情况?

淘宝为什么使用 HBase 及如何优化的

<http://www.aboutyun.com/thread-6940-1-1.html>

阐述问题:

- 1.大数据量, 而且数据增量不可预测, 采用什么方案比较合适?
- 2.hbase 有什么优点和缺陷?
- 3.hbase 采用什么模型保证数据不丢失?
- 4.hbase 的 Meta 表损坏以及 split 方面的 bug 会造成什么情况?
- 5.什么情况下会关闭 jobtracker?
- 6.淘宝为了保障服务从结果上的可用, 都采取了什么措施?
- 7.hbase 中 split 为什么存在风险?

HBase 工程师线上工作经验总结----HBase 常见问题及分析

<http://www.aboutyun.com/thread-6929-1-1.html>

阐述问题:

- 1.HBase 遇到问题, 可以从几方面解决问题?
- 2.HBase 个别请求为什么很慢? 你认为是什么原因?
- 3.客户端读写请求为什么大量出错? 该从哪方面来分析?
- 4.大量服务端 exception, 一般原因是什么?
- 5.系统越来越慢的原因是什么?
- 6.Hbase 数据写进去, 为什么会没有了, 可能的原因是什么?
7. regionserver 发生 abort,遇到最多是什么情况?
- 8.从哪些方面可以判断 HBase 集群是否健康?
- 9.为了加强 HBase 的安全性, 你会采取哪些措施?

HBase 在搜狐内容推荐引擎系统中的应用

<http://www.aboutyun.com/thread-7297-1-1.html>

Facebook 针对 hbase 的优化方案分析

<http://www.aboutyun.com/thread-7180-1-1.html>



www.aboutyun.com

对于云技术、大数据爱好者、可以关注：**about** 云腾讯认证空间，每天更新最新内容，经典文章

<http://user.qzone.qq.com/278595437/main>

about 云官方群 **39327136**、**371358502**
云计算爱好者群

惯用手机可以关注微信：



云资源、云技术、疑问解答邮件订阅地址：

http://list.qq.com/cgi-bin/qf_invite?id=a52796b9458ddfd62e055bf899ec187af5860bbdcbf5ec9

淘宝云技术入门、hadoop、openstack 及其它视频：

<http://aboutyun.taobao.com/?v=1>