



about 云 V1.5 整理共享资源

此篇是 about 云

在

about 云资源汇总 V1.2: 包括 **hadoop,openstack,nosql**,虚拟化

<http://www.aboutyun.com/thread-5928-1-1.html>

about 云资源汇总指引 V1.3: 包括 **hadoop,openstack,nosql**,虚拟化

<http://www.aboutyun.com/thread-6150-1-1.html>

about 云资源汇总指引 V1.4:包括 **hadoop,openstack,nosql**,虚拟化

<http://www.aboutyun.com/thread-6730-1-1.html>

基础上的一个继续

本文由 about 云整理完成，about 云不会以个人名义共享资料给网友。

本文链接

<http://www.aboutyun.com/thread-8203-1-1.html>

hadoop 系列

hive 在腾讯分布式数据库时间分享

<http://www.aboutyun.com/thread-6811-1-4.html>



阿里服务架构及 **hadoop** 体系资源分享

<http://www.aboutyun.com/thread-6812-1-4.html>



阿里Hadoop集群架构及服务体系

梁李印（无影）

阿里巴巴集团-海量数据



hbase phoenix sql 入门

<http://www.aboutyun.com/thread-7079-1-3.html>

本书讲解了：

- Phoenix 最值得关注的一些特性有：

嵌入式的 JDBC 驱动，实现了大部分的 java.sql 接口，包括元数据 API

可以通过多行键或是键/值单元对列进行建模

完善的查询支持，可以使用多个谓词以及优化的扫描键

DDL 支持：通过 CREATE TABLE、DROP TABLE 及 ALTER TABLE 来添加/删除列

版本化的模式仓库：当写入数据时，快照查询会使用恰当的模式

DML 支持：用于逐行插入的 UPSERT

VALUES、用于相同或不同表之间大量数据传输的 UPSERT

- 官方下载地址
- 如何安装 phoenix
- 如何创建数据库
- 如何查询数据

.....



Hadoop 权威指南（第三版）英文版

<http://www.aboutyun.com/thread-6903-1-3.html>

书籍介绍

《Hadoop 权威指南（第 3 版）》的内容包括：使用 Hadoop 分布式文件系统（HDFS）保存大数据集；使用 MapReduce 运行分布式计算；使用 Hadoop 的数据和 I / O 构件实现压缩、数据完整性、序列化（包括 Avro）和持久化；了解常见的陷阱和高级特性，以编写实用的 MapReduce 程序；设计、构建和管理专用的 Hadoop 集群——或者在云中运行 Hadoop；使用 Sqoop 从关系型数据库载入数据到 HDFS；使用 Pig 查询语言进行大规模数据处理；使用 Hadoop 的数据仓库系统 Hive 分析数据集；利用 HBase 处理结构化和半结构化数据，以及利用 ZooKeeper 构建分布式系统.....

作者介绍:

(美) 怀特, Tom White, Cloudera 工程师兼 Apache 软件基金会成员, 自 2007 年 2 月起成为 ApacheHadoop 代码提交者。他为 oreilly.com、java.net 和 IBMdeveloperWorks 撰写了大量文章, 还经常在业界会议上作 Hadoop 主题演讲。



CDH 4.6 hive 相关下载

<http://www.aboutyun.com/thread-7213-1-3.html>

../	11-Mar-2014	17:39	-
0.10.0-cdh4.2.0/	10-Mar-2014	23:36	-
0.10.0-cdh4.2.0-SNAPSHOT/	11-Mar-2014	17:39	-
0.10.0-cdh4.2.1/	10-Mar-2014	23:36	-
0.10.0-cdh4.2.1-SNAPSHOT/	11-Mar-2014	17:39	-
0.10.0-cdh4.2.2/	11-Mar-2014	17:39	-
0.10.0-cdh4.3.0/	11-Mar-2014	17:39	-
0.10.0-cdh4.3.0-SNAPSHOT/	10-Mar-2014	23:36	-
0.10.0-cdh4.3.1/	11-Mar-2014	17:39	-
0.10.0-cdh4.3.2/	24-Mar-2014	16:46	-
0.10.0-cdh4.4.0/	11-Mar-2014	17:39	-
0.10.0-cdh4.4.0-SNAPSHOT/	10-Mar-2014	23:36	-
0.10.0-cdh4.5.0/	11-Mar-2014	17:39	-
0.10.0-cdh4.5.0-SNAPSHOT/	10-Mar-2014	23:36	-
0.10.0-cdh4.5.0.1/	17-Mar-2014	23:53	-
0.10.0-cdh4.5.0.1-SNAPSHOT/	10-Mar-2014	23:36	-
0.10.0-cdh4.5.0.2-SNAPSHOT/	01-Apr-2014	00:30	-
0.10.0-cdh4.6.0/	11-Mar-2014	17:39	-
0.10.0-cdh4.6.0-SNAPSHOT/	10-Mar-2014	23:36	-
0.10.0-cdh4.7.0-SNAPSHOT/	10-Mar-2014	23:37	-
0.11.0-cdh5.0.0-beta-1/	11-Mar-2014	17:39	-
0.11.0-cdh5.0.0-beta-1-SNAPSHOT/	10-Mar-2014	23:38	-
0.11.0-cdh5.0.0-SNAPSHOT/	10-Mar-2014	23:37	-
0.12.0-cdh5.0.0-beta-2/	11-Mar-2014	17:39	-
0.12.0-cdh5.0.0-beta-2-SNAPSHOT/	10-Mar-2014	23:39	-
0.12.0-cdh5.0.0-SNAPSHOT/	10-Mar-2014	23:38	-
0.12.0-cdh5.1.0-SNAPSHOT/	10-Mar-2014	21:40	-
0.8.1-cdh4.0.0/	11-Mar-2014	17:39	-
0.8.1-cdh4.0.0-SNAPSHOT/	10-Mar-2014	23:40	-
0.8.1-cdh4.0.1/	11-Mar-2014	17:39	-
0.8.1-cdh4.1.0-SNAPSHOT/	10-Mar-2014	23:40	-
0.8.1-cdh4b2-SNAPSHOT/	10-Mar-2014	23:40	-
0.9.0-cdh4.1.0/	11-Mar-2014	17:39	-
0.9.0-cdh4.1.0-SNAPSHOT/	10-Mar-2014	23:40	-
0.9.0-cdh4.1.1/	11-Mar-2014	17:39	-
0.9.0-cdh4.1.1-SNAPSHOT/	10-Mar-2014	23:40	-
0.9.0-cdh4.1.2/	11-Mar-2014	17:39	-
0.9.0-cdh4.1.2-SNAPSHOT/	10-Mar-2014	23:40	-



MapReduce2.0 源码分析与实战编程

<http://www.aboutyun.com/thread-7240-1-3.html>

MapReduce2.0 源码分析与实战编程



Storm 相关 20 文档与相关包

<http://www.aboutyun.com/thread-7285-1-3.html>

Storm 相关脚本说明文档

Storm 相关问题整理

storm 项目-流数据监控 3 《实例运行》

storm 项目-流数据监控 4 《MetaQ 接口》

storm 项目-流数据监控 5 《zookeeper 统一配置》

<http://www.aboutyun.com/thread-7509-1-2.html>

目录.....1

1. 前言.....1

2. 约定.....1

3. 安装步骤.....2

3.1. 配置/etc/hosts.....2

3.2. 设置 myid.....2

3.3. 修改 conf/zoo.cfg.....2

4. 启动 ZooKeeper 集群.....3

5. 安装验证.....3

Apache Hadoop 2.2.0 API 文档分享

<http://www.aboutyun.com/thread-7428-1-2.html>

所有类 (Apache Hadoop Main 2)

所有类 (Apache Hadoop Main 2)

常量字段值 (Apache Hadoop Main 2)

已过时的列表 (Apache Hadoop Main 2)

API 帮助 (Apache Hadoop Main 2)

Apache Hadoop Main 2.2.0 API

索引 (Apache Hadoop Main 2)

嵌套列表 (Apache Hadoop Main 2)

嵌套 (Apache Hadoop Main 2)

类分层结构 (Apache Hadoop Main 2)

序列化表格 (Apache Hadoop Main 2)

org

上一个 下一个 框架 无框架 所有类

常量字段值

目录

org.apache.*

org.apache.*

org.apache.hadoop.mapreduce.DistributedCache

限定符和类型	常量字段	值
public static final String	CACHE_ARCHIVES	"mapred.cache.archives"
public static final String	CACHE_ARCHIVES_SIZES	"mapred.cache.archives.filesizes"
public static final String	CACHE_ARCHIVES_TIMESTAMPS	"mapred.cache.archives.timestamps"
public static final String	CACHE_FILES	"mapred.cache.files"
public static final String	CACHE_FILES_SIZES	"mapred.cache.files.filesizes"
public static final String	CACHE_FILES_TIMESTAMPS	"mapred.cache.files.timestamps"
public static final String	CACHE_LOCALARCHIVES	"mapred.cache.localArchives"
public static final String	CACHE_LOCALFILES	"mapred.cache.localFiles"
public static final String	CACHE_SYMLINK	"mapred.create.symlink"

Hadoop-2.4.0 分布式安装手册

<http://www.aboutyun.com/thread-7508-1-2.html>

1. 前言
2. 部署
 - 2.1. 机器列表
 - 2.2. 主机名
 - 2.2.1. 临时修改主机名
 - 2.2.2. 永久修改主机名
 - 2.3. 免密码登录范围
3. 约定
 - 3.1. 安装目录约定
 - 3.2. 服务端口约定
4. 工作详单
5. JDK 安装
 - 5.1. 下载安装包
 - 5.2. 安装步骤
6. 免密码 ssh2 登录
7. Hadoop 安装和配置
 - 7.1. 下载安装包
 - 7.2. 安装和环境变量配置
 - 7.3. 修改 hadoop-env.sh

.....



hive 高级编程分享

<http://www.aboutyun.com/thread-5325-1-2.html>

hive 高级编程:

- (1)hive components
- (2)mapReduce
- (3)Hive QL
- (4)Hive 优化
- (5)sql 优化



Yarn 框架代码详细分析

<http://www.aboutyun.com/thread-7425-1-2.html>

3.1.2. MRAppMaster 收集 JobHistory.....	12
3.1.3. JobHistoryServer.....	12
3.2. RM 调度器.....	13
3.2.1. 简述.....	13
3.2.2. FairScheduler 代码分析.....	14
3.2.3. FairScheduler 资源预分配.....	14
3.2.4. FairScheduler 抢占资源.....	15
3.2.5. FairScheduler container 分配.....	15
3.3. MRAppMaster 分配器.....	17
3.3.1. 代码分析.....	17
3.3.2. 任务周期管理及资源分配.....	17
3.3. shuffle.....	19
3.5. NM 的资源下载.....	21
3.6. Metrics 解析.....	22
3.7. YARN 中 WEB 框架解析.....	23
3.7.1. 代码分析.....	23
3.7.2. WEB 的一些功能.....	24
四、配置.....	27
4.1. 配置解析.....	27
4.2. DEBUG 配置.....	27
4.2. 安全配置.....	28



hadoop 大数据

<http://www.aboutyun.com/thread-7226-1-2.html>

BingoCloud 动态生成 hadoop 集群，并在运行中调整集群规模_标清

下面视频为英语视频，有翻译无字幕：

《大数据 大影响--对话大数据 2013 论坛》--维克托教授的演讲_标清

《大数据 大影响--对话大数据 2013 论坛》--维克托教授的演讲_标清



hadoop 图形处理文档下载

<http://www.aboutyun.com/thread-7863-1-2.html>

近年来，云计算在 IT 行业掀起了新一轮技术革新浪潮。云计算是一种新兴的计算模型，它是并行计算、分布式计算、网格计算的综合发展，以简单、透明服务的形式提供无限制的计算资源。Hadoop 实现的开源云平台提供了并行计算模型 MapReduce、分布式文件系统 HDFS 和分布式数据库 HBase 等。随着数字图像数据量不断增长，单机模式的图像处理已逐渐不能满足用户需求。文中提出了利用 Hadoop 云平台实现海量图像的并行化处理，设计了基于 MapReduce 图像处理的类型和格式，实现了图像处理的并行化。



hadoop 在电信上网日志处理的应用架构及在金融行业的应用架构分析

<http://www.aboutyun.com/thread-7694-1-1.html>

Hadoop 技术在电信运营商 上网日志处理中的应用架构

方建国

一、电信运营商上网日志处理的现状

在移动互联网如此普及的今天，每天会产生大量的上网日志，这些上网日志由于数据量巨大，产生后只能被保留 3 天，就因为存储空间等原因被丢弃。目前，电信运营商只能基于 CDR（详细通话记录）为主的客户行为分析可能缺失了大量的客户行为有效信息。例如，两个通话行



redis 学习笔记整理及 MongoDB 文档下载

<http://www.aboutyun.com/thread-7852-1-2.html>

- 一、. redis 环境搭建. 2
- 二、. redis 学习笔记之数据类型. 3
- 三、. redis 学习笔记之排序. 11
- 四、. redis 学习笔记之事务. 16
- 五、. redis 学习笔记之 pipeline 20
- 六、. redis 学习笔记之发布订阅. 23
- 七、. redis 学习笔记之持久化. 28
- 八、. redis 学习笔记之主从复制. 30
- 九、. redis 学习笔记之虚拟内存. 31
- 1NoSQL 简述... 4
- 2MongoDB 简介... 4
- 3 术语介绍... 5
- 4MongoDB 资源消耗... 6

.....
21.1 连接操作相关的命令... 32

- 21.2 对 value 操作的命令... 32
- 21.3 对 String 操作的命令... 33
- 21.4 对 List 操作的命令... 33
- 21.5 对 Set 操作的命令... 34
- 21.6 对 zset (sorted set) 操作的命令... 35
- 21.7 对 Hash 操作的命令... 35
- 21.8 持久化... 36
- 21.9 远程服务控制... 36



hadoop 大数据系列文档下载

<http://www.aboutyun.com/thread-7430-1-2.html>
ANTLR 技术及其在 HQL 语言编译中的应用
APDPlat 应用级产品开发平台简介
flume+kafka+storm 搭建
Google - Bigtable A Distributed Storage System for Structured Data
Google - Spanner Google's Globally-Distributed Database
Hadoop 集群搭建文档
Hadoop 教程 Hive 执行过程
hadoop 最简单部署
HBase The Definitive Guide
HBase 用例分析
Hbase
kafka-and-storm-event-processing-in-realtime-131023085422-phpapp01
Language Implementation Patterns
Lustre 部署软硬件依赖
lustre 性能优化方案
Manning.Hadoop.in.Practice.Oct.2012
MetaQ 调研及安装部署文档
NewSQL - How will the database incumbents respond to NoSQL and NewSQL
NewSQL - Scalable SQL and NoSQL Data Stores
NuoDB - Technical-Whitepaper
O'Reilly 2012.8 第一版 Getting Started with Storm
Oreilly.Getting.Started.with.Storm.中文版
O'Reilly.High.Performance.MySQL.3rd.Edition.Mar.2012
Storm - Getting Started with Storm
the+definitive+antlr+reference--building+domain+specific+languages
VMware_View_医疗桌面虚拟化方案介绍
ZooKeeper A simple totally ordered broadcast protocol
ZooKeeper 在携程的使用及前景
大数据相关框架讲义 (1-11) 杨尚川
发送、接受、转发包处理结构框图
分布式系统原理介绍

流数据监控 1 《设计文档》

流数据监控 2 《代码解析》

实时处理方案架构(作者-落枫)

中文分词算法 之 基于词典的正向最大匹配算法

hadoop2 hbase hive 包下载

<http://www.aboutyun.com/thread-7718-1-2.html>

CDH4-Hadoop_Hive_HBase 安装及 hadoop2.0.0-CDH4.2.0 系列手工安装指南文档下载

<http://www.aboutyun.com/thread-7480-1-2.html>

下面文档还是比较全的，阅读了里面的内容，ssh 登录，可能会遇到问题，大家可以参考[hadoop 学习通过虚拟机安装 hadoop 完全分布式集群](#)中 ssh 的配置

BIEE --商业智能，大数据分享

<http://www.aboutyun.com/thread-7227-1-2.html>

BIEE --1 商业智能，大数据，hadoop 视频学习，biee 学习第一期

BIEE --6 商业智能，大数据，hadoop 视频学习，biee 学习第二期

BIEE --6 商业智能，大数据，hadoop 视频学习，biee 学习第三期

BIEE --6 商业智能，大数据，hadoop 视频学习，biee 学习第四期

BIEE --6 商业智能，大数据，hadoop 视频学习，biee 学习第五期

spark 概述与编程模型

<http://www.aboutyun.com/thread-8074-1-2.html>



pig 实战技 pig 学习总结文档下载

<http://www.aboutyun.com/thread-8101-1-2.html>

Pig 实战-- pig 是 hadoop 项目的一个拓展项目

<1>. Pig 简介

[pig](#) 是 hadoop 项目的一个拓展项目，用以简化 hadoop 编程（简化的程度超乎想象啊），并且提供一个更高层次抽象的数据处理能力，同时能够保持 hadoop 的简单和可靠性。

<2>. 安装 pig

2.1 下载 pig: [点击下载](#) pig 安装包

2.2 解压下载完成的 pig 安装包:

```
xuqiang@ubuntu:~/hadoop/src/pig$ tar zxvf pig-0.8.1.tar.gz
```

2.3 设置环境变量

```
xuqiang@ubuntu:~$ vim .bashrc
```

```
export PATH=~/hadoop/src/pig/pig-0.8.1/bin/:$PATH
```

```
export PIG_HOME=~/hadoop/src/pig/pig-0.8.1/
```

```
export HADOOP_CONF_DIR=~/hadoop/src/hadoop-0.21.0/conf
```

```
export PIG_CLASSPATH=~/hadoop/src/hadoop-0.21.0/conf
```

这里需要说明的是 pig 是能够运行在两种模式下: local 模式和 mapreduce 模式, 变量 HADOOP_CONF_DIR 主要是为了在 mapreduce 模式下使用。



hadoop、spark、hive、hortonworks 等 20 大数据资料文档下载

<http://www.aboutyun.com/thread-7843-1-1.html>

<input type="checkbox"/>	Ganglia系统监控_IT168文库.pdf	6M
<input type="checkbox"/>	尹培森 - talk-slide-final-xusen.pdf	2.2M
<input type="checkbox"/>	Apache Mahout Cookbook-.pdf	6.3M
<input type="checkbox"/>	HWX.Qubole.Hive.UDF.Guide.1.0.pdf	494KB
<input type="checkbox"/>	大数据利器.xls	1.6M
<input type="checkbox"/>	ScalaTutorial中文版.pdf	334KB
<input type="checkbox"/>	学习Hadoop记录125页.doc	4.6M
<input type="checkbox"/>	hadoop大数据入门常见问题.doc	36KB
<input type="checkbox"/>	ZooKeeper-3.4.6分布式安装指南.pdf	130KB
<input type="checkbox"/>	Pig-Cheat-Sheet.pdf	5.3M
<input type="checkbox"/>	Hive.docx	58KB
<input type="checkbox"/>	about云.txt	393B

推荐系统实践电子书分享

<http://www.aboutyun.com/thread-7864-1-1.html>

hadoop 大数据，大家比较感兴趣，这里给大家分享一本来自网络的电子书，书中内容讲的还是比较全面的：

包括：

用户行为分析

推荐系统冷启动

利用用户标签数据

推荐系统实例

目 录

第 1 章 好的推荐系统..... 1	2.5.1 基础算法..... 64
1.1 什么是推荐系统..... 1	2.5.2 基于 LFM 的实际系统的例子..... 70
1.2 个性化推荐系统的应用..... 4	2.5.3 LFM 和基于邻域的方法的比较..... 72
1.2.1 电子商务..... 4	2.6 基于图的模型..... 73
1.2.2 电影和视频网站..... 8	2.6.1 用户行为数据的二分图表示..... 73
1.2.3 个性化音乐网络电台..... 10	2.6.2 基于图的推荐算法..... 73
1.2.4 社交网络..... 12	第 3 章 推荐系统冷启动问题..... 78
1.2.5 个性化阅读..... 15	3.1 冷启动问题简介..... 78
1.2.6 基于位置的服务..... 16	3.2 利用用户注册信息..... 79
1.2.7 个性化邮件..... 17	3.3 选择合适的物品启动用户的兴趣..... 85
1.2.8 个性化广告..... 18	3.4 利用物品内容信息..... 89
1.3 推荐系统评测..... 19	3.5 发挥专家的作用..... 94
1.3.1 推荐系统实验方法..... 20	第 4 章 利用用户标签数据..... 96
1.3.2 评测指标..... 23	4.1 UGC 标签系统的代表应用..... 97
1.3.3 评测维度..... 34	4.1.1 Delicious..... 97
第 2 章 利用用户行为数据..... 35	4.1.2 CiteULike..... 98
2.1 用户行为数据简介..... 36	4.1.3 Last.fm..... 98
2.2 用户行为分析..... 39	4.1.4 豆瓣..... 99
2.2.1 用户活跃度和物品流行度的分布..... 39	4.1.5 Hulu..... 99
2.2.2 用户活跃度和物品流行度的关系..... 41	4.2 标签系统中的推荐问题..... 100
2.3 实验设计和算法评测..... 41	4.2.1 用户为什么进行标注..... 100
2.3.1 数据集..... 42	4.2.2 用户如何打标签..... 101
2.3.2 实验设计..... 42	4.2.3 用户打什么样的标签..... 102
2.3.3 评测指标..... 42	4.3 基于标签的推荐系统..... 103
2.4 基于邻域的算法..... 44	4.3.1 实验设置..... 104
2.4.1 基于用户的协同过滤算法..... 44	4.3.2 一个最简单的算法..... 105
2.4.2 基于物品的协同过滤算法..... 51	4.3.3 算法的改进..... 107
2.4.3 UserCF 和 ItemCF 的综合比较..... 59	4.3.4 基于图的推荐算法..... 110
2.5 隐语义模型..... 64	4.3.5 基于标签的推荐解释..... 112
	4.4 给用户推荐标签..... 115
	4.4.1 为什么要给用户推荐标签..... 115

图灵社区会员 臭豆腐(StinkBC@gmail.com) 专享 尊重版权

hadoop 源码分析文档下载

<http://www.aboutyun.com/thread-7653-1-1.html>

- [Hadoop 二次开发必备,Hadoop 源码分析\(一\)](#)
- [Hadoop 二次开发必备, Hadoop 源码分析 \(二\)](#)
- [Hadoop 二次开发必备, Hadoop 源码分析 \(三\)](#)
- [Hadoop 二次开发必备, Hadoop 源码分析 \(四\)](#)
- [Hadoop 二次开发必备, Hadoop 源码分析 \(五\)](#)
- [Hadoop 二次开发必备, Hadoop 源码分析 \(六\)](#)

- [Hadoop 二次开发必备,Hadoop 源码分析\(七\)](#)
- [Hadoop 二次开发必备, Hadoop 源码分析 \(八\)](#)
- [Hadoop 二次开发必备, Hadoop 源码分析 \(九\)](#)
- [Hadoop 二次开发必备, Hadoop 源码分析 \(十\)](#)
- [Hadoop 二次开发必备, Hadoop 源码分析 \(十一\)](#)
- [Hadoop 二次开发必备, Hadoop 源码分析 \(十二\)](#)
- [Hadoop 二次开发必备,Hadoop 源码分析\(十三\)](#)
- [Hadoop 二次开发必备, Hadoop 源码分析 \(十四\)](#)
- [Hadoop 二次开发必备, Hadoop 源码分析 \(十五\)](#)
- [Hadoop 二次开发必懂: 源码分析 MapReduce 概论及 MapTask\(1\)](#)
- [Hadoop 二次开发必懂: 源码分析 MapTask 辅助类\(2\)](#)
- [Hadoop 二次开发必懂: 源码分析 Task 的内部类和辅助类\(3\)](#)
- [Hadoop 二次开发必懂: 源码分析 mapreduce.lib.partition/reduce/output\(4\)](#)



让你了解 **hadoop** 生态系统的一本书

<http://www.aboutyun.com/thread-7453-1-1.html>

内容还是比较全的: 总共 18 章 :

详细介绍了:

hadoop、mapreduc、HDFS、Hbase、Mahout、Pig、Zookeeper、Avro、Chukwa。

第 1 章 Hadoop 简介 /1

1.1 什么是 Hadoop/2

1.1.1 Hadoop 概述 /2

1.1.2 Hadoop 的历史 /2

1.1.3 Hadoop 的功能与作用 /2

1.1.4 Hadoop 的优势 /3

1.1.5 Hadoop 的应用现状和发展趋势 /3

1.2 Hadoop 项目及其结构 /3

1.3 Hadoop 的体系结构 /6

1.3.1 HDFS 的体系结构 /6

1.3.2 MapReduce 的体系结构 /7

篇幅有限，更多可以查看本贴



hadoop 的 CDH4 及 CDH5 系列文档下载分享

<http://www.aboutyun.com/thread-6739-1-1.html>

CDH4-High-Availability-Guide







CDH4-Installation-Guide

CDH4-Quick-Start

CDH5-Installation-Guide

CDH5-Quick-Start

CDH5-Release-Notes

-  CDH4-High-Availability-Guide.pdf
-  CDH4-Installation-Guide.pdf
-  CDH4-Quick-Start.pdf
-  CDH5-Installation-Guide.pdf
-  CDH5-Quick-Start.pdf
-  CDH5-Release-Notes.pdf



看看老外的视频，怎么教你安装 **Hive**、**Sqoop**

<http://www.aboutyun.com/thread-7693-1-1.html>

我们都看的是国内的视频，国外的视频质量确实高，他们流利的英语，有点难懂。

但是因为你自己的场景，不听他讲话，也大概知道他做什么，所以这个确实锻炼英语的最好时机。

下面分别是安装 hive 与 sqoop。



hadoop2.4 最新版本各种包下载

<http://www.aboutyun.com/thread-7688-1-1.html>

安装文档：

[hadoop2.4.0 伪分布安装配置](#)

[CentOS6 hadoop2.4 完全分布式安装文档](#)

编译包：**hadoop-2.4.0-src.tar.gz** 百度网盘：

链接：<http://pan.baidu.com/s/1jGyZUDK> 密码：kfnr

360 网盘：

<http://yunpan.cn/QiuGKPf6ihGwz> 访问密码 3d7e

64 位下载包下载：

hadoop-2.4.0-64bit.tar.gz

百度网盘：

链接：<http://pan.baidu.com/s/1hqEDe2S> 密码：224f

链接：<http://pan.baidu.com/s/1gdl95OB> 密码：h6dc

360 网盘

<http://yunpan.cn/QiuGA5BtVWbWk> 访问密码 13c1

hadoop-eclipse-plugin-2.4.0 插件包:

链接: <http://pan.baidu.com/s/1jGJX9Im> 密码:



HDFS 定义、架构基本原理及 HDFS2.0 介绍视频下载

<http://www.aboutyun.com/thread-7687-1-1.html>

HDFS 部署及 HDFS 命令的使用

HDFS 定义、架构基本原理及 HDFS2.0 介绍

理解 hadoop 集群及网络拓扑, HDFS 关键流程



百度网盘:

理解 **hadoop** 集群及网络拓扑, **HDFS** 关键流程

链接: <http://pan.baidu.com/s/1c0vitxe> 密码: 0y24

HDFS 部署及 **HDFS** 命令的使用

链接: <http://pan.baidu.com/s/1sj79fwp> 密码: s7xl

HDFS 定义、架构基本原理及 **HDFS2.0** 介绍

链接: <http://pan.baidu.com/s/1hqj4RxM> 密码: d7cw



HDFS 必读书籍: HDFS-Hadoop 分布式文件系统深度实践

<http://www.aboutyun.com/thread-7454-1-1.html>

目 录

第 1 章 HDFS HA 及解决方案	1
1.1 HDFS 系统架构	2
1.2 HA 定义	3
1.3 HDFS HA 原因分析及应对措施	4
1.3.1 可靠性	4
1.3.2 可维护性	5
1.4 现有 HDFS HA 解决方案	5
1.4.1 Hadoop 的元数据备份方案	6
1.4.2 Hadoop 的 SecondaryNameNode 方案	7
1.4.3 Hadoop 的 Checkpoint ode 方案	7
1.4.4 Hadoop 的 BackupNode 方案	8
1.4.5 DRDB 方案	9
1.4.6 FaceBook 的 AvatarNode 方案	10
1.5 方案优缺点比较	10
第 2 章 HDFS 元数据解析	13
2.1 概述	14
2.2 内存元数据结构	14
2.2.1 INode	15
2.2.2 Block	16
2.2.3 BlockInfo 和 DatanodeDescriptor	17
2.2.4 小结	17
2.2.5 代码分析——元数据结构	18
2.3 磁盘元数据文件	24
2.4 Format 情景分析	27
2.5 元数据应用场景分析	45



hadoop、hive、storm 文档、电子书籍分享

<http://www.aboutyun.com/thread-7431-1-1.html>

Hadoop 权威指南(第 2 版)

HADOOP 权威指南++中文版

Hadoop 应用开发技术详解

Hive 高级编程

hive 与 hbase 整合经验谈

storm 环境搭建 v1

站在 hadoop 上看 hive



spark 系列:文档、视频、jar 包

<http://www.aboutyun.com/thread-7619-1-1.html>

1 Andy Konwinski - The Spark Project Today and What's Next

2 陈超 - Spark 内核探索与性能优化

- 3 John Busch - Spark 在闪存中的优化
- 4 孙元浩 - Transwarp Inceptor-基于 Spark 的内存分析和挖掘引擎
- 5 顾荣 - YAFIM: 基于 Spark 的并行化频繁项集挖掘算法
- 6 易岚 - Spark 实际应用模式的经验共享
- 7 邢刚 - Spark 在中国移动省公司试点应用
- 9 明风 - 基于 Spark Graphx 的大规模用户图计算和应用
- 10 连城 - Catalyst_Spark a Chain Reaction
- 11 王健宗 - 面向实时计算和 ETL 处理的 Shark 应用
- 12 尹绪森 - talk-slide-final-xusen
- 13 薛伟 - 快速模型更新在精准广告推荐中的应用

1 Andy Konwinski - The Spark Project Today and What's Next.pdf	592.33 KB
2 陈超 - Spark内核探索与性能优化.pdf	1.89 MB
3 John Busch - Spark在闪存中的优化.pdf	3.24 MB
4 孙元浩 - Transwarp Inceptor-基于Spark的内存分析和挖掘引擎.pdf	2.68 MB
5 顾荣 - YAFIM : 基于Spark的并行化频繁项集挖掘算法.pdf	742.20 KB
6 易岚 - Spark实际应用模式的经验共享.pdf	1.02 MB
7 邢刚 - Spark在中国移动省公司试点应用.pdf	1.68 MB
9 明风 - 基于Spark Graphx的大规模用户图计算和应用.pdf	1004.97 KB
10 连城 - Catalyst_Spark a Chain Reaction.pdf	1.25 MB
11 王健宗 - 面向实时计算和ETL处理的Shark应用.pdf	1.75 MB
12 尹绪森 - talk-slide-final-xusen.pdf	2.19 MB
13 薛伟 - 快速模型更新在精准广告推荐中的应用.pdf	738.67 KB



五、spark 的编程语言是 Scala30 文档分享

<http://www.aboutyun.com/thread-7116-1-1.html>

Scala 是一种多范式的编程语言，支持命令式、函数式和面向对象的编程范式。从面向对象的角度来看，Scala 中的每个值都是一个对象。同理，从函数式编程的角度来看，每个函数也都是一个值。Scala 还是一种静态类型语言，其类型系统表达能力强且安全。

此外，Scala 还是一种虚拟机语言，Scala 编译器生成字节码，使用 JRE2 直接在 Java 虚拟机（JVM）上运行。这样，Scala 可以在几乎任何支持 JVM 的地方运行（需要增加 Scala 运行时库），并使用已有的 Java 库和 Java 代码。

篇幅有限部分内容如下：

- 1.Scala 讲座
- 2.[Lift.Cookbook(2013.6)].Richard.Dallaway.文字版
- 3.Actors in Scala
- 4.Arduino in Action
- 5.Artima.Akka.Concurrency.2013
- 6.AW.Scala for the Impatient.2012
- 7.BeginningScala
- 8.Dart_in_action
- 9.english-Programming Scala-铅笔书-p226

10.Functional_Programming_in_Scala

11.IBM: 面向 Java 开发人员的 Scala 指南-p178

链接: <http://pan.baidu.com/s/1kT1aoXT> 密码: 98jd



淘宝云百度云去哪酒店实时搜索等资料下载

<http://www.aboutyun.com/thread-7160-1-1.html>

- [1] 阿里分布式数据库服务原理与实践
- [1] 百度媒体云技术及架构
- [1] 互联网系统的稳定性保证：微博的实践
- [1] 去哪儿酒店实时搜索技术分享
- [2] Nashorn 与 JDK8——动态语言在 JVM 上的高性能实现
- [2] 后端-大型网站 SEO 优化实践
- [2] 基于 OpenStack 构建网易云主机服务
- [2] 架构模式与实践漫谈
- [2] 淘宝文本中的语义分析以及技术展望
- [2] 淘宝消息中间件技术演变
- [2] 天猫个性化推荐架构



淘宝系列、云计算文档下载

<http://www.aboutyun.com/thread-8122-1-1.html>

淘宝网图片存储系统架构

对于图片存储，淘宝自己开发了 TFS 并且开源。想解决图片存储的问题，可以借鉴淘宝的经验

可以参考: [淘宝网图片存储系统架构](#)

高并发环境下的数据产品架构设计

Agenda

- 什么是大数据产品
- 数据产品的业务场景
- 技术架构
 - 高性能网站
 - 应用计算
 - 云计算
- 走过的一些弯路
- Q & A

淘宝云梯分布式计算平台整体架构



淘

目录

系统架构

数据同步方案

调度系统

元数据应用



Spark 源码研究分享：整体架构和 rdd 计算流程视频

<http://www.aboutyun.com/thread-8009-1-1.html>



openstack 系列

云里雾里云计算

<http://www.aboutyun.com/thread-8201-1-1.html>

第 1 章 云里雾里的云计算.....3

序..... 3

1.1 云计算要解决什么问题? 3

.....

第 3 章 相关论文资料..... 91



openstack 开发，Python 系列最全文档书籍下载

<http://www.aboutyun.com/thread-7853-1-1.html>

新手指导线路指导：

书籍：python 简明教程---》征服 python-语言基础与典型应用--->python 核心编程第二版---》python 网络编程--->Django 讲解---->python 源码剖析

工具：wingide.rar

python 写界面：pyqt4.exe

●●●●●●●●●●●●●●●●



台湾辅仁大学--Python 视频分享

<http://www.aboutyun.com/thread-8173-1-1.html>



老外 python 视频 30 讲、python100 例、python100 习题等系列汇总

<http://www.aboutyun.com/thread-8095-1-1.html>

老外 python 视频 30 讲，根据实际需求，练练检测自己的英语吧

[Python 教程]-1.-.Installing_Python_video

[Python 教程]-2.-.Numbers_and_Math_video

[Python 教程]-3.-.Variables

.....

[Python 教程]-28.-.Default.Parameters

[Python 教程]-29.-.Multiple.Parameters

[Python 教程]-30.-.Parameter.Types

[Python 教程]-31.-.Parameter.Types



openstack 相关文档分享

<http://www.aboutyun.com/thread-6741-1-1.html>

openstack

openstack-basic-install-apt-grizzly

openstack-install-guide-yum-havana

Openstack 管理功能测试文档



混合云管理平台 CloudForms 简介

<http://www.aboutyun.com/thread-8008-1-1.html>

本译文为 openstack-install-guide-apt-debian-havana 的第四章——身份认证服务

详见附件



在 Debian 上部署 OpenStack 官方文档翻译 1

<http://www.aboutyun.com/thread-7255-1-1.html>

在 Debian 上部署 OpenStack 官方文档翻译 2

<http://www.aboutyun.com/thread-7665-1-1.html>

在 Debian 上部署 OpenStack 官方文档翻译 3

<http://www.aboutyun.com/thread-7961-1-1.html>

本译文为 openstack-install-guide-apt-debian-havana 的第四章——身份认证服务



OpenStack 开发之 Python 资料大全汇总

<http://www.aboutyun.com/thread-7950-1-1.html>

下面是其资源汇总：

- 1、[openstack 编程，Python 系列文档下载](#)
- 2、[Python 核心编程下载分享](#)
- 3、[openstack 编程：Python 标准库分享](#)
- 4、[openstack 编程：Python 学习手册 第3版](#)
- 5、[Python Cookbook（第2版）中文版](#)
- 6、[openstack 开发，Python 系列最全文档书籍下载](#)

.....



OpenStack Icehouse 版本官方安装教程-英文版

<http://www.aboutyun.com/thread-7617-1-1.html>

OpenStack Icehouse 版本官方安装教程-英文版哦，如果英语有些吃力的话，可以试着网上翻译。

这里仍然延续着以前的安装教程，采取的是三节点安装。

.....



openstack-H 版 Centos6.4 下安装单网卡安装

<http://www.aboutyun.com/thread-7426-1-1.html>

说明：

openstack 的安装的资料已经不少了，这里分享的是单网卡，对于只是安装过一次 openstack 的新手，可以多参考几份资料，可以看到其中差异，也能够明白很多参数的意义。

.....



CloudComputing 相关英文文档

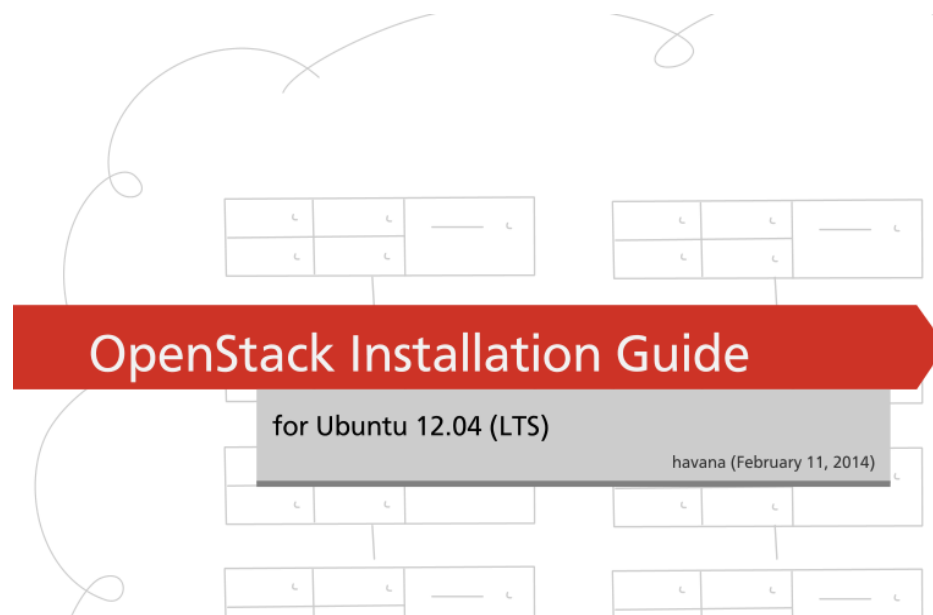
<http://www.aboutyun.com/thread-7096-1-1.html>

Cloud.Application.Architectures).George.Reese
Cloud.Computing Automating the Virtualized
cloud-computing-implementation-management-and-security
CloudComputingWebBasedApplications



openstack 安装英文版 openstack-install-guide-ubuntu12_04-apt-trunk

<http://www.aboutyun.com/thread-7052-1-1.html>



openstack 入门、建设公有云、私有云、商业模式、云存储等系列文档下载

<http://www.aboutyun.com/thread-8199-1-1.html>

本文链接

<http://www.aboutyun.com/thread-8203-1-1.html>

- OpenStack on NSA.pptx
- openstack_nova_ubuntu12.04安装文档.docx
- openstack部署开发高可用手册.docx
- openstack入门线路指导.pdf
- OpenStack商业模式探讨v1.pdf
- OpenStack完整安装手册(CentOS6.2)_by_yz.doc
- openstack云平台分析与比较.pptx
- 程辉-OpenStack建设公有云平台实践.pdf
- 邓昶-OpenStack实践与思考.pdf
- 刘继伟：Openstack构建企业私有云.pdf
- 使用Ubuntu安装OpenStack私有云-stack.sh安装.pdf



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.qqzone.qq.com/278595437/main>

about 云官方群 **39327136**、**371358502**

云计算爱好者群

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>