knowledge system

# Current Plan

## 学习中

Day:

1. Databricks Blog

书籍：

1. CoolplaySpark

2.'Reilly.Effective.Akka.Aug.2013

3. Akka Essentials

4. Hadoop技术内幕深入解析YARN架构设计与实现原理(董西成)

5. Hadoop权威指南（第四版）

6. Learning Real-time Processing with Spark Streaming

参考资料：

## 待学习

技术：

1. Docker

框架：

1. Strata ： Strata+Hadoop World大会

Apache Flink

Apache Samza

Google Cloud Data Flow

StreamSets

Tensor Flow

Apache NiFi

Druid

LinkedIn WhereHows

Microsoft Cognitive Services

书籍：

数据库系统：

 1. 数据库系统概念

2. 数据库：原理、性能与编程

2. O'Reilly ： Building Real-Time Data Pipelines

平台：

1. 百度开放云  : https://bce.baidu.com/doc/BAE/QuickGuide.html

## 完成列表

## 学习内容

框架：

1. Akka

2. Kafka

3. Spark

4. Hadoop

5. Hive

6. Flume

技术：

1. Docker

平台：

1. Databricks ： 社区版

2. Cloudera ：CM，CDH

3. Amazon ： Kinesis

4. 数据科学家集成工作台

# Bigdata Projects

## dr-elephant

来自LinkedIn公司的开放源代码软件

https://github.com/linkedin/dr-elephant

https://github.com/linkedin/dr-elephant/wiki

https://groups.google.com/forum/#!forum/dr-elephant-users

性能监控和调优工具的项目-开源项目Dr.Elephant：是Hadoop和Spark 性能监控和调优工具。它会自动收集所有的指标，针对指标执行分析任务，并用简易的方式将其展示出来。它的目标是提高开发人员的工作效率，简化任务调优工作以更好的提高集群性能。它使用一组可插拔、可配置、基于规则的启发式（heuristics）来分析Hadoop和Spark任务并给出关于任务性能的详细报告信息，然后，基于该结果给出如何让任务执行得更为有效建议。

比较：

1. 大多数Hadoop的优化工具，无论是开源还是专有的，旨在收集系统资源指标和监控集群资源。

## BlueData平台

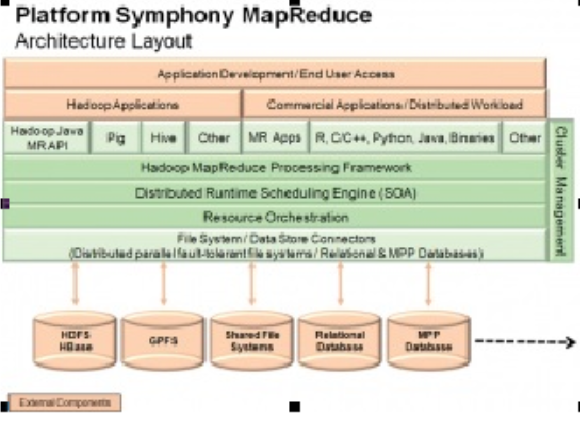
介绍文档：http://wenda.chinahadoop.cn/question/2950#rd?sukey=e9670b423d2de698983b65a386814f0eeb03ca723a1fbe27101e6fd1a0ea6af911702406af2c722d0dadaec220e9ebf7

 BlueData公司发言人表示，启动BlueData正在寻求应对EPIC软件平台的挑战，简称弹性私人即时集群。该软件支持ApacheSpark处理引擎，以及来自Cloudera和Hortonworks的Hadoop分布。此外，BlueData平台使用Docker容器技术，其封装的应用及其配套软件作为一个独立的单元，对Hadoop的部署提供多租户的方式。

参考资料：http://www.d1net.com/bigdata/news/409537.html

## Platform MapReduce

Platform MapReduce提供了企业级可管理性和可伸缩性、高资源利用率和可用性、操作便利性、多应用支持以及一个开放分布式系统架构，其中包括对于Hadoop分布式文件系统（HDFS）和Appistry Cloud IQ的即时支持，稍后还将支持更多的文件系统和平台，这将确保企业更加关注将MapReduce应用程序转移至生产环境中。



## Apache Hadoop

### MapReduce

### HDFS

### YARN

### Common

## Apache Spark

### 00 基础

### 01 Spark 入门

### 02 SPARK 核心组件

#### Spark SQL

Spark SQL 源码解析

Catalyst

Spark SQL 应用实战

#### Spark MLlib

类库设计文档

底层类库等

jira: issues

SPARK-1856：Standardize MLlib interfaces

SubTask - Pipeline and Parameters

##### 论文与书籍

MLlib: Machine Learning in Apache Spark

http://arxiv.org/pdf/1505.06807v1.pdf

Deliver-the-Power-of-ML.pdf

### 03 SPARK 调优

### 04 SPARK 应用开发案例实战

### 05 生态系统

### 06 Spark 源码解析

### 07 Spark 第三方

### 08 参考资料

## Apache Ambari

Apache Ambari是对Hadoop进行监控、管理和生命周期管理的开源项目。它也是一个为Hortonworks数据平台选择管理组建的项目。Ambari向Hadoop MapReduce、HDFS、 HBase、Pig, Hive、HCatalog以及Zookeeper提供服务。

新的仪表盘和可视化数据也被添加到Ambari，这是Hortonworks公司开发的一个开源的Hadoop管理工具。而推出的Ambari2.2.2版本将包括预建的仪表盘，可以让Hadoop的系统管理员对使用的跨集群资源和集群整体健康指标实现可视化。

## StackIQ Rock+ Big Data

StackIQ Rock+ Big Data是一款Rocks的商业流通集群管理软件，该公司已加强支持Apache Hadoop。Rock+支持Apache、Cloudera、Hortonworks和MapR的分布，并且处理从裸机服务器来管理Hadoop集群配置的整个过程。

## Zettaset Orchestrator

Zettaset Orchestrator是端到端的Hadoop管理产品，支持多个Hadoop的分布。Zettaset吹捧Orchestrator的基于UI的经验和MAAPS（管理、可用性、自动化、配置和安全）的处理能力。

## Apache Apex

Spark与Storm注意：Apache Apex已然横空出世

http://developer.51cto.com/art/201604/510227.htm

## Seahorse

https://seahorse.deepsense.io/index.html

Introduction

Seahorse is an open-source visual framework allowing you to create [**[](https://spark.apache.org/)**](https://spark.apache.org/)applications in a fast, simple and interactive way. Creating Spark applications with Seahorse is as easy as dragging and dropping operation on the canvas, all while connected to any Spark Cluster (YARN, Mesos, Standalone) or to a bundled local Spark.

# 图

## cayley

An open-source graph database

https://github.com/google/cayley

Cayley是Google开源的图数据库，收到Freebase和Google的知识图谱（Google's Knowledge Graph）的启发，其目标是成为链接数据和图形状数据的工具箱，譬如语言Web，社交网络等等方面。

## GraphLab

https://turi.com/learn/

https://turi.com/products/create/open\_source.html

https://github.com/turi-code/SFrame

## Resource manage

资源统一管理系统或者资源统一调度系统：这类系统设计动机是解决以下两类问题：

参考资料：

浅谈Borg/YARN/Mesos/Torca/Corona一类系统http://dongxicheng.org/mapreduce-nextgen/borg-yarn-mesos-torca-corona/

多集群下资源共享方案介绍 http://dongxicheng.org/mapreduce-nextgen/multi-clusters-sharing-resources/

统一资源管理与调度平台（系统）介绍 http://dongxicheng.org/mapreduce-nextgen/mesos\_vs\_yarn/

### Apache Mesos

#### Mesos at Twitter

参考资料：

论文 - Mesos: A Platform for Fine-Grained Resource Sharing in the Data Center - http://mesos.berkeley.edu/mesos\_tech\_report.pdf

## H2O

## Apache Storm

## Apache Flume

## Dremel

交互式SQL查询系统

## HaLoop

一种迭代MapReduce的系统

会在各步骤中都以一种高效率的方式对数据进行分区

## MillWheel

支持连续地流式处理

## F1

A distributed sql database that scales

一个快速但没有容错的SQL查询系统

## Apache Kylin

开源的分布式分析引擎

提供Hadoop智商的SQL查询接口及多维分析（OLAP）能力以支持超大规模数据，最初由eBay Inc. 开发并贡献至开源社区。

提供亚秒级的交互式分析能力

Extreme OLAP Engine for Big Data

### Caravel 可视化工具

sCaravel 可视化工具与Apache Kylin间的整合。

# Monitor

## OS提供的监控工具

## JVM提供的监控工具

## 开源监测工具

http://www.csdn.net/article/1970-01-01/2814389

### collectd

### Nagios

### Ganglia

### Graphite

### Pingdom

### Munin

### osquery

Facebook开源操作系统监控工具

# Performance optimization

## perfj

https://github.com/coderplay/perfj

PerfJ is a wrapper of linux perf for java programs.

In order to profile java programs, you need a profiler that can sample stack traces. There has historically been two types of profilers:

System profilers: like Linux perf, which shows system code paths (eg, JVM GC, syscalls, TCP), but not Java methods.

JVM profilers: like hprof, LJP, and commercial profilers. These show Java methods, but usually not system code paths.

## naver/pinpoint

https://github.com/naver/pinpoint

## 诊断工具

你可以在市场上已有的各种性能诊断、追踪，或是APM工具之间随意选择。

免费的Dynatrace Personal License

Java本身也提供了各种出色的工具，例如Java Mission Control等等

在使用这些跟踪工具时，通常不需要对代码进行任何修改，因为他们都利用了JVMTI（JVM Tooling Interface）以捕获代码层面的信息，甚至能够跨远程的各层次进行调用追踪，这一点对于分布式、面向（微）服务的应用来说非常实用。你所要做的就是修改你的JVM启动命令行选项，以加载这些工具。

某些工具的开发商还提供了与IDE的集成功能，你只需简单地表示“在运行时开启XYZ性能诊断功能”。我在YouTube上做了一个简单的视频指南，演示了如何对在Eclipse中启动的应用进行追踪。（https://www.youtube.com/watch?v=unrey8wfq-M&list=PLqt2rd0eew1bmDn54E2\_M2uvbhm\_WxY\_6&index=14）

来源：

诊断Java代码中常见的 数据库性能热点问题应该这么做

http://mt.sohu.com/20160427/n446384040.shtml

### Dynatrace Personal License

### Java工具：Java Mission Control

### 扩展：JVMTI（JVM Tooling Interface）

## 书籍

《性能之巅:洞悉系统、企业与云计算》

《软件性能详解与案例分析》

# Security

## Book: Hadoop security

## 黑客与安全

### Metasploit

Metasploit是一款开源的安全漏洞检测工具,可以帮助安全和IT专业人士识别安全性问题,验证漏洞的缓解措施,并管理专家驱动的安全性进行评估,提供真正的安全风险情报。

# Search Engine

## Solr

## Lucene

## ElasticSearch

## 分词技术

### Jcseg

Jcseg是基于mmseg算法的一个轻量级开源中文分词器，同时集成了关键字提取，关键短语提取，关键句子提取和文章自动摘要等功能，并且提供了最新版本的lucene, solr, elasticsearch的分词接口。 <http://git.oschina.net/lionsoul/jcseg>

# Distributed Messaging System

## ZeroMQ

## Akka

## Kafka

# Language

## Java

### j-zerocopy

### NIO

内存映射文件（MEMORY MAPPED FILE）

http://javarevisited.blogspot.hk/2012/01/memorymapped-file-and-io-in-java.html

http://blog.csdn.net/hongchangfirst/article/details/11599369

http://www.linuxtopia.org/online\_books/programming\_books/thinking\_in\_java/TIJ314\_029.htm

## Scala

## Python

# Project Manage

## Git

\*\*\* git init : 初始化一个仓库，会在目录下生成一个.git目录；

\*\*\* git status : 查看当前状态；

\*\*\* git log : 查看提交日志；

-> git log --graph --pretty=format:'%Cred%h%Creset -%C(yellow)%d%Creset %s %Cgreen(%cr)%Creset' --abbrev-commit --date=relative

注意: 进入 log 后，用vi命令进行操作，如q退出；

\*\*\* git add ... : 增加一个文件到暂存空间；

-> git add newfile git rm --cached ...

\*\*\* git commit : 将提交到暂存空间中的修改提交到仓库； -- record changes torepository : git commit -m"message"

-> git commit -a -m"message" :针对全部已经跟踪的文件 —— 对已经跟踪的文件有效 :

-> -v : verbose 打印详细信息

-> - diff信息 :

- > -m : 设置提交信息，没有写的话会弹出窗口，在窗口中添加提交信息，此时可以多行提交 —— git必须设置提交信息

\*\*\* git commit --amend : 修改上次提交

-> git commit --amend --reset-author : 修改作者信息 # 显示、配置远程地址

-> add - 增加一个远程地址

-> -v : 显示详细信息 -> git remote add origin URL - 添加一个origin的地址URL

-> git remote - 显示当前远程地址 :

\*\*\* git push

-> git push origin master - push 到 origin 的 master 分支上 \*\*\* gti fetch

## Svn

## Maven

## Sbt

# 通用工具

## Netcat

a small utility found in most Unix-like systems

nc -lk 9999

## Log tools

Log4j

Logback

# File System / Storage System

## Tachyon / Alluxio

## Orange FS

## Cluster FS

## Ceph

## NFS

# AI-ML

机器学习项目

<http://www.kdnuggets.com/2015/12/top-10-machine-learning-github.html>

## DeepLearning

### 课程

<https://classroom.udacity.com/courses/ud730/lessons/6370362152/concepts/63798118150923>

## tensorflow

https://github.com/tensorflow/tensorflow

## 参考

### AI, Deep Learning, and Machine Learning: A Primer

http://a16z.com/2016/06/10/ai-deep-learning-machines/

## DeepDive

http://www.cad.zju.edu.cn/home/vagblog/?p=3010

DeepDive是一个可以通过机器学习技术利用特定领域知识，并且通过用户反馈提高分析质量的开源知识提取系统。

其目标为帮助用户从数据中提取实体和关系，并对事实经行推理。

DeepDive可以处理结构化/非结构化、干净/带噪声的数据，并将结果输出至数据库。

主要面向从互联网非结构化数据中抽取结构化信息，做一系列后处理，构建知识库并抽取关系等。

## Algorithms

### 排序算法

#### EdgeRank算法

参考资料：

[1] https://en.wikipedia.org/wiki/Criticism\_of\_Facebook#News\_Feed\_and\_Mini-Feed

[2] http://edgerank.net/

[3] http://marketingland.com/edgerank-is-dead-facebooks-news-feed-algorithm-now-has-close-to-100k-weight-factors-55908

[4] https://research.facebook.com/serkan-piantino

[5] https://research.facebook.com/publications/machinelearning

[6] http://www.dailydot.com/technology/facebook-news-feed-algorithm-edgerank/

[7] https://www.facebook.com/business/news/News-Feed-FYI-Showing-More-High-Quality-Content

[8] https://pdfs.semanticscholar.org/daf9/ed5dc6c6bad5367d7fd8561527da30e9b8dd.pdf

[9] http://www.slate.com/articles/technology/cover\_story/2016/01/how\_facebook\_s\_news\_feed\_algorithm\_works.html

[10] https://code.facebook.com/posts/861999383875667/recommending-items-to-more-than-a-billion-people/

[11] http://zhihu.com/question/20553088/answer/78721851

[12] http://weibo.com/1111681197/D0qNcDlcc

[13] http://techcrunch.com/2014/04/03/the-filtered-feed-problem/

#### Facebook NewsFeed

关于 Facebook NewsFeed，看这一篇就够了！

http://mp.weixin.qq.com/s?\_\_biz=MzA4OTk5OTQzMg==&mid=301747355&idx=1&sn=afc8439c0f7e1404620b2e2c9b0c7fef#rd

EdgeRank，顾名思义，就是对边（Edge）进行排序（Rank）。

每一个朋友的每一个操作（Action）都可能以新鲜事（News）的方式呈现在用户面前，比如朋友发布了一条新鲜事，朋友赞了朋友的朋友一条新鲜事，朋友评论了朋友的朋友一条新鲜事，朋友给自己的照片加了一个标签，等等。稍微用脚后跟想一想就知道，按照这样罗列的话，每个用户得看多少新鲜事？于是 FB 就想在一条动态呈现给一个用户之前，先预估一下这个用户对这条新鲜事感兴趣程度。在 FB 的定义下，一个 Action 就是一个 Edge，所以这套量化兴趣的算法就叫做 EdgeRank。

EdgeRank 主要有三个因素在起作用：

·         亲密度（Affinity Score）

·         边的权重（Edge Weight）：反映了产生一条新鲜事的成本，成本越高的，权重越大

·         新鲜程度（Time Decay）：FB 用了一个类似指数衰减的函数来量化动态的新旧程度。

2011年之后，Facebook 内部就不再提 EdgeRank 算法了，因为用户数和广告主的飞增，导致 NewsFeed 的排序算法必须要更上一层楼。

2011年之后的 NewsFeed 排序算法，全面转向了机器学习，用优化理论来决定每个因素的权重，将人从繁杂的策略量化中解放出来。

# 调度

## Stolos

https://github.com/sailthru/stolos

Stolos is a task dependency scheduler that helps build distributed pipelines. It shares similarities with Chronos, Luigi, and Azkaban, yet remains fundamentally different from all three.

Stolos是一个任务以来调度器，可以帮助构建分布式的pipelines。

类似的其他项目有 Chronos、Luigi和Azkaban，Stolos和这些项目有相似之处，但也有完全不同的地方。

# 虚拟现实（VR）

# Company & Universities

## Hortonworks

### HDP

### Ambari

## Cloudera

### CDH

### Impala

交互式查询

## Twitter

### Storm

### Heron

## Amazon

### Amazon Kinesis

http://aws.amazon.com/cn/kinesis/

凭借 Amazon Kinesis 服务，您可以轻松地在 AWS 云中处理实时流数据。

Amazon Kinesis 是一种在 AWS 上流式处理数据的平台，提供了多种强大的服务，让您可以轻松地加载和分析流数据，同时还可让您根据具体需求来构建自定义流数据应用程序。Web 应用程序、移动设备、可穿戴设备、行业传感器和许多软件应用程序和服务都可能生成大量的流数据（有时达到每小时数 TB），需要对其进行连续地收集、存储和处理。Amazon Kinesis 服务让您能以较低的费用实现该目的。

Amazon Kinesis Firehose   --  将大量流数据轻松加载至 AWS

Amazon Kinesis 分析         --  使用标准 SQL 轻松分析流数据

Amazon Kinesis Streams    --  构建自己的自定义应用程序来处理或分析流数据

#### Amazon Kinesis Firehose

轻松将大量流数据加载至 AWS。

#### Amazon Kinesis Analytics

使用标准 SQL 轻松分析流数据。

#### Amazon Kinesis Streams

构建自己的自定义应用程序来处理或分析流数据。

### Dynamo

亚马逊的key-value模式的存储平台

### Amazon Redshift

处理SQL的DBMS

### S3

## CMU（卡内基梅隆大学）

### GraphLab

机器学习系统

## 腾讯

Torca（腾讯搜搜）

## Facebook

Corona at Facebook

## Google

### Borg at Google

sAccording to a recent story in WIRED, it's the 10+ year old resource orchestration software used to control Google's datacenter infrastructure and deploy/manage running applications

参考资料 ：

http://www.quora.com/What-is-Borg-at-Google

### Pregel

图计算处理系统

bulk-Synchronous parallel（BSP）并行迭代图计算模型

它会将中间状态保存在内存中

## Microsoft

### Azure

### Dryad

微软Dryad分布式并行计算平台解析

## berkeley

https://inst.eecs.berkeley.edu

## Databricks

### rxin

https://rxin.github.io/

## Linkedin

https://engineering.linkedin.com/blog

https://github.com/linkedin

https://engineering.linkedin.com/kafka/running-kafka-scale

http://www.confluent.io/resources

http://docs.confluent.io/2.0.1/

https://cwiki.apache.org/confluence/display/KAFKA/Kafka+papers+and+presentations

http://zqhxuyuan.github.io

# CEP（Complex Event Process）

## Esper

原来：http://esper.codehaus.org/

现在：https://github.com/codehaus/

esper的官网 http://www.espertech.com/

CEP即Complex Event Process，中文意思就是“复杂事件处理”。听起来好像很复杂，实际上就是基于事件流进行数据处理，把要分析的数据抽象成事件，然后将数据发送到CEP引擎，引擎就会根据事件的输入和最初注册的处理模型，得到事件处理结果。

参考资料：

Esper-技术简介 :  http://www.cnblogs.com/qlee/archive/2011/06/22/2086550.html

# Resources

## Cgroups

# deployment & integrate

## 虚拟技术

## DevOpts

## Docker

# OS

## 书籍

## Linux 操作系统经典

深入理解linux虚拟内存管理

深入Linux内核架构

深入理解linux内核

linux内核设计的艺术图解linux操作系统架构设与实现原理

莱昂氏unix源代码分析

Linux内核设计与实现

深度探索linux操作系统

linux内核完全剖析

linux内核修炼之道

linux内核精析

cpu源代码分析与芯片设计及linux移植 pdf

linux运维之道 pdf

# References

## 论文

<http://cs.stanford.edu/people/chrismre/papers/deepdive_vlds.pdf>

<http://www.kdd.org/kdd2016/papers/files/Paper_801.pdf>

## 书籍

Spark大数据分析实战

机器学习实战

Spark快速大数据分析

Spark机器学习

Python金融大数据分析

## 网站

人工智能网 http://www.ailab.cn/

数据科学日报 http://www.lovedatasci.com/

http://tensorflow.org/

https://www.datarobot.com/

https://www.kaggle.com/

https://www.alteryx.com/

浙江大学可视分析小组博客

http://www.cad.zju.edu.cn/home/vagblog/?p=3010

斯坦福

http://deepdive.stanford.edu/index.html

## 微信

DeepTech深科技

资料查询：

1. https://www.quora.com

https://www.quora.com/What-is-secondary-sort-in-Hadoop-and-how-does-it-work

组织：

http://tech.meituan.com ：美团点评技术团队

全面：

http://www.lovedatasci.com/：数据科学家日报

https://segmentfault.com

机器学习：

http://www.52nlp.cn/ 我爱自然语言处理

书籍：

http://www.allitebooks.com/

 麻省理工科技评论：

https://www.zan-shang.com/projects/56c40c3fe4b0aaa60e31cbd3#rd

论文：

http://www.vldb.org/

http://citeseerx.ist.psu.edu/

## NLP相关

ACL  ：  http://www.aclweb.org/

ACL Wiki  ： http://aclweb.org/aclwiki/index.php?title=Main\_Page

EuroMatrix ： http://www.euromatrix.net/

JHU Workshop ： http://www.clsp.jhu.edu/workshops/

LDC   http://www.ldc.upenn.edu/

Moses   http://www.statmt.org/moses/

nlper   http://nlpers.blogspot.com/

NLPJob   http://www.nlpjob.com/

SRILM   http://www.speech.sri.com/projects/srilm/

Statistical Machine Translation   http://www.statmt.org/

Text Analysis   http://textanalysisonline.com/

Text Mining     http://textminingonline.com/

Text Processing   http://textprocessing.org/

Text Summarization   http://textsummarization.net/

机器翻译档案计划   http://www.mt-archive.info/

欧洲议会平行语料库   http://www.statmt.org/europarl/

自然语言处理与信息检索共享平台   http://www.nlpir.org/

## 公开课

http://blog.coursegraph.com/

http://52opencourse.com/

<https://classroom.udacity.com>

一、 卷积神经网络教程

最好的卷积神经网络入门教程之一是Udacity的Deep Learning，来自Google的讲师，课件在tensorflow/tensorflow项目中。

视频课程地址 https://classroom.udacity.com/courses/ud730/lessons/6370362152/concepts/63798118150923

实验课件地址 https://github.com/tensorflow/tensorflow/blob/master/tensorflow/examples/udacity/1\_notmnist.ipynb

可以本地跑docker也可以用nbviewer直接看iPython Notebook:

http://nbviewer.jupyter.org/github/tensorflow/tensorflow/blob/master/tensorflow/examples/udacity/1\_notmnist.ipynb

二、bigdatauniversity 的Spark

https://www.bigdatauniversity.com.cn/courses/BigDataUniversity/BD097EN/2015\_11/about

## 世界名牌大学

http://weibo.com/p/1001603984127807359083

美国的开放课件运动，有不少人非常感兴趣，以下为开放网络课程的大学及网站地址：

一、伯克利

加州大学伯克利分校

http://webcast.berkeley.edu/courses.php

作为美国第一的公立大学，伯克利分校提供了许多优秀教授的播客和视频讲座，可以跟踪最新的讲座。想看教授布置的作业和课堂笔记，可以点击该教授的网页，通常，他/她都会第一堂课留下网址。实在不行，用google搜搜吧！

伯克利的视频都是.rm格式，请注意转换。

二、麻省

麻省理工学院

http://ocw.mit.edu/OcwWeb/web/courses/courses/index.htm

麻省理工是免费开放教育课件的先驱，计划在今年把1800门课程的课件都放在网站上，提供课程与作业的ＰＤＦ格式下载。

麻省理工只提供少数的视频讲座。坐过学生上麻省有一个绝对优势，麻省理工在中国大陆和中国台湾都建立了镜像网站，把麻省的课程都翻译成立中文。鉴于PDF格式，推荐使用FoxIt　Reader。

www.core.org.cn（中国大陆）推荐

www.myoops.org（中国台湾）

三、卡耐基梅隆

http://www.cmu.edu/oli/

卡耐基梅隆针对初入大学的大学生，提供１０门学科的课程视频。与其他大学的免费课程一样，非卡耐基梅隆的学子能学习课程，但是为了使学生能够及时了解自己的课程进度，卡耐基梅隆建议造访者在网站上注册，建立自己的资料库。这样一来，你得在有限的时间内完成一门课程，还要参加几次考试，当然，即使你得了100分，卡耐基梅隆也不会给你开证明，更不会给你学分。

四、犹他

犹他大学

http://ocw.usu.edu/front-page/Courese\_listing

犹他大学类似于麻省理工，提供大量的课程课件

五、塔夫茨

塔夫茨大学

http://ocw.tufts.edu

塔夫茨大学也是“开放式教育课程”的先驱之一，初期提供的课程着重在本校专长的生命科学、跨领域方法、国际观点以及对美国地区性、全国性社群服务的基础理论。

六、公开

英国公开大学

http://openlearn.open.ac.uk/course/index.php

英国十几所大学联合起来，组建了英国公开大学。有一部分课程是对注册学生开放的，但是有一批很好的课程是免费的，并提供视频。每门课还设立了论坛，在社区中，大家发表意见，提供其他的学习资源，互相取经。在这个网站里，最能锻炼自学者的能力，因为你要不停地淘，才能找到宝贝。

七、约翰霍普金斯

约翰霍普金斯

http://ocw.jhsph.edu/topics.cfm

只有极少数人能够进入约翰霍普金斯大学就读，但是，现如今有动机的人不用花一分钱，便能通过网站获得该校的前沿知识。约翰霍普金斯提供了本学院最受欢迎的课程，包括青少年健康、行为和健康、生物统计学、环境、一般公共卫生、卫生政策、预防伤害、母亲和儿童健康、心理卫生、营养、人口科学、公共卫生准备和难民卫生等。

八、Connexions

http://cnx.rice.edu

CNX.org由莱斯大学开发，号称是课程资源免费共享图书馆。与其他大学不同的事，CNX邀请教授学者建立自己的社区，把自己的最新成果公布于世，接受大家的评价。可以说CNX开辟了大学资源共享的新天地，尤其适合自学能力超强的大学生。有些课程有中文版。

九、索菲亚

索菲亚大学

http://sofia.ocw.cn/gallery

无论是想当一名管理者、作家、评论员、还是要从事设计和IT业的人，索菲亚大学的免费课程肯定让你受益匪浅。索菲亚大学提供了8门学科的课程，其中《企业网络安全实战》已翻译成中文。

十、华盛顿

华盛顿大学

http://www.cs.washington.edu/education/course-webs.html

华盛顿大学的计算机工程学比较强，相关的几百门课程都已经放到网上。不但本科生能找到所需要的课程，连研究生也能淘到宝贝。该网站还提供特色讲座，比如：妇女、计算机与合作。课程不但提供讲座介绍、课堂笔记、有些课程还提供视频。

牛津、斯坦福、耶鲁大学联合网站

http://www.alllearn.org

哥伦比亚大学

http://ci.columbia.edu/ci

伯克利音乐学院

http://www.berkleeshares.com

杜克大学法律中心

http://www.law.duke.edu/cspd/lectures

圣母大学

http://ocw.nd.edu

英国格雷莎姆学院

http://www.gresham.ac.uk/default.asp

加州大学Irvine分校

http://ocw.uci.edu

富布莱特学校

http://ocw.fetp.edu.vn/home.cfm

日本东京大学

http://ocw.u-tokyo.ac.jp/english

日本早稻田大学

http://www.waseda.jp/ocw/index.html

日本大阪大学

http://ocw.osaka-u.ac.jp/index.php

法国巴黎高科

http://graduateschool.paristech.org

再给个看大学视频播客的

斯坦福大学

http://itunes.stanford.edu

加州大学伯克利分校

http://itunes.berkeley.edu

普渡大学

http://boilercast.itap.purdue.edu:1013/Boilercast

美国西南理工

http://pocast.swtc.edu/lecture/index.php

加州大学洛杉矶分校

http://www.bruincast.ucla.edu

西肯塔基大学

http://blog.wku.edu/podcasts

Dupage学院

http://www.cod.edu/multimedia/podcast/CODcast/Welcome.html

纽约城市大学

http://podcast.york.cuny.edu/lectures

莱斯大学

http://webcast.rice.edu

加州大学圣地亚哥分校

http://podcast.ucsd.edu

剑桥大学

http://mediaplayer.group.cam.ac.uk/main/Podcasts.html

美国大学华盛顿法律学院

http://www.wcl.american.edu/podcasts

杜克大学法律学院

http://www.law.duke.edu/webcast

乔治敦大学

http://webcast.georgetown.edu

芝加哥大学商学院

http://www.chicagogsb.edu/multimedia/podcast

波士顿学院

http://frontrow.bc.edu

哈佛商学院在线

http://www.hbsp.harbard.edu/b02/en/hbr\_ideacast.jhtml;jsessioned=NVHF0YFBS5ZCGAKRGWDR5VQBKE0YIISW

威斯康星麦迪逊大学

http://havenscenter.org/audio/audio.htm

约翰霍普金斯大学

http://www.johnshopkins.edu/podcasts.index1.html

伦敦政治经济学院

http://www.lse.ac.uk/rescources/podcasts/Default.htm

普林斯顿大学

http://uc.princeton.edu/main/index.php

英国泰晤士报MBA

http://uc.princeton.edu/main/index.php

耶鲁大学

http://www.yale.edu/opa/podcast/

\*本文整理自网络，版权归原作者所有，如对版权有异议，请后台联系，议定合作或删除。

## 常用网址

https://www.quora.com/What-is-secondary-sort-in-Hadoop-and-how-does-it-work

## Apache

1. 发布目录：http://www.eu.apache.org/dist/

## 人

### Brendangregg

<http://www.brendangregg.com/index.html>