

海量数据处理中的云计算

C3. Hadoop体系概述

北京邮电大学信息与通信工程学院

2013年春季学期

上节问题

- 问题：
 - 对Google的三大核心技术任意一个的最重要技术特点进行研究
- 要求：
 - 详细讲解该特点，并说明为什么是最重要的

本节目录

- Hadoop发展历程
- 基于Hadoop的海量数据处理架构
- Hadoop技术组件介绍
- Hadoop实战

- 课程设计主题讨论
 - 问题定义与描述
 - 为什么需要使用Hadoop ?
 - 计划使用的Hadoop具体技术
 - 方案构想与预期效果

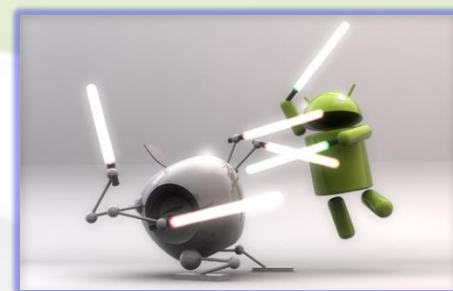
Hadoop的由来

- 当老二干不过老大时，怎么办？
- 免费+开源



IBM

vs.



vs.



Hadoop的由来

- Doug **Cutting** + Google论文 + Yahoo !

Doug Cutting

Nutch, Lucene创始人

Hadoop创始人



Google

挣扎在如何使Nutch满足Web-Scale要求 — 2002

2003 — 发表GFS论文

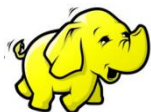
参考GFS论文，在Nutch中实现NDFS — 2004 — 2004 — 发表MapReduce论文

参考MapReduce论文，实现并移植Nutch — 2005



加入Yahoo!并扩充，建立Hadoop项目

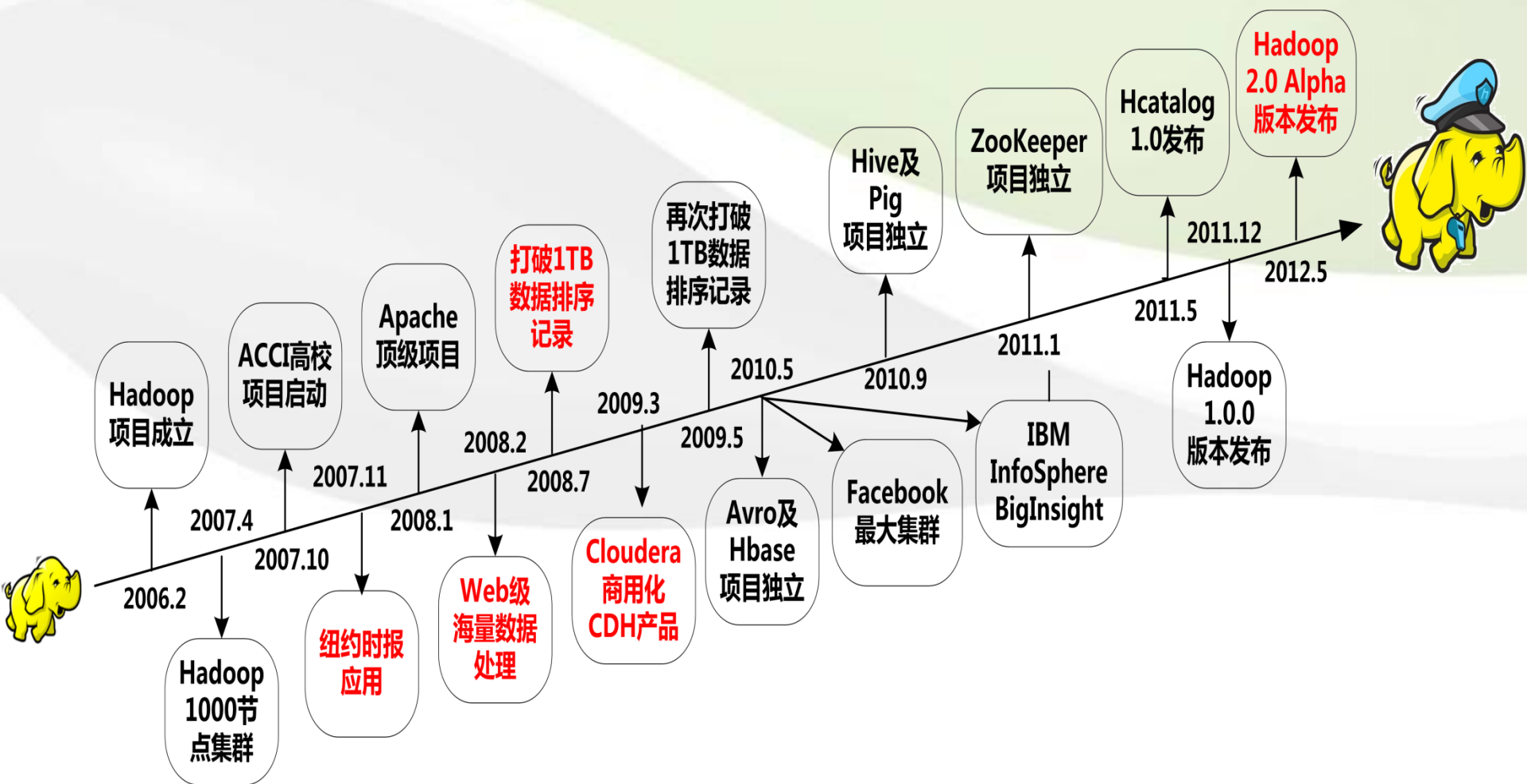
2006 2006 — 发表BigTable、Chubby论文



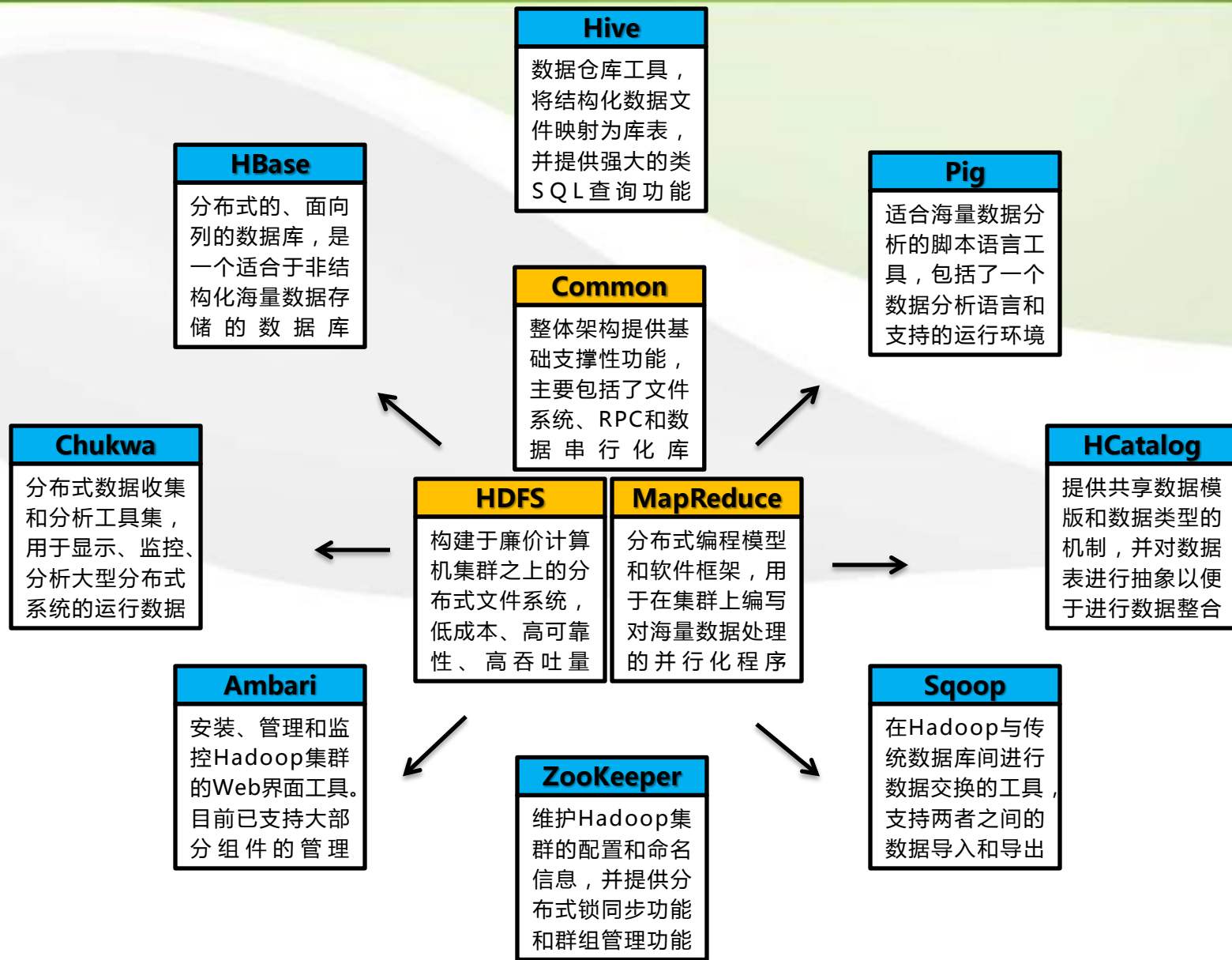
Hadoop成为Apache顶级项目 — 2008



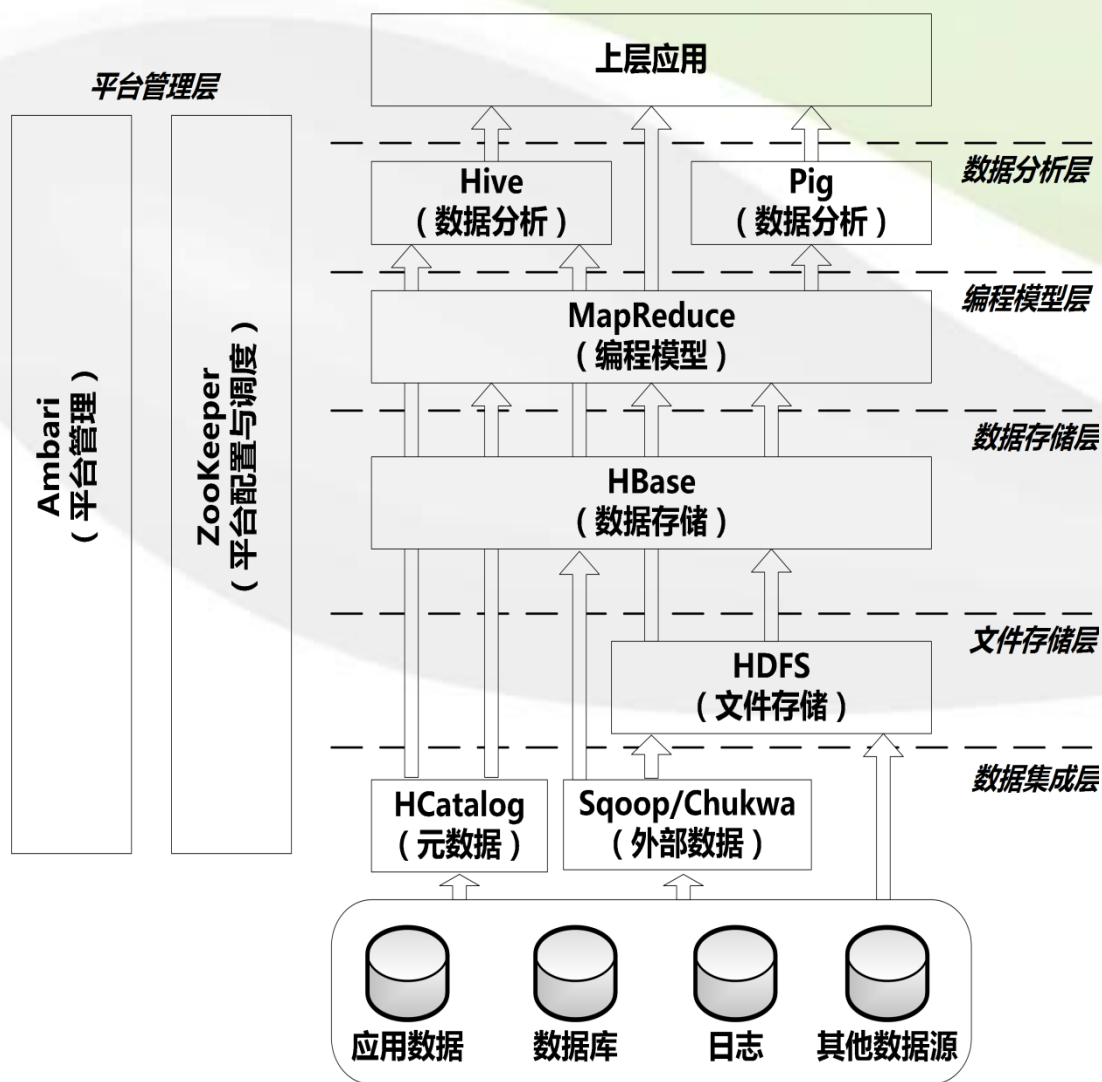
Hadoop发展历程



Hadoop技术族

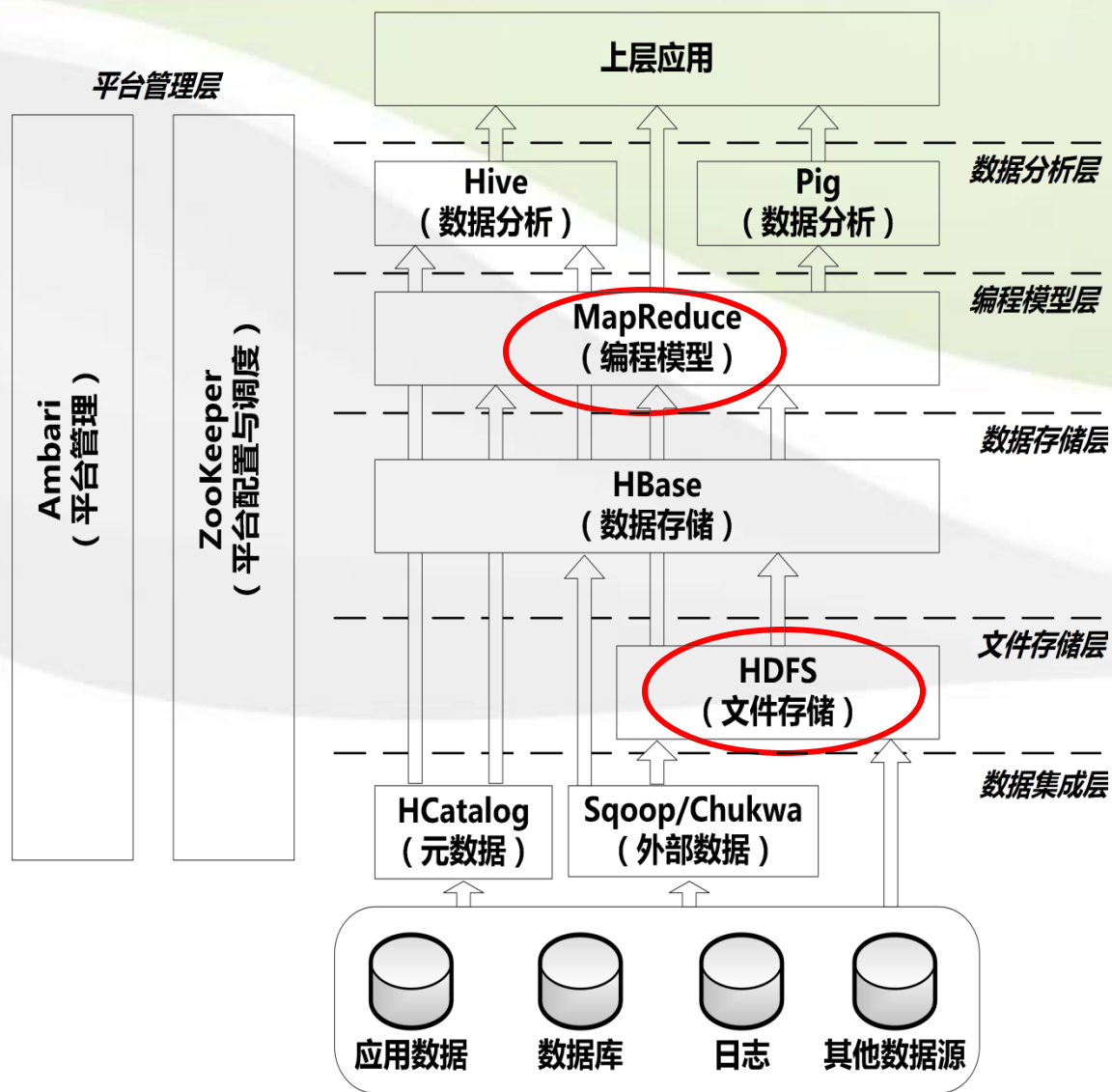


基于Hadoop的海量数据处理框架

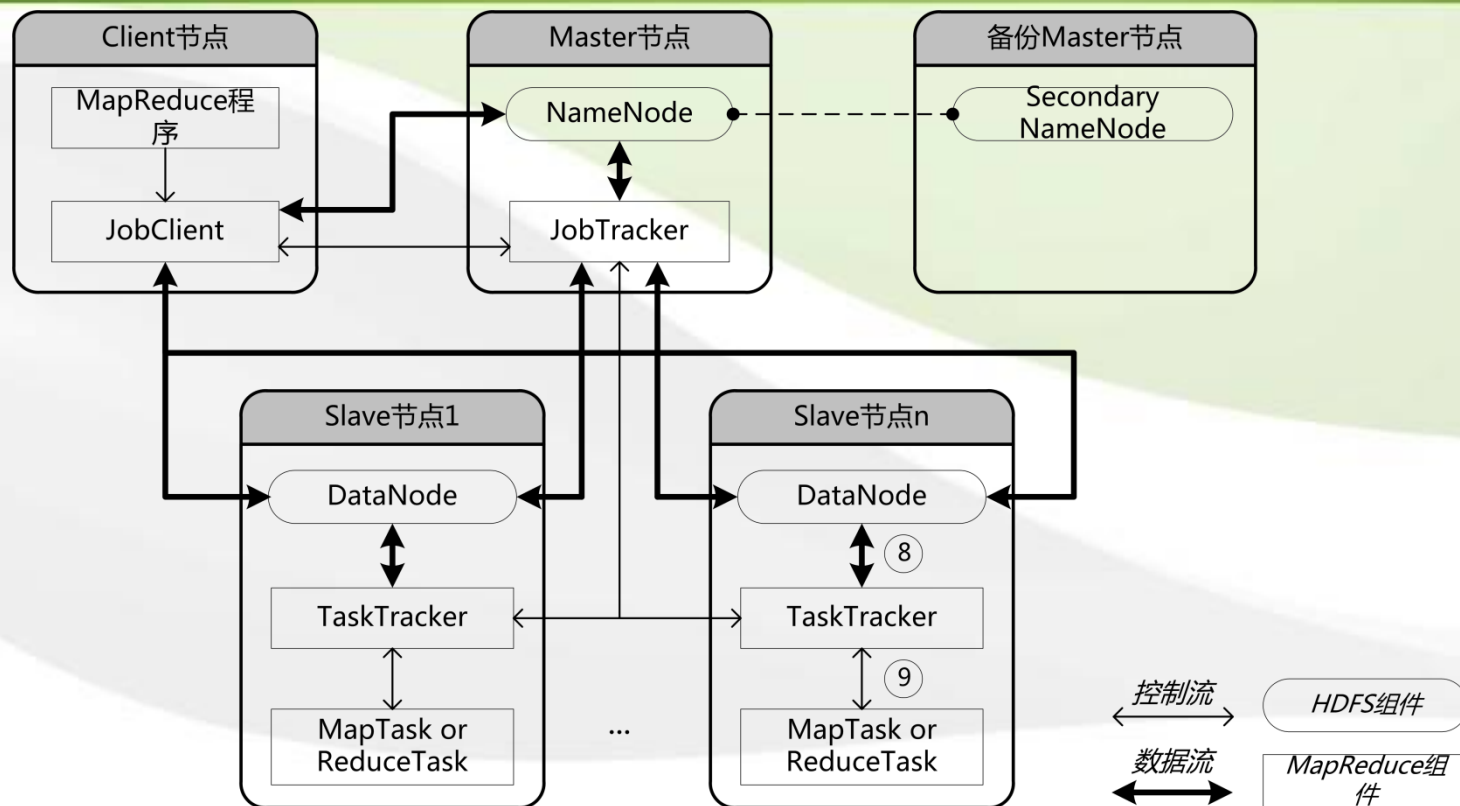


- **平台管理层**：确保整个数据处理平台平稳安全运行的保障，包括配置管理、运行监控、故障管理、性能优化、安全管理等在内的功能。
- **数据分析层**：提供一些高级的分析工具给数据分析人员，以提高他们的生产效率。
- **编程模型层**：为大规模数据处理提供一个抽象的并行计算编程模型，以及为此模型提供可实施的编程环境和运行环境。
- **数据存储层**：提供分布式、可扩展的大量数据表的存储和管理能力，强调的是在较低成本的条件实现海量数据表的管理能力。
- **文件存储层**：利用分布式文件系统技术，将底层数量众多且分布在不同位置的通过网络连接的各种存储设备组织在一起，通过统一的接口向上层应用提供对象级文件访问服务能力。
- **数据集成层**：系统需要处理的数据来源，包括私有的应用数据、存放在数据库中的数据、被分析系统运行产生的日志数据等，这些数据具有结构多样、类型多变的特点。

Hadoop核心 – HDFS与MapReduce



Hadoop核心 – 基础架构



• NameNode (NN)

- 管理者，命名空间，维护文件树及文件和目录的元数据
- 命名空间镜像文件、编辑日志文件、数据块与DN的对应

• Secondary NN

- NameNode备份
- 与NN定时通信
- 保存元数据快照
- 故障切换

• DataNode (DN)

- 与NameNode定时通信
- 保存元数据快照
- 故障切换

• JobClient

- 基于接口库编写的客户端程序
- 提交MR作业

• JobTracker

- 控制协调者
- 协调作业的执行
- 仅有一个

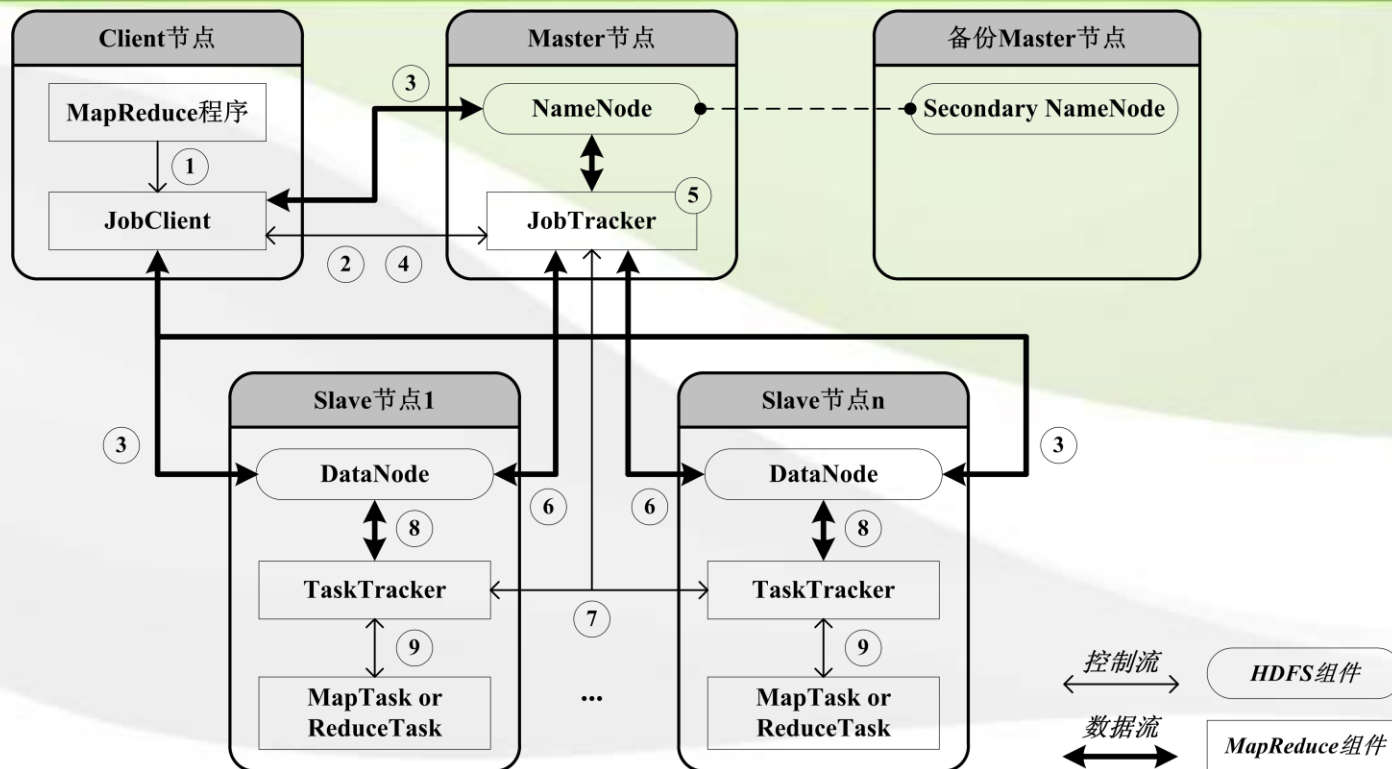
• TaskTracker

- 任务控制
- 一个或多个Map或Reduce任务
- 维持心跳

• MapTask , ReduceTask

- 具体执行Map和Reduce任务的程序

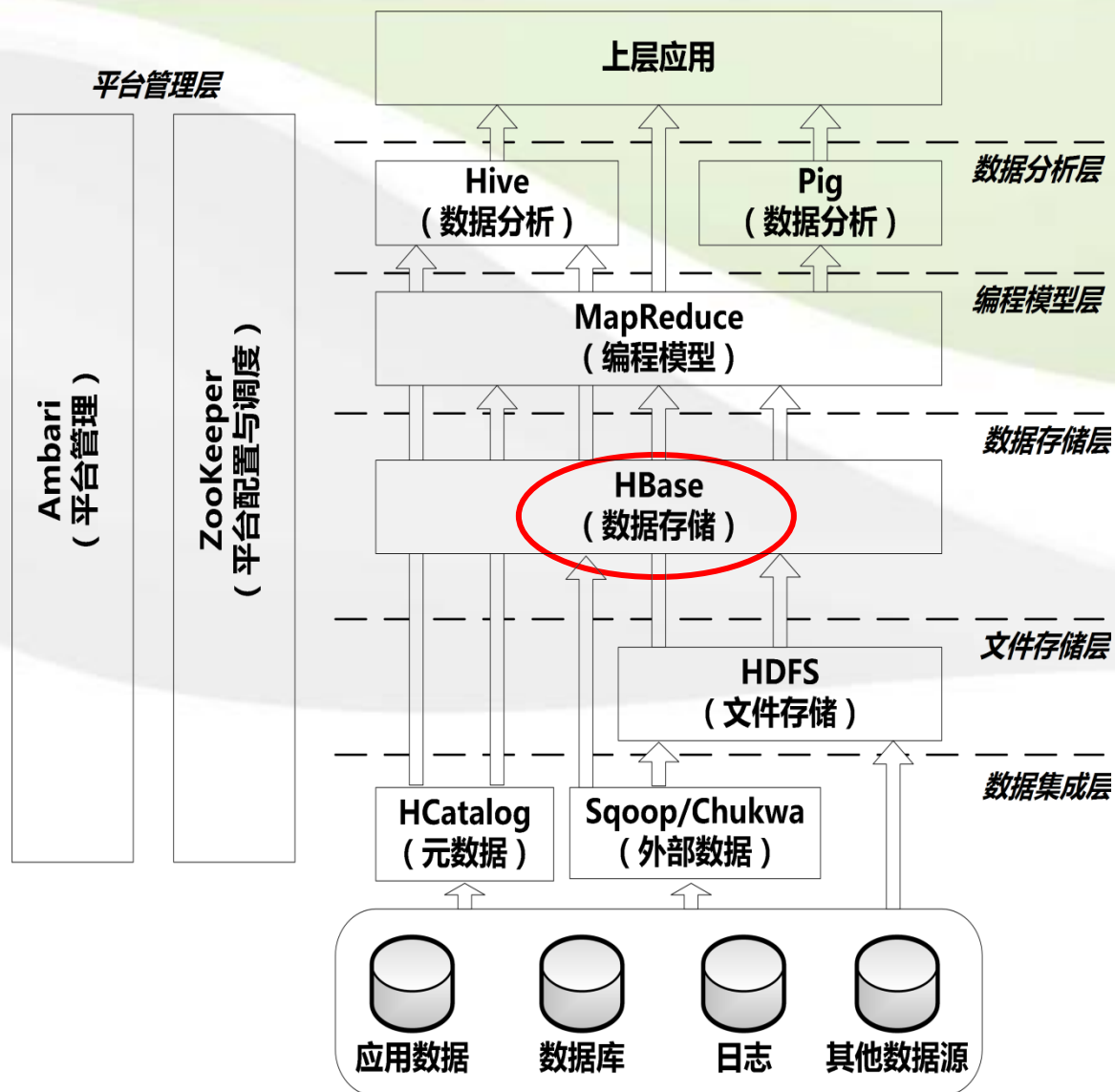
Hadoop核心 – 运行机制



- ① MapReduce程序启动作业
- ② JobClient从JobTracker获得作业ID
- ③ JobClient切分数据文件并存入HDFS (多份)
- ④ JobClient向JobTracker提交此作业
- ⑤ JobTracker将作业放入队列中等待调度
- ⑥ JobTracker从HDFS中取出输入数据，并根据输入数据创建对应数量的Map任务和Reduce任务

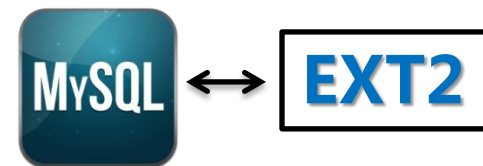
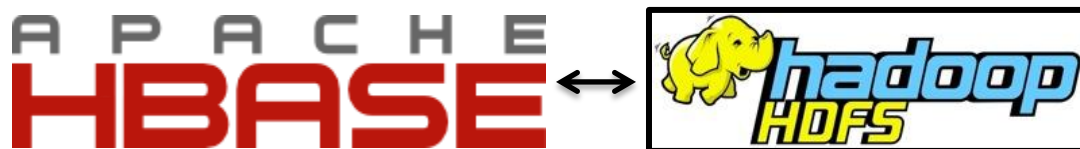
- ⑦ JobTracker将Map任务和Reduce任务分配到空闲的TaskTracker节点
- ⑧ TaskTracker从HDFS取出数据存入本地磁盘，并启动一个TaskRunner准备运行任务
- ⑨ TaskRunner在新Java虚拟机中创建出MapTask或ReduceTask进行运算，MapTask和ReduceTask会定时向TaskRunner报告进度，直到任务完成

Hadoop数据存储 – HBase



Hadoop数据存储 – Why HBase ?

- 来源：Fay Chang, **Jeffrey Dean**, et. al., **Bigtable**: A Distributed **Storage System** for Structured Data, Seventh Symposium on Operating System Design and Implementation, 2006.
- Why HBase ? (<http://hbase.apache.org/book/architecture.html#arch.overview>)
 - **HDFS** is a **distributed file system** that is well suited for the storage of large files. It's documentation states that it is not, however, a general purpose file system, and does **not provide fast individual record lookups** in files.
 - **HBase**, on the other hand, is built on top of HDFS and provides **fast record lookups** (and updates) for **large tables**.
 - This can sometimes be a point of conceptual confusion. HBase internally puts your data in indexed "**StoreFiles**" that exist on HDFS for high-speed lookups.



Hadoop数据存储 – 感受HBase

文件

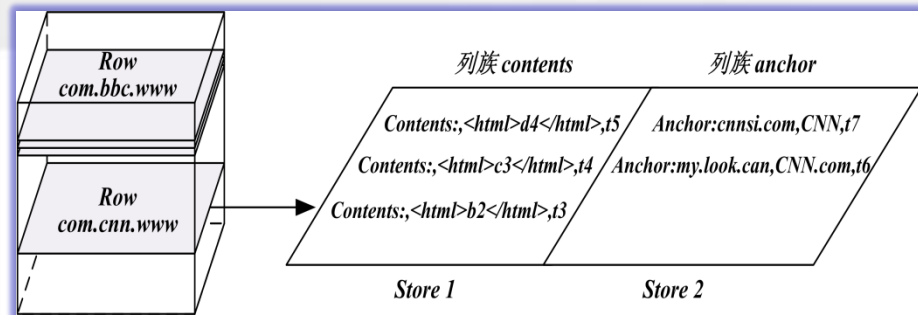
1UTC时间	2记录时间	3流开始时间	4流结束时间	6IMEI	7IMSI	8NSAPI	9APN
1328637497.116240	2012-2-8 1:58	1328637496.734460	1328637497.116240	3562810462199400	460008033221470	6	cmnet
1328637479.945410	2012-2-8 1:57	1328637475.599330	1328637479.945410	3580590123929200	460004003238107	5	cmwap
1328637497.460850	2012-2-8 1:58	1328637494.477920	1328637497.460850	3576490418359000	460028750209074	7	cmnet
1328637497.678810	2012-2-8 1:58	1328637496.679290	1328637497.678810	123050000265435	460006151753402	5	cmwap
1328637497.397070	2012-2-8 1:58	1328637494.994230	1328637497.397070	3520060476705710	460001881324031	5	cmnet
1328637478.219940	2012-2-8 1:57	1328637477.959980	1328637478.219940	3521890410124690	460020331605452	5	cmwap
1328637479.786740	2012-2-8 1:57	1328637478.454070	1328637479.786740	3519880490368710	460002407220818	5	cmwap
1328637497.408050	2012-2-8 1:58	1328637494.372890	1328637497.408050	3531840362073860	460028157806762	5	cmnet
1328637497.144830	2012-2-8 1:58	1328637496.831110	1328637497.144830	3564090467725200	460009617903834	5	cmnet

表

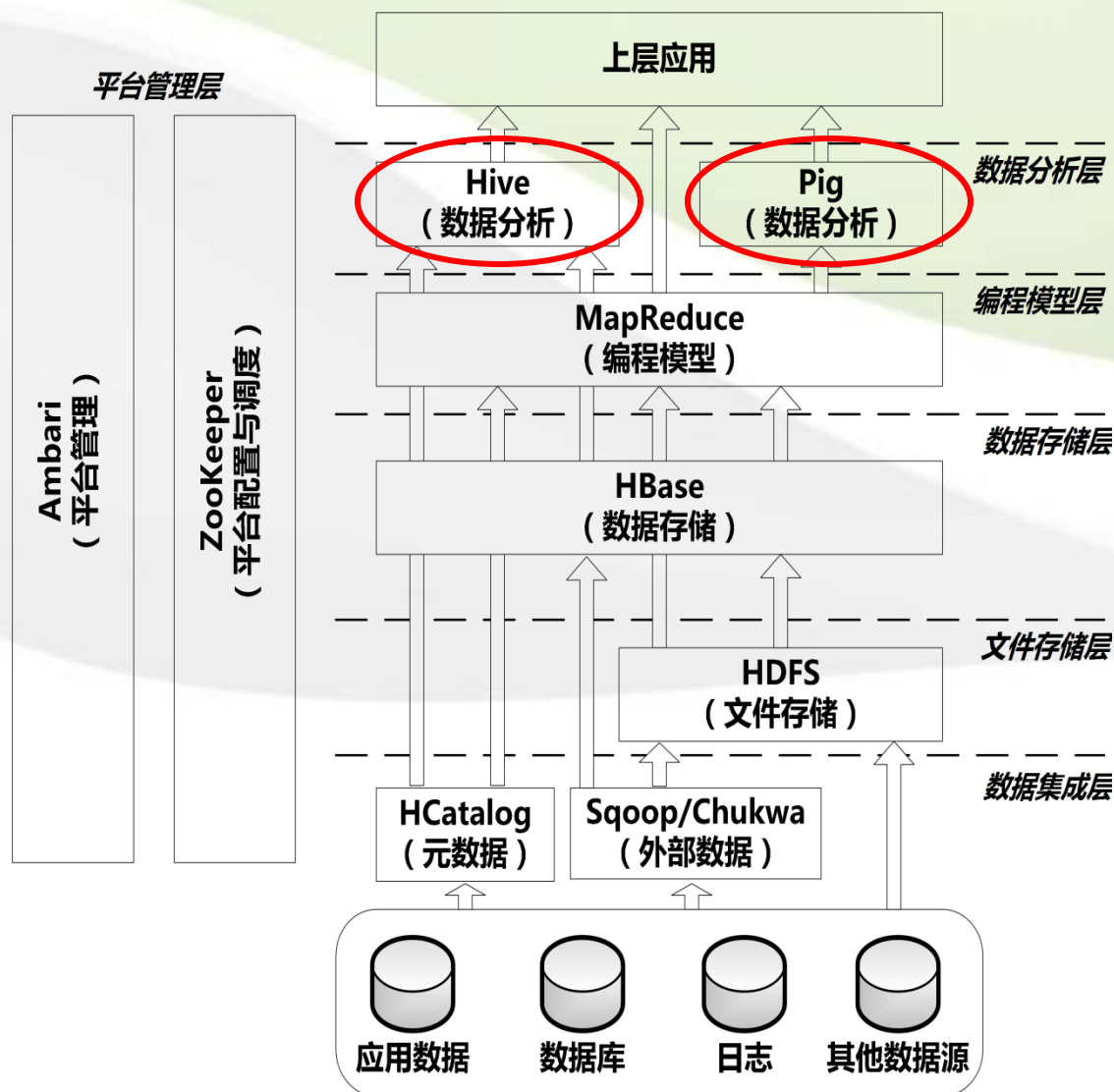
表

RowKey	contents:	anchor:
com.bbc.www	<html>a1</html> [t ₁]	anchor:com.bbc.www = "BBC" [t ₂]
com.cnn.www	<html>b2</html> [t ₃] <html>c3</html> [t ₄] <html>d4</html> [t ₅]	anchor:cnnsi.com = "CNN" [t ₆] anchor:my.look.ca = "CNN.com" [t ₇]

行关键字 版本 (时间戳) 列关键字 限定词 值 数据元



Hadoop数据分析 – Hive和Pig



Hadoop数据分析 – Hive

- 来源：Ashish Thusoo, Joydeep Sen Sarma, et al., **Facebook**, “Hive: A **Warehousing** Solution over A Map-Reduce Framework” , Proceedings of the VLDB Endowment, Aug. 2009.
- Why Hive ? (<http://hive.apache.org/>)
 - Hive is a **data warehouse** system for Hadoop that facilitates easy data summarization, **ad-hoc queries**, and the analysis of large datasets stored in Hadoop compatible file systems. Hive provides a mechanism to **project structure** onto this data and query the data using a SQL-like language called **HiveQL**.
 - 数据库 vs. 数据仓库：存取（面向事务） vs. 分析（面向主题）



Hadoop数据分析 – 感受Hive

代码

```
#include "mapreduce/mapreduce.h"
class WordCounter : public Mapper {
public:
    virtual void Map(const MapInput& input) {
        const string& text = input.value();
        const int n = text.size();
        for (int i = 0; i < n; ) {
            while ((i < n) && isspace(text[i])) i++;
            int start = i;
            while ((i < n) && !isspace(text[i])) i++;
            if (start < i)
                Emit(text.substr(start, i-start), "1");
        }
    }
    REGISTER_MAPPER(WordCounter);
};
```

```
class Adder : public Reducer {
public:
    virtual void Reduce(ReduceInput* input) {
        int64 value = 0;
        while (!input->done()) {
            value += StringToInt(input->value());
            input->NextValue();
        }
        Emit(IntToString(value));
    }
    REGISTER_REDUCER(Adder);
};
```

```
int main(int argc, char** argv) {
    ParseCommandLineFlags(argc, argv);
    MapReduceSpecification spec;
    for (int i = 1; i < argc; i++) {
        MapReduceInput* input = spec.add_input(i);
        input->set_format("text");
        input->set_filepattern(argv[i]);
        input->set_mapper_class("WordCounter");
    }
    MapReduceOutput* out = spec.output();
    out->set_filebase("/gfs/test/freq");
    out->set_num_tasks(100);
    out->set_format("text");
    out->set_reducer_class("Adder");
    spec.set_machines(2000);
    spec.set_map_megabytes(100);
    spec.set_reduce_megabytes(100);
    MapReduceResult result;
    if (!MapReduce(spec, &result)) abort();
    return 0;
}
```

SQL

```
SELECT * FROM log WHERE date > '2012-12-01'
DISTRIBUTE BY date SORT BY date ASC, host ASC;
```

Hadoop数据分析 – Pig

- 来源：Christopher Olston, Benjamin Reed, Utkarsh Srivastava, et al., **Yahoo!**, "Pig Latin: A Not-so-foreign **Language for Data Processing**", ACM SIGMOD, 2008.
- Why Pig ? (<http://pig.apache.org/>)
 - Apache Pig is a platform for analyzing large data sets that consists of **a high-level language** for expressing **data analysis programs**, coupled with infrastructure for evaluating these programs.
 - Pig's language layer currently consists of a textual language called **Pig Latin**.
 - Pig vs. Hive : 数据准备 vs. 数据呈现



Hadoop数据分析 – 感受Pig及对比

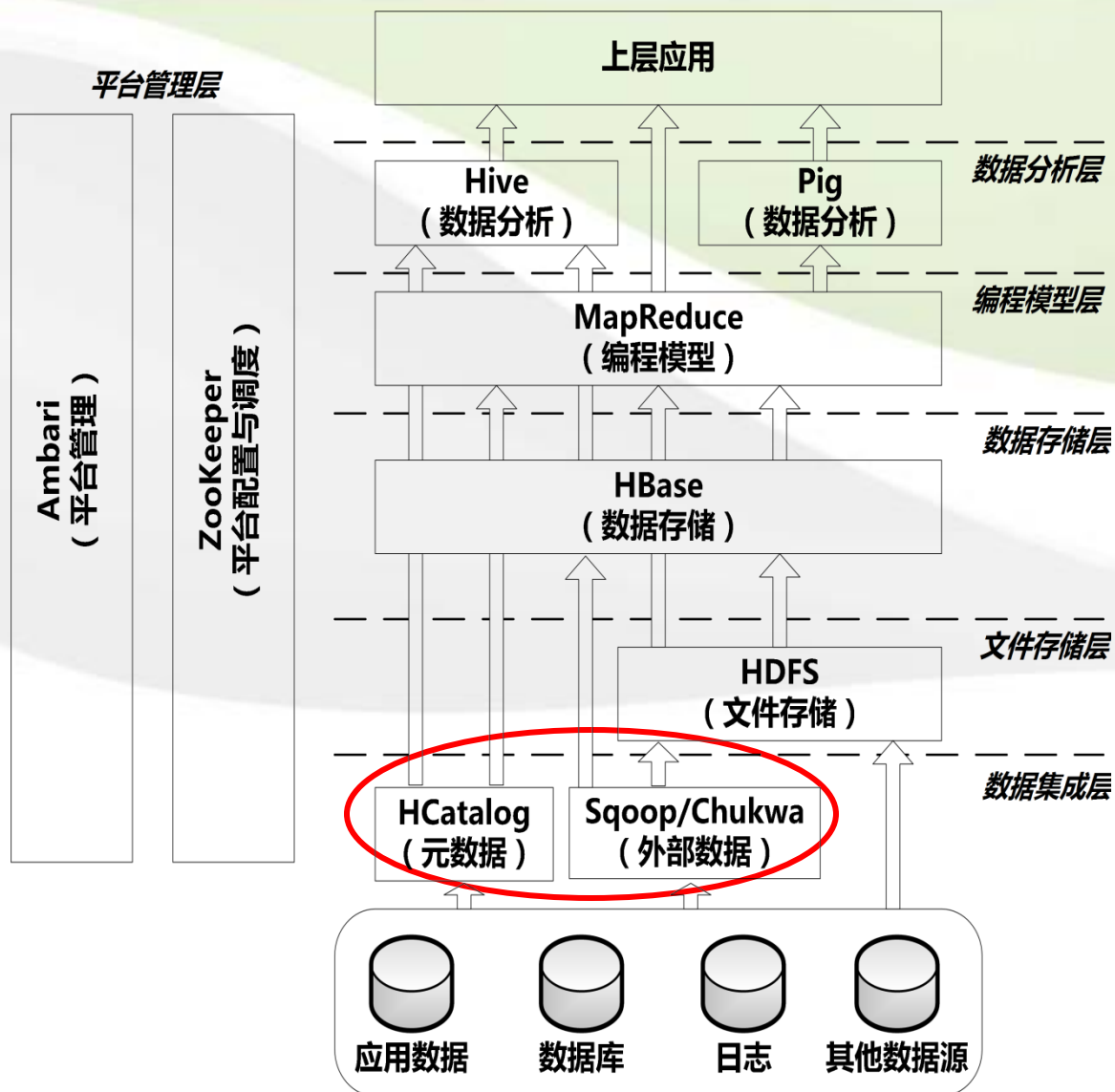
Hive

```
SELECT category, AVG(pagerank) FROM urls  
WHERE pagerank > 0.2  
GROUP BY category HAVING COUNT(*) > 106
```

Pig

```
> good_urls = FILTER urls BY pagerank > 0.2;  
> groups = GROUP good_urls BY category;  
> big_groups = FILTER groups BY COUNT(good_urls) > 106;  
> output = FOREACH big_groups GENERATE  
category, AVG(good_urls.pagerank);
```

Hadoop数据集成



Hadoop数据集成 – HCatalog

- 来源：Hortonworks，2010年3月提出，2011年5月发布。



- Why HCatalog ? (<http://incubator.apache.org/hcatalog/>)
 - Apache HCatalog is a table and storage **management service** for data created using Apache Hadoop.
 - Providing a **shared schema** and data type mechanism.
 - Providing a **table abstraction** so that users need not be concerned with where or how their data is stored.
 - Providing **interoperability** across data processing tools such as **Pig, Map Reduce, and Hive**.



Hadoop数据集成 – Sqoop

- 来源：Aaron Kimball为Hadoop开发的一个小模块（2009年），并由Cloudera支持成为Apache开源项目。

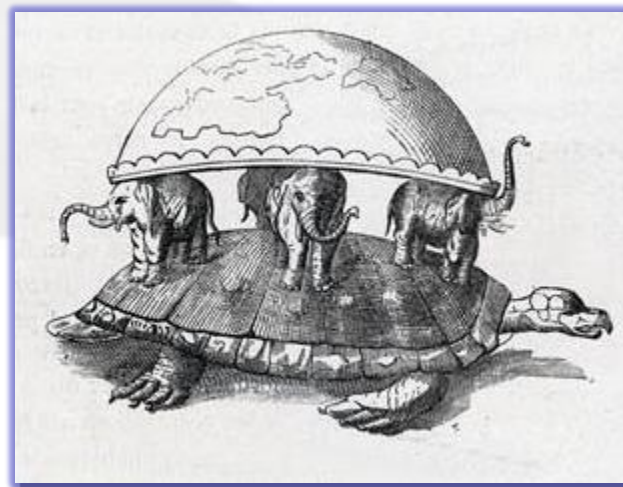


- Why Sqoop ? (<http://sqoop.apache.org/>)
 - Apache Sqoop is a tool designed for efficiently **transferring bulk data** between Apache Hadoop and structured datastores such as relational databases.



Hadoop数据集成 – Chukwa

- 来源：由Yahoo! 2008年提供的模块扩展出的Apache项目。
- Why Chukwa ? (<http://sqoop.apache.org/>)
 - Chukwa is an open source **data collection system** for monitoring large distributed systems.



Hadoop在产业界



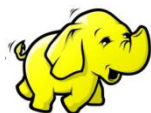
Hadoop在学术界



Hadoop实战



单机Hadoop运行与开发环境



Hadoop



Linux



Cygwin虚拟环境



Windows操作系统

eclipse

Hadoop运行与开发环境搭建

- 目标：
 - 建立Windows下的单机Hadoop环境
 - 建立Eclipse开发环境
 - 调试Hadoop代码
- 步骤：
 - 安装和配置Cygwin
 - 安装和配置Hadoop
 - 运行示例程序验证Hadoop安装
 - 安装和配置Eclipse下的Hadoop开发环境
- 参考：
 - QQ群共享：附录A（基于Cygwin的Hadoop环境搭建）.pdf

作业

- 完成：
 - 建立Windows下的单机Hadoop环境
 - 建立Hadoop的Eclipse开发环境
 - 运行WordCount程序
- 要求：
 - 所有同学在下周四下午6点前将安装过程及运行WordCount程序结果截图，整理成文档，发送到：liujun@bupt.edu.cn
 - 邮件标题：海量数据处理作业（姓名）
 - 文档名：海量数据处理作业_hadoop安装_姓名
 - 下节课现场演示Hadoop环境、开发环境和WordCount程序结果



课程设计主题讨论

- 问题定义与描述
- 为什么需要使用Hadoop ?
- 计划使用的Hadoop具体技术
- 方案构想与预期效果

