

Hadoop在海量网页检索中的应用 之我见

即刻搜索 研发部系统平台组
何鹏 hepeng@jike.com
2011-12-2

主要内容

- 即刻搜索
 - Hadoop与即刻搜索
 - 即刻Hadoop特性
 - 正在进行的工作
 - 即刻搜索Hadoop集群
-

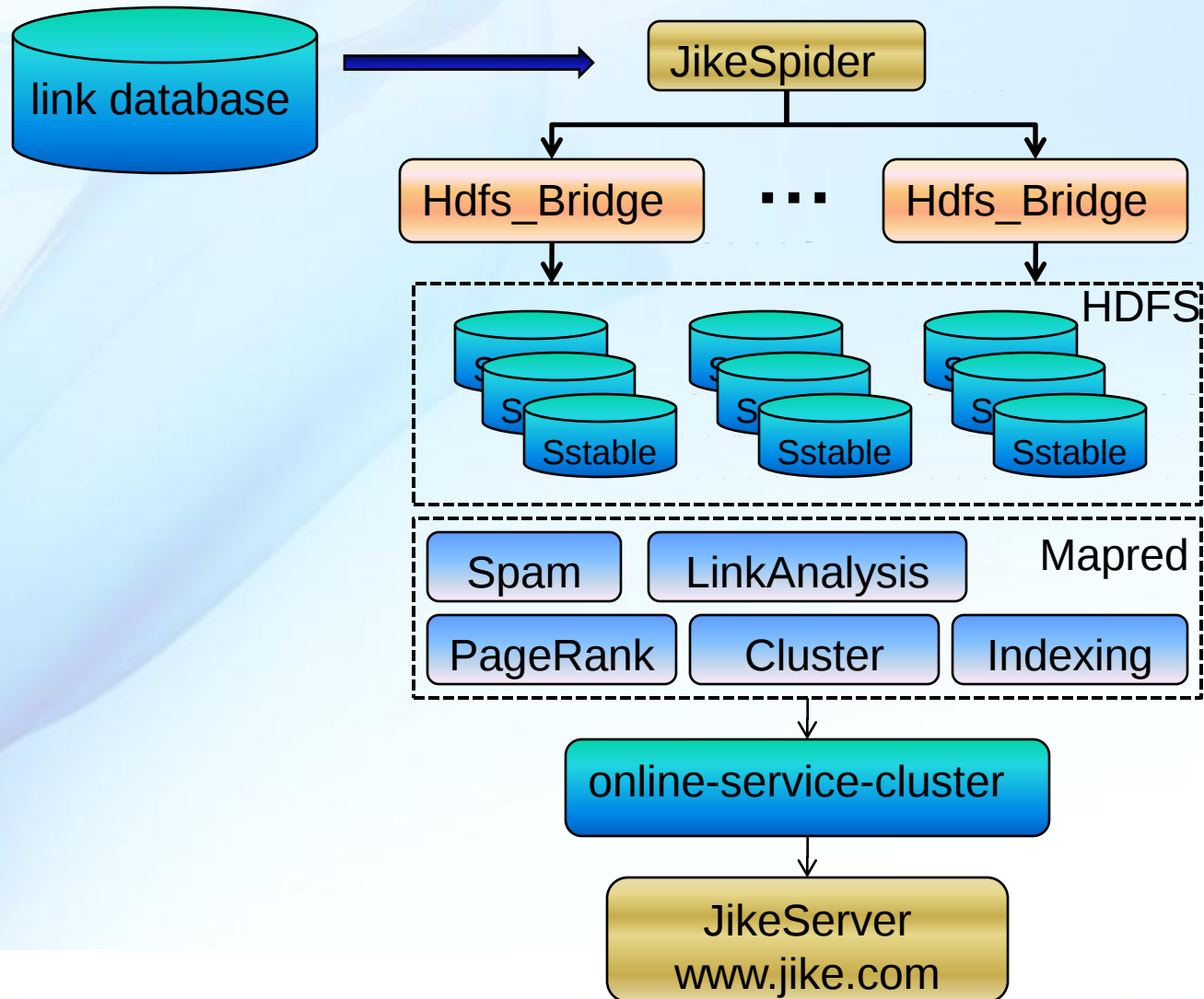
即刻搜索介绍

- 2011年6月20即刻上线



- www.goso.cn → www.jike.com
- 数据规模
 - 200+亿
- 200+亿的背后？

Hadoop与即刻搜索



已完成的工作

Hdfs & Mapred相关:

Hdfs_Bridge

Flush sstable to hdfs

支持sstable mapred

Hadoop pipes相关:

多路径输入

多路径输出

本地调试

部分优化

多表join功能

代码级别优化

sstable

- sstable介绍
 - 全称为sorted strings table;
 - 存储对象为key/value对, 并且是根据key进行过排序;
 - 支持多种压缩--gzip, lzo
 - Why sstable not sequencefile?
 - 随机查找(random seek)
 - 读写性能
 - 一致性(搜索pipe line流程)
-

What is Hdfs-Bridge?

- Hdfs-Bridge
 - 中间件
 - 写缓存，满足爬虫的快速写
 - 自动flush sstable文件
 - 原理
 - 写转化为内存写
 - 用DFS直接flush
 - Why not hadoop-thrift-server?
 - 多次不必要的序列化和反序列化
 - 速度不能满足
-

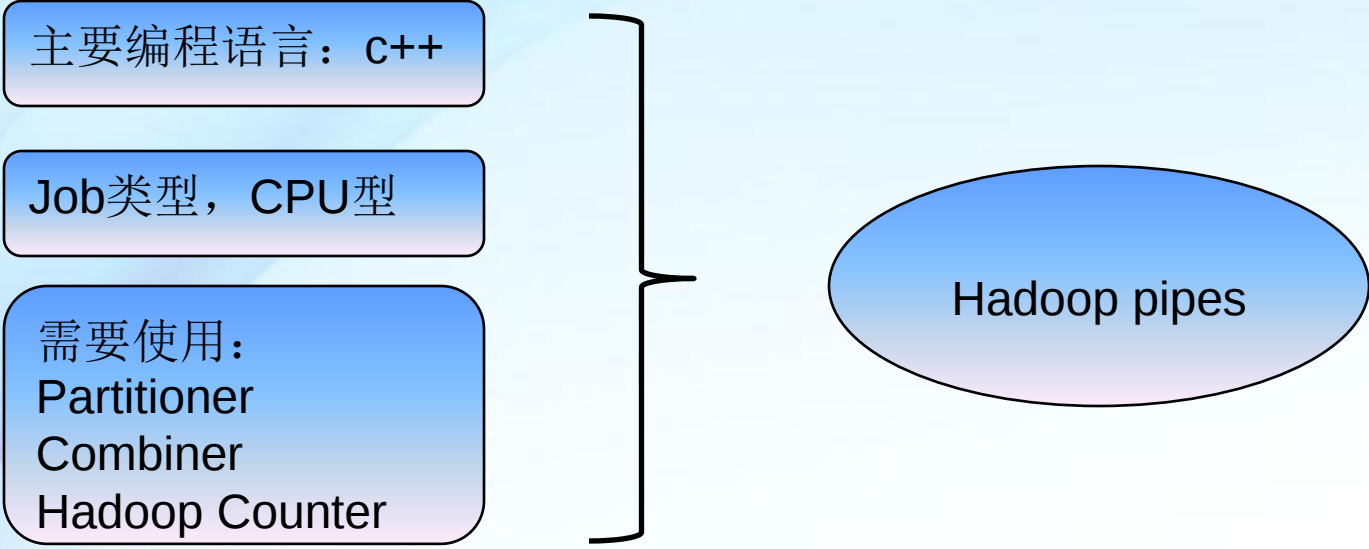
Pipes

- Why?

主要编程语言: C++

Job类型, CPU型

需要使用:
Partitioner
Combiner
Hadoop Counter



Hadoop pipes

Jike Hadoop Pipes Features

多路径输入

多路径输出

修改hadoop pipes的通信协议，对于原始的emit(key,value)，改为emit(type, key, value)。

本地调试

单个task执行的模拟

部分优化

代码级别优化，string及buffer相关

Pipes Wrapper

```
class BasicMapper {
public:
    typedef ::mapreduce::TaskContext TaskContext;
    explicit BasicMapper() {
        map_context_.reset(new MapContext());
    }
    virtual ~BasicMapper() {}
    virtual void OnCreate(TaskContext* context) {}
    virtual void OnFirstMap(MapContext* context) {}
    virtual void OnLastMap(MapContext* context) {}
    virtual void Map(MapContext* context) = 0;
    .....
protected:
    const std::string& GetInputValue() {
        return map_context_->GetInputValue();
    }
    const std::string& GetInputkey() {
        return map_context_->GetInputKey();
    }
    template <typename T>
    bool ValueToThrift(T *object) {
        return map_context_->ValueToThrift(object);
    }
    .....
    DISALLOW_COPY_AND_ASSIGN(BasicMapper);
};
```

Programmer's Perspective

```
class SampleMap: public BasicMapper {
public:
    virtual ~SampleMap() {}
    virtual void Map(mapreduce::MapContext* context) {
        string key = GetInputKey();
        string value = GetInputValue();
        Object obj_val;
        ValueToThrift(&obj_val); // deserialized
        ..... // other processing
        context->Emit(key, newvalue);
        // context->Emit(type, key, newvalue)
    }
};
```

Programmer's Perspective

```
class SampleRed: public BasicReducer {
public:
    virtual void Reduce(mapreduce::MapContext* context) {
        string key = GetInputKey();
        while (NextValue()) {
            string value = GetInputValue();
            ..... // other processing
        }
        context->Emit(key, newvalue);
        // context->Emit(outputdir, key, newvalue)
    }
};
```

```
REGIST_MAP_RED(SampleMap, SampleReduce)
REGIST_MAP_RED_PAR_COM(SampleMap,
SampleRed, SamplePar, SampleCom)
```

Usage

Submit a job

```
./bin/sample_map_red_binary \  
--auto_run=true \  
--num_mapper=100 \  
--num_reducer=50 \  
--input_format=sstable \  
--output_format=sstable \  
--input_path=/test/input/data \  
--output_path=/test/output \  
... // other parameters,like input  
output path
```

Debug

```
./bin/sample_map_red_binary \  
--local_run=true \  
--num_mapper=100 \  
--num_reducer=50 \  
--input_format=sstable \  
--output_format=sstable \  
--input_path=/test/input/data \  
--output_path=/test/output \  
... // other parameters,like input  
output path
```



Pipes_Wrapper makes
programmer's life better!

正在进行的工作

- 作业调度

- 问题:

部分大型作业，多个task分配到同一台机器，导致改机器负载过大，拖慢整个作业进度；极端状况下，内存打爆，出现OOM。

- 原因

任务调度分配不合理。

- 解决方案:

tasktracker收集CPU、内存、磁盘及网络等相关信息，汇报给jobtracker;

调度器调度任务时，将CPU、内存、磁盘、网络等信息纳入考虑之中；

正在进行的工作

- 集群能源管理
 - 机器整体：cpu，io，磁盘较长时间处于idle，可整机进入省电模式
 - 模块：单独关闭idle较长的模块
- 集群特点
 - 白天、夜晚集群利用率不同
 - 数据分属性
 - 热数据 -- 近期内使用较多的
 - 冷数据 -- 短期内较少访问和使用的数据

Jike Hadoop Cluster

- 多个集群，规模不一
- 搜索相关集群：
 - 集群规模
 - 500+台
 - 配置：
 - 2cpu 6 core + 48G内存 + 11 x 2T硬盘
 - 总容量：10PB
 - 使用率：78.15%
- 每天处理数据量
 - 读取平均约500T，峰值>1P
 - 写平均约300T，峰值>500T

即刻搜索，期待您的加入！

Thank You! Q & A ?



[将即刻设为首页](#)

[关于我们](#) [诚聘英才](#) [征求意见](#)

版权所有©2011 人民搜索 保留所有权利 [京ICP证110419](#) [京公网安备110105010684](#) [免责声明](#)