分布式流计算框架

摘要

该项目设计成为一个处理通用数据,分布式的,高扩展性的,容错的流数据计算框架.可方便应用在各种建立在流式数据处理的场合,提供持续的高效的流数据处理服务. 系统主要通过管道(processing element. PE)处理数据流.PE 在接到数据流后对数据进行处理:

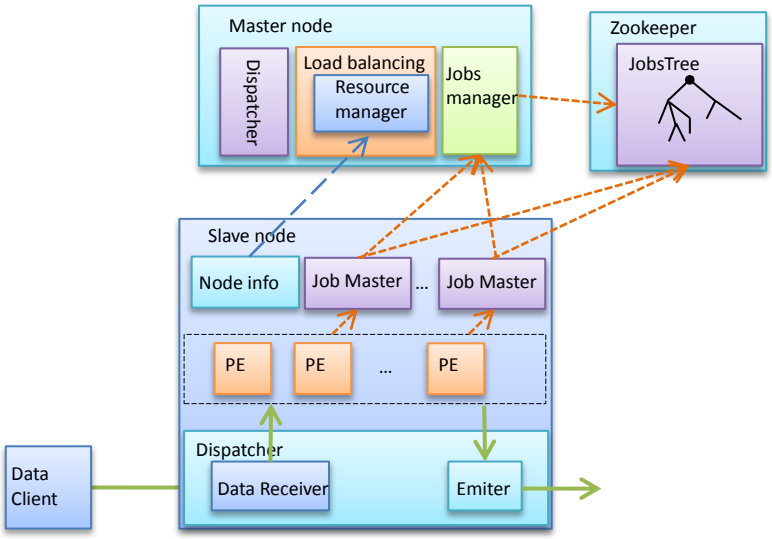
1. 生成需进一步数据的数据发送给后续 PE 模块.
2. 生成数据结果发送给相关应用程序.

该项目借鉴了基于 MapReduce 的分布式计算框架的经验.采取 primary-slave 模式,分为 Master 和 Snode (slave node)两种节点.Master 节点主要负责系统资源的调度与分配.Snode 节点主要复制具体的流数据任务的整个运行周期的管理和跟踪.

数据流处理主要通过任务 Job 的形式运行在 Snode 节点中.

1. 系统结构

1.1. Master node &Slave node&zookeeper



系统主要分为 Maseter Node,Slave Node,Client ,Zookeeper4 大部分:

Master Node:

1) Dispatcher: 负责接收客户端创建不同数据处理应用(job)的请求并指派某个 Slave

node 作为 Job-master.

2) Load balancing & resource manager: 负责接收所有 slave Node 的当前状态汇报,包括 cp,memory,job account 等.在 job-master 请求创建后续 PE 运行节点时,根据负载均衡算法分配合适的运行节点.
3) Job manager: 负责所有 Job 的基本状态维护,提供 Job 查询状态查询接口等运维相关接口.

Slave Node:

1) Node info: 负责向 master node 汇报当前节点的 cpu,memory,job account 信息
2) Job master: 负责单个 Job 的整个运行周期的维护.包括 job 中的各个 PE 的创建,PE 状态维护.
3) PE: 根据定义处理接收的数据,生成后续数据发送给后续 PE 或者提交结果.
4) Dispatcher:
   a) Data receiver: 接收客户端发送的数据,根据数据 key 发送给相应 Job 的 PE 模块处理.
   b) Emiter: 接收 PE 的发送请求,发送数据给后续 PE 或者提交结果.

Client: 创建流数据处理 Job,接收数据源发送的数据, 分发数据.

Zookeeper: 维护所有 Job 的状态信息.

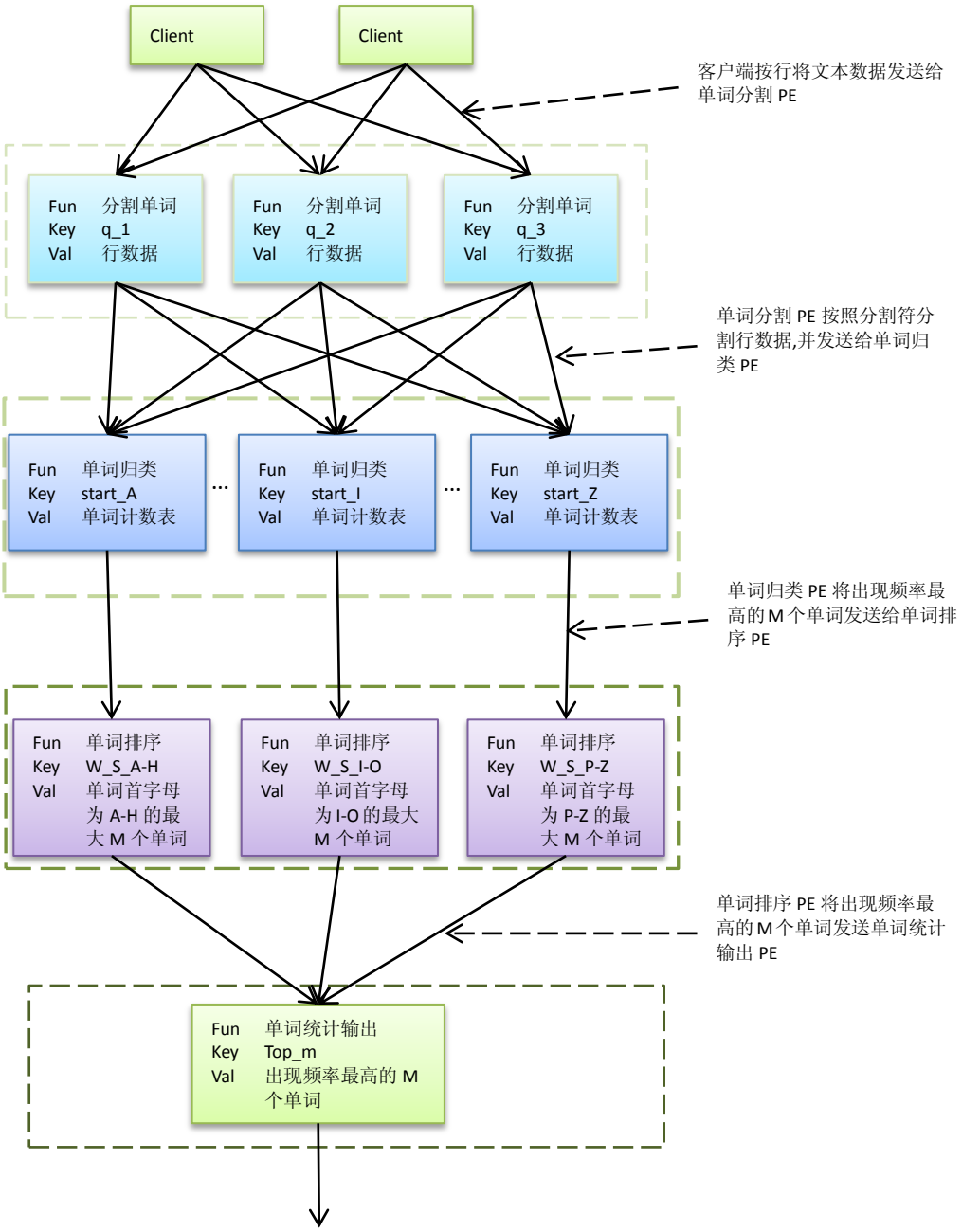
1.2. Job, Job-master,PE(Task)

流数据处理分析是通过 Job 的方式运行.单个 Job 在运行周期内是通过 Job-master 模块进行管理和跟踪的.Job 的状态包括,prepare,running,succeded,failed,killed.

1) Prepare: job-master 根据客户端请求的流数据处理任务生成 PE 链.
2) Running: PE 开始处理流数据.
3) succeded,failed,killed: Job 终结时的运行结果.

例如对某段文本中 N 个单词的出现频率最高的 M 个单词的统计任务,该任务具体如下:

1) 客户端按行依次将整段文本发送给单词分割 PE
2) 单词分割 PE 解析行数据,分割单词数据,发送给单词归类 PE
3) 单词归类 PE 根据单词的首个字母发送给单词排序 PE
4) 单词排序 PE 定时将出现次数最高的 M 个单词发送给单词统计 PE
5) 单词统计 PE 接收数据,统计单词出现次数最高的 M 个单词,并输出结果.


Client
Client
客户端按行将文本数据发送给单词分割 PE
Fun 分割单词
Key q_1
Val 行数据
Fun 分割单词
Key q_2
Val 行数据
Fun 分割单词
Key q_3
Val 行数据
单词分割 PE 按照分割符分割行数据,并发送给单词归类 PE
Fun 单词归类
Key start_A
Val 单词计数表
Fun 单词归类
Key start_I
Val 单词计数表
Fun 单词归类
Key start_Z
Val 单词计数表
单词归类 PE 将出现频率最高的 M 个单词发送给单词排序 PE
Fun 单词排序
Key W_S_A-H
Val 单词首字母为 A-H 的最大 M 个单词
Fun 单词排序
Key W_S_I-O
Val 单词首字母为 I-O 的最大 M 个单词
Fun 单词排序
Key W_S_P-Z
Val 单词首字母为 P-Z 的最大 M 个单词
单词排序 PE 将出现频率最高的 M 个单词发送单词统计输出 PE
Fun 单词统计输出
Key Top_m
Val 出现频率最高的 M 个单词

2. JOB