大并发 实时数据处理解决方案：

1. 架构

整体系统采用分布式结构搭建，系统功能包括：

* 对外提供海量数据的载入，测试结果会看到具体指标项
* 海量数据查询功能

此外系统本身具有如下特点：

* 可扩展性，集群上面的节点有自定义的角色划分，并且各个角色可以实现动态搜索， 根据业务量增大或者缩小，可以动态添加或者减少节点数量。
* 负载均衡：这里负载均衡的目的是为了减少单台服务器压力过大（尤其在网卡，内存方面）的情况。这里提供两种负载均衡：第一种针对后端倒入数据的节点：采用均衡 (默认是roundrobin)方式动态发现，分派预定块数据。第二种针对前端结构请求端节点，也是采用均衡 (默认是roundrobin)方式动态发起请求。
* 故障转移：针对后端数据导入节点，系统采用定义了物理度量值功能，如果单节点不符合条件（比如down机，或者正在处理数据导致cpu，网络，内存等饱和）分发节点会动态选择其他候选节点来处理。

下图是系统的业务导向：



客户端可以发起标准的http-rest请求来调用系统，此外也可以使用客户端程序来调用系统。在部署架构上来讲，如果使用http-rest方式，需要在分布式系统前段再部署一层http负载均衡软件（比如nginx,haproxy等等）。如果使用客户端程序的话那么不需要加装http负责均衡层，客户端程序本身自带负载均衡策略。

数据库端我们目前采用hadoop-hbase作为存储解决方案，可以使用其它系统动态替换。

系统分为以下4种角色：



系统角色划分

**备注：各个角色的之间通讯都使用异步通讯方式，极大提升了cpu的利用率以及系统的吞吐率**

* 业务节点(master) ：是系统的业务前段，负责对外接受数据，分派任务给工作节点。

这里简单图示下master节点：



* 工作节点(worker) ：是系统的具体工作节点，负责处理具体的业务，这里当然就是导入数据啦。

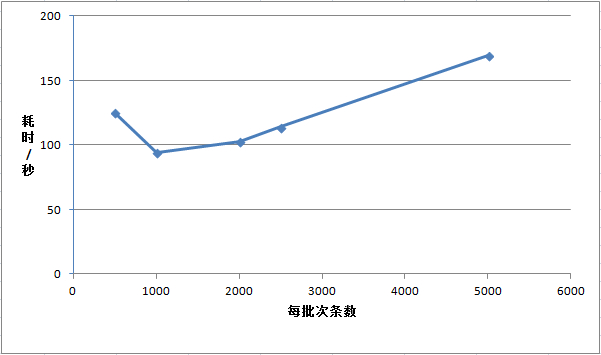
这里简单图示下worker节点：



* 管理节点(seed)：集群内部管理节点，必须存在一个以上的seed节点，这样集群才能正常运行。运维人员可以查看该节点数据来观察集群的健康状况。
* 监控节点(monitor)：集群管理节点，可以监控集群节点的状况（包括：活动，停顿，在工作等等情况）系统默认采用日志输出的方式，但也可以接入第三方软件实现动态监控。

**备注：以上角色定义都是自定义，可以根据需要自定义其它角色**

1. 测试结果



这里展示的使用20万测试数据（根据业务需要的模拟数据）的测试结果，测试方式采用客户端http-rest方式，分别采用每500，1000，2000，2500，5000一个数据包发送系统对外总耗时情况。从图上看出1000条批次的耗时100秒左右是最优的。