# ArangoDB集群模式测试文档

## 测试环境:

两台服务器

10.61.2.206

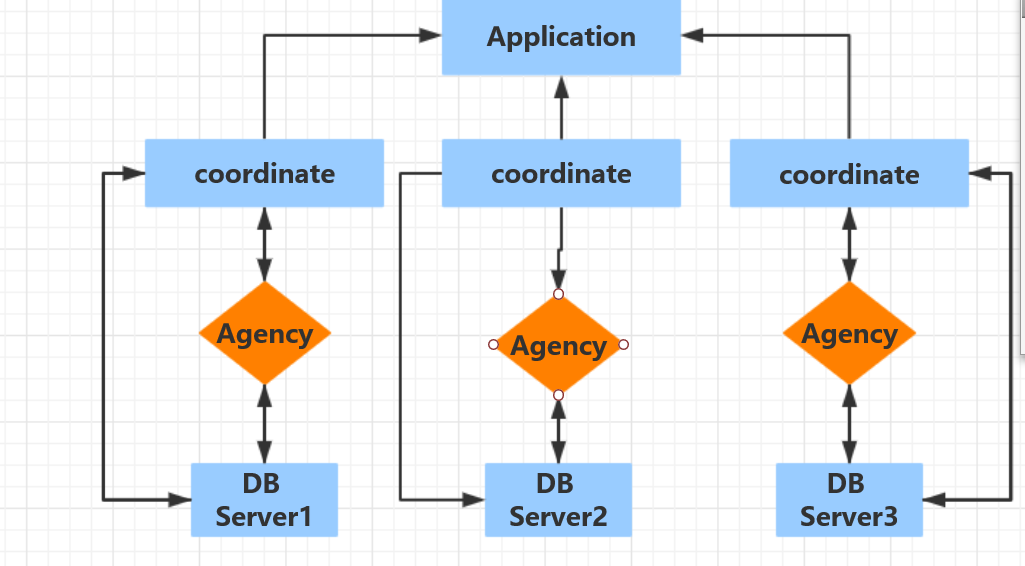
10.61.2.127

所用端口

协调器层开放三个端口：10.61.2.206：8529、10.61.2.127：8529、10.61.127：8534

使用mmfile存储引擎进行测试，结点数为4千万，边集为2亿

设计的集群架构为：



## MMFILE存储引擎

导入大图后占用内存情况

10.61.2.206 占用内存30G左右，10.61.2.127占用60G左右

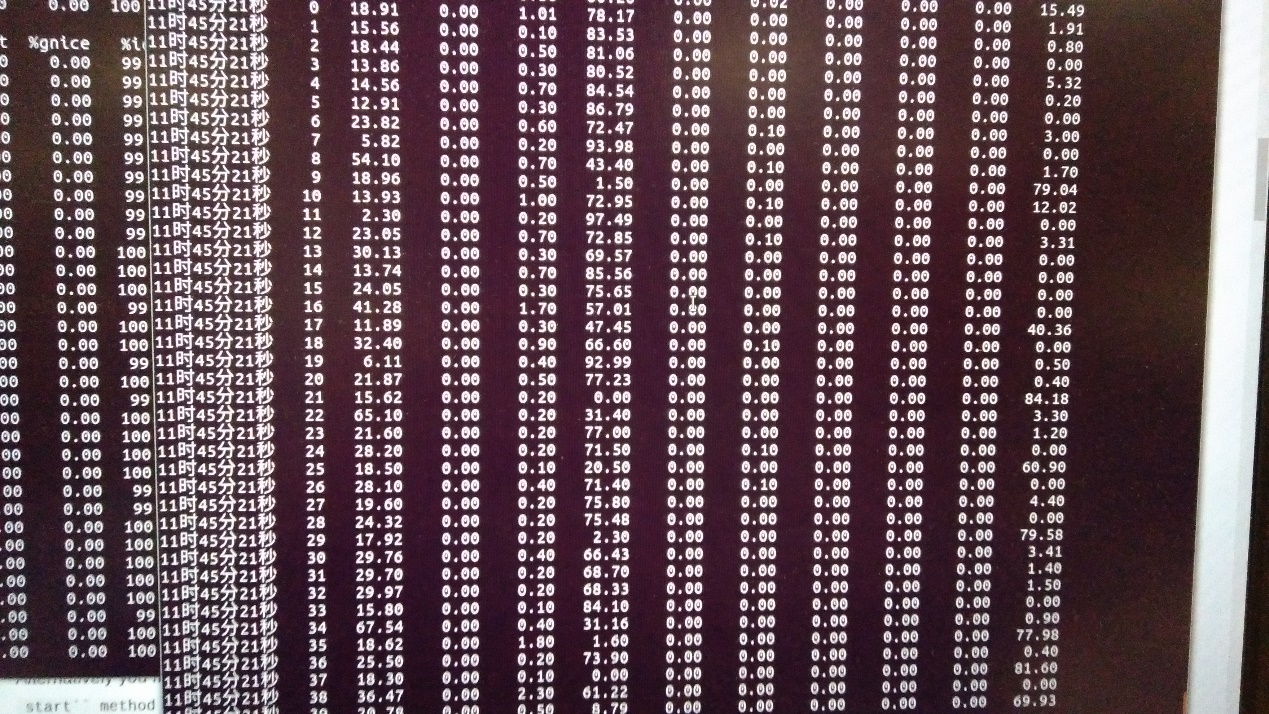
使用AQL的bfs遍历，尝试不用的开始点进行遍历，发现时间收敛于120秒左右

BFS具体时间截图：

使用AQL的Dfs遍历，尝试不用的开始点进行深度为10的遍历，发现时间也收敛于120秒左右，稍微慢于BFS

DFS具体时间截图：

进行图遍历时的CPU占用（10.6.2.127）：



集群模式下使用Pregel框架跑图算法需要对图事先做划分，划分方法为将链接到同一个结点的所有出度边存储到这个结点的同一个分片上

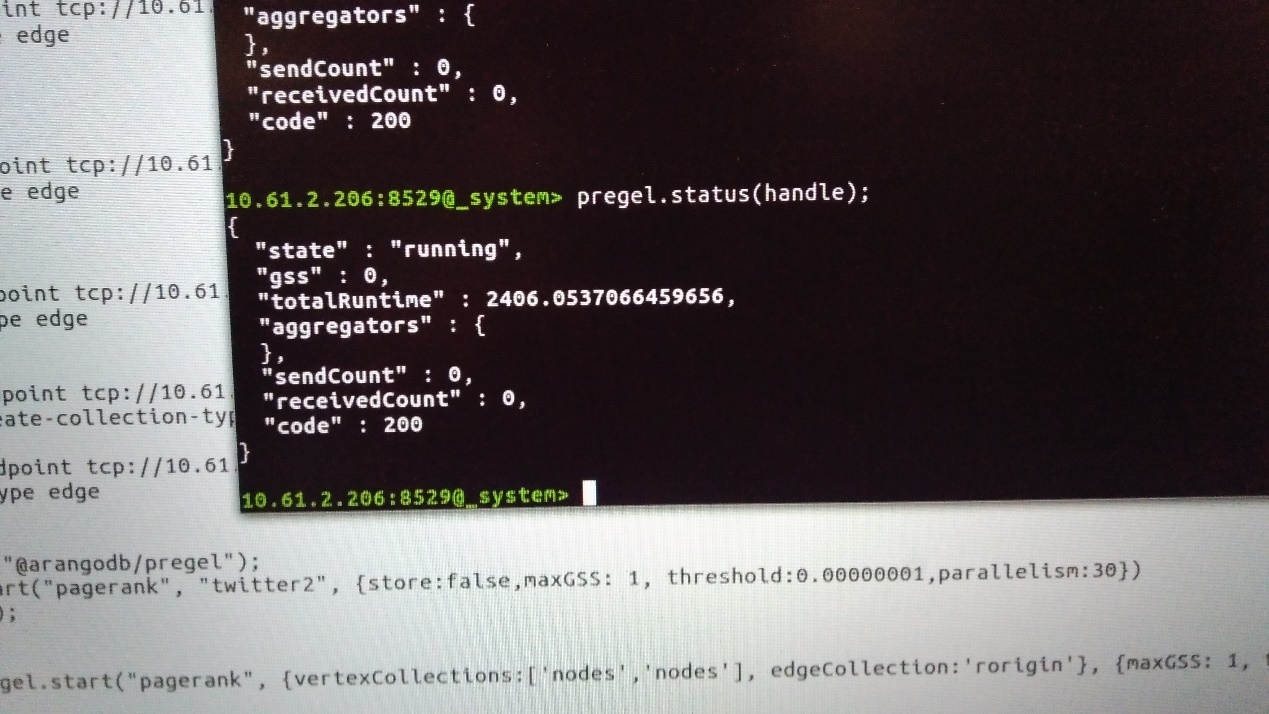
本次测试按照官方文档的方法进行划分，原文如下：

To enable iterative graph processing for your data, you will need to ensure that your vertex and edge collections are sharded in a specific way.

The pregel computing model requires all edges to be present on the DB Server where the vertex document identified by the \_from value is located. This means the vertex collections need to be sharded by '\_key' and the edge collection will need to be sharded after an attribute which always contains the '\_key' of the vertex.

Our implementation currently requires every edge collection to be sharded after a "vertex" attributes, additionally you will need to specify the key distributeShardsLike and an **equal** number of shards on every collection. Only if these requirements are met can ArangoDB place the edges and vertices correctly.

使用arangodb中集成的pregel框架运行Pagerank算法迭代一次时间开销超过40分钟，并未能跑出具体时间，慢于单机模式，其中阈值设置为0.000001



进行pagerank的CPU占用情况,初始时刻CPU有明显占用率升高，但20秒左右之后CPU占用率为1%以下