

KYLIN-5349

现状

job 并发检查

- 启动或新建的时候每个项目下会初始化一个 job schedule 线程池, 此时使用的系统级配置 `kylin.job.max-concurrent-jobs`
- 在执行 job 的时候会二次验证执行的数量不能超过 `kylin.job.max-concurrent-jobs` 配置的值
- 所以每个项目都是同样的线程池大小

构建子分区接口

- 在构建子分区接口中用配置 `kylin.job.max-concurrent-jobs` 来控制 segment 的任务并发数

```
1 private void checkConcurrentSubmit(int partitionSize) {  
2     int runningJobLimit = getConfig().getMaxConcurrentJobLimit();  
3     int submitJobLimit = runningJobLimit * 5;  
4     if (partitionSize > submitJobLimit) {  
5         throw new KylinException(JOB_CONCURRENT_SUBMIT_LIMIT, submitJobLimit)  
6     }  
7 }
```

Dev Design

job 并发检查

- 初始化 job schedule 线程池时,使用项目级配置 `kylin.job.max-concurrent-jobs`
- 执行 job 二次验证时,同样使用项目级配置 `kylin.job.max-concurrent-jobs`
 - `FetcherRunner` 处理 job 最上面, 检查当前线程池的大小与项目配置项的关系
 - 线程池 `maximumPoolSize` == 项目配置项: 不操作
 - 线程池 `maximumPoolSize` < 项目配置项: 修改当前线程池的 `corePoolSize` 和 `maximumPoolSize`
 - 线程池 `maximumPoolSize` > 项目配置项 并且 线程池工作数量 `activeCount` <= 项目配置项: 修改当前线程池的 `corePoolSize` 和 `maximumPoolSize`

- 即扩容时直接修改线程池的大小, 缩容时只有执行任务的work数量小于配置项时才能修改线程池的大小

构建子分区接口

- 在判断时获取项目级别配置, project信息从外部传进来

限制:

- 缩容时不会立即触发修改线程池大小, 会过段时间修改 (因为二次验证的控制, 当前job执行完成后就会修改)

Test

项目级别添加 `kylin.job.max-concurrent-jobs` 设置为 1, 同时运行 3个, 1个 running, 2个 pending

▶ 恢复 ◀ 重启 || 暂停 × 终止 🗑 删除

<input type="checkbox"/>	类型 ▾	对象 ▾	数据范围	状态 ▾	任务开始时间
<input type="checkbox"/>	▶ 刷新数据	test2	全量加载	<div></div>	2022-10-
<input type="checkbox"/>	▶ 刷新数据	test1	全量加载	<div></div>	2022-10-
<input type="checkbox"/>	▶ 刷新数据	test	全量加载	<div></div> 50%	2022-10-

共 3 条 20条/页 < 1 > 前往 1 页

`kylin.job.max-concurrent-jobs` 修改为 2, 等待约 30s
(`kylin.job.scheduler.poll-interval-second` 控制), 2 个 running, 1 个 pending

批数据任务列表

▶ 恢复 ◀ 重启 || 暂停 × 终止 🗑 删除

🔍 搜索任务对象或任务

<input type="checkbox"/>	类型 ▾	对象 ▾	数据范围	状态 ▾	任务开始时间 ▾	总耗时 ▾
<input type="checkbox"/>	▶ 刷新数据	test2	全量加载	<div></div>	2022-10-25 20:47:55 GMT+8	0 mins
<input checked="" type="checkbox"/>	▶ 刷新数据	test1	全量加载	0%	2022-10-25 20:47:51 GMT+8	0 mins
<input type="checkbox"/>	▶ 刷新数据	test	全量加载	<div></div> 50%	2022-10-25 19:24:13 GMT+8	85.94 mins

共 3 条 20条/页 < 1 > 前往 1 页

任务 ID: 5812b427-ce4fa787-

对象: test1

状态: **RUNNING**

等待时间: 2.39m

持续时间: < 0.01m

任务详情

🔍 检测资源

等待时间: 2.39m

`kylin.job.max-concurrent-jobs` 再次修改为 1，2 个 running，1 个 pending，即 running 的任务不会受到影响

类型	对象	数据范围	状态	任务开始时间
刷新数据	test2	全量加载	<div></div>	2022-10-25 20:47:55 GMT+8
刷新数据	test1	全量加载	<div>10%</div>	2022-10-25 20:47:51 GMT+8
刷新数据	test	全量加载	<div>50%</div>	2022-10-25 19:24:13 GMT+8