

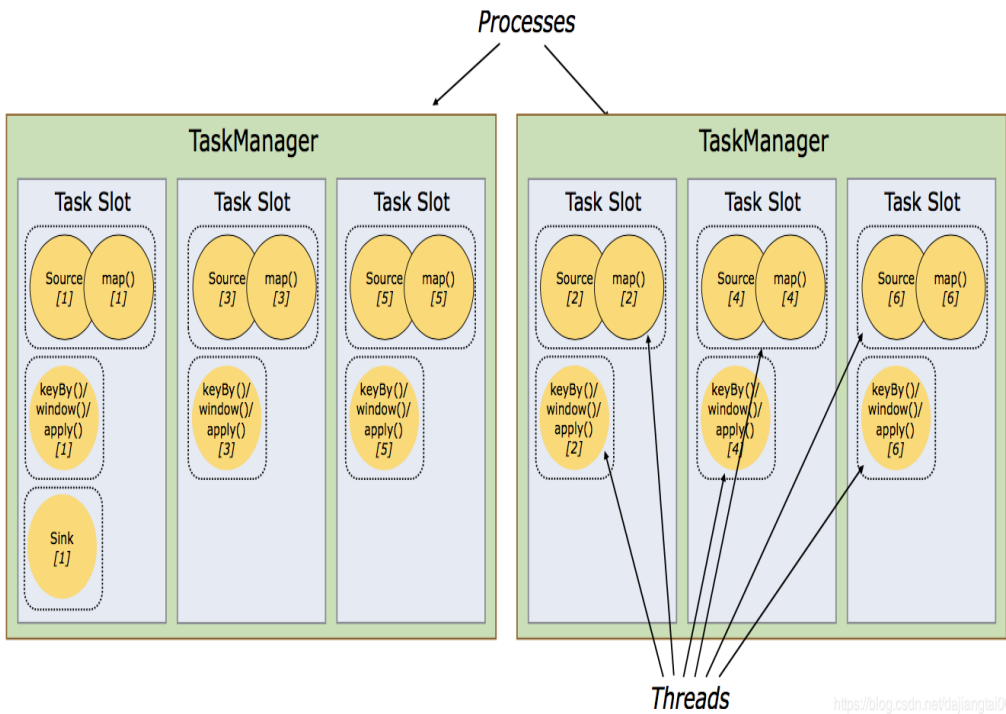
flink随手笔记之Slot分配与共享

原创 大数据研习社 发布于2019-03-18 16:02:19 阅读数 1706 ☆ 收藏

说明：本文为《Flink大数据项目实战》学习笔记，想通过视频系统学习Flink这个最火爆的大数据计算框架的同学，推荐学习CSDN官网课程：

Flink大数据项目实战：<http://t.cn/ExrHPI9>

共享Slot



默认情况下，Flink 允许subtasks共享slot，条件是它们都来自同一个Job的不同task的subtask。结果可能一个slot持有该job的整个pipeline。

允许slot共享有以下两点好处：



举报

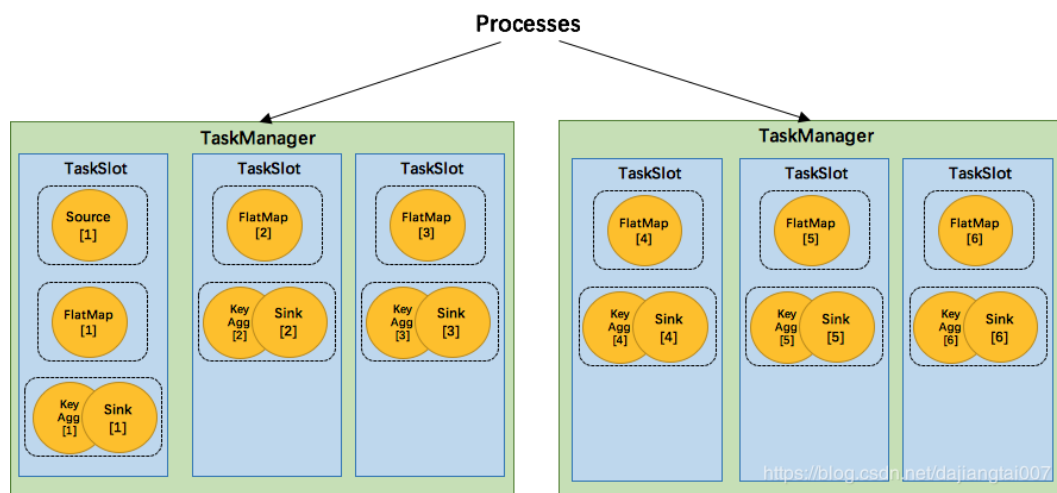
展开

<https://blog.csdn.net/dajiangtai007>

1.Flink集群需要的任务槽与作业中使用的最高并行度正好相同(前提，保持默认SlotShringGroup)。也就是说我们不需要再去计算一个程序总共会起多少个task了。

2.更容易获得更充分的资源利用。如果没有slot共享，那么非密集型操作source/flatmap就会占用同密集型操作 keyAggregation/sink 一样多的资源，将task的2个并行度增加到6个，能充分利用slot资源，同时保证每个TaskManager能平均分配到重的subtasks。

2.2共享Slot实例



将 WordCount 的并行度从之前的2个增加到6个（Source并行度仍为1），并开启slot共享（所有operator都在default共享组），将得到如上图所示的slot分布图。

首先，我们不用去计算这个job会其多少个task，总之该任务最终会占用6个slots（最高并行度为6）。其次，我们可以看到密集型操作 keyAggregation/sink 被平均地分配到各个 TaskManager。

2.3 SlotSharingGroup(soft)

SlotSharingGroup是Flink中用来实现slot共享的类，它尽可能地让subtasks共享一个slot。

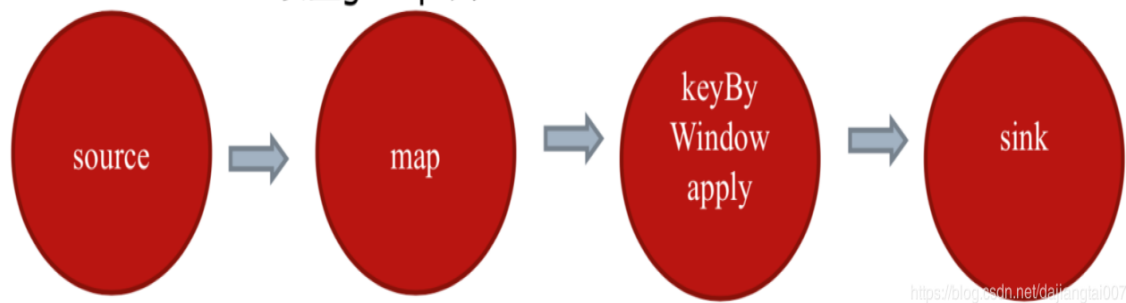
保证同一个group的并行度相同的sub-tasks 共享同一个slots。算子的默认group为default(即默认一个job下的subtask都可以共享一个slot)

为了防止不合理的共享，用户也能通过API来强制指定operator的共享组，比如：someStream.filter(...).slotSharingGroup("group1");就强制指定了filter的slot共享组为group1。怎么确定一个未做SlotSharingGroup设置算子的SlotSharingGroup什么呢(根据上游算子的group 和自身是否设置group共同确定)。适当设置可以减少每个slot运行的线程数，从而整体上减少机器的负载。



举报

设置group 为test

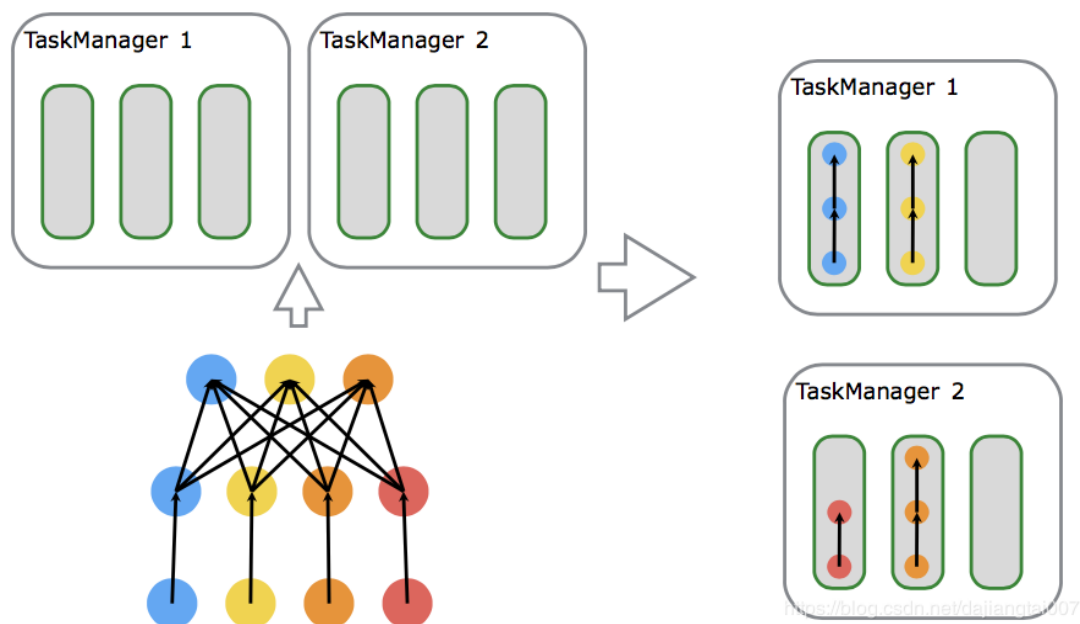


2.4 CoLocationGroup(强制)

CoLocationGroup可以保证所有的并行度相同的sub-tasks运行在同一个slot，主要用于迭代流(训练机器学习模型)。

3. Slot & parallelism的关系

3.1 Slots & parallelism



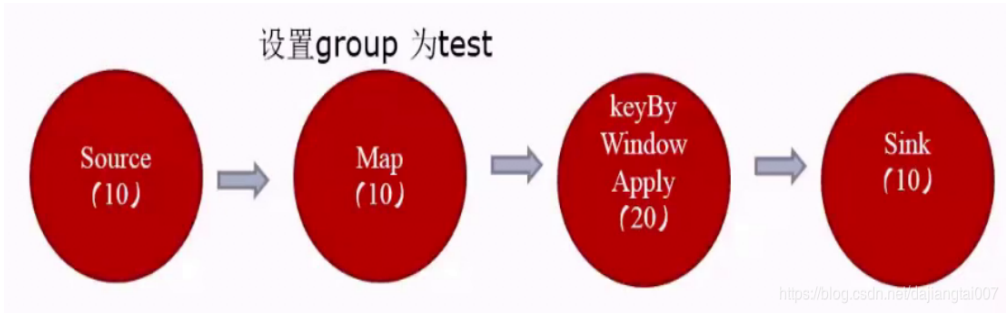
如上图所示，有两个TaskManager，每个TaskManager有3个槽位。假设source操作并行度为3，map操作的并行度为4，sink的并行度为4，所需的ots数与job中task的最高并行度一致，最高并行度为4，那么使用的Slot也为4。

3.2如何计算Slot

如何计算一个应用需要多少slot?



举报



如果不设置SlotSharingGroup，那么需要的Slot数为应用的最大并行度数。如果设置了SlotSharingGroup，那么需要的Slot数为所有SlotSharingGroup的最大并行度之和。比如已经强制指定了map的slot共享组为test，那么map和map下游的组为test，map的上游source的组为默认的default，此时default组中并行度为10，test组中最大并行度为20，那么需要的Slot=10+20=30。



获取更多大数据学习资料、视频、技术交流

请加QQ群：732021751

加群备注：flink

<https://blog.csdn.net/dajiangtai007>

👍

🔗

💬

☆

📄


⏪

⏩

🏆

👤

👍 点赞 ☆ 收藏 🔗 分享 ...



大数据研习社

发布了71 篇原创文章 · 获赞 200 · 访问量 27万+

Flink中slot的一点理解

参考：<https://ci.apache.org/projects/flink/flink-docs-release-1.3/setup/config.html#configuring-taskman...>

阅读数 6087

博文 来自：[a6822342的博客](#)

私信

举报

关注

