

批处理任务还是逐个执行

今天，我们的公寓供水出现了一些问题，暂时无法提供水源。用餐结束后，看着家人和我留下的一堆碗碟，我思考了几个问题。

其中一个问题是如何在没有供水的情况下让洗碗机继续工作。它可能被设计为可以连接到一个水桶。此外，水管的连接头应该具有灵活性，以便轻松从公共水源切换到私人的自制水桶。

另一个问题是应该批量处理还是逐个处理任务。我们可以每顿饭后洗碗，或者等上一两天再洗。

这让我联想到编程。我们可以批量处理任务或逐个处理任务。

批量处理任务会带来一个明显的问题：它需要更多的资源。随着洗涤的延迟，碗碟的数量会增加，并且需要更多的内存空间来积累数据以延后处理。

在现实生活中，处理空间或物品的数量是有上限的。例如，洗碗机最多只能处理二十个碗碟，就像程序在计算机上的内存有限，或者道路上能通过的车辆数量有限。

另一个问题是如何分割任务。我们应该一次分割一个项目，还是三个项目？

对于碗碟或汽车，简单地将每个项目作为一个单元。这意味着一个碗就是一个碗，一辆车就是一辆车。通常，它们无法被分解成更小的部分。尽管还有一些例外，比如一辆大卡车可以载运多辆车；一辆大卡车可以被分解为一个单元和多辆车通过道路。

在编程中，灵活性要大得多。即使是插入或更新 SQL，也可以被分解成更小的部分，更不用说下载任务、DFS 搜索或查询了。

好了，现在我们已经思考了处理单元。那么，问题是我们应该在一个批次中处理多少个单元。可以是从一个到总数之间的任何数字。

这里的问题是，一个任务的批次数量是固定的还是动态的。对于生成式 AI，输入文本的总字符数是灵活的。它有一些上下文限制或输入限制，但在其限制范围内是灵活的。

使用洗碗机时，其内部空间是有限的。在这个限制内，碗碟的数量是灵活的。我们通常会尽可能多地放入机器中。

对于程序来说，数据库一次处理的批量 SQL 有一个限制。在这个限制内，它可以处理的 SQL 数量是灵活的。但我们应该考虑将 SQL 从客户端传递到数据库服务器的网络任务，用户可以等待的时间，以及如果批次中的一个单元任务失败会发生什么。

因此，对于一个批次中应该处理多少个单元任务的问题，我们应该考虑任务的目标、下游消费者或处理器的限制以及失败的概率。

这种思维方式可以应用于许多事情。基本上有两个问题需要考虑：什么是单元任务，以及一个批次中应该处理多少个单元。通过思考这些问题，我们可能会找到一个最佳解决方案。