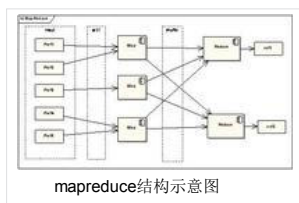


MapReduce

百科名片

[求助编辑](#)



MapReduce是一种编程模型，用于大规模数据集（大于1TB）的并行运算。概念"Map（映射）"和"Reduce（化简）"，和他们的主要思想，都是从函数式编程语言里借来的，还有从矢量编程语言里借来的特性。他极大地方便了编程人员在不会分布式并行编程的情况下,将自己的程序运行在[分布式系统](#)上。当前的[软件](#)实现是指定一个Map（映射）函数，用来把一组键值对映射成一组新的键值对，指定并发的Reduce（化简）函数，用来保证所有映射的键值对中的每一个共享相同的键组

目录

- [映射和化简](#)
- [分布和可靠性](#)
- [用途](#)
- [参考](#)

映射和化简

[编辑本段](#)

简单说来，一个映射函数就是对一些独立元素组成的概念上的列表（例如，一个测试成绩的列表）的每一个元素进行指定的操作（比如前面的例子里，有人发现所有学生的成绩都被高估了一分，他可以定义一个“减一”的映射函数，用来修正这个错误。）。事实上，每个元素都是被独立操作的，而原始列表没有被更改，因为这里创建了一个新的列表来保存新的答案。这就是说，**Map**操作是可以高度并行的，这对高性能要求的应用以及并行计算领域的需求非常有用。

而化简操作指的是对一个列表的元素进行适当的合并（继续看前面的例子，如果有人想知道班级的平均分该怎么做？他可以定义一个化简函数，通过让列表中的元素跟自己的相邻的元素相加的方式把列表减半，如此递归运算直到列表只剩下一个元素，然后用这个元素除以人数，就得到了平均分。）。虽然他不如映射函数那么并行，但是因为化简总是有一个简单的答案，大规模的运算相对独立，所以化简函数在高度并行环境下也很有用。

分布和可靠性

[编辑本段](#)

MapReduce通过把对数据集的大规模操作分发给网络上的每个节点实现可靠性：每个节点会周期性的把完成的工作和状态的更新报告回来。如果一个节点保持沉默超过一个预设的时间间隔，主节点（类同**Google File System**中的主服务器）记录下这个节点状态为死亡，并把分配给这个节点的数据发到别的节点。每个操作使用命名文件的原子操作以确保不会发生并行线程间的冲突；当文件被改名的时候，系统可能会把他们复制到任务名以外的另一个名字上去。（避免副作用）。

化简操作工作方式很类似，但是由于化简操作在并行能力较差，主节点会尽量把化简操作调度在一个节点上，或者离需要操作的数据尽可能近的节点上了；这个特性可以满足**Google**的需求，因为他们有足够的带宽，他们的内部网络没有那么多的机器。

用途

[编辑本段](#)

在**Google**，**MapReduce**用在非常广泛的[应用程序](#)中，包括“分布**grep**，分布排序，web连接图反转，每台机器的词矢量，web访问日志分析，反向索引构建，文档聚类,机器学习，基于统计的机器翻译...”值得注意的是，**MapReduce**实现以后，它被用来重新生成**Google**的整个索引，并取代老的ad hoc程序去更新索引。

MapReduce会生成大量的[临时文件](#)，为了提高效率，它利用**Google**文件系统来管理和访问这些文件。

其他实现

Nutch项目开发了一个实验性的**MapReduce**的实现[2]。

中国人民大学信息学院开发了国内第一个**MapReduce**在线测试平台。

参考

[编辑本段](#)

Dean, Jeffrey & Ghemawat, Sanjay (2004). "MapReduce: Simplified Data Processing on Large Clusters". Retrieved Apr. 6, 2005

"Our abstraction is inspired by the map and reduce primitives present in Lisp and many other functional languages." -"MapReduce: Simplified Data Processing on Large Clusters"

扩展阅读:

- 1 <http://zh.wikipedia.org/w/index.php?title=MapReduce&variant=zh-cn>
- 2 <http://cloudcomputing.ruc.edu.cn/>

开放分类:

编程, [互联网](#), [计算机](#), [google](#), [并行计算](#)

“MapReduce”相关词条:

[什么是相关词条](#) [我来完善](#)

208
本词条对我有帮助

百度百科中的词条内容仅供参考，如果您需要解决具体问题（尤其在法律、医学等领域），建议您咨询相关领域专业人士。

[添加到收藏](#)

分享到:

合作编辑者

[林春彬](#), [百科ROBOT](#), [billfu21](#), [aooyou](#), [zxq2005ce](#), [zhoupo86](#), [xampp](#), [hanjunjie1986](#) [更多](#)
如果您认为本词条还需进一步完善, 百科欢迎您也来参与 [编辑词条](#)在开始编辑前, 您还可以先学习[如何编辑词条](#)

如想投诉, 请到[百度百科投诉中心](#); 如想提出意见、建议, 请到[百度百科吧](#)。