并行化MATLAB在天河一号上的实现

Implementation of Parallelized MATLAB on TianHe 1A

MathWorks 公司　虞旭东

MATLAB 和 Simulink 简介

MATLAB 和 Simulink 是美国 MathWorks 公司的软件产品。MATLAB 是著名的以矩阵计算为基础的科学计算语言，具有高效、易用、图形化等特征，配合大量的工具箱，可以帮助各行各业的科研人员、工程师、分析师快速解决各种复杂的应用数学问题；Simulink 是 MathWorks 基于 MATLAB 语言开发的工业仿真平台，被广泛地应用于汽车、航空航天、能源交通、电子通讯等多个领域。

MATLAB 和 Simulink 的并行化

计算量巨大以致需要并行化是一个非常常见的需求。MathWorks 自 2006 年起开始支持并行化 MATLAB/Simulink，到现在的 MATLAB R2012B 版本，已经在上百个 HPC 中心实现了 MATLAB/Simulink 并行化，还将 MATLAB 的 HPC 服务扩展到了如亚马逊 EC2、EGI（European Grid Infrastructure）、SharcNet、TaraGrid 等云计算或网格计算中心。

2012 年 4 月份，MathWorks 在天河 1A 超级计算机上成功测试部署了 MATLAB 并行化方案，并通过位于某大学的远端客户机向天河 1A 提交作业并成功实现了 SOA（Service Oriented Architecture）的仿真并行化。

MATLAB 和 Simulink 并行化的语言概念

MATLAB 和 Simulink 在 CPU 上并行化的基础是 MPI，使用的是阿贡国家实验室的标准 MPICH2；在 GPU 上 MathWorks 则使用 nVidia 的 CUDA。

MathWorks 在 MPI 和 CUDA 基础上发展了一套独特的平行化语言系统。

MATLAB 的并行语言主要可以分为 3 个部分：

1 工具箱内置并行化

MATLAB 提供了包含了大量数学函数的大量的工具箱给各行各业。MathWorks 现已并行化了 14 个工具箱。

2 任务并行语言结构

任务并行语言指的是并行化 MATLAB 各个进程各自完成各自的任务，各个进程之间并无太多的数据交换。

MATLAB 主要提供了两套任务并行语言语法，createJob/createTask 和 parfor。

createJob/createTask 是 MATLAB 并行语言中唯一不依赖于 MPI 的语法，非常类似于一般的分布式计算方式，首先在 MATLAB 中使用

```
j=createJob
```

来创建一个作业，然后用

```
createTask(j,@func,1,{5})
```

在 j 的句柄下创建多个作业，func 是用户函数，1 是输出变量数，{5} 是输入数据。最后提交作业：

```
j.submit
```

parfor 则是 MATLAB 在 MPI 基础上开发的语法，即 parallel for。用户可以简单地将 for 循环语句代替为 parfor，程序中的循环会被自动地并行化，由于使用了 MPI 作为通讯协议，parfor 可以跨机运行。

3 数据并行语言结构

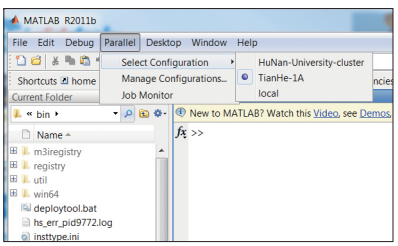
数据并行语言结构提供了基于 MPI 进程间大量互换数据的 MATLAB 语法，基础语法是 spmd（single program multi data），并在其结构中使用并行进程间数据交换的操作。

在 spmd 语法结构中，用户可以使用类 MPI 的语法来进行数据交换操作，如 labsend 和 labreceive。但更为常见的是使用 MATLAB 的并行数据结构 codistributed array，很好地解决了数据量过大、单机无法处理的问题。

并行化 MATLAB 用户体验

MATLAB 在并行化上提供了 2 套方案，一套是为用户工作站设计的 MATLAB 并行工具箱（Parallel Computing Toolbox，简称 PCT）；一套是为计算集群、网格和计算云设计的 MATLAB 分布式计算服务器（MATLAB Distributed Computer Server，简称 MDCS）。

用户使用 PCT 可以将并行代码运行在多核的用户工作站，以充分利用工作站计算资源；而 MDCS 则是服务器端软件。无论是本地使用 PCT 多核计算还是提交到集群让 MDCS 计算，用户代码和提交方式不需要做



MATLAB图形界面中可进行不同的配制选择

任何改变，只需要在 MATLAB 图形界面中选择不同的 configuration 设置即可。

MATLAB 并行化既提供了批处理的作业提交方式，也提供了交互式（interactive）的作业提交方式。无论计算机集群使用何种调度器，MATLAB 的 PCT/MDCS 都可以提供 SOA（Service Oriented Architecture）的服务。

一个典型的并行化 MATLAB 用户程序开发可以通过以下步骤来实现：

· 首先开发串行程序。

· 在用户 12 核的本地工作站上，使用交互式并行界面：

matlabpool open 12

用户可以在此交互式并行界面中，用 MATLAB 并行化语言，实时编写调试，将串行程序并行化。难度远比在计算机集群批处理作业要容易。

· 将编好的程序用批处理或交互式提交到远程计算机集群。

并行化 MATLAB 在天河一号上的实现

1 天河一号超级计算机设计的独特性

集成 MDCS 与天河一号的难点主要有 3 个：

· 天河一号集群使用基于 SLURM（Simple Linux Utility for Resource Management）的自制调度器；

· 天河一号集群中大多数节点是无盘工作站；

· 很多天河一号 MATLAB 用户都使用 WINDOWS 操作系统进行远程工作，而天河调度器不支持 WINDOWS 。

2 解决方案

针对天河一号超级计算机，我们对 PCT/MDCS 的部署方案进行了调整和重新设计。

首先，对 MDCS 的安装进行了调整，使之能够在无盘工作站运行良好。其次，在客户端的 MATLAB 中，内置了基于 Java 的 ssh 客户端。基于此 MATLAB SSH，使用 sftp 协议，将客户端的 MATLAB 运行文件结构和天河一号的共享文件系统进行了文件结构镜像。最后，使用 ssh 协议，在天河一号的登录节点进行远程作业提交，并使用文件结构镜像，将运算结果传输回客户端。

3 测试结果

我们在某大学的计算中心设置了 MATLAB/PCT 的客户端，使用 VPN 远程连接到天河一号系统，提交远程作业客户端的 MATLAB 并行程序正确地在天河一号系统中的 MDCS 上运行并返回结果。

我们进行了任务并行程序的测试，程序是基于 parfor 的蒙特卡罗仿真，运算速度基本达到了相对 worker 数目的线性加速。

后续工作及展望

通过 PCT/MDCS 和天河一号的集成，我们基本完成了 MATLAB 并行化的 IT 架构配置，有一些后续工作尚未完成。

1 GPU 的并行化设置

天河一号的许多计算节点都配置了 NVIDIA 的 Tesla GPU，我们需要将 MATLAB 现有的 GPU 运算方案在天河一号上配置并将之并行化。

2 基于 RDMA (Remote Direct Memory Access) 协议的 MPI 在 MDCS 上的实现

由于标准 MPICH2 是基于以太网 TCP/IP 协议，不能很好地利用天河一号上的 Infiniband 的低延迟高速网络的特性，我们需要使用基于 RDMA 协议的 MPI，如 Intel MPI 或 MVAPICH2，可以大大加快数据并行语言程序的运算速度。

3 MATLAB 编译应用程序与天河一号上 MDCS 的连接

MathWorks 为 MATLAB 提供了基于 Java 和 .net 的编译器，用户可以将 MATLAB 程序编译为独立于 MATLAB 环境的可执行程序。我们同样可将 MATLAB 并行化语言进行编译，并需要在天河一号上配置 MDCS 的 Java 运行环境，使 MATLAB 编译应用程序与天河一号可以直接连接。

（责编　三丰）

公　示

根据新闻出版总署《关于开展新闻记者证核发情况自查工作并重申有关规定的紧急通知》（《2009》299 号）、《新闻记者证管理办法》、《关于 2009 年换发新闻记者证的通知》、《关于期刊申领新闻记者证的有关通知》、《关于广播电影电视新闻单位申领新闻记者证的通知》要求，我单位航空制造技术杂志社已对申领记者证人员的资格进行严格审核，现将我单位已领取或拟领取新闻记者证人员名单进行公示，公示期 2013 年 3 月 15 日 ~3 月 25 日。举报电话为 010-83138953。

已领取新闻记者证名单：

刘柱　记者证统一编号 K11438701000002

拟领取新闻记者证名单：

刘振敏