

轻量级企业分布式计算系统构建方案

dungeonsnd@gmail.com

v0.1

2015/7/3

一 背景

本方案的目的是把很多企业的空闲机器资源利用起来，构建一个轻量级的企业分布式计算系统，使企业在有限资源的情况下也能拥有较强的计算能力。有个较强的计算能力之后，就可以为自身甚至为外界提供无偿/有偿的计算服务。

由于移动互联网时代的到来，越来越多的移动 app 投入市场。而这些移动手机、平板、电视及穿戴设备受硬件及电量限制，往往不能进行密集的 CPU 计算，所以这些计算就需要放在服务器端。这有区别传统的 PC 客户端程序，它们往往使用胖客户端来分担服务器的计算量。

很多企业有很多 PC 机(实体的或虚拟的)，但是大部分机器的绝大部分时间资源(如 CPU)是处于闲置状态。比如很多企业有较多的产品线，而不同的产品线有各种机器环境，如生产环境、准生产环境、测试环境、开发环境、管理环境，甚至是办公环境等。这样算下来，一个小型企业可以有上百台机器可提供计算，一个千人左右的中型软件类公司可能提供数千台可提供计算的机器。如何把这个机器的计算能力收集并整合起来提供统一的高效计算服务，是很多企业面临的问题。当然，整合的过程中，对所有处理结点机器可以进行灵活调度，不能使用结点上原有的应用卡顿，这才是分布式计算框架要解决的核心问题。显然，要实现这样的整合就需要一个简洁、轻量、可扩展和易维护的基础框架。本文即是解决此问题的一种探索。

二 方案

作者对市面上已有的分布式计算框架(如 Hadoop、Storm、Spark)。另外，作者多年前也了解到有个项目叫 BOINC (<http://www.equn.com/wiki/BOINC>)。它是特性简介如下，

所有分布式计算项目中均需涉及到任务分发、结果回收和验证、积分授予等相关的处理，BOINC 作为一个通用的分布式计算平台，提供了一个考虑周全的软件应用框架，使得基于 BOINC 平台的计算项目在开发时，仅需要关注项目研究内容相关的方面。也因为 BOINC 平台只是一个软件框架，并不对计算项目提供服务器硬件及网络等资源，具体计算项目的服务器仍是项目方自行提供的。

BOINC 平台采用了传统的客户端—服务端构架，服务端部署于计算项目方的服务器，客户端部署于志愿者的计算机，两者之间通过标准的互联网协议进行通讯。BOINC 基于 GNU 的 LGPL 协议发布，服务端一般部署于 Linux 操作系统，客户端则支持所有主流的操作系统，如 Windows、Linux、Mac OS X、FreeBSD 等。作者暂时没有深入研究学习这些现有框架，所以不了解能否或已经实现了本文的目标。

不过作者根据实际需要进行思考，觉得要实现上述的资源整合功能，下图所描述系统架构应该是一种有效探索研究。

如果现有的框架无法实现本文提出的需求，那么可以从下图描述的架构出去探索。如果现有开源项目已经实现了这些功能，我们只需要研究这些项目，并可能稍加改造就可以为我们所用，那就再好不过了。



