



大数据技术与应用专业规划教材

AWS-教育部-AWS产学合作专业综合改革项目规划教材



云计算与大数据 技术理论及应用

◎ 林伟伟 彭绍亮 编著



清华大学出版社



大数据技术与应用专业规划教材

AWS-教育部-AWS产学合作专业综合改革项目规划教材

云计算与大数据 技术理论及应用

◎ 林伟伟 彭绍亮 编著



清华大学出版社
北京

内 容 简 介

为了更好地帮助读者掌握云计算、大数据的技术原理和应用方法,本书使用了作者在云计算与大数据相关项目研发实践中总结的大量编程实例和实际应用开发案例。本书主要剖析了分布式计算技术,Google云、亚马逊云及阿里云技术原理,云存储技术,Hadoop和Spark技术原理;并从技术应用开发实践方面给出大量编程实例和应用开发案例,具体包括P2P应用程序开发,云计算任务调度和能耗优化资源调度算法,大数据分析计算应用案例,生物医药大数据计算案例。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

云计算与大数据技术理论及应用/林伟伟,彭绍亮编著. —北京:清华大学出版社,2019
(大数据技术与应用专业规划教材)
ISBN 978-7-302-52445-8

I. ①云… II. ①林… ②彭… III. ①云计算—数据处理—教材 IV. ①TP393.027 ②TP274

中国版本图书馆CIP数据核字(2019)第043590号

责任编辑:贾 斌
封面设计:刘 键
责任校对:李建庄
责任印制:丛怀宇

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦A座

邮 编:100084

社 总 机:010-62770175

邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

课件下载: <http://www.tup.com.cn>, 010-62795954

印 装 者:三河市铭诚印务有限公司

经 销:全国新华书店

开 本:185mm×260mm 印 张:30.25

字 数:736千字

版 次:2019年7月第1版

印 次:2019年7月第1次印刷

印 数:1~1500

定 价:89.00元

产品编号:074693-01

前言

FOREWORD

背景与内容规划

随着云计算与大数据应用的快速增长,云计算与大数据技术已逐步成为当前及未来信息处理的基础性技术。因此,在该领域急需大量的相关人才与研发人员,本书正是为了适应这一新的发展趋势和需求而编写的。与现有的教材和图书主要以阐述技术原理不同,为了更好地帮助读者深入理解技术原理和应用研发方法,本教材以应用需求为背景剖析这些技术的原理和应用方法,使用了我们在云计算与大数据相关研究和项目开发实践中总结的大量编程实例和实际应用开发案例,从理论上剖析技术原理本质和从实践上解析技术应用方法。本书主要内容涉及传统分布式计算的基本原理、基本开发技术与方法,云计算的技术原理与编程技术,云存储技术,大数据技术原理与平台架构、应用开发技术与应用案例。本书可为计算机相关专业的本科生、研究生和专业技术人员提供丰富、全面的分布式计算、云计算、大数据技术的知识体系和研发实践技术,也能使相关专业科研人员进一步从事相关研究打下良好基础,并对云计算、大数据等新技术的研究与应用起到较好的推动作用。

本书主要内容包括:分布式计算范型技术原理与编程技术,Google云、亚马逊云及阿里云技术原理,云存储技术,Hadoop和Spark技术原理与平台。本书从技术应用开发实践方面给出大量编程实例和应用开发案例,具体包括客户/服务器程序开发、P2P应用程序开发、云计算任务调度算法、云计算能耗优化资源调度算法、3个大数据分析计算应用案例和3个生物医药大数据计算案例。全书共分12章,各章之间的层次关系如下:



教学资源与使用方法

本书配套了 PPT 课件和课后习题参考答案,使用本书进行教学的教师可以到清华大学出版社网站 www.tup.tsinghua.edu.cn 申请,或发送邮件至 linww@scut.edu.cn 或 lin_w_w@qq.com 向作者索取本书相关教学资源。

本书可以作为计算机相关专业的本科高年级学生和研究生教材,学生最好在学习过操作系统、计算机网络、面向对象编程语言之后学习本课程。全书内容可根据不同的教学目的和对象进行选择。但根据本书的定位,建议每章讲授最低学时分配如下:

章 名	建议重点讲授章节	建议学时
第 1 章	所有小节	2
第 2 章	所有小节	8
第 3 章	所有小节	6
第 4 章	4.1,4.3,4.4,4.6 节	8
第 5 章	5.1,5.2 节	4
第 6 章	6.2.1,6.2.2,6.3,6.4,6.5 节	10
第 7 章	所有小节	3
第 8 章	所有小节	4
第 9 章	所有小节	2
第 10 章	所有小节	2
第 11 章	所有小节	2
第 12 章	所有小节	2

此外,本书的教学应该有相应的实验教学内容,建议实验课程的学时数不少于理论课程学时数的三分之一。

致谢

本书受到国家超级计算长沙中心和深圳鹏城实验室的支持,感谢国家重点研发计划 2018YFC0910405,2017YFB0202602,2017YFC1311003,2016YFC1302500,2016YFB0200400,2017YFB0202104;国家自然科学基金 61872084,61772205,61772543,U1435222,61625202,61272056;化学生物传感与计量学国家重点实验室基金等项目,和湖南智超医疗科技有限公司、北京以利天诚科技有限公司(www.ylitech.com)、亚马逊公司等单位和专家的支持。

华南理工大学 林伟伟

2019年6月于广州

第 1 章 绪论	1
1.1 分布式计算概念	1
1.1.1 定义	1
1.1.2 优缺点	1
1.1.3 经典的分布式计算项目	2
1.2 分布式计算模式	4
1.2.1 单机计算	5
1.2.2 并行计算	5
1.2.3 网络计算	6
1.2.4 对等计算	6
1.2.5 集群计算	7
1.2.6 网格计算	7
1.2.7 云计算	7
1.2.8 雾计算	8
1.2.9 边缘计算	9
1.2.10 大数据计算	9
1.3 CAP 定理	11
1.3.1 CAP 定理历史	11
1.3.2 CAP 定理应用	12
1.3.3 CAP 问题的实例	13
习题	14
第 2 章 分布式计算编程基础	15
2.1 进程间通信	15
2.1.1 进程间通信概念	15
2.1.2 IPC 原型与示例	16
2.2 Socket 编程	17
2.2.1 Socket 概述	17
2.2.2 流式 Socket 编程	18

2.3	RMI 编程	25
2.3.1	RMI 概述	25
2.3.2	RMI 基本分布式应用	26
2.4	P2P 编程	35
	习题	44
第 3 章	云计算原理与技术	47
3.1	云计算概述	47
3.1.1	云计算起源	47
3.1.2	云计算的概念与定义	48
3.1.3	云计算与分布式计算	49
3.1.4	云计算分类	51
3.2	云计算关键技术	54
3.2.1	体系结构	54
3.2.2	数据存储	56
3.2.3	计算模型	58
3.2.4	资源调度	59
3.2.5	虚拟化	60
3.3	Google 云计算原理	61
3.3.1	GFS	61
3.3.2	MapReduce	61
3.3.3	BigTable	63
3.3.4	Dremel	66
3.4	亚马逊云服务	69
3.4.1	亚马逊云平台存储架构	69
3.4.2	EC2、S3、SimpleDB 等组件	70
3.5	基于亚马逊云的大数据分析案例	76
3.5.1	亚马逊云平台存储架构	76
3.5.2	亚马逊云的 Web 服务器日志大数据分析案例	79
3.6	阿里云	93
3.6.1	飞天开放平台架构	93
3.6.2	开放云计算服务 ECS	96
3.6.3	开放存储服务 OSS 和 CDN	97
3.6.4	开放结构化数据服务 OTS	99
3.6.5	关系型数据库(RDS)	101
3.6.6	开放数据处理服务(ODPS)	101
	习题	103

第 4 章 云计算编程实践	104
4.1 CloudSim 体系结构和 API 介绍	104
4.1.1 CloudSim 体系结构	104
4.1.2 CloudSim 3.0 API 介绍	110
4.2 CloudSim 环境搭建和使用方法	113
4.2.1 环境配置	114
4.2.2 运行样例程序	114
4.3 CloudSim 扩展编程	117
4.3.1 调度策略的扩展	118
4.3.2 仿真核心代码	120
4.3.3 平台重编译	124
4.4 CloudSim 的编程实践	125
4.4.1 CloudSim 任务调度编程	125
4.4.2 CloudSim 网络编程	132
4.4.3 CloudSim 能耗编程	135
4.5 MultiRECloudSim	147
4.5.1 MultiRECloudSim 体系结构和原理	147
4.5.2 MultiRECloudSim 的 API	153
4.5.3 MultiRECloudSim 的使用方法	156
4.6 云环境任务调度编程实践	170
4.6.1 云计算的资源管理	170
4.6.2 云任务调度模拟实验	173
习题	180
第 5 章 云存储技术	182
5.1 存储基础知识	182
5.1.1 存储组网形态	182
5.1.2 RAID	187
5.1.3 磁盘热备	194
5.1.4 快照	195
5.1.5 数据分级存储概念	196
5.2 云存储概念与技术原理	197
5.2.1 分布式存储	198
5.2.2 存储虚拟化	204
5.3 对象存储技术	208
5.3.1 对象存储架构	208
5.3.2 传统块存储与对象存储	209
5.3.3 对象	209

5.3.4	对象存储系统组成	211
5.4	存储技术趋势	213
5.4.1	存储虚拟化	213
5.4.2	固态硬盘	213
5.4.3	重复数据删除	214
5.4.4	语义化检索	214
5.4.5	存储智能化	214
5.4.6	混合存储系统	215
	习题	215
第6章	大数据技术原理与平台	216
6.1	大数据概述	216
6.1.1	大数据产生的背景	216
6.1.2	大数据的定义	216
6.1.3	大数据的4V特征	217
6.2	大数据存储平台	217
6.2.1	HDFS	217
6.2.2	HBase	226
6.2.3	Cassandra	237
6.2.4	Redis	245
6.2.5	MongoDB	251
6.3	大数据计算模式	259
6.3.1	MapReduce	259
6.3.2	Spark	264
6.3.3	流式计算	272
6.4	典型大数据分析管理平台	278
6.4.1	Cloudera Impala	279
6.4.2	Hortonworks Data Platform	281
6.4.3	HadoopDB	298
6.5	大数据并行计算编程实践	300
6.5.1	基于MAPREDUCE程序实例(HDFS)	300
6.5.2	基于MAPREDUCE程序实例(HBase)	307
6.5.3	基于Spark的程序实例	311
6.5.4	基于Impala的查询实践	316
6.6	大数据研究与发展方向	318
6.6.1	数据的不确定性与数据质量	318
6.6.2	跨领域的数据处理方法的可移植性	319
6.6.3	数据处理的时效性保证——内存计算	319
6.6.4	对于流式数据的实时处理	320

6.6.5 大数据应用.....	321
6.6.6 大数据发展趋势.....	323
习题.....	324
第 7 章 实时医疗大数据分析案例.....	326
7.1 案例背景与需求概述	326
7.1.1 背景介绍.....	326
7.1.2 基本需求.....	326
7.2 设计方案	328
7.2.1 ETL	328
7.2.2 非格式化存储.....	329
7.2.3 流处理.....	329
7.2.4 训练模型与结果预测.....	329
7.3 环境准备	329
7.3.1 节点规划.....	330
7.3.2 软件选型.....	331
7.4 实现方法	332
7.4.1 使用 Kettle/Sqoop 等 ETL 工具,将数据导入 HDFS	332
7.4.2 基于 Spark Streaming 开发 Kafka 连接器组件	338
7.4.3 基于 Spark MLlib 开发数据挖掘组件.....	345
7.5 不足与扩展	349
习题.....	350
第 8 章 保险大数据分析案例.....	351
8.1 案例背景与需求概述	351
8.1.1 背景介绍.....	351
8.1.2 基本需求.....	351
8.2 设计方案	354
8.2.1 基于 GraphX 的并行家谱挖掘算法	354
8.2.2 基于分片技术的随机森林算法.....	356
8.2.3 基于内存计算的 FP-Growth 关联规则挖掘算法	359
8.3 环境准备	360
8.4 实现方法	365
8.4.1 基于 GraphX 的并行家谱挖掘	365
8.4.2 基于分片技术的随机森林模型用户推荐.....	367
8.4.3 基于 FP-Growth 关联规则挖掘算法的回归检验	371
8.4.4 结果可视化.....	376
8.5 不足与扩展	381
习题.....	382

第 9 章 基于 Spark 聚类算法的网络流量异常检测	383
9.1 基本需求与数据说明	383
9.1.1 基本需求	383
9.1.2 数据说明	384
9.2 设计方案	386
9.2.1 聚类问题描述	386
9.2.2 系统整体架构和算法设计	386
9.2.3 数据预处理	387
9.2.4 聚类算法	388
9.2.5 聚类质量评估算法	388
9.2.6 检测算法	389
9.3 实现方法和程序设计	389
9.3.1 搭建 Spark 集群实验平台	390
9.3.2 程序运行说明	390
9.3.3 数据预处理	391
9.3.4 基于 R 的数据分析和可视化	392
9.3.5 聚类算法	394
9.3.6 聚类质量评估	394
9.3.7 异常检测	395
9.4 结果展示	396
9.4.1 Spark 平台说明与作业提交演示	396
9.4.2 聚类算法及其质量评估	397
9.4.3 有效性分析	398
9.4.4 示例说明	399
9.5 展望	399
习题	400
第 10 章 基于 Hadoop 的宏基因组序列比对计算	401
10.1 相关背景介绍与基本需求	401
10.1.1 相关背景	401
10.1.2 基本需求	404
10.2 设计方案	404
10.2.1 串行程序分析	404
10.2.2 并行程序设计	405
10.3 实现方法	406
10.3.1 自定义 Hadoop Streaming Inputformat	406
10.3.2 修改 SOAPaligner 程序的输入文件函数	408
10.4 环境建立和实验数据说明	410

10.4.1	案例环境	410
10.4.2	实验数据	410
10.5	结果展示	411
10.5.1	测试方法	411
10.5.2	测试结果和分析	412
习题	412
第 11 章	基于细胞反应大数据的生物效应评估计算	413
11.1	相关背景介绍与基本需求	413
11.1.1	相关背景	413
11.1.2	基本需求	414
11.2	设计方案	414
11.2.1	基本思路	414
11.2.2	设计框架	415
11.3	环境建立和实验数据说明	416
11.3.1	案例环境	416
11.3.2	实验数据	417
11.4	实现方法	418
11.4.1	算法分析	418
11.4.2	基因谱两两比对——富集积分矩阵并行化计算	422
11.4.3	基因谱聚类分析——KMedoids 算法并行化	428
11.5	结果展示	429
11.5.1	基因谱两两比对——计算富集积分矩阵实验分析	429
11.5.2	基因谱聚类实验分析	431
习题	432
第 12 章	基于 Spark 的海量宏基因组聚类问题分析计算	433
12.1	相关背景介绍与基本需求	433
12.1.1	相关背景	433
12.1.2	基本需求	442
12.2	问题分析与设计方案	444
12.2.1	问题分析	444
12.2.2	设计方案	446
12.3	实现方法	446
12.3.1	基于 Spark 的相似基因对问题的实现	446
12.3.2	利用 LSH 加速相似基因对算法	447
12.3.3	基因图的生成	450
12.3.4	图的基本性质分析	451
12.3.5	基因图聚类	451

- 12.4 环境建立和实验数据说明 454
 - 12.4.1 案例环境 454
 - 12.4.2 实验数据 454
- 12.5 结果展示 454
 - 12.5.1 LSH 方法精确度分析 454
 - 12.5.2 可扩展性分析和加速效果分析 456
 - 12.5.3 基因图顶点的度分布和连通性分析 458
 - 12.5.4 基因图聚类结果分析 459
 - 12.5.5 总结 461
- 习题 462
- 参考文献 463

第 1 章

绪 论

本章首先介绍分布式计算的定义、优缺点和经典的分布式计算项目等,然后概述分布式计算的相关模式,包括单机计算、并行计算、网络计算、对等计算、网格计算、云计算、雾计算、边缘计算和大数据计算,最后重点对 CAP 定理进行了详细介绍和讨论。本章讨论的分布式计算相关概念为后续章节内容的理解打下基础。

1.1 分布式计算概念

1.1.1 定义

分布式计算是一门计算机科学,主要研究对象是分布式系统。在介绍分布式计算概念前,首先简单了解什么是分布式系统。简单地说,一个分布式系统是由若干通过网络互联的计算机组成的软硬件系统,且这些计算机互相配合以完成一个共同的目标(往往这个共同的目标称为“项目”)。分布式计算的一种简单定义是,在分布式系统上执行的计算。

更为正式的定义为,分布式计算是一门计算机科学,它研究如何把一个需要非常巨大的计算能力才能解决的问题分成许多小的部分,然后把这些小的部分分配给许多计算机进行处理,最后把各部分的计算结果合并起来得到最终的结果。本质上,分布式计算是一种基于网络的分而治之的计算方式。

1.1.2 优缺点

在 WWW 出现之前,单机计算是计算的主要形式。自 20 世纪 80 年代以来,由于受 WWW 流行的刺激,分布式计算得到飞速发展。分布式计算可以有效利用全世界联网机器的闲置处理能力,帮助一些缺乏研究资金的、公益性质的科学研究,加速人类的科学进程。

下面详细介绍分布式计算的优点:

1) 高性价比。分布式计算往往可以采用价格低廉的计算机。今天的个人计算机比早期的大型计算机具有更出众的计算能力,体积和价格不断下降。再加上 Internet 连接越来越普及且价格低廉,大量互连计算机为分布式计算创建了一个理想环境。因此,分布式计算相对传统的小型机和大型机等计算具有更好的性价比。

2) 资源共享。分布式计算体系反映了计算结构的现代组织形式。每个组织在面向网络提供共享资源的同时,独立维护本地组织内的计算机和资源。采用分布式计算,组织可以非常有效地汇集资源。

3) 可伸缩性。在单机计算中,可用资源受限于单台计算机的能力。相比而言,分布式计算有良好的伸缩性,对资源需求的增加可通过提供额外资源有效解决。例如,将更多支持电子邮件等类似服务的计算机增加到网络中,可满足对这类服务需求增长的需要。

4) 容错性。由于可以通过资源复制维持故障情形下的资源可用性,与单机计算相比,分布式计算提供了容错功能。例如,可将数据库备份复制维护到网络的不同系统上,以便在一个系统出现故障时,还有其他备份可以访问,避免服务瘫痪。尽管不可能构建一个能在故障面前提供完全可靠服务的分布式系统,但在涉及和实现系统时最大化系统的容错能力,是开发者的职责。

然而无论何种形式的计算,都有其利与弊的权衡。分布式计算发展至今,仍然有很多需要解决的问题。分布式计算最主要的缺点有:

1) 多点故障。分布式计算存在多点故障情形。由于设计多个计算机,且都依赖于网络通信,因此一台或多台计算机的故障,或一条或多条网络链路的故障,都会导致分布式系统出现问题。

2) 安全性低。分布式系统为非授权用户的攻击提供了更多机会。在集中式系统中,所有计算机和资源通常都只受一个管理者控制,而分布式系统的非集中式管理机制包括许多独立组织。分散式管理使安全策略的实现和增强变得更为困难;因此,分布式计算在安全攻击和非授权访问防护方面较为脆弱,并可能会非常不幸地影响到系统内的所有参与者。

1.1.3 经典的分布式计算项目

1. WWW

WWW 是到目前为止最大的一个分布式系统,WWW 是环球信息网(World Wide Web)的缩写,中文名字为“万维网”,“环球网”等,常简称为 Web。它是一个由许多互相链接的超文本组成的系统,通过互联网访问。在这个系统中,每个有用的事物,称为一个“资源”;并且由一个全局“统一资源标识符”(URI)标识;这些资源通过超文本传输协议(Hypertext Transfer Protocol,HTTP)传送给用户,而后者通过单击链接获得资源。万维网并不等同互联网,万维网只是互联网所能提供的服务之一,是靠着互联网运行的一项服务。

WWW 是建立在客户机/服务器模型之上的。WWW 是以超文本标注语言(标准通用标记语言下的一个应用)与超文本传输协议为基础的,能够提供面向 Internet 服务的、一致的用户界面的信息浏览系统。其中,WWW 服务器采用超文本链路链接信息页,这些信息

页既可放置在同一主机上,也可放置在不同地理位置的主机上;而链路由统一资源定位器(URL)维持,WWW客户端软件(即WWW浏览器)负责信息显示与向服务器发送请求。

2. SETI@home

SETI@home(Search for Extra Terrestrial Intelligence at Home,寻找外星人),是一个利用全球联网的计算机共同搜寻地外文明的项目,本质上它是一个由互联网上的多个计算机组成的处理天文数据的分布式计算系统。SETI@home是由美国加州大学伯克利分校的空间科学实验室开发的一个项目,它试图通过分析阿雷西博射电望远镜采集的无线电信号,搜寻能够证实地外智能生物存在的证据,该项目参考网站为 <http://setiathome.berkeley.edu/index.php>。

SETI@home是目前因特网上参加人数最多的分布式计算项目。SETI@home程序在用户的个人计算机上,通常在屏幕保护模式下或后台模式运行。它利用的是多余的处理器资源,不影响用户正常使用计算机。SETI@home项目自1999年5月17日开始正式运行。至2004年5月,累积进行了近 5×10^{21} 次浮点运算,处理了超过13亿个数据单元。截至2005年关闭之前,已经吸引了543万用户,这些用户的计算机累积工作243万年,分析了大量积压数据,但是项目没有发现外星文明的直接证据。SETI@home是迄今为止最成功的分布式计算试验项目。

3. BOINC

BOINC(Berkeley Open Infrastructure for Network Computing,伯克利开放式网络计算平台)是美国加利福尼亚大学伯克利分校于2003年开发的一个利用互联网计算机资源进行分布式计算的软件平台。BOINC最早是为了支持SETI@home项目而开发的,之后逐渐成了最为主流的分布式计算平台,为众多的数学、物理学、化学、生命科学、地球科学等学科的项目所使用。如图1-1所示,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,服务端一般由数据库服务器、数据服务器、调度服务器和WEB门户组成;客户端部署于志愿者的计算机,一般由分布在网络上的多个用户计算机组成,负责完成服务端分发的计算任务。客户端与服务端之间通过标准的互联网协议进行通信,实现分布式计算。

BOINC是当前最为流行的分布式计算平台,提供了统一的前端和后端架构,一方面大为简化了分布式计算项目的开发,另一方面,对参加分布式计算的志愿者来说,参与多个项目的难度也大为降低。目前,已经有超过50个的分布式计算项目基于BOINC平台,BOINC平台上的主流项目包括有SETI@home、Einstein@Home、World Community Grid等。更详细的介绍可参考该项目网站 <http://boinc.ssl.berkeley.edu/>。

4. 其他的分布式计算项目

除了以上3个最经典的分布式系统外,还有很多其他的分布式计算项目^[3],它们通过分布式计算构建分布式系统和实现特定项目目标。

- Climateprediction.net: 模拟百年以来全球气象变化,并计算未来地球气象,以对付未来可能遭遇的灾变性天气。
- Quake-Catcher Network(捕震网): 借由日渐普及的笔记本计算机中内置的加速度计,以及一个简易的小型USB微机电强震仪(传感器),创建一个大的强震观测网。

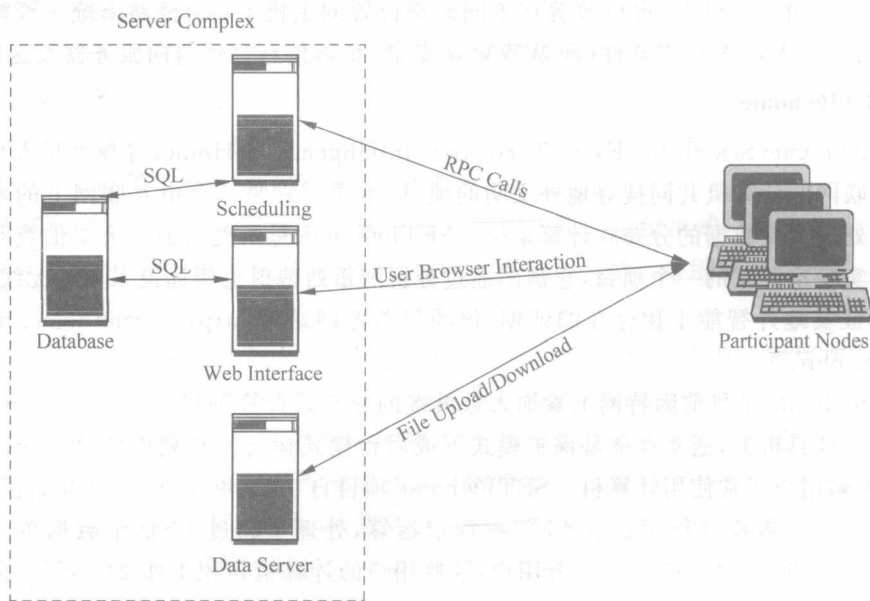


图 1-1 BOINC 的体系结构

可用于地震的实时警报或防灾、减灾等相关的应用上。

- World Community Grid(世界社区网格): 帮助查找人类疾病的治疗方法, 和改善人类生活的相关公益研究, 包括艾滋病、癌症、流感病毒等疾病及水资源复育、太阳能技术、水稻品种的研究等。
- Einstein@Home: 于 2005 年(世界物理年)开始的项目, 旨在找出脉冲星的引力波, 验证爱因斯坦的相对论预测。
- FightAIDS@home: 研究艾滋病的生理原理和相关药物。
- Folding@home: 了解蛋白质折叠、聚合以及相关疾病。
- GIMPS: 寻找新的梅森素数。
- Distributed.net: 成立于 1997 年, 是互联网的第一个通用分布式计算项目。2002 年 10 月 7 日, 以破解加密术著称的 Distributed.net 宣布, 在经过全球 33.1 万名电脑高手共同参与, 苦心研究了 1726 天, 于 2002 年 9 月 25 日破解了以研究加密算法而著称的美国 RSA 数据安全实验室开发的 64 位密钥——RC5-64 密钥。目前正在进行的是 RC5-72 密钥。

1.2 分布式计算模式

随着互联网与移动互联网应用的快速发展, 出现很多新的分布式计算模式与范型, 如云计算、雾计算、大数据计算等。这些新型计算模式或新技术, 本质上是分布式计算的发展和延伸。与分布式计算相关的计算模式有很多, 下面讨论一下单机计算、并行计算、网络计算、对等计算、集群计算、网格计算、云计算、雾计算和边缘计算等, 以便更好地区分和理解各种分布式计算模式的概念。