第四届“梦想加速”项目竞赛题

**项目A：**

1. **项目名称：**Cache 日志解析
2. **描述：**Cache日志的输出格式是可配置的，输出的内容千变万化；但要快速地反向解析出内容中的待定量，却是项有点挑战的事。请使用Java（或C/C++）实现一种Cache日志的快速解析方法。已经部分输出格式如下(以%打的都是待定量)：

①、#Fields: %>a %ui %un [%tl] "%rm %ru HTTP/%rv" %Hs %<st "%{Referer}>h" "%{User-Agent}>h" %tr %<sb "%{Content-Length}<h" "%{Range}>h" %Ss %Sh %rb %ts.%03tu %<a %hc\_tag %lp Via:"%{Via}>h"

②、#Fields: %>a %ui %un [%tl] %rm "%ru" HTTP/%rv %Hs %<st "%{Referer}>h" "%{User-Agent}>h" %tr %<sb "%{Content-Length}<h" "%{Range}>h" %Ss %Sh %rb %<a %{htc1s}>h %lp Via:"%{Via}>h"

③、#Fields: %>a [%tl] "%rm %ru HTTP/%rv" %Hs %<st "%{Referer}>h"

④、#Fields: %unix-time %>a - - %date %time %code %<st %rd %tr %rm %url "%{User-Agent}>h" "%{Referer}>h" %{Cookie}>h "%url"

⑤、#Fields: %ts.%03tu %6tr %>a %Ss/%03Hs %<st %rm %ru %un %Sh/%<A %mt

1. **难度：**易
2. **总体要求及考核标准（总和100分）**
3. 基本点：

①、能解析出提供的样例日志文件(gz格式）中的各待定量（文件内容的头几行中标明了日志格式），输出解析结果并能说明其正确性。（50分）

②、输出各项指标数据，如每秒解析的行数，每秒解析的总字节数等。（20分）

1. 拓展点：

①、解析指标达到每秒50MB。（10分）

②、实现某种字串匹配算法。（10分）

1. **提交材料：**项目报告（Word或PPT）、制作开发的应用程序、程序源代码
2. **提交至邮箱：**shenzx@chinanetcenter.com、1104316817@qq.com
3. **收获：**
4. 编程能力提升。
5. 学习抽象化思维和理论。
6. 了解字串匹配算法知识。
7. **涉及技术知识：**
8. 可参考字串匹配算法知识。
9. Java/C++/C编程能力。
10. 锻炼抽象思维。

**项目B：**

1. **项目名称：**统计量存储格式转换
2. **描述：**网卡上的流量值是不断累加的，直到溢出，如果要获取每分钟的流量值需要两次获取网卡上的流量，并相减。类此的情况很多，请使用C++（或Java）实现统计量存储格式的转换，输入为不断累加的标量序列，输出为按时间的增量标量序列。要求具有较高的通用性和性能。可阅读相关的开源代码学习, 但最终代码需要自己编写, 不能使用第三方库。
3. **难度：**易
4. **总体要求及考核标准（总和100分）**
5. 基本点：

①、支持统计标量格式转换。（20分）

②、支持大量标量序列同时转换。（20分）

1. 拓展点：

①、支持累计标量序列 [输入端] 不连续。（30分）

②、支持累计标量序列 [输入端] 错序。（30分）

1. **提交材料：**项目报告（Word或PPT）、制作开发的应用程序、程序源代码
2. **提交至邮箱：**linsy@chinanetcenter.com、1104316817@qq.com
3. **收获：**
4. 标准文档阅读能力。
5. 编程能力的提升。
6. **涉及技术知识：**
7. C++（Java）编程能力。

**项目C：**

1. **项目名称：**分布式通讯框架
2. **描述：**使用Java实现server/client的通讯框架，server/client均可多点部署，client可向一个或多个server发送请求，server可向所有client发送广播。。
3. **难度：**中
4. **总体要求及考核标准（总和100分）**
5. 基本点：

①、client连接到server，实现请求和响应。（10分）

②、新server可连接到已有server，实现server集群，server间可做请求和响应。（10分）

③、client可向指定server或所有server发送请求，并得到预期各数的响应。（20分）

1. 拓展点：

①、有新server加入时，client可感知，并在当前连接的server网络异常时，可自动重连其他的server。（20分）

②、各server可以向其他所有的server和client发送消息广播。（20分）

③、实现消息订阅，不同client可向server集群订阅不同主题的消息，server根据主题准确向指定的client推送数据。（20分）

1. **提交材料：**项目报告（Word或PPT）、制作开发的应用程序、程序源代码
2. **提交至邮箱：**liuz@chinanetcenter.com、1104316817@qq.com
3. **收获：**
4. 网络通讯原理理解。
5. 编程能力提升。
6. 分布式、可扩展理解。
7. **涉及技术知识：**
8. 网络通信原理。
9. Java编程能力。
10. 分布式系统原理。

**项目D：**

1. **项目名称：**基于 IPv6 地址的复杂运算
2. **描述：**随着 IPv4 的枯竭，IPv6 马上面临大规模商用，地址长度从 IPv4 的 32 位变成了 128 位，地址数量庞大无比，过去的一些算法已不再适用。请在 Linux 下，用 C/C++ 设计一套数据结构、算法，满足 IPv4 和 IPv6 的基本运算。
3. **难度：**中
4. **总体要求及考核标准（总和100分）**
5. 基本点：

①、识别常见的 IPv6、IPv4 地址格式。例如：FE80:: 、FE80::/32 、210.34.0.0 、210.34.0.0/16。（10分）

②、从字典库中查询 IPv6、IPv4 单个地址、地址段的属性。字典库的格式（为地址段）：FE80::/32 属性1； FE81::/64 属性2； 210.34.0.0/16 属性3。（15分）

③、2 个字典库的差异比对。字典库的格式，同上。（15分）

④、2 个字典库做减，即字典库1减去字典库2中属性相同的地址段。字典库的格式，同上。（10分）

⑤、2 个字典库做加，即字典库1覆盖字典库2中属性不同的地址段，包括字典库2中不存在的地址段。字典库的格式，同上。（10分）

1. 拓展点：

①、附带详细设计文档。（10分）

②、合理的模块化编码。（15分）

③、良好的代码风格、质量。（15分）

1. **提交材料：**项目报告（Word或PPT）、制作开发的应用程序、程序源代码
2. **提交至邮箱：**qiuwk@chinanetcenter.com、1104316817@qq.com
3. **收获：**
4. 增加 Linux 下 C/C++开发能力。
5. 增强文档编写能力。
6. 引导理论知识转化为实际成果。
7. **涉及技术知识：**
8. Linux 基础知识。
9. 常用的数据结构。
10. 模块化设计理念。

**项目E：**

1. **项目名称：**Linux内核定时器的应用层实现优化
2. **描述：**借鉴内核定时器的优点，应用层实现定时器。
3. **难度：**难
4. **总体要求及考核标准（总和100分）**
5. 基本点：

①、Linux内核定时器的实现研究。（20分）

②、Linux应用层定时器的实现研究。（20分）

③、Linux内核定时器的应用层实现。（20分）

1. 拓展点：

①、定时器的精度评价及其改进。（20分）

②、定时器的负载评价及其改进。（20分）

1. **提交材料：**PPT、应用程序、程序源代码
2. **提交至邮箱：**chenwensheng@chinanetcenter.com、1104316817@qq.com
3. **收获：**
4. Linux 内核源码。
5. Linux 内核定时器的各种实现原理，方法。
6. Linux应用层各种定时器的实现方法。
7. Linux程序性能改进调优方法。

**项目F：**

1. **项目名称：**基于DPDK的报文处理模块
2. **描述：**DPDK：Intel® Data Plane Development Kit，是X86平台报文快速处理的库和驱动的集合，模仿于OCTEON专用网络处理器，实现的一套网络处理包KIT，定位在不需要专用网络处理器，但通用网络处理器对数据处理性能又不能满足需求的场景。
3. **难度：**难
4. **总体要求及考核标准（总和100分）**
5. 基本点：

①、结合DPDK实现ARP基本功能。（50分）

②、实现对接收到的ARP请求报文，根据配置的匹配条件，以指定MAC地址进行响应。（30分）

（i）条件匹配域包括APR报文源/目的IP地址，源MAC地址；

（ii）指定的MAC地址包括任意指定的MAC地址或已知直连网络内的主机IP地址,模块自动获取到的MAC地址；

（iii）具备可视化管理、配置接口。

1. 拓展点：

①、TCP/IP协议栈移植。（15分）

②、NAT功能实现。（5分）

1. **提交材料：**项目报告（Word或PPT）、程序及源代码
2. **提交至邮箱：**zhoujz@chinanetcenter.com、1104316817@qq.com
3. **收获：**
4. 网络数据抓包，协议分析技能。
5. 较为全面立体的网络体系结构学习。
6. 了解网卡硬件部分关键编程模型和特性，接地气的网络知识。
7. **涉及技术知识：**
8. 熟悉Linux开发环境，了解内核驱动开发。
9. 较强C/C++编程能力。
10. 网络知识（参考TCP/IP详解卷1)。

**项目G：**

1. **项目名称：**基于Nginx的内容优化插件
2. **描述：**移动互联网的兴起，导致多样的网络环境、多种类型的终端设备和多样的应用，希望根据这些参数进行相应的内容适配，例如：同样访问一张图片，不同网络环境或不同分辨率的终端设备或不同的应用场景，可以返回不同分辨率的图片；视频等也是类似。
3. **难度：**难
4. **总体要求及考核标准（总和100分）**
5. 基本点：

①、做成Nginx的一个插件，分感知和适配两块。感知网络环境、终端设备和应用。（30分）

②、支持常用图片格式的自适配。（20分）

③、支持常用视频格式的自适配。（20分）

1. 拓展点：

①、支持HTML的自适配。（30分）

1. **提交材料：**项目报告（Word或PPT）、制作开发的应用程序、程序源代码
2. **提交至邮箱：**dongsj@chinanetcenter.com、1104316817@qq.com
3. **收获：**
4. 基于Nginx的插件编程。
5. 常用图片格式的适配转换。
6. 常用视频格式的适配转换。
7. HTML内容适配。
8. **涉及技术知识：**
9. Nginx插件。
10. 常用图片、视频格式。
11. HTML内容适配技术。

**项目H：**

1. **项目名称：**网络点餐，优选最近的厨师
2. **描述：**

**背景说明：**A城有6个区，每个区拥有30个乡镇，每个乡镇拥有50个村，每个村拥有100个小组；A城所有区的名称各不相同，所有乡镇的名称各不相同，所有村的名称各不相同，所有小组的名称也各不相同；总有10万个厨师居住在A城，每个厨师都会做不超过20道菜。每个厨师会做的菜有可能相同也可能不同。每个厨师都有一个从业编号，从0-99999；最近全国都在大力发展互联网+, A城的市政府也不甘落后，提出想要建立一套网上点餐系统，方便A城的公民点餐；期望是A城的公民在网上系统点餐，点好以后会有厨师的助手将菜送到约定的地址。网上点餐系统的其中一个模块被层层转包给了某无证程序员小明。

**该模块具体要求：**

**功能要求：**

1. 能据送餐目的地优选离该地最近的厨师，提供相关厨师的报价和好评情况等信息；假定路人甲住在A城的B镇C村D小组，他想吃蛋炒饭，系统首先查找跟他同小组的厨师，如果同小组有厨师会做蛋炒饭，就会将同小组厨师的报价和好评信息提供给路人甲，让他去选择一个厨师下单；如果同组没有合适的厨师就找同村，同村没有就找同镇，同镇没有就找同区，同区没有就找本市，本市也没有就告知找不到；
2. 厨师会做的且愿意做的菜，价格经常发生变化，需要允许在系统上调整。

**性能要求：**

1. A城国际化大都市，常住人口超过2000万，每天还有大量游客涌入。高峰期点餐，可能会有每秒百万级别的查询量；
2. 修改单个厨师信息需要在2秒内完成；且高峰期修改不能影响查询性能；
3. 由于经过层层转包，建设经费有限，只能提供单台比较弱的服务器来支持这个优选的计算，基础配置信息为CPU: 主频2.0GHZ，物理核心2个，超线程后线程数最多4个；内存8G；硬盘500G。

**补充说明：**

1. 厨师信息包含：厨师的从业编号（00000-99999），所在的小组（字符串，长度不超过50字节），会做的菜的菜名（字符串，长度不超过50字节），每道菜的报价（从0-100000的整数），以及每道菜的评分（从0-100的整数）；
2. 城市行政区域划分信息：区名（字符串，不超过10个字节），乡镇名（字符串，不超过10个字节），村名（字符串，不超过10个字节），小组名（字符串，长度不超过50字节）；
3. 数据持久化存储方式不限，可以是文件也可以是数据库；允许将持久化数据先加载到内存然后对外提供服务，但是需要保证内存数据和持久化数据的一致性；
4. 不需要页面，只需要提供几个函数以及测试程序：

1）提供查询函数，根据送餐目的地所在的小组名（字符串），以及单个菜名，查找到最近的厨师的信息；

2）提供增，删，改函数，允许根据从业编号，修改厨师提供的菜名和价格；

3）测试程序包含功能和性能的测试；

4）如果有用到的话，提供加载函数，用于将持久化数据加载到内存中。

1. 编程语言不限。
2. **难度：**难
3. **总体要求及考核标准（总和100分）**
4. 基本点：

①、查询函数实现。（5分）

②、增删改函数实现。（10分）

③、测试程序实现。（5分）

④、能支持每秒百万以上查询请求。（10分）

⑤、修改单个厨师信息不超过2秒。（5分）

⑥、相关程序内存使用不超过4G。（10分）

⑦、在支持每秒100万以上的查询请求的同时，增删改单个厨师信息不超过2秒。（5分）

1. 拓展点：

①、支持每秒支持100万以上的查询请求的基础上，每多支持10万请求量加5分。（25分）

②、内存使用不超过4G的基础上，每减少400M，加2分。（10分）

③、支持每秒100万查询请求所用的CPU逻辑核心数越少越好，在4个CPU核心线程的基础上，每减少1个加5分。（15分）

1. **提交材料：**项目报告（Word或PPT）、制作开发的应用程序、程序源代码
2. **提交至邮箱：**caoc@chinanetcenter.com、1104316817@qq.com
3. **收获：**
4. 如何在合理控制内存使用量的基础上写出高性能的程序。
5. 如何合理解决冲突（增删改查的冲突），使之不影响性能。