Contribution ID: **11** Type: **分会报告**

高能物理实验数据跨域访问缓存系统

Tuesday, 4 July 2017 17:40 (20 minutes)

摘要：高能物理计算环境具有实验数据量巨大，实验数据全球共享，实验数据长期稳定存储三个主要特征。针对这三个特征，海量的高能物理实验数据需要跨域访问处理完成实验操作，跨域实验数据访问分析也成为高能物理实验中不可或缺的一个环节。针对全球化的海量实验数据共享，高能物理计算主要使用了网格技术进行跨域数据共享，如为欧洲大型强子对撞机实验设计了全球最大的网格系统WLCG（Worldwide LHC Computing Grid）。
网格技术将跨域站点的存储、计算资源共享形成一个具有巨大存储能力的分布式数据网络。网格系统具有十分庞大的规模和数据共享能力，但是也有这很大的局限性。首先网格系统中需要为实验作业预先分配计算、存储资源，使得系统资源难以充分利用。其次，系统中的数据文件需要全部传输至本地后计算分析，调度十分不灵活并且基于文件的传输中有大量无用事例，形成了大量的资源浪费。最后网格系统使用 GridFTP 协议进行跨域数据传输，GridFTP 协议复杂需要多端口配置，而且无法穿透防火墙，运行维护代价较高。
除网格计算系统外，目前高能物理计算环境中使用的跨域数据访问系统，主要有分布式 EOS 进行小范围站点间的数据共享，CVMFS(CERN Virtual Machine File System) 进行高能物理实验软件共享。分布式 EOS 采用 replica 策略进行数据文件跨域传输，不同客户端通过 IP 标示，被选择与最近服务器站点进行通信，若该服务端有相应数据文件则传输至客户端，若不存在服务器从其他服务器端拉取文件后再传输至客户端。CVMFS 设计目标为高能物理实验软件的跨域分发，基于 FUSE 将 web 目录以本地磁盘的方式挂载到本地，客户端只需要在第一次访问时下载必要的库，基于 HTTP 协议进行数据传输，它支持只读模式，在应用层设置了 cache 缓存加速 web 目录读速度。无论是分布式 EOS 或者网格技术都有自己固定的应用场景，受到应用和设计的限制，都无法实现基于事例级的实验数据跨域高速访问。
本文中设计了针对事例级高能物理实验数据的跨域访问缓存系统，系统中设计了本地缓存服务器进行跨站点数据缓存。物理学家进行实验作业分析时，不需要将整个 DST 文件下载到本地。将事例请求发送至缓存服务器后，缓存服务器向远程站点发送请求，之后以事例为级别进行 HTTP 多流传输至本地缓存，并返回至客户端。对客户端来说，所有操作都是在缓存服务器上进行，远程站点是透明化的。缓存服务器提供了按需访问、动态调度的新型高能物理数据跨域访问模式，系统访问及传输以事例问单位，大大的减少了资源浪费，提供了作业处理效率。同时缓存系统提供了统一数据管理、远程站点统一文件视图，为用户提供了本地化操作模式。缓存系统中设计了用户操作日志分析模块，以 syslog 模式抓取用户对于数据分析的记录，通过近期数据分析，实现数据预取来增强系统读性能。在整个缓存系统模块中应用了多进程并发处理机制，实现高效的用户消息处理模式和高性能的读写调度架构。系统中客户端与服务器端通信都采用了高能物理计算中通用的 XROOTD 架构，具有较强的普适性与通用性，更好的与高能物理实验分析作业相结合。
作为一种新型的跨域访问系统架构，有效的解决了传统基于文件处理的资源浪费和效率低下问题，同时缓存服务器将远程站点的数据以本地化的模式提供给用户，提供了便捷高效的数据处理模式。整个系统为高能物理跨域计算提供了新型的架构，在高能物理计算环境中具有较好的应用发展前景。

Primary author: Mr 徐, 琪 (中国科学院高能物理研究所)

Co-authors: Mr CHENG, yaodong (IHEP); Dr 陈, 刚 (中国科学院高能物理研究所); Dr 程, 耀东 (中国科学院高能物理研究所); Ms 王, 聪 (中国科学院高能物理研究所)

Presenter: Mr 徐, 琪 (中国科学院高能物理研究所)

Session Classification: 科学数据管理与信息化 I

Track Classification: 科学数据管理技术与系统