

CDAS: 一种高效的存储联盟数据访问系统

Wednesday, 17 July 2019 15:15 (25 minutes)

高能物理实验通常会生成海量实验数据, 随着越来越多的实验装置投入运行, 实验数据的增长也越来越快, 这些数据在分布式站点上存储, 通过分布式数据管理系统组织和管理。随着资源的不断优化, 针对新的应用场景和技术需求, 传统的分布式数据管理系统面临着很多挑战, 比如全局命名空间的快速构建、异地站点数据的高效访问和存储等。针对这些问题, 本文提出了一种高效的存储联盟数据访问系统 CDAS。

CDAS 作为分布式站点间的数据缓存和访问系统, 采用边缘计算的思想, 允许在远程站点动态地构建数据联盟内的全局命名空间, 并且支持灵活的数据缓存与同步、数据去重与压缩存储等机制, 在联盟中实现了全局命名空间的快速构建、异地数据的高效访问以及存储空间的高利用率。

CDAS 由四部分组成, 分别是元数据服务器、缓存服务器、存储优化引擎以及数据访问接口, 四个部分的具体描述如下:

元数据服务器将拥有 XROOTD、HTTP 或 S3 等协议的分布式站点目录进行聚合, 并记录文件的各类信息, 按需实时地为用户提供联盟内全局命名空间, 使用户无需考虑数据的实际存储位置, 同时隐藏站点协议的多样性, 最终实现全局命名空间的无差异视图;

缓存服务器用于对客户端文件实体以及文件元数据信息的缓存和同步, 文件实体的缓存和同步以数据块为单位, 对于用户本地缓存中已存在的数据块不再进行重复的传输和存储, 同时支持灵活的本地站点新文件的同步回传策略。通过缓存和同步机制, 最终加快联盟内的数据访问速度和全局命名空间构建速度;

存储优化引擎基于数据重删机制和数据压缩机制, 目的是优化数据实际存储所需空间。其中, 数据重删机制用于降低数据块之间的数据冗余, 采用在线定长块去重的策略, 将单个数据块与唯一编码相对应, 在数据块传输之前预先传输数据块编码至目标站点, 存在相同编码则认为存在相同数据块, 此时不再进行数据块传输, 确保目标站点已存储的数据块不再重复传输至该站点, 规避了同一数据块在同一站点冗余存储所造成的时间和资源浪费; 同时, 数据压缩机制用于降低数据块内部的数据冗余, 保证传输到站点的新数据块在压缩后存储, 有效降低新数据块所占用的存储空间, 进一步提高联盟内站点存储资源利用率。

数据访问接口提供命令行和 FUSE 客户端, 保留用户习惯并且屏蔽联盟内数据访问流程的复杂性与异构性, 提供本地化的统一访接口, 最终实现分布式站点间透明的数据共享。

本文中的存储联盟由分布在成都、稻城和北京的存储站点组成, 站点中存储了 LHAASO 实验产生的海量数据。基于 LHAASO 实验数据的测试表明, CDAS 能够将分布在成都、稻城和北京的存储数据聚合成统一的全局命名空间, 缓存机制能够将数据访问性能提高 10 倍以上, 存储优化引擎使得原始实验数据的存储空间占有率降低 50% 左右。

Primary authors: Dr CHEN, Gang (IHEP); XU, Qi (UCSA, IHEP); Dr CHENG, Yaodong (IHEP); 符, 世园 (中国科学院高能物理研究所)

Presenter: 符, 世园 (中国科学院高能物理研究所)

Session Classification: 核电子学与探测技术 II

Track Classification: 科学数据存储与管理技术 II