

Бурмако Ю.Н.

Белорусский государственный экономический университет, г. Минск

Использование grid-структур в обработке данных с LHC. Проект Worldwide LHC Computing Grid

Большой адронный коллайдер (Large Hadron Collider, LHC), построенный Европейским центром ядерных исследований (CERN) недалеко от Женевы, является крупнейшим в мире ускорителем элементарных частиц. Ежегодно в рамках этого проекта будет генерироваться около 15 Пбайт данных. После запуска ускорителя в 2009 г. более чем 7 тысячам ученым из 500 исследовательских центров со всего мира потребуется обеспечить доступ к экспериментальным данным. Анализ экспериментальных данных, собираемых с детекторов LHC – ATLAS, CMS, ALICE и LHCb, включая их сравнение с моделированными данными, потребует огромных вычислительных мощностей. Для этого была разработана модель хранения и обработки данных на базе grid-технологий – Worldwide LHC Computing Grid. Такая модель имеет несколько ключевых преимуществ перед традиционными моделями:

- управление затратами на поддержку и обновление необходимых ресурсов для таких вычислительных задач намного легче в распределенном окружении;
- надежность распределенной децентрализованной системы значительно выше. Репликация данных (множество копий) и автоматическое распределение задач по ресурсам гарантирует баланс нагрузки и облегчает доступ к данным всем пользователям. Охват всех часовых поясов также облегчает круглосуточный мониторинг и поддержку.

Задача проекта WLCG – создать и поддерживать инфраструктуру для хранения и анализа данных, которой будет пользоваться все сообщество ученых, участвующих в LHC. В рамках проекта была разработана четырехуровневая модель распределенного хранения и обработки данных с LHC. Tier-0 (Уровень-0) в этой модели – CERN, где будет производиться первичная реконструкция событий, калибровка и хранение копий полных баз данных. Поток данных с детекторов выдает около 300 ГБ/с, затем отфильтровываются “интересные события”, в результате чего “сырые данные” составляют поток около 300 МБ/с. После первоначальной обработки данные будут размещаться в центрах Tier-1 (крупных региональных центрах). Здесь будет производиться полная реконструкция событий, хранение актуальных баз данных по событиям, создание и хранение наборов анализируемых событий. Центры Tier-1 предоставляют экспериментальные данные центрам Tier-2 (небольшим региональным центрам), где будет производиться репликация и хранение наборов анализируемых событий. Ученые получают доступ к данным и средствам их обработки благодаря третьему уровню вычислительных ресурсов Tier-3, который может состоять как из локальных кластеров, так и индивидуальных компьютеров, которые, тем не менее, могут являться частью LCG.

На сегодняшний день к проекту подключились около 50 стран с суммарными вычислительными ресурсами около 140 тыс. процессоров и ресурсами хранения около 110 петабайт. Несмотря на внушительные цифры этого еще недостаточно. Участвующие научные организации прилагают все усилия, чтобы довести параметры локальных вычислительных ресурсов, включенных в LCG, до требуемых.

Список литературы

1. Worldwide LHC Computing Grid [Electronic resource]. – Electronic data. – CERN, 2009. – Mode access: <http://lcg.web.cern.ch/LCG/>.
2. Кореньков В., Ужинский А. Архитектура сервиса передачи данных в grid / В. Кореньков, А. Ужинский // Открытые системы [Электронный ресурс]. – 2008. – №02 – Режим доступа: <http://sqlmag.ru/os/2008/02/4926522/>.