Р10-2004-167

Д. А. Архипкин, Ю. Р. Зулькарнеева

ОРГАНИЗАЦИЯ РАБОЧЕЙ ВЫЧИСЛИТЕЛЬНОЙ СРЕДЫ ЭКСПЕРИМЕНТА **STAR** В **ОИЯИ**. РЕЗУЛЬТАТЫ И ОПЫТ РАБОТЫ ПЕРВЫХ ДВУХ ЛЕТ

Архипкин Д. А., Зулькарнеева Ю. Р. Р10-2004-167

Организация рабочей вычислительной среды эксперимента STAR в ОИЯИ. Результаты и опыт работы первых двух лет

Приведено описание организации рабочей вычислительной среды физического эксперимента STAR в ОИЯИ. Данный подход, основанный на использовании распределенной файловой системы ASF, реализован на базе вычислительного мини-кластера в отделе НЭОСТАР ЛФЧ. Приводится описание кластера, принцип его работы и схема построения, примеры выполняемых расчетов. Результаты работ, выполненных на этом кластере, указывают на широкие возможности концепции распределенных вычислений в целях продуктивного и своевременного участия в обработке, анализе и моделировании экспериментального материала.

Работа выполнена в Лаборатории физики частиц ОИЯИ.

Сообщение Объединенного института ядерных исследований. Дубна, 2004

Arkhipkin D. A., Zulkarneeva Yu. R. P10-2004-167

Organization of the STAR Experiment Software Framework at JINR. Results and Experience from the First Two Years of Work

The organization of STAR experiment software framework at JINR is described. The approach being based on the distributed file system ASF was implemented at the NEOSTAR minicluster at LPP, JINR. An operation principle of the cluster as well as its work description and samples of the performed analysis are also given. The results of the NEOSTAR minicluster performance have demonstrated broad facilities of the distributed computing concept to be employed in experimental data analysis and high-energy physics modeling.

The investigation has been performed at the Laboratory of Particle Physics, JINR.

Communication of the Joint Institute for Nuclear Research. Dubna, 2004

ВВЕДЕНИЕ

На протяжении ряда лет физики ОИЯИ принимают активное участие в сооружении детектора STAR в BNL и проведении с его помощью экспериментов на коллайдере RHIC. Однако для полноценного и эффективного участия в обработке и анализе экспериментальных данных необходимо располагать реальными возможностями для выполнения хотя бы части из отмеченных выше задач, находясь непосредственно в Дубне. Это означает, что такой локальный дубненский кластер должен обеспечивать следующие возможности:

– доступ к базе экспериментальных данных коллаборации STAR и его вычислительной среде;

– обработку экспериментальных данных и проведение расчетов по их моделированию в соответствии со стандартами, принятыми в коллаборации;

– представление результатов этой обработки в соответствии со стандартами коллаборации, а также осуществление других действий;

– обеспечение доступа в центральный репозиторий исходных текстов коллаборации STAR.

Для реализации указанных возможностей был создан кластер ЛФЧ STAR, удовлетворяющий всем этим требованиям. Ниже дается его краткое описание, принцип работы и схема построения.

1. КРАТКИЕ СВЕДЕНИЯ О РАСПРЕДЕЛЕННОЙ ФАЙЛОВОЙ СИСТЕМЕ AFS (ANDREW FILE SYSTEM)

Стандартный подход к организации работы с программным окружением коллаборации заключается в копировании необходимых файлов и последующей синхронизации/обновлении этих файлов различными способами (rsync, другие программы). Данный подход дает возможность создания независимой копии программного окружения (это хорошо), но также он накладывает существенные ограничения на коллективную работу над файлами и исходными текстами программного обеспечения (это плохо). В частности, приходится искать способ «обратной синхронизации», когда изменения, внесенные в такую локальную копию, необходимо отразить в центральном хранилище, что влечет за собой большие трудозатраты на организацию и поддержку такой синхронизации. С другой стороны, непосредственная работа с центральным

хранилищем влечет за собой замедление работы ввиду низкой скорости связи между удаленным кластером и центральным хранилищем.

При организации коллективного доступа физиков ЛФЧ к вычислительной среде коллаборации STAR применяется подход, основанный на использовании широко известной распределенной файловой системы AFS [1]. AFS позволяет найти компромисс между преимуществами локальной копии всей системы и удаленной работой с центральным хранилищем.

Концепция AFS. AFS облегчает работу над файлами независимо от того, где эти файлы расположены. Пользователи AFS не обязаны знать какой сервер хранит данный файл, и администраторы могут перемещать данные с машины на машину, не прерывая пользовательский доступ. Пользователи работают с файлами, используя один и тот же путь, а AFS находит нужные файлы автоматически, так же как это происходит на локальной файловой системе одиночного компьютера. Хотя AFS и предоставляет удобный доступ к файлам, это не создает проблем с безопасностью доступа. В AFS используется продуманная схема защиты данных.

Клиент-сервер вычисления. AFS использует модель клиент-сервер. В модели такого вида есть два типа машин. Серверы хранят данные и выполняют различные операции для клиентских машин. Клиентские машины выполняют вычисления для пользователей и предоставляют доступ к данным с серверных машин. Некоторые компьютеры работают одновременно как клиентские и серверные машины.

Распределенная файловая система. AFS — это распределенная файловая система, которая объединяет файловые системы нескольких серверных машин, предоставляя легкий доступ к файлам, хранящимся на удаленном файл-сервере, как к локальным файлам. Распределенная файловая система имеет два основных преимущества перед обычной централизованной файловой системой.

1. Повышенная доступность. Копия популярного файла, например исполняемый модуль приложения, может храниться на нескольких серверах одновременно. Выход из строя одного сервера или даже нескольких не приведет к потере доступа к данному файлу. Вместо этого запросы пользователей будут переназначены доступным серверам. В случае централизованной файловой системы потеря связи с центральным файл-сервером означает прекращение работы всей системы.

2. Повышенная эффективность. В распределенной файловой системе нагрузка равномерно распределена между несколькими небольшими серверами в отличие от большого (и обычно более дорогого) файл-сервера централизованной системы.

Кэширование данных (буферизация). Использование файлов системы AFS возможно только при работе с клиентской AFS машины. Когда запрашивается файл, кэш-менеджер (часть AFS) запрашивает этот файл с подхо-

дящего AFS-сервера и сохраняет (кэширует) локальную копию этого файла на локальном диске клиентской машины. Приложения клиентской машины используют данную локальную (кэшированную) копию запрошенного файла. Это увеличивает общую скорость работы, потому что использовать файл локально гораздо быстрее, чем пересылать файлы по сети при каждом запросе.

Так как приложения работают с локальной копией, то никаких изменений в центральном хранилище не произойдет, пока файл не будет закрыт. В этом случае кэш-менеджер передаст все изменения назад файл-серверу (серверам), где они заменят соответствующие части первоначального файла.

Если доступ к файл-серверу прекращается по каким-то причинам (перебои в работе сети и т. д.), то можно продолжать работу с локальным файлом, но никакие изменения не будут сохранены, пока сервер не возобновит свою работу.

Итак, все файлы коллаборации (исходные тексты и исполняемые модули) находятся на серверах AFS, а файлы, запрашиваемые при работе удаленного кластера, сохраняются в кэше локальных клиентских компьютеров для ускорения доступа.

Таким образом, решены все проблемы синхронизации как прямой, так и обратной (работа ведется напрямую с центральным хранилищем), и проблема медленного доступа к центральному серверу (промежуточная буферизация внутренними средствами файловой системы).

2. УСТАНОВКА ПРОГРАММНОГО ОКРУЖЕНИЯ ЭКСПЕРИМЕНТА STAR С ПОМОЩЬЮ AFS НА ВЫЧИСЛИТЕЛЬНОМ КЛАСТЕРЕ ЛФЧ STAR

Набор пакетов программ, необходимый для работы с экспериментальными данными STAR и/или их моделирования, включает в себя:

- RedHat linux 8.0 (*www.redhat.com*);
- OpenAFS client (*www.openafs.org*);
- CVS client (*www.cvshome.org*) [2];
- MySQL server (*www.mysql.com*) [3].

Рекомендуемое средство распределения нагрузки — Sun Grid Engine (SGE) (*http://gridengine.sunsource.net*).

Все компоненты общедоступны и не требуют приобретения лицензий для использования.

Установка программного окружения с помощью AFS-client. Эта операция производится следующим образом. Необходимо скопировать файлы */afs/rhic.bnl.gov/star/group/templates/cshrc* и */afs/rhic.bnl.gov/star/group /templates/login* к себе в домашнюю директорию как *.cshrc* и *.login*:

% cp */afs/rhic.bnl.gov/star/group/templates/cshrc* ∼*/.cshrc*,

% cp */afs/rhic.bnl.gov/star/group/templates/login* ∼*/.login*.

При следующем входе пользователя в систему переменные окружения PATH, MANPATH, LD_LIBRARY_PATH будут включать в себя надлежащую информацию STAR, а также появятся несколько специфичных для STAR переменных окружения (сокращенный список):

– ROOTSYS — системная директория ROOT;

– STAR_LEVEL — версия программного обеспечения *dev/new/pro/old*;

– STAR_VERSION — версия библиотеки *SL98d/SL98c/SL98b/SL98a*;

– STAR_SYS — архитектура (*i386_linux2*, *sun4x_56*, ...);

– STAR_LIB — путь к текущим библиотекам STAR library;

– STAR_BIN — путь к исполняемым программам (StAF, geant, ...);

– CVSROOT — репозиторий STAR CVS;

– CERN_ROOT — путь к текущей версии CERN library, используемой в STAR.

Далее необходимо создать ссылку с */opt/star на /afs/rhic.bnl.gov/@sys/opt /star*.

Нужно учесть, что трансляция этого пути зависит от того, как AFS воспримет *'@sys'* на вашей клиентской машине. Для того чтобы не произошло ошибки, необходимо установить параметр sysname (настраивается в */etc/sysconfig/afs*). Для текущего стандарта STAR это установка RedHat 8:

AFS_POST_INIT="/usr/bin/fs sysname -newsys i386_redhat80".

Рекомендуемый размер кэша (буфера) AFS 2 Гбайта. При меньшем размере буфера наблюдаются периодические замедления работы в связи с частым обновлением данных из центрального хранилища STAR.

Настройка CVS-client. Установка CVS client ничем не отличается от установки стандартного клиента CVS, поэтому здесь не рассматривается. Необходимая для работы с репозиторием STAR переменная окружения CVSROOT устанавливается автоматически (см. выше).

Зеркалирование базы данных STAR (MySQL mirroring [4]). Для обеспечения равномерной нагрузки на центральный сервер баз данных STAR и существенного ускорения доступа к базе данных хорошей практикой является создание зеркалированной офф-лайн-базы для каждой рабочей группы, работающей удаленно.

Зеркалирование базы данных STAR MySQL достигается стандартными средствами MySQL. Для создания зеркала следует обратиться к текущему эксперту по базам данных коллаборации STAR (на 16.10.2004 это Майкл де Филлипс). Следует напомнить, что на каждой клиентской машине, входящей в кластер, необходимо указать, где и как сконфигурирован локальный сервер баз данных (зеркало). Для этого необходимо установить переменную окружения $STDB_SERVERS, в которой указать путь к конфигурационному XML-файлу следующего формата:

#
dbServers.xml
#
<StDbServer>
<server>*your server name*</server>
<host>*your_server_name.jinr.ru*</host>
<port>3306</port>
<socket>*/tmp/mysql.sock*</socket>
<databases>Calibrations</databases>
</StDbServer>

(параметры подставляются исходя из деталей установки сервера, произведенной экспертом STAR).

После этого все MySQL-запросы, относящиеся к калибровке, будут переадресованы на локальный сервер. Более подробное описание можно найти здесь (AFS должен быть уже настроен): *$STAR/StDb/servers/README*, где $STAR — это переменная окружения, указывающая на программную среду STAR.

Технические требования к скорости работы сети, необходимые для нормального функционирования системы. Для полноценной работы с программным обеспечением STAR необходимо не менее 1,5 Мбайт/с между *star.bnl.gov* и *jinr.ru*. Меньшая скорость связи приведет к недопустимо большим задержкам в работе с программным обеспечением.

Необходимые параметры компьютеров локального кластера. Для задач разработки программного обеспечения хватит и самого слабого компьютера. Для обработки данных необходимо иметь как минимум 256 Мбайт оперативной памяти, а для проведения моделирования Монте-Карло как минимум 512 Мбайт.

После выполнения всех вышеописанных действий пользователь получает возможность полноценного использования программного окружения STAR на локальной машине. Становятся доступными для использования следующие операции.

• Создание и компилирование программ с использованием **любой** версии программного обеспечения STAR.

• Проведение полноценных MC-симуляций в среде STARsim (бывшая StAF).

• Использование версии ROOT, адаптированной под нужды эксперимента STAR.

• Прямой доступ к центральному CVS-репозиторию коллаборации STAR (если есть пароль доступа).

• Ускоренная работа с офф-лайн-базой данных STAR, содержащей **всю** актуальную на данный момент информацию.

Уместно указать на преимущества *клиентского* использования AFS перед другими средствами зеркалирования.

• Все заботы по компиляции и обновлению версий программ STAR выполняет коллаборация, нет необходимости отслеживать изменения и вносить в локальную копию.

• Все изменения доступны в режиме реального времени (нет задержек при внесении изменений). AFS сама отслеживает обновление файлов и загружает измененные версии строго по необходимости.

• Обеспечен прямой доступ в хранилище исходных текстов CVS. Нет необходимости в дополнительной синхронизации локального хранилища.

• Экономия дискового пространства: на диске хранится только актуальная и используемая в данный момент информация, не надо хранить копии всех версий программного обеспечения со дня образования коллаборации.

Из более чем 50 исследовательских групп, образующих коллаборацию STAR, эта технология успешно опробована и применяется в следующих институтах — членах коллаборации STAR:

• Wayne State University, USA;
• Parallel Distributed Systems Facility, NERSC Systems, USA;
• Indiana University, USA;
• Institute of Physics, Czech Republic;
• Instituto de Fisica, San Paulo, Brasil;
• Joint Institute for Nuclear Research, Russia (НЭОСТАР, ЛФЧ).

3. ПРИМЕРЫ РАБОТ, ВЫПОЛНЕННЫХ В НЭОСТАР ЛФЧ НА БАЗЕ МИНИ-КЛАСТЕРА STAR

Начиная с конца 2002 г. по настоящее время на созданном кластере ЛФЧ STAR были выполнены различного рода работы, связанные с обработкой реальных экспериментальных данных, полученных в экспериментах с детектором STAR на RHIC в сеансах 2001–2004 г., с моделированием ряда процессов, а также осуществлена разработка и тестирование новых программных продуктов для нужд коллаборации STAR. Ниже описаны примеры некоторых из этих работ.

2002 год.

1. Работа с реальными экспериментальными данными (RUN II):

• реконструкция событий в Au–Au-соударениях при 200 ГэВ;

• адаптация программы нейронной сети Neuro Net для разделения сигналов электронов от адронов в условиях их регистрации с Barrel Electromagnetic Calorimeter (BEMC):
http://www.star.bnl.gov/cgi-bin/cvsweb.cgi/StRoot/StEmcUtil/neuralNet/.

2. Моделирование Монте-Карло: тестирование на мини-кластере ЛФЧ STAR с использованием пакетов GSTAR, StAF.

3. Изучение рабочих характеристик SMD-подсистемы STAR BEMC: *http://www.star.bnl.gov/∼dmitry/UPC/*.

2003 год.

1. Обработка экспериментальных данных, полученных в RUN III: реконструкция событий, получение и анализ инклюзивных спектров электронов с Pt = 1,5–12,0 ГэВ/*c*, образованных в d–Au-соударениях при 200 ГэВ: *http://www.star.bnl.gov/protected/heavy/dmitry/2004/electrons/index.html.*

2. Моделирование Монте-Карло:

• 200k электронных событий: StAF + STAR geant framework (GSTAR);

• 500k электронных событий (пионы и протоны): StAF + GSTAR.

3. Разработка и тестирование программного обеспечения:

• разработка StRoot класса C++ для выделения электромагнитного сигнала детектора STAR BEMC и его подсистем SMD, PSD;

• разработка STAR database online browser для обращения и работы с базами данных установки STAR в режиме он-лайн: *http://online.star.bnl.gov/Browser/*.

4. Разработка программы работы с базами данных калориметра BEMC в режиме он-лайн (STAR BEMC database online browser): *http://www.star.bnl.gov/∼dmitry/EMC_DB1.1/*.

2004 год.

1. Обработка экспериментальных данных RUN III, IV: регистрация мягких прямых фотонов, образующихся в Au–Au- и d–Au-соударениях при 200 ГэВ методом их конверсии в газе STAR TPC (отработка методики выделения фотонов): *http://wnnuk107.jinr.ru/∼dmitry/*.

2. Моделирование Монте-Карло GSTAR+STARsim:

• розыгрыш событий с фотонами в области Pt = 20–100 МэВ (*http://wnnuk107.jinr.ru/∼julia/soft_gammas/*) и Pt = 2–6 ГэВ (*http://wnnuk107.jinr.ru/∼julia/conv_study/*);

• моделирование Au–Au-соударений при 62 и 200 ГэВ, HIJING + STARsim: *http://www.star.bnl.gov/∼julia29/physics_gamma_pi0/*

3. Реконструкция реальных и смоделированных событий на установке STAR: STAR Big Full Chain (bfc)+ root4star.

4. Разработка и тестирование программного обеспечения для:

• конверсии фотонов на элементах конструкции STAR SVT и в газе TPC: *http://wnnuk107.jinr.ru/∼julia/conv_study/*;

• разработки офф-лайн-программного обеспечения подсистемы BEMC;

• усовершенствования STAR database browser.

ЗАКЛЮЧЕНИЕ

На основе только части результатов работ, из упомянутого выше списка, были представлены два сообщения на «Quark Matter» 2004: характеристики STAR BEMC [4] и измерение инклюзивных спектров электронов с большими Pt, образованных в d–Au-соударениях при $\sqrt{s} = 200$ ГэВ [5]. Из вышеприведенных примеров работ можно видеть, что локальный мини-кластер ЛФЧ STAR предоставляет физикам ОИЯИ, **находящимся непосредственно в Дубне**, достаточно широкие возможности для продуктивного и своевременного участия в процессах обработки, анализа и моделирования экспериментального материала, получаемого коллаборацией STAR.

Вышеописанная технология универсальна и потому применима в любом эксперименте, использующем распределенные вычисления. Основной ее выигрыш состоит в уменьшении сложности настройки рабочей среды до минимума и снижении затрат на поддержку системы в рабочем состоянии, что немаловажно в современных условиях. Использование этой технологии в решениях типа GRID облегчает проблему установки и обновления программного обеспечения на локальных узлах сети, а также практически исключает конфликты программного обеспечения различных коллабораций.

Благодарим Р. Я. Зулькарнеева за поставленную задачу, поддержку работы на всех ее этапах и конструктивное обсуждение результатов.

Авторы благодарят Жерома Лоре (Jerome Lauret, Software and Computing Leader at STAR BNL) и Майкла де Филлипса (Michael DeFillips, Online Computing and Database Leader at STAR BNL) за консультацию по вопросам организации программного обеспечения и помощь в тестировании мини-кластера ЛФЧ STAR в отделе НЭОСТАР.

ЛИТЕРАТУРА

1. *Campbell R.* Managing AFS: The Andrew File System. Upper Saddle River: Prentice Hall, 1998.
2. *Vesperman J.* Essential CVS. Sebastopol: O'Reilly & Associates, 2003.
3. *Zawodny J. D., Balling D. J.* High Performance MySQL. Sebastopol: O'Reilly & Associates, 2004.
4. *Arkhipkin D. et al.* Performance of the STAR Barrel Electromagnetic Calorimeter. «Quark Matter» 2004, LBNL USA, Jan. 11–17, 2004.
5. *Suaide A.* Inclusive Electron Distribution at Hight pt in d–Au and pp Collisions at RHIC. «Quark Matter» 2004, LBNL USA, Jan. 11–17, 2004.

Получено 27 октября 2004 г.

Редактор *О. Г. Андреева*

Подписано в печать 12.01.2005.
Формат 60 × 90/16. Бумага офсетная. Печать офсетная.
Усл. печ. л. 0,5. Уч.-изд. л. 0,6. Тираж 290 экз. Заказ № 54725.

Издательский отдел Объединенного института ядерных исследований
141980, г. Дубна, Московская обл., ул. Жолио-Кюри, 6.
E-mail: publish@pds.jinr.ru
www.jinr.ru/publish/