

КОНСОЛИДИРОВАННАЯ СИСТЕМА ПРЕДСТАВЛЕНИЯ СОСТОЯНИЯ СРЕДЫ СУПЕРКОМПЬЮТЕРНЫХ ВЫЧИСЛЕНИЙ ИММ УРО РАН

Д.А. Усталов, А.Ю. Берсенов, А.В. Созыкин, Д.Г. Ермаков

Введение. Суперкомпьютер — сложная гетерогенная вычислительная система, объединяющая большое количество программного и аппаратного обеспечения. Мониторинг состояния такого вычислителя — критически важная задача, поддержка решения которой обеспечивается при помощи специализированных информационных систем.

Информационные системы, выполняющие сбор, хранение, обработку и визуализацию различных параметров состояния аппаратного и программного обеспечения — важнейшая часть инфраструктуры любого суперкомпьютерного центра (СКЦ).

Нередко при интенсивном развитии инфраструктуры суперкомпьютерного центра выполняется установка и эксплуатация нескольких несовместимых систем мониторинга, предназначенных для разных задач.

Проблематика. Использование такого подхода к сбору данных затрудняет получение оперативной информации о состоянии аппаратного и программного обеспечения вычислительного центра, значительно ограничивает возможности применения методов интеллектуального анализа данных для повышения качества обслуживания пользователей вычислительного центра, а также усложняет подготовку отчетности перед вышестоящими организациями.

В СКЦ ИММ УрО РАН работают две системы мониторинга: 1) собственная система, собирающая сведения об управлении задачами, нагрузке на ЦПУ, ГПУ, память, сеть и СХД; 2) система Zabbix, собирающая данные об энергопотреблении и температурном состоянии серверов [1].

Помимо этого, некоторые сведения о состоянии вычислительных ресурсов поступают в смешанном виде из иных источников: по электронной почте, через информационные ресурсы внутренней сети СКЦ ИММ УрО РАН, а также путём ручного ввода команд на извлечение данных в терминальном режиме. Это затрудняет не только оперативное слежение за состоянием суперкомпьютера «УРАН» и принятие верных решений о развитии вычислительных ресурсов СКЦ ИММ УрО РАН, но и значительно усложняет получение обратной связи от пользователей.

Оперативной информацией являются сведения об общем количестве решаемых на вычислителе задач в заданный момент времени с указанием пользователей и организаций, объёма выделенной оперативной памяти, показателей загрузки центральных процессоров, данных о температурном состоянии узлов суперкомпьютера.

Невозможно мгновенно оценить и наглядно представить, какой пользователь сегодня использует вычислительные ресурсы наиболее интенсивно. Устраивает ли его качество предоставляемых услуг? Не сталкивается ли он с проблемами охлаждения при чрезмерной нагрузке суперкомпьютера?

Получение ответа на каждый из таких вопросов занимает продолжительное время: от нескольких минут до нескольких часов при необходимости взаимодействия различных структурных подразделений ИММ УрО РАН [2].

Цели и задачи. Для решения проблемы предлагается создать консолидированную систему представления данных о состоянии вычислительной системы, наглядно отображающую всю необходимую оперативную информацию в виде графиков, таблиц и изображений.

Цель. Создать систему, представляющую оперативные данные о состоянии суперкомпьютера «УРАН» в режиме реального времени. Обеспечить функционирование системы не только на системах с традиционными устройствами ввода (клавиатура, мышь), но также и на сенсорных дисплеях, планшетных компьютерах, прочих мультимедийных системах. Это позволит снизить время на анализ операционной деятельности вычислительного центра в несколько раз за счёт автоматизации и визуализации данных.

Задачи: 1) собрать данные с существующих систем мониторинга, провести тесную интеграцию с целью получения данных в режиме реального времени; 2) агрегировать данные из этих систем, представить в непротиворечивом и структурированном виде; 3) выполнить визуализацию этих данных.

Реализация системы в виде Web-приложения позволит авторизованному пользователю получать необходимую информацию из любой точки Земли, откуда имеется доступ в Интернет.

Архитектура системы. Принято решение выполнять отправку данных с каждого вычислительного узла при помощи специализированной программы, работающей в фоновом режиме. С целью обеспечения масштабируемости системы и повторного использования данных, выполняется их отправка не конечному получателю, а в высокопроизводительную очередь сообщений. Автономная система представления данных получает сообщения из очереди в режиме реального времени, выполняет их обобщение, и передаёт агрегированную информацию своим клиентам — Web-браузерам. Архитектурная схема представлена на рис. 1.

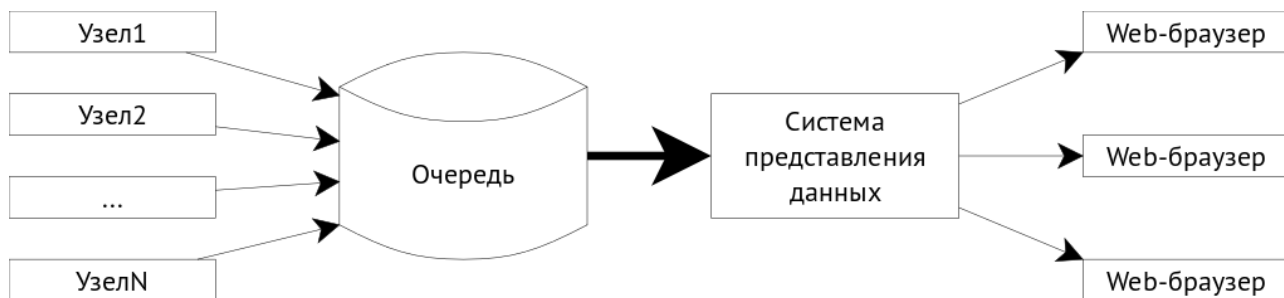


Рис. 1. Архитектура консолидированной системы мониторинга

Реализация. В процессе инженерной реализации системы использовано только свободное программное обеспечение. Программы для сбора и отправки данных разработаны на языке программирования Python. В качестве очереди сообщений использован RabbitMQ — AMQP-брокер от компании VMware. Web-приложение, выполняющее обобщение данных и их передачу в Web-браузер, разработано на языке программирования Ruby. Благодаря технологии Web-сокетов удалось реализовать визуализацию данных по мере их поступления.

Серверное программное обеспечение ориентировано на использование UNIX-совместимых операционных систем. Архитектура и реализация системы не накладывают существенных ограничений на клиентское программное обеспечение: требуется любая операционная система, с которой совместимы браузеры Google Chrome, Mozilla Firefox, Opera, и др.

Графический интерфейс системы выполнен в виде Web-страниц на основе стека технологий HTML5 и представлен на рис. 2 и рис. 3.

Кластер «УРАН»

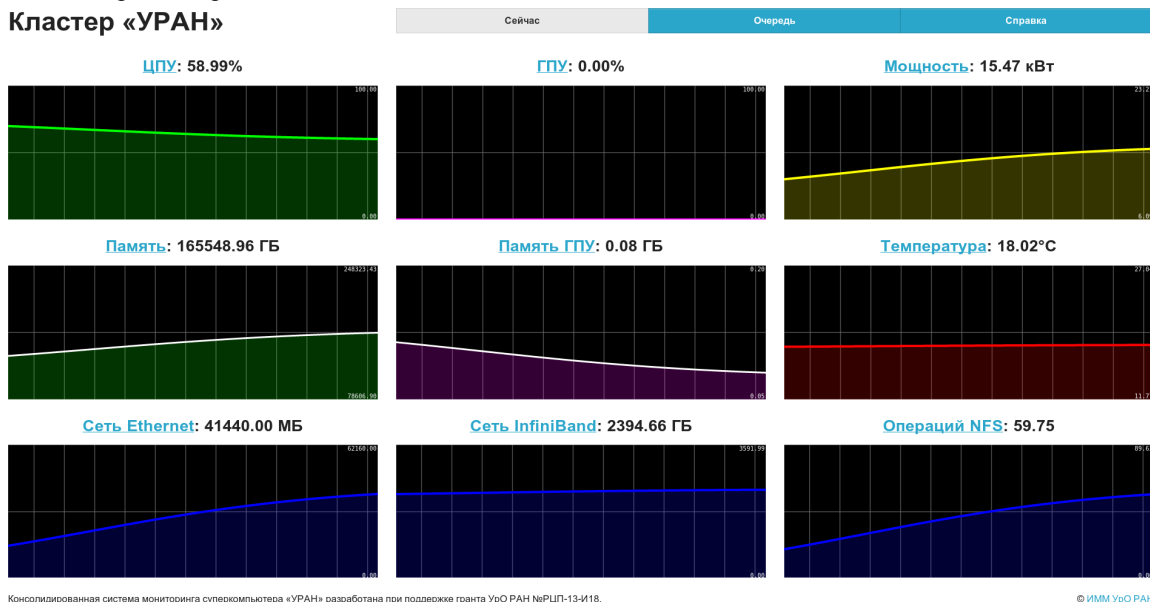


Рис. 2. Экран с оперативной информацией по кластеру «УРАН»

Архитектура системы очередей позволяет потребителям данных «подписываться» на сообщения от определённых источников. Для оценки влияния количества потребителей на количество переданных сообщений был создан ещё один потребитель данных и эксперимент был повторён. Результаты представлены на рис. 5.

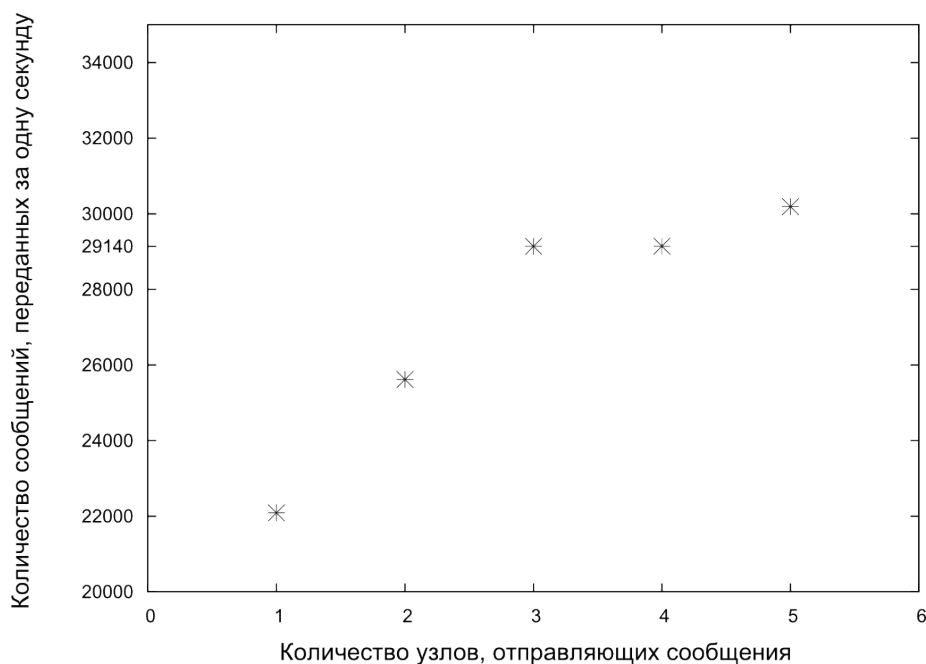


Рис. 5. Оценка пропускной способности в конфигурации «много отправителей — два получателя»

Узким местом оказался процессор на сервере очередей. Ожидается, что при использовании более современного процессора производительность значительно возрастёт.

Таким образом, при текущем состоянии инженерной инфраструктуры СКЦ ИММ УрО РАН, имеется возможность за одну секунду предоставлять данные о двадцати тысячах узлов одновременно и без задержек.

Заключение. Представлен подход к интеграции сведений о состоянии среды суперкомпьютерных вычислений. Предложена архитектура информационной системы, выполняющей консолидацию данных, поступающих от различных инструментов мониторинга. Описано практическое применение данного подхода и архитектуры в СКЦ ИММ УрО РАН. Продемонстрирована масштабируемость системы до двадцати тысяч отслеживаемых узлов при текущем состоянии инженерной инфраструктуры СКЦ ИММ УрО РАН.

Преимуществами предложенного подхода являются возможность неинвазивной интеграции и обобщения информации от систем мониторинга, работающих в вычислительном центре. Среди недостатков подхода можно отметить очевидную необходимость разрабатывать процедуру сбора данных для каждого конкретного источника данных.

В качестве направлений дальнейшей работы можно отметить 1) обеспечение встроенной поддержки интеграции с популярными системами мониторинга; 2) использование преимуществ технологий Semantic Web и Linked Data для развития формата представления данных [3].

Данная работа поддержана грантами УрО РАН №12-П-1-2012, №12-С-1-1001/3, №РЦП-13-И18. При проведении работ был использован суперкомпьютер «УРАН» ИММ УрО РАН.

ЛИТЕРАТУРА:

1. А.В. Созыкин, М.Л. Гольштейн, М.А. Чернокутов "Система оперативного мониторинга температуры и энергопотребления суперкомпьютера «УРАН»" // Программные продукты и системы, №4, 2011.
2. Параллельные вычисления в УрО РАН. URI: <http://parallel.uran.ru/>
3. Д.А. Усталов "Семантический подход к описанию конфигурации среды облачных вычислений" // Научный сервис в сети Интернет: поиск новых решений: Труды Международной суперкомпьютерной конференции (17–22 сентября 2012 г., г. Новороссийск). — М.: Изд-во МГУ, 2012. — 752 с.