

Система мониторинга загрузки суперкомпьютеров ЮУрГУ

А.Ю. Сафонов, П.С. Костенецкий, К.В. Бородулин, Ф.М. Мелехин

Южно-Уральский государственный университет

В Лаборатории суперкомпьютерного моделирования Южно-Уральского государственного национального исследовательского университета разработана и активно развивается система мониторинга и обработки данных с суперкомпьютеров. Система позволяет администраторам и руководителям лаборатории получать различные отчеты о состоянии суперкомпьютеров и предоставляет инструменты для быстрого разрешения возникающих проблемных ситуаций. Работа посвящена проблемам сбора и отображения статистических данных, получаемых администраторами с суперкомпьютеров ЮУрГУ.

В современном суперкомпьютерном центре одной из самых важных задач являются задачи наблюдения за эффективностью использования вычислительных ресурсов и своевременного реагирования на возможные неполадки и проблемы на суперкомпьютере. Не менее важно также постоянно следить за количеством задач в очереди, составе запрашиваемых ресурсов и иметь возможность оценить динамику изменения этих и других статистических данных за определенный период. В Лаборатории суперкомпьютерного моделирования Южно-Уральского государственного национального исследовательского университета разработана и активно развивается система сбора и отображения статистических данных о загрузке суперкомпьютеров [1].

На суперкомпьютере «Торнадо ЮУрГУ» установлено две параллельные системы хранения данных: Panasas ActiveStor 11 и Axus YB. СХД кластера используют параллельные файловые системы PanFS и Lustre соответственно. Пользователям вычислительного кластера по умолчанию доступна СХД Panasas, а Lustre используется для отдельных расчетов, требующих выделения до 40 Тб места на хранилище либо требующих iSCSI доступ (при решении задач в ОС Windows).

СХД Panasas имеет собственную встроенную систему мониторинга с широкими возможностями и надстройки над ней не требуются, в то же время мониторинг СХД на базе Lustre ведется постоянно при помощи Nagios. Хранилище на базе файловой системы Lustre как правило вызывает ошибки двух видов: 1) переход в режим «Только чтение» при аппаратных неисправностях или ошибках в файловой системе; 2) деградация производительности при заполнении хранилища более чем на 85%. Отслеживание данных неполадок осуществляется при помощи Nagios.

Одним из способов организации мониторинга является подход на основе ситуационного экрана суперкомпьютера. Основная задача ситуационного экрана при анализе данных — это предоставление актуальной информации о состоянии системы: отказы оборудования или компонентов программного обеспечения должны быть немедленно отображены на ситуационном экране [2]. Однако в разных суперкомпьютерных центрах кроме решения основных задач существуют собственные требования к системе мониторинга и информационным сервисам.

Необходимость разработки собственной системы была вызвана потребностью в регулярном формировании множества специфических отчетов о загрузке суперкомпьютеров и о деятельности структурного подразделения университета. При возникновении новых требований к отчетам, процедура формирования дополнительных статистических выборок и графиков встраивается в функционал системы и в дальнейшем повторное создание аналогичного отчета многократно упрощается и ускоряется. Система строит графики и помогает системным администраторам формировать отчеты о загрузке суперкомпьютеров. Администраторам доступна функция просмотра информации о пользователях и их расчетах на суперкомпьютере, о текущей загрузке кластеров, количестве задач в очереди, информации о работающих и неработающих узлах. Также возможно автоматическое формирование отчетов о загрузке суперкомпьютеров за определенный период. Рядовым пользователям суперкомпьютера доступна информация о сво-

бодных узлах в очереди, о задачах данного пользователя, как о завершенных, так и о выполняющихся.

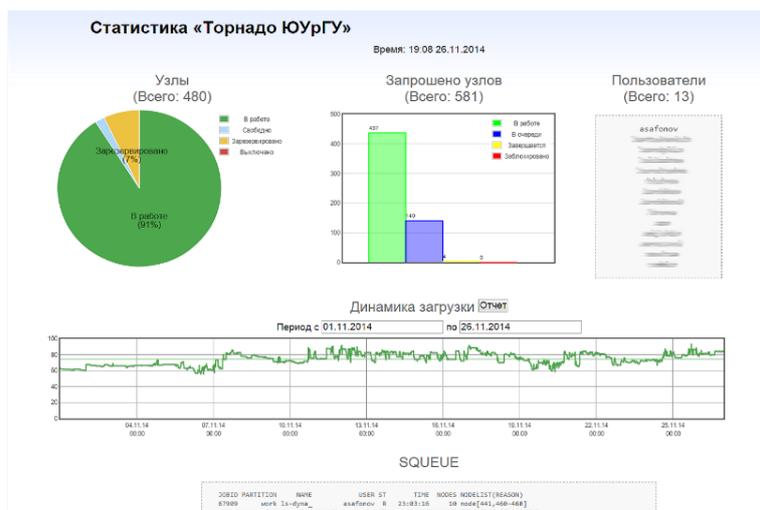


Рис. 1. Веб-интерфейс системы мониторинга загрузки суперкомпьютеров

Очередь задач SLURM настроена на хранение данных о задачах в базе данных MySQL. С помощью Django Framework информация о загрузке суперкомпьютера извлекается из базы, обрабатывается, после чего генерируется веб-страница с подготовленными для визуализации данными. Поскольку Django Framework построен по технологии MVT (Model-View-Template), он позволяет быстро разрабатывать гибкие и надежные веб-приложения, сосредотачиваясь непосредственно на реализации внутренней логики.

На рис. 1 показана главная страница веб-интерфейса системы мониторинга загрузки суперкомпьютеров ЮУрГУ. На данной странице собрана основная информация, интересующая руководителя и системных администраторов Лаборатории суперкомпьютерного моделирования ЮУрГУ, как например:

- 1) статистика по узлам в процентном соотношении (работающие, свободные, зарезервированные для определенных пользователей, выключенные);
- 2) статистика по количеству задач и узлов, запрошенных в рамках этих задач:
 - a. количество уже выполняющихся задач;
 - b. количество задач, находящихся в очереди на расчет;
 - c. количество заблокированных задач (например, превышено максимальное суммарное количество запрашиваемых узлов по всем задачам в очереди; или запрошена лицензия, которая в данный момент уже используется в другом расчете);
 - d. узлы, которые освобождаются и в скором времени вернутся обратно в очередь;
- 3) пользователи, которые находятся «онлайн» в данный момент времени (то есть запустившие терминал удаленного доступа и работающие в нем);
- 4) динамика загруженности суперкомпьютера за определенный период;
- 5) вывод консольной команды «sudo squeue», показывающей подробную информацию о всех задачах, работающих в данный момент с очередью задач.

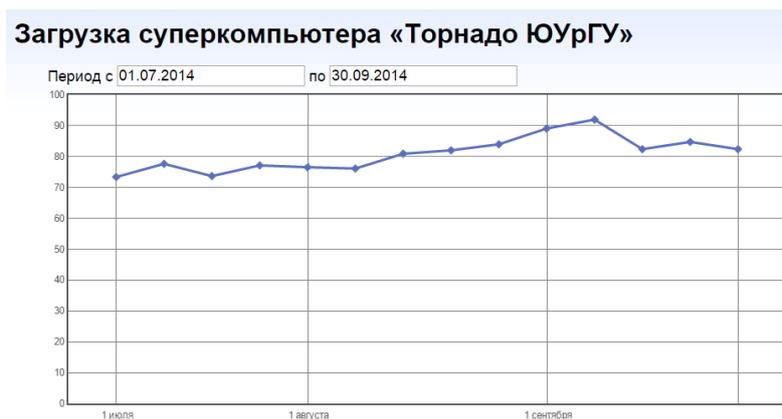


Рис. 2. Пример графика загрузки суперкомпьютера по неделям для квартального отчета

Состояние узлов «Торнадо ЮУрГУ»:

Показать только проблемные узлы
Показано: 480/480

Имя узла	Состояние	Информация sinfo
node001	sshd	
node002	sshd	
node003	sshd	
node004	sshd	
node005	sshd	
node006	sshd	
node007	sshd	
node008	sshd	
node009	sshd	
node010	sshd	
node011	sshd	
node012	sshd	
node013	sshd	
node014	sshd	
node015	sshd	
node016	sshd	
node017	sshd	Maintenance
node018	sshd	

Рис. 3. Пример отчета о текущем состоянии узлов суперкомпьютера

Кроме вышеописанных возможностей, система позволяет формировать отчеты с ограниченным доступом. На рис. 2 можно увидеть динамику изменения загруженности суперкомпьютера за определенный период, на рис. 3 представлен отчет о текущем состоянии узлов суперкомпьютера, а на рис. 4 – пример вывода информации о текущих и завершившихся расчетах выбранного пользователя суперкомпьютера за указанный период.

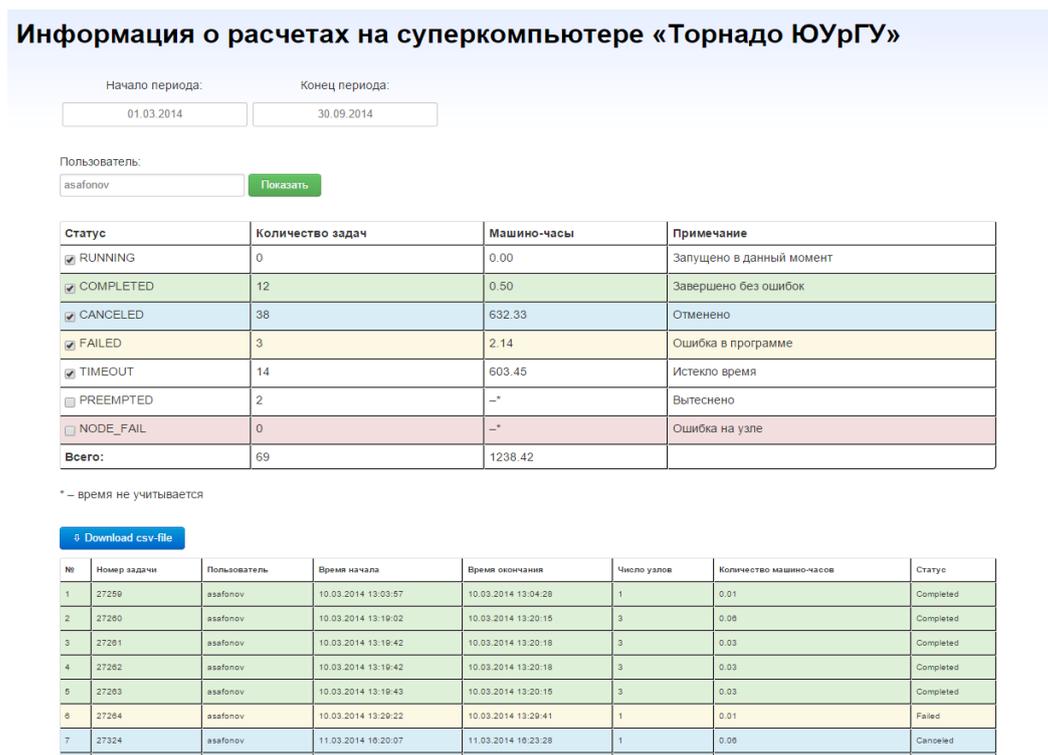


Рис. 4. Пример вывода информации о текущих и завершившихся расчетах выбранного пользователя суперкомпьютера за указанный период

Часто перед администратором стоит задача организовать формирование новых видов отчетов о загрузке суперкомпьютеров. Обычно это вызвано появлением новых требований к отчету подразделения в рамках программы развития Национального исследовательского университета, либо появлением новых сервисов на базе суперкомпьютеров. Создание системы многократно упростило данную задачу, позволив свести задачу к написанию нового модуля в рамках системы мониторинга. Например, в 2015 году был добавлен учет ресурсов, занятых под систему Персональный Виртуальный Компьютер [3,4,5], а также учет загрузки динамически выделенных частей суперкомпьютера [6].

Литература

1. Сафонов А.Ю., Костенецкий П.С. Система сбора и отображения статистики о загрузке суперкомпьютеров ЛСМ ЮУрГУ / Параллельные вычислительные технологии (ПаВТ'2015): труды международной научной конференции (30 марта – 3 апреля 2015 г., г. Екатеринбург). Челябинск: Издательский центр ЮУрГУ, 2015. С. 516.
2. Воеводин В. В. Ситуационный экран суперкомпьютера // Открытые системы. 2014. № 3.
3. Козырев В.И., Костенецкий П.С. Опыт использования VDI-системы «Персональный виртуальный компьютер» в ЮУрГУ // Научный сервис в сети Интернет: поиск новых решений: Труды Международной суперкомпьютерной конференции (17-22 сентября 2012 г., г. Ново-российск). М.: Изд-во МГУ, 2012. С. 285-286.
4. Костенецкий П.С., Семенов А.И., Соколинский Л.Б. Создание образовательной платформы «Персональный виртуальный компьютер» на базе облачных вычислений // Научный сервис в сети Интернет: экзафлопсное будущее: Труды Международной суперкомпьютерной конференции. М.: Издательство МГУ, 2011. С. 374–377.
5. Костенецкий П.С., Семенов А.И. Организация виртуальных персональных компьютеров студентов на базе суперкомпьютера // Параллельные вычислительные технологии (ПаВТ'2011): Труды международной научной конференции (28 марта – 1 апреля 2011 г., г. Москва). Челябинск: Издательский центр ЮУрГУ, 2011. С. 699.

6. Мелехин Ф.М., Бородулин К.В., Костенецкий П.С. Разработка системы динамического разделения вычислительных ресурсов суперкомпьютера на изолированные части // Научный сервис в сети Интернет: многообразие суперкомпьютерных миров: Труды Международной суперкомпьютерной конференции (22-27 сентября 2014 г., Новороссийск). М.: Изд-во МГУ, 2014.

A monitoring system for supercomputers of SUSU

Alexander Safonov, Pavel Kostenetskiy, Kirill Borodulin and Fedor Melekhin

Keywords: supercomputer, management, slurm, mysql

A new system capable of monitoring and processing data for the Supercomputer Simulation Laboratory of South Ural State University is described. This system allows administrators and managers to create many different reports and provides tools to robust reaction to problems of an urgent character. This article is devoted to issues concerning with collecting and visualizing statistical data obtained from supercomputers.