Обеспечение оперативного контроля и эффективной автономной работы Суперкомпьютерного комплекса МГУ*

А.С. Антонов, Вад.В. Воеводин, А.А. Даугель-Дауге, С.А. Жуматий, Д.А. Никитенко, С.И. Соболев, К.С. Стефанов, П.А. Швец

Научно-исследовательский вычислительный центр
Московского государственного университета имени М.В. Ломоносова (НИВЦ МГУ)

В НИВЦ МГУ разрабатывается система для обеспечения оперативного контроля и поддержки эффективного автономного функционирования суперкомпьютерных комплексов. Данная система внедряется в Суперкомпьютерном центре МГУ. В работе описывается опыт установки, настройки и эксплуатации системы для контроля работы суперкомпьютера «Чебышев».

1. Введение

Система Octotron, разрабатываемая в НИВЦ МГУ имени М.В. Ломоносова, предназначена для обеспечения оперативного контроля функционирования суперкомпьютерных комплексов [1, 2]. Система отслеживает наступление нештатных ситуаций в работе всех компонентов комплекса. При возникновении таких ситуаций система выполняет определенный набор действий. Оперативное реагирование на сбои различного рода позволяет минимизировать их негативные последствия, тем самым обеспечивая эффективную автономную работу комплекса.

Кратко напомним основные понятия и идеи, заложенные в систему Octotron. Система работает на основе модели суперкомпьютерного комплекса, представленной в виде мультиграфа [3]. Вершины графа в модели соответствуют физическим или логическим компонентам суперкомпьютера, работу которых необходимо контролировать (вычислительные узлы, источники бесперебойного питания, очереди, лицензии на ПО и т.д.), а дуги – связям между компонентами («состоит из», «обеспечивает электропитанием», «соединены сетью Infiniband» и т.д.). С каждой из вершин связан набор атрибутов – характеристик состояния компонентов (температура процессора, объем памяти, число заданий в очереди и т.д.). Значения атрибутов поставляются штатными системами мониторинга суперкомпьютера либо получаются обращением к внешним интерфейсам компонентов. При изменении значений атрибутов срабатывают зависимые от них правила, определяющие факт наличия нештатной ситуации. В случае определения такой ситуации вызывается определенный набор действий, т.е. выполняется реакция.

Новизна подхода, реализованного в системе Octotron, заключается в использовании модели суперкомпьютерного комплекса в качестве входных данных контролирующей системы: фактически, суперкомпьютер начинает самостоятельно контролировать собственную работу. Среди функциональных аналогов можно назвать решения известных вендоров (HP OpenView [4], IBM xCAT [5]), китайскую систему Iaso [6], созданную специально для суперкомпьютера Tianhe-2, и российскую разработку «Система Автоматического Отключения Оборудования» (САОО) [7] компании Т-Платформы. Упомянутые разработки являются закрытыми и не являются универсальными: возможности первых двух систем во многом завязаны на аппаратуру собственного производства, для Iaso необходима модификация ядра ОС и установка дополнительных компонентов ПО, система САОО контролирует только инфраструктуру вычислительного комплекса.

В настоящее время система Octotron проходит апробацию в Суперкомпьютерном комплексе МГУ. Система уже находится в штатной эксплуатации на суперкомпьютере «Чебышев». С начала 2015 года система запущена в опытную эксплуатацию на суперкомпьютере «Ломоносов», где она будет параллельно работать совместно со штатной системой автоматического отключения оборудования. Данная статья посвящена особенностям настройки и эксплуатации системы Octotron для обеспечения оперативного контроля работы суперкомпьютера «Чебышев».

* Работа выполняется при финансовой поддержке РФФИ, грант №12-07-33047.

2. Модель суперкомпьютера «Чебышев»

Модель суперкомпьютера «Чебышев» (625 вычислительных узлов, 5 000 процессорных ядер) содержит примерно 10 228 вершин, 25 698 дуг и 205 044 атрибута. В модели отражены следующие компоненты суперкомпьютера [3]:

- система электропитания (источники бесперебойного питания, модули с батареями);
- система охлаждения (холодильные установки, воздушные кондиционеры, мониторинг среды);
- управляющая часть (узлы доступа, очереди задач);
- вычислительная часть (шасси, узлы, диски, память);
- файловая система (зависит от типа ФС);
- сеть Ethernet (коммутаторы, порты);
- сеть Infiniband (коммутаторы, менеджер сети).

Следующие типы связей реализованы в модели суперкомпьютера «Чебышев»:

- содержит;
- охлаждает;
- соединены сетью Ethernet;
- соединены сервисной сетью;
- соединены сетью Infiniband;
- включает в себя;
- обеспечивает электропитанием.

На рис. 1 приведен фрагмент модели суперкомпьютера «Чебышев», содержащий компоненты, связанные дугами типа «содержит». Количество вершин графа и дуг между ними велико, однако суперкомпьютер имеет в целом довольно регулярную структуру, поэтому модель содержит множество изоморфных подграфов. На рисунках подобные подграфы объединяются в один подграф, при этом дуге, ведущей к этому подграфу, приписывается число – количество объединенных подграфов.

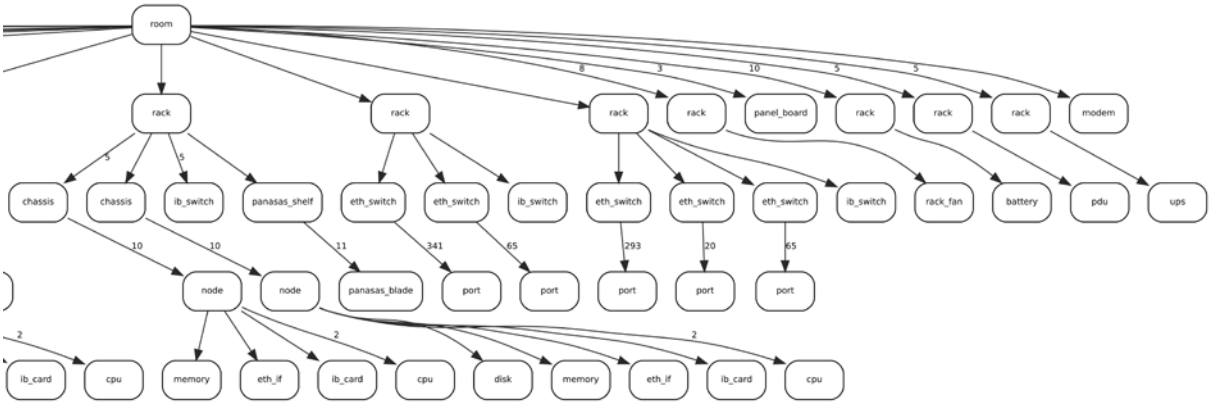

Рис. 1. Фрагмент модели суперкомпьютера «Чебышев»: физические связи между компонентами

Аналогичным образом отображаются фрагменты модели, описывающие охлаждение вычислительного комплекса (рис. 2) и электропитание его компонентов (рис. 3).

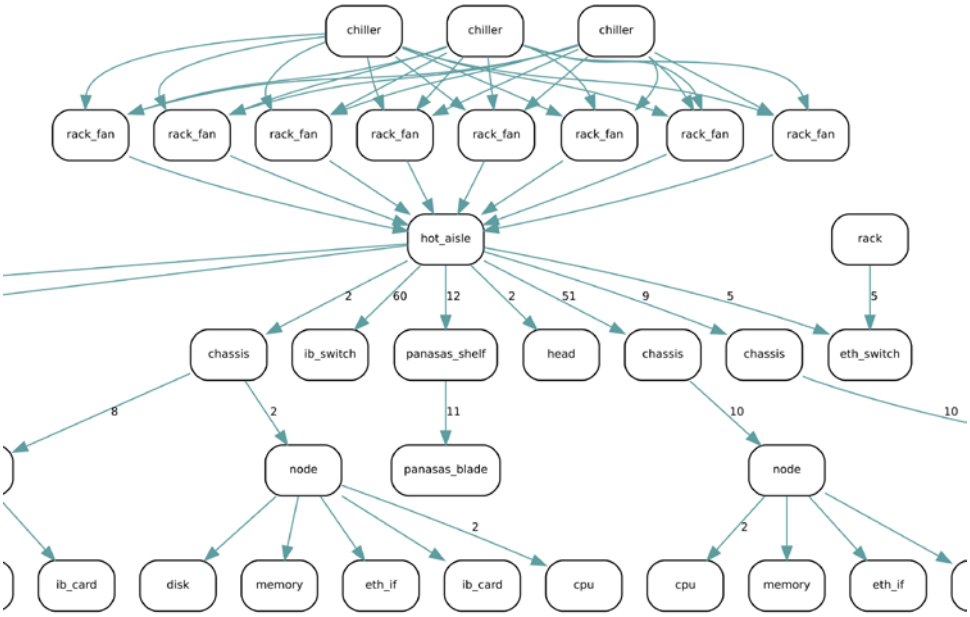

Рис. 2. Фрагмент модели суперкомпьютера «Чебышев»: охлаждение

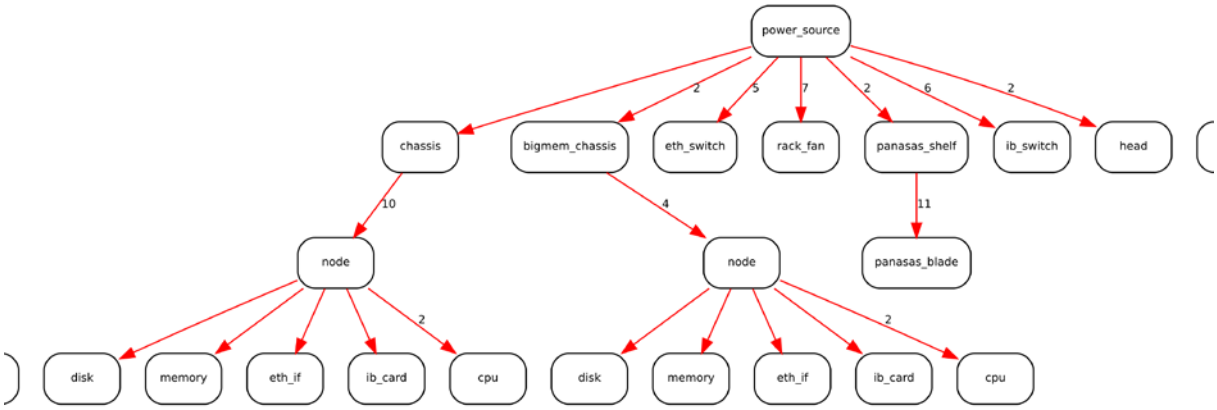

Рис. 3. Фрагмент модели суперкомпьютера «Чебышев»: электропитание

Разработка метода визуализации графа модели потребовала выполнения отдельного исследования. Подсистема визуализации графа на текущий момент позволяет получить компактные изображения, удобные для визуальной проверки корректности модели, для большинства типов связей, за исключением тех, которые образуют циклы в графе (как, например, связь «соединены сетью Infiniband»).

3. Источники данных о состоянии суперкомпьютера

Основным поставщиком данных о работе компонентов суперкомпьютера «Чебышев» является система мониторинга на основе collectd. С инфраструктурного оборудования данные поступают по протоколу SNMP. Все собираемые данные периодически импортируются в Octotron. Заметим, что сама система Octotron жестко не привязана к какой-либо конкретной системе мониторинга. Использована может быть любая система мониторинга, а для ее совместной работы с системы Octotron потребуется разработка несложного модуля импорта данных. Подчеркнем также, что Octotron самостоятельно не опрашивает оборудование и не обращается к датчиками напрямую. В системе есть средства получения сигналов SNMP traps, отправляемые поддерживающим стандарт SNMP оборудованием в критических случаях. Для «Чебышева» таким образом определяются критические сбои в системе электропитания.

С головных машин суперкомпьютера «Чебышев» раз в 10 минут снимаются следующие данные:

- число активных ssh-сессий пользователей;

- число активных лицензий на ПО;
- количество задач в каждой из очередей: общее, ожидающих, готовящихся к запуску, выполняющихся, завершенных;
- число процессоров: общее, доступных для запуска задач, заблокированных;
- баланс счета GSM-модема, подключенного к одной из головных машин и использующегося для рассылки экстренных SMS-оповещений.

Данные, собираемые раз в 10 минут со всех вычислительных узлов:

- температура внутри узла;
- температура каждого процессора;
- идентификатор выполняющейся на узле задачи;
- состояние файловой системы;
- состояние памяти (общий объем, объем свободной/занятой памяти);
- состояние карты Infiniband (счетчики переданных/принятых пакетов, ошибки);
- состояние карты Ethernet (ошибки);
- другие системные данные: средняя загруженность узла, количество процессов-зомби и т.д.

Для узлов с жесткими дисками (около 100 шт.) дополнительно собирается информация SMART о состоянии HDD. Кроме того, один раз в час на каждом узле проверяется работа сервиса ssh, видимость в сети Infiniband и работоспособность MPI.

Для общей файловой системы на основе Panasas раз в 10 минут собираются общие данные (объемы свободного/занятого пространства, производительность), а также статус и загруженность каждого blade-модуля (всего их 132). С той же периодичностью собираются данные с коммутаторов Ethernet.

Информация о работе климатической системы вычислительного комплекса собирается чаще – 1 раз в минуту. Она включает в себя значения с нескольких датчиков температуры и влажности воздуха в помещении, а также состояние каждого из 8-ми кондиционеров (температура воздуха и охлаждающей жидкости на входе и на выходе, различные предупреждения). С той же периодичностью собираются данные с пяти источников бесперебойного питания – около 60-ти параметров с каждого: состояние внешнего питания, собственных режимов, аккумуляторных батарей и т.д.

4. Правила определения сбоев в работе суперкомпьютера

Правила в системе Octotron представляют собой функции, имеющие доступ к атрибутам вершин графа модели суперкомпьютера. Правила могут получать доступ к значениям атрибутов соседних вершин по заданному типу связи. Octotron позволяет формировать правила нескольких типов:

- сравнение значения датчика с константой. Пример: выход температуры компонента за пределы заданного порога;
- наличие аварийных значений одновременно у нескольких датчиков. Пример: сообщение о повышении температуры несколькими датчиками горячего коридора;
- значения датчиков у смежных (по графу модели) компонентов не соответствуют друг другу. Пример: порты на двух концах связи Ethernet находятся в разных режимах;
- сохранение определенного уровня значения датчика в течение заданного промежутка времени. Пример: уровень загруженности узла может ненадолго превысить штатные значения, но достаточно продолжительный высокий уровня загруженности свидетельствует о проблеме на узле;
- получение ошибочных значений датчика несколько раз подряд. Пример: узел более трех раз подряд не проходит проверку доступа по ssh.

Все сбои в работе суперкомпьютера, определяемые правилами, имеют свой уровень критичности. В настоящее время мы используем 4 уровня: Info (информация), Warning (предупреждение), Danger (опасность), Critical (авария). Сбои уровня Critical в основном связаны с повышением температуры воздуха в помещении, горячем коридоре и на компонентах. Подобные

сбои могут нанести существенные повреждения оборудованию и помещениям вычислительного комплекса. К сбоям уровня Danger отнесены ситуации, существенно затрагивающие работу всех пользователей суперкомпьютера, например, отказы системы хранения данных, проблемы с очередями, а также незначительные отклонения в работе климатической и энергетической инфраструктуры. Сбои уровня Warning – локальные проблемы на узлах.

Для контроля работы суперкомпьютера «Чебышев» в настоящий момент используется около 160 правил. Вот некоторые из них:

- баланс счета GSM-модема близок к порогу отключения;
- сбои в работе двух или трех холодильных установок;
- значительный рост ошибок на сетевых интерфейсах;
- слишком малое количество пользовательских сессий на головной машине;
- слишком большое число заблокированных узлов;
- рассинхронизация времени на узлах;
- значение LoadAVG на «свободном» узле превышает значение 3.

5. Методы реагирования на сбои

Если правило определило сбой в работе суперкомпьютера, система может выполнить некоторую реакцию. По умолчанию для каждого события производится запись о нем в лог-файл и отправка e-mail администраторам. Информация о событиях уровня Critical дублируется по SMS. В критических случаях (срабатывание пожарной сигнализации, резкое повышение температуры в помещении, низкий уровень заряда аккумуляторных батарей при питании от них) суперкомпьютер может быть автоматически выключен. В случае недоступности вычислительных узлов по протоколу ssh или неработоспособности на них сервисов MPI они автоматически выводятся системой из счетного поля.

На практике множество сбоев, обнаруживаемых системой Octotron в ходе работы суперкомпьютера, не имеют критического характера и позволяют продолжать штатное функционирование вычислительного комплекса. Чтобы уменьшить интенсивность потока сообщений администраторам суперкомпьютера, в системе предусмотрена возможность блокировки повторяющихся оповещений. Она применяется в абсолютно типичных ситуациях: проблема не является критичной, о ней известно администраторам, но по каким-либо причинам устранить ее не удается.

Еще один полезный для администратора суперкомпьютера сервис – ежедневный дайджест событий, рассылаемый по e-mail. На рис. 4 приведен пример дайджеста с перечнем событий, случившихся на суперкомпьютере «Чебышев» 11 ноября 2014 г. В секции «reported events» перечислены актуальные обнаруженные сбои, в секции «suppressed events» – сбои, для которых была отключена отправка отдельных сообщений.

```
*** reported events ***

13, DANGER, "ntpd drift on node is too big"
2, DANGER, "ems sensor: front temp is very high"

4, WARNING, "bad system temp on node"
1, WARNING, "zombies present on node for last 1000 seconds"

13, RECOVER, "ntpd drift on node is ok"
3, RECOVER, "system temp on node is ok"
2, RECOVER, "ems sensor: front temp is back to normal"

*** suppressed events ***

8, DANGER, "too many free cpus"
```

Рис. 4. Пример дайджеста событий за сутки

6. Дальнейшие планы и заключение

Развитие системы ведется одновременно в нескольких направлениях. Так, статистика сбоев суперкомпьютера представляет исключительно ценный материал для анализа и прогнозирования его поведения. Методы определения типичных сбоев могут отчуждаться и тиражироваться; крайне перспективным направлением представляется создание коллективного банка неисправностей суперкомпьютерных комплексов и методов реагирования на них. Наличие в основе системы модели суперкомпьютера позволяет реализовать различные средства визуализации сбоев для оперативного поиска и устранения неисправностей. Сама по себе разработка модели суперкомпьютера является нетривиальным процессом, который, однако, может быть автоматизирован [8]. Важно, что вся подобная функциональность может быть реализована с помощью подключаемых к системе модулей, в то время как ядро системы Ocotron уже полнофункционально, удовлетворяет заданным при разработке требованиям и выполняет поставленные задачи.

Система Octotron разрабатывается с конца 2012 г., и уже сейчас она успешно эксплуатируется в Суперкомпьютерном комплексе МГУ. На текущий момент она полностью контролирует работу суперкомпьютера «Чебышев». Ведутся интенсивные работы по внедрению системы на суперкомпьютере «Ломоносов».

Разработанная система доступна под открытой MIT лицензией [9, 10].

Литература

1. Разработка принципов построения и реализация прототипа системы обеспечения оперативного контроля и эффективной автономной работы суперкомпьютерных комплексов / Антонов А.С., Воеводин Вад В., Воеводин Вл В., Жуматий С.А., Никитенко Д.А., Соболев С.И., Стефанов К.С., Швец П.А. // Вестник УГАТУ. — 2014. — Т. 18, № 2. — С. 227–236.
2. Соболев С.И. Суперкомпьютер в штатном режиме // Открытые системы. — 2014. — № 8.
3. Швец П.А., Воеводин В.В., Соболев С.И. Об одном подходе к моделированию суперкомпьютерных комплексов // Научный сервис в сети Интернет: многообразие суперкомпьютерных миров: Труды Международной суперкомпьютерной конференции (22-27 сентября 2014 г., г. Новороссийск). — Изд-во МГУ Москва, 2014. — С. 197–204.
4. HP OpenView, http://www.openview.hp.com/solutions/ams/ams_bb.pdf
5. xCAT, An extreme cluster/cloud administration toolkit, http://sourceforge.net/p/xcat/wiki/Main_Page/
6. Lu K. et al. Iaso: an autonomous fault-tolerant management system for supercomputers //Frontiers of Computer Science. – 2014. – Т. 8. – №. 3. – С. 378-390
7. Программное обеспечение компании Т-Платформы, http://www.t-platforms.ru/products/software.html
8. Воеводин Вад В., Стефанов К.С. Автоматическое определение и описание сетевой инфраструктуры суперкомпьютеров // Вычислительные методы и программирование: Новые вычислительные технологии (Электронный научный журнал). — 2014. — Т. 15, № 3. — С. 560–568.
9. Полный исходный код Octotron: https://github.com/srcc-msu/octotron_core
10. Рабочее окружение Octotron для создания модели на языке Python: https://github.com/srcc-msu/octotron