

НРС cloud-сервисы: реальность и перспективы

В середине апреля 2010 г. холдинг "Т-Платформы" представил широкой аудитории в России разработку в области системного программного обеспечения для суперкомпьютеров экзафлопного уровня производительности — операционную систему Clustrx®. Благодаря этой разработке, холдинг "Т-Платформы" уже в конце 2010 г. планирует начать предложение НРС-сервисов в России. О перспективах этого направления SN рассказал директор по развитию компании T-Massive Computing, входящей в состав холдинга "Т-Платформы", Виктор Советов.

Введение

По прогнозам мировых экспертов, суперкомпьютеры преодолеют экзафлопный рубеж производительности лишь через 10–11 лет, однако разработкой программного обеспечения для таких систем специалисты озаботились уже сейчас. В прошлом году был создан международный проект по созданию программного обеспечения для экзавычислений (International Exascale Software Project), а в этом году по инициативе этого проекта страны Большой Восьмерки пришли к соглашению выделить 10 млн евро на разработку системного и прикладного программного обеспечения для экзафлопных систем.

Между тем, отечественное программное обеспечение для инсталляций такого уровня уже разрабатывается, и первые достижения были представлены широкой аудитории в России на международной конференции "Параллельные вычислительные технологии 2010". Разработка носит название Clustrx® и представляет собой операционную систему для суперкомпьютеров с кластерной архитектурой, в которой впервые все средства мониторинга и управления машиной полностью поддерживаются операционной системой, а не вынесены в отдельный пакет, в отличие от других представленных на рынке решений. Эта ОС позволяет обслуживать весь жизненный цикл суперкомпьютера без необходимости установки сторонних программных продуктов и создается для того, чтобы обеспечить эффективную работу суперкомпьютеров различного уровня производительности: от терафлопного до экзафлопного. Сейчас под управлением Clustrx® работает мощнейший в Восточной Европе суперкомпьютер "Ломоносов", установленный в Московском Государственном Университете.

ОС Clustrx® разрабатывается компанией T-Massive Computing, входящей в состав холдинга "Т-Платформы". В разработку этого программного обеспечения холдинг вложил свыше \$1 млн. Новый программный пакет будет поставляться как в составе

высокопроизводительных решений компании "Т-Платформы", так и в качестве отдельного продукта. Кроме того, в ближайших планах разработчиков — создание версии Clustrx® для персональных суперкомпьютеров.

НРС-сервисы в России?

SN. Какие задачи требуется решить для того, чтобы НРС облачные сервисы стали реальностью?

В.С. Я вижу здесь ряд проблем на пути достижения этой цели.

Во-первых, поскольку высокопроизводительные вычисления первоначально появились в научной среде, то им изначально присущ ряд "детских проблем и родовых травм". В частности, не решено много вопросов, связанных с безопасностью вычислений. Сейчас, если клиент хочет посчитать коммерческую задачу, для него очень важно, чтобы ни его исходные данные, ни его модель, ни данные просчета не утекли на сторону, будь то провайдер НРС-услуг (в лице админов) или какое-то другое "заинтересованное" лицо. В настоящее время это очень важный момент, который сдерживает развитие всего этого рынка. Но мы планируем, что уже в конце этого года будем "выкатывать" решения, которые будут тестироваться коммерческими клиентами.

Второй проблемой, сдерживающей продвижение НРС-сервисов, является использование прикладного ПО. Сейчас, если клиент (или провайдер НРС-услуг) хочет использовать какой-либо пакет, к примеру Nastran, для инженерных расчетов, он должен купить лицензию на обоих основаниях и следовать всем пунктам этой лицензии, что очень неудобно для бизнеса. То есть, она может быть привязана к аппаратуре, количеству процессоров, длительности использования и т.д. Мы сейчас на уровне операционной системы вводим механизмы учета, которые позволят перейти к модели аренды

использования программного обеспечения, а не его покупки. Т.е. производитель софта предоставляет свой код/библиотеки провайдеру для продажи его через аренду. Это, во-первых, избавляет провайдера от необходимости платить за лицензии, которые могут и не использоваться или попросту простаивать, а, во-вторых, провайдер может спокойно набрать полный пакет коммерческого софта и предлагать клиентам, что они захотят. И клиент будет платить только за использование коммерческого кода соответственно времени его использования, характеру кода и др., на что будет, конечно, влиять бизнес-модель производителей самого софта.

И, наконец, третья задача, которую необходимо решить для запуска НРС-сервисов, это предсказуемость поведения приложения. Дело в том, что сейчас программы в основном пишутся так, что, если даст сбой хотя бы один узел, то придется пересчитывать либо всю задачу, либо существенную ее часть. Поэтому ни один провайдер не может гарантировать то, что пользовательские задачи будут выполняться какое-то предсказуемое время. Дело в том, что задача может выполняться неделю, две недели и даже месяцы, например, моделирование океанских течений. Вместе с тем, при сбое на любом узле необходимо будет пересчитывать большой кусок, который может занять несколько дней. Поэтому все, кто стоит в очереди на этот ресурс, сдвигаются. И для бизнеса это очень плохо. В настоящее время мы активно разрабатываем соответствующее ПО, которое позволит нам решить и эту задачу. И в следующем году мы уже собираемся предложить разработчикам НРС-моделей несколько механизмов обеспечения надежности выполнения приложений, например, через фиксированные контрольные точки. Резервировав определенное количество запасных узлов, которые будут запущены в случае краха тех, которые используются на задаче, продолжить вычисление, потеряв как можно меньше данных и ре-

зультатов, и сделать это прозрачным для программистов. Этот механизм будет напрямую поддерживаться нашей операционной системой.

Решение этих трех проблем может взорвать НРС-рынок.

Все это должно привести к тому, что появится общедоступная библиотека компонент прикладного ПО, доступ к которым может продаваться как сервис (в том числе).

SN. В этой связи возникает вопрос, насколько производители прикладного софта готовы перейти на бизнес-модель, по сути, товарного кредита?

В.С. На самом деле у них нет другого выхода. Или их будут вытеснять конкуренты, которые перейдут на эту модель раньше, а практически в любой сфере есть несколько конкурирующих пакетов. Либо они будут терять сегмент рынка. Дело в том, что массовые параллельные вычисления дешевеют, становятся все более и более привлекательными для небольших заказчиков из промышленного сектора, из различных дизайн-бюро и т.д. Т.е. из дорогой научной игрушки, которую обеспечивало государство, сейчас акцент все больше смещается в коммерческий сектор.

Поэтому, если они хотят получать деньги за свой код, то это вполне адекватная модель. И, на мой взгляд, они пойдут на это с большой охотой. Естественно, понадобится какое-то время, чтобы убедиться, что учетная система работает так, как предполагается, т.е. этот вопрос, прежде всего, доверия. Естественно, что все расчетные схемы они будут получать со стороны провайдера, т.е. провайдер будет посылать отчеты типа сколько, когда, на скольких узлах, какими пользователями их софт был запущен и, соответственно, с ними рассчитываться, получая на этом свою какую-то маржу.

SN. Т.е. софт будет привязан к конкретной системе?

В.С. Да, конечно, он всегда привязан. Все коммерческое ПО не может функционировать на широком диапазоне систем. Обычно в нем используется достаточно сложная математика и программисты сосредоточены на том, чтобы сделать оптимальным и быстродействующим именно расчетный модуль. Т.е. они “опираются” на какую-то подложку из операционной системы, которая работает на узле, совершенно конкретных библиотек и т.д. Т.е. получается, что весь комплект софта, начиная от операционной системы уровня вычислительного узла, библиотеками и самим приложением “наверху” должен поставляться одним “куском”.

При этом требования к аппаратной составляющей заключаются только в том, чтобы она была совместимой. Т.е., если софт собран с оптимизацией под Intel-процессоры, то узел должен их использовать.

SN. Ряд разработчиков прикладного НРС ПО при поставках его в Россию в силу определенных причин ограничивают его по масштабируемости, например, свыше 200 узлов. Как, по Вашему мнению, будет что-

либо меняться в этом направлении при развитии облачных НРС-сервисов?

В.С. Безусловно, да. Поскольку наличие скоростных каналов связи сильно нивелирует эти ограничения. Т.е., если какой-либо поставщик прикладного ПО урезает его масштабируемость при поставках на какую-либо территорию, то ничего не мешает российскому провайдеру открыть сервисную компанию в Европе или в США. И при наличии достаточно широких каналов для загрузки и выгрузки данных эти ограничения становятся совершенно бессмысленными. К этому можно добавить еще и тот фактор, что удельная стоимость вычислений все время падает и если на каких-либо территориях решат использовать какие-то открытые аналоги этого прикладного ПО и даже при больших затратах по сопровождению, т.е. необходимо поставить больше аппаратных средств, чтобы обеспечить ту же производительность, в силу того, что, как правило, коммерческий софт гораздо больше оптимизирован. Общая стоимость такой системы или услуги все равно не будет такой критичной.

SN. Какие виды НРС-сервисов Вы планируете предлагать?

В.С. Среди основных сервисов следующие:

- расчет готовых моделей на наших мощностях;
- разработка и расчет моделей на наших мощностях;
- хранение неограниченное время моделей на наших ресурсах;
- возможность на основе предоставляемого интерфейса самостоятельного изменения исходных параметров моделей клиентом и запуск их для расчетов на наших мощностях;
- возможность передачи в аренду готовых моделей для использования третьими компаниями.

Отдельно хотелось бы остановиться на последнем сервисе. К примеру, если разработчик этой модели институт, то он может предлагать ее в качестве сервиса. Т.е. другие компании могут запускать ее со своими данными, платить за ее использование сервис-провайдеру, а он определенную часть уже будет отчислять поставщику модели.

SN. Планируется ли предлагать в качестве сервиса еще и верификацию разработанных моделей?

В.С. Тестировать модели на уровне поставщика сервисов не совсем корректно и, по всей видимости, этим будут заниматься независимые компании. Т.е., для того, чтобы включить какую-либо модель в список общедоступных для продажи, она, естественно, должна пройти верификацию двумя-тремя независимыми организациями. На нее будут получены соответствующие сертификаты, и клиент уже будет сам решать в какой степени ей можно доверять. Возможно, что мы заключим ряд договоров для этого с некоторыми отраслевыми организациями.

SN. Что-то Вы можете сказать про особенность аппаратных компонент НРС-систем, планируемых для использования в качестве сервисов?

В.С. Если раньше мы имели дело, в основном, с гомогенными системами, то сейчас переходим к гетерогенным архитектурам, когда несколько вычислителей общего назначения объединяются со специализированными, например, на базе IBM Cell (такое объединение уже есть, например, в НРС-кластере “Ломоносов”) или/и графических спецпроцессоров, а также различных ускорителей, например, с использованием программируемых матриц (для реконфигурируемых алгоритмов). Однако софта для этих специализированных вычислителей наработано еще достаточно.

SN. Что можно сказать о разрабатываемой Вами операционной системе для НРС-кластеров?

В.С. Операционная система изначально проектировалась для нужд НРС, что еще, по сути, отсутствовало в Linux. И только сейчас наметился тренд разделения ОС собственно для кластера и ОС для вычислительного узла. Т.е. в данном случае это совершенно разные вещи. Наша ОС управляет ресурсами кластера, при этом она может грузить на вычислительные узлы какие угодно ОС, какие угодно библиотеки и т.д. У нее есть свои выделенные узлы, свои средства масштабирования, виртуализации и т.д. Она создавалась для управления кластерами любого масштаба, вплоть до экзафлопсов. В настоящее время мы без труда поддерживаем самые большие петафлопсные машины, которые есть на рынке. Это достигается благодаря выбранной архитектуре и позиционированию на масштабируемость и управляемость с самого начала.

Большое количество кода было написано с самого нуля потому, что то, что было в оригинальном Linux, не было ориентировано для нужд НРС. Можно добавить, что Linux стал основной ОС для высокопроизводительных вычислений исключительно по причине своей доступности, бесплатности и очень простой возможности для модификации из-за открытости кода. До недавнего времени такие решения были далеки от оптимальности, надежности и др. Был взрывной рост, который был спровоцирован дешевизной архитектуры на базе Linux. Сейчас, когда мы перешли определенный порог масштабируемости до уровня петафлопсной, вопросы управляемости стали во главу угла и, соответственно, потребовался другой подход к системному ПО. А когда возникла идея предлагать еще и НРС-сервисы, то требования к нему возросли еще в большей степени.

Наша операционная система является распределенной без каких-либо выделенных узлов и единой точки отказа в расчете, что она должна работать на ненадежных узлах, дисках, сети и т.д. Одной ОС мы можем поддерживать миллионы вычислительных узлов.■