Анализ безопасности данных высокопроизводительных вычислений в облачных структурах

authors[1]

[1]universities

15 апреля 2017 г.

Аннотация

Научные вычисления часто требуют большого количества ресурсов для проведения экспериментов. Обычно для этого приходилось обращаться к таким решениям, как кластеры и суперкомпьютеры, трудных для установки, содержания и использования. Облачные вычисления предоставляют ученым новую модель использования вычислительной инфраструктуры. Вычислительные ресурсы, хранилища, уже готовые решения могут быть динамически выделены по требованию и слиты с уже существующей инфраструктурой. Но переход в облачные структуры порождает обеспокоенность в надежности третьей стороны. С точки зрения безопасности появляется ряд неоцененных рисков и задач, ухудшается качество традиционных средств защиты. Данные проблемы уже появлялись при зарождении облачной технологии и были предложены некоторые решения. Данная статья рассматривает возможные риски при переносе научного приложения в облачные структуры, сравнивает задачи и возможные решения, использованные раннее для приложений и предлагает действенные версии реализаций этих решений для масштабирования их до уровня высокопроизводительных вычислений.

1 Введение

Высокопроизводительным вычислениям(HPC) необходима большая мощность для ресурсоемких приложений в области научных исследований, инженерных симуляций, задач бизнеса. С эпохи мейнфремов, затем во времена грид-вычислений, наиболее доступные вычислительные ресурсы, которые могли быть использованы в HPC, были внутреннего пользования (on-premise). Сейчас мы живем в эпоху облачных вычислений. On-premise требует больших инвестиций в закупку оборужования, программного обеспечения и организации инфраструктуры. Это может стать проблемой для среднего и малого бизнеса. Так же это требует ресурсов для поддержания всей системы, хотя возможно, большее время она будет недозагружена. С облаком пользователи могут получать необходимые ресурсы по требованию. Причем не только для основной архитектуры, но и могут запрашивать дополнительные ресурсы, если того требует нагрузка. Более того, облачные дата центры могут выдерживать нагрузку при обработки больших объемой данных. Из этих преимуществ очевидна причина тренда перевода HPC в облако.

Тем не менее, одной из основных причин организаций не мигрировать свои приложения в облако является безопасность. В частности, поднимается проблема конфеденциальности и владения данных, пользователи задаются вопросом, кому я доверяю свои вычисления и данные. Касаемо научного сообщества, в некоторых областях существуют строгие законы о перемещении данных вне юрисдикции определенных властей. Таким образом, для возможности перемещения приложения в облако, необходимо, чтобы облачная структура обладала определенными функциями обращения с ценными данными.

Данная статья рассматривает возможности организации безопасности данных в условиях HPC приложения. Сравнивает различные подходы решения этой задачи в облачных хостингах, предлагает возможности их применения и адаптации в HPC.

В разделе 2 рассматривается специфика HPC приложений, в частности необходимость масштабирования на большое число ресурсов. В разделе 3, поднимается вопрос безопасности, требования, которые должны быть удовлетворены для ее обеспечения. В разделе 4 разбираются решения крупных облачных провайдеров в области безопасности. В разделе 5 предлагаются варианты реализации различных методов организации безопасности в HPC окружении.

2 Специфика HPC

Одним из путей определения HPC это выделение основных требований HPC приложения по сравнению с обычным веб или десктоп приложении. HPC приложениям необходимо больше, иногда значительно больше, ядер процессора, чем в обычном сервере, больше памяти, выше пропускная способность ввода/вывода. Большинство современных HPC приложений требуют паралеллизма - либо посредством развертывания кластеров или грид-систем, состоящих из обычных серверов с возможностью горизонтального масштабирвания, либо созданием специализированного суперкомпьютера или системы с высоким числом ядер процессора, объемом памяти, высокой пропускной способности сети. Последний вариант больше относится к большим, спонсируемым государством компаниям или лабораториям. Но большая часть HPC приложений принадлежит бизнес-сектору. Им больше подходит первый вариант. Поэтому кластеры и грид-системы являются основным средством развертывания и применения HPC приложений. На практике, оказалось, что организациям выгодно и удобно использовать одну систему для многих приложений. Это вылилось в повсеместную эксплуатацию разделяемой инфраструктуры, то есть системы, на чьих серверах могут выполняться вычисления нескольких организаций.

Существуют несколько категорий HPC приложений.

- Слабосвязанные грид-вычисления
- Тесно связанные распределенные вычисления
- Обработка большого объема данных

2.1 Слабосвязанные грид-вычисления

Приложения данной категории могут быть распределены пр огромному числу ядер, сервером в грид-системе, но слабо зависят от производительности отдельного узла системы или межсерверного соединения. Данные приложения обычно спроектированны с учетом возможных падений некоторых узлов или сетевого соединения. Также они поддерживают динамическое добавление новых узлов системы.

2.2 Тесно связанные распределенные вычисления

Приложения этого типа зачастую разворачиваются на небольшом количестве мощных сервером, сильно зависят от их производительности, учитывают топологию системы. Также предполагают возможное расширение системы.

2.3 Обработка большого объема данных

Данные приложения требуют быстрого, надежного доступа к данным. Объем этих данных настлько велик, что они не хранятся там же где производятся вычисления. Поэтому они требуют надежных отказоустойчивых хранилищ данных

Таким образом, из особенностей HPC приложений можно выделить высокие требования к масштабируемости и гибкости системы, причем масштабируемость как горизонтальную, так и вертикальную, возможность эффективно разделять ресурсы для нескольких задач, быстрое реагирование за запросы по изменению архитектуры системы (перемещение данных и вычисления на другие сервера).

3 Безопасность в HPC

Дальнейшие разделы не дописаны. Выписаны основные идеи, которые я хочу туда вложить

Здесь рассмотреть, какая безопасность и зачем нужна. Задать требования к системе.

Пункты рисков:

- Безопасность данных- данные подвергается опасности утечки или потери из-за разделяемых ресурсов, плохого API.
- Из-за возможности утечки данных поднимается вопрос нарушения конфеденциальности в важных научных данных
- Целостность данных
- Контроль за распространением данных по федеральным законам.
- Безопасность результатов и процесса вычислений. Их целостность, конфеденциальность.
- Контроль распределения вычислений, чтобы вычисления выполнялись только на доверенных серверах
- Контроль сети???

Акцентируемся на данных.

4 Системы безопасности облачных провайдеров

Рассмотреть амазон. По azure я ничего не нашел. По digital ocean можно рассмотреть их советы по настройке инфраструктуры, но там вряд ли полезнее амазона.

5 Реализация методов безопасности в HPC

Предложение архитектуры с выделенными сервисом шифрования.

Организация данного сервиса с помощью контейнеров или виртульной машины. Для безопасности ВМ - хорошо, контроль за ядром системы, контейнеры менее безопасны, но не с точки зрения данных.

Описания данного сервиса, что должен делать. Как масштабировать его на много машин(). Возможно ли масштабировать его? Масштабировать с контейнерами проще, с ВМ дольше, но возможен hot migrate. Последнее вряд ли применимо в HPC, так как проводится неавтоматически.

6 Выводы