

ОБЛАЧНЫЕ ВЫЧИСЛЕНИЯ В СО РАН – ВОЗМОЖНОСТИ ПРИМЕНЕНИЯ И РЕАЛИЗАЦИЯ.

Потапов В.П., Пястунович О.Л.
Институт угля СО РАН г. Кемерово
pvp@kemsc.ru

Аннотация

Рассматривается подход по созданию системы облачных вычислений, используя в качестве основы телекоммуникационную среду СО РАН. Анализируются типы услуг в среде облачных вычислений, анализируются возможности реализации конкретных проектов СО РАН. В качестве примера приводится пример реализации конкретной задачи по обработке сейсмической информации с использованием технологии облачных вычислений. Формулируется вывод о перспективности данного направления и возможные пути реализации систем облачных вычислений.

Введение

Обычно под облачными вычислениями понимается комплекс требований по предоставлению расчетных или иных услуг (сервисов) через Интернет. В отличие от классических моделей вычислений, реализуемых на конкретной программно-аппаратной среде, облачная модель определяет клиентов, сервис, управляемый контент и виртуальные машины. Каждый из элементов облака имеет конкретные функции, а их объединение образует то или иное облако. Достоинством такого подхода является то, что облака можно строить как в глобальных, так и в локальных сетях, ориентируясь на их конкретную специфику и типы услуг ими представляемых, что кажется особенно важным для закрытых корпоративных систем. В облаках появляется возможность представления и управления вычислительной мощностью, перераспределяя конкретным пользователям в задаваемое время. Их использование также дает возможность представлять конкретное программное обеспечение, решая тем самым множество вопросов связанных с лицензированием программных продуктов. Это конечно не касается

широко распространенных офисных систем, но в тоже время дорогостоящие программные продукты могут быть использованы в режиме облака. Важной особенностью является динамичность облака, когда в любое время можно изменить его конфигурацию, подключая дополнительные сервисы и организуя интерактивный доступ к ним. При этом один сервер может быть преобразован в множество виртуальных машин тем самым снимая ограничения клиент-серверной архитектуры.

Сегодня, когда существуют достаточно развитые межрегиональные сети, объединяемые в телекоммуникационные структуры и использующие Интернет в качестве рабочей среды, становится важным вопрос о эффективности работы таких систем. В большинстве случаев эти сети используются с одной стороны для коммуникаций между организациями (системы телеконференций с выходом во внешний мир), либо для получения информации из глобальной сети, либо, достаточно редко для проведения удаленных расчетов, когда не хватает вычислительных мощностей. В тоже время сегодня, хотя и немногие организации, даже в СО РАН начинают использовать облачный подход, ориентируясь опять же на те специфические услуги, которые необходимы небольшому количеству исследователей. Многие институты имеют уникальные базы данных, к которым бы с удовольствием обращались исследователи всего мира, но такие случаи все же единичны.

В рамках создания единой телекоммуникационной структуры СО РАН, которая сформировалась за последние годы наверное настало время поставить вопрос о некоторой ее модификации, например используя облачный подход.

Рассмотрим более подробно аспекты реализации такого подхода.

Программное обеспечение как сервис (SaaS-Software as a Service)

Для этого вида представляемой услуги необходимо иметь некоторый реестр, имеющегося сегодня в СО РАН программного обеспечения. В

большинстве случаев это конкретные вычислительные модули, чаще всего работающие в режиме одиночного доступа и производить расчеты по ним можно только при консультации разработчика и его непосредственной помощи, Открытым остается вопрос подготовки исходных данных для расчетов и их местоположение, форматы обмена и т.п. Приведем конкретный пример.

Очень важные расчеты по очистке вод закрытых шахт на протяжении многих лет приходилось делать нанимая разработчиков модели, которую они рассчитывали в закрытом режиме, особенным образом подготавливая передаваемые данные и выдавая результаты расчетов, которые потом сравнивались с экспериментом. Совершенно понятно, что в отделении можно найти множество таких примеров и формируя облачный сервис надо много решить как в организационном, так и в правовом направлении. Иными словами необходимо убедить разработчиков, что именно облачная организация предоставит им больше возможностей и отвлечет от рутинной работы, хотя это вопрос, многие держатся за свои модели, предпочитая зарабатывать на них. Но сегодняшний все более ужесточаемый контроль со стороны государства, наверное, заставит многих из них изменить свое отношение, а этому как раз и должны способствовать облачные вычисления. Кроме этого грамотно организованный сервис должен способствовать устранению дублирования разработок, когда одни и те же вычислительные модели разрабатываются в различных организациях, а потом возникают споры у кого они лучше.

Рабочее место как услуга (Workspace as a Service)

Подобная услуга представляет собой конкретное программное обеспечение, направленное на решение конкретной задачи, либо класса задач и в основном предназначается для реализации удаленных расчетов. Пользователь при этом по умолчанию принимает правила подготовки данных, представления расчетов, загрузки баз данных и т.п. Этот вид сервиса можно считать частным случаем **SaaS**.

Платформа как сервис (PaaS Platforme as a Service)-

обычно представляется связующее программное обеспечение, доступ к базам данных на основе некоторой СУБД. Примером такого сервиса может служить библиотека MPI на вычислительных кластерах, интерфейс к базам геоданных на основе Oracle Spatial и т.п. Данный вид сервиса уже доступен в телекоммуникационной среде СО РАН. Например из Кемерово проводились геомеханические расчеты на Иркутском кластере. Сегодня разрабатываются модули по решению задач затопления шахт, подземной газификации и т.п. Наверное, это на сегодняшний день наиболее употребительный вид сервиса в системе. Вопрос стоит сегодня только в случайности связей, которые формируются в основном по личной договоренности. Здесь необходимо разработать некий регламент имеющихся услуг и способы представления клиентам. Сразу отметим, что такой вид сервиса, скорее всего, будет корпоративным и действовать внутри СО РАН.

Инфраструктура как сервис (Infrastructure as a Service, сокр. IaaS) – включает в себя аппаратные средства и технологию для компьютерных вычислений и хранения данных, операционные системы и другую инфраструктуру, которые предоставляются не как локальные ресурсы, а опосредованно – через обращение к сервисам, размещенным на стороне провайдера. Имеется также модель Аппаратные средства как сервис (Hardware as a Service, HaaS), но она скорее является подтипом модели IaaS. Наверное, в структуре телекоммуникационной структуры СО РАН это наименее подготовленный вид сервиса, так как он должен формироваться интегрально и на основе предыдущих типов услуг. Данный вид сервиса в облаке требует значительных организационных, правовых и финансовых усилий и для его осмысления необходимо провести ряд работ по региональным центрам в рамках единой идеологии облачных вычислений. Последовательность действий по созданию такого сервиса включает в себя

ревизию и формирование двух предыдущих видов сервиса по центрам, с последующей интеграцией на уровне ИВТ СО РАН.

Встает законный вопрос, а зачем вообще нужна организация вычислительных облаков, если уже имеется некоторая структура, которая в определенной степени удовлетворяет требованиям большинства исследователей.

Прежде всего, облачные вычисления позволят повысить эффективность работы в системе, предоставить общие информационные ресурсы, гибко интегрировать их для решения различных задач, набирая и комбинируя их в виртуальных кластерах данных, быстро задействовать новые мощности, оптимизировать инвестиции. Кроме этого, существенным преимуществом такой структуры должно стать представление новых уникальных информационно-вычислительных услуг, интегрируя в облаке работы Сибирского отделения. Не исключается и синергетический эффект, когда интеграция отдельных ресурсов позволит получить качественно новые услуги, которые не может представить одна из организаций СО РАН. Для примера, интегрируя хранилище данных по угольной промышленности Кузбасса, базу спутниковых снимков ИВТ СО РАН и вычислительные геомеханические модули ИГД СО РАН, мы можем за короткое время создать уникальную глобальную систему геомеханического мониторинга крупного горно-промышленного региона.

Такой подход не меняет общую концепцию развития системы телекоммуникаций, а придает ей, более законченную форму, которая легко может изменяться в зависимости от приоритетов развития самого отделения в целом.

Как уже было отмечено выше, облачная структура системы позволит сэкономить и на программном обеспечении, так как для отдельных пользователей, особенно это касается пользователей геоинформационных систем, аренда уже имеющихся активов не потребует создавать собственные центры по сбору и обработке, а пользователь может не задумываться о

приобретении необходимого софта и аппаратных средств, экономя на этом значительные суммы. Здесь правильно определиться с необходимыми программами и приобрести их централизованно, обеспечив к ним соответствующий доступ.

Создание облака даст пользователям возможность работы с наиболее современными системами, при этом не заводя в штате квалифицированных программистов и системных администраторов, так как основная ответственность, кстати и финансы, лягут на плечи провайдера услуг. При этом специализированный сервис, а именно этим и должны заниматься институты отделения, достанется именно им, однако у них “не будет болеть голова” о сопровождении системы, а ее внешние клиенты будут получать надежные, устойчивые, обновленные решения для эффективного выполнения стоящих перед ними задач.

Оптимизация корпоративных данных, вычислительных кластеров, к которым будет организован сервисный доступ, позволят избавиться от фрагментации самой телекоммуникационной сети, и более четко прогнозировать капиталовложения на ее структуру. Доступ к ресурсам облака организуется с помощью стандартных программных средств и весь вопрос состоит в том, что необходимо создать соответствующие метаописания процессов и данных в облаке, но как раз эти работы в СО РАН являются одними из наиболее развитых и практически все организации, занимающиеся телекоммуникационными ресурсами к этому готовы и нужно лишь их соответствующее организационное обеспечение.

Отдельно необходимо рассмотреть вопросы безопасности, а также интеллектуальной собственности, которая будет реализовываться в облаке. С нашей точки зрения это очень важные вопросы и их решение должно основываться с одной стороны на уже имеющихся системах, имеющихся в телекоммуникационной структуре СО РАН. А с другой стороны, основная часть этих вопросов будет сосредотачиваться у провайдера, который имея наиболее квалифицированные кадры, будет решать их по мере поступления.

Сложной проблемой является интеллектуальная собственность, но здесь именно облачная организация и позволит решать конфликты на основе того, что сама собственность остается у разработчика, а не передается другим пользователям, как это часто бывает и приводит к сокращению жизненного цикла самой разработки, из-за отсутствия сопровождения, в результате чего интерес к продукту пропадает. В облаке передается лишь сервис, да и чаще всего клиенту именно он и нужен. Кроме того он не платит за саму разработку, что иногда составляет существенные суммы, и ему не нужно обучать соответствующих специалистов, для сопровождения системы.

Одним из преимуществ облачной системы является то, что с ее помощью можно интегрировать знания, накопленные специалистами различных наук и формировать соответствующие метазнания, которые также представляются в форме некоторых услуг. Для примера можно указать интеграционные мультидисциплинарные проекты СО РАН, которые так и остаются чисто академическим решением и о них мало кто знает, но помещенные соответствующим образом в структуру облака они могут представлять значительный интерес для всего научного сообщества. Сформировав соответствующие метазнания, которые бы позволили работать в режиме “ON DEMAND”, можно получить уникальную структуру облачного сервиса.

Достоинством данного подхода является также и то, что на первом этапе можно выстраивать систему независимых облаков, организовав затем иерархическую или сетевую структуру, состоящую из множества подоблаков, каждое из которых может быть либо проблемно-ориентированным, либо отражать специфику конкретного регионального центра.

При этом, с точки зрения технологий создания необходимого программного обеспечения, элементы облачного сервиса можно организовывать как некоторые WEB-порталы, расширяя впоследствии список услуг за счет использования специального программного

обеспечения, доступ к которому представляется опять же через Web-интерфейс. Учитывая, что системы тарификации и биллинга в Интернете достаточно развиты, то вопросы проведения взаиморасчетов между клиентами системы и центром облака реализуются на этой основе также достаточно просто.

Реализации геомеханической задачи с использованием облачного подхода

В 2010г в рамках реализации реализации конкретной геомеханической задачи с использованием облачного подхода в Институте угля и углехимии в рамках проводились работы по созданию телекоммуникационной среды СО РАН.

Постановка задачи. Кемеровская область как один из крупнейших горнодобывающих регионов России обладает сложной горно-геологической обстановкой, на которую существенное влияние оказывают технологии и методы ведения горных работ. Как следствие на территории Кузбасса периодически происходят землетрясения разной силы, оползни и другие сейсмические события, приводящие к нарушению нормального функционирования шахт и разрезов. Это, в свою очередь, влечет за собой многомиллионные финансовые потери, как для владельцев предприятий, так и для бюджета области. В связи с этим особую актуальность приобретает вопрос о мониторинге сейсмической активности на территории региона. Но, не смотря на это, в Кузбассе существуют проблемы с получением сейсмоданных необходимых для определения миграции центров сейсмозергвыделения.

Сотрудниками Института угля СО РАН был выполнен поиск источников, предоставляющих информацию о сейсмических событиях для территории Кемеровской области. На основании его результатов был выбран сайт Центра данных Института геофизических исследований Национального ядерного центра Республики Казахстан (Центр данных ИГИ НЯЦ РК, Республика Казахстан, www.kndc.kz). Получаемые этим центром данные

обрабатываются и публикуются на сайте в открытом доступе с предоставлением возможности их загрузки по стандартным протоколам через сеть Интернет.

Принимая во внимание географическую удаленность источника данных от средств их обработки, было принято решение о разработке информационной системы с применением технологии облачных вычислений. В этом случае ключевые компоненты (источники данных, аналитические модули и средства интерпретации данных) объединяются в «облако» доступное в рамках глобальной сети Интернет. Пользователь получает возможность решать интересующие его задачи, общаясь исключительно с web-интерфейсом и не вникая в особенности реализации вычислительного ресурса и пространственное распределение его элементов.

Промоделируем решение задачи мониторинга миграции центров сейсмозерговывделения с использованием «облака». После прохождения процедуры авторизации сервер приложений отправляет запрос картографическому серверу на генерацию и отображение электронной карты. Затем пользователь с помощью специального инструмента выделяет интересующую его территорию, для которой будет построена траектория миграции. На основе действий пользователя формируется запрос для получения данных с сайта Центра данных ИГИ НЯЦ РК в виде бюллетеней. Каждый сейсмологический бюллетень представляет собой текстовый файл, содержащий записи о сейсмических событиях в формате IASPEI Seismic Format (ISF), состоящие из таких полей как дата, время, географическое местоположение, азимут, глубина, магнитуда и так далее. Загруженные сейсмобюллетени проходят предварительную обработку, в ходе которой из них вычлняются события, попадающие в границы Кемеровской области, отсекаются избыточные для решения поставленной задачи данные. Промежуточные результаты передаются в виде матрицы параметров вычислительному кластеру для расчетов центров сейсмозерговывделения по каждому календарному месяцу. Полученные данные являются дискретными,

содержат географические координаты долгота/широта и, следовательно, могут быть отображены на электронной карте. Поэтому они передаются на уровень web-интерфейса для последующей отрисовки в виде полилинии траектории движения центров сейсмоэнерговыведения на электронной карте.

Используя возможности созданного «облака» была построена траектория миграции центра сейсмопроявлений для Араличевского геолого-экономического района, которая представлена на рис. 1. Хорошо видно, что при движении центр сейсмоэнерговыведения огибает правобережные районы города Новокузнецка. Данное поведение объясняется расположением промышленной застройки, которая создает напряжения в массивах горных пород, способствующие возникновению сейсмических событий. Территория между районами города свободна от застройки и, по-видимому, служит зоной разгрузки.

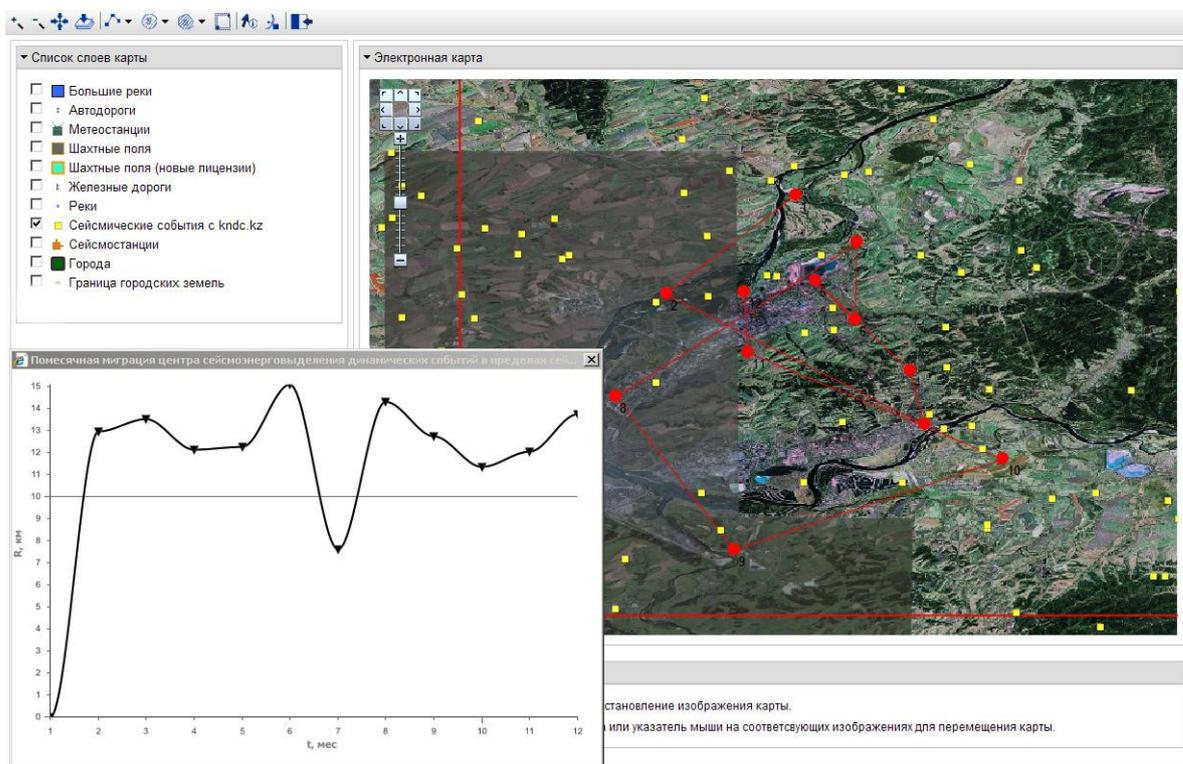


Рис. 1. Траектория движения центра сейсмоэнерговыведения в районе города Новокузнецка

Выводы.

Сегодня, с учетом уже имеющейся телекоммуникационной структуры СО РАН, созданы уникальные условия для организации системы облачных вычислений, реализующих различные виды сервисных услуг. При этом можно выделить чисто внутренний сервис, работающий для организаций СО РАН, однако наибольший интерес могут представлять, услуги, реализуемые на внешнем уровне, это позволит с одной стороны организовать внедрение разработок СО РАН, а с другой стороны даст возможность создавать системы уникальных корпоративных расчетов для сторонних организаций.