УДК 20.53.19, 28.23.13

И.И. Холод

Единая информационно-аналитическая платформа на базе защищенных облачных технологий

Аннотация. В данной статье описывается архитектуру построения облачной платформы для интеграции гетерогенных информационных систем на базе защищенных технологий. Такой подход позволяет интегрировать информацию и применять к ней современные методы анализа. Использование защищенных технологий позволяет использовать платформу в структурах государственных органов, включая Министерство обороны.

Ключевые слова и фразы: облачные технологии, защищенная облачная платформа, интеллектуальный анализ данных, data mining.

Введение

В настоящее время во многих органах государственной власти функционируют по нескольку автоматизированных информационно-аналитические системы, разработанных в рамках различных ОКР. Данные системы разрабатывались разными предприятиями, в разные годы и в результате имеют существенные различия, в части:

- аппаратных платформ: используются ПЭВМ и сервера закупленные в период с 90-х годов прошлого века, до настоящего времени;
- операционных систем (ОС): используются ОС семейства Windows (Windows NT, 2000 и др.), МС ВС, AstraLinux и др.
- систем управления базами данных (СУБД): используется Oracle, PostgreSQL, Линтер и др.;
- архитектур: клиент-серверные, терминальные, «настольные», тонкого клиента и др..
- протоколов взаимодействия: FTP, SOAP и др.

Независимая разработка каждой системы приводит к дублированию в них:

- информационной составляющей в части: нормаивно-справочной информации, характеристик объектов и т.п.
- прикладной составляющей в части: сбора и ведения информации, решения функциональных задач, подготовки аналитических отчетов и др.

Такая ситуация приводит к следующим проблемам в эксплуатации систем:

- сложность замены (модернизации) как аппаратных, так и программных средств в связи с их разнородностью;
- повышенные требования к квалификации эксплуатирующего персонала в связи с необходимостью знаний различных технологий (ОС, СУБД и др.);
- быстрое устаревание аппаратной составляющей и не возможность ее замены в связи с сильной зависимостью от программного обеспечения;
- сложность обеспечения безопасности информации в связи с использованием различных средств безопасности и различным уровнем сертификации;
- не равномерная загрузка вычислительных средств;
- высокое электропотребление при простоях систем;
- невозможность получения аналитики по совокупности информации, хранящейся в разных системах.

Проблемы усугубляются с разработкой каждой новой системы, т.к разрабатываются они, как правило независимо от уже существующих, тем самым внося свою лепту в разнородность используемых средств.

Мировой и отечественный опыт

В мировой практике для решения данных проблем применяют, бурно развивающиеся, технологии облачных вычислений. Эти технологии представляет собой модель обеспечения доступа по требованию к общему пулу конфигурируемых вычислительных ресурсов (например, вычислительным серверам, устройствам хранения данных, системам и сервисам — как вместе, так и по отдельности), которые могут быть оперативно предоставлены и освобождены с минимальными эксплуатационными затратами и/или обращениями к поставщику [1]:

В настоящее время выделяют несколько основных моделей облачных вычислений [1].

- программное обеспечение как услуга (SaaS — Software as a Service);
- платформа как услуга (PaaS — Platform as a Service);
- инфраструктура как услуга (IaaS — Infrastructure as a Service);
- рабочее место как услуга (WaaS — Workplace as a Service).

Среди мировых лидеров в области разработки технологий облачных вычислений можно назвать компании: Amazon [2], Google Inc [3], Microsoft [4] и др.

Они построены на платформе Apache Hadoop [5] и предоставляют масштабируемую инфраструктуру для выполнения созданных пользователем (по определенным правилам) приложений..

Существуют решения на базе облачных вычислений и для министерства обороны США. В качестве примера можно привести систему Army Private Cloud (APC2) [6]. Данная система на базе облачных вычислений предоставляет сервисы по заключению и отслеживанию выполнения контрактов с МО США. Данный сервис доступен как подразделениям МО, так и частным компаниям.

Из отечественных решений можно выделить национальную облачную платформу [7] — комплекс интегрированных информационных систем, предназначенный для предоставления органам исполнительной власти различного уровня, органам местного самоуправления, коммерческим организациям и физическим лицам прикладные сервисы в сфере здравоохранения, образования, безопасности, жилищно-коммунального хозяйства, имущественно-земельных отношений. Для предприятий малого и среднего бизнеса созданы сервисы для организации работы офиса, управления взаимоотношениями с клиентами, учёта торговых и производственных операций.

Все перечисленные облачные сервисы обладают двумя существенными недостатками:

1. требуют загрузку данных в свое хранилище, что потенциально повышает уровень секретности и требования к обеспечению безопасности;
2. используют программные средства, не разрешенные для использования в государственных структурах и в особенности МО РФ.

Эти недостатки делают невозможным их применение для разработки в области государственных структур и в частности в военной технике.

Однако принципы и технологии, заложенные в них, можно успешно использовать при построении защищенного «облака» в интересах МО РФ. Примером такого развития являются проекты Глобус [8] разработки компаний ОАО Северное конструкторское бюро и проект SeaCloud компании ОАО «СиПроект» [9] .

Глобус – защищенная облачная платформа, построенная на базе ОС Astra Linux Special Edition и системы с открытым кодом OpenStack. Данное решение позволяет строить облачные платформы уровня IaaS для виртуализации в них различных АС.

SeaCloud – платформа также ориентирована на предоставления сервисов уровня IaaS. Использует защищенную ОС Astra Linux Special Edition. Концепция SeaCloud предполагает создание виртуальных образов функционирующих рабочих мест и серверов и размещение их в «об-

лаке». За счет фиксации контрольных сумм снятых образов предполагается сохранение целостности и валидности полученных на эти системы сертификатов.

Обе системы, несмотря на использование ОС Astra Linux Special Edition имеют общие недостатки:

- на данный момент не сертифицированы (не имеют сертификатов не по НДВ не по одному классу защищенности);
- не решают проблемы разграничения доступа к информации с различным уровнем защиты;
- ограничиваются уровнем IaaS не предоставляя пользователю дополнительных сервисов (например, аналитических сервисов).

Архитектура защищенной облачной платформы

Построение защищенной облачной платформы должно решать следующие задачи:

- размещение существующих и вновь разрабатываемых АС на вычислительных ресурсах облака в защищенной среде;
- централизованный доступ к единой НСИ, актуальным характеристикам контролируемой информации и др.;
- проверка доступа, достоверности и др. виды обработки едиными унифицированными алгоритмами;
- решение аналитических и расчетных задач, в том числе статический и интеллектуальный анализ для выработки рекомендаций;
- разграничение доступа к информации в зависимости от грифа информации и уровня доступа пользователя и др.

Первая задача во многом решается существующими средствами систем Глобус и SeaCloud. Остальные задачи требуют новых решений. Для этого предлагается построение защищенной облачной платформы с единым информационным пространством в соответствии с архитектурой, представленной на рисунке 1.

Данная платформа будет включать в себя два уровня:

- физический уровень;
- виртуальный уровень.

На физическом уровне будет размещаться аппаратно-программное обеспечение, реализующее необходимые функции для работы виртуального уровня.

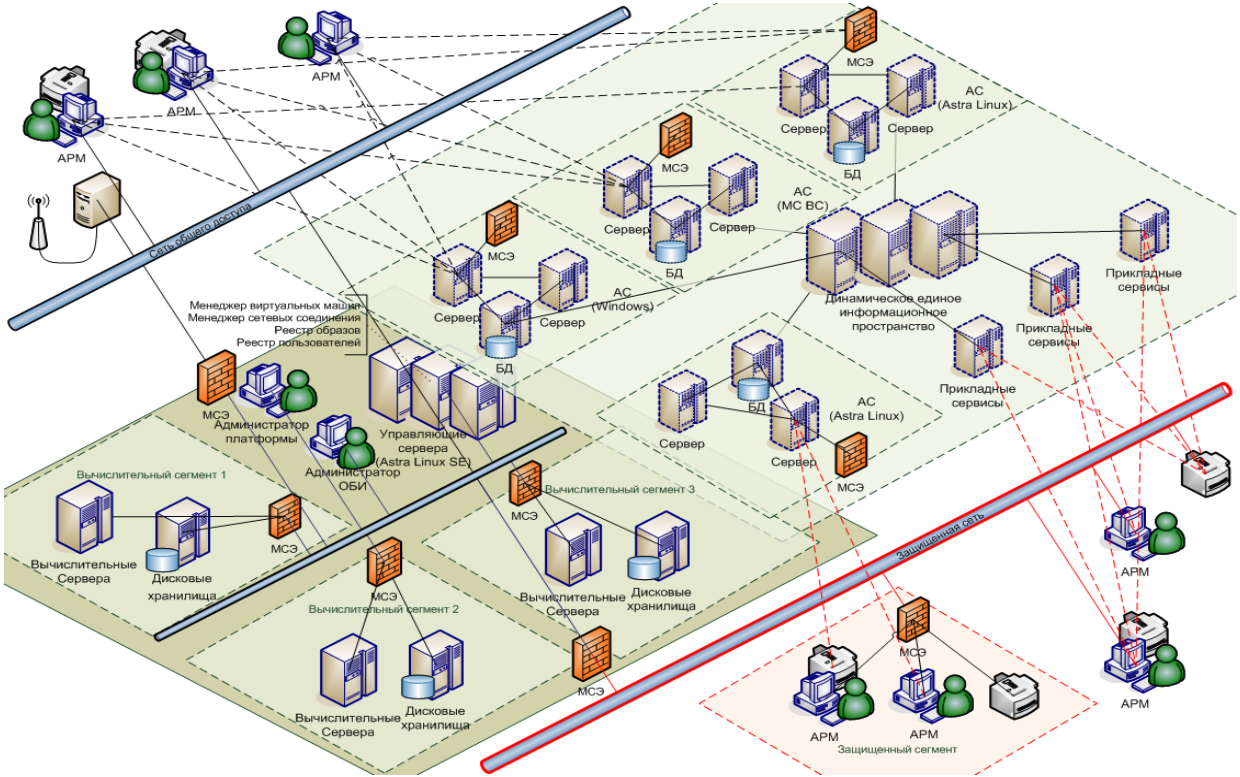

Рисунок 1. Архитектура интегрированной информационно-аналитической защищенной платформы.

Аппаратное обеспечение будет включать в себя следующие элементы:

- управляющие сервера – предназначенные для работы на них специального программного обеспечения (СПО) обеспечивающего работу платформы (в качестве такого СПО может быть использована платформа Глобус);
- рабочее место администратора платформы – отвечающего за управление всей платформой;
- рабочее место администратора ОБИ - отвечающего за настройку прав доступа к ресурсам платформы;
- вычислительные сегменты – предназначенные для выполнения на них виртуальных машин с прикладными системами, между которыми должен быть разграничен доступ на физическом уровне по соображениям безопастности.

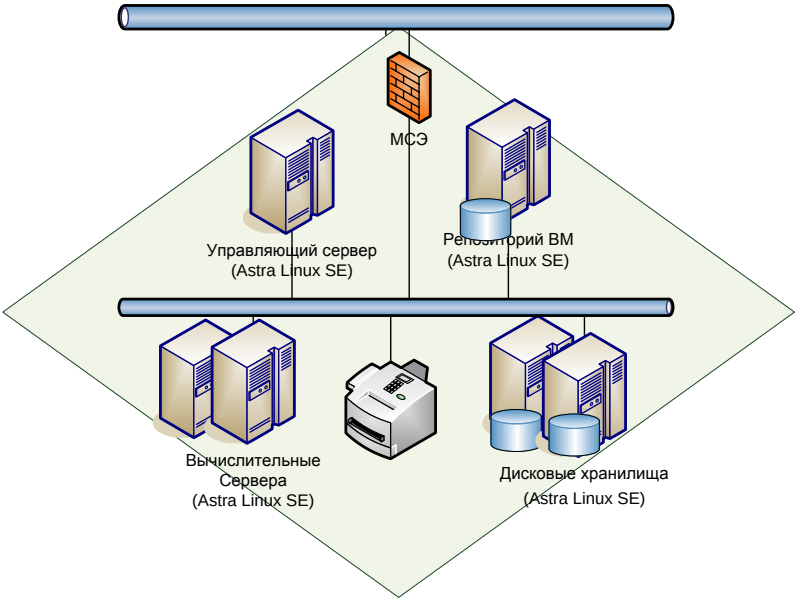

Рисунок 2. Архитектура вычислительного сегмента

Каждый вычислительный сегмент (рис. 2) должен включать в себя:

- межсетевой экран – обеспечивающий разграничение доступа к ресурсам сегмента;

- управляющий сервер – отвечающий за управление виртуальными машинами внутри сегмента;
- репозиторий виртуальных машин – хранящий образы виртуальных машин с прикладными системами;
- вычислительные сервера, на которых будут выполняться виртуальные машины с прикладными системами;
- дисковые хранилища, на которых будет сохраняться информация имеющая гриф соответствующий сегменту.

Виртуальный уровень будет включать в себя виртуальные машины, выполняющиеся на доступных вычислительных серверах. Такие виртуальные машины могут быть как образами уже эксплуатируемых систем, так и виртуальны машины, созданные по требованиям для вновь разрабатываемых систем.

Для сохранения функционирования существующих систем они должны быть перенесены в «облако» на виртуальные машины, настроенные в соответствии с требованиями предъявляемыми данными системами к аппаратной части. Для обеспечения целостности, а также актуальности сертификатов выданных на системы, они должны сниматься целиком в виде виртуальных образов. При этом виртуализации должны подвергаться серверные машины: сервера приложений, базы данных, межсетевые экраны и т.п. Сетевая конфигурация между реальными машинами так же должна сохраняться и воспроизводиться на виртуальном уровне в «облаке».

С клиентских машин доступ к ним будет осуществляться (в зависимости от архитектуры системы) или в терминальном режиме или через протоколы, реализованные в системах. При этом доступ от клиентских рабочих мест к соответствующим виртуальным машинам должен будет регламентироваться как на виртуальном уровне (в соответствии с политикой безопасности, определенной для данной системы), так и на физическом уровне, с точки зрения доступа пользователей в соответствующие вычислительные сегменты. Для систем, работающих с грифом информации «С» и выше, рабочие места должны подключаться к облаку или через выделенные или защищенные каналы передачи информации.

При таком подходе абонентами «облака» будут выступать подчиненные подразделения, подведомственные предприятия и институты, а также другие подразделения. Они будут являться как поставщиками информации в «облако», так и ее потребителями. Таким образом, в «облаке» будет интегрироваться вся информация, относящаяся к данному государственному органу. При этом физически информация будет размещаться на разных сегментах облака, но объединенных высокоскоростными каналами (за счет их локального компактного размещения). Такая

интеграция дает возможность значительно повысить аналитические возможности облака.

Динамическое информационное пространство

Для решения аналитических задач, с учетом физически разделенного хранения информации, предлагается построение динамического единого информационного пространства на виртуальном уровне (рис. 1). Такое пространство не будет аккумулировать всю информацию в одном месте. Информация будет храниться в исходных местах. Динамическое пространство будет выполнять роль посредника между ними и аналитическими сервисами, запрашивающими необходимую информацию. Решение данной задачи осложнятся следующим проблемами, которые возникают при интеграции информации из нескольких гетерогенных источников:

- неоднозначность – в разных системах одни и те же сущности имеют различные способы кодирования (например, номера КА), используют различные единицы измерений и т.д.
- дублирование – наличие одинаковой информации в разных источниках;
- противоречие – наличие различной информации в разных источниках об одном объекте или событии;
- разнородность информации – информация, хранящаяся в разных источниках, имеет разные форматы, принципы представления и хранения.

При создании классических хранилищ данных [] для решения данных проблем используются технологии извлечения, преобразования и загрузки данных (ETL – Extract Transform Load) [1, 2]. Стандартные ETL системы извлекают информацию из исходной базы данных, преобразуют ее в формат, поддерживаемый интегрированным хранилищем, а затем загружают в него преобразованную информацию. При этом конфигурация и настройки извлечения, преобразований и загрузки описываются заранее и хранятся в виде метаданных процесса трансформации. Они включают в себя следующее:

- конфигурация извлечения – список данных подлежащих извлечению, правила извлечения (например, порядок) и т.п.;
- конфигурация преобразований – алгоритмы преобразований (например: очистка, нормализация, пересчеты единиц измерений и т.п.), настройки преобразований и т.п.;

- конфигурация загрузки – правила сопоставления атрибутов исходной информации и атрибутов принимающего хранилища.

В отличие от стандартного ETL процесса для построения динамического информационного пространства предлагается замена последнего этапа, загрузки данных, на этап запроса данных. На этом этапе формируется запрос на получения данных в формате (в терминах) приемника. Данный запрос должен преобразовываться (переводиться) к запросу в терминах источника информации. Извлеченная в соответствии с данным запросом информация должна быть извлечена и преобразована в соответствии с заданными настройками, как и в стандартном процессе.

Процесс извлечения и преобразования будет описываться для каждого источника информации. Таким образом, работа динамического информационного пространства будет выглядеть, так как она представлена на рисунке 3.

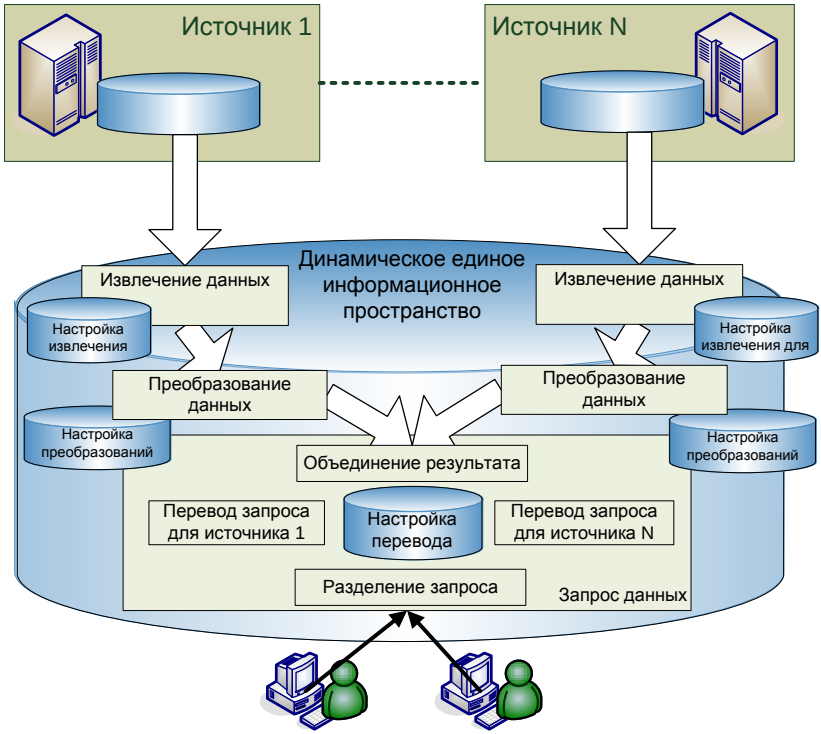

Рисунок 3. Архитектура единого-информационного пространства

Для работы с несколькими БД на этапе формирования запроса, должны выполняться следующие действия:

- разделение запроса – формирование на основе запроса пользователя нескольких запросов к нескольким источникам данных;
- перевод запроса с терминологии приемника на терминологию соответствующего источника;
- слияние результатов – объединение результатов полученных в результате извлечения и преобразования от каждого источника.

Динамическое единое информационное пространство будет включать в себя программное обеспечение выполняющее формирование запросов к источникам информации, извлекающее данные из них и преобразующее извлеченную информацию. Для каждого источника должны разрабатываться свои модули: извлечения данных, преобразования данных и перевода запросов. Они должны работать в соответствии с настройками заданными для этих источников. Модули разделения запросов и объединения результатов должны быть общими для всех источников информации и работать в соответствии с настройками перевода.

Основным недостатком такого подхода является низкая скорость выполнения запросов. Задержки возможны вследствие, преобразования запросов, преобразования данных, передачи данных по сетям. Однако в случае, когда источники информации размещаются в «облаке» данная проблема нивелируется.

Заключение

В статье описывается архитектура облачной платформы построенной на базе защищенных средств, что позволяет использовать ее в государственных структурах, включая Министерство обороны. Описывается подход виртуализации существующих у заказчика автоматизированных информационных систем и их размещение в облаке. Для выполнения аналитических задач над объединенной информацией предлагается создание динамического информационного пространства, осуществляющего трансформацию запросов пользователей к нему в запросы к исходным источникам информации. Это позволяет решить проблемы с разграничением хранения информации имеющей различные грифы, а также проблемы с актуализацией доступной информации. За счет размещения исходных информационных систем в едином вычислительном

кластере устраняется основной недостаток такого подхода – низкая скорость доступа к данным.

Список литературы

[1] Mell Peter, Grance Timothy. The NIST Definition of Cloud Computing. Recommendations of the National Institute of Standards and Technology. Computer Security Division Information Technology Laboratory National Institute of Standards and Technology. Gaithersburg. 2011
[2] Amazon Elastic MapReduce. Developer Guide. http://s3.amazonaws.com/awsdocs/ElasticMapReduce/latest/emr-dg.pdf
[3] Google Developers. Google BigQuery. https://developers.google.com/bigquery/what-is-bigquery?hl=ru.
[4] SQL Server Data Mining. Home. http://www.sqlserverdatamining.com/ssdm/
[5] Tom White. Hadoop: The Definitive Guide. T. White. USA.: O'Reilly Media, 2009. pp 526.
[6] Army Private Cloud (APC2) http://h10131.www1.hp.com/public/contract-vehicles/apc2/
[7] Национальная облачная платформа. https://www.o7.com/#main
[8] http://лаборатория50.рф
[9] http://www.seaproject.ru/products/seacloud
[10] И.И. Холод. Архитектура «облака» интеллектуального анализа данных на основе библиотеки алгоритмов с блочной структурой. ИЗВЕСТИЯ СПбГЭТУ «ЛЭТИ» 2014 г., № 6, С. 34-40

Об авторе:

Холод Иван Иванович

Заместитель главного конструктора АО "НИЦ СПб ЭТУ», к.т.н., доцент кафедры вычислительная техника Санкт Петербургского государственного электротехнического университета «ЛЭТИ» им. В.И. Ульянова (Ленина)

e-mail: iiholod@mail.ru