

В.И. Гриценко, А.А. Урсатьев

Информационные технологии: тенденция, пути развития

Показано воздействие *Business Intelligence* на информационные технологии такие, как технологии предоставления услуг на основе сервисной модели в среде *Grid u Cloud Computing*. Рассмотрен термин «облачный компьютеринг, вычисления в облаке, или облачный сервис».

The influence of *Business Intelligence* on the information technologies, such as the provision of services based on a service model in the *Grid and Cloud Computing* environment is shown. The term «cloud computing, calculation in a cloud, or a cloud service» is considered.

Показано вплив *Business Intelligence* на інформаційні технології такі, як технології надання послуг на основі сервісної моделі в середовищі *Grid* та *Cloud Computing*. Розглянуто термін «хмарний комп'ютеринг, обчислення в хмарі, або сервіс у хмарах».

Введение. К информационным технологиям (ИТ) авторы обращались [1] в период становления информационного общества, когда вопросы взаимодействия информационных ресурсов в гетерогенной распределенной среде были крайне актуальны, а требования к современным ИТ в части интероперабельности особенно обсуждаемы. За прошедшее время произошло переосмысление *Web*-технологии и завершилось ее развитие от механизма предоставления информации до оказания разного рода услуг как то доставка программ конечному пользователю, платформ для желающих создавать приложения, инфраструктуры и т.п. Возник интерес к аналитическим информационным системам (ИС), способным работать в широком информационном пространстве с данными, поступающими из разных источников.

***Business Intelligence* – эволюция, новые ИТ**

То, что в настоящее время в зарубежных и русскоязычных научно-технических изданиях уделяется достаточно большое внимание системам хранения данных, хранилищам данных, в том числе приближенных к реальному времени (активные хранилища), сопутствующим им ИТ, обусловлено, по крайней мере, двумя причинами. Первая из них – прогноз неукоснительного роста производимой мировым сообществом (промышленность, экономика, бизнес и др.), обрабатываемой и нуждающейся в хранении информации, и нежелание теперь в первую очередь бизнеса, наряду с научными и отраслевыми про-

изводствами, допускать любую потерю данных [2, 3]. Вторая – *Business Intelligence* и рост понимания того, что в условиях неустойчивой рыночной конъюнктуры и обостряющейся конкуренции, присущих мировому рынку, бизнес-анализ не может замыкаться в рамках одной организации, а по общему признанию должен опираться на информационное поле, представленное документами различного вида и образованное территориально или географически распределенными корпоративными данными, а также принимать во внимание данные во всем доступном временном срезе.

Приведем несколько примеров. В отраслях (нефтегазовый сектор, электроэнергетика и др.) наблюдается возрастание требований, предъявляемых к достоверности, оперативности сбора и обработки информации, на основе которой принимаются стратегические бизнес-решения. Сведения, необходимые для эффективного решения управленческих задач (данные о клиентах, продуктах и текущих маркетинговых мероприятиях) хранятся в многочисленных разрозненных информационных системах, а отсутствие связей между ними и различные маркировки товаров и услуг приводят к тому, что корпоративным аналитикам приходится вручную сопоставлять и сравнивать данные. При составлении отчетов о расходах менеджеры компании сталкиваются со сходными проблемами, например с отличием в системах кодирования затрат по филиалам концерна [4].

Для преодоления упомянутых сложностей было создано несколько объединенных (*federated*) хранилищ данных [1, 5], куда загружалась информация о продуктах, клиентах и продажах, поступающая от всех отделений корпорации. Менеджеры, ответственные за общее развитие бизнеса, смогли легко определить прибыли и объемы продаж для каждого продукта и для любого региона. Кроме того, представилась возможность отслеживать тенденции, корректировать текущие маркетинговые кампании и планировать новые. Хранилища также использовались для данных, которые ежемесячно выгружались из локальных систем. Они предназначались для заполнения матрицы глобальных издержек, используемой для повышения скорости составления отчетности о расходах, усиления прозрачности и выявления возможных путей сокращения затрат. Объединяя информацию о расходах путем преобразования локальных диаграмм отчетности в глобальную структуру, издержки сравнивались по странам и регионам, а также сопоставлялись с результатами деятельности и выявлялось распределение затрат [5].

До внедрения технологии федеративных хранилищ данных подобные сравнения были сопряжены с существенными сложностями: требовались значительные усилия по сбору информации, а собственно процесс был довольно продолжительным, в результате чего к моменту проведения анализа многие данные устаревали. Теперь для получения информации по каждой стране потребовалось от двух до трех дней, а технология хранилищ данных позволила упростить и ускорить сопоставление данных [5].

Пример из области электроэнергетики [5]. Процесс торговли электроэнергией во многом схож с функционированием фондовой биржи: каждые полчаса производители энергоносителей должны сообщать на центральный коммутатор, по какой цене и в каком количестве они готовы поставлять свою продукцию. Если продукт предлагается по слишком высокой цене, – товар не будет пользоваться спросом; по слишком низкой – чревато убытками. Таким образом, определение конкурентоспособной цены было задачей первостепенной важности. Процесс принятия решения осложнялся тем, что ежедневно

все электростанции получали огромное количество сообщений: более четырех миллионов информационных единиц в день. В хранилище эти данные должны находиться в течение семи лет.

Доступ к данным в реализованном хранилище осуществлялся через корпоративный интранет – работники компании, загрузившись на внутренние *Web*-страницы, получали информацию об объеме текущего и планируемого производства, а также о ценах. Кроме того, пользователи могли обмениваться результатами анализа данных, отправляя их по внутренней сети. С помощью хранилища данных практически мгновенно – каждые пять минут – отслеживались малейшие изменения цен на рынке.

Следует обратить внимание в обзоре [5] на ключевые слова «технология хранилищ данных и интегрируемые решения *Business Intelligence*», которые в обоих примерах позволяли успешно решать различные задачи, стоящие перед отраслями производства.

Уместно также сослаться на работу [6], в которой, в частности, говорится, что упустив признаки экономического спада, возможно, из-за охвата лишь ограниченного объема данных, аналитики потребовали качественно новых инструментов, способных решать такие глобальные задачи, как выявление нерегулярных тенденций в мировой и региональной экономике, бизнесе и на потребительских рынках. По мнению ряда аналитиков, к 2012 году более 35 процентов из пяти тысяч крупнейших компаний мира будут регулярно терпеть убытки из-за отсутствия своевременного доступа к актуальным данным и ошибочных прогнозов, вызванных работой с недостаточным объемом информации.

Так что же такое *Business Intelligence (BI)*? Однозначного четкого определения *BI* не существует [7–9]. Приведем некоторые варианты. Это:

- анализ данных, связанных исключительно с финансовой деятельностью предприятия;
- информационное обеспечение бизнеса, причем в самом широком смысле – интеллект, разведка¹ и другое, в общем, все то, что нужно для принятия решений;

¹ Разведка или «добыча» данных (*data mining*) представляет собой процесс обнаружения корреляции, тенден-

- все известные методы анализа от подготовки отчетов до поиска текстов и данных, область действия которых выходит за рамки финансовой аналитики;

- отождествление с системами поддержки принятия решений *DSS*¹ (*decision support system*);

- толкование данных, подготовка отчетов, построение моделей и выработка рекомендаций и прогнозов;

- знания, добытые о бизнесе с использованием различных аппаратно-программных технологий, позволяющие превращать данные в информацию, а затем информацию – в знания.

Из приведенного в [8] определения *BI* на него возлагаются задачи агрегации, интеграции и интерпретации данных из разнородных источников, с тем, чтобы превратить их в удобную для принятия решений информацию. Этот вывод сделан на основе анализа процесса управления бизнесом в контексте окружения для *BI*:

- обнаружение и детализация проблемы;
- постановка задачи и идентификация данных, необходимых для ее решения;

- анализ данных, в том числе пробный анализ *EDA* (*exploratory data analysis*), анализ, позволяющий построить гипотезы *SDA* (*structured data analysis*) и определить потоки данных, а также другие виды анализа;

- *BI* – толкование данных, подготовка отчетов, построение моделей и выработка рекомендаций и прогнозов;

- принятие решений.

Определение *BI* «данные – информация – знания» [7] привлекает своей завершенностью и логикой. Здесь данные воспринимаются как ре-

ций, шаблонов, связей и категорий. Она выполняется путем тщательного исследования данных с привлечением статистических и математических методов, предназначенных для нахождения представлений, интуитивно понятных специалистам, знающим бизнес-процессы и которые могут предсказать результат или значение определенных ситуаций, используя исторические или субъективные данные [7, 10].

¹ *DSS* основана на технологиях хранилища данных и *BI*. Система определяет методы и средства доступа и оперативного анализа информации в терминах предметной области [7].

альность, которую следует регистрировать, хранить и обрабатывать – это «сырые данные». Информация – то, что человек в состоянии понять о реальности, а добытые знания используются для принятия решений. В процессе организации информации для получения знания необходимо применять хранилища данных, а для представления этого знания пользователям необходимы инструменты *BI*. Информация сама по себе не очень подходит для принятия решений в виду ее значительного и возрастающего с каждым годом объема. Средства бизнес-интеллекта и хранилищ данных призваны находить в данных и информации то существенное, что реально присовокупляется к нашим полезным знаниям. Они не заменяют человека, а используют для формирования гипотез интуицию, основанную на его подсознании и личном опыте. В широком смысле слова *BI* определяет:

- процесс превращения данных в информацию и знания о бизнесе для поддержки принятия улучшенных и неформальных решений;

- информационные технологии (методы и средства) сбора данных, консолидации информации и обеспечения доступа бизнес-пользователей к знаниям;

- знания о бизнесе, добытые в результате углубленного анализа детальных данных и консолидированной информации.

В основе технологии *BI* лежит организация доступа конечных пользователей и анализ структурированных количественных по своей природе данных и информации о бизнесе. *BI* порождает итерационный процесс бизнес-пользователя, обеспечивающий доступ к данным и их анализ, а значит, проявление интуиции, формирование заключений, нахождение взаимосвязей с целью эффективных положительных изменений предприятия.

Архитектура *BI*-технологии состоит из инфраструктуры и прикладных сервисов. Инфраструктурный слой включает в себя информационные ресурсы и сети. На этом слое данные собираются, интегрируются и становятся доступными. Хранилище данных – один из возможных компонентов инфраструктурного слоя. *BI*-средства не обязательно должны работать в ин-

фраструктуре хранилища данных, но в этом случае проблема очистки и согласования данных возлагается именно на них (*BI*). Если данные должным образом не преобразованы, не очищены и не консолидированы, то никакие возможности *BI*-инструментов или приложений не смогут повысить их достоверность. Проблемы могут возникнуть из-за несогласованности метаданных. В рамках большой корпорации эти вопросы решаются на инфраструктурном уровне путем создания корпоративного хранилища данных и централизованного управления метаданными. Для использования *BI* в оперативных системах может потребоваться оперативный склад данных *ODS* (*operational data store*).

Прикладные сервисы, ответственные за функциональность, включают в себя все *BI*-сервисы, такие как механизмы запросов, анализа, генерации отчетов и визуализации, а также средства безопасности и метаданные [7]. В широко распространенном варианте *BI* (доступ к данным, располагаемым в хранилищах, генерация запросов и отчетов, отчетность и оперативная¹ аналитическая обработка *OLAP* (*online analytical processing*))² (рис. 1) эффективность принимаемых решений в немалой степени зависит от способности ряда функциональных подсистем предприятия, таких как системы управления отношениями с клиентами *CRM* (*customer relationship management*), управления внутренними и внешними ресурсами предприятия *ERP* (*enterprise resource planning*), управления эффективностью бизнеса *CPM* (*corporate performance management*) и других, взаимодействовать между собой. Призванные решить

¹ Оперативность анализа не связана с реальным временем обновления данных в системе. Эта характеристика относится к времени реакции *OLAP*-системы на запросы пользователей.

² В состав *BI* могут входить следующие средства анализа: пакеты статистического анализа, анализ временных рядов и оценки рисков; средства моделирования; пакеты для нейронных сетей; средства нечеткой логики и экспертные системы. Также входят средства для графического оформления результатов: средства деловой и научно-технической графики; «приборные доски», средства аналитической картографии и топологических карт; средства визуализации многомерных данных [7].

проблему интероперабельности³ программные продукты, реализующие интегрированную информационную среду предприятия *EAI* (*enterprise application integration*) путем объединения функциональных подсистем, баз данных, хранилищ данных и прочего, в большинстве случаев – частные решения. Подсистемы работают на собственном наборе данных, не могут обмениваться данными в режиме реального времени и не образуют единую систему [11].

Основные инструменты *BI* [7] все еще остаются наиболее распространенными, удовлетворяя большинство потребностей. Однако *BI* (см. рис. 1.) опирается, как правило, на хранилища данных, обеспечивающие доступ к уже использованному оперативным данным для «исторического анализа» и генерации отчетов. Помимо этого хранилища используют для получения данных ограниченное число бизнес-процессов и систем, а также поддерживают только структурированные данные. В своих требованиях к *BI*-системам все чаще высказывается желание работать со «свежими данными», в дополнение к традиционным стратегическим решениям. С одной стороны, регламент выполнения *ETL*-процессов заполнения хранилища огромными объемами данных⁴, с другой – бизнес больше не хочет ждать, пока данные поступят в хранили-

³ Понятие интероперабельности здесь употреблено в широком смысле слова: техническая интероперабельность – это совместимость систем на техническом уровне, включая протоколы передачи данных и форматы их представления; семантическая интероперабельность – свойство информационных систем, обеспечивающее взаимную употребимость полученной информации на основе общего понимания системами ее значения [12].

⁴ *OLAP*-инструменты, как правило, работают с многомерными БД, время формирования которой существенно зависит от объема загружаемых в нее данных. Несмотря на то, что многомерные БД (МБД) являются наиболее подходящими для оперативной аналитической обработки, сейчас эту возможность встраивают в реляционные СУБД или расширяют их *OLAP* для реляционных БД (*ROLAP*). Последние имеют преимущество по масштабируемости и гибкости, но проигрывает по производительности многомерному *OLAP* (*MOLAP*). Для поддержки МБД используются *OLAP*-серверы [13], оптимизированные для многомерного анализа и поставляемые с аналитическими возможностями.

ще, а затем появятся в отчетах. Теперь, как никогда, возрастает значение своевременности предоставления информации. Временной лаг в несколько часов или дней для помещения оперативных данных в хранилище с целью их последующего анализа в определенных ситуациях уже считается недопустимым – полная и точная информационная картина сегодня нужна в реальном времени непосредственно в ходе выполнения бизнес-процесса. Поэтому все более актуальны активные хранилища [9, 14], или хранилища «реального» времени, позволяющие эффективно решать задачи оперативной отчетности и аналитики «реального» времени.

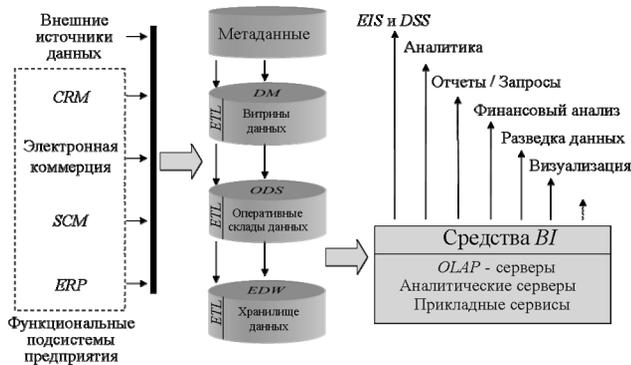


Рис. 1. Схема превращения данных в информационные продукты: *EIS* (executive information systems) – информационная система руководителя; *ETL* (extraction, transformation and loading) – технология извлечения, преобразования и загрузки служат для перегрузки данных из одних программных компонентов в другие, включая их очистку на этапе загрузки в хранилище (выявление дубликатов, ошибочных данных), согласование данных, получаемых из различных источников, трансформирование и др.; *DM* (data marts) хранят агрегированные, необходимые для аналитики оперативные данные; *ODS* (operational data store) – обеспечивает выполнение анализа данных сразу после ее обновления в функциональных подсистемах

Стратегия интеграции хранилищ. Информационное пространство

Эффективность хранилища зависит не только от возможности аккумулировать все данные предприятия в одном месте, а от способности предоставлять их всем заинтересованным подразделениям организации. Это, в первую очередь, стратегия интеграции хранилища данных в корпоративную архитектуру организации, чтобы было доступно обрабатывать важную для принятия решения информацию. Эффективное использование информации из активного храни-

лища для поддержки принятия тактических решений возможно, когда средства бизнес-анализа тесно взаимодействуют с приложениями оперативной обработки транзакций *OLTP* (online transaction processing) в рамках всего предприятия. Это позволит не только улучшить собственно процесс получения данных, но и обеспечить принятие решений практически в режиме реального времени. В момент регистрации бизнес-события в системе финансовых расчетов среда принятия решений (активное хранилище данных) уведомляется об этом событии (рис. 2).



Рис. 2. Активное хранилище и интеграция корпоративных приложений

Когда аналитические приложения, поддерживающие процесс принятия тактических решений, определяют необходимость предпринять некие действия, ставится в известность система *OLTP*, ответственная за соответствующие финансовые операции. Интеграция этих процессов дает возможность организовать «замкнутый цикл» принятия решений. Данные, передаваемые из среды финансовых расчетов в активное хранилище, будут инициировать решения, передаваемые обратно в оперативные системы финансовых расчетов для выполнения [14].

Системы *BI* традиционно взаимодействовали со структурированными данными из относительно ограниченного пула корпоративных данных, что существенно сужало область действия поиска информации для принятия решений. По оценке [15] методы *BI*, основанные на обработке структурированных данных из хранилищ, позволяют использовать не более 20 про-

центров от общего количества всех доступных данных. Очевидно, что решения, принимаемые в современном бизнесе, должны базироваться на более широком информационном поле, учитывающем не только многочисленные неструктурированные данные (текстовые документы, электронные письма, *Web*-страницы, мультимедиа и др.), но и хранилища и источники данных вне компаний. Это позволит аналитикам существенно расширить кругозор и перейти от обработки статистики к выявлению тенденций. Распределенность информации по множеству различных репозитариев, включая транзакционные базы данных, хранилища, библиотеки документов, папки электронной почты и прочее, где данные хранятся в разных форматах и привязаны к определенным приложениям или бизнес-процессам, также ограничивают компании в их возможностях анализа информации. Однако для расширения сферы влияния *BI* на все доступное информационное поле одной какой-либо системы, даже корпоративного уровня, недостаточно.

Сегодня формируется новое поколение архитектуры хранилищ данных для поддержания оперативного бизнес-анализа. *IBM* и *Microsoft* реализуют стратегии интеграции программных средств бизнес-интеллекта, ставя своей целью создание нового поколения ПО, которое будет обрабатывать как структурированные, так и неструктурированные данные [9]. В распоряжении компании *IDS* имеется фундаментальная технология, позволяющая переносить неструктурированные данные в структурированную среду и осуществлять их интеграцию в рамках этой среды [16]. На основе усовершенствованной версии *DB2 Warehouse* [16] *IBM* представила свою стратегию создания динамических информационных хранилищ (*Dynamic Warehousing*) [17], которая позволит анализировать информацию не только в ее «историческом прошлом», но и в ее актуальном настоящем, получая точные оперативные данные из самых разных источников в контексте текущей бизнес-операции и извлекая из них глубинную для бизнеса суть. Динамическое информационное хранилище определяется следующими ключевыми моментами:

- возможностью извлекать и использовать знания из неструктурированной информации;
- поддержкой доступа в реальном времени к агрегированной, очищенной информации;
- встроенной аналитикой, которую можно использовать в рамках бизнес-процесса;
- широким набором интегрированных средств реализации концепции «информация по требованию» (*information on demand*).

Динамическое хранилище данных представляет собой интегрированную с ядром *DB2* среду, расширяющую возможности собственно хранилища и предоставляющую информационные сервисы. С одной стороны – это объединение функциональности сервера БД, оптимизированного для оперативной обработки транзакций, и специализированных возможностей современного хранилища данных, в частности архитектуры *Shared Nothing* (без разделения), в которой для выполнения запроса назначается выделенный пул виртуальных ресурсов, средств поддержки масштабируемости и высокой производительности, а также встроенной аналитики. С другой – инфраструктура интегрированных информационных сервисов, таких как текстовый поиск и анализ, интеграция информации, управление процессами и др. (рис. 3).



Рис.3. Упрощенная архитектура динамического хранилища данных

Так в систему интегрирован инструментарий перемещения и трансформации данных, предназначенный для их загрузки в хранилище и подготовки для дальнейшей аналитической обработки, что снижает сложности и расходы, обычно связанные с выполнением этих операций.

Следующий уровень образуют средства оптимизации производительности хранилища: разбиение данных на разделы (*data partitioning*) в рамках одного сервера или кластера, обеспечивающее линейную масштабируемость хранилища; управление рабочей нагрузкой, гарантирующее высокий приоритет запросам наиболее критичных приложений; сжатие данных, позволяющее оптимизировать использование дискового пространства и повысить скорость обработки запросов.

На верхнем уровне хранилища находятся встроенные средства *BI*, которые можно использовать как в составе приложений (*in-line analytics*), так и в качестве отдельного сервиса. Все операции бизнес-аналитики выполняются непосредственно в хранилище, исключая необходимость в экспорте данных для анализа. Встроенные аналитические средства не заменяют высокоуровневый *BI*-инструментарий, который может быть использован на предприятии, но дополняет его, предоставляя возможность выполнять операции по разведке (добыче) и анализу данных непосредственно в контексте текущих бизнес-операций и приложений.

Информационный сервис текстового поиска и анализа (см. рис. 3), направленный на выявление полезных знаний из неструктурированной информации (электронная почта, комментарии и иллюстративный материал, голосовые и текстовые сообщения коммуникационных систем), выполнен на платформе интеллектуальной обработки контента (*omnifind analytics edition*). Последняя предоставляет полнофункциональный интерфейс, сочетающий в себе средства поиска, текстового анализа и визуализации и позволяет выявлять значимые для бизнеса данные из взаимосвязанной структурированной и неструктурированной информации. Эти возможности могут быть использованы для контроля над ситуацией в бизнесе, управления событиями и отслеживания изменений, имеющих тенденцию к развитию во времени.

Интеграция информации – средства агрегирования, очистки и преобразования информации из различных источников, а также предоставление этой информации в качестве сервиса

(*IBM Information Server*); управление процессами – средства извлечения информации в контексте выполняемых действий бизнес-процесса и ее анализ для совершенствования принятия решений и оптимизации процессов (программное обеспечение управления бизнес-процессами семейств *IBM FileNet u WebSphere*); специализированные по отраслям бизнес-модели – для повышения эффективности аналитики в конкретных индустриях (*IBM industry data models*).

Таким образом, сохраняя традиционные функции агрегирования и очистки операционных данных для глубокого исторического анализа и выявления проблем, возможностей и тенденций, современные хранилища данных должны в реальном масштабе времени обеспечивать бизнес-приложения актуальной аналитической информацией и уметь работать с разными типами информации, в том числе неструктурированной. Фактически от современного хранилища требуется поддержка одновременно совершенно разных типов рабочих нагрузок, включая выполнение запросов критичных приложений, требующих немедленного отклика, формирование традиционных отчетов для решения тактических и стратегических задач, обработку эпизодически возникающих непредсказуемых запросов от приложений, поддерживающих различные этапы бизнес-процессов, а также сопровождение новых аналитических потребностей традиционных транзакционных систем. Из этого следует, что хранилище данных должно стать динамичным, способным обеспечивать выполнение задач разных типов и различных требований к уровню обслуживания [17].

Майкрософт предлагает недорогую платформу для *BI* на базе *SQL Server, SharePoint* и *Office*. Первый включает в себя средства *ETL*, генерации отчетов, аналитики и разведки данных, *SharePoint* – приборные панели (*dashboard*), средства для демонстрации показателей (*scorecard*) и *Power Pivot* позволяет собирать данные из разных источников.

Около 80 процентов всех усовершенствований платформы *SQL Server 2008 R2* связано с бизнес-аналитикой. В то же время часть обновлений платформы нацелена на повышение мас-

штабируемости, эффективности и управляемости решений. Один экземпляр сервера теперь может использовать до 256 логических процессоров и до 2 Тбайт оперативной памяти (ранее можно было одновременно задействовать только 64 процессора). Улучшена работа *SQL Server* в виртуализованных средах под управлением гипервизора *Microsoft Hyper-V*. Виртуализованный экземпляр программы теперь можно непосредственно во время работы переместить с одного физического сервера на другой. Обновлены средства администрирования СУБД. Новая консоль *Utility Control Point* позволяет из одной точки управлять политиками группы серверов.

Важным новшеством по эффективности аналитических решений стала технология *StreamInsight*. Суть ее заключается в том, что она позволяет обрабатывать и анализировать поступающие данные сразу, без предварительной очистки и загрузки в хранилище. Такое свойство технологии *StreamInsight* обеспечивает ей широкую востребованность в телекоммуникационной сфере, где операторам необходимо анализировать каждое тарифицируемое событие в сети, чтобы исключить возможную утечку доходов. Специальные программные фильтры выделяют из потока данных только значимые для конкретного бизнес-процесса сведения, что дает возможность минимизировать объем хранимой исторической информации без потери качества, сокращает время выявления различных закономерностей и повышает скорость принятия решений.

Ключевой особенностью платформы *Microsoft SQL Server 2008* станут мощные встроенные инструменты *BI*, расширяющие возможности привычных для аналитиков продуктов *Microsoft Excel 2010* и *SharePoint Server 2010*. Встроенный в платформу *OLAP*-сервер с развитыми возможностями многомерного анализа данных может обрабатывать информацию в реальном времени и обеспечивает логическое представление данных в бизнес-терминах. Также в нем имеется набор интеллектуальных алгоритмов для задач прогнозирования и выявления скрытых закономерностей.

Модуль *Power Pivot for Excel 2010* устанавливается как надстройка к электронным табли-

цам и обеспечивает загрузку в *Excel* неструктурированных данных из разных источников и приложений. Пользователь сможет самостоятельно описывать эти данные в различных бизнес-терминах, задавая между ними внутренние связи, добавляя собственную информацию и расчетные формулы. Этот инструмент позволяет *Excel* при помощи *SQL Server* работать с таблицами в миллионы строк.

Подсистема *MDS (master data services)* выполняет управление структурой данных, реализует подсистему бизнес-правил, отслеживает версии данных и выполняет их защиту от несанкционированного изменения, а также позволяет наполнять справочники данных в ручном режиме [18, 19].

Технологии предоставления услуг

Итак, не имеющие доступа к распределенным корпоративным данным изолированные хранилища сегодня не актуальны. Будущее за интегрированными хранилищами, в реальном времени отслеживающими появление бизнес-событий и тесно взаимодействующими с системами принятия решений. С другой стороны – весьма значимые затраты на реализацию такого проекта (распределенное хранилище данных, средства *BI*, персонал ИТ-подразделения и др.) могут привести к тому, что совокупная стоимость владения, скорее всего, будет доступна только для крупных производств (компаний). Предприятиям среднего и малого бизнеса сделать доступными системы бизнес-аналитики с обязательными для них хранилищами данных можно с использованием технологии предоставления услуг аналогично тому, как это было реализовано в ВЦКП (вычислительный центр коллективного пользования). Технология, по которой пользователю с удаленного терминала предоставлялся доступ не только к вычислительным мощностям, но и к накапливаемой в сетях разнообразной информации, распределенной по всем ЭВМ сети. Доступ в подобные распределенные базы (банки) данных осуществлялся сетевыми СУБД. В общем случае пользователь получал и обрабатывал на ВЦ требуемую информацию [20, 21]. Новый виток спирали кардинально отличается сетями и средствами телекоммуни-

каций, имеющихся в бытность ВЦКП, да и концепция предлагаемых по сети услуг иная [12]. Она основана на сервис-ориентированной архитектуре (SOA), предусматривающей структуру из трех элементов (рис. 4): провайдеры служб, размещающие информацию о своих услугах, потребители услуг и реестр услуг. Взаимодействие между различными приложениями и технологиями взял на себя *Web* (рис. 5). *Web*-службы (*Web-service*)¹ [12, 22] – это набор основанных на *XML* спецификаций, обеспечивающих универсальный метод технического описания услуг и взаимодействия с ними.

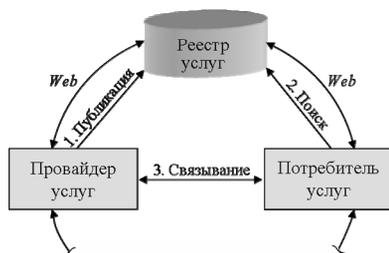


Рис.4. Сервис-ориентированная архитектура

Основная рекомендация *SOA* – слабая связанность² – придает вновь разрабатываемому ПО способность к быстрой адаптации приложений в непредусмотренных условиях и зара-

нее не оговоренными группами пользователей. Определяющим фактором адаптивности будет встроенная готовность компонентов информационных систем к взаимодействию с другими приложениями – к предоставлению и потреблению электронных сервисов машинного уровня [12, 23, 24]. Физически *Web*-служба представляет собой фрагмент программного обеспечения, называемый *агентом*. Реализуя абстрактную функциональность услуги, агент способен передавать и принимать сообщения. Один и тот же сервис может быть обеспечен разными агентами. Программно-стандартизованный способ межмашинного взаимодействия в сети распределенных гетерогенных систем реализуется на основе открытых стандартов: *SOAP*, *WSDL* и *UDDI* [11, 26, 27]. Платформа *Web*-служб, основанная на стеке этих стандартов, разделяется на три части – коммуникационные протоколы, описания служб и поиск услуг (см. рис. 5).

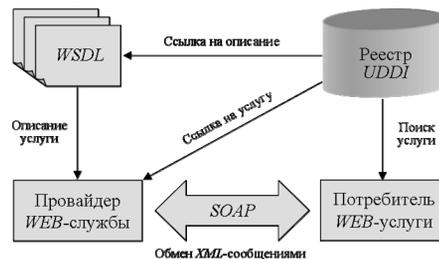


Рис. 5. *WEB*-служба:

UDDI (*universal description, discovery and integration* – универсальное описание, информация и интеграция) – глобальный каталог – регистр описаний услуг, предоставляемых *Web*-службой; *WSDL* (*web services description language*) – язык формата *XML*, описывает интерфейс *Web*-услуг в виде абстрактного определения в терминах сообщений, обмен которыми происходит при обращении к службе, т.е. раскрывает формализованное взаимодействие клиент-служба. *SOAP* (*simple object access protocol* – протокол доступа к объектам) – использует *XML* как способ представления информации для обмена сообщениями между ИС, корпоративными приложениями и др.

Опираясь на *SOA* и *Web*-службы созданы такие размещаемые прикладные услуги по требованию (запросу): ПО как услуга *SaaS*³ (*software-as-a-service*) или ПО по требованию *SoD* (*software on demand*); динамическая доставка данных по подписке *DaaS* (*data-as-a-service*); *DISaaS*

¹ Присутствие *Web* в этом термине неубедительно объясняется использованием *HTTP* в качестве протокола транспорта сообщений между услугой и его потребителем; иной связи с *Web* нет.

² Благодаря требованию слабой связанности, услуги обретают мобильность, способность перемещаться с одного сервера на другой, не требуя согласования и координации со всеми потребителями. Достигается это гибкостью интерфейса – ключевого элемента *SOA*. Эти свойства интерфейса делают его нейтральным к специфике реализации услуги, которая определяется аппаратной платформой, ОС и языком программирования. Подобный нейтралитет обеспечивает универсальность взаимодействия в разнородной среде, а услуги, интегрированные посредством таких интерфейсов, слабо связаны (*loose coupling*). Поскольку слабая связанность существенно снижает необходимость ручной координации изменений в информационной инфраструктуре со всеми взаимодействующими сторонами, сама возможность многосторонней интеграции становится вероятнее [25].

³ Устоявшийся термин *SaaS IDC* предлагает переименовать в *AaaS* (*application as-a-service*).

(*data integration system as a service*) – интеграция данных как сервис; *PaaS* – услуга, позволяющая разрабатывать, тестировать и внедрять пользовательские приложения; *IaaS* – услуга, которая обычно предоставляет унифицированные аппаратные и программные ресурсы, в некоторых случаях и на инфраструктурном уровне для установки ПО с оплатой по мере использования – доступ к ресурсам по модели временной оплаты (*pay-as-you-go*) и др. Помимо этих услуг могут быть доступны и востребованы *IdaaS* (*identification as service* – управление идентификацией), *HPaaS* (*high performance as a service* – высокопроизводительные вычисления) и другое специализированное ПО.

В основе *SaaS* лежит принцип подписки: программное обеспечение не продается как продукт, а предоставляется в аренду, оплата зависит от числа пользователей, объема транзакций и других количественных показателей. Возможность взять в аренду в первую очередь снимает вопрос о необходимости капитальных инвестиций в инфраструктуру – покупку оборудования, а чтобы воспользоваться *SaaS*, не нужно иметь ничего, кроме инфраструктуры, обеспечивающей доступ в *Internet*. Главные отличительные свойства модели *SaaS* таковы:

- ПО работает на стороне провайдера;
- условия использования сочетают в себе правила лицензирования¹ и хостинга;
- доступ к программе осуществляется через любой браузер или тонкий клиент;
- программа подстраивается под специфические требования пользователя, а не однократно конфигурируется специальным способом.

Таким образом, модель *SaaS*² имеет очевидные преимущества перед классической моде-

¹ В отличие от традиционного лицензирования программных продуктов, за «программы по запросу» клиент платит по принципу подписки, обязуясь вносить регулярные платежи. Считается, что основное преимущество *SaaS* – низкая стоимость владения и запуска в эксплуатацию.

² По данным работы [29], в зависимости от количества пользователей и вида ПО можно сэкономить до 85% совокупной стоимости владения решением в течение трех лет. Также услуга дает возможность перевести часть затрат на ИТ из капитальных в операционные.

лью распространения корпоративного ПО: она экономичнее, не требует установки и последующей поддержки ПО в среде обитания пользователя, вложения значительных средств защиты данных, поиск и содержание ИТ-персонала, который будет поддерживать и развивать инфраструктуру. Поэтому необходимость оптимизации бизнеса и, как следствие, поиска новых решений в среде территориально и географически распределенных корпоративных данных стимулировала интерес к доступным без больших инвестиций аналитическим информационным системам, позволяющим работать с разнообразными платформами и устройствами и обрабатывать большие объемы оперативных данных из самых разных источников. С ориентацией не в последнюю очередь на средний и малый бизнес³ предложено использовать бизнес-аналитику, ранее применяемую только крупными компаниями, по запросу – *Business Intelligence as a Service (BlaaS)* или *BI in the Cloud, BI Service, Grid Computing BI, On-Demand BI, SaaS BI*, что собственно одно и то же. Аналитика по запросу означает относительную простоту использования, сокращение затрат и сроков внедрения [9, 28].

Не вдаваясь в подробности об инструментарии *BI*-по требованию, отметим лишь, что *BlaaS* прогнозируют развитие в области *Web*- и финансовой аналитики, анализа рисков и угроз мошенничества, социальных сетей *SNA (social network analysis)*, отражающих множество мнений и позволяющих использовать эффект «мудрости толпы», событийного маркетинга и сравнительного анализа производительности. Вполне возможно, что будущее *BlaaS* связано с моделью динамической доставки данных по подписке *DaaS* с рынка данных. Едино в одном – наибольший эффект достигается при интеграции с традиционными инструментами бизнес-анализа [9].

Наличие современных технологий (*SOA* – *Web*-службы – *SaaS*), высокоскоростных кана-

³ Рассмотреть приобретение ПО по подписке в России готовы 78% компаний малого бизнеса, поскольку это дешевле (36%), позволит сфокусироваться на основном бизнесе (31%) и повысит надежность ИТ (30%) [29].

лов связи и высокопроизводительных коммуникаций¹ позволило выйти за рамки высокоэнергоемких ВЦКП или, как теперь принято говорить, центров обработки данных (ЦОД) и использовать компьютеры Глобальной сети наподобие того, как это было сделано в *Grid*. Слово «наподобие» употребили потому, что *Grid*² это физическая среда на физических машинах, предназначенная для решения задач конкретного класса, а технологии предоставления ПО либо иных услуг по требованию (*on demand*) имеют более широкий спектр приложений.

Сравнивая *Grid* [11, 31] с *WWW*, можно отметить, что *Grid* выходит за рамки совместного пользования информацией, предоставляя возможности распределенных вычислительных мощностей и систем хранения данных. Изначально *Grid*³ использовался академическими струк-

¹ К примеру, коммутирующая платформа *Arista 7500* имеет наивысшую производительность из всех современных решений, поддерживающих 10 *Gigabit Ethernet*. В *Arista 7500* имеется 384 порта, к которым можно подключать серверы и накопители *iSCSI*; скорость передачи – 5,76 млрд пакетов в секунду. Реализован виртуальный выход очередей *Virtual Output Queueing (VOQ)*, поддерживающий отдельные очереди для каждого физического выхода, что исключает возможность столкновения и взаимных блокировок потоков данных. Скорость коммутации *Arista 7500 VOQ* составляет 1,25 Тбит/с, что дает потенциал для увеличения скорости работы портов до 40 и 100 Гбайт/с. На *Arista 7500* работает специализированная сетевая ОС *EOS*, ее отличительная особенность от обычных серверных ОС в том, что она не монолитна. В *Arista EOS* – архитектура со множеством разделяемых процессов – состояние сети отделено от работы процессов, что позволяет без нарушения работоспособности системы в целом осуществлять восстановление после сбоев, обновление ПО и средств обеспечения безопасности. В дополнение к этому протокол процесса, функции управления и драйверы устройств могут работать не в ядре, а в пользовательском адресном пространстве. В результате существенно повышается стабильность работы ядра (стандартное ядро *Linux*) и открывается возможность для расширения функциональности ОС и обеспечения надежности коммуникаций [30].

² *Grid*-компьютинг: отдельный компьютер становится частью виртуальной вычислительной среды, полностью ретворяясь в глобальной ИТ-инфраструктуре [31].

³ *Globus Toolkit*, свободно распространяемый инструментарий, – фактический стандарт конструирования *Grid*-систем.

турами в задачах распределенной аналитической обработки больших объемов информации. Поддержка приложений электронного бизнеса (моделирование в области дизайна, сборки и управления жизненным циклом производства таких сложных объектов, как самолеты или автомобили, моделирование финансовых ситуаций и др.) расширила сферу применения *Grid* и сблизила с технологиями и стандартами, обеспечивающими взаимодействие на уровне служб. Эту функцию выполняет открытый стандарт *OGSA (open grid services architecture)*. Он интегрирует *Web*-службы и услуги в области *Grid*-вычислений, ресурсы распределенных, гетерогенных, динамических средств, что служит основой для использования *Grid*-приложений в коммерческих целях. Модель *OGSA* включает в себя три стандарта *Web*-служб: *SOAP*, *WSDL* и *WS-Inspection*⁴ (рис. 6) [31, 32].

На *Grid* повлияло стремление как можно эффективнее использовать дорогостоящие распределенные вычислительные ресурсы, сделать их динамичными и однородными. Поэтому архитектура сфокусирована на интеграции уже существующих ресурсов, включая оборудование и ПО, операционные системы, локальные средства, обеспечивающие управление и безопасность. В результате создана «виртуальная организация», ресурсы которой, переведенные в логическую форму, могут потребляться члена-

⁴ Корпорации *IBM* и *Microsoft*, ранее принимавшие участие в разработке технологий *Web*-служб, представили стандарт *WS-Inspection (web services inspection language)*, который дополняет *UDDI*, есть логическим его продолжением и полностью с ним совместим. Как и в случае с *UDDI*, все *web*-решения в рамках новой технологии основаны на *XML*, однако, как утверждается, в основу услуг будут положены более совершенные алгоритмы. Если *UDDI*, строго говоря, представляет собой универсальный справочник, то *WS-Inspection* существенно расширяет возможности поиска. Спецификация нового стандарта *WS-Inspection* дополняет глобальную технологию каталогов *UDDI* благодаря оптимизации процесса непосредственного обнаружения служб на *Web*-сайтах, отсутствующих в реестрах *UDDI*.

Стандарт *WS-Inspection* обмена информацией об услугах, предоставляемых друг другу *Web*-службами, инициирован в первую очередь требованиями обработки информации международных баз данных для предприятий электронной коммерции [33–35].

ми только этой организации. Эти ресурсы находятся под управлением администрации *Grid*, создавая таким образом распределенный пул ресурсов. Пользователь получает необходимые ему, возможно, очень большие по объему ресурсы, за которые он расплачивается по мере их использования.

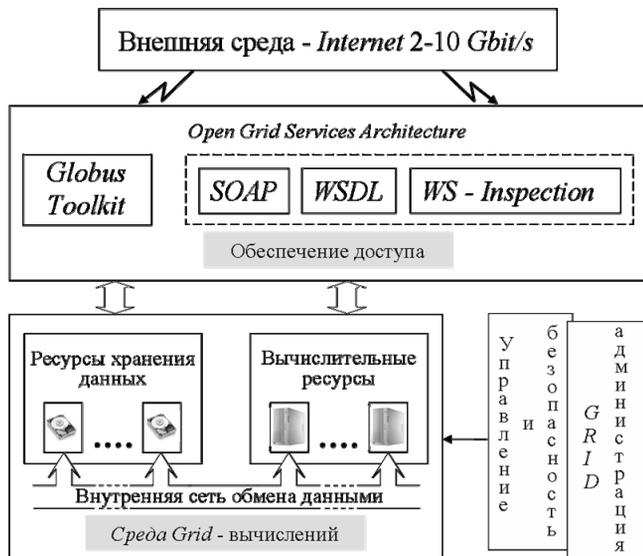


Рис. 6. *Grid* – кластер

Cloud Computing, или распределенная самоуправляемая компьютерная среда как услуга

Идея предоставления разнообразных услуг по требованию с повременной оплатой получила свое дальнейшее развитие в новой парадигме компьютерных систем *Cloud Computing*¹ [*klaud* – облако], предлагающей оперировать не

¹ Своим появлением термин *cloud computing* обязан Эрику Шмидту, который, работая в должности директора по технологиям *Sun Microsystems*, пришел к выводу о родственности сети и компьютера, позже растрогажированном в девизе *Sun Microsystems* «Сеть – это компьютер». Этот девиз – доведенная до предельного лаконизма мысль, высказанная Шмидтом еще в 1993 году: «Когда сеть станет такой быстродействующей, как процессор, компьютер как таковой перестанет существовать, он распространится по сети». Суть новой модели компьютерных систем в том, что сервисы, поддерживающие данные и архитектуру, размещены на удаленных серверах. Данные также находятся на этих серверах, на них же выполняются необходимые вычисления... И если в вашем распоряжении соответствующий браузер и соответствующие права доступа, то вы можете получить доступ к этому облаку независимо от используемого устройства» [9].

терминами конкретных компьютеров, а понятием *услуга*. Ранее словосочетания *Cloud Computing* не существовало, а сегодня новая парадигма компьютерных систем активно обсуждается. Именно парадигма, ведь как таковых «вычислений в облаке» в материальном воплощении не существует. *Cloud Computing* – это такая же метафора, как и другие модели организации вычислений:

- *Cluster Computing*. Объединение физических вычислительных узлов, находящихся в одном административном домене, в единый логический компьютер средствами локальных сетей с высокой пропускной способностью;
- *Grid Computing*. Объединение физических вычислительных узлов, находящихся в разных административных доменах, в единый логический компьютер средствами глобальных коммуникационных технологий;
- *Utility Computing*. Развитие идеи разделения ресурсов. Прежде предметом распределения служили время и другие ресурсы компьютера, а теперь его распространили на множество распределенных в сети компьютеров;
- *Distributed Computing*. Выполнение разных частей программы одновременно на двух (или более) взаимодействующих между собой по сети компьютерах.

По мнению авторов, целесообразно употребить словосочетание *распределенный компьютеринг, распределенный вычислительный ресурс как услуга* или *распределенная самоуправляемая компьютерная среда как услуга* вместо «облачный компьютеринг», как наиболее соответствующий смысловому наполнению понятия предоставляющей услуги распределенной самоуправляемой компьютерной среде. Непременное условие существования этой среды – виртуализация² – одна из двух компонент парадигмы. Вторую компоненту составляют технологии, позволяющие создавать приложения *SaaS*, распространяемые в виде прикладных услуг по требованию и также в виде услуг по требова-

² Виртуализация – создание пула ресурсов (серверов, систем хранения, ПО и др.) для их автоматического распределения и совместного использования.

нию предоставлять необходимую для работы этих приложений платформу *PaaS* и инфраструктуру *IaaS* [36, 37]. В рамках инфраструктуры распределенной компьютерной среды физические ресурсы заключены в большое количество масштабируемых специализированных пулов, сконфигурированных для динамического предоставления и переналадки по требованию в режиме реального времени. Все операционные элементы, включая базовые приложения, контент и все уровни ресурсов, абстрагированы один от другого. Устранение явных зависимостей между ресурсами и приложениями упрощает управление, а поскольку среда в целом характеризуется высокой гибкостью и надежностью, она способна развиваться по мере появления новых технологий, что позволяет минимизировать негативное влияние жестких привязок и изолированных мощностей, проявляющееся при изменении инфраструктуры [38].

Распределенный компьютеринг возможен при хорошо организованной и высоконадежной сетевой инфраструктуре, обеспечивающей качество обслуживания не только индивидуального, но и корпоративного уровня. Это должна быть высшая фаза распределенных компьютерных систем. Только тогда его можно будет воспринимать как некий сверхкомпьютер, точнее, сеть, рассматриваемую как единый компьютер. Такая сеть будет включать в себя разнообразные серверы и системы хранения. Ресурсы сети должны быть виртуализованы, что обеспечит динамическое масштабирование и избавит пользователей от привязанности к определенным физическим серверам. Подобный подход освободит тех, кто развертывает какое-то ПО, от необходимости думать, как и на каком физическом оборудовании оно будет работать. Доступ к ресурсам, разделяемым по запросу, осуществляется через *Internet*. Оплата ресурсов осуществляется по факту их использования (рис. 7).

Среда распределенного компьютеринга должна обладать следующими атрибутами:

- сервисной моделью, определяющей тип и характеристики услуг, предоставляемых распределенным компьютерингом потребителю;

- высокой автоматизацией процесса предоставления услуг по запросу потребителя и возможностью самообслуживания потребителя в рамках предоставляемых услуг;

- возможностью динамического масштабирования объема предоставляемой услуги, например увеличения количества процессорной мощности, предоставляемой потребителю для обработки пиковых нагрузок;

- эластичностью и разделяемостью, т.е. способностью перераспределять имеющиеся ресурсы между потребителями и давать возможность прозрачного для потребителей расширения пула доступных ресурсов;

- возможностью учета потребления ресурсов [39].

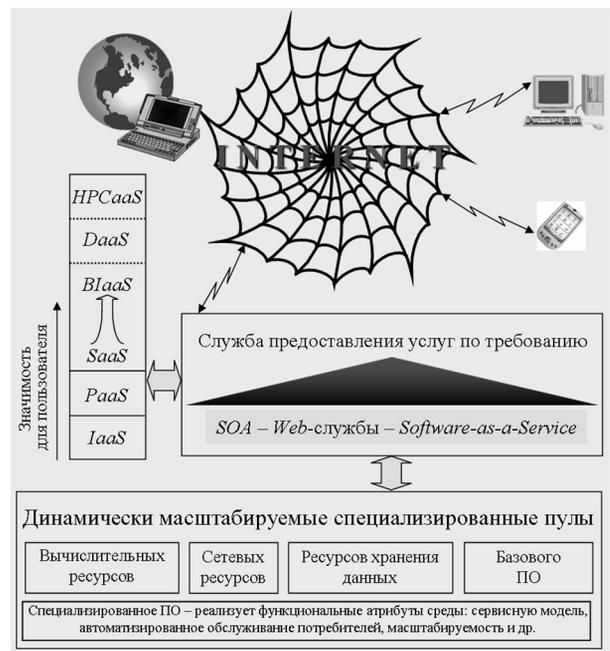


Рис. 7. Распределенная самоуправляемая компьютерная среда как услуга

Распределенный компьютеринг может быть частным (закрытым) или общедоступным (публичным) [38, 39]. Частный – это когда модель вычислений в среде распределенного компьютеринга реализуется на ресурсах, имеющихся в распоряжении организации, для обслуживания внутренних потребностей, т.е. происходит эксклюзивное использование ресурсов данным предприятием. При этом серверы, рабочие станции, устройства хранения, сети, базовая программ-

ная инфраструктура, интегрируются в частную компьютерную среду с помощью специализированного программного обеспечения, позволяющего реализовать функциональные атрибуты среды: сервисную модель, автоматизированное обслуживание потребителей, масштабируемость и др. Сервисная модель частной среды распределенного компьютеринга определяется внутренними потребителями среды, и тем, какие именно ИТ-услуги им требуются.

При этом внутри организации открываются новые возможности:

- благодаря автоматизации накладные расходы на оказание ИТ-сервиса резко падают, снижается время ожидания предоставления ресурсов;
- частная среда распределенного компьютеринга способствует эффективному распределению ресурсов внутри организации, динамически перераспределяя нагрузку между физическими системами ЦОД;
- появляется возможность отслеживать реальное потребление ИТ-ресурсов внутри компании и распределять затраты на поддержку и расширение базовой инфраструктуры между потребителями на основании бизнес-ценности.

Любой закрытый ЦОД, принадлежащий крупному предприятию, может быть назван закрытым распределенным компьютерингом, если он использует преимущества модели унифицированных ресурсов, которая благодаря виртуализации позволяет интерпретировать процессорные мощности, хранилища и сетевые каналы как пул гомогенных ресурсов.

При переносе нагрузок в общедоступную (публичную) среду распределенного компьютеринга появляется возможность гибкого распределения вложений на ИТ-обслуживание предприятия между капитальными – расширением собственной вычислительной среды для поддержки постоянных критических услуг, – и операционными, направляемыми на оплату ресурсов публичной вычислительной среды, используемой для размещения временных или других специфических нагрузок [39].

Возможно также объединение обеих моделей компьютеринга в «гибридную среду», но общая тенденция [38] состоит в том, что в конеч-

ном счете останется лишь публичная среда распределенного компьютеринга. В любом случае все ИТ-решения будут предоставляться по сервисной модели.

Примером общедоступной среды распределенного компьютеринга служит платформа *Windows Azure* для предприятий [40]. Для всех потребностей компании в ПО *SaaS* выглядит правильным выбором. Это – бизнес-модель доставки ПО, по которой приложение размещается у провайдера или третьей стороны и становится доступным клиентам по подписке. Клиенты *SaaS* используют выполняемое в инфраструктуре провайдера ПО на основе оплаты текущих расходов. Эта инфраструктура и конфигурация ПО невидимы пользователям, а значит, клиенты должны довольствоваться только предлагаемой функциональностью. *SaaS* построена на многопользовательской архитектуре, и контексты пользователей всегда логически отделяются один от другого.

Однако каждая компания имеет уникальную ИТ-среду, где могут использоваться различные технологии – от устаревших до специфических для конкретной области. Найти *SaaS*-услугу, удовлетворяющую потребностям специфической области, например бизнеса, зачастую невозможно, поэтому компаниям все равно приходится заниматься разработкой своих приложений. *Platform as a Service* как раз и отвечает потребностям компаний, желающих создавать и выполнять собственные приложения как услуги. К таким компаниям могут относиться независимые поставщики ПО (*ISV*), провайдеры сервисов, корпоративные ЦОД и все, кому необходимы собственные приложения. В случае *PaaS* предлагаются серверы приложений с почти бесконечной масштабируемостью благодаря огромному пулу ресурсов, а также необходимые вспомогательные услуги, в том числе защита, хранилища, инфраструктура интеграции и средства разработки полной платформы.

Провайдер услуг предлагает предварительно сконфигурированную, виртуализованную среду серверов приложений, в которой могут развертывать свои приложения отделы разработок предприятий. Так как провайдеры сервисов уп-

равляют оборудованием (взяв на себя его модернизацию, поддержку и др.), а также гарантируют минимальные простои серверов приложений, участие ИТ-специалистов предприятия сводится к минимуму. Разработчики создают приложения и помечают их дескрипторами ресурсов. При развертывании соответствующий механизм связывает с таким приложением необходимые инфраструктурные возможности, объявленные в дескрипторах приложения. Ресурсы могут включать в себя сетевые конечные точки, балансировщики нагрузок, процессорные ядра, память и зависимое ПО. Масштабирование по запросу в сочетании с управлением оборудованием и серверами приложений освобождает разработчиков от необходимости заботиться об инфраструктуре и позволяет им сосредоточиться на создании приложений. *PaaS*, как правило, подходит для совершенно новых приложений, так как устаревшие зачастую нуждаются в обширной переработке для соответствия правилам выполнения в изолированных программных средах.

В качестве еще одного примера приведем решение Майкрософт [39] для частной среды (*Hyper-V Cloud*) сервисной модели *IaaS* предоставления инфраструктуры как услуги. При такой модели предполагается, что потребитель формулирует свой запрос к среде распределенного компьютеринга в терминах необходимых ему вычислительных ресурсов: количества процессоров, оперативной памяти, дискового пространства, сетевых коммуникаций и базового ПО. Самообслуживание в рамках такого сервиса предполагает возможность самостоятельного создания нужной инфраструктуры для бизнес-приложения в виде виртуальных машин, установки и настройки прикладного ПО. При этом среда распределенного компьютеринга имеет механизм обеспечения доступности предоставленной инфраструктуры и способы учета потребления. Основа такого решения – серверная виртуализация на базе гипервизора *Hyper-V*, позволяющая отделить с помощью виртуальных машин нагрузки от физического оборудования и предоставляющая необходимую гибкость для выделения ресурсов, динамического перерас-

пределения нагрузок, обеспечения высокой доступности. Важнейшая роль принадлежит инструментам управления *System Center*, позволяющим реализовать автоматизацию предоставления инфраструктурного сервиса.

Компания *NetApp* продемонстрировала [41] свое решение в области эффективного и безопасного использования ресурсов системы хранения в средах распределенного компьютеринга. ПО *NetApp® MultiStore®* позволяет быстро создавать отдельные разделы и скрытые логические разделы в единой системе хранения данных (СХД). Все разделы виртуальной СХД отделены один от другого, что позволяет нескольким пользователям совместно использовать один и тот же ресурс хранения без ущерба для конфиденциальности и безопасности данных.

Тома данных постоянно приходится изменять, перемещать и удалять. *NetApp® Data Motion* обеспечивает непрерывную мобильность данных. Миграция данных осуществляется в реальном времени без плановых простоев. Повышение эффективности СХД достигается использованием технологии виртуализации, дедубликации и гибкого выделения ресурсов. Это позволяет сократить потребность в дисковом пространстве на 80 процентов.

NetApp предлагает полностью интегрированную комплексную систему защиты данных и восстановления после сбоев на основе уникальных технологий, переносящих задачу обеспечения безопасности данных с сервера на систему хранения. Благодаря этому обеспечивается последовательная защита данных в инфраструктуре распределенного компьютеринга.

Компания *Oracle* распространяет по модели *SaaS* бизнес-приложение *CRM on Demand*, в настоящее время идет работа над 20–30 продуктами, предназначенными для продажи в форме *SaaS* [42].

В системах с большим количеством массивов данных, клиентов, пользователей и приложений добиться удобства администрирования ресурсов хранения, а также высокого уровня надежности системы и непротиворечивости данных можно только путем автоматизированного управления СХД. Не менее важен мониторинг

в режиме реального времени имеющегося дискового пространства, показателей производительности системы и соответствия политикам в неоднородной среде хранения.

Как видим, переход к средам распределенного компьютеринга предполагает разработку и продвижение технологий, распространяемых в виде прикладных услуг по требованию (*SaaS*, *IaaS* и др.), как в ЦОД заказчиков через построение частных сред, так и в ЦОД сервис-провайдеров, предоставляющих услуги публичных размещенных сред. ЦОД, способные работать в условиях распределенного компьютеринга, должны не только гарантированно обеспечивать время безотказной работы, требуемый уровень безопасности и необходимое качество обслуживания, но и динамически предоставлять ресурсы. Такие ЦОД похожи на фабрики, в которых приложения и сервисы соответствуют целевому назначению центра по «производству» конкретного спектра ИТ-услуг.

Таким образом, появление и быстрое развитие распределенного компьютеринга вызвано необходимостью за приемлемые деньги создавать решения с неограниченными возможностями масштабирования [9]. Бизнес-модель *SaaS* позволяет сохранить инвестиции – приложения самостоятельно существуют в распределенной компьютерной среде до момента их востребования потребителями. Пользователи при этом экономят на инфраструктуре, так как нет необходимости заботиться о приложениях, расходовать средства на информационную безопасность и поддержку требований различных регламентирующих нормативов, потенциально при этом получая доступ к теоретически неограниченным по масштабам вычислительным ресурсам, заведомо превышающим самую крупную сетевую корпоративную ИТ-конфигурацию, т.е. в распределенной компьютерной среде аккумулируются мощности, чтобы предоставить более экономичное масштабируемое решение и избавить пользователя от необходимости работать исключительно на собственном оборудовании.

Перспективы становления и вопросы безопасности

По данным [36, 43] со ссылкой на *IDC* и *Gartner*, в 2009 г. рынок «облачных» сервисов

составил \$17,4 млрд, к 2013 г. он достигнет \$44 млрд – 10% всего ИТ-рынка, а к 2014 г. доходы от них достигнут \$148,8 млрд. Половина этого объема приходится на *SaaS*, но к 2013 г. доля этого компонента снизится до 39% в пользу платформ *PaaS* и инфраструктур *IaaS*, предоставляемых в виде услуг. По прогнозам *Gartner*, в 2010 г. на Западную Европу придется 23,8% рынка «облачных» услуг, на долю Японии – 10%. В 2014 г. на долю Великобритании будет приходиться 29% рынка «облачных» услуг, а на долю Японии – 12%.

Отмечено, что хотя интерес к распределенному компьютерингу ощутимо вырос, во многих организациях по-прежнему вызывают беспокойство идеи «облачных» вычислений, «облачных» продуктов и «облачных» сервисов. Основную проблему для ИТ-руководителей представляют вопросы безопасности, к другим – относятся доступность услуг, эффективность и надежность их поставщиков.

Полагают, что к 2012 г. 20% компаний в мире откажутся от собственных ИТ-ресурсов. Но есть и другая тенденция – к этому же году 45% всех хранимых данных компаний будут в той или иной степени конфиденциальными. А не опасно ли будет бизнесу работать в среде распределенного компьютеринга, помещая туда свои данные?

Размывание сетевого периметра, некогда ограничивавшего корпоративные информационные ресурсы, добавляет риски по защите и сохранению данных. Первый шаг в процессе перехода к распределенному компьютерингу – виртуализация, т.е. переход от физических серверов к виртуальным. Исчезает понятие «периметра сети», исключается возможность применения каких-либо средств защиты в одной точке. Кроме того, теряется возможность сегментировать сеть межсетевым экраном, и возникают проблемы с использованием аппаратных средств защиты, поскольку большая часть трафика перемещается внутри физических серверов – между виртуальными машинами. Помимо сказанного, коммуникационные решения, предоставляющие мобильным пользователям возможность доступа к важным данным в лю-

бом месте в любое время, создает дополнительные риски для информационной безопасности, так как доступ к сетям зачастую осуществляется из точек, не защищенных традиционными корпоративными межсетевыми экранами.

Утверждается [43], что при правильном подходе к выбору провайдера распределенного компьютеринга существует возможность даже повысить уровень защиты своих данных. Основные моменты подразумевают сохранность хранимых данных, их защиту при передаче, аутентификацию и изоляцию пользователей. Ключевые моменты – использование технологий шифрования информации, доступность данных пользователю только после аутентификации, использование клиентом индивидуальной виртуальной машины и др.

Высказывается такое мнение, что провайдер предоставляет более высокий уровень безопасности, чем тот, который компании могут обеспечить внутри собственной инфраструктуры: обслуживая предприятия с оборотом в миллиарды долларов, провайдеры таких услуг сделают все от них зависящее для обеспечения максимально безопасной среды. Трудно не принять это во внимание, как нельзя и уповать на то, что данные будут пребывать в полной безопасности в особенности после случая с атакой вирусом-террористом *Stuxnet* Бушерской АЭС в сентябре 2010 г. [44]. Эту точку зрения подтверждают обсуждения в рамках форума «Мир ЦОД 2011» [45], где среди экспертов не было единого мнения относительно того, чем опасны «облака». Проблема информационной безопасности, безусловно, существует и ее решение лежит в целевом взаимодействии как провайдеров, так и потребителей «облачных» услуг.

В ходе обзора авторами был высказан ряд иных взглядов и подходов в части оценки и использования новых информационных технологий обработки данных и информации. Совокупность этих аналитических выводов показала свою значимость при разработке новых интеллектуальных информационных технологий, вытекающих из государственных научно-технических программ [46–58].

1. Гриценко В.И., Урсатьев А.А. Распределенные информационные системы. Состояние. Перспективы развития // УСиМ. – 2003. – № 4. – С. 11–21.
2. Преймсбергер К. Общий взгляд IBM на хранение данных // PC WEEK/UE. – 2008. – № 12. – С. 20–21.
3. Миллард Э. Хранение данных: тенденция 2008 г. и на будущее // Там же. – С. 21.
4. Примеры использования технологии Хранилищ данных в энергетической отрасли. – <http://www.iso.ru/journal/articles/202.html>, 2002.
5. Гир Дэвид. Федеративный доступ к базам данных // Открытые системы. – 2003. – № 6. – С. 69–71.
6. Волков Д. Тихий передел // Там же. – 2010. – № 04. – С. 1.
7. Артемьев В. Что такое *Business Intelligence*? // Там же. – 2003. – № 04. – С. 20–26.
8. Черняк Л. Что *Business Intelligence* предлагает бизнесу // Там же. – С. 33–38.
9. Черняк Л. Бизнес-аналитика как сервис // Там же. – 2010. – № 04. – С. 16–23.
10. Базы данных. Интеллектуальная обработка информации / В.В. Корнеев, А.Ф. Гареев, С.В. Васютин и др. – М.: Нолидж, 2000 – 352 с.
11. Гриценко В.И., Урсатьев А.А. Распределенные информационные системы широкого применения. Концепция. Опыт разработки и внедрения. – К.: Наук. думка, 2005. – 317 с.
12. Фейгин Д. Концепция SOA // Открытые системы. – 2004. – № 06. – С. 14–18.
13. Архипенков С., Голубев Д., Максименко О. Хранилища данных. От концепции до внедрения. – М.: Диалог-Мифи, 2002. – 528 с.
14. Бробст С. Интеграция или изоляция? // Открытые системы. – 2010. – № 4. – С. 24–25.
15. Волков Д. Опционы будущего // Там же. – 2003. – № 4. – С. 1.
16. Инмон Б. DW 2.0: хранилища данных следующего поколения // Там же. – 2007. – № 5. – С. 20–26.
17. Дубова Н. Динамическое хранилище данных // Там же. – С. 28–32.
18. Вышла новая версия *SQL Server 2008* // Там же. – 2010. – № 04. – С. 6.
19. Елисеев И. Ответность своими руками // Сети. – <http://www.osp.ru/news/articles/2010/19/13002034/>
20. Глушков В.М. Введение в АСУ. – К.: Техніка, 1974. – 317 с.
21. Глушков В.М. Основы безбумажной информатики. – М.: Наука, 1982. – 552 с.
22. Web-службы // Открытые системы. – 2002. – № 11. – С. 1–78.
23. Волков Д. Конкретная форма SOA // Там же. – 2004. – № 06. – С. 26–31.
24. Как разобраться в SOA / Н. Голд, Э. Мохан, К. Найт и др. / Там же. – 2004. – № 06. – С. 32–37.
25. Дубова Н. SOA: подходы к реализации // Там же. – С. 19–25.

26. *Web Services Glossary*. – <http://www.w3.org/TR/2003/WD-ws-gloss-20030808>.
27. *Web Services Architecture*, W3C Working Draft 8, August 2003. – www.w3.org/TR/2003/WD-ws-arch-20030808.
28. Черняк Л. *SaaS* – конец начала // Открытые системы. – 2007. – № 10. – С. 42–46.
29. Сысойкина М. Облачные сервисы в России: слово или дело? // Мир ПК. – 2011. – № 1. – С. 71–73.
30. Черняк Л. ОС для облачных сетей // Открытые системы. – 2010. – № 7. – С. 20–21.
31. *Grid* // Там же. – 2003. – № 1. – С. 1–78.
32. *Решения IBM* на базе *GRID*-технологий. – <http://www.ibm.com/grid>
33. *WS-Inspection*. – http://www.ibm.com/news/ru/2002/07/04_07_01.html, 2002.
34. William A. Nagy (IBM), Keith Ballinger (Microsoft) *The WS-Inspection and UDDI Relationship*. – <http://www-106.ibm.com/developerworks/webservices/library/ws-wsi-luddi.html>, 2001.
35. *WS-Inspection* – новое XML-решение для электронной коммерции от IBM и Microsoft. – <http://www.uvtb.ru/scripts/servinfo.dll?GetObject&OID=963527&OT=3&POID=-1&encoding=windows-1251>, 2001.
36. *Вычислительные облака* // Открытые системы. – 2008. – № 7. – С. 1–78.
37. *Безопасные облака* // Там же. – 2010. – № 1. – С. 1–78.
38. Андерсен К., Линдберг П., Потапкин С. *Облако из Скандинавии* // Там же. – 2008. – № 7. – С. 36–37.
39. *Инфраструктура* облачных вычислений Майкрософт. – <http://www.microsoft.com/virtualization/ru/ru/cloud-computing.aspx>, 2010.
40. Коммалапати Х. Платформа *Windows Azure* для предприятий // *MSDN Magazine* – Февр. – 2010. – <http://msdn.microsoft.com/ru-ru/magazine/ee309870.aspx>
41. Семинар компании «Техносерв Украина» «Инновационные решения Fujitsu и NetApp по построению инфраструктуры хранения и обработки данных». – Киев, 28 окт. 2010 г.
42. Oracle строит гигантский ЦОД // Открытые системы. – 2010. – № 04. – С. 6.
43. Варганич Е. Безопасная миграция в «облака» // Сети и телекоммуникации. – 2010. – № 11. – С. 56–57.
44. Шабуров О. Король червей // Мир компьютерной автоматизации: встраиваемые компьютерные системы. – 2011. – № 2. – С. 58–60.
45. Мир ЦОД 2011: Чем опасны облака? / Материалы конф. «Мир ЦОД 2011». Статьи на диске «Мир ПК» // Мир ПК – июль 2011. – С. 3.
46. Гриценко В.И. Общество знаний: проблемы становления и развития // УСиМ. – 2004. – № 4. – С. 5–13.
47. Шлезингер М., Главач В. 10 лекций по статистическому и структурному распознаванию. – К.: Наук. думка, 2004. – 545 с.
48. *Інформаційні технології в біології та медицині* / В.І. Гриценко, А.Б. Котова, М.І. Вовк та ін. – Там же, 2007. – 384 с.
49. Файнзильберг Л.С. Информационные технологии обработки сигналов сложной формы. Теория и практика. – Там же, 2008. – 333 с.
50. *Оптимизационные задачи* структурного распознавания образов // УСиМ. – 2011. – № 2. – С. 1–97.
51. Гриценко В.И., Файнзильберг Л.С. Трансфер наукоемких технологий и изделий цифровой медицины для определения функционального состояния сердца человека в амбулаторных и домашних условиях // Матеріали II міжнар. форуму «Трансфер технологій та інновацій». – Київ: МОН України, 2008. – С. 275–283.
52. Файнзильберг Л.С. ФАЗАГРАФ® – эффективная информационная технология обработки ЭКГ в задаче скрининга ишемической болезни сердца // Клиническая информатика и телемедицина. – 2010. – Т. 6. – Вып. 7. – С. 22–30.
53. Гриценко В.И., Вовк М.И., Котова А.Б. Мозг в ракурсе интеллектуальных информационных технологий // Кибернетика и вычислительная техника. – 2010. – № 162. – С. 3–9.
54. Вовк М.И. Биоинформационная технология управления движениями человека // Кибернетика и вычислительная техника. – 2010. – № 161. – С. 42–52.
55. Пат. України № 51137. Спосіб голосового управління цифровим диктофоном / Т.К. Вінцюк, В.І. Гриценко, О.І. Павлов, П.А. Стасевич, Г.М. Тертичний. – Опубл. 12.07.2010, Бюл. № 13.
56. Пат. України № 51139. Спосіб введення-виведення інформації голосом в комп'ютерних та телекомунікаційних системах / Т.К. Вінцюк, В.І. Гриценко, О.І. Павлов, П.А. Стасевич, Г.М. Тертичний. – Опубл. 12.07.2010, Бюл. № 13.
57. Пат. України № 50036. Спосіб описування та розпізнавання мовленнєвих сигналів / Т.К. Вінцюк, В.І. Гриценко. – Опубл. 25.05.2010, Бюл. № 10.
58. Пат. України № 50040. Спосіб пофонемного розпізнавання злитого мовлення / Т.К. Вінцюк, В.І. Гриценко. – Опубл. 25.05.2010, Бюл. № 10.

Поступила 23.08.2011
Тел. для справок: (044) 526-4159 (Київ)
E-mail: aleksei@jirtc.org.ua
© В.И. Гриценко, А.А. Урсатьев, 2011