



Архитектура доставки информации SAS

Введение для руководителей ВЦ и АСУ

Москва 2002

Архитектура доставки информации SAS.....	1
<i>Введение для руководителей ВЦ и АСУ</i>	1
Представляем SAS.....	4
Интегрированные, масштабируемые, адаптивные архитектуры.....	4
Сектор рынка аналитических приложений	5
Различие между оперативными и аналитическими приложениями.....	5
Нахождение компромисса.....	6
Хранилища данных.....	6
Доступ к базам данных источников	7
Инструментальные средства ETL	7
Очистка данных.....	8
Архитектура метаданных	8
SAS Среда для разработчика.....	9
Язык программирования SAS	9
Аналитические процедуры	10
Добыча данных.....	10
Web-приложения.....	11
Windows-приложения.....	11
Настольные средства формирования отчетов	11
Открытая архитектура SAS.....	12
Клиент - серверные и распределенные вычислительные архитектуры.....	12
Архитектура поддержки взаимодействия SAS IOM	13
Публикация и подписка	14
Портал доставки информации SAS Information Delivery Portal.....	15
Архитектура портала SAS	17
Масштабируемость.....	18
Обработка в Системе SAS	18
Базы данных для аналитических приложений.....	20
Управление данными и производительностью	20
Журналирование	20

Сегментация и разбиение файлов	20
Проектирование таблиц	21
Инструменты создания моделей данных.....	21
Репозитарий хранилища	22
Подготовка отчетов и обработка запросов	22
Интерактивные средства подготовки отчетов и OLAP	23
Специализированный анализ	24
Пакетная обработка и форматы.....	24
Эксплуатационные моменты	25
Безопасность	25
Разграничение функций.....	25
Списки контроля доступа.....	26
Шифрование	26
Планирование процессов.....	27
Резервное копирование и восстановление	27
Управление конфигурацией приложений SAS	27
Контроль целостности	28
Интеграция корпоративной архитектуры.....	28
Обработка сообщений	28
Интеграция ERP при помощи технологии RFC	28
Проблема EAI: Интерфейс «спагетти»	30
SAS Integration Technologies: среда для создания EAI.....	31

Представляем SAS

Представить Архитектуру Доставки информации SAS в одном документе и добиться ясности изложения, - задача непростая. Рассказывая в этой брошюре о технологиях SAS, мы ориентировались на директоров информационных служб, которые ранее не сталкивались с SAS или поверхностно знакомы с компанией, но которые рассматривают перспективы использования разработок SAS в рамках проектов создания Информационного Хранилища (Data Warehouse) или аналитических систем (Business Intelligence). В брошюре описываются основные компоненты ИТ, которые могут быть использованы при внедрении решений SAS.

Разработки SAS ведут свою историю от анализа результатов измерений и экспериментов, подтверждения научных гипотез и статистического анализа. Для решения подобных задач необходимо извлечь информацию из исходных данных. Это требует соответствующего инструментария для обработки данных.

Информационная же отрасль изначально занимается поддержкой автоматизации процессов, циклическим сбором данных и подготовкой отчетов на основании имеющейся информации (т.е. сбором и обработкой данных). В информационной отрасли для описания процесса сбора информации обычно используется другая терминология. Кроме того, отраслевая модель управления, как правило, отлична от той, что используется у традиционных пользователей SAS. Тенденции к увеличению объемов данных и росту требований к системам аналитической обработки на коммерческой арене создают дополнительные возможности для использования SAS в рамках основных бизнес-процессов.

Интегрированные, масштабируемые, адаптивные архитектуры

В большинстве отраслей экономики средние и крупные предприятия являются сложно структурированными, т.е. состоящими из множества компаний и бизнес-подразделений. По мере роста предприятий, корпоративных слияний и поглощений возникают барьеры, препятствующие интеграции ИТ и повышающие расходы предприятий. Среди причин можно назвать:

- Несовместимые инфраструктуры
- Разнородное оборудование
- Разнородные пакеты приложений
- Неоднородная архитектура данных
- Большое количество различных СУБД

Еще более усложняет картину количество и разнообразие киосков данных, обычно имеющих на крупном предприятии. Их количество растет по мере того, как у стратегически важных приложений (в частности, у приложений добычи данных (data mining) и управления взаимоотношениями с клиентом (CRM)) появляется все больше пользователей.

Идея, стоящая за SAS Enterprise Computing, достаточно проста: предоставить нужную информацию нужным людям, когда они этого захотят, в том виде, в каком они этого захотят, по минимальной стоимости и с максимальной эффективностью. Программное обеспечение SAS специализируется на анализе данных и добыче данных (data mining). SAS лидирует также в области информационных хранилищ, хранения больших объемов информации и предоставлении результатов анализа.

Сектор рынка аналитических приложений

Отраслевой сектор аналитических приложений (business intelligence) разительно отличается от сектора персональных систем, сектора ERP-систем и стандартных приложений, и сектора систем управления реляционными базами данных (RDBMS), в которых доминируют соответственно Microsoft, SAP и Oracle. Такие ERP-системы, как R/3, и базы данных, выступают в роли источников данных для информационных хранилищ и аналитических приложений. А персональные продукты Microsoft, как правило, являются клиентом для просмотра результатов анализа. Аналитические системы занимают в информационной архитектуре положение между приложениями ERP на фундаменте баз данных и потребителем информации. Эти средства преобразуют «сырую» информацию в полезные знания.

Различие между оперативными и аналитическими приложениями

Оперативные и аналитические приложения имеют различные эксплуатационные параметры и характеристики. Не следует забывать о контрасте исходных посылов, которыми руководствовались создатели систем оперативной (транзакционной) обработки.

Характеристика	Оперативная OLTP система	Аналитическое приложение
Режим	Обновление данных	Чтение данных
Доступ к данным	Произвольный доступ. Несколько строк.	Последовательный доступ. Набор данных целиком
Транзакции	Огромное количество простых транзакций.	Малое количество сложных запросов.
Профиль	Интенсивный ввод/вывод	Интенсивные вычисления
Ввод	Проверка, интенсивный диалог и взаимодействие с сотнями пользователей.	Большой объем пакетных извлечений данных из разнообразных источников.
Проектирование	От процессов	От бизнес-задач

Нахождение компромисса

ИТ подразделения могут использовать множество подходов и компромиссных вариантов, чтобы разрешить противоречия в требованиях, приведенных выше.

- Для работы с различными данными можно использовать различные приложения и платформы (как правило, системы для научных исследований и разработки обычно отделены от коммерческих и промышленных систем).
- Если данные одни и те же, можно организовать отложенное создание отчетов в ночное время, когда пользователи в большинстве своем разошлись по домам.
- Если специальный запрос должен быть выполнен в то же самое время, что и обновления в режиме реального времени, вы можете извлечь копию данных накануне и запустить специальный запрос на этой копии, чтобы не возникло чрезмерной нагрузки на подсистемы ввода/вывода информации.
- В конце концов, вы придете к использованию технологий построения информационных хранилищ, чтобы объединить перечисленные выше методы и воспользоваться другими. Хранилища, в конечном счете, позволяют разрешить конфликтные ситуации преимущественно путем интеграции операционных приложений и аналитических приложений в когерентную информационную архитектуру.

Технологии и продукты SAS с легкостью управляют аналитическими приложениями. Это будет показано более наглядно, когда мы перейдем к рассмотрению используемых в SAS языка и подхода к обработке данных.

Хранилища данных

Формально хранилища данных были определены как автономная среда для поддержки аналитической обработки информации, которая является предметно-ориентированной, вариантной во времени и интегрированной. Отделение аналитической среды от оперативных источников предполагает следующее:

- **Извлечение, трансформация и загрузка или ETL.** ETL (Extraction, Transformation, Loading) – начальный этап обработки, служащий для заполнения хранилища данных. Слово «извлечение» не вполне отражает суть дела, поскольку предполагает удаление информации. Однако типичный ход процесса выглядит следующим образом:
 - Доступ к исходной базе данных.
 - Чтение и копирование данных.

- Дописывание копии к согласованной по времени таблице хранилища данных.

SAS может читать «родные» каталоги СУБД и метаданные приложений, если речь идет о широко распространенных ERP-системах, таких, как SAP R/3. Перед сохранением данных они могут быть проверены, согласованы и трансформированы (т.е. различные кодификаторы и НСИ могут быть сведены к единой кодировке информационного хранилища в целях обеспечения согласованности отчетов).

- **Классификация и контроль версий.** Это процесс, в рамках которого данные группируются в соответствии со стандартными классификационными и иерархическими принципами для подготовки отчетности. Следует отметить, что для приложений добычи данных (data mining), адекватные взаимосвязи и иерархии не известны до тех пор, пока не произведен процесс собственно добычи данных. В качестве примера можно привести определение сегментации рынка.
- **Агрегирование.** Агрегирование обычно производится путем определения вычисляемых переменных и выполнения правил расчета в таблицах метаданных. Это обеспечивает согласованное качество отчетов.
- **Отчетные киоски данных.** Создаются киоски данных по каждой предметной области, функции или подразделению, доступные для средств подготовки отчетов или выполнения запросов.
- **Проверка.** При проверке используются таблицы НСИ (например, справочники клиентов).

Доступ к базам данных источников

SAS поддерживает стандарты и протоколы для доступа к различным базам данных. С тем, чтобы избежать необходимости написания интерфейсов для реляционных баз данных, SAS предоставляет набор средств доступа, которые представляют среды СУБД в таком виде, как если бы они были таблицами SAS. Для популярных ERP систем, например, SAP R/3, система доступа поддерживает уровень метаданных системы и её API. Этот набор средств доступа позволяет разработчику легко создавать аналитические приложения, используя единую среду для всей цепочки доставки информации. Разработчик получает возможность доступа к множеству источников информации в виде баз данных и файлов, а также возможность создавать хранилище данных и осуществлять анализ и подготовку отчетов.

Инструментальные средства ETL

Большинство инструментальных средств ETL имеет дело только с этим этапом процесса создания хранилища данных. Они генерируют код, основанный на правилах проверки и преобразования информации, которые зафиксированы в этом инструментальном средстве. Такая операция часто приводит к необходимости недокументированных манипуляций, если требуется сложная обработка вызванным модулем. Следует отметить, что в системе SAS подобное невозможно, поскольку код и преобразования непосредственно связаны в рамках интегрированной среды SAS.

Возможности SAS по трансформации и обработки больших входных потоков данных не ограничиваются SQL запросами к СУБД. Мощный язык обработки данных – 4GL SAS – позволяет производить обработку массивов данных целиком, а не построчно, как в SQL. Это многократно

повышает производительность ETL процессов. Кроме того, SAS может легко создавать множество выводных таблиц из исходных таблиц за один проход, делая тем самым процесс ETL и проще, и быстрее.

Очистка данных

Быстрый рост объемов данных, поступающих из систем электронного бизнеса и электронной коммерции, CRM и приложений обеспечения качества, call-центров и внутренних операций стимулировал эволюцию систем управления данными. Потребность очищать имеющиеся данные, повышать их качество и ценность оказалась на первом плане. Лучший способ очистить данные - удалить нестандартные данные, дубликаты и неверные данные. В дополнение к этому, можно повысить ценность существующих данных посредством их консолидации и дополнения (например, к именам пользователей могут быть добавлены сведения социально-демографического характера). Избыточные или дублированные данные - проблема, касающаяся не только старых систем, имеющихся в эксплуатации, но и сравнительно новых, в частности, Web-порталов, сайтов электронной коммерции и сетевых аукционов. На многих сайтах имеется множество имен или счетов для одного и того же пользователя, равно как и одно и то же изделие может фигурировать в различных списках. В некоторых случаях могут быть по-разному сконфигурированы инсталляции SAP R/3, что приводит к трудностям при интерпретации обобщенных отчетов.

Однажды идентифицированные, эти дублируемые записи могут быть удалены, модифицированы или объединены в один исходный блок контактной информации.

Архитектура метаданных

Текущая архитектура метаданных SAS - "лучевая" (hub-and-spoke), причем в центре находится SAS/Warehouse Administrator - среда управления метаданными хранилища, которая хранит следующую информацию:

- Определение источников данных и задач.
- Определение оборудования, средств коммуникации и различных СУБД для распределенных гетерогенных хранилищ.
- Описание процессов ETL.
- Правила агрегирования данных и вычислений.
- Текущие последовательности загрузки хранилища и обработки данных.
- Процессы, происходящие в киосках данных.
- Метаданные для отчетов.
- Структуры для работы HOLAP – гибридная среда для оперативной аналитической отчетности.
- Пользователи и группы пользователей для важных аналитических сообществ.

Говоря о перспективах развития, необходимо отметить, что SAS является членом группы разработчиков стандартов Object Management Group (OMG). В ближайших версиях Warehouse Administrator будет обеспечена максимальная поддержка нового стандарта XML.

В 9 версии SAS концепция централизованного администрирования будет существенно усовершенствована посредством приложения Management Console - единого центра управления всей цепочкой поставки информации.

Management Console позволит интегрировать все аналитические сообщества и средства размещения, в дополнение к средствам управления хранилищем данных, описанным выше.

Средства доступа SAS к ERP-системам относятся к категории 'Knowledge Solutions', т.к. позволяют получать дополнительную бизнес-информацию из метаданных ERP-систем. Комплексная модель SAP R/3 рационализируется, упрощается и последовательно описывается, чтобы облегчить процесс быстрого кодирования запросов. Как правило, создание отчета в этой среде происходит намного быстрее, чем при использовании языка ABAP компании SAP.

SAS Среда для разработчика

Компания SAS лидирует в разработке полных решений (end-to-end) и предоставляет платформы разработки приложений для формирования всех компонентов, которые могут потребоваться для создания таких решений (например, ETL, аналитические процедуры, добычу данных, средства подготовки запросов для конечного пользователя, средства подготовки отчетов и web-приложения).

Язык программирования SAS

Язык программирования 4GL SAS представляет собой мощную среду разработки четвертого поколения для создания аналитических приложений. SAS используется как для создания структур данных, так и для работы с ними. Необходимо отметить, что это в корне отличается от пользовательских средств подготовки отчетов, созданных на основе SQL. Дело в том, что SQL разрабатывался, как язык запросов к реляционной базе данных и слабо подходит для задач, далеких от транзакций. Напротив язык SAS ориентирован на обработку таблиц целиком, а не на работу с отдельными записями.

Производительность труда разработчика, пользующегося средствами SAS, весьма высока по сравнению со специалистами, использующим SQL, при этом возможности и гибкость, а также эффективность полученного кода оказываются несравнимо выше.

4GL SAS позволяет выполнять следующие операции.

- Манипуляции с таблицами.
- Написание кодов запросов без ограничений, присущих SQL.
- Выполнение аналитических и статистических процедур.
- Выполнение операций объединения с большей эффективностью, нежели SQL.
- Генерация SQL в случае необходимости.

- Поддержка запросов к реляционным СУБД (причем, если это возможно, используя «родной» SQL той базы, к которой идет обращение).
- Использование более эффективного управления хранением данных для временного или постоянного хранения.

Аналитические процедуры

SAS представляет собой полнофункциональную среду четвертого поколения с полным набором функций, что устраняет необходимость в программировании циклов чтения-записи для обработки таблиц или файлов. Такое программирование было необходимо при работе со средами третьего поколения.

Кроме того, SAS предоставляет большое число готовых статистических и аналитических процедур, которые имеют статус отраслевого стандарта. И, в дополнение к этому, SAS располагает мощными подпрограммами агрегирования, которые могут подготавливать несколько сводных справок за один проход без сортировки. Широта диапазона функций означает возможность использования единого языка, языка SAS. При этом не надо выходить из среды разработки, для того, чтобы выполнить сложные манипуляции в более развитой среде.

Добыча данных

Средства добычи данных (data mining) с полным правом можно назвать визитной карточкой SAS. В рамках этого небольшого документа невозможно дать полное описание решения SAS для добычи данных. Главное, что следует знать, администратору по информационным технологиям - средства добычи данных SAS интегрируют в себе все, что нужно пользователю для анализа и представления результатов на каждом этапе процесса добычи данных:

- средства хранения данных,
- средства оперативной аналитической обработки (OLAP),
- визуализации данных,
- деревья решений,
- прогнозирование,
- нейронные сети,
- методы кластеризации,
- методы прогнозирования
- и другие аналитические инструменты.

Эти, основанные на правилах, инструменты добычи данных выявляют устойчивые закономерности или исключительные значения, а также воспроизводят способность человека осмысливать пространственную, географическую и визуальную информацию. Благодаря интуитивному графическому интерфейсу нет необходимости тратить силы, изучая, как работает инструментарий, прежде чем приступить к его использованию. Сложные технологии добычи данных выполняются в среде, полностью свободной от кодирования, что дает разработчикам возможность

сосредоточиться на визуализации данных, выявлении новых закономерностей и формулировке новых вопросов.

Информацию по этим вопросам можно найти также по адресу www.sas.com/technologies/data_mining/index.html.

Web-приложения

AppDev Studio, среда разработки SAS на базе Web, позволяет организациям еще более повысить возврат от инвестиций в хранилища данных за счет расширения возможностей доставки информации из хранилища по сетям intranet. При этом можно охватить значительно большую аудиторию при минимальных ценах внедрения.

AppDev Studio - это мощная среда разработки, поддерживающая метод drag-and-drop. В ней можно создавать настраиваемые приложения для Web-анализа и OLAP-приложения. Это программное обеспечение позволяет использовать технологию гибридного OLAP на основе метаданных SAS/Warehouse Administrator, и предоставить пользователю результат посредством технологии JSP.

В дополнение к AppDev Studio, продукт SAS/IntrNet позволяет организовать доставку полезной информации, выработанной приложениями SAS в виде динамических Web страниц. Иными словами, пользователю нужен только браузер для того, чтобы обращаться к данным со своего компьютера.

Web-приложения можно так же развертывать как .asp страницы при помощи технологии SAS Integration Technologies.

Windows-приложения

Интерфейсы тонкого клиента, такие как Enterprise Guide, предоставляют быстрый и прозрачный доступ к серверам поддержки принятия решений SAS и хранилищам данных. Это, в свою очередь, дает пользователю возможность создавать запросы и формировать отчеты всего за несколько нажатий на клавиши мыши. Enterprise Guide - это клиентское приложение на базе Windows, вырабатывающее код SAS для тех пользователей, задачи которых предъявляют высокие требования к ресурсам.

На самом деле, приложения, написанные при помощи любого инструментария, совместимого со стандартом DCOM, такого, например, как Microsoft Visual Basic или Delphi, могут получать доступ к службам SAS. Более подробно это изложено в разделе "Открытая архитектура SAS" в данном документе.

Настольные средства формирования отчетов

Настольные средства формирования отчетов (Desktop Reporting Tools) традиционно представляют собой системы "тактического значения", входящие составной частью в реализации ERP. Они используются в ситуациях, когда пользователю необходимо представить некий логический «фасад», в котором можно выполнять кодирование запроса или отчета. Хранилище существует не всегда. В лучшем случае присутствуют некие зачаточные структуры, поддерживающие оперативную отчетность. Запросы адресованы либо базе данных ERP-OLTP непосредственно, либо хранилищу

текущих данных ODS (operational data store). Как правило, степень свободы ограничена рамками модели, в которой фактические структуры данных отображаются в логический бизнес-вид. Архитектуры систем, основанные на этих инструментах, чаще всего используются для подготовки отчетов в ERP-системах, наподобие SAP R/3. До сих пор ничто не давало повода усомниться в работоспособности этого процесса, несмотря на то, что он может вызвать и серьезные проблемы, связанные с производительностью ввода/вывода и управлением конфигурацией. Кроме того, такие средства не поддерживают сложных аналитических функций.

Настольные средства могут получать доступ к данным и службам SAS посредством драйверов ODBC; однако, возникнет необходимость в интерфейсах к метаданным. Лучше всего начинать разработку аналитических систем с выработки архитектуры хранилища данных.

Открытая архитектура SAS

SAS - это полностью открытая система. Все версии SAS, начиная с 8, предоставляют полные и открытые технологии для организации распределенных вычислений корпоративного уровня.

Под открытостью здесь понимается следующее.

- Стронние клиенты могут запрашивать данные и вычислительные службы SAS.
- Клиенты SAS могут запрашивать службы сторонних производителей.
- Вся многозвенная обработка может быть выполнена на базе стандартных отраслевых протоколов, регламентирующих использование Web и распределенные вычисления

Так, например, для работы средств CRM требуется доступность знаний, правил и данных, являющихся результатом работы аналитических системами. Благодаря технологии SAS Integration Technologies, можно обеспечить доступ к таким данным из оперативных сред (например, для использования оператором центра обработки запросов) в реальном времени.

Список стандартов, поддерживаемых SAS, приведен в приложении А.

Клиент - серверные и распределенные вычислительные архитектуры

Технологии интеграции в SAS делают возможной быструю разработку полных систем, за счет предоставления независимым разработчикам приложений доступа к мощным серверам SAS по стандартным отраслевым протоколам.

Поддержка стандартов на распределенные объекты DCOM открывает сервер SAS для любой совместимой со стандартами интерфейсной системой любого независимого разработчика на любой платформе. Архитектура интеллектуальных бизнес-средств на базе Web, может опираться на стандарты CGI, ASP или JAVA.

Технологии SAS Integration Technologies предоставляют существующим пользователям SAS возможность контролируемого переноса ранее созданных клиент-серверных приложений SAS в

многозвенные архитектуры. Кроме того, SAS/IntrNet позволяет разрабатывать приложения для обработки информации, которые обращаются к данным SAS через динамические Web страницы.

Архитектура поддержки взаимодействия SAS IOM

Модель Integrated Object Model (IOM), поддерживаемая в SAS Integration Technologies, служит для предоставления распределенных объектных интерфейсов на базе программного обеспечения SAS. Она имеет такие функции, как процедурный язык скриптов, поддержку данных, файловых систем, результатов и форматирование данных. IOM дает возможность использовать стандартные языки программирования, инструментарии программирования и протоколы обеспечения взаимодействия для разработки клиентских программ, обращающихся к этим службам на сервере IOM.

Эта модель предполагает, что SAS выполняется на сервере IOM. Пользователи и приложения запрашивают объект Workspace на сервере. Запросы на Workspace поступают на генератор объектов, который выделяет новое рабочее пространство. Главное преимущество создания приложений на базе компонентов при помощи модели IOM состоит в том, что это позволяет разъединить бизнес-логику приложения и презентационные методы. Независимость этих двух составляющих дает возможность повторно использовать функциональность приложения, придавая ему различный «внешний вид» в зависимости от потребностей пользователя, следовательно, расширяется круг пользователей данного приложения.

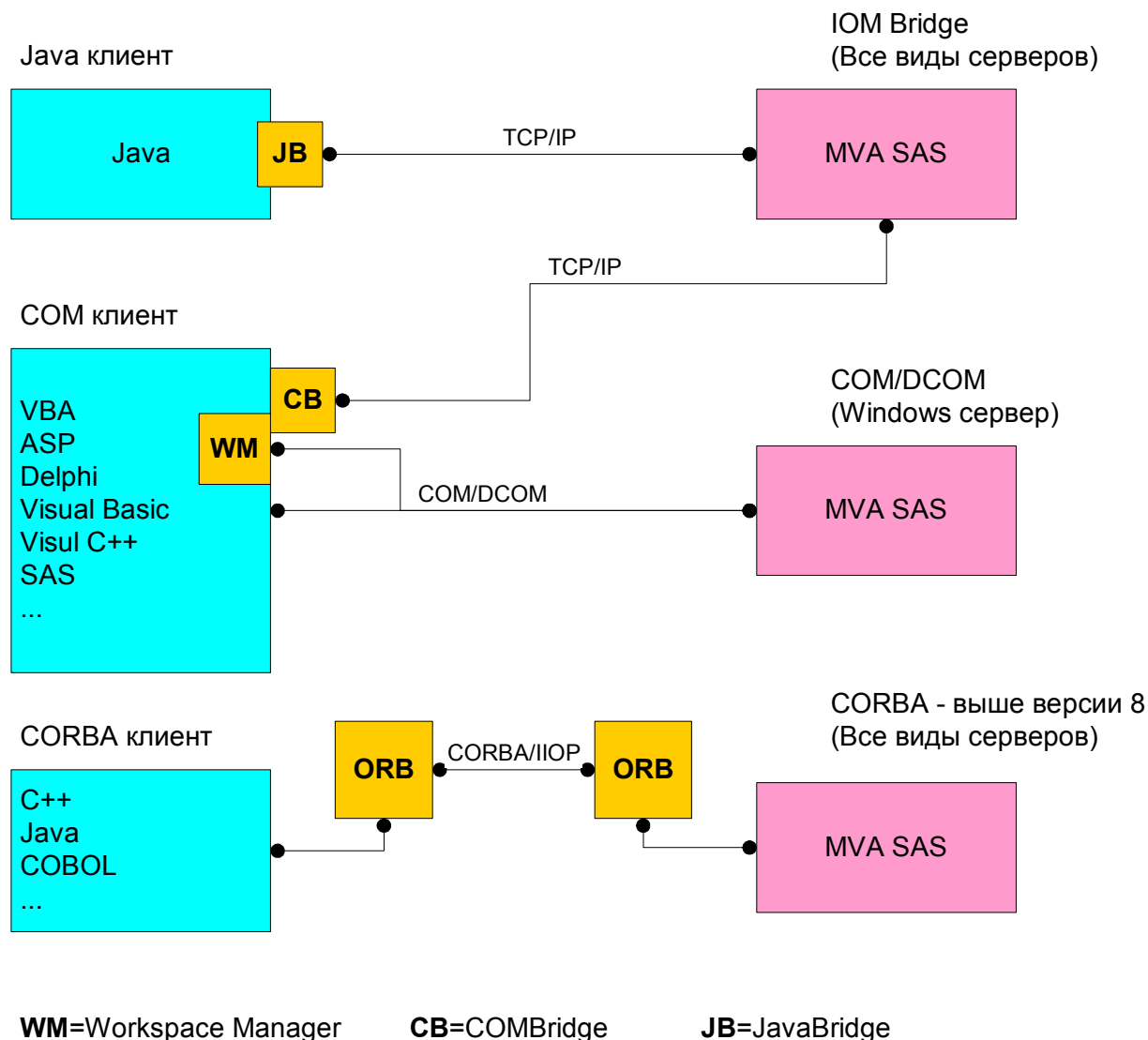


Рисунок 1. Архитектура поддержки взаимодействия SAS IOM

Публикация и подписка

Информационные архитектуры, обеспечивающие принудительную доставку результатов выполнения запросов или формирования отчетов, бесспорно, лучше всего подходят для крупномасштабных проектов. Однако по мере роста числа пользователей и отчетов, а также по мере увеличения объемов пакетов с требуемыми результатами, может возникнуть необходимость в оболочке «публикации и подписки», с более формальной структурой. SAS предоставляет мощные средства доставки информации по принципу публикации и подписки на базе протокола Lightweight Directory Access Protocol (LDAP). Эти средства позволяют создавать автоматические процессы быстрой доставки информации требуемым пользователям. Таким образом, ИТ открывают поле деятельности для разработки специальных функций подготовки отчетов, и возникает новая сфера

применения LDAP – обеспечение безопасной, основанной на стандартах масштабируемости на уровне предприятия и выполнения непосредственного «сверления» метаданных хранилища данных.

Портал доставки информации SAS Information Delivery Portal

Портал доставки информации SAS предоставляет пользователям единое рабочее пространство, из которого они могут обращаться к внутренним и внешним Web-серверам, выполнять специализированные запросы, запросы OLAP, а также просматривать содержимое каналов публикации при помощи персональных средств, например, Word, Excel и Acrobat.

Портал представляет собой промежуточный слой, приложение JAVA, предоставляющее многопользовательские, многозадачные службы кэширования и поддержки безопасности клиентам через Web-браузер. Настольные клиенты практически не требуют поддержки, архитектура отличается высокой масштабируемостью. Портал может запускать как функции SAS, так и функции систем других разработчиков.

Специалистам, опирающимся в своей работе на знания, приходится иметь дело со стремительно растущими объемами информации из всё большего числа источников. При этом нельзя найти двух специалистов, которые могли бы пользоваться абсолютно одинаковой информацией. Даже когда речь идет об одинаковом информационном наполнении, форм, в которой пользователи хотели бы получать его, великое множество. В портале эти проблемы решены за счет предоставления таким специалистам единой точки доступа к информации и необходимых им для работы инструментов.

SAS Information Delivery Portal делает многообразные функциональные возможности аналитических средств SAS доступными многочисленным квалифицированным специалистам, ответственным работникам и потребителям информации через единый настроенный Web-интерфейс. Таким образом, они могут полностью сосредоточиться на реальной работе с информацией, не отвлекаясь на попытки найти ее.

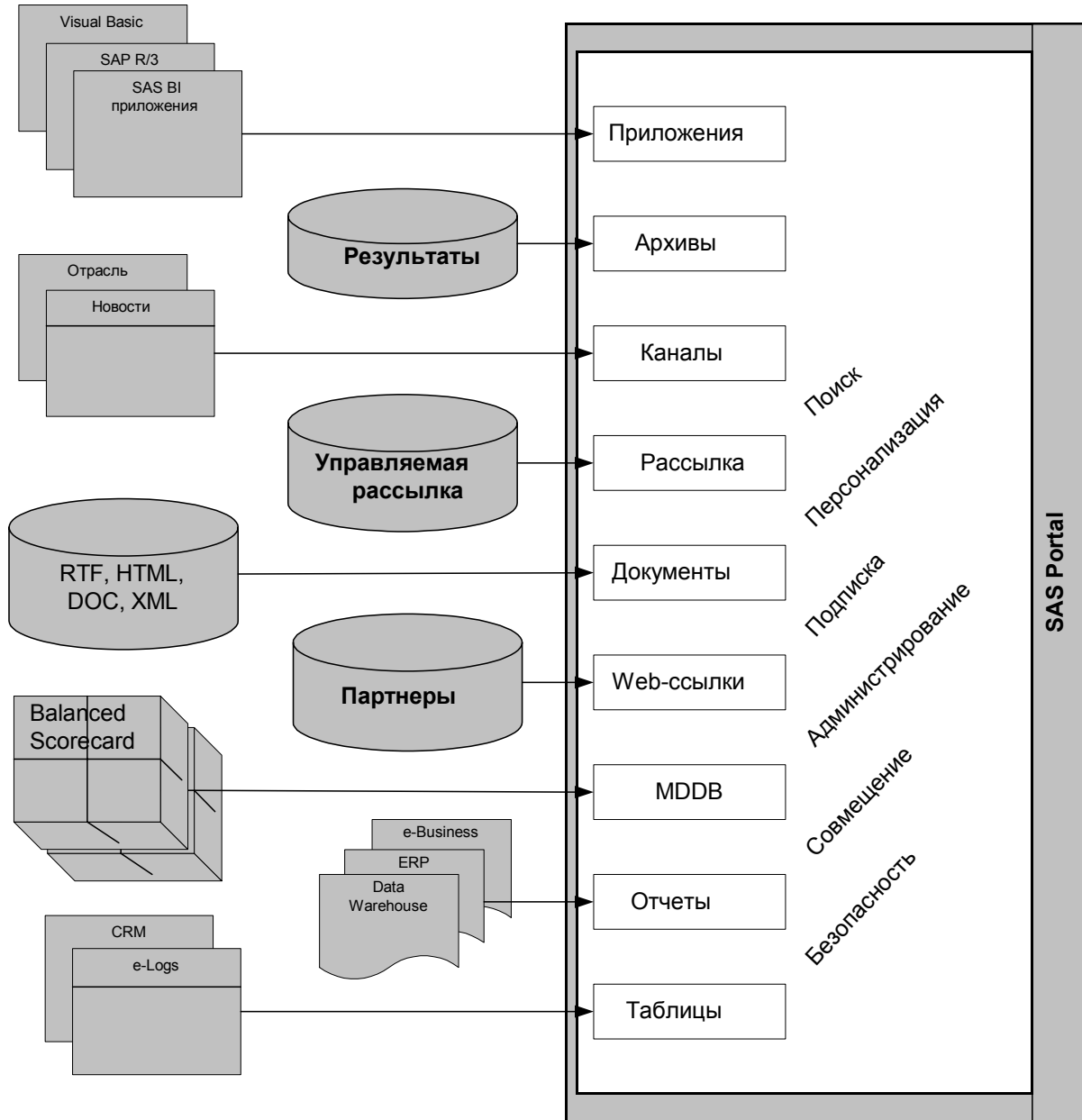


Рисунок 2. Портал доставки информации SAS

Архитектура портала SAS

Портал предоставляет возможности администрирования из единой точки и централизованного управления доступом к каналам публикации и подписки. Пользователи могут быть распределены по функциональным рабочим группам. Портал является настраиваемым, он поставляется с поддержкой функций OLAP и может быть сконфигурирован для работы с настольными системами Microsoft Exchange и Lotus. Портал легко интегрируется в существующую стратегию совместной работы. Благодаря возможности публиковать данные, получаемые не из системы SAS, в структуре каналов публикации, портал можно использовать для сортировки и распространения документов по всему предприятию. Кроме того, портал открывает новый потенциал стандарта каталогов LDAP. Схемы метаданных хранилища данных могут быть представлены в каталоге LDAP, что обеспечивает беспрецедентную возможность - полную прозрачность данных хранилища для Web-клиентов. Портал, кроме того, можно рассматривать, как способ предоставить пользователям интерфейс к средствам публикации и подписки SAS.

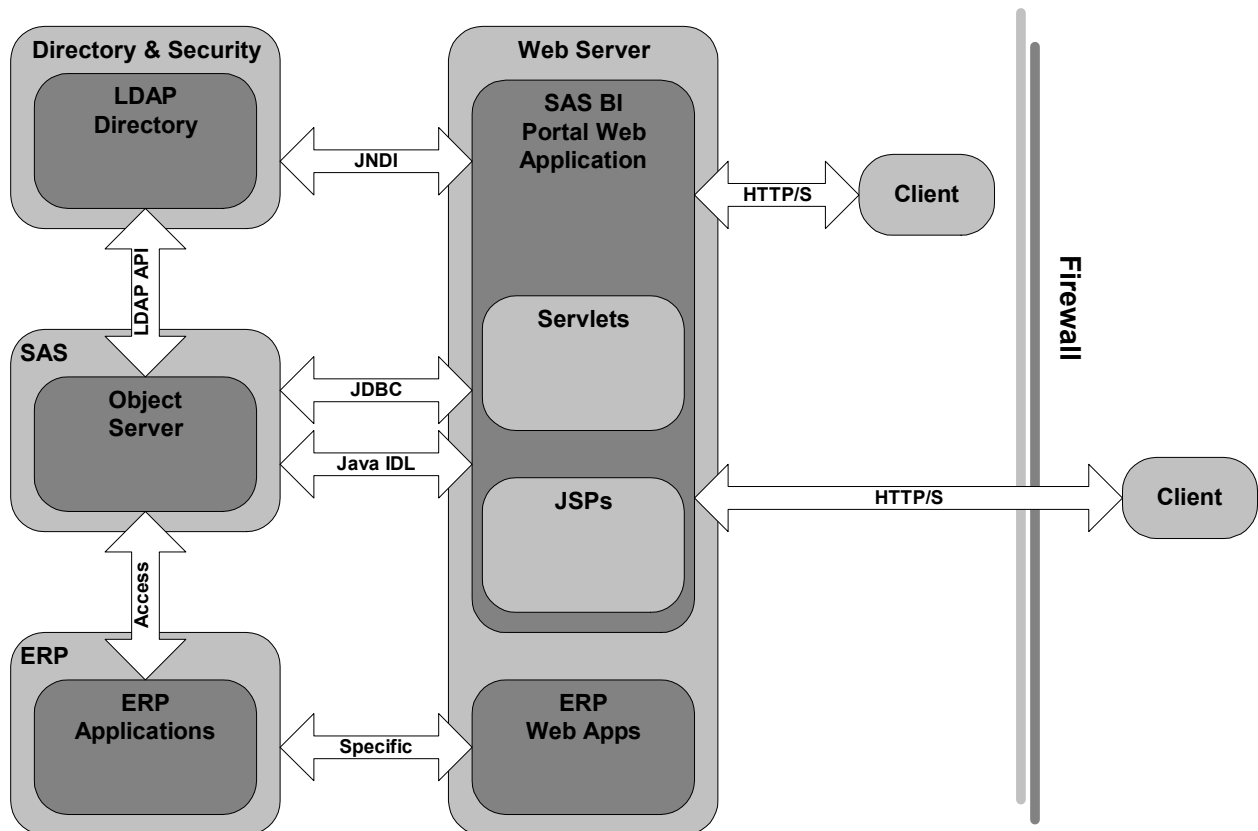


Рисунок 3. Архитектура портала SAS

Масштабируемость

Система может масштабироваться по двум показателям - числу пользователей и объему данных. Для того чтобы добиться масштабируемости среды доставки информации необходимо рассмотреть целый ряд особенностей, отличающих среду доставки информации от OLTP. В среде доставки информации:

- Транзакции, которым разрешено только чтение, не требуют блокировок на уровне строк, поддержки откатов и т. д.
- Каждая транзакция (запрос) может потребовать доступа к большим объемам данных, а не к небольшому числу строк.
- Запросы имеют непредсказуемую и динамичную природу, они могут также потребовать сложной обработки с большой загрузкой центрального процессора.

Обработка в Системе SAS

SAS - это интегрированная среда разработки и выполнения приложений. Конструкции языка SAS переводятся в размеченный код (tokenized code). На этапе обработки данных SAS вырабатывается машинный код, оптимизированный для операционной системы и центрального процессора. При этом обеспечивается минимальная нагрузка, создаваемая компилятором, поскольку в запросах, как правило, обрабатывается, большое число строк.

Процедуры SAS тоже представляют собой откомпилированные программы, в которых осуществляется синтаксический разбор управляющих предложений и выполнение соответствующих действий. Таким образом, в SAS обеспечивается высокая эффективность разработки и высокая скорость выполнения. Поток логически связанных задач часто называют сессией или сеансом SAS. Подразумевается, что сессии SAS "принадлежат" одному пользователю или клиенту, в то время как наборы данных SAS могут совместно использоваться несколькими пользователями в рамках нескольких сессий. Практика показывает, что такой подход позволяет поддерживать несколько сот одновременно работающих пользователей, которые обращаются к хранилищу данных с тонких клиентов. Многозадачность достигается благодаря поддержке выполнения Web-приложений средствами промежуточного слоя, например, SAS/IntrNet, или путем использования портала SAS, который также гарантирует защиту и предоставляет средства кэширования. Архитектуры такого рода стимулируют крупномасштабные внедрения систем на базе тонких клиентов.

Когда приложение должно поддерживать тысячи одновременно работающих пользователей, требуется истинная многопоточность. В это случае можно порекомендовать использовать высокопроизводительный сервер данных SAS. Хранилище данных сети Intranet на базе SAS High Performance Data Server SPDS и SAS/IntrNet предоставят поддержку очень крупных внедрений.

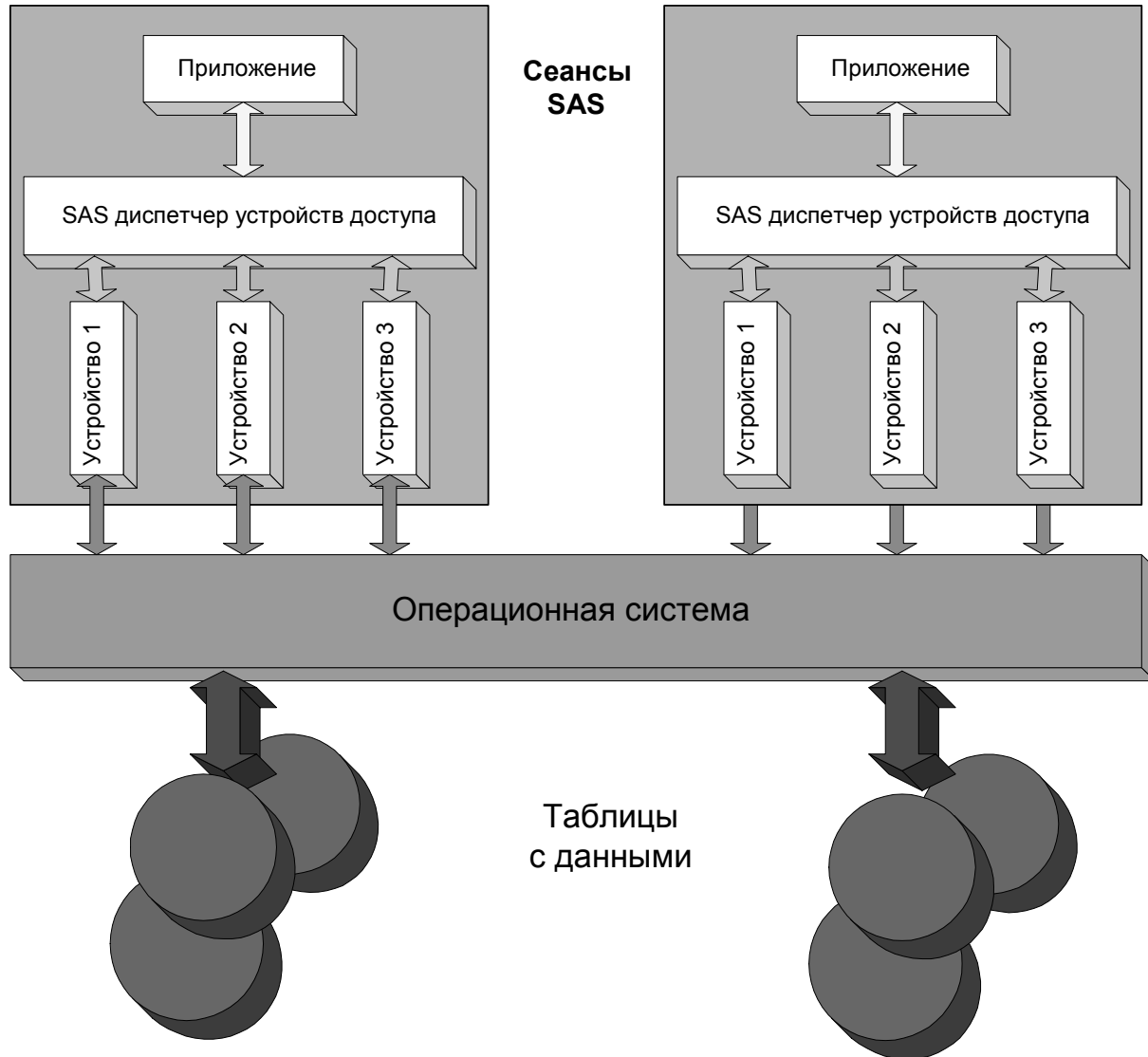


Рисунок 4. Сеансы SAS.

Базы данных для аналитических приложений

Технология систем управления базами данных (СУБД) разрабатывалась на основе опыта работы с управляемыми конфигурациями взаимосвязанных таблиц с поддержкой языков программирования четвертого поколения. Начиная в этого этапа эволюции СУБД, можно проследить путь развития двух типов семейств OLTP:

- баз данных OLTP, поддерживающих высокую скорость обновления для приложений с большим числом пользователей, и
- систем OLAP, или баз данных бизнес-знаний, оптимизированных для поддержки небольших групп запросов, для которых характерны, в основном, операции только чтения и множественные итерации циклов OLAP.

Первые, как правило, требуют таких функций, как журналирование и поддержка целостности на уровне ссылок на случай отката и восстановления. Последние, ориентированные на бизнес-знания, со временем развились в еще более специализированные структуры, например, многомерные базы данных.

Управление данными и производительностью

SAS является информационно-аналитической базой данных и может управлять совокупностью отдельных наборов данных (или таблиц) как единой интегрированной базой данных. Таблицы организованы в библиотеки. В каждой таблице данные хранятся в виде строк и столбцов. При необходимости можно определить индексы; тем не менее, на практике индексы используются редко, поскольку большинство задач аналитической обработки требуют чтения или объединения таблиц целиком. Две, приведенных ниже, особенности характеризуют SAS и объясняют за счет чего достигается высокая производительность в аналитических системах, созданных на базе SAS.

Журналирование

Так как журналирования не производится, заполнение хранилища данных или добавление данных в таблицы выполняется очень быстро. В случае сбоя, который вызывает некорректное завершение процесса, быстрее выполнить работу заново, чем производить откат по журналу. На практике некорректные завершения происходят чаще при работе с OLTP, поскольку внутренние управляющие таблицы реляционной СУБД быстро переполняются, если только не запрограммировано большое число контрольных точек повторного запуска. SAS поддерживает неограниченное число контрольных точек повторного запуска. Процессы SAS ETL можно сравнить с утилитой массовой загрузки РСУБД.

Сегментация и разбиение файлов

В целях оптимизации доступа по чтению к группе взаимосвязанных таблиц, в средах OLTP считается нормой специально сегментировать и распределять таблицы. Но если обработка предусматривает чтение целых таблиц, делать этого не рекомендуется. В SAS таблицы хранятся

целиком, при этом таблицы SAS имеют фиксированную длину записи. Такая структура гарантирует высокую производительность центрального процессора и позволяет эффективно поддерживать большое число столбцов, что играет существенную роль при добыче данных.

В результате применения этих методик, загрузка таблиц SAS и обработка запросов к ним выполняется на несколько порядков быстрее, чем в традиционных РСУБД, при объемах данных, превышающих 100 Гбайт.

Целостность на уровне ссылок и триггеры в хранилище данных не применяются. Целостность обычно обеспечивается за счет корректной разработки ETL-процессов. В случае необходимости, можно периодически выполнять специальные процессы для обеспечения качества данных, которые выявляют отклонения в данных, такие как "висячие" или повторные записи. На этапе проектирования физической структуры таблиц нормой считается отказ от поддержки нормальных форм, т.е. денормализация.

Проектирование таблиц

Проектирование таблиц обычно предполагает два этапа - проектирование логической и физической структуры. Общепринято, что в зависимости от того, с какой точки зрения рассматриваются сущности данных - с точки зрения аналитического приложения, или оперативного приложения, между этими сущностями выявляются разные взаимосвязи. В операционной базе данных таблицы, как правило, имеют один первичный ключ, и, возможно один или два вторичных. Ключи поддерживают наиболее распространенные маршруты доступа и возвращают небольшое число собственных строк (например, комплектующие под заказ). В логической структуре используются формальные логические группы размерностей, или характеристик, и аналитических переменных, причем последние являются численными значениями и подлежат статистической обработке и анализу. Концептуальное представление обычно изображается звездообразной или многолучевой схемой. Необходимо обратить внимание, что при проектировании физической схемы как аналитической, так и операционной базы, считается обычной практикой отказ от нормальных форм логической модели, введение внешних ключей для повышения эффективности работы на физическом уровне. В действительности для некоторых приложений по добыче данных полный отказ от нормальных форм необходим, поскольку, по определению, взаимосвязь между характеристиками заранее неизвестна.

Инструменты создания моделей данных

Инструменты создания моделей данных существенным образом сокращают время, необходимое для разработки логической модели данных. Роль этих инструментов особенно велика на этапах анализа требований и создания проекта логики приложения. Материалы, вырабатываемые этими инструментами, часто экспортируются как файлы в формате RTF и затем включаются в документы Microsoft Word, которые используются разработчиками в обсуждениях проекта во время рабочих встреч. И все же часто, по причинам, приведенным выше, язык определения данных (data definition language, DDL) не в состоянии описать во всех тонкостях и создать требуемую физическую структуру. При этом не имеет значения, какая выбрана РСУБД. Схема «звезда» требуется весьма часто, поэтому было бы естественно ожидать, чтобы инструменты вырабатывали именно ее.

Однако маловероятно, чтобы любой инструментарий мог создавать для SAS оптимально настроенную структуру HOLAP. Такой логический куб может быть реализован, например, как комбинация нескольких различных физических наборов данных разных форматов, скажем, MDDB¹, SAS, SPDS² и PCУБД. Эти наборы могут, в свою очередь, размещаться на различных машинах, в различных операционных системах.

Репозитарий хранилища

Для хранения данных в SAS используются различные репозитории, наиболее подходящие для того или иного приложения. Для аналитического приложения, не предъявляющего высоких требований к ресурсам, оптимальной может быть стандартная PCУБД в сочетании с SAS/ACCESS. Для высокопроизводительного приложения OLAP, в SAS предусмотрена выделенная MDDB. Приложениям EIS или webEIS может потребоваться доступ к различным комбинациям структур. В обработку средствами OLAP могут быть вовлечены самые различные структуры - таблицы SAS, таблицы MDDB и таблицы внешних систем, отличных от SAS, при этом вся конфигурация управляется диспетчером данных HOLAP через метаданные. Метаданные обеспечивают независимость разработки структур данных и приложений, т.е. приложение может не знать к каким форматам данных будет осуществлен запрос.

Для достижения максимальной производительности можно порекомендовать сервер данных SAS Scalable Performance Data Server (SPDS). Масштабируемый сервер данных SPDS представляет собой многопользовательский сервер, обеспечивающий поиск и выдачу данных в хранилищах данных. SPDS отличается высокой масштабируемостью по производительности при работе с таблицами большого объема. В его состав входит централизованный сервер имен и защищенный сервер данных, идентификация пользователя на котором производится при помощи идентификатора и пароля. Такая структура обеспечивает высокую готовность сервера. SPDS использует также возможности симметричной многопроцессорной архитектуры (SMP) за счет целевого использования потоков для выполнения параллельных или перекрывающихся нитей обработки.

Подготовка отчетов и обработка запросов

Рынок средств подготовки отчетов и обработки запросов исключительно многообразен. Многие инструменты OEM-производителей претендуют на предоставление полного спектра услуг широкому кругу пользователей. Средства для формирования отчетов можно условно разделить на несколько групп.

¹ MDDB (*Multidimensional Database*) – Многомерная база данных; формат хранения в SAS многомерных кубов.

² SPDS (*Scalable Performance Data Server*) – Масштабируемый сервер данных. Средство оптимального хранения и параллельной обработки больших объемов данных SAS. Разработано для многопроцессорных систем, поддерживающих SMP -архитектуру.

- **Интерактивные средства подготовки отчетов.** Некая центральная группа (например, подразделение ИТ или информационный центр) выполняет проектирование, стандартизацию и выработку графика подготовки отчетов и обработки запросов, которые доставляются большому числу пользователей.
- **Специализированный анализ и OLAP.** Во многих организациях есть группы аналитиков, работа которых опирается на сложный анализ. Таким пользователям необходим интерактивный доступ ко всем возможностям SAS.
- **Корпоративные Web-клиенты.** Главным требованием, для удовлетворения которого применяется этот сценарий, состоит в предоставлении простого в использовании клиентского приложения большому числу пользователей, имеющих самое различное географическое положение. Таким пользователям нужна периодическая помощь аналитического приложения. Функциональность приложения может включать в себя интерактивные специализированные запросы или OLAP. Под web-клиентами понимаются как пользователи Internet, так и пользователи Intranet. Причем, в случае с информационным хранилищем, вторые встречаются чаще, т.к. хранилище обычно содержит конфиденциальную информацию.

Суть подхода SAS состоит в том, чтобы применять наиболее подходящие средства и инструменты для решения той или иной задачи.

Интерактивные средства подготовки отчетов и OLAP

SAS содержит ряд средств для интерактивного составления табличных и графических отчетов пользователями. Эти средства не требуют специальных глубоких программистских знаний и навыков. Отчеты могут быть сохранены в формате HTML и подготовлены для публикации на web-сервере. Распространять отчеты можно как в интерактивном, так и в пакетном режиме. Результаты готовы для просмотра многочисленными пользователями и пользовательскими группами.

Системы OLAP реализуются либо как клиент-серверное приложение SAS, либо как приложение на базе Web. Большинство клиентов предпочитают выбирать опцию "Хранилище данных с поддержкой Intranet".

Известно, что производительность OLAP-приложения в значительной степени зависит от объема данных и способа хранения, ROLAP или MOLAP. В подходе SAS, HOLAP, разработчик может выбрать комбинацию обеих технологий, что означает большую гибкость. HOLAP создает масштабируемую среду для многомерных приложений. Прежде, разработчикам таких приложений, приходилось выбирать между различными технологиями хранения данных в базе данных. Решение HOLAP объединяет эти технологии в SAS, образуя аналитическую архитектуру, отличающуюся масштабируемостью и многомерностью.

Повышение производительности достигается за счет поддержки использования серверных вычислительных служб. Кэш, позволяющий повторно использовать результаты недавно обработанных запросов, может существенно повысить эффективность за счет более быстрой инициализации многомерных объектов. Когда требуется обеспечить масштабируемость OLAP,

HOLAP может распределить данные по многочисленным источникам данных. Благодаря использованию многочисленных источников данных можно организовать хранение некоторых данных по звездообразной схеме, частично в наборах данных SAS и частично в MDDB. При этом не исключено, что все источники физически расположены на совершенно разных машинах. Разработчик получает большой "простор для маневра", повышает гибкость системы хранения OLAP. Это означает, что разработчик может выбрать наиболее подходящий тип системы хранения и размещение приложений, повторно использовать результаты уже проделанной работы и искать подходящую машину в сети для доставки требуемых результатов. Главное преимущество HOLAP состоит в том, что впоследствии вам не придется перестраивать гигантский куб, когда возможности существующего хранения со временем будут исчерпаны. Результаты OLAP-анализа можно экспортировать в другие приложения, например, в Excel, или опубликовать на web.

Специализированный анализ

Профессиональному пользователю SAS, решающему сложные задачи, предоставляются дополнительные возможности. Высокоэффективный Windows-клиент Enterprise Guide (EG) предоставляет всеобъемлющие и мощные аналитические и статистические функции, средства обработки запросов, и в дополнение к ним оболочку управления проектом и публикации результатов. Аналитические службы предоставляются при помощи клиента на ПК или через распределенные серверы SAS. Enterprise Guide представляет собой, по сути, интерфейс пользователя ко всем функциям и возможностям системы SAS – от преобразования данных и формирования табличных отчетов, до статистических исследований и решения задач прогнозирования и оптимизации. EG может также выступать в роли обучающей системы для разработчика, т.к. автоматически генерирует код программ на языке SAS, который может быть сохранен для последующего пакетного запуска.

Пакетная обработка и форматы

Все отчеты создаваемые Системой SAS могут быть получены как интерактивно, так и как результат пакетной обработки по заданному регламенту. Это гарантирует, что пользователи будут получать отчеты, основанные на актуальной информации, вовремя.

SAS предоставляет разнообразные способы форматирования результатов. Система SAS поддерживает все многообразие форматов и форм представления, в которые нужно или желательно выводить результаты. Среди них можно назвать XML, HTML, PDF и RTF. SAS постоянно ведет разработку новых возможностей Системы SAS, что даст пользователям возможность значительно лучше управлять выводом аналитических результатов и составлением отчетов. Средствами SAS можно создавать отчеты, соответствующие архитектуре или внешнему оформлению Web-браузера. В действительности, только одна строка кода - это все, что нужно, чтобы отправить результаты выполненной SAS работы напрямую в вашу сеть Intranet.

SAS может помочь и в упаковке выходных данных, полученных не в системе SAS, в передаче файлов в форматах csv, HTML, RTF и PDF из исходных приложений, таких как SAP R/3 в оболочку SAS Portal. Опираясь на эти возможности можно выработать интегрированную стратегию, которая

гарантировала бы, что любой отчет или результат запросов, можно будет прочесть на стандартном настольном компьютере в локальной вычислительной сети.

Эксплуатационные моменты

Программы SAS хранятся в виде исходных кодов; следовательно, нет необходимости в управлении исходной, объектной и исполняемой версиями. Разработчик SAS определяет, нужно ли хранить промежуточные таблицы и создавать потоки обработки для повторного запуска на той или иной стадии или в случае аварийного завершения.

Безопасность

В Системе SAS поддерживаются три уровня защиты

- Защита на уровне операционной системы
- Защита на уровне таблиц данных SAS
- Защита на уровне таблиц данных базы данных.

Таблицы данных (SAS tables) и представления данных (SAS view), полученные на различных этапах, могут быть защищены паролем. В SAS поддерживаются различные уровни защиты по паролю. Для хранилищ данных в этом случае наиболее важен пароль на чтение. Модифицировать данные в хранилище могут только процессы загрузки данных, поэтому пароль на запись или модификацию доступен только этим процессам или операторам загрузки.

Многие операционные системы (например, все многопользовательские системы) предоставляют дополнительные возможности защиты для файлов. В этом случае доступ к файлам (таблицам данных SAS, плоским файлам и РСУБД) определяется идентификатором пользователя, который используется им для входа в операционную систему. В SAS защита на уровне операционной системы играет важную роль.

В РСУБД, как правило, защита данных осуществляется посредством дополнительных, частных систем защиты, и эта информация о безопасности используется SAS при организации хранения и при доступе к таблицам, хранящимся в РСУБД, из приложений, написанных в SAS. Все доступные системы защиты и соответствующие измерения взаимодействуют друг с другом и могут при необходимости использоваться внутри SAS.

Разграничение функций

Нередко необходимо предоставить различные функциональные возможности различным пользователям приложения. Например, исполнительный директор должен иметь возможность выполнять все функции и просматривать все данные. С другой стороны, специалисту, работающему в местном филиале, разрешен доступ только к данным, связанным с географическим регионом, где находится филиал. Кроме того, этому специалисту нельзя осуществлять операций «сверления» данных для определения показателей работы отдельных торговых агентов. Конечно, хотелось бы написать приложение только один раз, обеспечив, однако, чтобы оно выглядело и функционировало

по-разному в зависимости от того, кто его использует. Этого можно добиться либо запрашивая идентификатор пользователя и пароль при запуске приложения, либо определяя, каков идентификатор пользователя операционной систем. Приложения, разрабатываемые в SAS, можно создавать таким образом, чтобы они самостоятельно модифицировались в зависимости от идентификаторов пользователя. Такие приложения выглядят по-разному для разных пользователей.

Для поддержки разработки приложений и обеспечения большей гибкости, пользовательские идентификаторы можно группировать в группы. Тогда вместо использования идентификатора пользователя, для определения того, можно ли данному пользователю обращаться к тем или иным данным или функциональным возможностям, используется группа пользователя.

И, наконец, не все данные пригодны для тех или иных приложений. До сих пор речь шла об обеспечении или запрещении доступа к данным или приложениям разных пользователей или групп пользователей. Не менее важно, чтобы определенные данные были доступны вполне определенным приложениям. Для гарантии того, что такие данные недоступны извне приложения, все критически важные наборы данных, виды наборов данных или виды доступа к РСУДБ должны быть защищены паролем, при этом пароли должны быть доступны приложению, а никак не пользователю.

Списки контроля доступа

Защита на базе списков контроля доступа (access control list, ACL) поддерживается в приложениях на базе SAS webEIS и на сервере SAS SPDS (Scalable Performance Data Server).

Шифрование

Таблицы данных SAS могут быть зашифрованы таким образом, чтобы их нельзя было прочитать при простом просмотре с использованием того или иного средства чтения плоских файлов или с использованием генератора дампов. Однако специалисты, занимающиеся разработкой и созданием приложений, должны иметь в виду, что шифрование и декодирование таблиц данных SAS (или любых других таблиц любым другим программным обеспечением, предназначенным для этой цели) создает большую нагрузку и неблагоприятно отражается на производительности.

Помимо шифрования таблиц SAS Institute предлагает средства для шифрования сетевых протоколов для обеспечения защиты передачи данных по сети. В программном обеспечении SAS/SECURE используются криптографические службы, предоставляемые средствами шифрования Bsafe компании RSA и CryptoAPI компании Microsoft, которые подпадают под действие законодательства об экспорте ключей шифрования³. В качестве альтернативы SAS предлагает встроенные алгоритмы шифрования, входящие в состав программного обеспечения Base SAS.

³ Для выяснения доступности того или иного алгоритма шифрования в Вашей стране обратитесь в представительство SAS.

Планирование процессов

Метод планирования процессов, выполняемых в хранилище данных, как правило, зависит от типа и размера хранилища данных. Это можно пояснить на двух сценариях.

- **Хранилище небольшого или среднего размера, не имеющее привязок к регламенту работы оперативных систем** (например, финансовому инструменту, работа с которыми ведется пользовательскими подразделениями). В этом случае хранилище данных не связано с оперативными процессами (например, закрытие операционного дня), поэтому график его обновления поддерживается независимо. Это простейший способ планировать график обновлений с консоли администратора SAS/Warehouse Administrator. SAS/Warehouse Administrator предоставляет мощные встроенные функции планирования заданий.
- **Хранилища более крупного размера, а так же связанные с регламентом оперативных систем** (например, управленческие отчеты должны быть согласованы с бухгалтерскими книгами). В этом случае обновления должны быть тщательно синхронизированы и согласованы с финансовыми и бухгалтерскими процессами в компании. Лучше всего это делать в центре АСУ, где проще манипулировать взаимозависимостями между задачами. В данном сценарии можно порекомендовать экспортировать поток процессов SAS/Warehouse Administrator для ввода в стандартный пакет запуска заданий. Помимо этого SAS может предоставить интерфейс оператора, через который можно управлять как процессами в SAS, так и процессами вне SAS.

Резервное копирование и восстановление

Сейчас закрепилась хорошая практика делать резервные копии приложений перед загрузкой в хранилище данных, пересозданием киосков данных, содержащих отчеты, или выполнением агрегирования. Это сокращает время восстановления в случае аварийного окончания.

Сегментирование процессов в хранилище данных упрощает реализацию этого этапа. SAS может автоматически предоставлять поддержку рестарта в контрольных точках.

Управление конфигурацией приложений SAS

Управление конфигурацией – это процесс контролируемого перехода между средами разработки, тестирования и выполнения. Интегрированная среда разработки или архитектура на базе ядра SAS означает, что приложения SAS должны обрабатываться как одно целое при переходе приложения между этими средами. Нельзя перемещать одно задание или модуль; с точки зрения традиционных ИТ, должна перемещаться библиотека программ целиком. Хорошим аналогом в данном случае является SAP R/3. Необходимо инсталлировать и протестировать новый релиз целиком перед заменой предыдущей версии на новую (хотя и в R/3 предусмотрена возможность установления заплаток). Как правило, это не составляет трудности, если принять концепцию в принципе. SAS готова предоставить утилиты поддержки перехода для создания такой конфигурации управления, которая необходима клиенту.

Контроль целостности

Хотя в SAS поддерживаются средства поддержания физической целостности, в среде "доставки информации" контроль целостности лучше всего осуществлять при выполнении этапа преобразований и очистки данных - процесса ETL. Контроль целостности на этом этапе упрощает подготовку отчета об ошибках и предотвращает добавление в базу данных неполных пакетов

обновлений. Кроме того, это повышает эффективность, поскольку при этом не нужно создавать индексы исключительно для контроля выполнения ограничений, а также добавление строки не выливается при нарушении ограничения в сложную операцию отката всей транзакции, создающую большую нагрузку на систему. Кроме того, можно использовать утилиту непосредственной массовой загрузки, если таковая применима, и если не нужно соблюдать физические ограничения базы данных.

Интеграция корпоративной архитектуры

В современных сложных системах уровня предприятия, гигантские объемы ценных данных хранятся на многообразных взаимосвязанных платформах, в базах данных и приложениях. Создание целостной системы подготовки отчетов и коммуникаций между системами представляет собой сложнейшую задачу. ERP позволяет решить эту задачу только частично. Настоящие гибкие средства нельзя просто купить в магазине. Их необходимо проектировать. Система SAS предоставляет исчерпывающие и полностью открытые технологии, поддерживающие интеграцию предприятия.

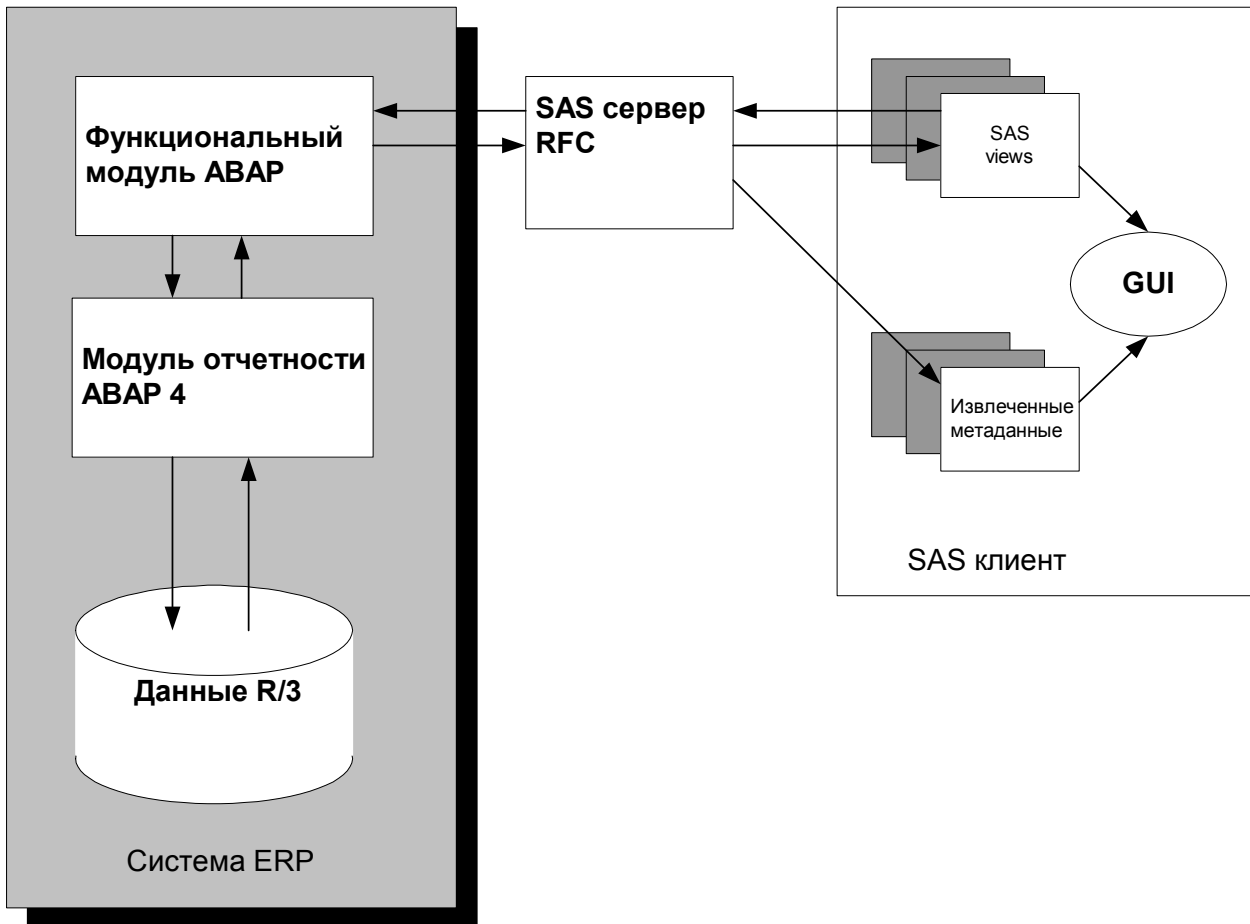
Обработка сообщений

Архитектура крупных предприятий во многом определяется стандартами обработки сообщений, обеспечивающими надежные коммуникации внутри приложения. Для упрощения интеграции в стандартные оболочки обработки сообщений SAS версии 8 и более поздние версии поддерживают работу с серверами сообщений MQSeries и MSMQ. Эта поддержка означает, что SAS может играть роль приложения-источника и целевого приложения в рамках средств интеграции, обмениваясь транзакциями в ERP-системах. Все транзакции данных могут выполняться концентратором Enterprise Architecture Integration (EAI) или хранилищем данных. Перед загрузкой с хранилище сообщения могут накапливаться в области ODS, при необходимости это может быть реализовано на базе событий. Другой вариант, при работе с SAS/ACCESS to ERP, можно использовать технологии RFC.

Интеграция ERP при помощи технологии RFC

SAS/Warehouse Administrator - это лидирующий в отрасли инструмент для управления несколькими разнородными физическими средами. Это программное обеспечение предоставляет единый вид на оперативные базы данных и базы данных, используемые в хранилищах.

Компания SAS лидирует в мире в разработке технологий, позволяющих извлекать "сырые", необработанные данные из систем поддержки текущих операций и помещать эти данные в среды, ориентированные на аналитические операции. Последнее поколение программного обеспечения SAS делает процесс доступа к многочисленным разрозненным источникам данных (в том числе DB2, BAAN, R/3, Sybase, ODBC, Informix, Ingres, Oracle и другим) еще более прозрачным. SAS



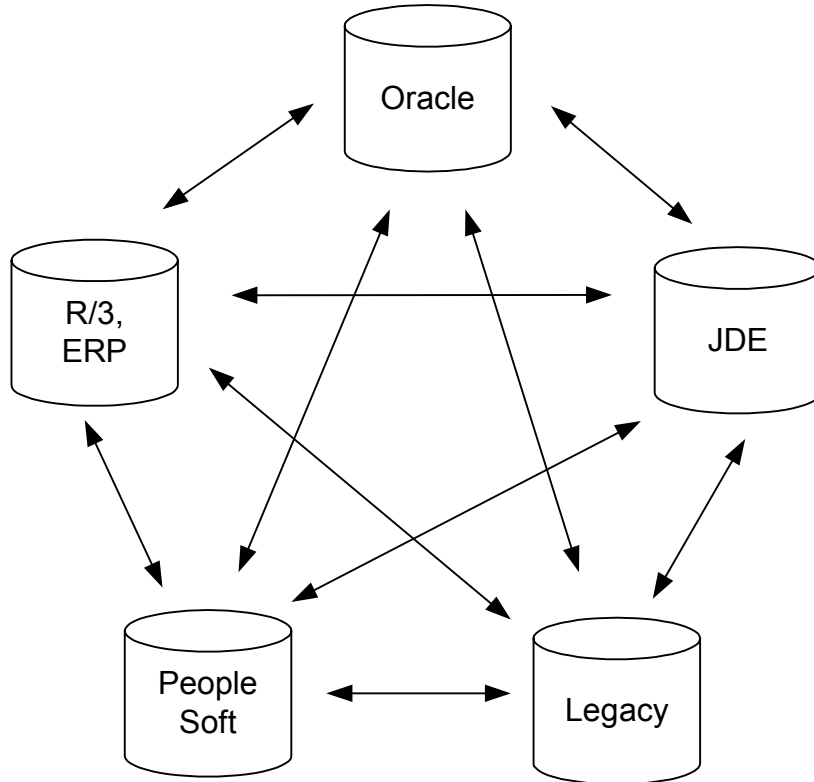
предоставляет уникальные и исключительно мощные возможности создавать и управлять видами данных, интегрирующими в себе многочисленные экземпляры в форматах ERP-систем, форматах производителей и РСУБД, сохраняя в то же время целостность метаданных исходной среды.

SAS предоставляет также улучшенные возможности доступа к хранилищу данных Business Warehouse системы SAP, что позволяет формировать полный вид предприятия, объединяющий в себе не только киоски данных SAP, но и иные данные. SAS поддерживает архитектуры обработки сообщений, и SAS/Warehouse Administrator может обмениваться информацией с промежуточным программным инструментарием EAI.

Рис. 5 Доступ ко множеству ERP систем

Проблема EAI: Интерфейс «спагетти»

Суть проблемы EAI в присутствии множества прямых связей между компонентами.



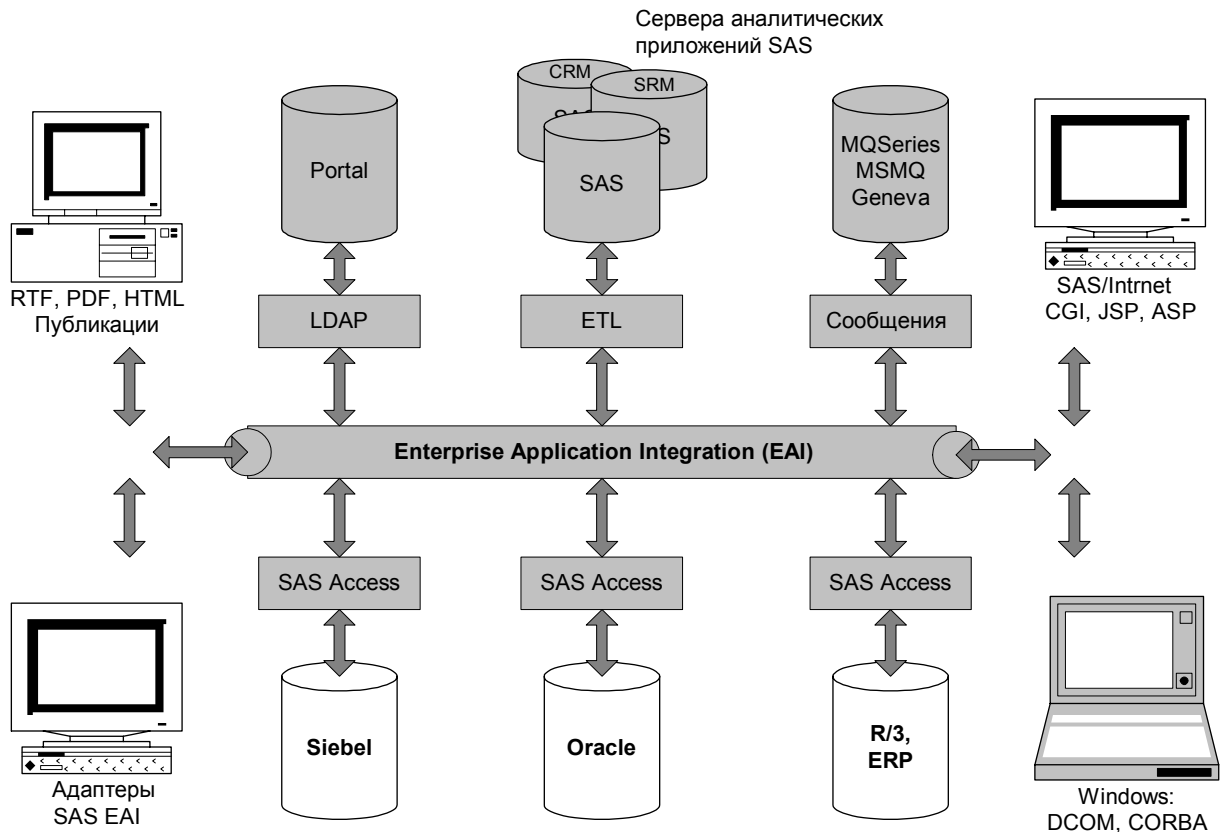
Зис 6. Интерфейс «спагетти»

SAS Integration Technologies: среда для создания EAI

Решение EAI основано на следующих компонентах:

- Структурированный набор протоколов для связи приложений.
- Системная архитектура, в которой однозначно определяются системы записей и репозитариев.
- Корпоративная "шина данных", которая способна поддерживать обработку сообщений и компоненты распределенных приложений.
- Набор адаптеров приложений.

За счет включения аналитических решений SAS и SAS Integration Technologies в оболочку EAI, обеспечивается действительная интеграция приложений уровня предприятия. Таким образом, можно организовать совместную работу в n-уровневых архитектурах с источниками данных операционных приложений и доставку результатов работы интеллектуальных бизнес-приложений на клиенты под управлением Windows или на базе браузеров в n-уровневой архитектуре



принудительной доставки и извлечения.

Рис 7. Среда SAS AEI