

Платформа Microsoft для бизнес-аналитики

Сергей Минюров

Октябрь, 2013

В области разработки интеграционных и аналитических решений продукты Microsoft является признанным лидером мирового рынка.

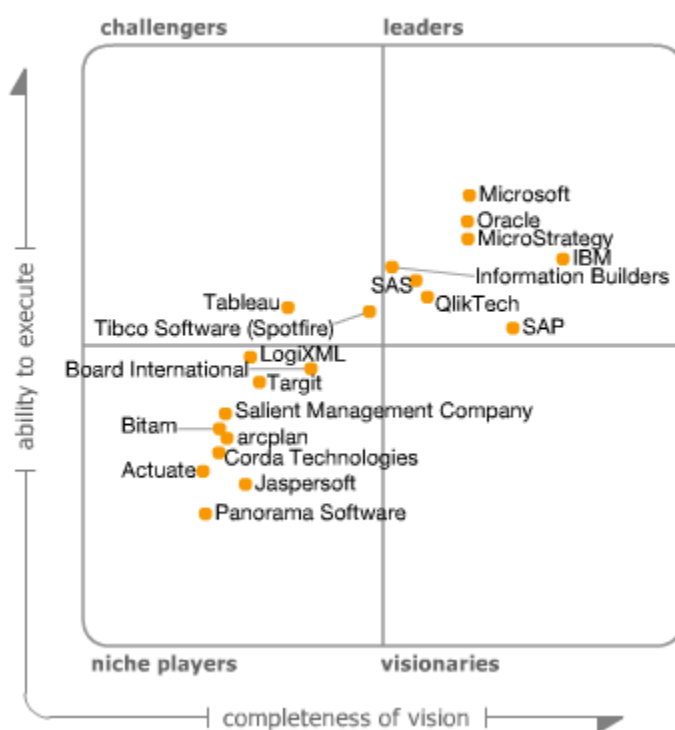


Рисунок 1. Microsoft признана Гартнер как технологический лидер рынка бизнес-аналитики

При использовании линейки продуктов Microsoft достигается максимальная совместимость и экономия при разработке технических решений, их внедрении и поддержке.

Продукты Microsoft, используемые для бизнес-аналитики

Используемые продукты Microsoft:

1. **Microsoft Windows Server** – операционная система как базовая технологическая платформа, обеспечивающая развертывание и поддержку программных решений, управление безопасностью и производительностью.

2. **Microsoft Office** – инструмент интерактивного анализа данных и разработки самими пользователями аналитических моделей данных и представлений.
 - **Microsoft Excel** – разработка пользовательских моделей анализа данных (PowerPivot) и визуальных аналитических представлений (сводные таблицы и диаграммы, цифровые панели).
 - **Microsoft Visio** – разработка визуальных аналитических представлений (стратегические карты, деловые процессы).
 - **Microsoft Access** – разработка пользовательских приложений по вводу первичных данных, автоматизации уровня команд и подразделений, оперативной отчетности.
3. **Microsoft Visual Studio** – инструмент разработки прикладных решений, в т. ч. в области бизнес-аналитики для профессиональных программистов. Для разработки решений используются следующие программные технологии (программные каркасы):
 - **ADO.NET** – технология доступа к реляционным базам данных, используется для разработки приложений, работающих на основе баз данных.
 - **Windows Communication Foundation (WCF)** – технология разработки сервисов и распределенных программ, используется для разработки сервисов (в т.ч. шины интеграции данных совместно с Windows Workflow Foundation или BizTalk), подключения приложений к шине интеграции данных, мастер-данным, сервисам SharePoint и т.д.
 - **Windows Workflow Foundation (WWF)** – технология разработки решений для автоматизации деловых процессов для прикладных программ либо приложений SharePoint.
 - **ASP.NET** – технология разработки публичных и корпоративных веб-приложений, в т.ч. приложений SharePoint.
 - Программные расширения на основе программных интерфейсов (API) для SQL Server, SharePoint и BizTalk.
4. **Microsoft Internet Information Server (IIS)** – веб-сервер для работы прикладных сервисов и приложений, в т.ч. Microsoft SharePoint и сервисов Microsoft SQL Server (см. далее).
5. **Microsoft SQL Server** – платформа для управления базами данных, интеграции, анализа и визуализации данных.
 - **СУБД (Database Engine)** – система управления реляционными базами данных, являющихся основой для работы прикладных программ, обеспечивающих первичный ввод данных и автоматизацию деятельности пользователей.
 - **Сервис интеграции данных (Integration Services, SSIS)** – платформа для разработки интеграционных решений по сбору, обработке и загрузке данных (ETL).
 - **Сервис анализа данных (Analysis Services, SSAS): OLAP, Tabular Mode, Data Mining** – платформа для разработки аналитических хранилищ данных на основе семантических моделей данных (BI semantic model).
 - **Сервис отчетности (Reporting Services, SSRS)** – платформа для разработки отчетов любого уровня сложности со встроенными возможностями публикации отчетов (в т.ч. в SharePoint), доставки по расписанию и конвертации в различные форматы.

- **Сервис качества данных** (Data Quality Services, DQS) – приложение для анализа качества данных и созданию баз знаний для автоматической проверки нормализации данных (с помощью сервиса интеграции).
 - **Сервис мастер-данных** (Data Master Services, DMS) – приложение для создания и управления едиными справочниками для информационных систем организации.
6. **Microsoft SharePoint** – портал для совместного доступа пользователей к оперативной и аналитической информации, их взаимодействия при совместном решении задач.
- **Excel Services** – сервис для публикации на SharePoint данных и визуальных представлений пользователями на основе Excel.
 - **PowerView** – инструмент для разработки пользователями диаграмм по анализу данных.
 - **Visio Services** – сервис для публикации на SharePoint диаграмм анализа данных и деловых процессов на основе Visio.
 - **PerformancePoint** – конструктор отчетов и диаграмм, цифровых панелей и стратегических карт для визуализации данных на основе многомерных и реляционных баз данных и пользовательских источников данных, разработанных в Excel.
 - **Access Services** – сервис для публикации на SharePoint баз данных для ввода данных, автоматизации работы команд и подразделений.
7. **Microsoft Azure** – публичные программные сервисы («облачная» платформа), обеспечивающие возможность удаленной работы с базами данных, отчетностью, офисными приложениями и порталом на SharePoint без развертывания собственной программно-технической инфраструктуры. Виды сервисов и сценарии работы:
- Ведение единых справочников для информационных систем, работающих в подразделениях и партнерских компаниях на основе сервиса данных SQL Azure.
 - Создание публичной интеграционной шины для автоматизации обработки заявок от населения и организаций, формирования публичной отчетности по результатам их обработки.
 - Публикация открытых данных для населения и организаций и публичный доступ к ним:
 - для информационных систем – подключение к сервису данных SQL Azure и выполнение запросов на чтение как к обычной реляционной базе данных;
 - для пользователей – разработка и предоставление доступа к информационным и аналитическим отчетам.
8. **Microsoft BizTalk** – программная платформа для интеграции приложений, обеспечения синхронизации данных с минимальными задержками и автоматизации сквозных деловых процессов, проходящих через различные подразделения и выполняющихся в разных компьютерных программах.

Разработка решений по бизнес-аналитике.

Классический (базовый) проект по бизнес-аналитике включает в себя следующие этапы (описывается логическая, а не временная последовательность работ):

1. Выбор и подключение к источникам данных: базы данных приложений, файлы и сервисы.
2. Анализ данных, оценка их качества и актуальности, выявление пересечений между разными источниками и приоритизация источников данных по качеству и актуальности.
3. Извлечение данных из источников и их сохранение во вспомогательной базе данных (Stage DB).
4. Очистка и консолидация данных во вспомогательной базе данных, загрузка в единое хранилище данных (Data Warehouse) или витрину данных (Data Mart).
5. Разработка сценариев анализа данных разными группами пользователей для поддержки принятия решений.
6. Построение аналитических моделей данных (PowerPivot или OLAP, Data Mining) и создание аналитических хранилищ данных.
7. Публикация данных для партнеров и публичных данных по деятельности компании.
8. Разработка решений по визуализации аналитических данных с помощью Excel, PowerView, PerformancePoint или Reporting Services.
9. Обучение пользователей по работе с представлениями аналитических данных, сбор пожеланий и предложений, дальнейшее развитие системы анализа данных.

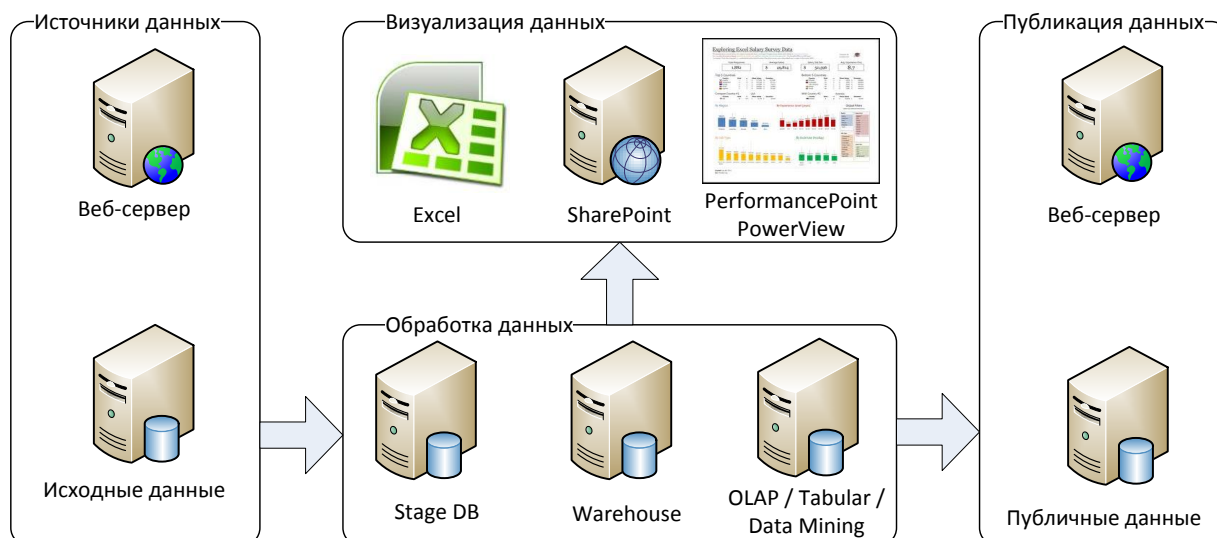


Рисунок 2. Технологические этапы обработки данных

Самая сложная и ответственная задача это обеспечение качества, полноты и актуальности данных. По этой задаче серьезная нагрузка ложится на аналитиков данных и разработчиков

интеграционных решений. Трудоемкость составляет 50-80% объемов работа по всему проекту.

Типичной ошибкой является недооценка важности, сложности и трудоемкости этой задачи, поспешность с построением конечных решений по визуализации данных. Это приводит к потере доверия пользователей к данным из-за их недостоверности и не актуальности, резкому повышению трудоемкости работ по визуализации аналитических данных, повторному выполнению задач и увеличению сроков и бюджета проекта в несколько раз.

Визуализация данных может строиться с использованием отдельных продуктов, либо как гибридное решение в зависимости от опыта команды разработчиков. Если имеются разработчики с опытом работы с SSRS, то можно построить полноценное решение на основе сервисов отчетности. Если есть опыт работы с PerformancePoint, то данный инструмент имеет собственные представления, и позволяет создавать гибридные визуальные решения, используя Excel или SSRS.

Если нет разработчиков с опытом работы с SSRS или PerformancePoint, то с помощью Excel достаточно просто создавать цифровые панели и аналитические отчеты и публиковать их для других пользователей с помощью SharePoint.

Более подробно этапы проекта по бизнес-аналитике описаны в следующей таблице.

Таблица 1. Этапы обработки данных

Этапы обработки данных	Технические решения
1. Сбор и консолидация данных	
1.1. Сбор данных от источников	<ul style="list-style-type: none"> • Веб-сервисы • Пакеты интеграции данных (SSIS) • Вспомогательная база данных (Stage DB)
1.2. Очистка данных: проверка значений атрибутов	<ul style="list-style-type: none"> • Вспомогательная база данных (Stage DB) • Сервис качества данных (DQS) • Пакеты интеграции данных (SSIS)
1.3. Формирование мастер-данных: стандартизация значений атрибутов и удаление дублей данных	<ul style="list-style-type: none"> • Вспомогательная база данных (Stage DB) • Сервис мастер-данных (MDS) • Пакеты интеграции данных (SSIS)
1.4. Синхронизация данных с источниками	<ul style="list-style-type: none"> • Веб-сервисы • Пакеты интеграции данных (SSIS)
2. Анализ данных	
2.1. Формирование хранилища данных: перенос данных из вспомогательной БД в хранилище данных	<ul style="list-style-type: none"> • Вспомогательная база данных (Stage DB) • Хранилище данных (Data Warehouse) • Пакеты интеграции данных (SSIS)

Этапы обработки данных	Технические решения
2.2. Формирование многомерной базы данных (OLAP-кубы) и табличных моделей данных	<ul style="list-style-type: none"> • Хранилище данных (Data Warehouse) • Многомерная база данных (SSAS) • Пакеты интеграции данных (SSIS)
2.3. Извлечение знаний и прогнозирование (Data Mining)	<ul style="list-style-type: none"> • Хранилище данных (Data Warehouse) • Многомерная база данных (SSAS OLAP) • Интеллектуальный анализ данных (SSAS Data Mining) • Пакеты интеграции данных (SSIS)
3. Визуализация данных	
3.1. Формирование стратегических карт и цифровых панелей для основных показателей эффективности	<ul style="list-style-type: none"> • Многомерная база данных (SSAS) • Microsoft Visio • Microsoft SharePoint • Microsoft Performance Point • Microsoft PowerView • Отчеты (SSRS)
3.2. Формирование витрин данных по направлениям деятельности и для отдельных подразделений	<ul style="list-style-type: none"> • Хранилище данных (Data Warehouse) • Многомерная база данных (SSAS) • Microsoft SharePoint • Microsoft Performance Point • Microsoft PowerView • Отчеты (SSRS) • Microsoft Excel
3.3. Обеспечение разработки персональных аналитических представлений и их совместного использования	<ul style="list-style-type: none"> • Многомерная база данных, табличные данные (SSAS) • Microsoft SharePoint • Microsoft PowerView • Отчеты (SSRS) • Microsoft Excel
4. Публикация данных	
4.1. Подготовка данных к публикации	<ul style="list-style-type: none"> • Хранилище данных (Data Warehouse) • Многомерная база данных (SSAS) • Пакеты интеграции данных (SSIS)
4.2. Размещение данных на публичном веб-сервере	<ul style="list-style-type: none"> • Пакеты интеграции данных (SSIS) • Веб-сервисы • Отчеты (SSRS)

Современный (расширенный) вариант проекта по бизнес-аналитике включает в себя также этапы и задачи по созданию мастер-данных (централизованные справочники) и интеграции приложений. Сложность данного подхода заключается в необходимости доработки унаследованных приложений.

При интеграции приложений решаются задачи на двух уровнях:

1. **Поток данных** (Data flow) – запрос, чтение данных и сохранение данных, преобразования данных в форматы разных приложений и в канонический формат единого хранилища данных.
2. **Поток работ** (Work flow) – обеспечение взаимодействия сотрудников, работающих в разных приложениях для выполнения общих задач в виде деловых процессов.

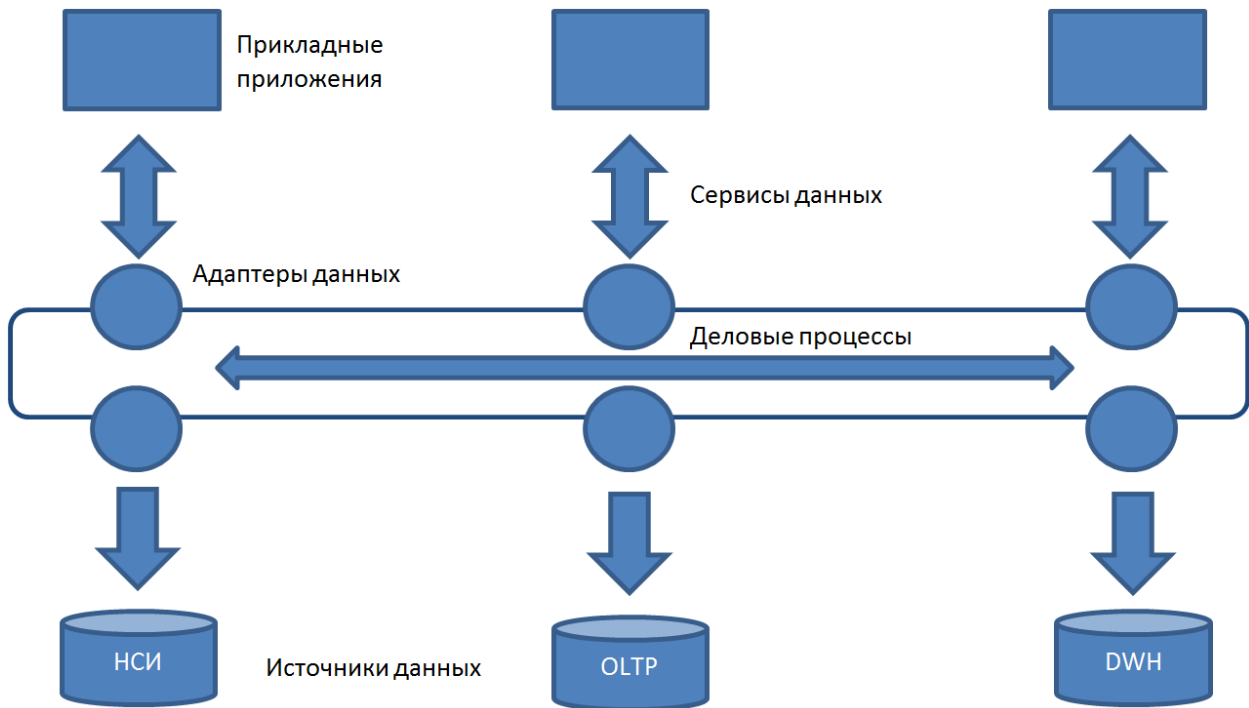


Рисунок 3. Концептуальная архитектура интеграции приложений

Технологической платформой для интеграции приложений являются Microsoft SharePoint и Microsoft BizTalk.