



ОНЛАЙН-ОБРАЗОВАНИЕ


Онлайн-образование

Не забыть включить запись!





Меня хорошо видно && слышно?

Ставьте  , если все хорошо
Напишите в чат, если есть проблемы

Применение SSIS в ETL

курс “SQL Server Developer”



Коробков Виктор

Правила вебинара



Активно участвуем



Задаем вопрос в чат или голосом



Off-topic обсуждаем в Slack



Вопросы вижу в чате, могу ответить не сразу

Цели вебинара | После занятия вы сможете

1 Создавать ETL пакеты

2 Обработать данные в SSIS среде

3 Применять дасрас файлы

Смысл | зачем вам это уметь

Для анализа данных, подготовки и переноса данных из OLTP среды в OLAP

Маршрут вебинара

1 ETL
пакеты

2 SSIS
среда

3 Дасрас - файлы

4 КХД

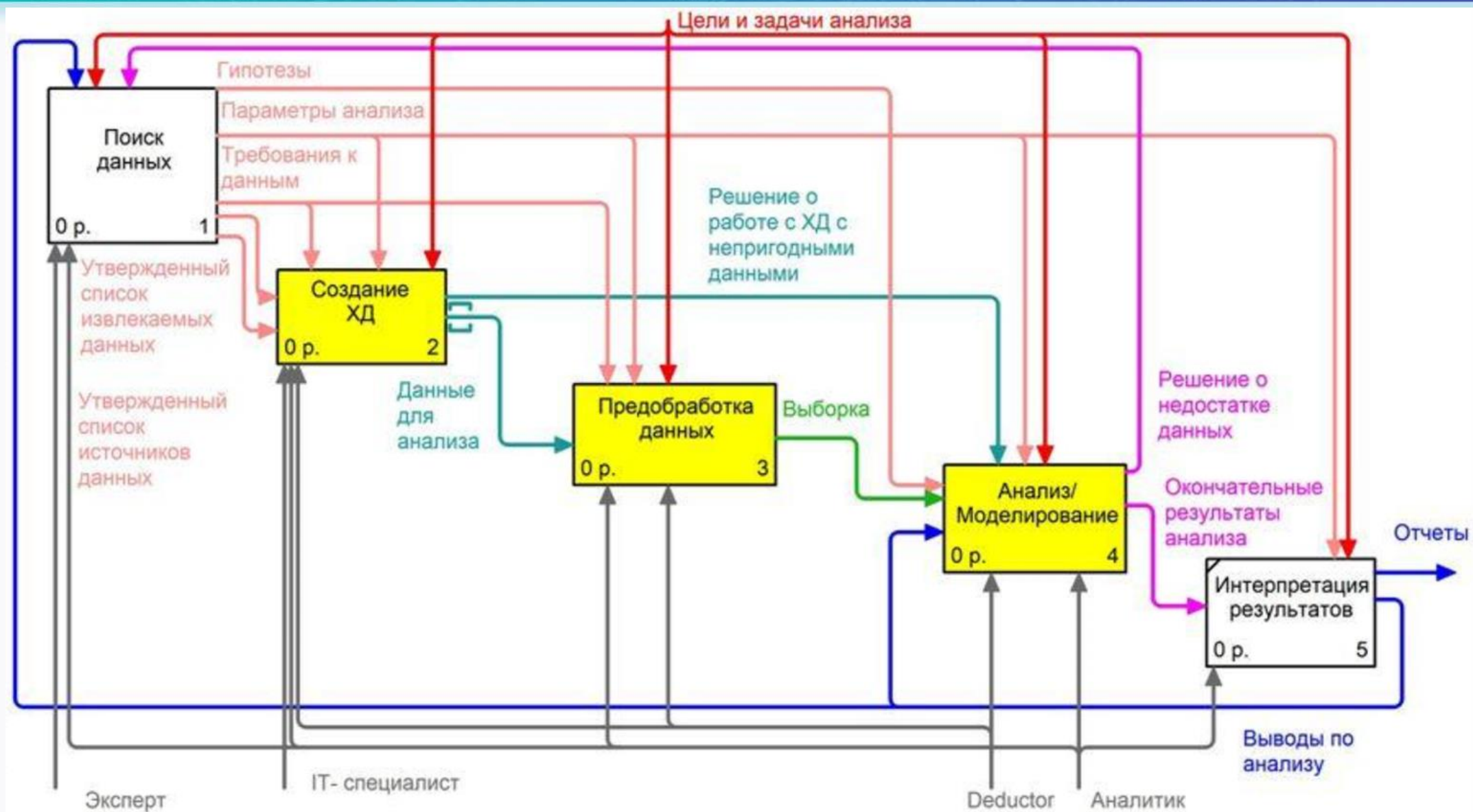
Анализ данных

Анализ данных - методы изучения процессов и оценки результатов.

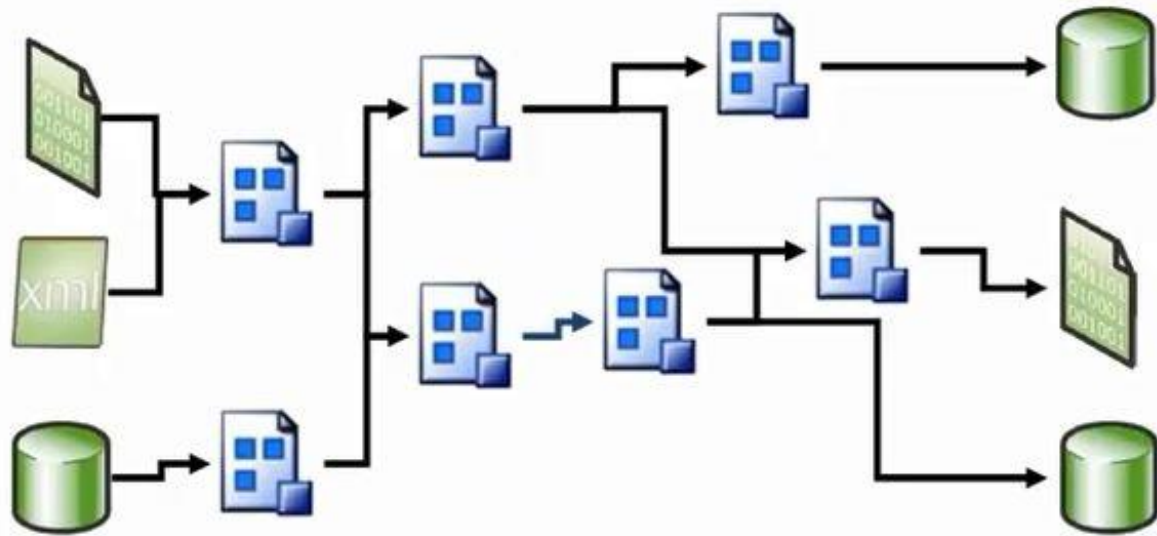
Роль анализа - подготовка информации для планирования и прогнозирования результатов деятельности.

Анализ данных должен быть постоянным и цикличным.

Анализ данных



ETL пакеты



Sources (Extract)

SQL Server, Flat Files, Excel, etc..

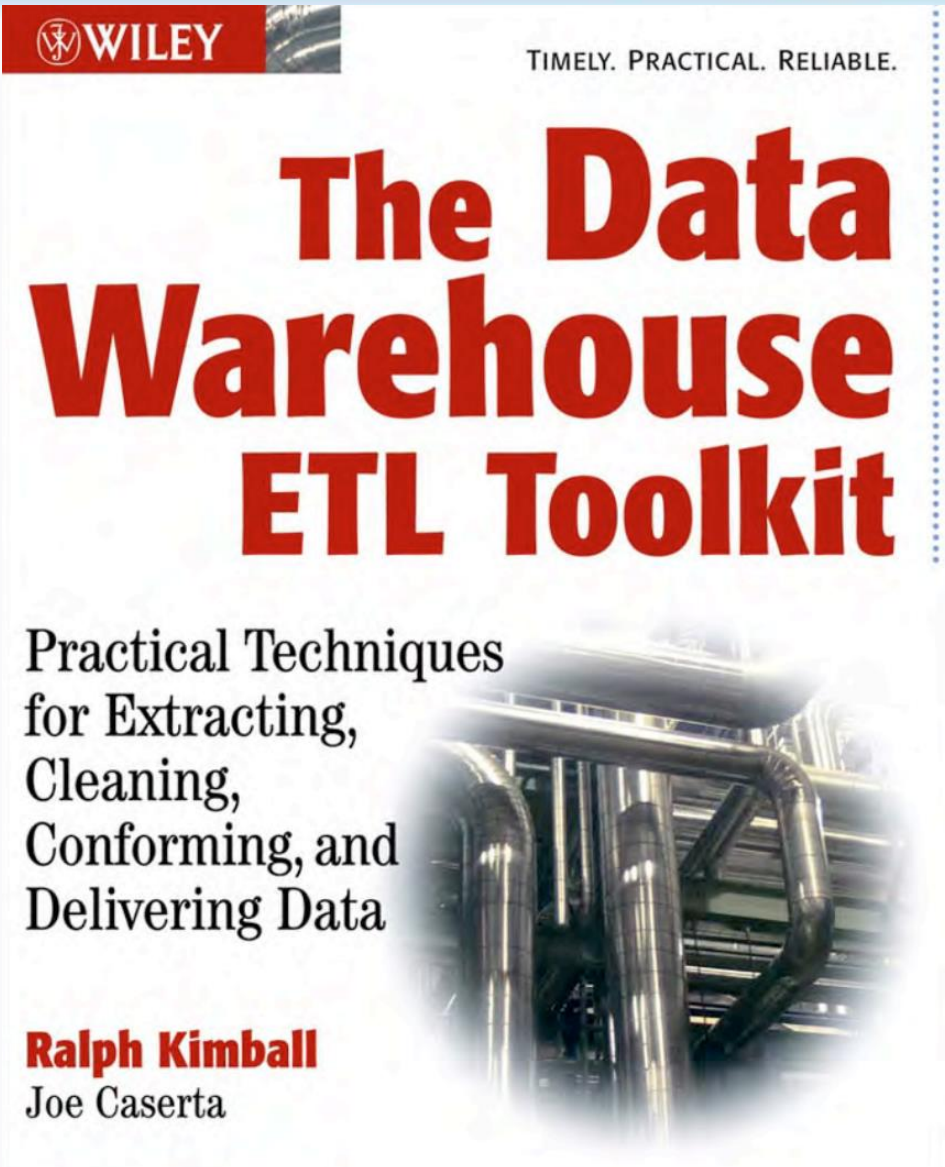
Transformations (Transform)

Data Conversion, Aggregate, etc...

Destinations (Load)

SQL Server, Flat Files, Excel, etc..

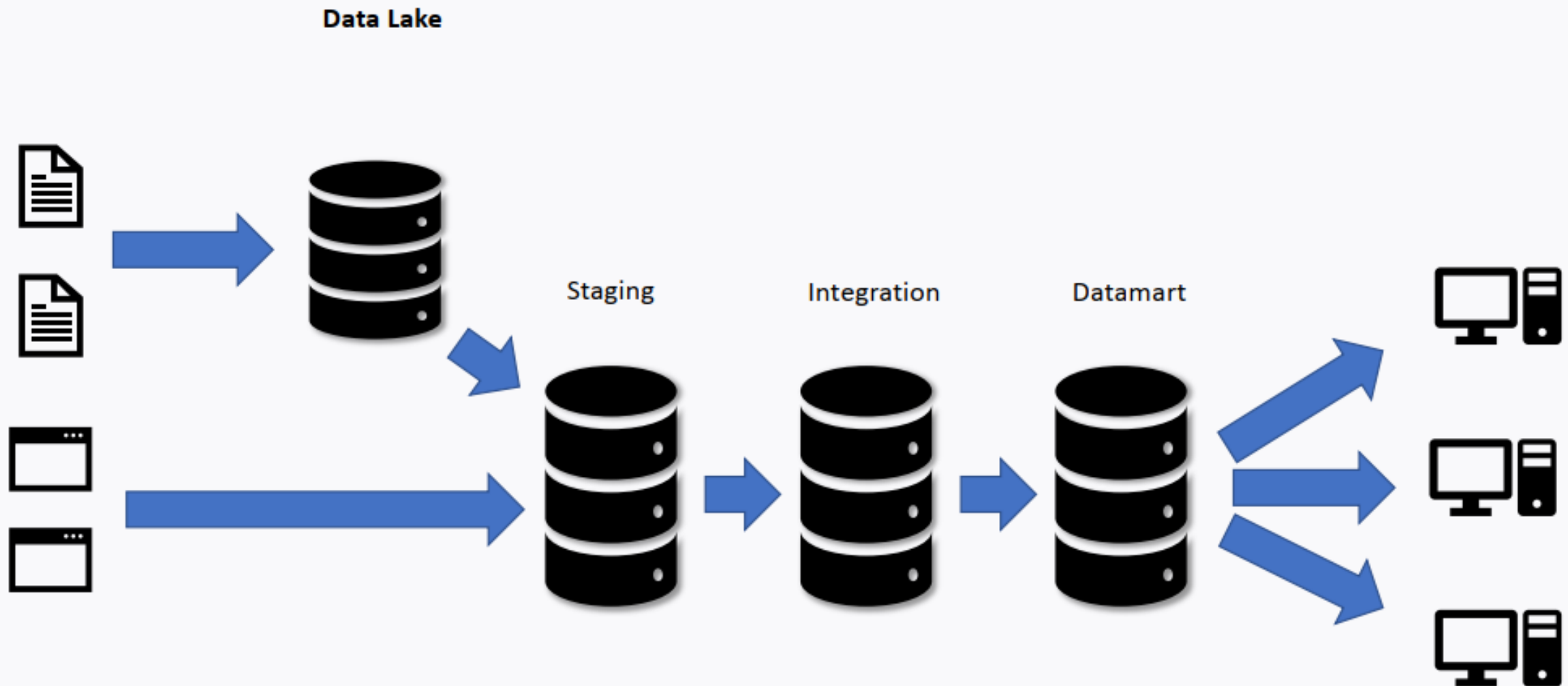
ETL пакеты



Операции, которые **ОБЯЗАТЕЛЬНО** должен совершать ETL процесс:

1. Загрузка данных в максимально удобном виде для аналитических приложений.
2. В процессе загрузки данные могут обогащаться дополнительной информацией.
3. Фиксировать и документировать происхождение данных.

Логическая модель

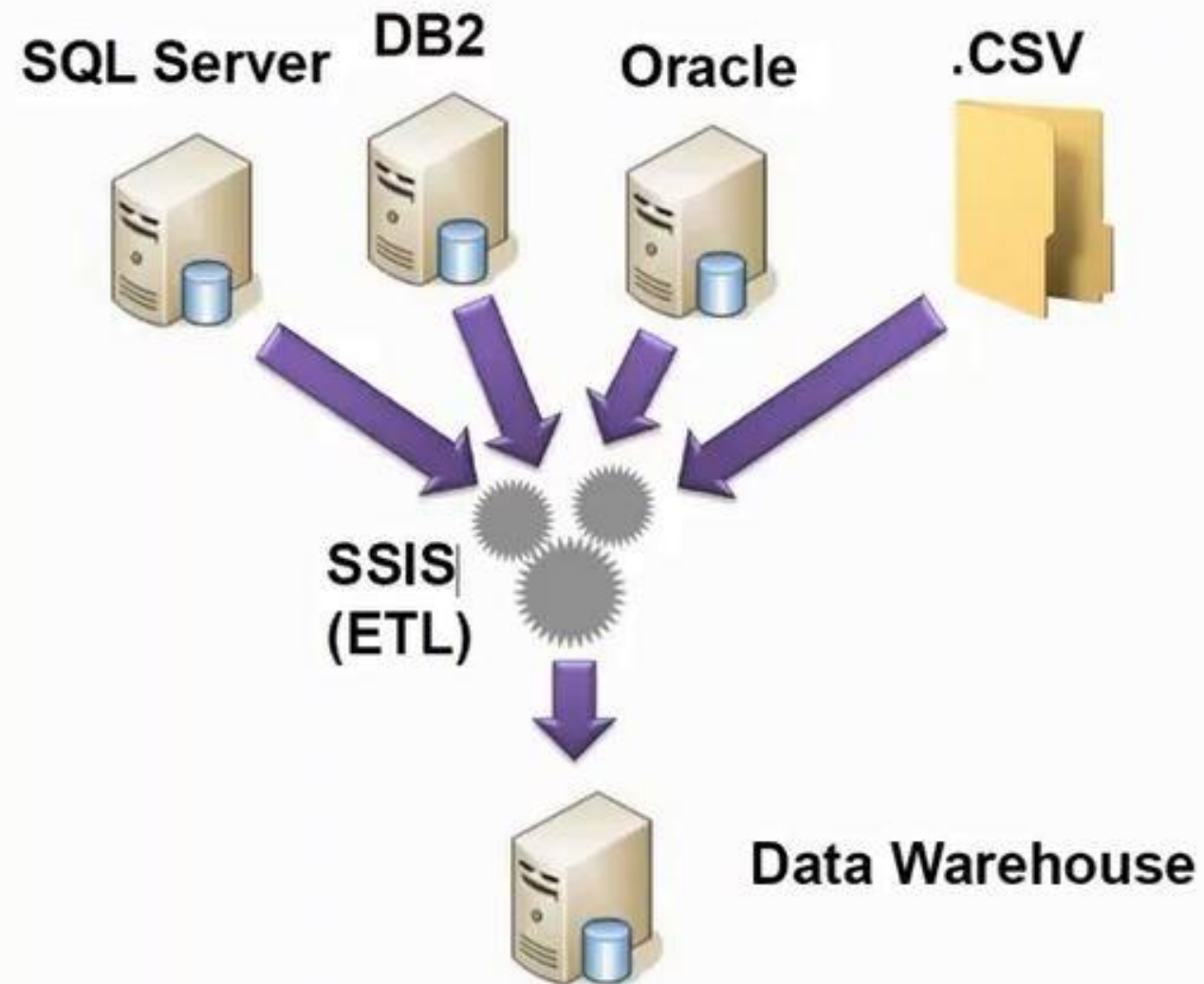


SSDT - SQL Server Data Tools

SQL Server Data Tools - средство разработки, позволяющее создавать реляционные базы данных SQL Server, базы данных SQL Azure, модели данных Analysis Services (AS), пакеты Integration Services (IS) и отчеты Reporting Services (RS).

Ссылка: <https://docs.microsoft.com/ru-ru/sql/ssdt/download-sql-server-data-tools-ssdt?view=sql-server-ver15>

SSIS B ETL

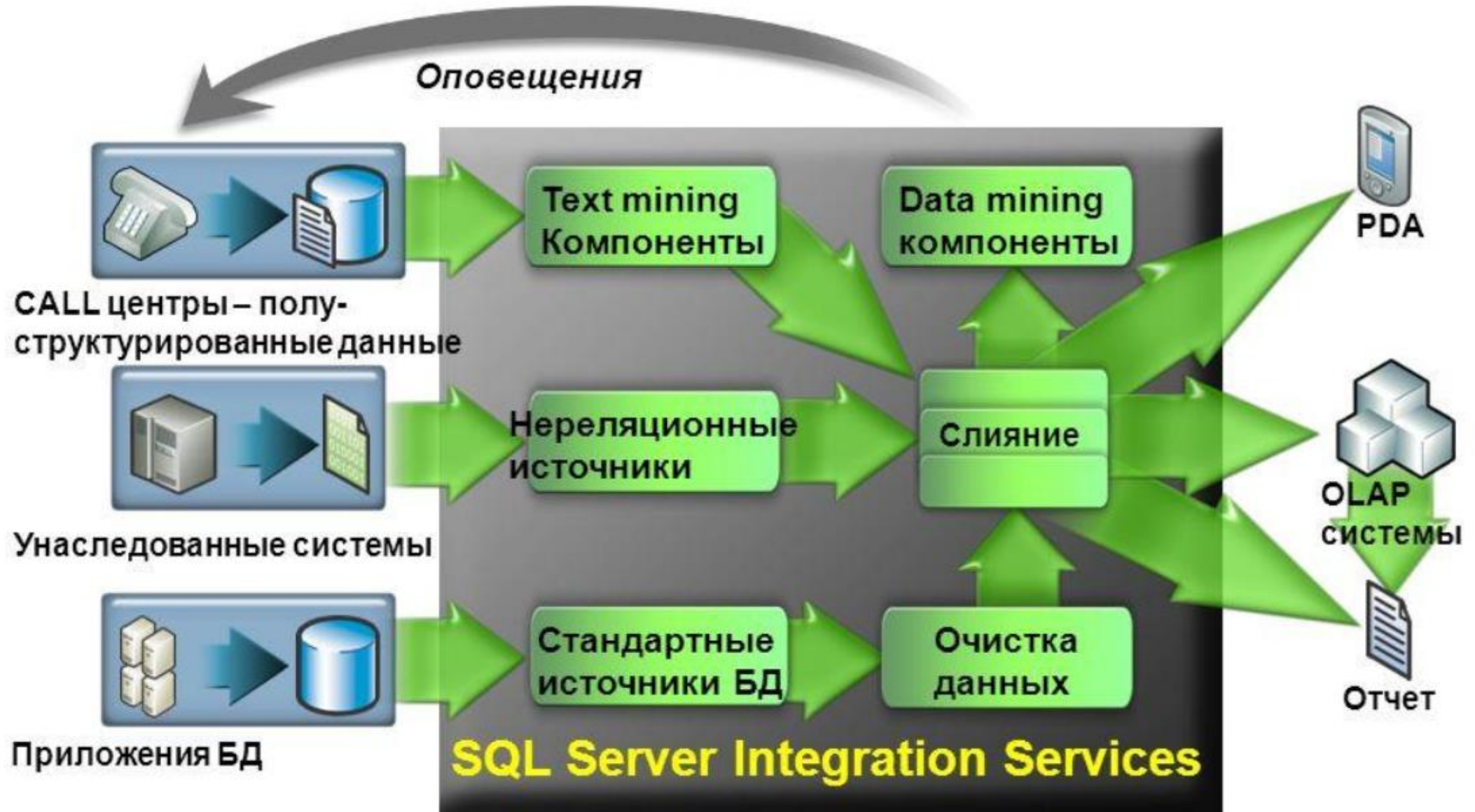


SSIS - SQL Server Integration Services

Службы *Integration Services* - платформа для построения высокопроизводительных решений интеграции данных и решений потока операций, включая операции извлечения, преобразования и загрузки (ETL) для хранилищ данных.

- Графические инструменты
- Мастера для построения и отладки пакетов
- Источники данных для извлечения данных
- Источники назначения для загрузки данных
- Преобразования для очистки, статистической обработки, слияния и копирования данных
- Задачи для выполнения функций потока операций
- Служба управления и администрирования пакетов
- API-интерфейсы для программирования объектной модели

SSIS - SQL Server Integration Services



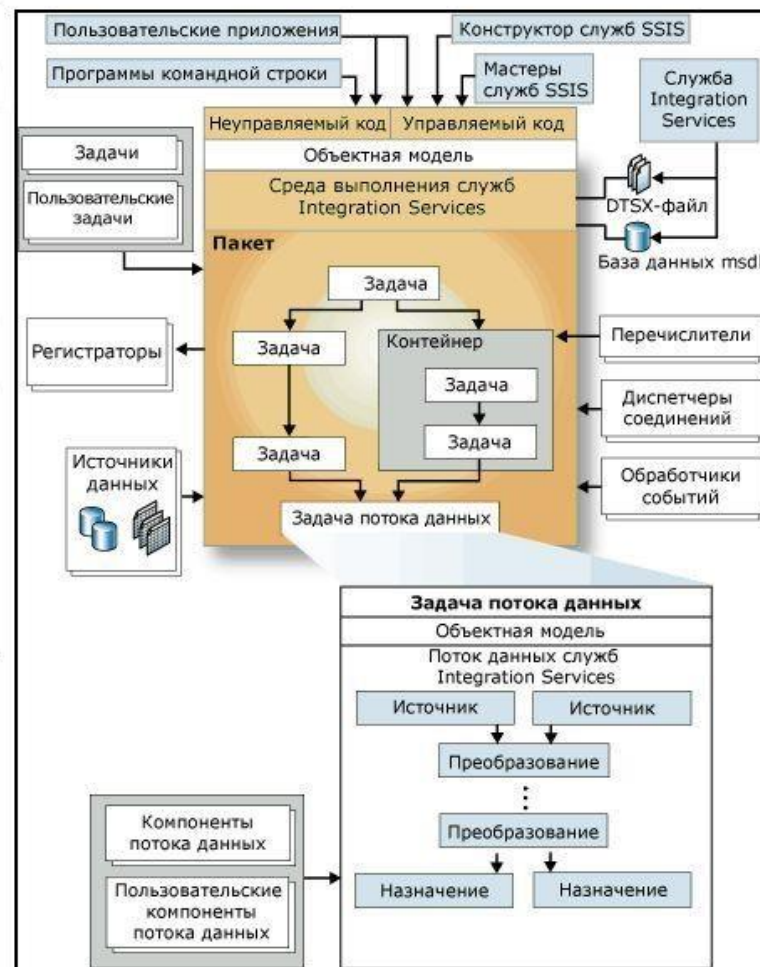
Архитектура SSIS

Термины

- Источник (и) - Sources
 - Приёмник(и) - Destinations
 - Преобразование данных - (Transformation)
- Время исполнения
- Пакет (Package)
 - Задача (Task)
 - Буфер (Buffer)
 - Труба (pipeline) потока данных

Конструктор служб SSIS

- Поток управления (Control Flow)
- Поток данных (Data Flow)
- Обработчики событий в пакете и объектов пакета (Event Handlers)
- Просмотр содержимого пакета
- Просмотр выполнения пакета



Функционал SSIS

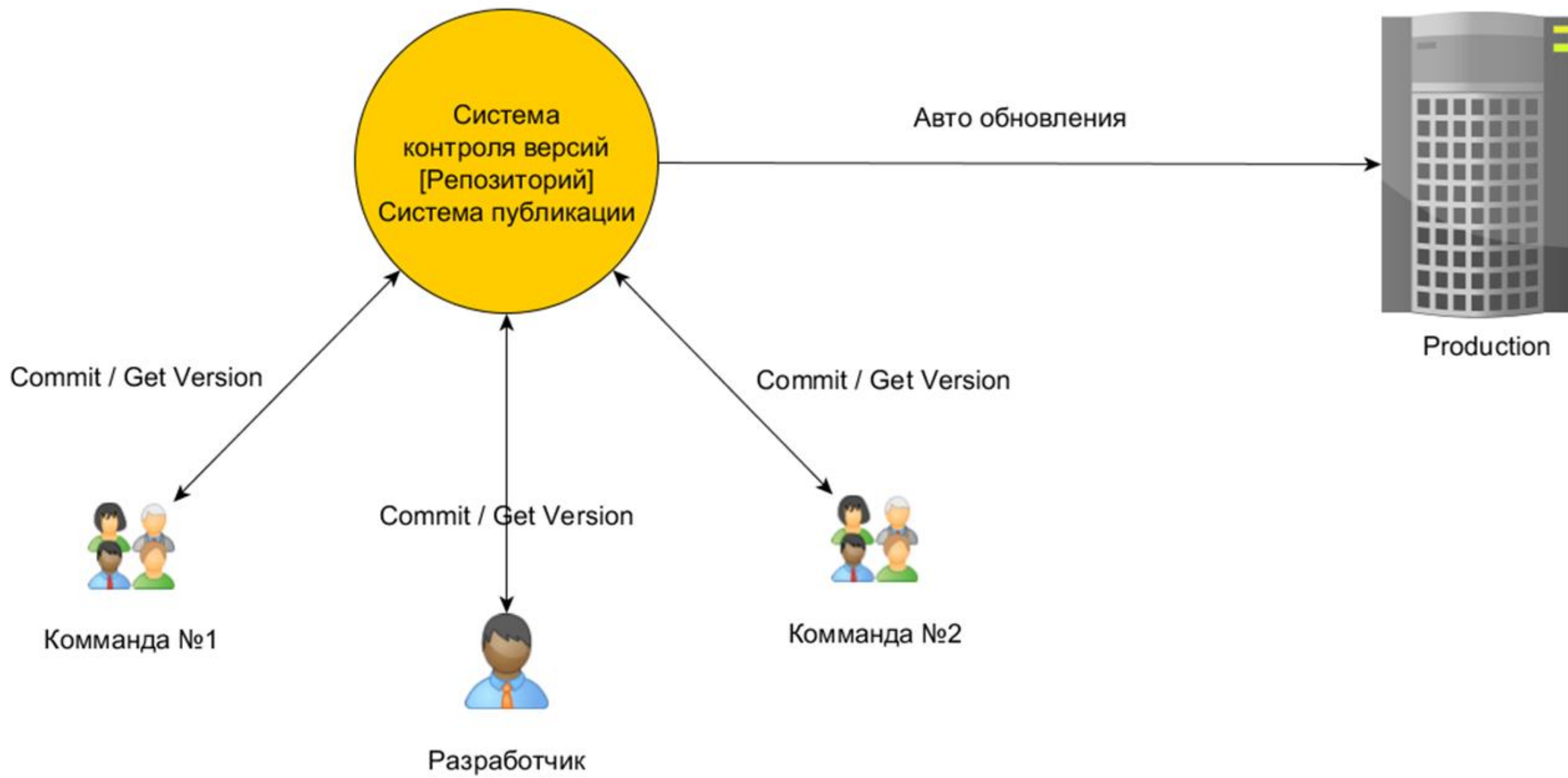
1. Параметры
2. Циклы
3. Переменные
4. Создание скриптов на Visual Basic или C#
5. Транзакции
6. Точки сохранения
7. Отслеживание и перехват ошибок
8. Документирование

Дас приложения

DACRAS файлы

Приложение уровня данных (DAC) - это автономная единица развертывания базы данных SQL Server, которая позволяет разработчикам и администраторам группировать объекты (таблицы, представления, функции ...) SQL Server в переносимый артефакт, называемый пакетом **приложения уровня данных** или файлом DACRAS.

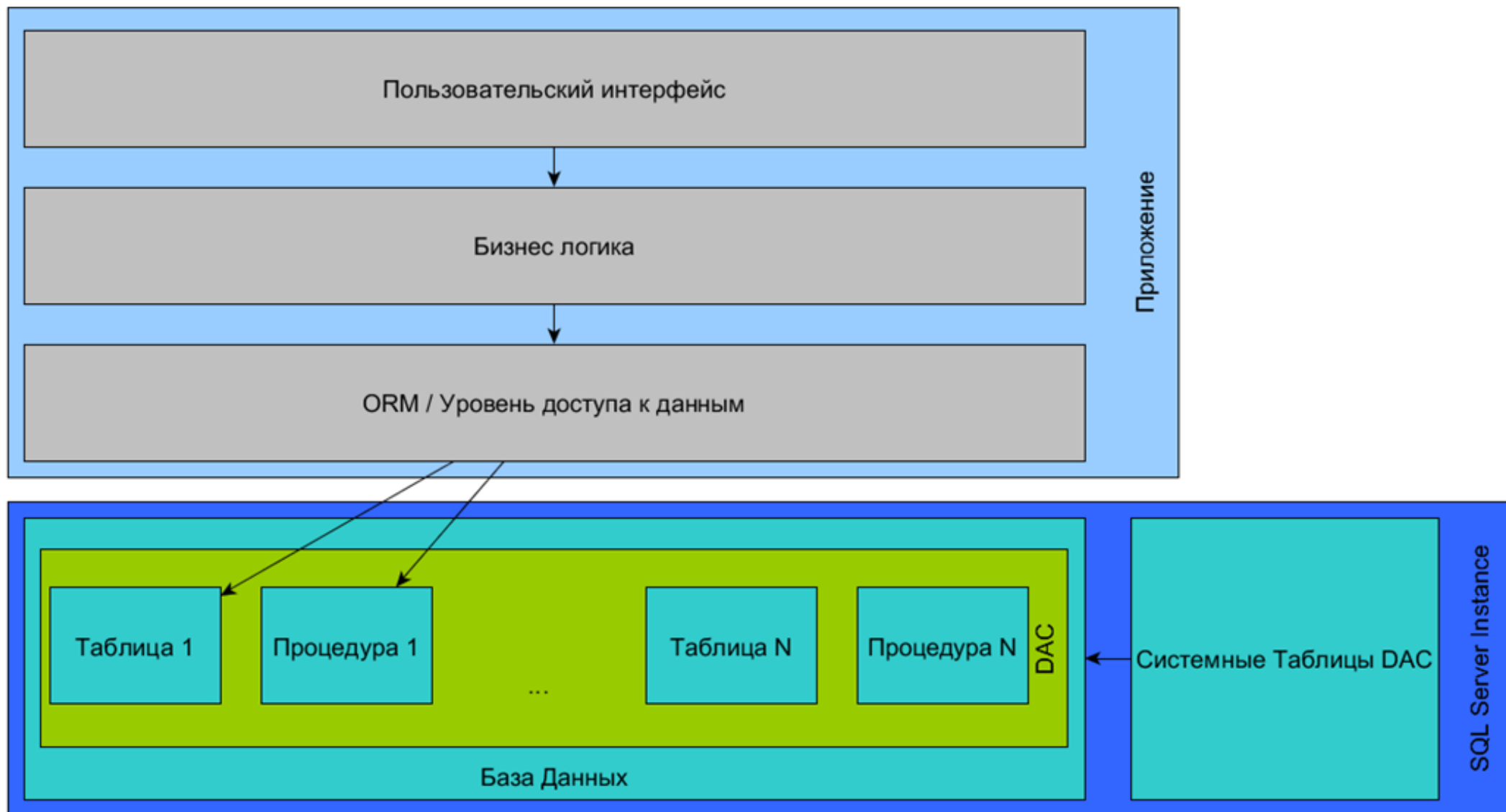
ДАСРАС файлы



ДАСРАС файлы

	Система контроля версий (код, схема)	Публикация
Программа	✓	✓
SQL Server База данных	✗	✗

ДАСРАС файлы



DACRAS файлы

	Система контроля версий	Публикация
Программа	✓	✓
SQL Server + DAC	✓	✓

Возможности DAC

1. Register
2. Delete
3. Driff Report
4. Extract
5. Publish
6. Compare Report
7. Update Script
8. Export
9. Import

Последовательность действий при публикации (Publish)

1. Выявление потерь
2. Сравнение схем исходного DAC пакета со схемой обновляемой базы данных
3. Определение зависимостей
4. Создание скрипта обновления с корректным порядком создания/обновления/удаления объектов
5. Выполнение скрипта перед развертыванием (PreDeploy)
6. Выполнение скрипта обновления
7. Выполнение скрипта после развертывания (PostDeploy)
8. Регистрация

Корпоративные хранилища данных

Inmon методология

Подход **Билла Инмона** основывается на том, что хранилище данных является централизованным хранилищем всех корпоративных данных.

При таком подходе организация сначала создает **нормализованную модель** хранилища данных.

Затем создаются витрины размерных данных на основе модели хранилища.

Kimball методология

Подход **Ральфа Кимбалла** основывается на важности **витрин данных**, которые являются хранилищами данных, принадлежащих конкретным направлениям бизнеса.

Хранилище данных — это просто **сочетание различных витрин данных**, которые облегчают отчетность и анализ.

Kimball — размерная (**ненормализованная**) модель данных, объединяющая таблицы фактов, схемы «звезда» и «снежинка».

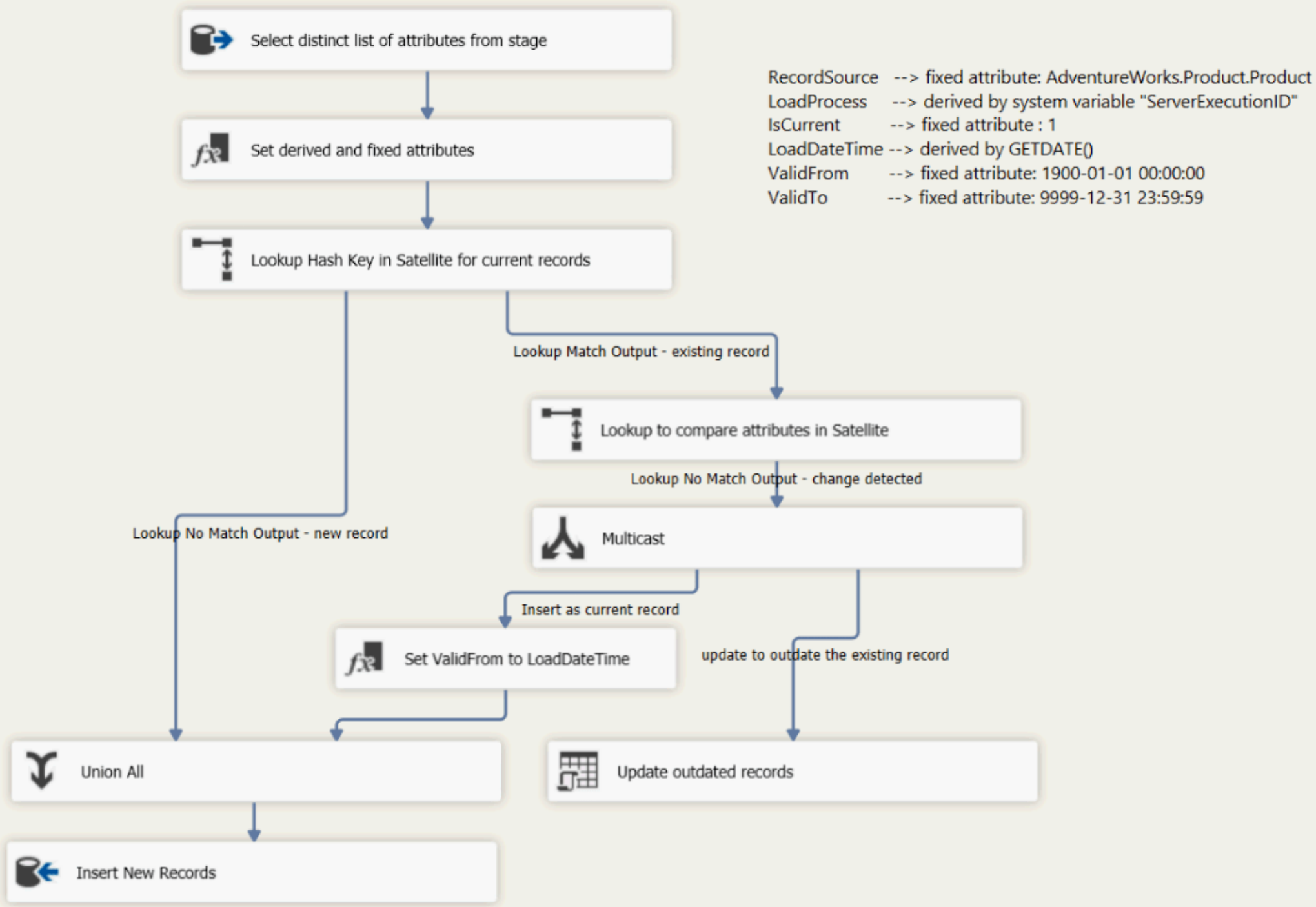
Методология была анонсирована в 1990 году.

Data Vault

Data Vault — гибридный подход, объединивший достоинства схемы «звезда» и 3-ей нормальной формы.

Впервые эта методология была анонсирована в 2000 году Дэном Линстедтом (Dan Linstedt).

В 2013 году, Дэн анонсировал версию 2.0, доработанную с учетом технологий (NoSQL, Hadoop) и новых требований, выставляемых к DWH.



Рефлексия



Отметьте 3 пункта, которые вам запомнились с вебинара



Что вы будете применять в работе из сегодняшнего вебинара?

The background of the entire image is an aerial photograph of a city with many skyscrapers, likely New York City. The image is overlaid with a semi-transparent blue layer. In the center of this blue layer, there is a white network pattern consisting of dots connected by thin lines. The text is written in white, bold, sans-serif font across the middle of the image.

Заполните, пожалуйста,
опрос о занятии по ссылке в чате

Спасибо за внимание!
Приходите на следующие вебинары



Коробков Виктор