

INFOSPHERE DATASTAGE

IBM InfoSphere DataStage - это инструмент ETL, входящий в пакет решений IBM Information Platforms и IBM InfoSphere. Он использует графическую нотацию для построения решений по интеграции данных и доступен в различных версиях, таких как Server Edition, Enterprise Edition и MVS Edition. Он использует архитектуру типа клиент-сервер. Серверы могут быть развернуты как в Unix, так и в Windows.

Это мощный инструмент интеграции данных, часто используемый в Data Warehousing проектах для подготовки данных к генерации отчетов.

IBM InfoSphere DataStage позволяет интегрировать большие объемы данных между многочисленными источниками данных и целевыми приложениями. IBM InfoSphere DataStage представляет собой главный модуль платформы интеграции данных InfoSphere Information Server, которая поддерживает универсальный доступ к данным локальной сети, облачной инфраструктуре, а также к структурированным и неструктурированным данным.

DataStage job заполняет одну или несколько таблиц в целевой базе данных. Количество DataStage Job, которые вы можете создать в проекте DataStage, не ограничено. DataStage jobs определяются с помощью DataStage Designer, но вы можете просматривать и редактировать некоторые свойства DataStage Job с помощью DataStage Manager.

Существует три основных типа DataStage jobs:

Server jobs: Они компилируются и запускаются на сервере DataStage. Server job будет подключаться к базам данных на других компьютерах по мере необходимости, извлекать данные, обрабатывать их, а затем записывать дан-

ные в целевое хранилище данных (target data warehouse).

Parallel jobs: компилируются и выполняются на сервере DataStage UNIX и могут выполняться параллельно на SMP, MPP и кластерных системах. Доступны только в Enterprise Edition.

Mainframe jobs: компилируется и запускается на мэйнфрейме, затем извлеченные данные загружаются в хранилище данных (data warehouse). Доступны только в Enterprise MVS Edition.

DataStage Jobs состоят из отдельных stages. Каждый Stage описывает источник данных, конкретный процесс или data mart. Например, один Stage может извлекать данные из источника данных, а другой преобразует их. Stages добавляются в работу и связываются вместе с помощью DataStage Designer.

Есть два типа Stages:

Built-in stages. Поставляется с DataStage. Используются для извлечения, агрегирования, преобразования или записи данных.

Plug-in stages. Дополнительные Stages, отдельно устанавливаются для выполнения специализированных задач, которые встроенные этапы не поддерживают.

InfoSphere DataStage – это отличный инструмент интеграции данных, позволяющий преобразовывать данные и перемещать их между операционными, транзакционными и аналитическими системами.

Список литературы

1. InfoSphere DataStage [Электронный ресурс]. - Режим доступа: <https://www.ibm.com/developerworks/ru/library/dm-1407governrest/index.html>. - Дата доступа: 14. 03. 2019.

Бардиер Вадим Анатольевич, студент 3 курса факультета информационных технологий и управления Белорусского государственного университета информатики и радиоэлектроники, bardiervadim97@gmail.ru.

Научный руководитель: Трофимович Алексей Федорович, старший преподаватель кафедры информационных технологий автоматизированных систем trofimovich_a_f@tut.by