

УДК 004.6

ПРОБЛЕМЫ ПРЕОБРАЗОВАНИЯ И ИНТЕГРАЦИИ ДАННЫХ ВО ВРЕМЯ ETL-ПРОЦЕССА

PROBLEMS OF DATA TRANSFORMATION AND INTEGRATION DURING THE ETL PROCESS

Горегляд Виктория Витальевна

студентка магистратуры

БГУИР, г. Минск, Республика Беларусь, viktoriya.goreglyad@gmail.com

Аннотация. Интеграция разрозненных информационных систем предприятия на сегодняшний день является наиболее рациональным способом построения его единой информационной среды. Поэтому очень важно знать основные проблемы, возникающие в процессе построения хранилища данных. В статье выполнен анализ текущей ситуации в области информационных технологий и проведен обзор актуальных проблем при интеграции и преобразовании данных в хранилище.

Abstract. Integration uncoordinated enterprise information systems is by far the most efficient way to build a unified information environment in the enterprise. Therefore, it is very important to know the main problems that arise in the process of building a data warehouse. The article analyzes the current situation in the field of information technologies and provides an overview of current problems in the integration and transformation of data into the storage.

Ключевые слова: хранилище данных, ETL, интеграция, проблемы, источники данных.

Keywords: data warehouse, ETL, integration, problems, data sources.

К настоящему времени во многих организациях накоплены колоссальные объемы данных, на основе которых можно решать самые разнообразные аналитические и управленческие задачи в любой сфере деятельности. Такие предприятия обладают несколькими базами данных, данные могут иметь разные представления, а иногда быть даже несогласованными. Более того, для оперативной аналитической обработки требуется привлечение внешних источников, которые тем более могут обладать разными форматами и требовать согласования. Именно поэтому возникает задача интеграции данных в единую информационную систему.

Хранилище данных (ХД) – это специальным образом организованный массив данных предприятия, обрабатываемый и хранящийся в едином аппаратно-программном комплексе, который обеспечивает быстрый доступ к оперативной и исторической информации, многомерный анализ, получение прогнозов и статистики[1].

Для интеграции и преобразования данных из различных источников в хранилища используются ETL-процессы. При использовании ETL исходные данные извлекаются, преобразовываются в соответствии с целевым форматом и загружаются в целевое ХД.

Следует понимать, что при интеграции информационных систем производится интеграция именно данных, и только потом техническая реализация способа передачи. В связи с этим, основной проблемой, возникающей при интеграции, является проблема, связанная с качеством данных. Поскольку системы обычно приобретаются или разрабатываются в разное время и для различных подразделений, привнесенное несоответствие еще больше осложняет интеграцию.

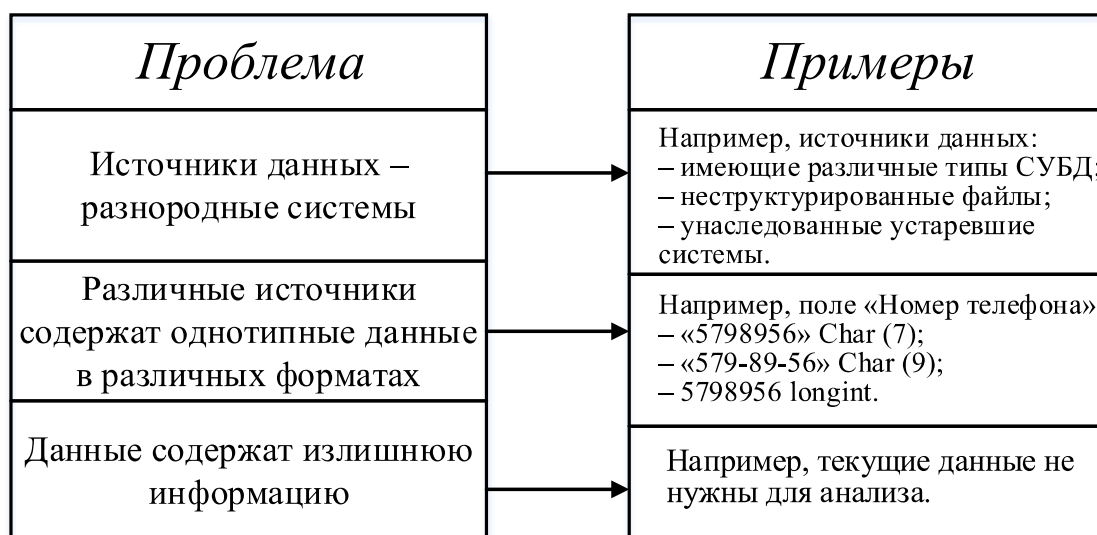


Рисунок 1 –Типовых проблем во время процесса интеграции

Неудачно проведенная интеграция становится источником ошибочной информации, которая попадает в отчеты и на основании которой в конечном итоге предпринимаются неверные действия[2]. На рисунке 1 представлены примеры типовых проблем во время процесса интеграции.

Для качественного процесса преобразования данных в рамках ETL необходимо выполнять следующие операции:

1. Преобразование структуры данных. Так как данные интегрируются из различных источников, то они могут отличаться своей структурной организацией (например, форматами, типами, кодировкой и так далее).

2. Агрегирование данных. Общим свойством всех этих источников является то, что они содержат данные с максимальной степенью детализации. Однако, для достоверного описания предметной области использование данных с максимальным уровнем детализации не всегда целесообразно. В результате агрегирования большое количество записей о каждом событии заменяется относительно небольшим количеством записей.

3. Перевод значений. Данные часто хранятся с использованием специальных кодировок, которые позволяют сократить избыточность данных и тем самым уменьшить объем памяти, требуемой для их хранения, то есть привести их в сокращенный вид. В этом случае перед загрузкой данных в БД требуется выполнить перевод таких сокращенных значений в более полные и понятные.

4. Создание новых данных. В процессе загрузки в БД может понадобиться вычисление некоторых новых данных на основе существующих. Создание новой информации на основе имеющихся данных тесно связано с процессом обогащения данных, который может производиться на этапе преобразования данных в ETL.

5. Очистка данных. Сбор данных производится из огромного числа источников и некоторые из них не содержат автоматических средств поддержки целостности, непротиворечивости и корректного представления данных. При переносе информации в БД можно столкнуться с потоками «грязных» данных, которые могут стать причиной неправильных результатов анализа. Поэтому в процессе ETL применяется очистка – процедура корректировки данных, которые не удовлетворяют определенным критериям качества (содержат пропуски, неправильные форматы, дубликаты и т. д.)[4].

Список литературы

1. Батьков, В.О. Анализ проблем современных хранилищ данных / В.О. Батьков // Труды Международного симпозиума «Надежность и качество». –2013. – Том 1.– С. 259 – 261.

2. Нейл Рейден. Интеграция данных – это моделирование их смысла //Intelligententerprise. – №22 (131). – 2005.

3. Талгатова З.Т. Анализ существующих моделей процессов ETL для хранения данных // Технические науки - от теории к практике: сб. ст. по матер. LIV междунар. науч.-практ. конф. № 1(49). – Новосибирск: СибАК, 2016.

4. Консолидация данных – ключевые понятия // Корпоративный менеджмент. <http://www.cfin.ru/itm/olap/cons.shtml>

УДК 517.9

РАЗРЕШИМОСТЬ УРАВНЕНИЙ ДИФФУЗИИ С ДРОБНЫМИ ПРОИЗВОДНЫМИ
SOLVABILITY OF DIFFUSION EQUATIONS WITH FRACTAL DERIVATIVES

Горлов Владимир Александрович

кандидат физико-математических наук, доцент 122 кафедры Средств связи (и авиационных комплексов связи) ВУНЦ ВВС «Военно-воздушная академия им. проф. Н.Е. Жуковского и Ю.А. Гагарина», г. Воронеж, Россия, gorlov_v_a@mail.ru

Макарова Алла Викторовна

кандидат физико-математических наук, преподаватель 206 кафедры математики ВУНЦ ВВС «Военно-воздушная академия им. проф. Н.Е. Жуковского и Ю.А. Гагарина», г. Воронеж, Россия, allagm@mail.ru

Аннотация. В статье изучается разрешимость уравнений диффузии с дробными производными. В частности, рассматривается уравнение диффузии с позиции применения дробных производных для его исследования.

Abstract. We investigate the solvability of diffusion equations with fractal derivatives. In particular, it's considered diffusion equation with the approach of fractal derivatives for its investigation.

Ключевые слова: дробная производная, нестационарные задачи, нелинейная диффузия.

Keywords: fractal derivative, non-stationary problems, nonlinear diffusion.

Многочисленные явления диффузии и тепломассопереноса существенно различаются с физической точки зрения, но могут быть описаны единым математическим аппаратом [1]. Сущность данного аппарата состоит в его комплексной природе, включая такие направления как статистика вынужденных или случайных блужданий, описание диффузии дифференциальными уравнениями в частных производных и элементы неравновесной термодинамики.

Для исследования процессов аномальной диффузии применяется несколько подходов, использующих переменные коэффициенты диффузии, корреляции дробного порядка, дробные лапласианы, скачкообразные блуждания [2]. Наиболее точные и интересные результаты можно получить с помощью уравнений дробных производных.

Математический аппарат интегродифференцирования дробного порядка позволяет описывать процессы в системах, для которых существенен учёт нелокальных свойств по времени и пространству. В данном случае производные дробного порядка - метод учёта эффектов памяти и пространственных корреляций.

Уравнения в дробных производных описывают эволюцию физической системы с потерями, причём дробный показатель производной указывает на долю состояний системы, сохраняющихся за всё время эволюции. Такого рода системы с “остаточной” памятью занимают промежуточное положение между системами, обладающими полной памятью, и марковскими системами.

Дифференциальное уравнение диффузии устанавливает связь между временным и пространственным изменением концентрации в любой точке среды, в которой происходит диффузионный процесс.

Заметим, что дифференциальное уравнение нестационарной диффузии относится к параболическому типу, а стационарной диффузии – к эллиптическому типу дифференциальных уравнений.