МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования НИЖЕГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ им. Р.Е.АЛЕКСЕЕВА

Институт радиоэлектроники и информационных технологий



Конспект

по теме

«<u>V Диаграммы Чена и Мартина. БД и ХД Сравнения ETL V»</u>

по дисциплине

«Базы данных»

РУКОВОДИТЕЛЬ:	
	_ Мисевич П. В.
(подпись)	(фамилия, и.,о.)
СТУДЕНТ:	
	Сухоруков В.А.
(подпись)	(фамилия, и.,о.)
	19-B-2
	(шифр группы)
Работа защищена «	
С оценкой	

Диаграммы E-R. Диаграммы Чена и Мартина.

Модель сущность-связь (ЕR-модель) — модель данных, позволяющая описывать концептуальные схемы предметной области.

ER-модель используется при высокоуровневом проектировании баз данных. С её помощью можно выделить ключевые сущности и обозначить связи, которые могут устанавливаться между этими сущностями.

Нотация Питера Чена

Эта нотация была представлена Питером Ченом, являющимся одним из основоположников реляционных баз данных, и долгое время применялась для графической интерпретации предметной области в терминах сущностей и связей, иллюстрирующих ее абстрактное представление на логическом и концептуальном уровнях.

Множества сущностей изображаются в виде прямоугольников, множества отношений изображаются в виде ромбов. Если сущность участвует в отношении, они связаны линией. Если отношение не является обязательным, то линия пунктирная. Атрибуты изображаются в виде овалов и связываются линией с одним отношением или с одной сущностью.

Особой ценностью для разработчика модели в нотации Чена является возможность ее использования на концептуальном уровне, когда еще нет четкого представления о структуре будущей базы данных, но есть описание объектов предметной области с их атрибутивным составом и пониманием связей между ними. Это позволяет на начальном этапе разработки базы данных представить структурную модель, используя терминологию предметной области, но при этом подготовленную для трансформации в модель базы данных.

Нотация Мартина

Модели такого типа менее громоздки по сравнению с моделями в нотации Питера Чена. Тем не менее, сложность предметных областей нередко мешает представлению всей модели в рамках единого рабочего пространства, что во многих средствах моделирования баз данных исправляется возможностью формирования и представления моделей базы данных в разрезе отдельных рабочих пространств, которые могут соответствовать функциональному делению предметной области или какому-либо другому фактору, уменьшающему количество рассматриваемых элементов модели базы данных.

Основу всей модели базы данных в нотации Мартина составляют элементы "Сущность", представляемые прямоугольником с указанием существительного в качестве названия сущности. В некоторых случаях допускается использовать словосочетания, обозначающие особенности данных, которые представляются описываемой сущностью.

При описании сущностей в нотации Мартина и многих других нотациях, в отличие от нотации Питера Чена, атрибуты указываются не рядом с сущностью в виде отдельных элементов модели, а в ней самой, и являются составной частью сущностей. При этом указание на ключевые атрибуты обозначается с помощью соответствующего подчеркивания.

Устанавливая связи между сущностями, в нотации Мартина ее смысловое наполнение можно обозначать единственной глагольной формой, имеющей смысл связи от "левой" сущности к "правой" сущности, представляя в качестве "левой" сущности ту, у которой кардинальность связи в верхней се границе равна "1". В случае установления связи многие — ко — многим "левой" сущностью является та, которая по логике модели является более значимой.

БД и ХД. Сравнение ETL V.

Системы оперативной обработки информации получили название **OLTP** (On-Line Transaction Processing – оперативная, то есть в режиме реального времени, обработка транзакций).

Транзакция — некоторый набор операций над базой данных, который рассматривается как единое завершенное, с точки зрения пользователя, действие над некоторой информацией, обычно связанное с обращением к базе данных.

Понимание преимуществ, которые способен дать интеллектуальный анализ, привело к появлению нового класса систем — **информационных систем поддержки принятия решений** (информационных СППР), ориентированных на аналитическую обработку данных с целью получения знаний, необходимых для разработки решений в области управления.

Хранилище данных — разновидность систем хранения, ориентированная на поддержку процесса анализа данных, обеспечивающая целостность, непротиворечивость и хронологию данных, а также высокую скорость выполнения аналитических запросов.

Отличия БД и ХД:

Свойство	БД	ХД
Изменение данных	Данные могут добавляться,	Допускается только пополнение,
	изменяться и удаляться	ранее добавленные данные
		изменяться не должны, что
		позволяет обеспечить
		хронологию
Периодичность	Часто, но в небольших	Редко, но в больших объемах
обновления	объемах	
Доступ к данным	Должен быть обеспечен	Должен быть обеспечен доступ к
	доступ ко всем текущим	историческим данным с
	данным	соблюдением их хронологии
Характер выполняемых	Стандартные, настроенные	Нерегламентированные,
запросов	заранее	формируемые аналитиком «на
		лету» в зависимости от
		гребуемого анализа
Время выполнения	Несколько секунд	До нескольких минут
запросов		
Время хранения данных	Как правило, не более года	Годы, десятилетия

Консолидация — комплекс методов и процедур, направленных на извлечение данных из различных источников, обеспечение необходимого уровня их информативности и качества, преобразование в единый формат, в котором они могут быть загружены в хранилище данных или аналитическую систему.

Источник данных — объект, содержащий структурированные данные, которые могут оказаться полезными для решения аналитической задачи. Необходимо, чтобы используемая аналитическая платформа могла осуществлять доступ к данным из этого объекта непосредственно либо после их преобразования в другой формат.

В основе конструкции консолидации лежит процесс ETL. Процесс ETL решает задачи извлечения данных из разнотипных источников, их преобразования к виду, пригодному для хранения в определенной структуре, а также загрузки в соответствующую базу или хранилище данных. Если у аналитика возникают сомнения в качестве и информативности исходных данных, то при необходимости он может задействовать процедуры оценки их качества, очистки или обогащения, которые также являются составными частями процесса консолидации данных.