

Моделирование данных (Data Modelling)

Моделирование данных заключается в определении и анализе требований к данным, необходимым для поддержки бизнес-процессов в компании.

Требования к данным первоначально записываются в виде концептуальной модели, которая представляет собой набор спецификаций данных и используется для обсуждения начальных требований заинтересованными сторонами. Затем концептуальная модель преобразуется в логическую, которая описывает структуры данных. Реализация одной концептуальной модели может потребовать использования нескольких логических моделей. И наконец, последний шаг в моделировании данных — преобразование логической модели в физическую. Моделирование данных определяет не только элементы данных, но также их структуры и отношения между ними.

Главная цель моделирования данных — дать возможность компании управлять ими как ресурсом. К основным задачам моделирования данных относятся:

- помочь бизнес-аналитикам и IT-специалистам использовать единую информационную модель компании при решении стоящих перед ними задач;
- осуществлять интеграцию ресурсов в информационную систему компании;
- разрабатывать и проектировать базы и хранилища данных.

Окончательных, завершенных моделей данных не существует, поскольку они должны изменяться вместе с изменениями в бизнесе. Для хранения различных версий модели данных обычно используют специальные репозитории, из которых модель можно извлекать, расширять и редактировать.

Различают два уровня моделирования данных: стратегический и аналитический:

- **Стратегическое моделирование** — это часть создания стратегии информационной системы компании, которая определяет ее общее представление и архитектуру.
- **Аналитическое моделирование** — моделирование данных в процессе системного анализа, когда логические модели данных создаются как часть разработки новых баз данных.

Существует ряд проблем, которые необходимо учитывать при моделировании данных:

1. Бизнес-правила, специфичные для определенного периода или процесса, жестко фиксируются в модели данных. Это означает, что даже небольшие изменения в бизнесе ведут к значительным изменениям в информационной системе компании. Поэтому модель данных должна быть достаточно гибкой, чтобы изменения в бизнесе могли быть отражены в ней относительно быстро и эффективно.

2. Типы сущностей в концептуальной модели компании не идентифицируются или идентифицируются неправильно. Это может привести к репликации данных и соответствующим затратам. Поэтому определения данных должны быть как можно более явными и понятными, чтобы минимизировать неправильное толкование и дублирование.
3. Модели данных для различных подсистем информационной системы компании могут существенно различаться. В результате для обмена данными между ними потребуются сложные интерфейсы, затраты на создание и поддержку которых могут достигать 25-70% от стоимости системы в целом. Поэтому модели данных должны разрабатываться с учетом минимизации стоимости организации их взаимодействия.
4. Игнорирование стандартов, которые обеспечивают соответствие моделей данных бизнес-потребностям, может привести к тому, что данные нельзя будет передать в электронном виде клиентам, поставщикам или партнерам, потому что структура и форматы данных не были стандартизированы.