

Проектирование Data Warehouse (DWH) - основы

Начало проектирования

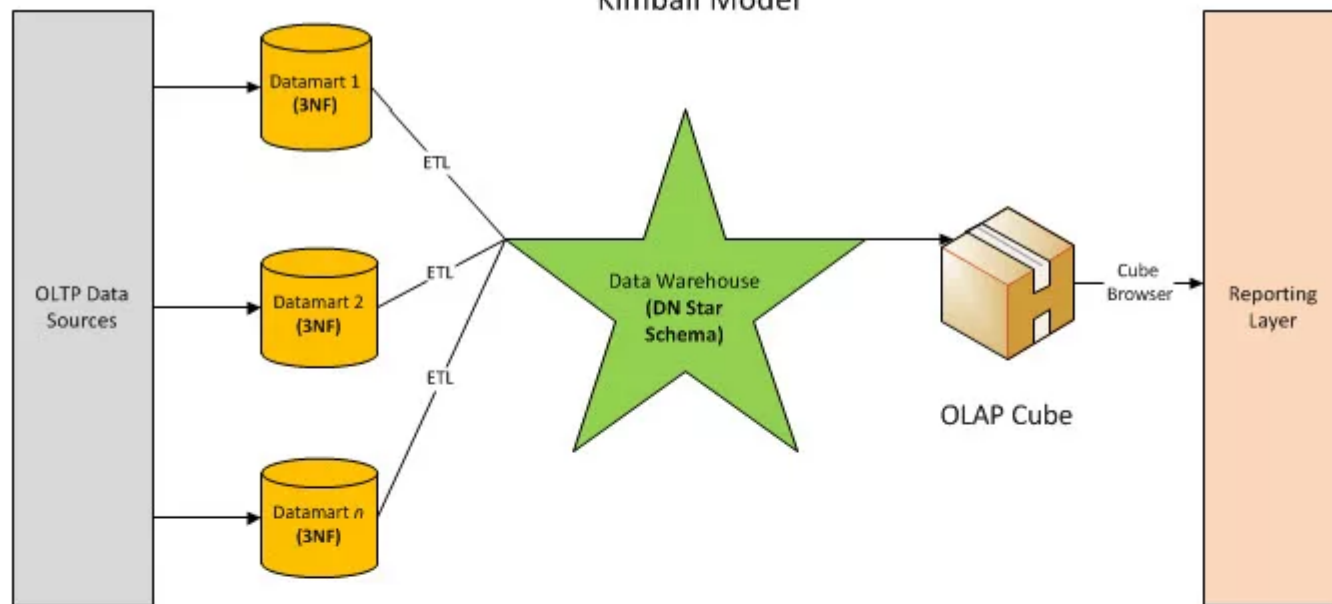
При проектировании хранилища данных (DWH) выбор начального слоя зависит от того, какие данные и для каких целей вам нужны. Можно было бы предположить, что разработку следует начинать со слоя staging (STG), так как это логическая точка входа данных в систему. Однако это не всегда эффективно, поскольку без понимания конечного использования данных можно тратить ресурсы на обработку информации, которая в конечном счете окажется ненужной.

Вместо этого, более целесообразно следовать потребностям бизнеса. **Подходы, предложенные Ральфом Кимбаллом и Биллом Инмоном, представляют две различные философии проектирования DWH, каждая из которых подчеркивает разные аспекты важности.**

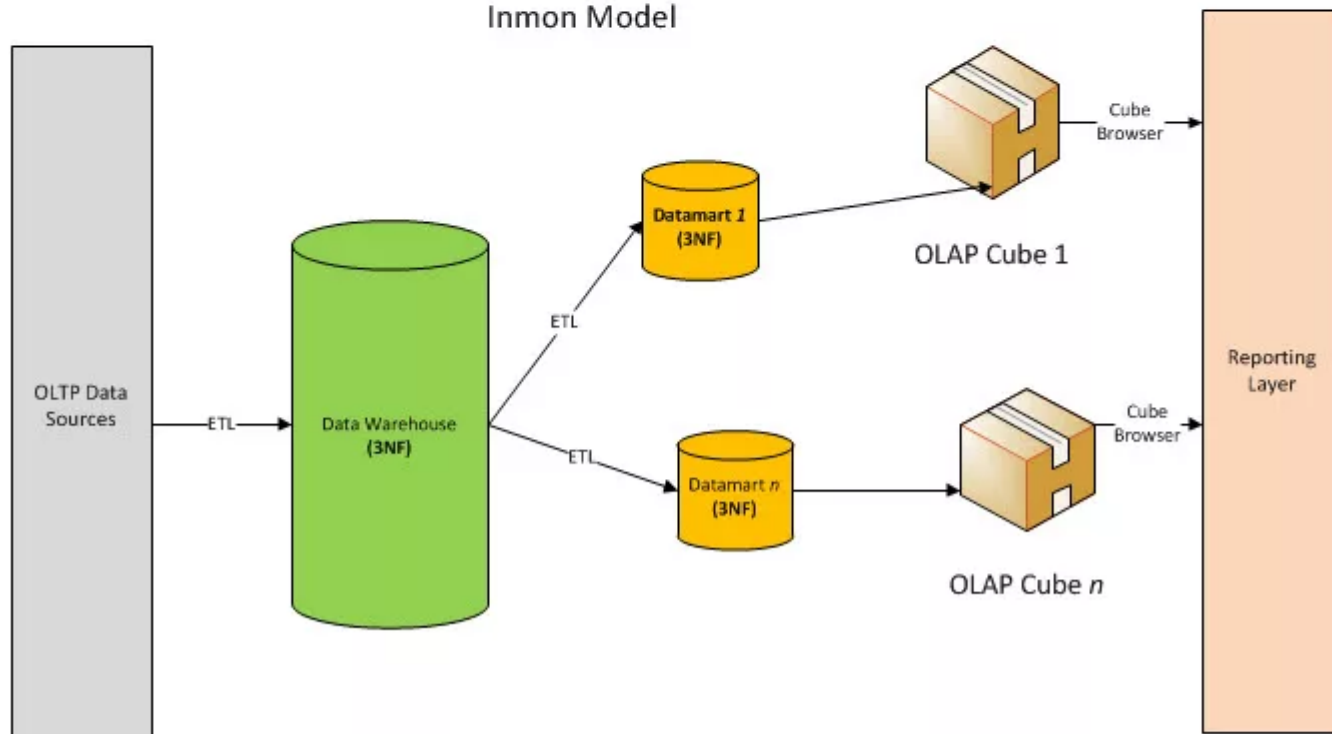
Подход Ральфа Кимбалла утверждает, что проектирование должно начинаться с витрин данных (CDM), поскольку они отражают конкретные потребности бизнеса. Сначала выясняется, какие данные нужны для витрин, а затем строится остальная часть хранилища, чтобы обеспечить эти потребности. Это практический и быстрый способ, который обеспечивает непосредственные результаты для бизнеса.

Подход Билла Инмона выдвигает на первый план корпоративную модель данных. Здесь акцент делается на всестороннем представлении предметных областей бизнеса в нормализованной форме в DWH. Именно из этой полной и интегрированной модели данных (DDS) затем формируются витрины. Этот подход требует больших первоначальных инвестиций времени и ресурсов, но обеспечивает более обширные и глубокие возможности анализа данных в долгосрочной перспективе.

Kimball Model



Inmon Model





Подход Ральфа Кимбалла

Преимущества:

- Ориентация на действие: Вы строите хранилище, исходя из того, что нужно здесь и сейчас, что обеспечивает быструю отдачу.
- Простая архитектура: Сложности упрощаются благодаря модели "звезда", делая систему более прозрачной для понимания и использования.
- Экономия ресурсов: Минимум людей способен обеспечить функционирование системы, что уменьшает затраты.

Недостатки:

- Частичная картина: В хранилище может не хватить некоторых данных, что может привести к необходимости последующих доработок.
- Трудности с масштабированием: Если потребности бизнеса меняются, могут потребоваться значительные изменения в структуре данных.
- Ограниченная гибкость: Система может быть не готова к неожиданным запросам, так как первоначально нацелена на узкий круг задач.

Представим, что вы открываете кафе и хотите сразу начать продавать кофе, потому что это то, что ваши клиенты хотят больше всего. Вы быстро настраиваете кофейную машину и начинаете обслуживание, не теряя времени на организацию полного меню. Таким образом, вы быстро отвечаете на спрос и оптимизируете процесс под нужды клиентов.

Но если вам внезапно потребуется добавить чай в меню, вам, возможно, придется полностью пересмотреть организацию вашего кафе, потому что изначально вы планировали продавать только кофе.

Подход Билла Инмона

Преимущества:

- Целостность: Система предлагает полную и объединенную картину данных, что полезно для комплексного анализа.

- Консистентность: Интеграция данных обеспечивает единообразие и надежность информации.
- Глубина анализа: Позволяет проводить детальный анализ данных, создавать множество разнообразных отчетов.

Недостатки:

- Сложность: Система требует сложного подхода к моделированию данных и может быть непонятна без специальных знаний.
- Высокие затраты: Начальное проектирование и настройка занимают много времени и ресурсов.
- Требуется специалистов: Для эффективного управления и поддержки хранилища нужен штат квалифицированных сотрудников.

Возвращаясь к аналогии с кафе, представьте, что вы сперва тщательно спланировали каждую деталь — от меню до организации кухни, чтобы убедиться, что можете предложить разнообразные напитки и блюда. Это дает вам полную картину того, как будет функционировать ваше кафе, и позволяет вам легко адаптировать меню в будущем.

Однако для открытия такого комплексного кафе потребуется больше времени и ресурсов, и вам нужно будет нанять более квалифицированный персонал, чтобы управлять более сложными процессами на кухне.