1. Что такое консолидация данных?

Консолидация — это объединение данных из одной или нескольких областей данных и вывод их в виде таблицы в итоговом листе. При консолидации данных они компонуются так, что их становится проще обновлять и обобщать на регулярной основе или при необходимости.

1. Какие есть способы консолидации данных?

Существует два основных способа консолидации данных.

1. Консолидация по расположению. Этот метод применяется, если данные из различных источников упорядочены одинаково и в них используются одни и те же подписи столбцов и строк (например, при наличии нескольких листов расходов, созданных из одного и того же шаблона).
2. Консолидация по категории. Этот метод применим, если данные из различных источников упорядочены по-разному, но и в них используются одни и те же подписи столбцов и строк (например, при наличии листов с данными о запасах для каждого месяца, которые имеют одинаковый макет, но содержат разные элементы или разное число элементов).
3. Что такое ETL?

ETL – аббревиатура от Extract, Transform, Load. Это системы корпоративного класса, которые применяются, чтобы привести к одним справочникам и загрузить в DWH и EPM данные из нескольких разных учетных систем.

1. Зачем нужна ETL система?

Проблема, из-за которой в принципе родилась необходимость использовать решения ETL, заключается в потребностях бизнеса в получении достоверной отчетности из того бардака, который творится в данных любой ERP-системы.

1. Как работает ETL система?
2. Процесс загрузки – Его задача затянуть в ETL данные произвольного качества для дальнейшей обработки, на этом этапе важно сверить суммы пришедших строк, если в исходной системе больше строк, чем в RawData то значит — загрузка прошла с ошибкой;
3. Процесс валидации данных – на этом этапе данные последовательно проверяются на корректность и полноту, составляется отчет об ошибках для исправления;
4. Процесс мэппинга данных с целевой моделью – на этом этапе к валидированной таблице пристраивается еще n-столбцов по количеству справочников целевой модели данных, а потом по таблицам мэппингов в каждой пристроенной ячейке, в каждой строке проставляются значения целевых справочников. Значения могут проставляться как 1:1, так и \*:1, так и 1:\* и \*:\*, для настройки последних двух вариантов используют формулы и скрипты мэппинга, реализованные в ETL-инструменте;
5. Процесс агрегации данных – этот процесс нужен из-за разности детализации данных в OLTP и OLAP системах. OLAP-системы — это, по сути, полностью денормализованная таблица фактов и окружающие ее таблицы справочников (звездочка/снежинка), максимальная детализация сумм OLAP – это количество перестановок всех элементов всех справочников. А OLTP система может содержать несколько сумм для одного и того же набора элементов справочников. Можно было-бы убивать OLTP-детализацию еще на входе в ETL, но тогда мы потеряли бы «аудиторский след». Этот след нужен для построения Drill-down отчета, который показывает — из каких строк OLTP, сформировалась сумма в ячейке OLAP-системы. Поэтому сначала делается мэппинг на детализации OLTP, а потом в отдельной таблице данные «схлопывают» для загрузки в OLAP;
6. Выгрузка в целевую систему — это технический процесс использования коннектора и передачи данных в целевую систему.
7. Какое прикладное назначение ETL?

Прикладное назначение ETL состоит в том, чтобы организовать такую структуру данных с помощью интеграции различных информационных систем. Учитывая, что BI-технологии позиционируются как «концепции и методы для улучшения принятия бизнес-решений с использованием систем на основе бизнес-данных», можно сделать вывод о прямой принадлежность ETL к этому технологическому стеку.

1. Какие задачи ETL-системы должны решать?
2. Привести все данные к единой системе значений и детализации, попутно обеспечив их качество и надежность;
3. Обеспечить аудиторский след при преобразовании (Transform) данных, чтобы после преобразования можно было понять, из каких именно исходных данных и сумм собралась каждая строчка преобразованных данных.
4. Что нужно знать для использования ETL?

Для достижения успеха при переносе данных из одной системы в другую крайне важно четко представлять процессы ETL, а также структуру исходного приложения и приложения назначения.