Введение

Гипотеза: Можно создать иерархическую структуру parent-child в Oracle Business Intelligence (Oracle BI), которая будет адекватно отображать предметную область "Подразделение" с сохранением ранее созданных связей и данных, а так же сможет стать основой при разработке иерархических структур parent-child для других предметных областей компании АО "Эр-Телеком Холдинг"

Задача 1. Анализ ПО, инструментов, уже имеющихся разработок в компании и за её пределами

* 1. **СУБД**

Данные в компании хранятся в реляционных базах данных, которые обеспечиваются СУБД Oracle.

Реляционная база данных – это связанная информация, представленная в виде двумерных таблиц. Двумерные таблицы состоят из строк и столбцов. У каждой строки есть уникальный идентификатор – это может быть одно поле или несколько – который называется первичным ключом. По нему легко выполнить поиск в таблице одной конкретной строки. Вся информация в базе хранится в виде строк – множества полей, привязанных к конкретному идентификатору, причем число этих полей совпадает с числом столбцов.[1]

Для доступа к реляционным базам данных используется декларативный язык программирования SQL.

SQL (Structured Query Language) — это структурированный язык запросов. На этом языке можно формулировать выражения (запросы), которые извлекают требуемые данные, модифицируют их, создают таблицы и изменяют их структуры, определяют права доступа к данным и многое другое. [2]

СУБД – это программное обеспечение, которое используется для создания и работы с базами данных. Главная функция СУБД – это управление данными (которые могут быть как во внешней, так и в оперативной памяти). СУБД обязательно поддерживает языки баз данных, а также отвечает за копирование и восстановление данных после каких-либо сбоев. [3]

В компании АО «ЭР-Телеком Холдинг» используется СУБД Oracle. В основе большей части Oracle лежит PL/SQL — язык программирования, который предоставляет процедурные расширения используемой в Oracle версии SQL, а также служит языком программирования инструментария Oracle. [4]

СУБД Oracle расположена на двух основных серверах: SA и DWH.

Сервер SA (System Administration) взаимодействует с другими источниками данных, обрабатывает данные, хранит данные короткие промежутки времени. Отсутствует историчность. Лишние данные и таблицы регулярно вычищаются. На этом сервере можно производить основную обработку данных и приведение к необходимому виду.

Сервер DWH позволяет хранить большой объем данных, на нем содержаться исторические данные по многим направлениям деятельности, включая блок продаж, HR-блок, блок выручки/затрат и т.д. Информация на DWH пополняется ежедневно, напрямую записывается в таблицы, без дополнительной обработки. Обработка данных на DWH производится в исключительных случаях.

В одном из источников предложено следующее определение для DWH:

DWH — это система данных, отдельная от оперативной системы обработки данных. В корпоративных хранилищах в удобном для анализа виде хранятся архивные данные из разных, иногда очень разнородных источников. Эти данные предварительно обрабатываются и загружаются в хранилище в ходе процессов извлечения, преобразования и загрузки, называемых ETL. Решения ETL и DWH — это одна система для работы с корпоративной информацией и ее хранения. [5.д https://mcs.mail.ru/blog/chto-takoe-dwh-i-pochemu-bez-nih-dannye-kompanii-bespolezny]

* 1. **OLAP кубы, метрики, измерения; ETL**

[6.д https://cyberleninka.ru/article/n/olap-tehnologii]

Куб OLAP представляет собой структуру данных, которая обеспечивает возможность быстрого анализа данных за рамками ограничений реляционных баз данных. Кубы способны отображать и суммировать большие объемы данных, также предоставляя пользователям доступ к любым точкам данных с возможностью поиска. Таким образом, данные могут быть сведены, фрагментированы и обработаны по мере необходимости для решения самых широкого спектра вопросов, относящихся к интересующей вас области пользователя.

Измерение примерно эквивалентно классу пакета управления. Каждый класс пакета управления имеет набор свойств, а каждое измерение — набор атрибутов, при этом каждый атрибут сопоставляется с одним свойством класса. Измерения позволяют выполнять фильтрацию, группирование и маркировку данных. К примеру, можно отфильтровать компьютеры по установленной операционной системе или сгруппировать людей по категориям, используя пол или возраст. Затем данные могут быть представлены в формате, где данные классифицируются по категориям и категориям, что позволяет более - глубоко анализировать анализ. Измерения также могут иметь естественные иерархии, позволяющие пользователям "детализировать" до более детального уровня детализации. К примеру, измерение даты обладает иерархией, позволяющей выполнять детализацию до уровня лет, затем — до уровней кварталов, месяцев, недель и отдельных дней.

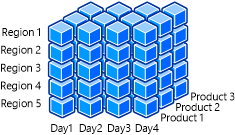


Рис.1 – OLAP куб

В этом рисунке 1 показан куб OLAP, содержащий измерения даты, региона и продукта.

Меры, метрики — это числовые значения, позволяющие пользователям создавать плоскостные и объемные срезы, выполнять агрегирование и анализ. Они являются одной из основных причин построения кубов OLAP на основании инфраструктуры хранилищ данных. При помощи служб SSAS можно создавать кубы OLAP, использующие бизнес-правила и вычисления для форматирования и отображения мер в настраиваемом формате. Большой объем времени разработки куба OLAP тратится на определение того, какие меры будут отображены, и каким образом они будут вычисляться.

Когда пользователь детализирует данные куба OLAP, он анализирует данные на другом уровне уплотнения. Уровень детальности данных повышается с каждой операцией детализации, что позволяет пользователю изучать данные на разных уровнях иерархии. По мере детализации пользователь переходит от общей информации к данным, имеющим более узкий фокус. [5]

На текущий момент в компании АО «ЭР-Телеком Холдинг» иерархия реализована не в том виде, в котором бы её хотели видеть конечные пользователи. Ниже на рис.2 представлена текущая иерархическая разбивка по разным уровням измерения OLAP-куба.

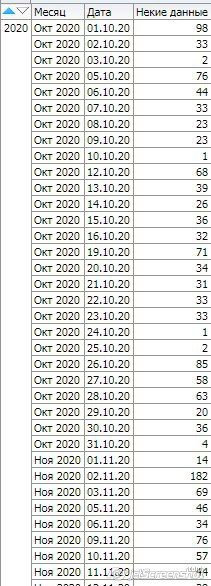


Рис.2 – Данные OLAP-куба в иерархической разбивке

На рис.2 представлены данные, которые можно посмотреть в разрезе года, месяца, дня. Их можно посмотреть одновременно во всех разрезах, но каждый из них будет выглядеть, как отдельный столбец. Конечные пользователи же хотят видеть иерархию на подобии файловой структуры (рис.3), которую можно свернуть/развернуть и в которой можно посмотреть метрики на каждом уровне иерархии.



Рис.3 – Пример файловой иерархической структуры

На рис.3 представлена файловая иерархическая структура, которая является прототипом того, что конечные пользователи хотят видеть в анализах вместо того, что представлено на рис.2.