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".

Далее в тексте «заказчик» равнозначно «компания 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.

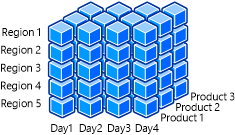


Рис.1 – OLAP куб

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» реализована уровневая иерархия. Она принципиально отличается от того, что хотели бы видеть конечные пользователи. Ниже на рис.2 представлена текущая уровневая иерархия по измерению «Каллендарь» OLAP-куба.

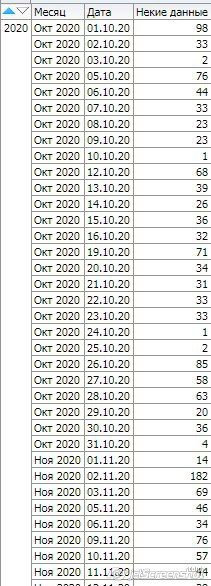


Рис.2 – Данные OLAP-куба в разрезе уровневой иерархии «Календарь»

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 иерархической структуры

На рис.3 представлена иерархическая структура, которая является прототипом того, что конечные пользователи хотят видеть в анализах вместо того, что представлено на рис.2. Видно, что элементы структуры могут содержать элементы нижнего уровня, разворачивать их, и, наоборот, прятать. При это расчетные метрики должны выводить для каждого уровня корректные значения, включающие как значения самого уровня, так и все внутренних.

* 1. **BI**

Business Intelligence (BI, Бизнес-аналитика) — это набор IT-технологий для сбора, хранения и анализа данных, позволяющих предоставлять пользователям достоверную аналитику в удобном формате, на основе которой можно принимать эффективные решения для управления бизнес-процессами компании. [6]

Все уровни пользователей, от сотрудников до учредителей, получают гибкий доступ к необходимой им управленческой отчетности, не прибегая к помощи IT-специалистов.

Oracle BI – это основной инструмент для аналитической работы с данными, который используют конечные пользователи, предоставленный компанией АО «ЭР-Телеком Холдинг». Именно с помощью этого инструмента строятся анализы, которые должны будут содержать иерархическую структуру. Заказчиком предоставлено обеспечение взаимодействия BI и серверов баз данных. Для сохранения корректного взаимодействия при описании новой структуры данных необходимо придерживаться регламентов компании.

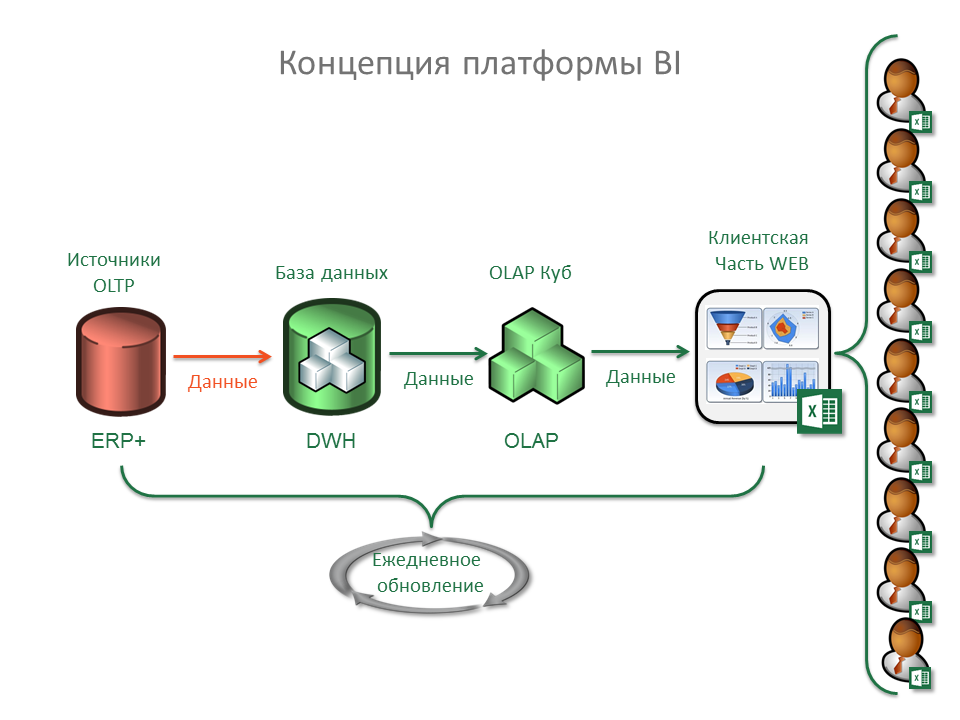
[](http://biweb.ru/wp-content/uploads/2016/09/BI.png)

Рисунок 2 – Концепция платформы BI

Концепция платформы BI, представленная на рисунке 2, состоит из следующих компонентов:

* ETL-инструменты: программы, позволяющие выполнять загрузку данных в DWH из различных учетных систем. Заказчиком в качестве ETL-инструмента предоставлена программа Oracle Date Integrator 11g.
* DWH-хранилище: полноценная база данных SQL для подготовки и хранения данных для аналитики.
* OLAP-кубы: технология, позволяющая делать в реальном времени (1-5 секунд) любые отчеты и проводить полноценный анализ данных.
* Клиентские приложения: как правило, для детального анализа данных и построения динамических отчетов пользователи используют сводные таблицы, подключенные к OLAP-кубам. В компании АО «ЭР-Телеком Холдинг» в качестве клиентского приложения выступает Oracle BI.

Основные преимущества BI:

* Скорость построения отчетов.
* Динамический анализ данных в любой детализации
* Быстрый анализ любого объема данных (технология OLAP).
* Автоматизация подготовки данных для отчетов и построения корпоративной отчетности.
* Консолидация данных (данные для отчетов могут быть в разных учетных системах).
* Анализ показателей План/Факта, Анализ выполнения различных KPI.
* Удобная визуализация данных на Web (при этом обновленные данные в отчет поступают автоматически).
* Единый и удобный доступ к аналитической отчетности для всех сотрудников через корпоративный BI портал.
* Уменьшение нагрузки на Учетные системы.

В целом — повышение общей управляемости и эффективности бизнеса. [Navid Mahlouji. Information Technologies for Business Intelligence// CiteseerX. – 2014. – С. 102 / [Электронный ресурс]. – Режим доступа: URL: https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.978.7425]

* 1. **Инструменты**

Для разработки измерения заказчиком были предоставлены следующие программы: Oracle Data Integrator 11g (ODI), Tool for Oracle Application Developers (TOAD), Oracle Business Intelligence Enterprise Edition (OBIEE). Заказчик строго ограничил использование каких-либо других инструментов, помимо указанных.

Далее приведено краткое описание каждой из используемых программ и их назначение.

Oracle Data Integrator (ODI) — это интеграционная платформа корпоративного уровня, которая обеспечивает извлечение, преобразование и загрузку данных из разнообразных источников: баз данных, файлов и других источников (например, LDAP каталогов или WEB-сервисов). Позволяет создавать автоматические пакеты сбора и обработки данных. [7]

Tool for Oracle Application Developers (TOAD) - инструмент для разработчиков приложений Oracle. Позволяет просматривать базы данных, выполнять PL/SQL код, использовать средства отладки. [8]

Oracle Business Intelligence Enterprise Edition (OBIEE) – инструмент для описания структуры BI. Позволяет описать структуру таблиц, их связи, пути поиска данных при построении анализов. Описанная в АТ структура является основной связи Oracle BI и DWH.

Структура данных в АТ состоит из трех частей:

Физический уровень – здесь хранится полное описание таблиц сервера: название, поля, типы полей, первичные ключи, внешние ключи и так далее.

Бизнес-уровень – здесь хранится описание структур запросов, а именно: к каким таблицам необходимо делать запрос в каком случае, если есть несколько таблиц агрегатов; на какую таблицу связей опираться при создании иерархического запроса; как считать метрики в разрезе измерений; и так далее.

Презентационный уровень – здесь хранится структура в таком виде, в каком видят её пользователи. Сюда выносятся необходимые метрики и измерения, группируются по предметным областям. [http://www.olap.ru/home.asp?artId=2344]

2. Что уже есть

На текущий момент в компании АО «ЭР-Телеком Холдинг» нет ни одного примера успешной реализации иерархической структуры в BI. Данная разработка является принципиально новой. На текущий момент в компании АО «ЭР-Телеком Холдинг» реализована уровневая иерархия, представленная на рис.2.

Для решения поставленной задачи было исследовано множество литературы. В статье «Oracle BIEE 11g – Parent Child Hierarchies – Multiple Modeling Methods» [<https://www.rittmanmead.com/blog/2010/11/oracle-bi-ee-11g-parent-child-hierarchies-multiple-modeling-methods/> - статья по созданию иерархии] продемонстрировано возможное решение поставленной задачи через иерархическую структуру в OBIEE. Описанная в статье структура не удовлетворяет условиям, поставленным регламентами компании, но доказывает, что существует возможность решения поставленной задачи.

Однако в статье не указано на чем строится данная иерархическая структура. В технической документации Oracle было найдено описание построения ключевых объектов и связей для создания parent-child иерархии. [https://docs.oracle.com/middleware/12211/biee/BIEMG/GUID-0424E7A7-C7DB-447A-B0C4-0BD6790888EA.htm#hpp\_l\_value\_dimension – документация по OBIEE].

На основе этих документов составлено описание будущей разработки, которые было бы понятно и разработчикам, и конечным пользователям, и заказчикам в рамках компании АО «ЭР-Телеком Холдинг». И указано её отличие от существующей уровневой иерархии, так как далее подразумевается переход от одной структуры к другой.

Создание многоуровневой иерархии, где каждый элемент, кроме корневых, имеет вышестоящий элемент. Данная иерархия называется parent-child («родитель-потомок»). В отличие от уровневой иерархий, все элементы измерения иерархии «родитель-потомок» находятся в одном логическом столбце.

В иерархии «родитель-потомок» родительский элемент элемента находится в другой строке того же логического столбца, на который указывает родительский ключ. Это отличается от иерархии на основе уровней, где родительский элемент элемента находится в другом логическом столбце в той же строке. Другими словами, навигация в иерархии «родитель-потомок» следует за значениями данных, а навигация в иерархии на основе уровней следует за структурой метаданных.