

Реализация технологии OLAP в BI-платформе Prognosz Platform

В данном документе описаны особенности реализации технологии OLAP в BI-платформе Prognosz Platform. Для выполнения OLAP-анализа в платформе используется инструмент «Аналитические запросы (OLAP)», который доступен как в настольном приложении, так и через web.

Общая информация

В таблице ниже представлена общая информация о реализации технологии OLAP в Prognosz Platform.

Характеристики	Описание
Модель хранения данных	ROLAP
Поддерживаемые СУБД	Oracle Database, MS SQL Server, Teradata, PostgreSQL, SQLite, OLEDB/ODBC-совместимые СУБД
Поддержка иерархий Parent-Child	Да
Поддержка иерархий с уровнями	Да
Поддержка виртуальных кубов/мультикубов	Да
Поддержка вычисляемых кубов	Да
Поддержка In-Memory	Да
Язык вычислений	Да (собственный язык Fore, MDX для интеграции с внешними серверами)
API для клиентского доступа, разработки приложений	.Net, web-сервис JSON/SOAP-RPC, COM, встроенный язык программирования
Клиентские приложения	Собственный клиент (настольное приложение, web), Excel (собственная надстройка)
Интеграция со сторонними OLAP-серверами	MS SQL Server Analysis Services, SAP BW (MDX, OLEDB for OLAP)
Поддерживаемые ОС	Windows, Linux

Многомерная модель

Перед началом анализа данных в платформе необходимо определить многомерную модель предметной области и настроить отображение объектов многомерной модели на реляционные объекты источника данных. Для этого в платформе сперва регистрируются реляционные объекты источника данных – таблицы, представления, определяются пользовательские запросы. Затем на основе реляционных объектов определяются объекты многомерной модели – измерения и кубы.

Для измерений определяется реляционный объект-источник и перечень атрибутов. Для куба определяется перечень фактов и список входящих в него измерений, а также выполняется привязка фактов куба и атрибутов измерений к таблице фактов.

Как правило, в источнике для моделирования данных используется схема «Звезда».

Каждое измерение – самостоятельный объект и может использоваться в нескольких кубах. Различные измерения в рамках одного куба также могут использовать данные из различных источников.

Поддерживается и альтернативный вариант формирования многомерной модели. Пользователь сразу же создаёт справочники с перечнем атрибутов и кубы на их основе, при этом платформа автоматически генерирует реляционные объекты хранения данных и привязки к ним. Такой вариант может подойти при работе «с нуля», когда готовый источник отсутствует, а также для быстрого прототипирования.

Выборка данных из источников

В клиентском инструменте анализа данных пользователь размещает измерения в строки и столбцы кросстаблицы, выбирает интересующий его факт, настраивает фильтрацию по отдельным элементам измерения. Для формирования отчёта платформа анализирует используемые измерения и факты, привязки к реляционным объектам, а также настройки куба (см. ниже), и на основании этой информации генерирует SQL-запрос для извлечения данных из источников.

Стратегии фильтрации данных

Стратегии фильтрации определяют, как платформа сформирует условие фильтрации WHERE в SQL-запросе на извлечение данных из источника. Стратегию фильтрации можно переопределить для каждого измерения куба. Доступны следующие опции:

- Не выполнять фильтрацию в запросе. Будут выбраны все элементы измерения.
- Использовать предикат IN.
- Использовать IN или BETWEEN в зависимости от процента/количества фильтруемых элементов. Для слишком больших множеств элементов будет использоваться BETWEEN, т.к. в некоторых СУБД есть ограничение на длину списков элементов, в т.ч. в IN.
- По умолчанию. Платформа выбирает предикат фильтрации IN или BETWEEN автоматически.

Стратегии агрегации данных

В настройках кубах и в отчётах есть возможность управлять способами агрегации данных.

В кубах доступны следующие опции:

- Способ агрегация фактов. Выбирается функция агрегации для фактов с одинаковыми значениями ключевых атрибутов измерений куба. Агрегация выполняется в SQL-запросе на стороне источника данных.
- Опция агрегация по измерениям целиком. Если для измерения не задано никакого фильтра (т.е. выбираются все данные по этому измерению), измерение не участвует в выборке при извлечении данных. Агрегация выполняется на стороне источника данных. Используются функции агрегации, заданные для фактов, либо функция, переопределённая в отчёте (см. ниже).
- Агрегация по иерархиям измерений для родительских элементов. Выполняется на стороне BI-сервера.

В отчётах доступна возможность выбора функции агрегации по т.н. фиксированным измерениям, которые не участвуют в строках/столбцах кросстаблицы а используются в качестве фильтра. Агрегация выполняется на стороне BI-сервера, за исключением случая, когда для выборки по измерению используются все элементы и в кубе включена опция агрегации по измерениям целиком.

Вычисления

Механизм вычислений делает инструментальный анализа более гибким. Платформа позволяет определять вычисляемые атрибуты для реляционных объектов, вычисляемые факты в кубах, вычисляемые элементы измерений для вычисляемых кубов. Во всех вычислениях используются выражения на собственном языке Fore. Вычисления выполняются на стороне BI-сервера.

Механизм виртуальных кубов позволяет объединять несколько отдельных кубов (предметных областей/витрин данных) в единый куб по общим измерениям.

Повышение производительности

Для повышения производительности работы с большими объёмами данных в платформе можно использовать следующие механизмы:

- Партицирование таблиц. Данный механизм настраивается на стороне источника данных (если поддерживается) и полностью прозрачен для платформы.
- Использование преагрегированных данных. Отдельные таблицы источника или материализованные представления могут содержать агрегаты для определенного уровня иерархии (например, агрегаты за годы). В этом случае при определении куба в платформе факт можно привязать к нескольким таблицам фактов, и ключевые атрибуты для соответствующих уровней измерения привязать к соответствующим таблицам фактов. При выборке данных в разрезе годов платформа сформирует корректный запрос к таблице с агрегатами.
- Расчёт многомерных кубов на сервере. Этот механизм позволяет на основе кубов-источников вычислять значения практически для любых координат целевого куба, при этом расчёт выполняется на стороне источника данных. Результат вычислений вставляется в таблицу фактов куба. Расчёт выполняется путём генерации хранимой процедуры.
- Использование In-Memory кэша кубов. Если данные из реляционных объектов определённого куба извлекаются не достаточно быстро (например, используются представления или запросы), платформа позволяет извлечь всю информацию данного куба в оперативную память. В дальнейшем при работе с данным кубом вся обработка будет выполняться на стороне BI-сервера, без обращения к источнику.
- Кэширования измерений. При работе с большими измерениями кэширование измерения на BI-сервере помогает избежать дорогостоящих операций извлечения данных из источника.

Параметризация

Данная возможность также значительно повышает гибкость инструмента. Все объекты платформы могут иметь параметры, которые можно обрабатывать. Например, возможно определить параметры для запроса, измерения, куба. Есть следующие способы задать значения параметров при открытии объекта:

- Значение по умолчанию.
- Значение, задаваемое пользователем. Например, заданное значение может использоваться в условии фильтрации SQL-запроса при открытии отчёта.
- Значение от управляющего объекта. Поддерживается в кубах. Выбранное измерение становится управляющим, и отмеченные в нём элементы передаются в качестве параметра в управляемые (параметризованные) измерения куба. Параметры самого куба также можно использовать для управления параметризованными измерениями.

Этот механизм позволяет динамически перестраивать измерения в процессе интерактивного взаимодействия с кубом и навигации по иерархиям измерений.

- Программно.

Безопасность

Платформа позволяет разграничить права доступа на реляционные и многомерные объекты, готовые отчёты. Для измерений поддерживается возможность задавать права доступа на отдельные элементы. Администратор может выполнять аудит всех операций пользователей.

Для аутентификации пользователей поддерживается интеграция с Active Directory, SAML, OAuth.