DOI: 10.17122/ntj-oil-2017-4-119-127

УДК 004.652

К.А. Бондаренко, И.А. Мосякова, П.И. Чермянин (ООО «РН-Уфанипинефть», г. Уфа, Российская Федерация)

НЕКОТОРЫЕ ПОДХОДЫ К ПРИМЕНЕНИЮ АНАЛИТИЧЕСКИХ СИСТЕМ ПРИ РАЗРАБОТКЕ И ЭКСПЛУАТАЦИИ НЕФТЯНЫХ МЕСТОРОЖДЕНИЙ

K.A. Bondarenko, I.A. Mosyakova, P.I. Chermyanin (RN-UfaNIPIneft LLC, Ufa, Russian Federation)

SOME APPROACHES TO ANALYTICAL SYSTEMS APPLICATION FOR DEVELOPMENT AND OPERATION OF OIL FIELDS

Введение

В настоящее время действуют ограничения на ресурсоёмкие расчёты на геолого-гидродинамических моделях, что привело к появлению достаточно тривиальных с математической точки зрения моделей, используемых производственным персоналом. При этом большое количество скважин и геолого-технологических мероприятий, проводимых на них, увеличение количества «срезов» отчётности привели к разрастанию объема данных, в результате чего существенно увеличилось время расчётов, нивелировав тем самым их основное преимущество - скорость получения прогноза.

Цели и задачи

Рассмотреть вопрос оптимизации процесса оперативного прогнозирования уровней добычи нефти и жидкости нефтегазодобывающего общества на интервале бизнес-планирования.

Методы

Исторически для прогноза уровней добычи используются статистические методы, которые обеспечивают достаточный уровень точности прогноза (менее 1 % за первый год) на интервале планирования 1-5 лет и адаптируются под новые сценарные условия.

Background

At present, there are restrictions on resource-intensive calculations on geological and hydrodynamic models, which led to the appearance of quite trivial models from the mathematical point of view used by production personnel. At the same time, a large number of wells and geological and technological measures carried out on them, an increase in the number of «slices» of reporting led to an increase in the amount of data, which significantly increased the calculation time, thus eliminating their main advantage - the speed of obtaining a forecast.

Aims and Objectives

Consider the optimization of the operational forecasting process of the oil and gas production levels for the oil and gas producing company in the business planning interval.

Methods

Historically, statistical methods are used to forecast production levels, which ensure a sufficient level of forecast accuracy (less than 1 % for the first year) in the planning interval 1-5 years and adapt to new scenario conditions.

Результаты

Представлены результаты анализа опыта применения OLAP (On Line Analytical Processing), в частности применение технологии обратной записи (Writeback) для оперативного планирования уровней добычи нефти.

Results

The results of analyzing the experience of using OLAP (On Line Analytical Processing) are presented, in particular, the use of Writeback Technology for operational planning of oil production levels.

Ключевые слова: прогноз добычи, бизнес-аналитика, хранилище данных, ETL, OLAP, многомерные базы данных, визуализация данных, оптимизация, нефть, жидкость

Key words: production forecast, business intelligence, data warehouse, ETL, OLAP, multidimensional databases, data visualization, optimization, oil, liquid

Разработка BI-решения расчёта уровней добычи нефти

Бизнес-аналитика (BI) - инструменты, используемые для извлечения, хранения, анализа и предоставления информации пользователям. Основная цель BI - оперативно обеспечивать пользователей необходимой корректной информацией. Технология BI выбрана исходя из оценки сложности реализации поставленной задачи, она оптимальна с точки зрения минимизации затрат на разработку решения и максимизации её функциональности (таблица 1).

При выборе технологической платформы для реализации требуемого функционала было проведено сравнение BI-платформ ведущих вендоров (таблица 2). В результате была выбрана платформа Microsoft SQL Server, т.к. эта платформа в наибольшей степени отвечает функциональным требованиям.

Компонентами платформы BI являются [1] (рисунок 1):

1. Службы интеграции данных (SQL Server Integration Services) - платформа для выполнения операций извлечения, преобра-

зования и загрузки данных (Extract Transform Load - ETL), обеспечивающая заполнение хранилища данных и его синхронизацию с данными различных источников;

2. Хранилище данных (SQL Server Database Engine) - масштабируемая, высокопроизводительная система управления базой данных (СУБД), способная хранить большие объёмы данных, образующиеся в результате консолидации данных в единое хранилище для анализа и построения отчётов.

3. Аналитическая БД (SQL Server Analysis Services) - обеспечивает возможность построения OLAP-решений, включая возможность расчёта ключевых индикаторов производительности (KPI). Применяется также для построения data-mining - решений, которые используют специализированные алгоритмы для выявления трендов и зависимостей данных.

4. Сервер отчётов (SQL Server Reporting Services) - инструмент для построения отчётов, предназначенный для создания, публикации и распространения детализированных бизнес-отчётов как для внутренних, так и для внешних целей.

Таблица 1. Сравнение технологий расчёта уровней добычи нефти

	Макросы Excel	Собственный код	PCУБД	Self-BI ¹	BI
Наличие средств трансформации и загрузки данных	✓	✓	✓	✓	✓
Производительность интеграции	✓	✓	✓	✓	✓
Производительность вычисления агрегатов	✓	✓	✓	✓	✓
Автоматическое вычисление агрегатов по различным иерархиям	✓	✓	✓	✓	✓
Наличие инструментов контроля качества данных	✓	✓	✓	✓	✓
Наличие средств визуализации	✓	✓	✓	✓	✓
Распространенность в Компании	✓	✓	✓	✓	✓
Трудоемкость реализации	✓	✓	✓	✓	✓
<p>Обозначение цветом: ✓ - низкий уровень; ✓ - средний уровень; ✓ - высокий уровень. ¹Self-BI - технология BI, реализованная десктопными приложениями, не требует специальных знаний, позволяет интегрировать и трансформировать данные, строить табличные модели, визуализировать данные и, опционально, публиковать визуализации на сервере.</p>					

При разработке хранилища данных был проведён анализ источников данных, используемых в бизнес-процессе. В качестве основного источника данных выступала так называемая модель, состоящая из 200 связанных между собой ссылками книг Excel, содержащих как исходные данные, так и результаты расчётов, представленные в различных разрезах и формах. После выделения источников данных и их атрибутивного состава были разработаны многомерная модель и реляционная ER-модель DWH [2].

Заключительным этапом разработки хранилища была разработка оптимальной физической модели с учётом профиля нагрузки при ETL и процессинге многомерной БД. На последующих этапах при разработке ETL и многомерной БД в физическую модель вносились изменения.

В рамках разработки решения были созданы процедуры извлечения, трансформации и загрузки (ETL) данных [3]. Процедуры ETL включали в себя параллельное извлечение данных из книг Excel во временную область (staging), их трансформацию к модели DWH и загрузку в постоянную область хранения (target). Основной задачей процесса преобразования являлась подготовка данных к хранению в том виде, который оптимизирован для быстрого исполнения запросов. Для гибкой процедуры загрузки данных использовались возможности параметризации пакетов с помощью книги MS Excel, описывающей данные и метаданные (перечень загружаемых файлов, диапазоны данных) [4]. Данная книга показывает, какие файлы относятся к конкретным измерениям, а измерения к ключам.

Таблица 2. Сравнение вендоров технологии BI

	Microsoft	Oracle	SAS	Tableau	Qlik
Выполнение на сервере	✓	✓	✓	✓	✓
Средства разработки ETL	✓	✓	✓	-	-
Создание многомерных моделей	✓	✓	✓	✓	✓
Агрегативные функции	✓	✓	✓	✓	✓
Поддержка иерархий	✓	✓	✓	✓	✓
Создание вычислительных мер	✓	✓	✓	-	-
Онлайн выполнения запросов с различными группировками и срезами	✓	✓	✓	✓	✓
Возможность визуального построения запроса из MS Excel, сводных таблиц	✓	✓	-	✓	✓
Возможность навигации по модели drill-down, drill-through, roll-up)	✓	✓	✓	✓	✓
Возможность проведения анализа «что-если»	✓	-	✓	✓	✓

Рисунок 1. Компоненты BI технологии

Такая архитектура решения позволяет:

- корректировать модель загрузки неквалифицированным в области информационных технологий (ИТ) пользователям;
- загружать модель в новые версии плана;
- указывать способ учёта проведения геолого-технологических мероприятий (ГТМ) на скважинах на базовом/бездействующем фонде;
- отдельно выделять типы ГТМ с множественным проведением на скважинах операций гидроразрыва пласта (МГРП);
- оперативно добавлять в модель новые месторождения, деление по видам проведённых ГТМ.

В процедурах ETL были применены различные техники управления первичными ключами таблиц измерений:

- генерируемые СУБД при операциях массовой вставки identity-суррогатные ключи - например, проектные кусты;
- генерируемые в ETL HiLo-суррогатные ключи - например, для каждого месторождения выделен определенный диапазон значений первичных ключей темпов;
- использование мастер-данных для таблиц измерений с применением гибкого поиска соответствий (fuzzy lookup) на этапе ETL по бизнес-ключу - например, в источниках месторождение «Месторождение» может быть записано как «Мест», «Мест-е», а также «М/р».

Применение средств параллелизма платформы BI позволило достичь выполнения процедур ETL на многопроцессорных системах за 10 минут. Обновление же данных VBA кодом в исходных книгах Excel занимало несколько часов, при этом, разумеется, возможности вертикального масштабирования ограничены, а горизонтального - вообще отсутствуют.

Разработка многомерной модели - важная задача, решаемая при проектировании разработки месторождения [5]. Правильно выстроенная модель

обеспечивает простоту поддержания DWH и многомерной БД, повышает масштабируемость и производительность решения.

Разработка многомерной модели данных включала в себя построение измерений, формирование групп мер, определение связей между группами мер и измерениями, создание вычисляемых мер в скрипте куба. Вычисляемые меры (показатели добычи нефти, жидкости и др.) были созданы на основе формул, заложенных в модель расчета, реализованную в исходных книгах Excel [6].

Использование OLAP-куба за счет поддержки измерений и иерархий обеспечивает возможность оперативного получения информации по любым интересующим срезам, не зависимо от уровня детализации.

В разработанном решении для оперативного пересчёта уровней добычи нефти под новые сценарные условия использовался анализ «что-если», реализуемый с помощью функции Writeback. Данный анализ позволяет производить пересчёт непосредственно в сводной таблице Excel, осуществляется это внесением и публикацией изменений в ячейках таблицы. Лучшие варианты расчётов можно сохранять и использовать для дальнейшего анализа.

Однако чрезмерное увеличение количества мер, доступных для обратной записи, может привести к катастрофической деградации производительности. Одним из преимуществ многомерных баз данных является скорость получения агрегированных данных, в основном, за счёт вычисления агрегатов на этапе процессинга DWH. Однако при проведении расчётов в вычисляемых мерах и запросах это преимущество может быть утеряно, так как при выполнении запросов дизайны агрегаций использоваться уже не будут и вычисления будут производиться во время выполнения запроса. Таким образом, меры, зависящие от мер обратной записи, уже не удастся реализовать как физические, что влечёт за собой создание вычисляемых мер и отказ от использования дизайна агрегаций во время процессинга. В зависимости от слож-

ности последующих вычислений на основе групп мер обратной записи и детальности запрашиваемых срезов производительность запросов может быть снижена в десятки раз.

Для исключения подобных ситуаций были проведены достаточно длительные работы по оптимизации производительности решения: пересмотр целесообразности обратной записи для групп мер, ограничения безопасности на обратную запись, изменения дизайна многомерной модели, дизайнов агрегаций, вычисляемых мер, кэширования вычислений и т.д. В результате проведенных работ время вычислений при обновлении данных (например темпов падения) составило несколько секунд против часов в решении на основе книг Excel. Кроме того, существенное преимущество в оперативности получения результата достигается даже без проведения оптимизаций производительности - десятки секунд против часов. Тем не менее авторы пришли к заключению, что такая задержка (десятки секунд) является недопустимой и приняли решение о проведении оптимизации.

Анализ статистических данных

Исторически подготовка исходных данных для расчётов уровней добычи нефти проводится в файлах Excel, что в условиях большого фонда скважин существенно ограничивает возможности анализа фактических данных, проведения многовариантных расчётов и контроля их качества.

Подготовка данных включает в себя анализ:

- фактических данных и построение прогнозных темпов падения базовой добычи жидкости и нефти (в т.ч. с разделением на основную и неустановившуюся базу);
- фактических данных и построение прогнозных темпов падения добычи жидкости и нефти по различным типам ГТМ;
- прогноз запускных параметров для различных типов ГТМ в зависимости от фактических данных, свойств пластов и стадии разработки месторождения;

- динамики фонда скважин (количество выбытий/прибытий, остановочные/запускные, средние дебиты жидкости и нефти по скважинам).

При подготовке данных сотрудник вручную, на основе отчётности добывающего общества и выгрузок из месячного эксплуатационного рапорта, анализирует статистические данные по падению добычи жидкости и нефти, а также по выбытию из эксплуатации фонда скважин.

В связи с трудоемкостью процесса, мониторинг падения добычи нефти и жидкости проводится на основе данных за один год, и выглядит как «пила» (рисунок 2).

В результате полученные темпы падения добычи жидкости и нефти используются при последующем планировании «пирогом», но в долгосрочной перспективе - уже на второй год, они дают некорректный прогноз.

Наиболее наглядно это видно при планировании добычи нефти от ГТМ (рисунок 3). Анализ фактических данных показал, что после первого полугодия падение дебита нефти лучше описывается не степенной, а экспоненциальной зависимостью.

За счет использования OLAP-куба появилась возможность проводить более глубокий анализ; например, при дизайне аналитического решения предусмотрена возможность мониторинга эффекта от проведенных ГТМ (расчет темпов падения прироста дебита жидкости и нефти) за весь период времени, когда имеется эффект, даже за рамками календарного года. Также возможно проводить расчет фактического темпа падения сквозной базовой добычи (по жидкости и по нефти) с заданного года по текущий момент.

Использование средств визуализации (интерактивных панелей мониторинга - дашбордов) дает возможность динамически разворачивать любой элемент отчёта и просматривать всплывающие подсказки с дополнительной информацией. Достоинством данного средства являются простота, динамичность, гибкость, наглядность и возможность публикации отчёта на сервере для всеобщего доступа.

Рисунок 2. Профиль добычи нефти

Рисунок 3. Функциональные зависимости снижения эффекта от проведения на скважинах геолого-технологических мероприятий

Выводы

1. Технология BI позволяет производить автоматический контроль вводимых параметров на этапе ETL во время трансформации и загрузки данных в модель. При раз-

работке SSIS-пакетов возможно использовать условия для проверки их достоверности, например наличие и корректность запускных дебитов добычи и темпов падения нефти для новых скважин [7].

2. Контроль конечных результатов, в частности степень выработки запасов нефти, осуществляется в многомерном кубе с помощью создания дополнительных вычислительных мер, например отклонение от БП, выработка запасов нефти. Также с помощью средств визуализации – дашбордов либо отчётов в Power BI (средство Self-BI от Microsoft) и регламентных отчётов возможно *визуализировать для ускорения процесса проверки результатов:*

- рейтинг отклонений прогнозных параметров добычи нефти от фактических;
- рейтинг кустов бурения, пластов или месторождений по технологическим параметрам (например накопленная добыча на скважину);
- заданные в расчете параметры выбытия скважин из эксплуатации, их сравнение с историческими данными;
- заданное в расчете количество ГТМ и запускные приросты и их сравнение с историческими данными.

3. Использование BI-решения позволяет добиться *оптимальных результатов* по следующим направлениям:

- получение фактических данных по добыче нефти в разрезе типов базовой добычи и приростов ГТМ, разделения приростов ГТМ по типам;
- анализ добычи нефти можно проводить как по месторождениям, так и по объектам разработки, по площадным объектам, либо сравнивать добычу нефти различными добывающими обществами;
- вариативность деления базовой добычи нефти на основную и неустановившуюся, без проведения на скважинах ГТМ;
- возможность получать результат с учётом и без учёта выбытия скважин из действующего фонда;

- возможность проводить анализ падения уровня добычи нефти в зависимости от стадии разработки месторождения, наличия системы поддержания пластового давления (ППД) и др.;
- прогноз перевода добывающих скважин в систему ППД, их выбытий по причине неудавшихся операций ГТМ, аварий, простоев и перевода добывающих скважин в консервацию (ликвидацию);
- возможность прогнозирования уровня добычи нефти с учётом водонефтяного фактора по конкретному фонду скважин.

4. Важным преимуществом BI-решения при планировании добычи нефти *для многопластовых месторождений является возможность контролировать остаточные запасы нефти*, что актуально при планировании добычи за горизонтом 5-летнего бизнес-плана.

5. Основной эффект от применения BI-решения заключается *в повышении оперативности анализа фактических данных и подготовки исходных данных для расчетов, снижении трудозатрат на расчёты уровней добычи и подготовку отчётов, проверку результатов*. За счет повышения производительности появляется возможность повышения вариативности и детализации расчетов (до объектов разработки). За счёт гибкости настройки и использования аналитической базы данных возможно добавлять новые контролируемые параметры разработки месторождений, детализировать расчёты по любым требуемым срезам, производить расчёты по различным вероятностным сценариям и др.

Как следствие, предлагаемый инструмент прогнозирования *позволяет получать более точные и вариативные прогнозы добычи нефти* и производить детальный анализ полученных результатов.

Список литературы

1. Kimball R., Ross M. The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. 3rd Ed. 2013. 601 p.
2. Harinath S., Zare R., Meenakshisundaram S., Carroll M., Guang-Yeu Lee D. Professional Microsoft SQL Server Analysis Services 2008 with MDX. 2009. 1017 p.

References

1. Kimball R., Ross M. The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. 3rd Ed. 2013. 601 p.
2. Harinath S., Zare R., Meenakshisundaram S., Carroll M., Guang-Yeu Lee D. Professional Microsoft SQL Server Analysis Services 2008 with MDX. 2009. 1017 p.

3. Ferrari A., Webb C., Russo M. *Expert Cube Development with Microsoft SQL Server 2008 Analysis Services*. 2009. 362 p.

4. Piasevoli T. *MDX with Microsoft SQL Server 2008 R2 Analysis Services Cookbook*. 2011. 480 p.

5. Caserta J., Kimball R. *The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*. 2004. 528 p.

6. Cote C., Lah M., Sarka D. *SQL Server 2017 Integration Services Cookbook*. 2017. 558 p.

7. Leonard A., Masson M., Moss J.M., Ufford M. *SQL Server 2012 Integration Services Design Patterns (Expert's Voice in SQL Server)* 1st ed. 2012. 423 p.

3. Ferrari A., Webb C., Russo M. *Expert Cube Development with Microsoft SQL Server 2008 Analysis Services*. 2009. 362 p.

4. Piasevoli T. *MDX with Microsoft SQL Server 2008 R2 Analysis Services Cookbook*. 2011. 480 p.

5. Caserta J., Kimball R. *The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*. 2004. 528 p.

6. Cote C., Lah M., Sarka D. *SQL Server 2017 Integration Services Cookbook*. 2017. 558 p.

7. Leonard A., Masson M., Moss J.M., Ufford M. *SQL Server 2012 Integration Services Design Patterns (Expert's Voice in SQL Server)*. 1st ed. 2012. 423 p.

Авторы

• Бондаренко Константин Анатольевич
ООО «РН-УфаНИПИнефть»
Эксперт
Блок технологий и исследований
Российская Федерация, 450103, г. Уфа,
ул. Бехтерева, д. 3/1
e-mail: BondarenkoKA@ufanipi.ru

• Мосякова Ирина Александровна
ООО «РН-УфаНИПИнефть»
Ведущий специалист
Департамента развития систем проектирования
Российская Федерация, 450103, г. Уфа,
ул. Бехтерева, д. 3/1
e-mail: MosyakovaIA@ufanipi.ru

• Чермянин Павел Игоревич
ООО «РН-УфаНИПИнефть»
И.о. директора департамента
развития систем мониторинга
Российская Федерация, 450103, г. Уфа,
ул. Бехтерева, д. 3/1
e-mail: ChermyaninPI@ufanipi.ru

The Authors

• Bondarenko Konstantin A.
RN-UfaNIPIneft LLC
Expert
Block of Technologies and Research
3/1, Bekhterev str., Ufa, 450103,
Russian Federation
e-mail: BondarenkoKA@ufanipi.ru

• Mosyakova Irina A.
RN-UfaNIPIneft LLC
Leading Specialist
of Design Systems Development Department
3/1, Bekhterev str., Ufa, 450103,
Russian Federation
e-mail: MosyakovaIA@ufanipi.ru

• Chermyanin Pavel I.,
RN-UfaNIPIneft LLC
Acting director of the Monitoring Systems
Development Department
3/1, Bekhterev str., Ufa, 450103,
Russian Federation
e-mail: ChermyaninPI@ufanipi.ru