

**УДК 004.6**

## **АНАЛИЗ ВОЗМОЖНОСТЕЙ ПЕРВИЧНОЙ ОБРАБОТКИ ДАННЫХ ПРИ ПОМОЩИ АНАЛИТИЧЕСКИХ СИСТЕМ**

**Гец О.В.** (Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

**Научный руководитель – к. техн. н. Иванов С.В.**

(Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

В работе рассматриваются особенности средств и методов Business Intelligence для применения в нефтедобывающих компаниях с вертикально-интегрированной иерархией. Выделены задачи, которые решаются при помощи обработки данных в данной отрасли. Предложены рекомендации по дальнейшему изучению при помощи методов и моделей математической статистики.

В современных компаниях, в том числе нефтяной отрасли, возрастает роль информационных технологий, а также набирает популярность направление цифровой трансформации. На уровне аналитических подразделений компаний для целей обработки данных высокое значение приобрели системы Business Intelligence. Несмотря на обилие сфер применения BI-систем, нефтедобывающая отрасль слабо рассматривается с точки зрения применения подхода Business Intelligence в научных источниках.

Перенос ETL-процессов на BI-платформы позволит искать скрытую информацию, строить предсказательную аналитику и проводить перекрестный анализ информации из совершенно несовместимых источников данных путем создания универсальной модели.

Использование стандартных средств ETL-процессов приводит к ряду проблем, к числу которых можно отнести разнородность форматов информационных систем-источников, сложность в разработке алгоритмов для создания «глобальной» формы данных, поддерживаемой в хранилище для создания запросов от бизнес-пользователей.

Целью работы выступает формирование подхода к движению потоков данных на нефтедобывающих предприятиях от систем-источников к бизнес-пользователям на основе методов Business Intelligence, к числу задач необходимо отнести анализ методов сбора и обработки данных в предметной области.

Для нефтедобывающей отрасли характерной особенностью является ухудшение структуры запасов углеводородов на месторождениях, находящихся в поздней стадии разработки. Соответственно, необходимо применять методы регулирования для получения оценки по ранее выполненным фактическим работам и выработать рекомендации по повышению их эффективности в планируемом периоде в разрезе месяца, квартала или года.

Использование KPI в вертикально-интегрированных нефтяных компаниях является одним из наиболее важных критериев эффективности работы, поскольку служат индикаторами соответствия уровня текущих показателей деятельности стратегическим целям организации.

Источниками данных являются внутренние информационные системы учета и хранения, а также любые табличные данные.

Важно отметить, что BI-системы не всегда имеют прямой доступ к базам данных. Использование представлений для выгрузок, а не excel-файлов, является более предпочтительным, поскольку вероятность возникновения ошибок ручного ввода, нарушение структуры исходных данных более низкая.

В рамках исследования реализован метод циклических переменных. Некоторые данные из базы необходимо выгружать как целиком в виде справочников, в которых нет полей, отвечающих за дату загрузки и обновления, так и периодически. В последнем случае используются циклические переменные: задается начало и конец временного интервала (чаще

всего ими выступают начало и конец предыдущего месяца, предшествующего отчетному периоду для расчета KPI).

Помимо полных и периодических данных, возможно использовать метод инкрементальной загрузки, что является удобным решением для выгрузок из данных по факту/плану. Разные дочерние организации вносят данные в свои показатели в разном организационном порядке и в разные временные рамки по различным месторождениям, что приводит к частой перезагрузке слоев приложений и, соответственно, увеличенной нагрузке на сервер. В таком случае необходимо загружать только измененные, либо новые записи из базы или хранилища данных.

Одной из особенностей данных в любой организации, в том числе и в нефтегазовой отрасли, является задание абонентов (в данном случае – месторождений) в виде родительско-дочерней иерархии с разным количеством уровней. Использование иерархических методов также важно использовать при создании BI-системы в данной сфере.

Применение BI-систем позволит сократить время обработки запроса и оптимизировать работу гибкой отчетности, в том числе при наличии каких-либо изменений в исходных базах данных.

В качестве результата проведенного исследования можно выделить, что использование систем BI позволит на каждом этапе ETL-процесса контролировать характер извлекаемых данных, использовать различные сценарии их обработки в зависимости от типа системы-источника, проводить анализ накопленной информации и строить прогноз изменения ключевых показателей.

Таким образом, в ходе исследования решалась задача сокращения количества-возможных ошибок при построении единой системы данных из различных источников, а также соответствия исходного запроса пользователя к получаемым им данным.

При этом важно понимать, что полученные результаты можно расширить при применении методов математической статистики, а также интеграцией с различными языками статистической обработки данных (например, R/Python).

Гец О.В. (автор)

Подпись

Иванов С.В. (научный руководитель)

Подпись