

## АРХИТЕКТУРА АНАЛИТИЧЕСКОЙ ПЛАТФОРМЫ ДЛЯ ВСТРАИВАЕМЫХ ИНТЕЛЛЕКТУАЛЬНЫХ ПОДСИСТЕМ

В.С. АБАТУРОВ<sup>1</sup>, А.Ю. ДОРОГОВ<sup>2</sup>, О.В. ЗАБРОДИН, И.В. РАКОВ

*Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» 197376, Россия,  
Санкт-Петербург, улица Профессора Попова, дом 5  
<sup>1</sup>vasilianich@yandex.ru, <sup>2</sup>vaksa2006@yandex.ru*

В настоящей работе предлагается вариант архитектурного решения аналитической платформы для встраиваемых интеллектуальных подсистем на основе СУБД PostgreSQL. Архитектура аналитической платформы подчинена требованиям SQL/MM и PMML. Показана схема формирования сценариев основных фаз извлечения знаний: фазы обучения, фазы тестирования и прикладной фазы. Приведены преимущества представленного архитектурного решения.

*Ключевые слова:* аналитическая платформа, извлечение знаний, унифицированный интерфейс.

Современные системы управления технологическими процессами требуют обработки и анализа больших объемов информации. Данная проблема давно стала критической в областях, непосредственно связанных с аналитической обработкой данных (Data Mining, искусственный интеллект, системы поддержки принятия решений, техническое зрение, мультимедиа технологии, и др.). В настоящее время рынок аналитических систем экспоненциально развивается. В этом процессе принимают участие такие крупные зарубежные компании как: IBM Cognos, MicroStrategy, Oracle, SAS, Microsoft, а также российские фирмы BaseGroup Labs, «Прогноз» [1].

Существующие стандарты Data Mining затрагивают основных аспекты построения аналитических систем: унификация интерфейсов, посредством которых любое приложение может получить доступ к функциональности аналитической платформы; разработка единого соглашения по хранению и передаче моделей Data Mining.

Среди существующего многообразия стандартов (разработанными различными международными организациями и в разное время) согласованными является пара SQL/MM [2] и PMML (Predicted Model Markup Language) [3]. Проведенный анализ показал, что стандарт SQL/MM достаточен для разработки архитектуры аналитической платформы встроеной в реляционную базу данных. Стандарт PMML полностью покрывает представление моделей алгоритмов и, кроме того, последняя версия стандарта допускает использование PMML моделей для информационных целей.

В настоящей работе предлагается вариант архитектурного решения аналитической платформы (рис. 1, а) для встраиваемых интеллектуальных подсистем на основе СУБД PostgreSQL. Взаимодействие аналитической платформы с приложениями осуществляется с помощью трех независимых интерфейсов: SQL/MM – расширение языка SQL для управления процессами извлечения знаний; SOAP – протокол коммуникаций между интернет-приложениями; DDS – открытый стандарт распределенного сервиса для систем реального времени.

Ключевую роль в данной архитектуре выполняет стандарт PMML, который применяется не только для представления и хранения моделей знаний, но и для создания интерфейса сообщений с аналитической платформой через каналы удаленного доступа.

Предложенная архитектура позволяет с помощью хранимых процедур СУБД PostgreSQL выполнять классические этапы извлечения знаний, включающие: обучающую фазу, фазу тестирования и фазу применения (рис. 1, б).

Фаза обучения – этап, на котором строится вычислительная модель интеллектуального анализа данных.

Фаза тестирования – этап, на котором осуществляется проверка качества предсказания на основе построенной модели.

Прикладная фаза – этап, на котором строка оперативных данных оценивается на основе обученной модели.

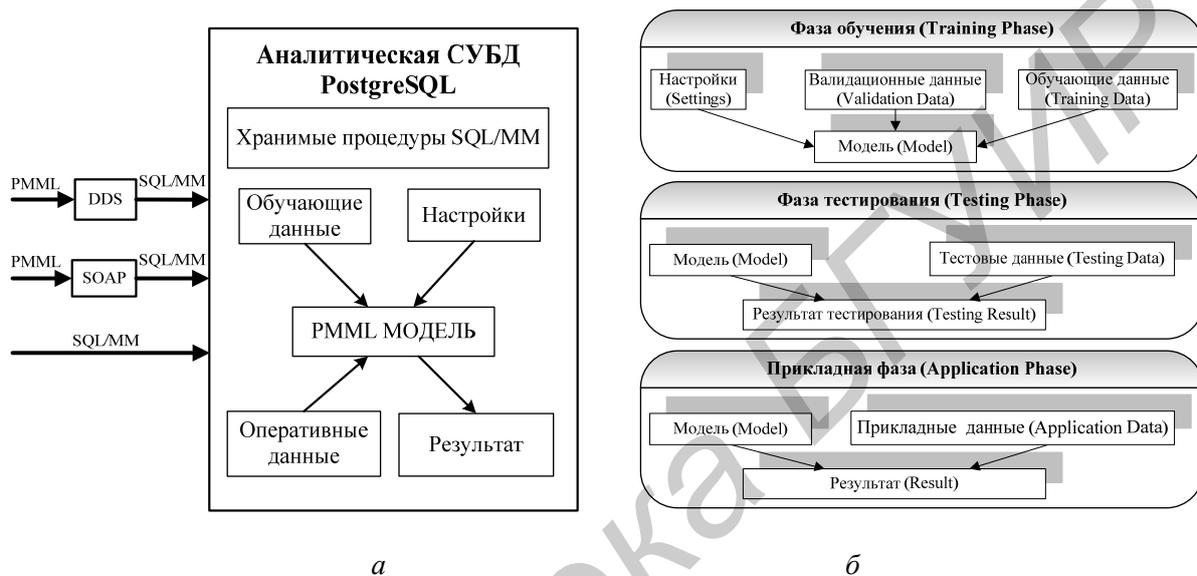


Рис. 1. Аналитическая платформа: а – архитектура аналитической платформы; б – фазы извлечения знаний

Поскольку в качестве аналитической СУБД была выбрана СУБД PostgreSQL, аналитическая платформа, реализованная в соответствии с предложенной архитектурой, будет иметь следующие тактико-технические характеристики:

- обработка масштабных массивов разнородной информации (до 32 TB);
- расширяемость и масштабируемость аналитики;
- многоплатформенность;
- контроль целостности данных;
- интерфейсы доступа к языкам программирования высокого уровня;
- триггеры и правила для управления процессами;
- система управления правами доступа и авторизации;
- параллельная обработка пользовательских сессий;
- шифрование трафика.

Преимуществами предложенной архитектуры аналитической платформы встроенной в базу данных является высокая гибкость применения алгоритмов извлечения знаний, а также простые возможности их расширения и масштабирования.

#### Список литературы

1. Data Mining Community. Top Resource [Электронный ресурс]. – Режим доступа: <http://www.kdnuggets.com>. Дата доступа: 23.01.14.
2. PMML Version 4.1, 2012, Data Mining Group (DMG) <http://www.dmg.org/>
3. ISO/IEC 13249-6-2006, SQL/MM Part 6.