**УДК 004.65, 004.41, 311.31**

**Басов И.Г., Шеремет Н.Н.**

**ПОСТРОЕНИЕ И ИСПОЛЬЗОВАНИЕ МОДЕЛИ ДАННЫХ ДЛЯ СИСТЕМ ЭКОНОМИЧЕСКИХ ПОКАЗАТЕЛЕЙ С ПРИМЕНЕНИЕМ НЕРЕЛЯЦИОННЫХ ПОДХОДОВ И OLAP-ТЕХНОЛОГИЙ**

**Аннотация**: *В статье рассмотрена проблема автоматизации процессов хранения и обработки данных системы экономических показателей, предложена модель представления экономических показателей в виде функций. Для реализации информационной системы предполагается использование ряда инструментов и технологий (OLAP, Apache Kylin, HBase).*

**Ключевые слова**: *система экономических показателей, параметризация экономического показателя, функция многих переменных, нереляционная распределенная база данных, распределенная платформа обработки данных, OLAP.*

***Введение.*** При оценке эффективности деятельности экономических субъектов, а также состояния отдельных отраслей экономики, важную роль играют значения определенных экономических показателей. Система экономических показателей представляет собой совокупность взаимосвязанных, систематизированных показателей, характеризующих экономику в целом, ее отрасль, регион, сферу экономической деятельности, группу однородных экономических процессов [1]. Состав и структура экономических показателей представляют один из значимых объектов изучения экономической науки.

Если в масштабах отдельного предприятия система может включать небольшое количество показателей и допускать непосредственный анализ специалистом без использования средств автоматизации, то при рассмотрении экономических показателей отрасли или экономики в целом система может иметь достаточно развернутую структуру с определенной иерархией и многочисленными связями, поэтому возникает необходимость в средствах автоматизированной обработки. Например, по состоянию на 1 августа 2000 года только Каталог статистических показателей Российской Федерации содержал около пятнадцати тысяч статистических показателей [2, раздел 1].

***1. Постановка задачи.*** Задача автоматизации процессов хранения и обработки данных в системе экономических показателей связана со следующими сложностями: большие объемы данных накладывают ограничения на используемые технологии хранения и способы обработки данных, требуют значительных вычислительных мощностей. Для решения этих проблем требуется создание универсального механизма хранения и работы с показателями и связями между ними, а также подбор технологий хранения и обработки данных с учетом их структуры и значительных объемов.

Требования к системе:

- разграничение прав доступа и безопасность данных;

- наличие интерфейса доступа к данным и функциям системы;

- оперативность выполнения запросов;

- гибкость;

- способность хранить и обрабатывать большие массивы данных;

- масштабируемость.

Под гибкостью в данном случае подразумевается способность без изменения структуры системы в целом и ее базы данных, в частности, обеспечивать возможность добавления новых показателей и их параметров, а также связей с другими показателями. Разрабатываемая система должна быть масштабируемой, чтобы обеспечить возможность  достижения необходимого уровня производительности.

***2. Модель данных.*** Одним из возможных подходов к организации данных является представление показателей в виде функций многих переменных. Значение такой функции - это значение показателя в условиях, соответствующих значениям параметров. Множество показателей динамично вследствие динамичности системы экономических показателей. Различие природы показателей обуславливает различие множеств их параметров. Параметризация показателей облегчает задачу задания связей между показателями и использование этих связей в дальнейших расчетах.

В качестве примера рассмотрим такой экономический показатель, как балансовая (валовая) прибыль предприятия. Пусть перед нами стоит задача сохранить значения этого показателя для всех предприятий республики за все отчетные периоды. Продолжительность отчетного периода - один месяц. При этом возникает целый ряд значений, таких как, например, показатель валовой прибыли предприятия А за август 2017 года или показатель валовой прибыли предприятия Б за апрель 2016 года. Очевидно, эти показатели отражают однородные данные, а их совокупность может быть использована для наблюдения по отрасли, к которой относятся эти предприятия, или в целом по республике. Таким образом, при рассмотрении валовой прибыли как одного из показателей, характеризующего экономику государства, необходимо снабдить его тремя параметрами: наименование предприятия, год и месяц, когда была произведена фиксация значения показателя. Таким образом, в данном примере конкретное значение показателя для некоторого предприятия в определенный год и месяц можно представить в виде точки в четырехмерном пространстве или значения функции трех переменных в точке, соответствующей данному предприятию, году и месяцу. Если зафиксировать значения двух параметров (предприятие и год), а значение оставшегося параметра (месяца) не будет задано, то результатом функции-показателя будет не одно значение, а одномерная совокупность (вектор из 12 значений показателя для заданного предприятия в заданный год). При свободных двух параметрах в результате получится двумерная совокупность (или матрица) значений и т.д. Например, нам нужно рассчитать балансовую прибыль предприятия А за 2017 год. Для этого нужно указать наименование предприятия А и год 2017, опустив месяц, значения элементов полученного вектора просуммировать.

Если кроме показателя балансовой прибыли предприятия добавить показатель “отчисления” с той же дискретизацией, то можно ввести расчетный показатель “чистая прибыль”, который будет равен разности показателей балансовой прибыли предприятия и отчислений предприятия и будет зависеть от тех же параметров: наименование предприятия, год и месяц.

***3. Технологии и особенности реализации.*** Динамичность системы экономических показателей и различия между множествами параметров показателей затрудняют хранение и обработку данных такой структуры в реляционной базе данных: возникает выбор между гибкостью системы и ее способностью обрабатывать большие объемы данных. Для обеспечения гибкости системы в условиях, когда множество показателей не определено полностью, необходимо в структуру базы данных на этапе проектирования закладывать возможность расширения множества показателей и их параметров, что вынуждает использовать специализированные нереляционные СУБД или реляционные СУБД, но с применением шаблона проектирования “open schema” [3] или таких специфических особенностей СУБД, как тип данных JSONB в PostgreSQL. Использование в реляционной СУБД перечисленных возможностей является отступлением от реляционной модели и замедляет обработку данных, что особенно критично при больших объемах. В такой ситуации индексирование не является решением, поскольку не способно в достаточной мере ускорить обработку данных при использовании шаблона проектирования "open schema" (из-за большого количества объединений таблиц, в том числе, рекурсивных). В случае применения индексирования по столбцам с типом данных JSONB в PostgreSQL скорость выборки также будет недостаточной, особенно во время операций записи, которые сопровождаются индексацией. Поэтому для решения поставленной задачи целесообразно использовать специализированную нереляционную СУБД, которая будет обладать необходимой гибкостью, отказоустойчивостью и масштабируемостью. Одной из таких СУБД является HBase - нереляционная распределенная система управления базами данных, разрабатываемая в рамках проекта Hadoop фонда Apache Software Foundation [4]. Данная СУБД не накладывает жестких ограничений на структуру хранимых данных и способна хранить и обрабатывать большие объемы данных.

Для увеличения скорости обработки запросов к данным и проведения вычислений значений показателей целесообразно использовать технологию OLAP, которая применяется для подготовки агрегированной информации на основе больших массивов данных, структурированных по многомерному принципу [5]. В качестве платформы для анализа данных можно использовать систему Apache Kylin.

Apache Kylin - распределенная платформа обработки больших объемов данных, предоставляющая интерфейс поддержки SQL и OLAP для их интерактивного анализа, что обеспечит выполнение условия о необходимости наличия доступа к данным и оперативности выполнения запросов. Наличие API ODBC, JDBC, REST API и возможности интеграции с существующими инструментами бизнес анализа (например, Tableau, PowerBI/Excel, MSTR, QlikSense, Hue, SuperSet)[6] облегчит создание на основе этого продукта информационной системы, удовлетворяющей требованиям поставленной задачи.

Технология Apache Kylin используется китайской телекоммуникационной компанией ZTE как основа системы, предоставляющей возможности OLAP в промышленности, банковской, финансовой и других сферах, компанией Meizu как часть статистической аналитической платформы и другими компаниями [7].

***Заключение.*** Таким образом, в работе предложена модель представления экономических показателей в виде функций, определенных на некотором множестве в многомерном пространстве, каждое измерение которого соответствует некоторому параметру. В качестве хранилища данных была выбрана СУБД HBase как наиболее полно соответствующая требованиям к системе и построенной модели. Применение платформы Apache Kylin упростит и ускорит обработку больших массивов данных и предоставит интерфейсы поддержки SQL и OLAP для интерактивного анализа.

**Библиографический список**

1) Борисов Е.Ф., Петров А.А., Березкина Т.Е. Экономика. 2-е издание. Учебник для бакалавров

2) Постановление Госкомстата РФ от 30.11.2000 N 118 "Об утверждении "Методологических положений по формированию и ведению Каталога статистических показателей (КСП)"

3) Свободная энциклопедия “Wikipedia” [Электронный ресурс: https://en.wikipedia.org/wiki/Entity-attribute-value\_model (дата обращения: 14.05.2018)]

4) Официальный сайт проекта “Apache HBase” [Электронный ресурс: http://hbase.apache.org/ (дата обращения: 14.05.2018)]

5) Интернет-проект «Корпоративный менеджмент» [Электронный ресурс:<https://www.cfin.ru/software/hyperion/essbase/what_is_olap.shtml> (дата обращения: 14.05.2018)]

6) Портал Национальной электронной библиотеки им. Н. Э. Баумана [Электронный ресурс:<https://ru.bmstu.wiki/Apache_Kylin> (дата обращения: 14.05.2018)]

7) Официальный сайт проекта “Apache Kylin” [Электронный ресурс:<http://kylin.apache.org/community/poweredby.html> (дата обращения: 14.05.2018)]

**Информация об авторах**

Басов Игорь Геннадьевич (Донецкая Народная Республика, Донецк) - младший научный сотрудник, ГУ "Институт экономических исследований" (Донецкая Народная Республика, г. Донецк, ул. Университетская, 77, office@econri.org).

Шеремет Николай Николаевич (Донецкая Народная Республика, Донецк) - младший научный сотрудник, ГУ "Институт экономических исследований" (Донецкая Народная Республика, г. Донецк, ул. Университетская, 77, office@econri.org)

**Basov I.G., Sheremet N.N.**

**DEVELOPMENT AND APPLICATION OF THE DATA MODEL FOR SYSTEMS OF ECONOMIC INDICATORS WITH APPLICATION OF NOSQL APPROACHES AND OLAP-TECHNOLOGY**

**Abstract**: *This paper discusses the problem of automating the processes of storing and processing data of the system of economic indicators, proposes a model for representing economic indicators in the form of functions. To implement the information system, it is expected to use a number of modern tools and technologies (OLAP, Apache Kylin, Apache HBase).*

**Keywords**: *system of economic indicators, parametrization of the economic indicator, a function of many variables,  non-relational distributed database, distributed analytics engine, OLAP.*

**About authors**

Basov Igor Gennadievich (Donetsk People's Republic, Donetsk) - junior researcher, SI "Economic Research Institute" (Donetsk People's Republic, Donetsk, Universitetskaya str., 77, office@econri.org).

Sheremet Nikolay Nikolaevich (Donetsk People's Republic, Donetsk) - junior researcher, SI "Economic Research Institute" (Donetsk People's Republic, Donetsk, Universitetskaya str., 77, office@econri.org)

**Bibliographic list**

1) Borisov E.F., Petrov A.A., Berezkina T.E. Economy. 2nd edition. A Textbook for Bachelors

2) Decree of the State Statistics Committee of the Russian Federation of November 30, 2000 N 118 "On approval of the" Methodological provisions for the formation and maintenance of the Catalog of Statistical Indicators (CSI) "

3) Free encyclopedia "Wikipedia" [Electronic resource: https://en.wikipedia.org/wiki/Entity-attribute-value\_model (date of circulation: 14.05.2018)]

4) The official website of the project "Apache HBase" [Electronic resource: http://hbase.apache.org/ (date of circulation: 14.05.2018)]]

5) Internet project "Corporate Management" [Electronic resource: https://www.cfin.ru/software/hyperion/essbase/what\_is\_olap.shtml (date of circulation: 14.05.2018)]

6) Portal of the National Electronic Library. NE Bauman [Electronic resource: https://ru.bmstu.wiki/Apache\_Kylin (date of circulation: 14.05.2018)]

7) The official website of the project "Apache Kylin" [Electronic resource: http://kylin.apache.org/community/poweredby.html (date of circulation: 14.05.2018)]