УДК 004.01

Степанова Т.А.
студент

Кацупеев А.А.
канд. тех. наук, доцент

Южно-Российский государственный политехнический университет (НПИ) имени М.И. Платова
Россия, г. Новочеркасск

СРАВНЕНИЕ ПРОИЗВОДИТЕЛЬНОСТИ СУБД В ПРОЦЕССЕ ПРОЕКТИРОВАНИЯ СППР «ЭЛЕКТРОННАЯ ЭКСПЛУАТАЦИОННАЯ ДОКУМЕНТАЦИЯ»

Системы поддержки принятия решений очень востребованы на крупных предприятиях, где речь идет о работе с большими объемами данных. Примером может служить широкое их развитие в таких отраслях, как приборостроение, авиационная и космическая промышленность. В связи с этим существует необходимость помогать аналитикам данных с сортировкой и анализом информации, и разработка системы поддержки принятия решений в данном случае позволяет повысить продуктивность сотрудников, актуализировать и обеспечить оперативное хранение документации и упрощенный доступ к ней. В статье же рассмотрены основные этапы проектирования системы, а также проведена сравнительная оценка скорости выполнения основных функциональных операций на базе СУБД PostgreSQL 9.6 и MySQL 5.1.

Ключевые слова: *информационная система, система поддержки принятия решений, эксплуатационная документация, контекстная диаграмма, декомпозиция, СУБД, схема функционирования, анализ данных, производительность процессора, хранение данных, экспертная система, функциональные операции.*

Данная система предназначена для анализа данных эксплуатационной документации (нетиповых инструкций), а также анализа действий оператора в соответствии с этими инструкциями. Данными для анализа могут являться: информация о сценариях СИППР, их отправке и получении, а также типе сценария, расположении документации и т.д.

Во время анализа предметной области было выполнено описание бизнес-процессов, представленное в виде диаграмм нотации IDEF0 [1]. На рисунке 1 показана контекстная диа-

грамма системы поддержки принятия решений, которая далее детализируется, т.е. осуществляется ее разбиение на диаграммы декомпозиции.

Рис. 1. Контекстная диаграмма

В декомпозиции контекстной диаграммы можно выделить следующие процессы: сбор информации (ввод в базу эксплуатационных данных), анализ данных (выбор метода анализа, обработка данных и вывод результатов), формирование экспертного решения, принятие решения (предполагается участие не только системы, но и пользователя, задействованного в принятии решения) и формирование отчёта (вместе с добавлением дополнительной информации и комментариев эксперта к результатам работы) [5].

Диаграмма декомпозиции блока «Сбор информации» показана на рисунке 2.

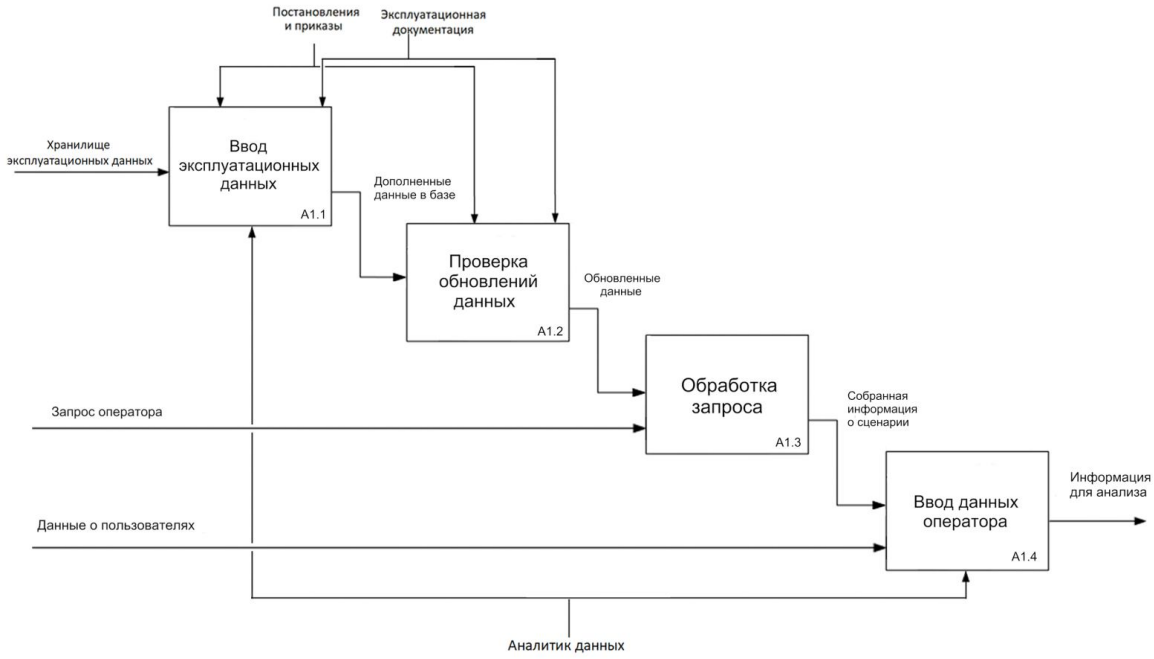

Рис. 2. Диаграмма декомпозиции блока «Сбор информации»

Диаграмма декомпозиции блока «Анализ данных» показана на рисунке 3.

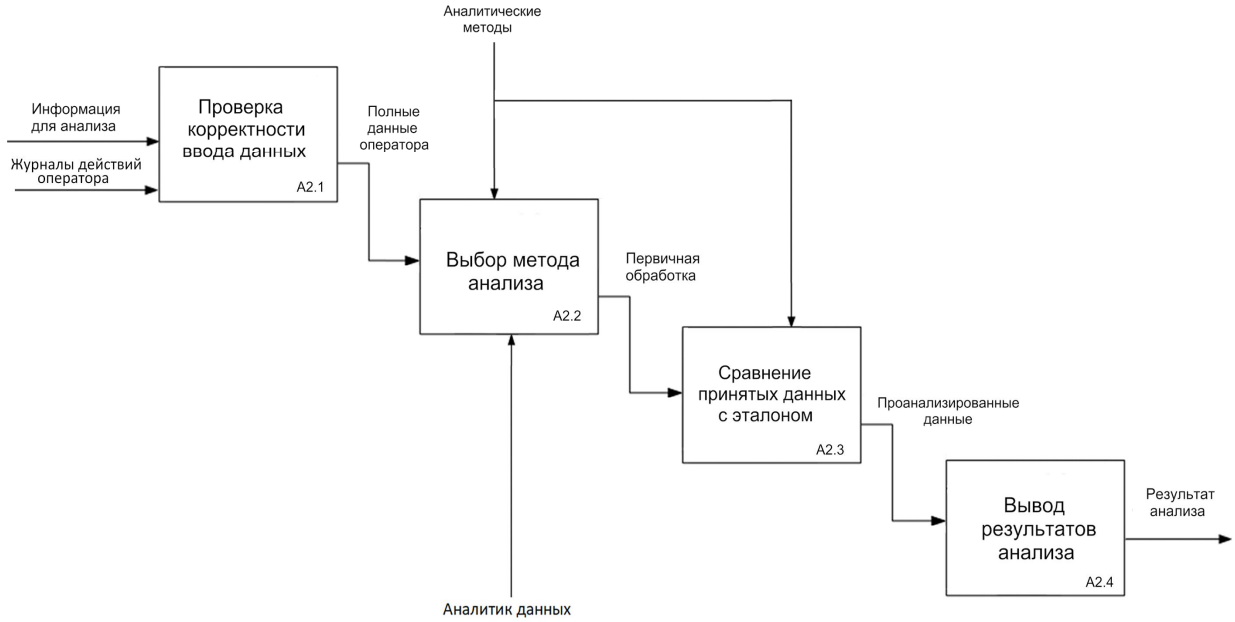

Рис. 3. Диаграмма декомпозиции блока «Анализ данных»

Диаграмма декомпозиции блока «Принятие решения» показана на рисунке 4.

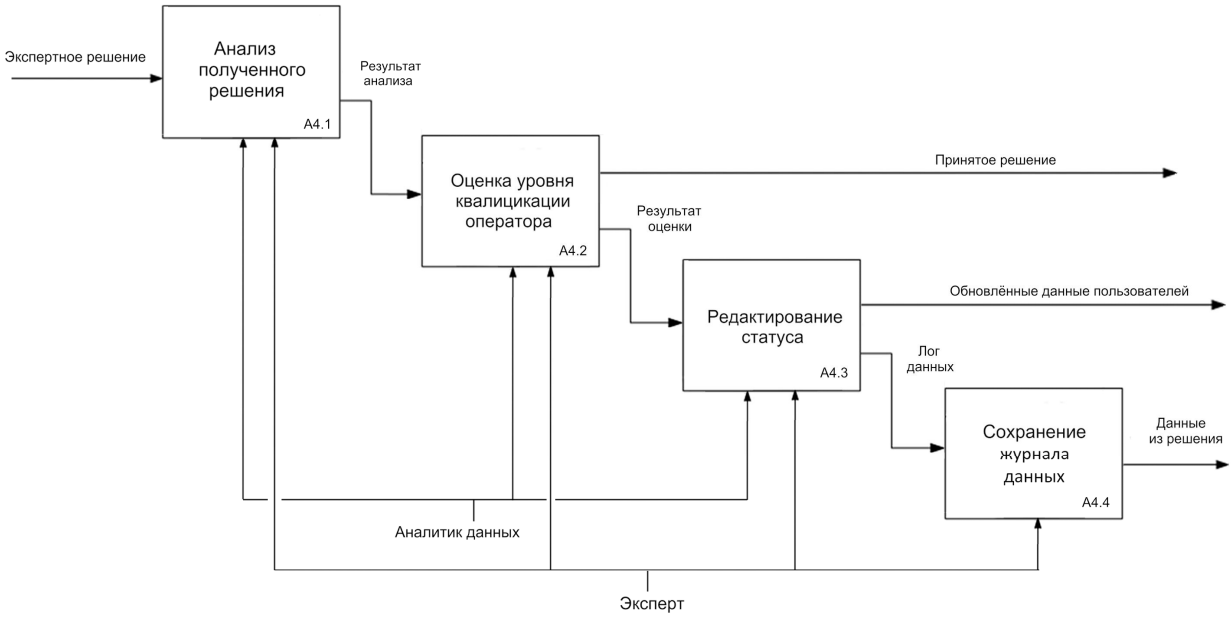

Рис. 4. Диаграмма декомпозиции блока «Принятие решения»

Диаграмма декомпозиции блока «Формирование отчёта» показана на рисунке 5.

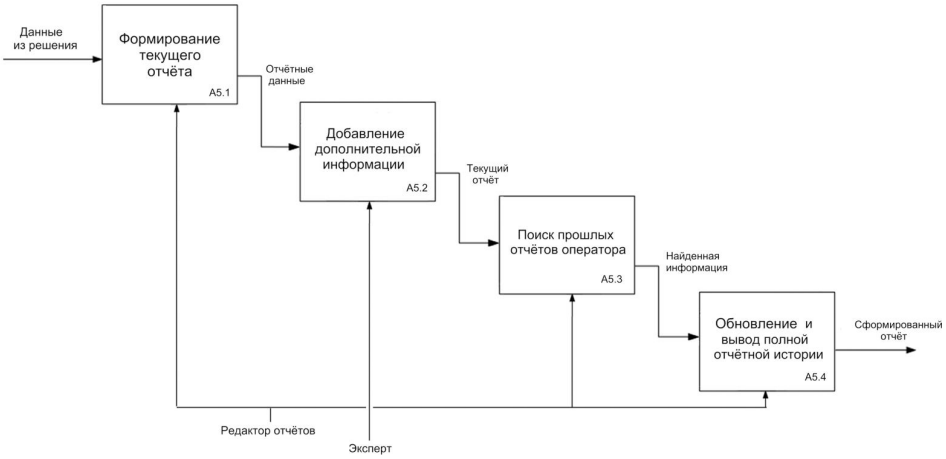

Рис. 5. Диаграмма декомпозиции блока «Формирование отчёта»

Главными критериями при выборе архитектуры построения ИС являются скорость ее действия, надежность, масштабируемость и безопасность [2]. Учитывая это, для реализации данной системы была выбрана двухуровневая архитектура «клиент-сервер».

На сервере располагается база данных, хранящая информацию об эксплуатационной документации, а на клиентской части находится инструментарий аналитика данных, позволяющий извлекать нужную информацию из хранилища, формировать отчёты и обновлять актуальную информацию [6].

Схема функционирования СИППР эксплуатационной документации представлена на рисунке 6.

Рис. 6. Схема функционирования

В общем случае, минимум компонентов необходимых для функционирования ИС:

– сервер – управляет работой всей системы и располагает в себе СУБД;

– СУБД – осуществляет ввод и обработку данных эксплуатационной документации и разных транзакций;

– хранилище данных – осуществляет хранение больших объемов данных, подлежащих анализу;

– клиент – получает информацию из БД и принимает решение согласно полученным данным [3].

При выборе подходящей СУБД важно учитывать скорость выполнения основных функциональных операций системы, поскольку речь идет о больших массивах данных [4]. Следует сформулировать критерии выбора и сравнить наиболее приемлемые и распространенные в качестве вариантов на реализацию в системе СУБД, из которых были взяты для оценки СУБД PostgreSQL 9.6 и MySQL 5.1 [7].

В качестве выходного параметра Y используется скорость выполнения основных функций пользователя до получения искомого аналитического результата. То есть рассматривается время обработки данных сервером от клиента, формулирование и передача запроса к базе данных, обработка результата запроса с помощью определенного алгоритма, а также формирование ответа и его передача пользователю.

Объектом изучения служит производительность сервера и разработанной информационной системы. Время на передачу данных по сети считается «узким» местом с точки зрения производительности, поэтому его можно не учитывать. В качестве оцениваемых функций реализуемой СППР были выбраны следующие:

- изменение значения одного исходного параметра (y_1, мс);
- анализ входящих параметрических данных и группировка информации; при расчете были учтены значения пяти классов пользователей (y_2, мс);
- автоматический расчет полученных в результате анализа значений; при расчете учитывается усредненные данные нескольких попыток (y_3, мс);
- автоматическое редактирование учетных записей пользователей в зависимости от результатов сценария. Проверка проводилась на основе двадцати тестовых записей (y_4, мс);
- визуализация результатов расчета с использованием встроенных графических элементов (y_5 , мс).

Варьируемые параметры:

- производительность серверного процессора – x_1;
- объем оперативной памяти сервера – x_2;
- пропускная способность памяти – x_3;
- производительность файловой системы – x_4.

Однако иногда вычислить значения некоторых из этих параметров может быть затруднительно. Например, для того, чтобы оценить скорость процессора, не может учитываться лишь частота, на которой он работает. Важную роль играет архитектура процессора, набор архитектурных инструкций, имеется ли кэш данных и т.д. Если говорить о пропускной способности памяти, то здесь важна не только частота работы или способность этого модуля работать в паре с другим. Фирма-изготовитель регулирует спецификацию и качество изготовления материнской платы, что может также влиять на частоту. В свою очередь, производительность файловой системы зависит от ее фрагментированности, типа, размера кэша, а также от того, имеются ли в системе функционирующие антивирусы и т.д. [8].

Учитывая перечисленные проблемы, оценка производительности процессора сервера, системы файлов и пропускной способности памяти требует использования индекса производительности. Его можно вычислить в тестировании при помощи утилиты Sandra Professional от фирмы SiSofhvare.

При уровне значимости а = 0,05 вычисляется табличное значение критерия Кохрена, чтобы, при сравнении с ним, оценить разброс полученных экспериментально значений. Оно равно 0,2658, а опытные значения критерия Кохрена показаны в таблице 1.

Таблица 1

Значения критериев Кохрена

Функциональные операции / СУБД	y_1	y_2	y_1	y_1	y_1
PostgreSQL 9.6	0,143	0,204	0,2505	0,1251	0,1769
MySQL 5.1	0,2371	0,2108	0,1964	0,2428	0,2085

Полученные экспериментальные значения критерия не превышают табличного, значит можно судить об однородности дисперсий и относительной точности выборки данных, и после обработки результатов опытов получаем следующие уравнения регрессии:

- для системы на базе СУБД PostgreSQL 9.6:

$$y_1 = 1383 - 566\,x_1 - 43x_2 - 150x_3 - 269x_4;$$
$$y_2 = 1659 - 547x_1 - 98x_2 - 164x_3 - 436x_4;$$
$$y_3 = 1197 - 341x_1 - 16x_2 - 149x_3 - 179x_4;$$
$$y_4 = 10409 - 4243x_1 - 341x_2 - 1633x_3 - 1219x_4;$$
$$y_5 = 14238 - 6073x_1 - 312x_2 - 1221x_3 - 985x_4.$$

- для системы на базе СУБД MySQL 5.1:

$$y_1 = 1851 - 991x_1 - 13x_2 - 306x_3 - 141x_4;$$
$$y_2 = 2913 - 1471x_1 - 22x_2 - 526x_3 - 664x_4;$$
$$y_3 = 1771 - 754x_1 + 28x_2 - 371x_3 - 219x_4;$$
$$y_4 = 13672 - 7469x_1 - 46x_2 - 2938x_3 - 1579x_4;$$
$$y_5 = 17256 - 8465x_1 - 25x_2 - 1861x_3 - 1352x_4.$$

Расчетные значения критерия Стьюдента, необходимые для проверки значимости факторов, приведены в таблице 2.

Таблица 2

Значения t-критерия Стьюдента

		t_0	t_1	t_2	t_3	t_4
PostgreSQL 9.6	1	24,27	9,32	3,79	1,98	4,72
	2	23,78	8,70	3,28	0,31	6,19
	3	16,69	4,88	3,24	0,28	1,57
	4	17,43	7,47	2,57	0,54	2,02
	5	23,51	10,04	2,51	0,32	3,58
MySQL 5.1	1	19,68	10,65	3,03	1,27	3,51
	2	16,59	8,23	3,13	0,46	1,64
	3	18,54	8,32	2,31	0,09	2,43
	4	18,36	9,81	3,06	0,75	1,09
	5	19,93	8,79	2,03	0,92	2,40

Для принятого выше уровня значимости, ниже табличного значения оказался критерий t_3 в большинстве уравнений, что позволяет судить об отсутствии значимости фактора «пропускная способность памяти», а также t_4 во втором и четвертом уравнениях для MySQL и t_4 в третьем уравнении для PostgreSQL, что снижает влияние фактора «производительность файловой системы».

Для информационной системы, построенной на базе СУБД PostgreSQL 9.6, уравнения регрессии приняли следующий вид:

$$y_1 = 1383 - 566\,x_1 - 43x_2 - 150x_3 - 269x_4;$$
$$y_2 = 1659 - 547x_1 - 98x_2 - 436x_4;$$
$$y_3 = 1197 - 341x_1 - 16x_2;$$
$$y_4 = 10409 - 4243x_1 - 341x_2 - 1219x_4;$$
$$y_5 = 14238 - 6073x_1 - 312x_2 - 985x_4.$$

Для СУБД MySQL 5.1:

$$y_1 = 1851 - 991x_1 - 13x_2 - 141x_4;$$
$$y_2 = 2913 - 1471x_1 - 22x_2;$$
$$y_3 = 1771 - 754x_1 + 28x_2 - 219x_4;$$
$$y_4 = 13672 - 7469x_1 - 46x_2;$$
$$y_5 = 17256 - 8465x_1 - 25x_2 - 1352x_4.$$

В результате проведенных расчетов можно сделать следующие выводы:

1. PostgreSQL 9.6 имеет относительно лучшие значения по сравнению с MySQL 5.1, если использовать эти СУБД для серверного хранения данных. В среднем, скорость выполнения основных функций для СУБД PostgreSQL 9.6 была в полтора раза больше. Таким образом, для данной системы он более предпочтителен.

2. Время работы над данными сильно зависит от скорости процессора, поскольку реализованные функции предполагают сложную оценку исходных данных. Поэтому процесс выбора аппаратного обеспечения, используемого для сервера приложений, нуждается в особом внимании относительно производительности сервера.

3. Фактор «пропускная способность памяти» не был значим для обработки функций, возможно потому, что в процессе работы её уровень оказался достаточным. Значение производительности файловой системы также оказалось малозначимо, и причины здесь могут быть схожими.

СПИСОК ЛИТЕРАТУРЫ

1. Знакомство с нотацией IDEF0 и пример использования [электрон. текстовые данные]. – Режим доступа: https://habr.com/company/trinion/blog/322832/.

2. Панфилов А.Н., Зуев В.А., Скоба А.Н. Методы и средства проектирования информационных систем и технологий / Юж.-Рос. гос. политехн. ун-т (НПИ) имени М.И. Платова. – Новочеркасск: Лик, 2016. – 60 с.

3. Бахтизин В.В., Глухова Л.А. Методология функционального проектирования IDEF0: учеб. пос. для студентов вузов. – М.: БГУИР, 2003. – 24 с.

4. Советов Б.Я. Архитектура информационных систем: учеб. для студ. учреждений – М.: Изд. центр «Академия», 2012. – 496 с.

5. Калянов Г.Н. Моделирование, анализ, реорганизация и автоматизация бизнес-процессов: учеб. пос. – М.: Финансы и статистика, 2007. – 240 с.

7. Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. Методы и модели анализа данных: OLAP и Data. – СПб.: БХВ-Петербург, 2004. – 336 с.

8. Радоуцкий В.Ю., Шульженко В.Н., Носатова Е.А. Основы научных исследований: учеб. пос. – Белгород: Изд-во БГТУ им. В.Г. Шухова, 2008. – 133 с.

9. Сидняев Н.И. Теория планирования эксперимента и анализ статистических данных: учеб. пос. – М.: Изд-во Юрайт, 2011. – 399 с.

Stepanova T.A.
student

Katsupeev A.A.
Candidate of Engineering Sciences, Associate Professor

South Russian State Polytechnic University (NPI) named after M.I. Platova
Russia, Novocherkassk

COMPARISON OF DBMS PERFORMANCE IN THE PROCESS OF DESIGNING DSS “ELECTRONIC ACCOMPANYING DOCUMENTATION”

Decision support systems are in great demand in large enterprises, where it comes to working with large amounts of data. An example is their widespread development in such industries as instrument making, aviation and space industry. In this regard, there is a need to help data analysts sort and analyze information, and the development of a decision support system in this case can increase employee productivity, update and ensure prompt storage of documentation and simplified access to it. The article also discusses the main stages of system design, as well as a comparative assessment of the speed of performing basic functional operations based on PostgreSQL 9.6 and MySQL 5.1 DBMS.

Key words: *information system, decision support system, operational documentation, context diagram, decomposition, DBMS, operation scheme, data analysis, processor performance, data storage, expert system, functional operations.*