

МОДЕЛИРОВАНИЕ И АНАЛИЗ РАБОТЫ РАСПРЕДЕЛЕННОЙ БАЗЫ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ ОБЪЕКТНО-ОРИЕНТИРОВАННОГО ПОДХОДА

Телятников А.О.

Донецкий национальный технический университет г. Донецк
кафедра автоматизированных систем управления
E-mail: Alexander.Telyatnikov@gmail.com

Abstract

Telyatnikov A.O. Modeling and the analysis of work of the distributed database with use of the object-oriented approach. The approach to the analysis of work DDB and an estimation of influence of changes brought in hardware and the software on its productivity, based on use of objective model DDB Is offered. Computing experiments for check of adequacy of model are lead.

Введение

Современный этап развития компьютерных информационных систем (КИС) можно охарактеризовать как переход от автоматизации отдельных задач к построению корпоративных КИС, созданию единого информационного пространства предприятия. В то же время наблюдается тенденция к децентрализации информационных систем, появлению распределенных КИС в состав которых входят распределенные базы данных (РБД). На сегодняшний день технологии распределенных СУБД достигли того уровня развития, когда на рынке уже имеются достаточно развитые и надежные коммерческие системы. Однако ввиду исключительной сложности РБД как объекта формального описания и расчета, до настоящего времени оценка ожидаемого эффекта от тех или иных изменений аппаратного и программного обеспечения а, значит и от капиталовложений, необходимых для их реализации, обычно базируется на общих рассуждениях, не подкрепленных численными расчетами, когда влияние отдельных факторов оценивается интуитивно. Поэтому создание моделей для анализа работы РБД и исследования влияния параметров технических средств и размещения данных на ее эффективность, является актуальной научной задачей.

Объектная модель РБД

Предлагается подход к анализу работы РБД и оценке влияния изменений аппаратного и программного обеспечения на ее эффективность, основанный на использовании разработанной ранее объектной модели РБД [1]. Данная модель построена как система взаимодействующих объектов ее типовых компонентов: узел РБД, канал передачи данных, приложение, запрос, таблица РБД. Модель позволяет рассчитывать различные параметры функционирования РБД, в том числе критерий эффективности – среднее время выполнения запросов и распространения обновлений в системе:

$$T = \frac{1}{n_q} \sum_{i=1}^{n_q} t_i^q + \frac{1}{n_u} \sum_{j=1}^{n_u} t_j^u ;$$

где n_q , n_u – количество запросов и обновлений порожденных функционированием системы;

t_{ij}^q – время выполнения запроса:

$$t_{ij}^q = t_{wt} + t_{tq} + t_{wp} + t_{p_{ij}} + t_{wt} + t_{ta} ;$$

t_{ij}^u – время выполнения обновления:

$$t_{ij}^u = t_{wt} + t_{tu} + t_{wp} + t_{p_{ij}} + t_{wt} + t_{tr};$$

t_{tq} – время передачи запроса:

$$t_{tq} = \frac{V_q}{\min(B_1)};$$

t_{ta} – время передачи ответа на запрос:

$$t_{ta} = \frac{V_a}{\min(B_1)};$$

$t_{p_{ij}}$ – время обработки запроса:

$$t_{p_{ij}} = \frac{K_j}{P_i} \cdot 60 (с);$$

t_{tu} – время передачи обновления:

$$t_{tu} = \frac{V_u}{\min(B_1)};$$

t_{tr} – время передачи сообщения о завершении обновления:

$$t_{tr} = \frac{V_r}{\min(B_1)};$$

t_{wt} – время ожидания передачи;

t_{wp} – время ожидания обработки;

V_q – объем запроса;

V_a – объем ответа на запрос;

P_i – производительность i -того узла передачи данных (тпмС);

K_j – количество транзакций ТРС-С соответствующее j -тому запросу;

V_u – объем обновления;

V_r – объем сообщения о завершении обновления;

B_1 – пропускная способность 1-го канала передачи данных.

Характеристика объекта экспериментальных исследований

Для проверки адекватности разработанной модели проведен ряд вычислительных экспериментов. В качестве объекта экспериментальных исследований выбрана распределенная компьютерная информационная система ЗАО ПО “Киев-Конти”. В состав компании “Киев-Конти” входят 4 фабрики: – три в Украине (Донецкая, Константиновская, Горловская) и одна в России (Курская). Компания имеет распределенную систему сбыта, состоящую из пяти филиалов (складов продукции), из них четыре в Украине – в г. Донецк, Киев, Львов, Николаев и один филиал в России в г. Воронеж, а также несколько региональных представительств.

Из-за территориальной распределенности структуры компании “Киев-Конти” ее компьютерная информационная система также имеет распределенную архитектуру, одним из

элементов которой является РБД, в состав которой входит 10 узлов: центральный узел (корпоративный сервер), по одному узлу на каждой фабрике и в каждом филиале. Данная РБД имеет размер и характеристики, позволяющие использовать ее в качестве объекта экспериментальных исследований для разработанной модели.

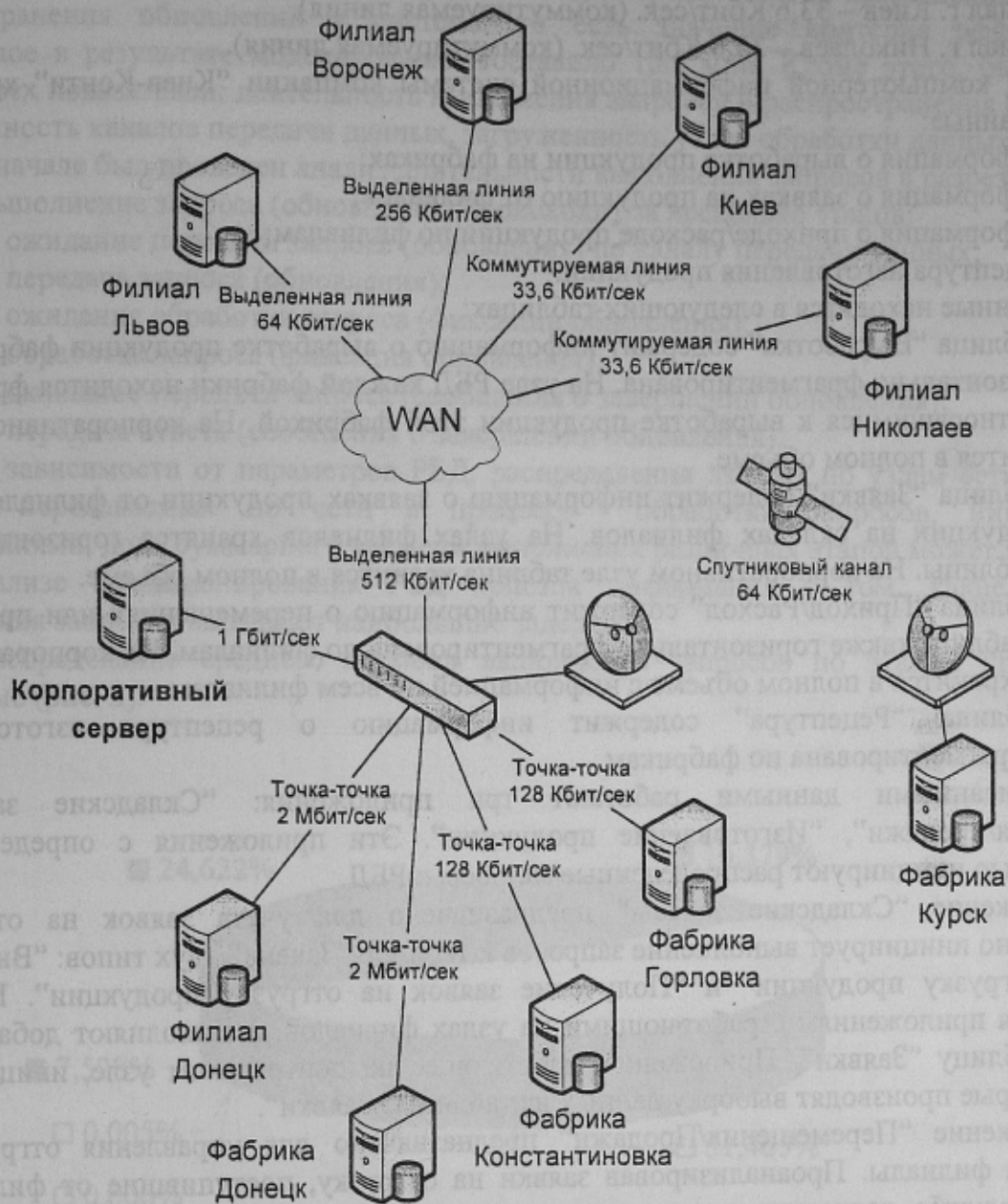


Рисунок 1 - Структура компьютерной сети ЗАО ПО "Киев-Кonti".

Рассмотрим структуру сети передачи данных компьютерной информационной системы ЗАО ПО "Киев-Кonti", приведенную на рис.1. Корпоративный сервер, расположенный в центральном офисе компании в г. Донецке, подключен к центральному коммутатору с помощью высокоскоростного соединения (1 Гбит/с). К этому же концентратору подключено сетевое оборудование, обеспечивающее следующие соединения:

1. Филиал г. Донецк – 2 Мбит/сек. (выделенная линия).
2. Фабрика г. Донецк – 2 Мбит/сек. (выделенная линия).
3. Фабрика г. Константиновка – 128 Кбит/сек. (выделенная линия).
4. Фабрика г. Курск – 64 Кбит/сек. (спутниковый канал).
5. Провайдер Internet – 512 Кбит/сек. (выделенная линия).

Все остальные филиалы, кроме филиала в г. Донецке, подключаются к корпоративному серверу посредством сети WAN и имеют следующие соединения с провайдерами Internet:

1. Филиал г. Воронеж – 256 Кбит/сек. (выделенная линия).
2. Филиал г. Львов – 64 Кбит/сек. (выделенная линия).
3. Филиал г. Киев – 33,6 Кбит/сек. (коммутируемая линия).
4. Филиал г. Николаев – 33,6 Кбит/сек. (коммутируемая линия).

В РБД компьютерной информационной системы компании “Киев-Конти” хранятся следующие данные:

- 1) информация о выработке продукции на фабриках;
- 2) информация о заявках на продукцию от филиалов;
- 3) информация о приходе/расходе продукции по филиалам;
- 4) рецептура изготовления продукции.

Эти данные находятся в следующих таблицах:

1. Таблица “Выработка” содержит информацию о выработке продукции фабриками. Таблица горизонтально фрагментирована. На узле РБД каждой фабрики находится фрагмент с данными, относящимися к выработке продукции этой фабрикой. На корпоративном узле таблица хранится в полном объеме.

2. Таблица “Заявки” содержит информацию о заявках продукции от филиалов и об остатках продукции на складах филиалов. На узлах филиалов хранятся горизонтальные фрагменты таблицы. На корпоративном узле таблица хранится в полном объеме.

3. Таблица “Приход/Расход” содержит информацию о перемещениях или продажах продукции. Таблица также горизонтально фрагментирована по филиалам. На корпоративном узле таблица хранится в полном объеме с информацией по всем филиалам.

4. Таблица “Рецептура” содержит информацию о рецептуре изготовления продукции. Фрагментирована по фабрикам.

С описанными данными работают три приложения: “Складские запасы”, “Перемещения/продажи”, “Изготовление продукции”. Эти приложения с определенной интенсивностью инициируют распределенные запросы к РБД.

Приложение “Складские запасы” предназначено для учета заявок на отгрузку продукции. Оно инициирует выполнение запросов к таблице “Заявки” двух типов: “Внесение заявок на отгрузку продукции” и “Получение заявок на отгрузку продукции”. Первые инициируются приложениями, работающими на узлах филиалов, и выполняют добавление записей в таблицу “Заявки”. Приложение, работающее на центральном узле, инициирует запросы, которые производят выборку данных из таблицы “Заявки”.

Приложение “Перемещения/Продажи” предназначено для управления отгрузками продукции на филиалы. Проанализировав заявки на отгрузку, поступившие от филиалов, центральная служба логистики принимает решение об отгрузках и сообщает филиалам информацию об отгруженной им продукции. Это выполняется с помощью запросов “Внесение отгрузки”. Приложение “Перемещения/Продажи”, функционирующее на узлах филиалов, инициирует запросы “Получение отгрузки” которые осуществляют чтение данных из таблицы “Приход/расход”.

Приложение “Изготовление продукции” предназначено для управления выработкой и рецептурой производства продукции. В конце каждой смены на фабриках в таблицу “Выработка” с помощью запросов “Внесение выработки” вносится информация об объемах продукции, изготовленной за смену. Производственный отдел делает выборку данной информации с помощью запросов “Получение выработки”.

В начале каждой смены на фабриках просматривается рецептура производства продукции, то есть происходит выборка данных из таблицы “Рецептура”, с помощью запроса

“Получение рецептуры”. Изменения в рецептуру изготовления продукции вносит производственный отдел используя запрос “Изменение рецептуры”.

Анализ результатов моделирования

С помощью разработанной модели были проведены вычислительные эксперименты с целью анализа параметров функционирования РБД. Среднее время выполнения запросов и распространения обновлений в системе, то есть значение критерия эффективности, полученное в результате моделирования составило 1,86 мин. Кроме этого, был проведен анализ трех показателей: длительность выполнения запросов и распространения обновлений, загруженность каналов передачи данных, загруженность узлов обработки данных.

Вначале был проведен анализ длительности выполнения запросов в разрезе этапов.

Выполнение запроса (обновления) происходит за несколько этапов:

- ожидание передачи запроса (обновления) по каналу передачи данных;
- передача запроса (обновления);
- ожидание обработки запроса (фиксации обновления);
- обработка запроса (фиксация обновления);
- ожидание передачи запроса (сообщения о завершении обновления);
- передача ответа (сообщения о завершении обновления).

В зависимости от параметров РБД, распределения данных по узлам сети и объемов данных, передаваемых по сети в процессе обработки запросов, инициируемых приложениями, доля суммарного времени выполнения различных этапов может различаться. При анализе функционирования РБД полезна информация о том, какие из этапов выполнения запросов вызывают наибольшие задержки.

Распределение среднего времени выполнения запросов по этапам приведено на диаграмме (рис. 2).

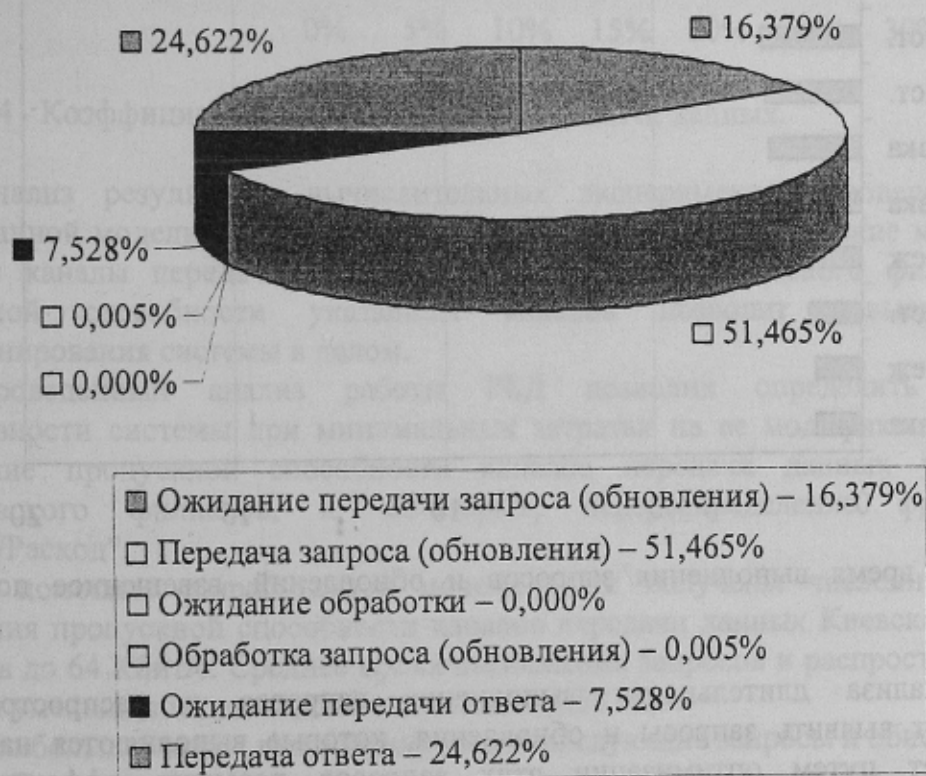


Рисунок 2 - Распределение среднего времени выполнения запросов (обновлений) по этапам.

На діаграмме видно, что наибольшие задержки вызывает передача обновлений и ответов на запросы. Обработка данных узлами РБД происходит быстро и практически не оказывает влияния на среднее время выполнения запросов и распространения обновлений.

На диаграмме (рис. 3) представлено среднее время выполнения запросов (обновлений), взвешенное по их удельному объему. Анализ этих данных показал, что наибольшее время выполнения имеют запросы, инициируемые на узлах "Филиал. Киев" и "Филиал. Николаев" и запросы, которые обращаются к таблицам, хранящимся на этих узлах. Это связано с тем, что при передаче данных используются медленные коммутируемые каналы, со скоростью передачи 33,6 Кбит/сек.

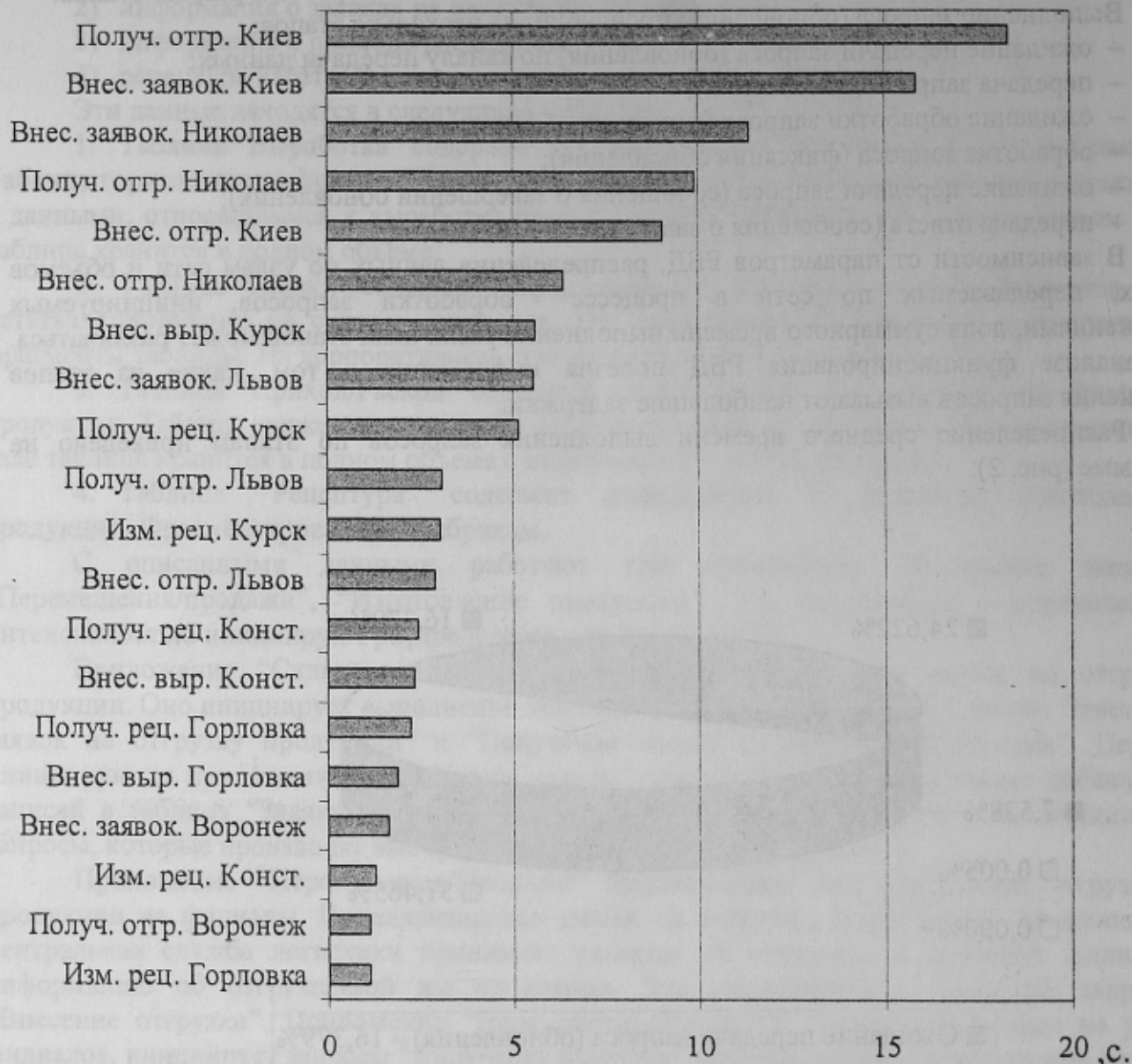


Рисунок 3 - Среднее время выполнения запросов и обновлений, взвешенное по их удельному объему.

Результаты анализа длительности выполнения запросов и распространения обновлений позволяют выявить запросы и обновления, которые выполняются наиболее долго. Это позволяет путем оптимизации этих запросов повысить эффективность функционирования РБД в целом.

Кроме этого, для анализа представляет интерес загрузка каналов передачи данных. Такой анализ позволит выяснить, на сколько загружены каналы и какие из них являются "узкими местами" системы.

На рис. 4 приведены коэффициенты загрузки каналов передачи данных. Наибольший коэффициент загрузки имеют каналы, соединяющие Киевский и Николаевский филиалы с Центральным узлом. Это объясняется тем, что по ним передаются достаточно большие объемы данных, но при этом они являются коммутируемыми и имеют невысокую пропускную способность.



Рисунок 4 - Коэффициенты загрузки каналов передачи данных.

Анализ результатов вычислительных экспериментов проведенных с помощью разработанной модели РБД позволил выявить так называемые "узкие места" системы. Ими являются каналы передачи данных Киевского и Николаевского филиалов. Повышение пропускной способности указанных каналов позволит повысить эффективность функционирования системы в целом.

Проведенный анализ работы РБД позволил определить пути повышения эффективности системы при минимальных затратах на ее модификацию. Это, во-первых, повышение пропускной способности каналов передачи данных узлов Киевского и Николаевского филиалов, и, во-вторых, перераспределение фрагментов таблицы "Приход/Расход".

С помощью разработанной модели были получены численные оценки влияния увеличения пропускной способности каналов передачи данных Киевского и Николаевского филиалов до 64 Кбит/с. Среднее время выполнения запросов и распространения обновлений при этом уменьшилось на 15,12% и составило 1,58 мин.

Наибольшее время выполнения имеют следующие запросы и обновления:

1. Получение отгрузки. Киев
2. Внесение заявок на отгрузку. Киев
3. Внесение заявок на отгрузку. Николаев

Рациональное перераспределение таблиц и фрагментов данных, к которым обращаются указанные запросы (обновления) потенциально приводит к улучшению работы РБД, в смысле уменьшения времени отклика системы. Моделирование РБД показало, что после удаления копии фрагментов таблицы "Приход/Расход" из узлов филиалов г. Киева и г. Николаева, когда данная таблица остается только на корпоративном сервере, среднее время выполнения запросов и распространения обновлений уменьшилось еще на 18,19% и составило 1,29 мин.

На рис. 5 представлены результаты моделирования РБД.

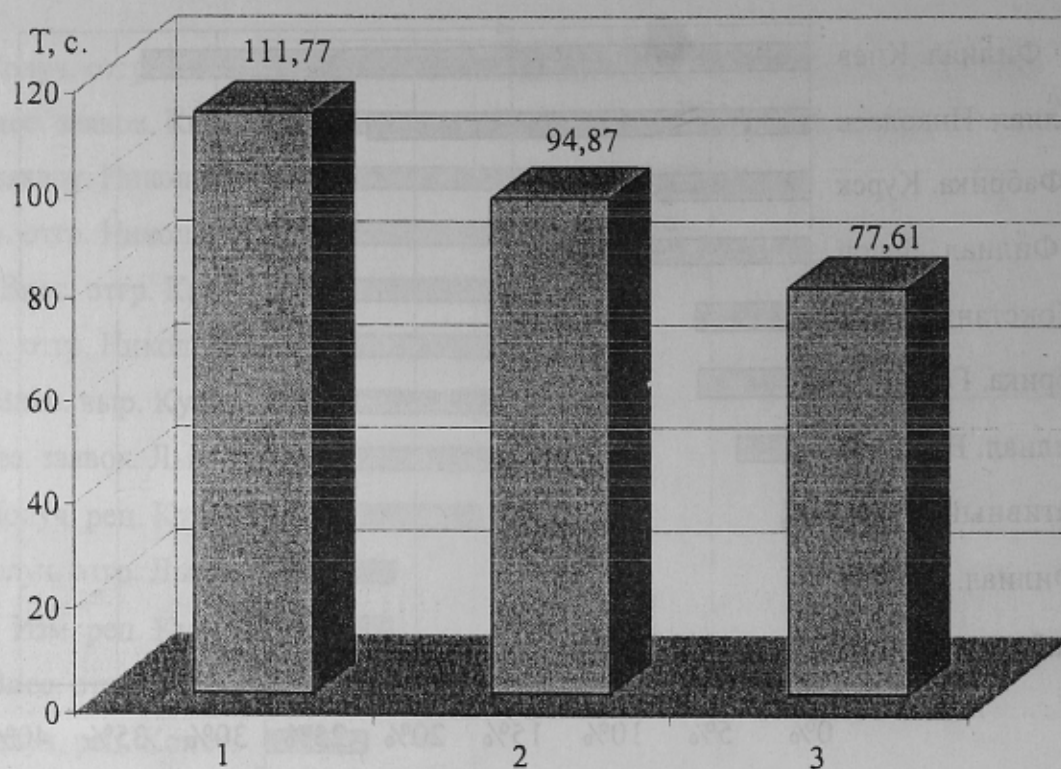


Рисунок 5 - Значения критерия эффективности при различных параметрах РБД.

1 – первоначальный вариант; 2 – после увеличения пропускной способности каналов Киевского и Николаевского филиалов до 64 Кбит/с; 3 – после перераспределения данных..

Выводы

Таким образом, с помощью разработанной модели были получены численные оценки влияния увеличения пропускной способности отдельных каналов передачи данных и перераспределения данных по узлам компьютерной информационной сети.

Разработанная модель может быть использована для получения оценок влияния тех или иных изменений аппаратного и программного обеспечения на эффективность работы РБД. Также модель может быть использована при оптимизации распределения данных по узлам информационной сети с использованием генетических алгоритмов [2].

Литература

1. Телятников А.О. Разработка объектной модели распределенной базы данных // Наукові праці ДонНТУ. Випуск 74. - Донецьк: ДонНТУ, 2004. – С. 192 - 200.
2. Лаздынь С.В., Телятников А.О. Оптимизация распределенных баз данных с использованием генетических алгоритмов // Вестник ХГТУ № 1(19). – Херсон: ХГТУ, 2004. – С. 236-239.