Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования

«Московский государственный технический университет имени Н. Э. Баумана»

На правах рукописи

**Григорьев Ю.А., Ермаков Е.Ю., Ермаков О.Ю., Паничкина А.А., Плужникова О.Ю.**

**МЕТОДИЧЕСКИЕ УКАЗАНИЯ-2021 К ВЫПОЛНЕНИЮ ПРОЕКТНО-ТЕХНОЛОГИЧЕСКОЙ ПРАКТИКИ (П-ТП)**

Москва – 2021

Оглавление

[Цель и задачи проектно-технологической практики (П-ТП). 4](#_Toc60745157)

[Концепция КСАМ 2.0 5](#_Toc60745158)

[Этапы выполнения П-ТП. 7](#_Toc60745159)

[Требования к защите результатов П-ТП. 7](#_Toc60745160)

[Начало работы в КСАМ 2.0. 7](#_Toc60745161)

[1. Раздел 1. Описать в КСАМ 2.0 схему базы данных, запросы и транзакции хранилища данных. 9](#_Toc60745162)

[1.1. Описание схемы БД. 9](#_Toc60745163)

[1.2. Ввод данных о таблицах и их атрибутах. 10](#_Toc60745164)

[1.3. Ввод данных о запросах. 11](#_Toc60745165)

[1.4. Ввод данных о транзакциях. 14](#_Toc60745166)

[Требования к отчёту по разделу 1. 14](#_Toc60745167)

[2. Раздел 2. Описать в КСАМ 2.0 конфигурации узлов и сетей, а также расположение БД и транзакций 16](#_Toc60745168)

[2.1. Ввод данных об узлах и сетей . 16](#_Toc60745169)

[2.2. Описание расположения БД и транзакций (пункты «Физические БД» и «Обращения к транзакциям») 17](#_Toc60745170)

[Требования к отчёту по разделу 2. 19](#_Toc60745171)

[3. Раздел 3. Выполнение вычислений, построение графиков (пункт «Параметры расчетов») 20](#_Toc60745172)

[Требования к отчёту по разделу 3. 23](#_Toc60745173)

[4. Выводы по П-ТП. 25](#_Toc60745174)

[ЛИТЕРАТУРА 25](#_Toc60745175)

[Приложение 1. Параметры наполнения базы данных хранилища данных 26](#_Toc60745176)

[Приложение 2. Запросы 30](#_Toc60745177)

[Приложение 3. Транзакции. 46](#_Toc60745178)

[Приложение 4. Характеристики узлов и сетей 48](#_Toc60745179)

[Приложение 5. Параметры обращений к транзакциям. 50](#_Toc60745180)

[Приложение 6. Варианты выполнения вычислений 51](#_Toc60745181)

[Приложение 7. Классификация кластеров 52](#_Toc60745182)

[Приложение 8. Расчёт числа записей в соединении нескольких таблиц  53](#_Toc60745183)

[Приложение 9. Расшифровка csv-файла 54](#_Toc60745184)

[Приложение 10. Интерфейс редактирования модели 56](#_Toc60745185)

[Приложение 11. Титульный лист и задание. 57](#_Toc60745186)

# Цель и задачи проектно-технологической практики (П-ТП).

**Цель П-ТП:** проанализировать временные характеристики функционирования автоматизированной информационной системы на этапе её проектирования (на примере аналитических запросов TPC-H к хранилищу данных) с помощью пакета КСАМ 2.0.

Тест TPC-H оценивает производительность систем поддержки принятия решений (СППР). Он состоит из набора сложных, бизнес-ориентированных запросов ad hoc. Данные в таблицах и запросы подобраны так, чтобы отражать некоторую усредненную по индустрии бизнес-активность. Типичные запросы составлены так, чтобы соответствовать основным типам запросов в СППР: ценообразование и скидки, управление прибылью, исследование предпочтений покупателей, исследование рынка и т. п.

В отличие от натурных экспериментов (например, TPC-H), требующих создание дорогостоящего стенда, наполнение базы данных и воссоздание рабочей нагрузки, пакет КСАМ 2.0 выполняет оценку времени выполнения ресурсоёмких запросов с помощью математических моделей. Это намного дешевле, что особенно важно на этапе проектирования АСОИУ, когда параметры системы только выбираются.

При выполнении П-ТП решаются следующие **задачи**:

1. Описание концептуального проекта (КП) проектируемой системы.
2. Описание технического проекта (ТП) проектируемой системы.
3. Проведение моделирования работы системы в заданных условиях.
4. Формирование предложений по выбору рациональной конфигурации анализируемой системы.

**Примечание. Перед выполнением П-ТП** **необходимо прочитать настоящие методические указания до конца.**

# Концепция КСАМ 2.0

На рисунке 1 представлена реализованная в КСАМ 2.0 схема описания модели компонентов проектируемой АСОИУ.



Рисунок 1 – Схема описания компонентов АСОИУ в КСАМ 2.0.

1 – таблицы используются для формирования базы данных,

2 - для базы данных описываются запросы,

3 - для базы данных описываются транзакции и в них включаются запросы,

4 - описываются узлы и сети будущей системы,

5 - определяется узел, где размещается база данных и выполняются все её транзакции, а также параметры обращения к транзакциям,

6 – определяются варьируемые параметры, и выполняется модельный эксперимент, результаты вычислений сохраняются в csv-файле.

Вид интерфейса модели в соответствии с перечисленными компонентами представлен на рисунке 2.



Рисунок 2 – Интерфейс модели в КСАМ 2.0

**В приложении 10 представлен общий интерфейс редактирования модели с поясняющими обозначениями.**

**Внимание! Используйте поле «Поиск по модели», чтобы выделить в дереве объекты только с требуемыми именами (или начальными префиксами).**

# Этапы выполнения П-ТП.

П-ТП выполняется в три этапа:

1. **Раздел 1.** Описание в КСАМ 2.0 схемы базы данных, запросов и транзакций хранилища данных.
2. **Раздел 2**. Описание в КСАМ 2.0 конфигураций узлов и сетей, а также расположения БД и транзакций.
3. **Раздел 3**. Выполнение расчётов, построение графиков, выбор наилучшей конфигурации узла, где хранится база данных.

## Требования к защите результатов П-ТП.

1. Продемонстрировать описание и работу модели, разработанной в среде КСАМ 2.0.
2. Предоставить один отчёт по П-ТП, содержащий 3 раздела (см. ниже).
3. Ответить на вопросы преподавателя.

##

## Начало работы в КСАМ 2.0.

Ниже представлена последовательность действий при входе в КСАМ:

1. URL адрес системы **http://ksam.bmstu.ru/**
2. В боковой меню выберите пункт «Войти» и заполните форму авторизации.
3. Смените пароль после первого входа в систему
4. В боковом меню (кнопка перед КСАМ 2.0) выберите пункт «Модели» и создайте новую модель (кнопка «Создать модель»)
5. Выберите слева «Анонимная модель» (щёлкнуть).
6. Щёлкните слева «Анонимная модель» и введите новое название модели.
7. Сохраните изменения локально (кнопка «Сохранить локально»)
8. Сохраните модель удаленно. Для этого в окне редактирования модели рядом с кнопкой «Сохранить локально» раскройте список (иконка V) и выберите пункт «Сохранить удаленно».

**Примечание**.

Чтобы просмотреть и/или изменить модель из облака выполните следующие действия:

1. В боковом меню (кнопка перед КСАМ 2.0) выберите пункт «Модели».
2. В списке «Модели в облаке» напротив нужной модели нажмите иконку V,
3. Выберите «Сохранить локально» или «Сохранить локально как…».

 **Внимание!** «Сохранить локально как…» позволит сохранить модель из облака **как новую** и под новым именем.

Если некоторая модель уже есть в списке «Локальные модели» и Вы ее **повторно** сохраняете из облака с помощью «Сохранить локально», то старая версия(локальная) замещается версией из облака.

**Важно! В процессе работы Вы можете потерять данные четырьмя способами:**

1. **После заполнения модели Вы не сохранили её удаленно и вышли из аккаунта.**
2. **Вы перезаписали локальную модель этой же моделью из облака, нажав кнопку «Сохранить локально».**
3. **Очистили КЭШ браузера.**
4. **Вы перезагрузили страницу браузера, не сохранив изменения локально или удаленно.**

**Будьте внимательны.**

# 1. Раздел 1. Описать в КСАМ 2.0 схему базы данных, запросы и транзакции хранилища данных.

## 1.1. Описание схемы БД.

Инфологическая схема базы данных, к которой обращаются анализируемые при выполнении П-ТП запросы, приведена на рисунке 2 (она взята из теста TPC-H).



Рисунок 2 – Инфологическая схема БД.

Более подробно схема базы данных и запросы теста TPC-H описаны в прилагаемом файле «tpch2.17.1.pdf».

## 1.2. Ввод данных о таблицах и их атрибутах.

**Важно! Параметры наполнения базы данных приведены в Приложении 1. Там же приведены варианты таблиц (и коэффициент SF), информацию о которых необходимо ввести. Номер Вашего варианта указан в задании на П-ТП.**

Последовательность действий при вводе данных о таблице и атрибутах:

1. Выберите локальную модель, раскройте слева модель (кнопка ‘V’), щёлкните «Схемы баз данных» и затем справа «+Добавить схему БД».

2. Щёлкните слева «новая схема БД», введите справа название схемы БД в поле «Название схемы БД».

3. Раскройте слева схему БД и затем выберите слева пункт «Таблицы», щёлкните справа кнопку «+Добавить таблицу»

4. Щёлкните слева «новая таблица», справа заполните следующие поля:

«Название таблицы»,

«Количество строк в таблице» (т.е. записей).

5. Выберите таблицу слева, щёлкните справа кнопку «+Добавить атрибут».

6. Щёлкните слева «новый атрибут», справа заполните следующие поля:

«Название атрибута»,

«Тип данных» (выбрать из списка),

«Размер (байт)» - это длина атрибута в байтах.

«Мощность атрибута» - это число различных значений атрибута. Заполняется только для ключей (PK, FK см. Приложение 1) и полей, по которым будет проводиться группировка (см спецификацию TPC-H пункт 4.2.2, 4.2.3 ).

«Первичный ключ - это флаг для идентификации первичного ключа.

7. Повторить пункты 5,6 для остальных атрибутов таблицы.

8. Повторить пункты 3-7 для остальных таблиц.

**Важно! После каждого шага сохраняйте внесенные данные с помощью кнопки «Сохранить локально».**

## 1.3. Ввод данных о запросах.

**Важно! Запросы описаны в Приложении 2. Там же приведены варианты запросов, информацию о которых необходимо ввести. Номер Вашего варианта указан в задании на П-ТП.**

Последовательность действий при вводе данных о запросах:

1. Раскройте схему БД слева и затем выберите пункт «Запросы», щёлкните справа кнопку «+Добавить запрос»

2. Щёлкните слева «новый запрос», справа сверху введите название запроса в поле «Название запроса».

3. Далее следует выполнить три шага:

3.1. Определить таблицы.

3.2. Определить связи.

3.3. Определить атрибуты.

Рассмотрим эти шаги подробнее.

3.1. Определить таблицы. Здесь указываются таблицы, участвующие в запросе.

3.1.1. Щёлкните иконку ‘V’ справа от имени таблицы, если она входит в запрос. Выбрать один из двух вариантов:

- «Добавить без псевдонима» - добавить таблицу в запрос без псевдонима (по сути псевдонимом будет считаться название таблицы);

- «Добавить с псевдонимом» - откроется окно для ввода псевдонима, в котором требуется указать уникальное в контексте данного запроса название псевдонима таблицы; можно назначить одной таблице несколько псевдонимов (это позволяет соединять таблицу с ней самой).

Внизу появляется список имён таблиц и/или псевдонимов таблиц, участвующих в запросе. Нажав на крестик, можно удалить таблицу или псевдоним из запроса. В дальнейшем исходную таблицу или её псевдоним будем называть просто таблицей.

Щёлкните кнопку «Следующий шаг», чтобы перейти к шагу «Определить связи»

3.2. Определить связи. Используется, если в запрос включено несколько таблиц.

3.2.1. Щёлкните кнопку «Добавить ДЖОИН». Откроется окно выбора левой и правой таблицы соединения. Щёлкните по пустому полю ниже строки «Левая таблица» и выберите левую таблицу. Аналогично выберите правую таблицу соединения. Щёлкните кнопку «Добавить». Таким способом можно добавить несколько пар соединяемых таблиц (см. примечание ниже).

3.2.2. Далее следует определить атрибуты, по которым производится соединение. Для этого нажмите иконку ‘V’ выбранного соединения и выделите те атрибуты, которые участвуют в JOIN. Условие, накладываемое на атрибуты соединения, здесь не задаётся (предполагается произвольное условие: равенство, больше, меньше и др.).

Щёлкните кнопку «Следующий шаг», чтобы перейти к шагу «Определить атрибуты».

3.3. Определить атрибуты. Здесь для каждой таблицы указывается эффективная селективность атрибутов (для подзапросов) и определяются атрибуты, участвующие в выборке, агрегировании, группировании и сортировки.

3.3.1. Щёлкните иконку ‘V’ после имени таблицы и для атрибута таблицы укажите:

- флаг «проекция», если атрибут указан за ключевым словом SELECT как отдельный атрибут, а не как агрегат,

- «условие» - вероятность P (эффективную селективность), если на атрибут накладывается ограничение в условии WHERE – вероятность определена в описании запроса (см. Приложение 2),

- флаги «сортировка» и/или «группировка», если атрибут используется за ключевым словом «order by» и/или «group by» (если в запросе конструкция order by указана для агрегата, то вместо агрегата установите флаг «сортировка» для первого атрибута, который указан в group by).

**Примечание**. В модели время сортировки для режима "group by ... order by" зависит от 1) числа записей (для реализации group by) и 2) от числа групп (для реализации order by) - см. описание математической модели в [1]. Поэтому для оценки времени неважно, по какому атрибуту проекции выполняется сортировка.

4. Определить агрегаты на выходе. Здесь указываются агрегаты (sum, avg, count, арифметическое выражение и др.), которые указаны в блоке SELECT (название агрегата м.б. любым, число агрегатов равно числу агрегатов за select, размер агрегата равен 4 байтам).

5. Повторите пункты 1-4 для всех запросов, указанных в задании.

**Важно! После каждого шага сохраняйте внесенные данные с помощью кнопки «Сохранить локально».**

**Примечание.**

Если в запросе соединяются более 2-х таблиц, то Join требуется добавить несколько раз (см. п. 3.2.1). Например, пусть в операторе select указано

FROM T1,T2,T4 ps1,T4 ps2 WHERE T1.a=T2.a and T2.b=ps1.b and ps1.c=ps2.c.

В этом случае следует описать таблицы T1, T2, T4 (псевдоним ps1), T4 (псевдоним ps2) и добавить следующие Join: Join(T1,T2), Join(T2, ps1), Join(ps1, ps2). Порядок соединения можно изменить. Пары соединяемых таблиц определяются по равенству (или другому условию) атрибутов соединения в условии where запроса.

## 1.4. Ввод данных о транзакциях.

**Важно! Варианты транзакций описаны в Приложении 3. Номер Вашего варианта указан в задании на П-ТП.**

Последовательность действий при вводе данных о транзакциях:

1. Раскройте схему БД слева и затем выберите пункт «Транзакции», щёлкните справа кнопку «+Добавить транзакцию».

2. Щёлкните слева «новая транзакция», и справа заполните следующие поля:

- «Название транзакции» (TP1, TP2 - на латинице, чтобы имя отображалось в csv-файле без искажений),

- для тех запросов, которые вошли в транзакцию, в поле «Количество повторений**»** укажите 1 (единицу),

- для подзапросов укажите галочку «Подзапрос» (для запросов Q81, Q91, Q71, Q41, Q21).

3. Повторите пункты 1-2 для остальных транзакций.

**Важно! После каждого шага сохраняйте внесенные данные с помощью кнопки «Сохранить локально».**

##

## Требования к отчёту по разделу 1.

**В отчёт не надо вставлять скриншоты.**

1. Для таблиц, указанных в запросах, приведите (из Приложения 1)

- названия таблиц и число записей в них,

- список атрибутов и их параметры для этих таблиц.

2. Приведите запросы транзакций Вашего варианта П-ТП из Приложения 2 (может быть изменённые и с дополнительными запросами – см. примечания к некоторым запросам).

3. Приведите расчёты, которые Вы выполнили, используя Приложение 8.

4. Укажите параметры всех транзакций для Вашего варианта П-ТП (из Приложения 3).

# 2. Раздел 2. Описать в КСАМ 2.0 конфигурации узлов и сетей, а также расположение БД и транзакций

## 2.1. Ввод данных об узлах и сетей .

**Важно! Характеристики узлов и сетей приведены в Приложении 4. Там же приведены варианты аппаратного комплекса. Номер Вашего варианта аппаратного комплекса указан в задании на П-ТП.**

Последовательность действий при описании узлов для выбранной модели:

1. Раскройте модель слева, выберите пункт «Узлы», щёлкните справа кнопку «+Добавить узел».

2. Щёлкните слева «новый узел», и справа заполните следующие поля:

- «Название узла»,

- «Тип узла» (кластер – это отдельная многопроцессорная станция, клиент – может включать несколько рабочих станций (машин)),

- «Количество машин в узле» (процессорных ядер в кластере или число рабочих станций клиента).

3. Для узла типа «кластер» указать

- Мод (модель кластера): SE,

- ОП (производительность ОП в Мбайт/с),

- Диск (производительность диска в Мбайт/с); значение производительности указано в колонке «Мбайт/с(диск)/коэффициент p» после знака равенства в Приложении 4; коэффициент p – это вероятность, что читаемого блока нет в ОП,

- Количество дисков (в кластере);

- Пропускной канал сети (производительность шины между процессорами в Мбит/с),

- ЦПУ (производительность ЦПУ в ГГц).

4. Повторить пункты 1-3 для остальных узлов аппаратного комплекса.

Последовательность действий при описании сети для выбранной модели (она соединяет рабочие станции и кластер):

1. Выберите пункт «Сети», щёлкните справа кнопку «+Добавить сеть».

2. Щёлкните слева «новая сеть», и справа заполните следующие поля:

- «Название сети»,

- «Скорость сети, Мбит/с»,

- указать узлы, которые объединяются в сеть.

**Важно! После каждого шага сохраняйте внесенные данные с помощью кнопки «Сохранить локально».**

## 2.2. Описание расположения БД и транзакций (пункты «Физические БД» и «Обращения к транзакциям»)

Последовательность действий при описании физической БД для выбранной модели:

1. Выберите пункт «Физические базы данных», щёлкните справа кнопку «+Добавить БД».

2. Щёлкните слева «новая физическая БД», и справа заполните следующие поля:

- «Название физической БД» (в узле кластера),

- «Тип базы данных» (выберите из списка тип «Реляционная база данных»),

- «Кластер размещения» (выберите из списка узел, который Вы определили как кластер, и где будут располагаться БД и транзакции),

- «Модель данных» (выберите из списка схему базы данных).

**Важно! Варианты параметров обращений к транзакциям (REQUEST)**

**приведены в Приложении 5. Номер Вашего варианта указан в задании на П-ТП.**

Последовательность действий при вводе обращений к транзакциям для выбранной модели:

1. Выберите пункт «Обращения к транзакциям», щёлкните справа кнопку «+Добавить обращение к транзакции».

2. Щёлкните слева «новое обращение к транзакции», и справа заполните следующие поля:

- «Название обращения» (на латинице, чтобы имя отображалось в csv-файле без искажений),

- «Тип обращения» (выберите из списка online или offline),

- «Частота запросов»:

а) для типа online – частота обращений к транзакции (кол/час), приходящихся на одну машину в узле,

б) для типа offline - число пакетов (определение см. ниже),

- «Узел» (для online – выбрать узел типа «клиент», для offline – выбрать узел типа «кластер»),

- «База данных» (выберите из списка физическую базу данных),

- «Транзакция» (выберите из списка транзакцию, к которой выполняется обращение).

3. Повторить пункты 1-2 для остальных обращений к транзакциям (REQUEST) Вашего варианта.

**Примечание**.

Для режима «online» каждая машина в узле (для клиента –каждая рабочая станция) обращается к транзакции с указанной частотой. Для типа «offline» моделируется пакетный режим работы системы. В этом случае задаётся число пакетов. В каждом пакете выполняются запросы, указанные в транзакции. Пакеты запускаются, как правило, в ночное время с разными параметрами входящих в них запросов и выполняются параллельно.

**Важно! После каждого шага сохраняйте внесенные данные с помощью кнопки «Сохранить локально».**

## Требования к отчёту по разделу 2.

**В отчёт не надо вставлять скриншоты.**

1. Приведите характеристики узлов и сетей вашего варианта аппаратного комплекса (из Приложения 4).

2. Приведите Ваш вариант параметров обращений к транзакциям (из Приложения 5).

# 3. Раздел 3. Выполнение вычислений, построение графиков (пункт «Параметры расчетов»)

**Важно! Варианты вычислений приведены в Приложении 6. Номер Вашего варианта указан в задании на П-ТП.**

Описание действий при вычислении на модели:

1. Раскройте модель слева, выберите пункт «Параметры расчетов», щёлкните справа кнопку «+Добавить параметр».

**Для обращения к транзакции типа «offline»**.

2. В появившейся справа форме заполните следующие поля:

- в поле «Тип объекта» выберите «Обращение к транзакции»,

-в поле «Объект» выберите обращение к транзакции типа «offline» из числа обращений, созданных в пункте 2.2,

- в поле «Поле» выберите «Частота» (число пакетов) и укажите нижнюю и верхнюю границы, а также шаг для числа пакетов транзакции (см. Примечания ниже).

3. Радом к кнопкой «Сохранить локально» щёлкните иконку ‘V’ и выберите «Начать расчет».

**Важно! Расчёты выполняются на сервере. Перед отправкой параметров модель автоматически сохраняется удаленно.**

**4. Через некоторое время (~ 1 минуту, а может быть и раньше) раскройте боковое меню (кнопка перед КСАМ 2.0) и щелкните «Результаты вычислений». В списке доступных расчетов появится новая позиция (сверху) с указанием имени модели, для которой были выполнены вычисления, времени и даты.**

5. Нажмите на требуемую позицию, чтобы скачать соответствующий csv-файл с результатами. Сохраните результаты (здесь Вам потребуются столбцы, связанные с запросами, которые вошли в транзакцию типа offline, см. Приложение 9).

6. Вернитесь в окно редактирования модели (кнопка «Назад» веб-браузера).

7. Удалите предыдущие параметры расчётов (кнопка справа).

8. Добавьте параметр расчёта (кнопка «+Добавить параметр»).

**Для транзакции типа «online»**.

9. В появившейся справа форме заполните следующие поля:

- в поле «Тип объекта» выберите «Узел»,

-в поле «Объект» выберите узел типа «pc» из числа узлов (клиентов), созданных в пункте 2.1,

- в поле «Поле» выберите радиокнопку «Кол-во машин» (число рабочих станций клиента) и укажите нижнюю и верхнюю границы, а также шаг для числа рабочих станций (см. Примечания ниже).

11. Выполните указанные выше пункты 3,4 для этого параметра.

12. Нажмите на требуемую позицию, чтобы скачать соответствующий csv-файл с результатами. Сохраните результаты (здесь Вам потребуются столбцы, связанные с запросами, которые вошли в транзакцию типа online, см. Приложение 9).

13. Удалите предыдущие параметры расчётов.

14. Измените характеристики узла и сети (пункты «Узлы» и «Сети») для второго варианта (всего их 3 для Вашего задания, см. Приложение 6) и повторите вычисления для этого варианта (пункты 1-13) для ранее назначенных границ варьируемых параметров. Далее повторите вычисления для третьего варианта характеристик узла и сети Вашего задания (см. Приложение 6).

15. Используя данные csv-файлов (Excel), постройте следующие графики:

а) зависимость времени выполнения обращения к транзакции (request) типа «offline» от числа пакетов (сумма времени соответствующих запросов); исследованные варианты конфигураций узла и сети должны выступать на графике в виде легенды; из соответствующих csv-файлов надо выбирать время выполнения запросов, которые вошли в обращение к транзакции типа «offline»;

б) зависимость времени выполнения обращения к транзакции (request) типа «online» от числа рабочих станций; исследованные варианты конфигураций узла и сети должны выступать на графике в виде легенды; из соответствующих csv-файлов надо выбирать время выполнения запросов, которые вошли в обращение к транзакции типа «online»;

в) графики зависимостей загрузок устройств (диска, процессора, сети) от числа рабочих станций (для request типа «online»).

.

Примечания!!

1. Границы интервала изменения числа пакетов и числа рабочих станций Вам необходимо подобрать вручную (см. выше пункты 2 и 9):

а) проведя ряд дополнительных модельных экспериментов, найдите интервал изменения числа пакетов, в котором время выполнения транзакции типа «offline» изменялось бы от 1000 сек до 10000 сек;

б) проведя ряд дополнительных модельных экспериментов, найдите интервал изменения числа рабочих станций, в котором время выполнения транзакции типа «online» изменялось бы от 1 сек до 100 сек; **следите, чтобы загрузки устройств не превышали 1 (см. Приложение 9).** Для некоторыхвариантов имеется пологий участок, где время изменяется очень мало. Это связано с небольшой загрузкой устройств (диска, процессора, сети). В этом случае увеличьте верхнюю границу числа рабочих станций.

2. Шаг изменения (см. выше пункты 2 и 9) следует выбрать так, чтобы число точек составило около 10 на каждую транзакцию (при большем числе точек возрастает время вычислений). Например, «1 1 10», или «10 10 100», или «20 20 200» и т.д.

3. При изменении характеристик узла и сети (см. пункт 14) параметры узла типа «pc» менять не надо.

4. Внимание! Если после исследования всех 3-х конфигураций (см. пункт 1 этого примечания) во всех конфигурациях загрузка какого-либо устройства превышает 1 или время выполнения транзакции превышает установленный порог (100 с. – для online, 10000 с. – для offline) даже при числе рабочих станций, равным 1, то некоторые исходные данные следует скорректировать, согласовав их с преподавателем. При этом следует скорректировать и многокритериальную задачу (см. ниже).

**Внимание! Если даже при одной рабочей станции загрузка какого-либо устройства (диск, процессор, сеть) превышает 1, то обычно загрузку можно уменьшить, уменьшив частоту обращения к транзакции (она может быть меньше 1), или увеличив число устройств, или увеличив скорость устройства, или уменьшив коэффициент p. Изменение надо согласовать с преподавателем. В этом и смысл моделирования: на модели можно подобрать конфигурацию на этапе проектирования системы ещё до закупки дорогостоящего оборудования. При изменении числа устройств необходимо скорректировать и стоимости исследуемых конфигураций (стоимости устройств можно взять из Интернет).**

## Требования к отчёту по разделу 3.

**В отчёт не надо вставлять скриншоты.**

1. Приведите характеристики узлов, которые Вы исследовали (Приложение 6 -> Приложение 4).

2. Приведите графики зависимостей времени выполнения транзакции от числа пакетов (для request типа «offline») и числа рабочих станций (для request типа «online»), а также графики зависимостей загрузок устройств (диска, процессора, сети) от числа рабочих станций (для request типа «online»); исследованные варианты конфигураций узла (кластера) и сети должны выступать на графике в виде легенды.

3. Определите, какая конфигурация кластера является лучшей и почему. Для этого решите многокритериальную задачу:

цена кластера (y1) → min (см. Приложение 4),

число рабочих станций (y2) → max,

число пакетов (y3) → max,

время выполнения транзакции (online) < 100 c (см. примечания выше),

время выполнения транзакции (offline) < 10000 c (см. примечания выше).

Для решения многокритериальной задачи выполните следующие действия:

- с помощью графиков определите допустимые решения (векторы критериев, т.е. альтернативы),

- сравните альтернативы, сведя многокритериальную задачу к однокритериальной:

K= 0.4 (max{y1i} / y1i) + 0.4 (y2i / min {y2i}) + 0.2 (y3i / min {y3i}) → max,

где

i – номер альтернативы,

y1i - цена кластера, y2i - число рабочих станций, y3i - число пакетов.

# 4. Выводы по П-ТП.

В выводах необходимо указать, какая из проанализированных конфигураций узла и сети является лучшей по производительности и почему.

## ЛИТЕРАТУРА

1.Григорьев Ю.А., Ермаков О.Ю. и др.Оценка временных характеристик автоматизированных систем обработки информации и управления с помощью пакета КСАМ 2.0: Учебно-методическое пособие / Ю.А. Григорьев, О.Ю. Ермаков, А.А. Паничкина, О.Ю. Плужникова. – М.: Издательство «Спутник +», 2019. – 56 с. (есть в библиотеке МГТУ)

# Приложение 1. Параметры наполнения базы данных хранилища данных

**Варианты таблиц (последнее число – это коэффициент наполнения SF).**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **1** S\_,P\_,PS\_,C\_,O\_,L\_, N\_,R\_,100 | **2** S\_,P\_,PS\_,C\_,O\_,L\_, N\_,R\_,200 | **3** S\_,P\_,PS\_,C\_,O\_,L\_, N\_,R\_,300 | **4** S\_,P\_,PS\_,C\_,O\_,L\_, N\_,R\_,400 | **5** S\_,P\_,PS\_,C\_,O\_,L\_, N\_,R\_,500 |

**Число записей в таблицах**

|  |  |  |
| --- | --- | --- |
| Название таблицы (рис. 2) | Число записей в таблице | Примечание |
| SUPPLIER(S\_) | SF\*10000 |  |
| PART(P\_) | SF\*200000 |  |
| PARTSUPP(PS\_) | SF\*800000 |  |
| CUSTOMER(C\_) | SF\*150000 |  |
| ORDERS(O\_) | SF\*1500000 |  |
| LINEITEM(L\_) | SF\*6001215 |  |
| NATION(N\_) | 25 |  |
| REGION(R\_) | 5 |  |

**Параметры атрибутов таблиц**

**Внимание! Мощность заполняется только для ключевых значений PK и FK (см. раздел 1.2, пункт 6). Мощность PK равна числу записей в таблице. Мощность FK соответствует числу записей в родительской таблице. Например, пусть SF=100. Тогда мощность атрибута S\_SUPPKEY (PK) в таблице SUPPLIER равна числу записей в этой таблице: SF\*10000=100\*10000=1000000. Мощность атрибута S\_NATIONKEY (FK) равна 25 (число записей в родительской таблице NATION). Мощности атрибутов PS\_PARTKEY (FK) и PS\_SUPPKEY (FK) таблицы PARTSUPP равны числу записей соответственно в таблицах PART (SF\*200000=20000000) и SUPPLIER (SF\*10000=1000000). И так далее.**

Таблица «SUPPLIER»

|  |  |  |  |
| --- | --- | --- | --- |
| Название атрибута | Длина (байты) | Тип | Примечание |
| S\_SUPPKEY  | 8  | long | PK |
| S\_NAME  | 25  | text |  |
| S\_ADDRESS  | 40  | text |  |
| S\_NATIONKEY  | 8  | long | FK |
| S\_PHONE  | 15  | text |  |
| S\_ACCTBAL  | 4  | decimal |  |
| S\_COMMENT  | 101  | text |  |
| Длина записи | 201 |  |  |

Таблица «PART»

|  |  |  |  |
| --- | --- | --- | --- |
| Название атрибута | Длина (байты) | Тип | Примечание |
| P\_PARTKEY  | 8 | long | PK |
| P\_NAME  | 55 | text |  |
| P\_MFGR  | 25 | text |  |
| P\_BRAND  | 10 | text |  |
| P\_TYPE  | 25 | text |  |
| P\_SIZE  | 4 | integer |  |
| P\_CONTAINER  | 10 | text |  |
| P\_RETAILPRICE  | 4 | decimal |  |
| P\_COMMENT  | 23 | text |  |
| Длина записи | 164 |  |  |

Таблица «PARTSUPP»

|  |  |  |  |
| --- | --- | --- | --- |
| Название атрибута | Длина (байты) | Тип | Примечание |
| PS\_PARTKEY  | 8 | long | FK |
| PS\_SUPPKEY  | 8 | long | FK |
| PS\_AVAILQTY  | 8 | integer |  |
| PS\_SUPPLYCOST  | 4 | decimal |  |
| PS\_COMMENT  | 199 | text |  |
| Длина записи | 227 |  |  |

Таблица «CUSTOMER»

|  |  |  |  |
| --- | --- | --- | --- |
| Название атрибута | Длина (байты) | Тип | Примечание |
| C\_CUSTKEY  | 8 | long | PK |
| C\_NAME  | 25 | text |  |
| C\_ADDRESS  | 40 | text |  |
| C\_NATIONKEY  | 8 | long | FK |
| C\_PHONE  | 15 | text |  |
| C\_ACCTBAL  | 4 | decimal |  |
| C\_MKTSEGMENT  | 10 | text |  |
| C\_COMMENT  | 117 | text |  |
| Длина записи | 227 |  |  |

Таблица «ORDERS»

|  |  |  |  |
| --- | --- | --- | --- |
| Название атрибута | Длина (байты) | Тип | Примечание |
| O\_ORDERKEY  | 8 | long | PK |
| O\_CUSTKEY  | 8 | long | FK |
| O\_ORDERSTATUS  | 1 | text |  |
| O\_TOTALPRICE  | 4 | Decimal |  |
| O\_ORDERDATE  | 4 | Date |  |
| O\_ORDERPRIORITY  | 15 | text |  |
| O\_CLERK  | 15 | text |  |
| O\_SHIPPRIORITY  | 4 | Integer |  |
| O\_COMMENT  | 79 | text |  |
| Длина записи | 138 |  |  |

Таблица «LINEITEM»

|  |  |  |  |
| --- | --- | --- | --- |
| Название атрибута | Длина (байты) | Тип | Примечание |
| L\_ORDERKEY  | 8 | long | FK |
| L\_PARTKEY  | 8 | long | FK |
| L\_SUPPKEY  | 8 | long | FK |
| L\_LINENUMBER  | 4 | integer |  |
| L\_QUANTITY  | 4 | decimal |  |
| L\_EXTENDEDPRICE  | 4 | decimal |  |
| L\_DISCOUNT  | 4 | decimal |  |
| L\_TAX  | 4 | decimal |  |
| L\_RETURNFLAG  | 1 | text |  |
| L\_LINESTATUS  | 1 | text |  |
| L\_SHIPDATE  | 4 | Date |  |
| L\_COMMITDATE  | 4 | Date |  |
| L\_RECEIPTDATE  | 4 | Date |  |
| L\_SHIPINSTRUCT  | 25 | text |  |
| L\_SHIPMODE  | 10 | text |  |
| L\_COMMENT  | 44 | text |  |
| Длина записи | 137 |  |  |

Таблица «NATION»

|  |  |  |  |
| --- | --- | --- | --- |
| Название атрибута | Длина (байты) | Тип | Примечание |
| N\_NATIONKEY  | 8 | long | PK |
| N\_NAME  | 25 | text |  |
| N\_REGIONKEY  | 8 | long | FK |
| N\_COMMENT  | 152 | text |  |
| Длина записи | 193 |  |  |

Таблица «REGION»

|  |  |  |  |
| --- | --- | --- | --- |
| Название атрибута | Длина (байты) | Тип | Примечание |
| R\_REGIONKEY  | 8 | long | PK |
| R\_NAME  | 25 | text |  |
| R\_COMMENT  | 152 | text |  |
| Длина записи | 185 |  |  |

# Приложение 2. Запросы

**Варианты запросов**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **1** Q1, Q3Q2, Q7 | **2** Q1, Q5Q4, Q9 | **3** Q1, Q6Q2, Q7 | **4** Q1, Q10Q4, Q9 | **5** Q3, Q5Q2, Q7 |
| **6** Q3, Q6Q4, Q9 | **7** Q3, Q10Q2, Q7 | **8** Q5, Q6Q4, Q9 | **9** Q5, Q10Q2, Q7 | **10** Q6, Q10Q4, Q9 |
| **11** Q1, Q3Q2, Q8 | **12** Q1, Q5Q4, Q8 | **13** Q1, Q6Q2, Q8 | **14** Q1, Q10Q4, Q8 | **15** Q3, Q5Q2, Q8 |
| **16** Q3, Q6Q4, Q8 | **17** Q3, Q10Q2, Q8 | **18** Q5, Q6Q4, Q8 | **19** Q5, Q10Q2, Q8 | **20** Q6, Q10Q4, Q8 |
| **21** Q1, Q3Q2, Q9 | **22** Q1, Q5Q4, Q7 | **23** Q1, Q6Q2, Q9 | **24** Q1, Q10Q4, Q7 | **25** Q3, Q5Q2, Q9 |
| **26** Q3, Q6Q4, Q7 | **27** Q3, Q10Q2, Q9 | **28** Q5, Q6Q4, Q7 | **29** Q5, Q10Q2, Q9 | **30** Q6, Q10Q4, Q7 |
| **31** Q1, Q3Q4, Q7 | **32** Q1, Q5Q2, Q9 | **33** Q1, Q6Q4, Q7 | **34** Q1, Q10Q2, Q9 | **35** Q3, Q5Q4, Q7 |
| **36** Q3, Q6Q2, Q9 | **37** Q3, Q10Q4, Q7 | **38** Q5, Q6Q2, Q9 | **39** Q5, Q10Q4, Q7 | **40** Q6, Q10Q2, Q9 |
| **41** Q1, Q3Q4, Q8 | **42** Q1, Q5Q2, Q8 | **43** Q1, Q6Q4, Q8 | **44** Q1, Q10Q2, Q8 | **45** Q3, Q5Q4, Q8 |
| **46** Q3, Q6Q2, Q8 | **47** Q3, Q10Q4, Q8 | **48** Q5, Q6Q2, Q8 | **49** Q5, Q10Q4, Q8 | **50** Q6, Q10Q2, Q8 |
| **51** Q1, Q3Q4, Q9 | **52** Q1, Q5Q2, Q7 | **53** Q1, Q6Q4, Q9 | **54** Q1, Q10Q2, Q7 | **55** Q3, Q5Q4, Q9 |
| **56** Q3, Q6Q2, Q7 | **57** Q3, Q10Q4, Q9 | **58** Q5, Q6Q2, Q7 | **59** Q5, Q10Q4, Q9 | **60** Q6, Q10Q2, Q7 |
| **61** Q1, Q3Q7, Q8 | **62** Q1, Q5Q8, Q9 | **63** Q1, Q6Q7, Q8 | **64** Q1, Q10Q8, Q9 | **65** Q3, Q5Q7, Q8 |
| **66** Q3, Q6Q8, Q9 | **67** Q3, Q10Q7, Q8 | **68** Q5, Q6Q7, Q8 | **69** Q5, Q10Q8, Q9 | **70** Q6, Q10Q7, Q8 |
| **71** Q1, Q3Q7, Q9 | **72** Q1, Q5Q7, Q8 | **73** Q1, Q6Q7, Q9 | **74** Q1, Q10Q7, Q8 | **75** Q3, Q5Q7, Q9 |
| **76** Q3, Q6Q7, Q9 | **77** Q3, Q10Q8, Q9 | **78** Q5, Q6Q7, Q9 | **79** Q5, Q10Q7, Q8 | **80** Q6, Q10Q7, Q9 |
| **81** Q1, Q3Q8, Q9 | **82** Q1, Q5Q7, Q9 | **83** Q1, Q6Q8, Q9 | **84** Q1, Q10Q7, Q9 | **85** Q3, Q5Q8, Q9 |
| **86** Q3, Q6Q7, Q8 | **87** Q3, Q10Q7, Q9 | **88** Q5, Q6Q8, Q9 | **89** Q5, Q10Q7, Q9 | **90** Q6, Q10Q8, Q9 |
| **91** Q1, Q3Q2, Q4 | **92** Q1, Q5Q2, Q4 | **93** Q1, Q6Q2, Q4 | **94** Q1, Q10Q2, Q4 | **95** Q3, Q5Q2, Q4 |
| **96** Q3, Q6Q2, Q4 | **97** Q3, Q10Q2, Q4 | **98** Q5, Q6Q2, Q4 | **99** Q5, Q10Q2, Q4 | **100** Q6, Q10Q2, Q4 |

**Важно.**

**1. Запросы Q2, Q4 с подзапросом в условии where необходимо предварительно преобразовать в запросы с соединением таблиц – как это происходит при оптимизации (см. Примечания к соответствующим запросам).**

**2. Запросы Q7, Q8, Q9 с подзапросом за ключевым словом from необходимо предварительно преобразовать (см. Примечания к соответствующим запросам).**

**1. Запрос Q1**

select l\_returnflag,

l\_linestatus,

sum(l\_quantity) as sum\_qty,

sum(l\_extendedprice) as sum\_base\_price,

sum(l\_extendedprice\*(1-l\_discount)) as sum\_disc\_price,

sum(l\_extendedprice\*(1-l\_discount)\*(1+l\_tax)) as sum\_charge,

avg(l\_quantity) as avg\_qty,

avg(l\_extendedprice) as avg\_price,

avg(l\_discount) as avg\_disc,

count(\*) as count\_order

from lineitem

where l\_shipdate <= date '1998-12-01' - interval '[DELTA]' day (3) // P=0,96

group by

l\_returnflag,

l\_linestatus

order by

l\_returnflag,

l\_linestatus;

**2. Запрос Q2**

select s\_acctbal,

s\_name,

n\_name,

p\_partkey,

p\_mfgr,

s\_address,

s\_phone,

s\_comment

from part,

supplier,

partsupp,

nation,

region

where p\_partkey = ps\_partkey

and s\_suppkey = ps\_suppkey

and p\_size = [SIZE] // P=1/50

and p\_type like '%[TYPE]' // P=30/150

and s\_nationkey = n\_nationkey

and n\_regionkey = r\_regionkey

and r\_name = '[REGION]' // P=1/5

and ps\_supplycost = (

select min(ps\_supplycost)

from partsupp, supplier,

nation, region

where p\_partkey = ps\_partkey

and s\_suppkey = ps\_suppkey

and s\_nationkey = n\_nationkey

and n\_regionkey = r\_regionkey

and r\_name = '[REGION]' // P=1/5

)

order by

s\_acctbal desc,

n\_name,

s\_name,

p\_partkey;

**Примечание к запросу Q2.**

Этот запрос необходимо преобразовать в запрос с соединением таблиц:

1. Дополнительно описать следующий запрос Q21 (см. подзапрос выше) и включить его в транзакцию, куда входит запрос Q2:

select p\_partkey, min(ps\_supplycost) as psmin

from **part**, partsupp, supplier,

nation, region

where p\_partkey = ps\_partkey

and s\_suppkey = ps\_suppkey

and s\_nationkey = n\_nationkey

and n\_regionkey = r\_regionkey

and r\_name = '[REGION]' // P=1/5

**and p\_size = [SIZE] // P=1/50**

**and p\_type like '%[TYPE]' // P=30/150**

 group by p\_partkey;

2. Описать в базе данных (пункт «Данные») дополнительную таблицу tablQ21, которая соответствует запросу Q21. Число записей можно рассчитать по формулам, которые приведены в Приложении 8.

3. Изменить исходный запрос Q2:

- добавить во from таблицу tablQ21,

- вместо равенства «and ps\_supplycost= (подзапрос select)» добавить в where следующие строки:

and part. p\_partkey=tablQ21.p\_partkey

and ps\_supplycost= tablQ21.psmin // P= 1/99900

**3. Запрос Q3**

select

l\_orderkey,

sum(l\_extendedprice\*(1-l\_discount)) as revenue,

o\_orderdate,

o\_shippriority

from

customer,

orders,

lineitem

where

c\_mktsegment = '[SEGMENT]' // P=1/5

and c\_custkey = o\_custkey

and l\_orderkey = o\_orderkey

and o\_orderdate < date '[DATE]' // P=0,49

and l\_shipdate > date '[DATE]' // P=0,51

group by

l\_orderkey,

o\_orderdate,

o\_shippriority

order by

revenue desc,

o\_orderdate;

**4. Запрос Q4**

select

o\_orderpriority,

count(\*) as order\_count

from

orders

where

o\_orderdate >= date '[DATE]' // в интервале

and o\_orderdate < date '[DATE]' + interval '3' month // P=0,04

and exists (

select \*

from

lineitem

where

l\_orderkey = o\_orderkey

and l\_commitdate < l\_receiptdate // P= 0,98

)

group by

o\_orderpriority

order by

o\_orderpriority;

**Примечание к запросу Q4.**

Этот запрос необходимо преобразовать в запрос с соединением таблиц:

1. Дополнительно описать следующий запрос Q41 (см. подзапрос выше) и включить его в транзакцию, куда входит запрос Q4:

select o\_orderkey, count(\*) as ocount

from

**orders,**

lineitem

where

l\_orderkey = o\_orderkey

and l\_commitdate < l\_receiptdate // P= 0,98

**and o\_orderdate >= date '[DATE]' // в интервале**

**and o\_orderdate < date '[DATE]' + interval '3' month// P=0,04**

 group by o\_orderkey;

2. Описать в базе данных (пункт «Данные») дополнительную таблицу tablQ41, которая соответствует запросу Q41. Число записей можно рассчитать по формулам, которые приведены в Приложении 8.

3. Изменить исходный запрос Q4:

- добавить во from таблицу tablQ41,

- вместо квантора «and exists (подзапрос select)» добавить в where следующие строки:

and orders. o\_orderkey =tablQ41.o\_orderkey

and tablQ41.ocount>0 // P=1

**5. Запрос Q5**

select

n\_name,

sum(l\_extendedprice \* (1 - l\_discount)) as revenue

from

customer,

orders,

lineitem,

supplier,

nation,

region

where

c\_custkey = o\_custkey

and l\_orderkey = o\_orderkey

and l\_suppkey = s\_suppkey

and c\_nationkey = s\_nationkey

and s\_nationkey = n\_nationkey

and n\_regionkey = r\_regionkey

and r\_name = '[REGION]' // P=1/5

and o\_orderdate >= date '[DATE]' // в интервале

and o\_orderdate < date '[DATE]' + interval '1' year // P= 0,15

group by

n\_name

order by

revenue desc;

**6. Запрос Q6**

select

sum(l\_extendedprice\*l\_discount) as revenue

from

lineitem

where

l\_shipdate >= date '[DATE]' // в интервале

and l\_shipdate < date '[DATE]' + interval '1' year // P= 0,14

and l\_discount between [DISCOUNT] - 0.01 and [DISCOUNT] + 0.01 // P= 0,3

and l\_quantity < [QUANTITY]; // P= 0,5

**7. Запрос Q7**

select supp\_nation,

cust\_nation,

l\_year, sum(volume) as revenue

from ( select n1.n\_name as supp\_nation,

n2.n\_name as cust\_nation,

extract(year from l\_shipdate) as l\_year,

l\_extendedprice \* (1 - l\_discount) as volume

from supplier,

lineitem,

orders,

customer,

nation n1,

nation n2

where s\_suppkey = l\_suppkey

and o\_orderkey = l\_orderkey

and c\_custkey = o\_custkey

and s\_nationkey = n1.n\_nationkey

and c\_nationkey = n2.n\_nationkey

and (

(n1.n\_name = '[NATION1]' and n2.n\_name = '[NATION2]') //P1=P2= 1/25

or (n1.n\_name = '[NATION2]' and n2.n\_name = '[NATION1]') //P3=P4=1/25

)

and l\_shipdate between date '1995-01-01' and date '1996-12-31' // P=0,29

) as shipping

group by

supp\_nation,

cust\_nation,

l\_year

order by

supp\_nation,

cust\_nation,

l\_year;

**Примечание к запросу Q7.**

Последовательность преобразования запроса:

1. Дополнительно описать запрос Q71, который соответствует подзапросу select, указанному за from, и включить его в транзакцию, куда входит запрос Q7 (заменить таблицы n1 и n2 на их декартово произведение n=n1×n2 и далее выполнять соединения с таблицей n; для этого требуется создать вспомогательную таблицу n).

2. Описать в базе данных (пункт «Данные») дополнительную таблицу tablQ71, которая соответствует запросу Q71. Число записей можно рассчитать по формулам, которые приведены в Приложении 8 (условие с вероятностями P1÷P4 применять к декартовому произведению n1 и n2; для этого необходимо рассчитать требуемую вероятность P и применить её к ключу таблицы n).

3. Заменить подзапрос select за ключевым словом from на таблицу tablQ71.

**8. Запрос Q8**

select o\_year,

sum(case

when nation = '[NATION]'

then volume

else 0

end) / sum(volume) as mkt\_share

from ( select extract(year from o\_orderdate) as o\_year,

l\_extendedprice \* (1-l\_discount) as volume,

n2.n\_name as nation

from part,

supplier,

lineitem,

orders,

customer,

nation n1,

nation n2,

region

where p\_partkey = l\_partkey

and s\_suppkey = l\_suppkey

and l\_orderkey = o\_orderkey

and o\_custkey = c\_custkey

and c\_nationkey = n1.n\_nationkey

and n1.n\_regionkey = r\_regionkey

and r\_name = '[REGION]' // P=1/5

and s\_nationkey = n2.n\_nationkey

and o\_orderdate between date '1995-01-01' and date '1996-12-31' // P=0,3

and p\_type = '[TYPE]' // P=1/150

) as all\_nations

group by

o\_year

order by

o\_year;

**Примечание к запросу Q8.**

Последовательность преобразования запроса:

1. Дополнительно описать запрос Q81, который соответствует подзапросу select, указанному за from, и включить его в транзакцию, куда входит запрос Q8.

2. Описать в базе данных (пункт «Данные») дополнительную таблицу tablQ81, которая соответствует запросу Q81. Число записей можно рассчитать по формулам, которые приведены в Приложении 8.

3. Заменить подзапрос select за ключевым словом from на таблицу tablQ81.

**9. Запрос Q9**

Select nation,

o\_year,

sum(amount) as sum\_profit

from (

select

n\_name as nation,

extract(year from o\_orderdate) as o\_year,

l\_extendedprice \* (1 - l\_discount) - ps\_supplycost \* l\_quantity as amount

from

part,

supplier,

lineitem,

partsupp,

orders,

nation

where

s\_suppkey = l\_suppkey

and ps\_suppkey = l\_suppkey

and ps\_partkey = l\_partkey

and p\_partkey = l\_partkey

and o\_orderkey = l\_orderkey

and s\_nationkey = n\_nationkey

and p\_name like '%[COLOR]%' // P= 5/91

) as profit

group by

nation,

o\_year

order by

nation,

o\_year desc;

**Примечание к запросу Q9.**

Последовательность преобразования запроса:

1. Дополнительно описать запрос Q91, который соответствует подзапросу select, указанному за from, и включить его в транзакцию, куда входит запрос Q9.

2. Описать в базе данных (пункт «Данные») дополнительную таблицу tablQ91, которая соответствует запросу Q91. Число записей можно рассчитать по формулам, которые приведены в Приложении 8.

3. Заменить подзапрос select за ключевым словом from на таблицу tablQ91.

**10. Запрос Q10**

select

c\_custkey,

c\_name,

sum(l\_extendedprice \* (1 - l\_discount)) as revenue,

c\_acctbal,

n\_name,

c\_address,

c\_phone,

c\_comment

from

customer,

orders,

lineitem,

nation

where

c\_custkey = o\_custkey

and l\_orderkey = o\_orderkey

and o\_orderdate >= date '[DATE]' // в интервале

and o\_orderdate < date '[DATE]' + interval '3' month // P=0,04

and l\_returnflag = 'R' // P=0,25

and c\_nationkey = n\_nationkey

group by

c\_custkey,

c\_name,

c\_acctbal,

c\_phone,

n\_name,

c\_address,

c\_comment

order by

revenue desc;

# Приложение 3. Транзакции.

**Варианты транзакций.**

Транзакция=

(запрос1, запрос2)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **1** ТР1= (Q1, Q2) ТР2**=** **(**Q3, Q7**)** | **2** ТР1= (Q1, Q4) ТР2**=**(Q5, Q9) | **3** ТР1= (Q1, Q2) ТР2**=** **(**Q6, Q7**)** | **4** ТР1= (Q1, Q4) ТР2**=** **(**Q10, Q9**)** | **5** ТР1= (Q3, Q2) ТР2**=** **(**Q5, Q7**)** |
| **6** ТР1= (Q3, Q4) ТР2= (Q6, Q9) | **7** ТР1= (Q3, Q2) ТР2**=**(Q10, Q7) | **8** ТР1= (Q5, Q4) ТР2**=**(Q6, Q9) | **9** ТР1= (Q5, Q2) ТР2**=**(Q10, Q7) | **10** ТР1= (Q6, Q4) ТР2**=**(Q10, Q9) |
| **11** ТР1= (Q1, Q2) ТР2= (Q3, Q8) | **12** ТР1= (Q1, Q4) ТР2**=**(Q5, Q8) | **13** ТР1= (Q1, Q2) ТР2**=**(Q6, Q8) | **14** ТР1= (Q1, Q4) ТР2**=**(Q10, Q8) | **15** ТР1= (Q3, Q2) ТР2**=**(Q5, Q8) |
| **16** ТР1= (Q3, Q4) ТР2**=**(Q6, Q8) | **17** ТР1= (Q3, Q2) ТР2**=**(Q10, Q8) | **18** ТР1= (Q5, Q4) ТР2**=**(Q6, Q8) | **19** ТР1= (Q5, Q2) ТР2**=**(Q10, Q8) | **20** ТР1= (Q6, Q4) ТР2**=**(Q10, Q8) |
| **21** ТР1= (Q1, Q2) ТР2**=**(Q3, Q9) | **22** ТР1= (Q1, Q4) ТР2**=**(Q5, Q7) | **23** ТР1= (Q1, Q2) ТР2**=**(Q6, Q9) | **24** ТР1= (Q1, Q4) ТР2**=**(Q10, Q7) | **25** ТР1= (Q3, Q2) ТР2**=**(Q5, Q9) |
| **26** ТР1= (Q3, Q4) ТР2**=**(Q6, Q7) | **27** ТР1= (Q3, Q2) ТР2**=**(Q10, Q9) | **28** ТР1= (Q5, Q4) ТР2**=**(Q6, Q7) | **29** ТР1= (Q5, Q2) ТР2**=**(Q10, Q9) | **30** ТР1= (Q6, Q4) ТР2**=**(Q10, Q7) |
| **31** ТР1= (Q1, Q4) ТР2**=**(Q3, Q7) | **32** ТР1= (Q1, Q2) ТР2**=**(Q5, Q9) | **33** ТР1= (Q1, Q4) ТР2**=**(Q6, Q7) | **34** ТР1= (Q1, Q2) ТР2**=**(Q10, Q9) | **35** ТР1= (Q3, Q4) ТР2**=**(Q5, Q7) |
| **36** ТР1= (Q3, Q2) ТР2**=**(Q6, Q9) | **37** ТР1= (Q3, Q4) ТР2**=**(Q10, Q7) | **38** ТР1= (Q5, Q2) ТР2**=**(Q6, Q9) | **39** ТР1= (Q5, Q4) ТР2**=**(Q10, Q7) | **40** ТР1= (Q6, Q2) ТР2**=**(Q10, Q9) |
| **41** ТР1= (Q1, Q4) ТР2**=**(Q3, Q8) | **42** ТР1= (Q1, Q2) ТР2**=**(Q5, Q8) | **43** ТР1= (Q1, Q4) ТР2**=**(Q6, Q8) | **44** ТР1= (Q1, Q2) ТР2**=**(Q10, Q8) | **45** ТР1= (Q3, Q4) ТР2**=**(Q5, Q8) |
| **46** ТР1= (Q3, Q2) ТР2**=**(Q6, Q8) | **47** ТР1= (Q3, Q4) ТР2**=**(Q10, Q8) | **48** ТР1= (Q5, Q2) ТР2**=**(Q6, Q8) | **49** ТР1= (Q5, Q4) ТР2**=**(Q10, Q8) | **50** ТР1= (Q6, Q2) ТР2**=**(Q10, Q8) |
| **51** ТР1= (Q1, Q4) ТР2**=**(Q3, Q9) | **52** ТР1= (Q1, Q2) ТР2**=**(Q5, Q7) | **53** ТР1= (Q1, Q4) ТР2**=**(Q6, Q9) | **54** ТР1= (Q1, Q2) ТР2**=**(Q10, Q7) | **55** ТР1= (Q3, Q4) ТР2**=**(Q5, Q9) |
| **56** ТР1= (Q3, Q2) ТР2**=**(Q6, Q7) | **57** ТР1= (Q3, Q4) ТР2**=**(Q10, Q9) | **58** ТР1= (Q5, Q2) ТР2**=**(Q6, Q7) | **59** ТР1= (Q5, Q4) ТР2**=**(Q10, Q9) | **60** ТР1= (Q6, Q2) ТР2**=**(Q10, Q7) |
| **61** ТР1= (Q1, Q7) ТР2**=**(Q3, Q8) | **62** ТР1= (Q1, Q8) ТР2**=**(Q5, Q9) | **63** ТР1= (Q1, Q7) ТР2**=**(Q6, Q8) | **64** ТР1= (Q1, Q8) ТР2**=**(Q10, Q9) | **65** ТР1= (Q3, Q7) ТР2**=**(Q5, Q8) |
| **66** ТР1= (Q3, Q8) ТР2**=**(Q6, Q9) | **67** ТР1= (Q3, Q7) ТР2**=**(Q10, Q8) | **68** ТР1= (Q5, Q7) ТР2**=**(Q6, Q8) | **69** ТР1= (Q5, Q8) ТР2**=**(Q10, Q9) | **70** ТР1= (Q6, Q7) ТР2**=**(Q10, Q8) |
| **71** ТР1= (Q1, Q7) ТР2**=**(Q3, Q9) | **72** ТР1= (Q1, Q7) ТР2**=**(Q5, Q8) | **73** ТР1= (Q1, Q7) ТР2**=**(Q6, Q9) | **74** ТР1= (Q1, Q7) ТР2**=**(Q10, Q8) | **75** ТР1= (Q3, Q7) ТР2**=**(Q5, Q9) |
| **76** ТР1= (Q3, Q7) ТР2**=**(Q6, Q9) | **77** ТР1= (Q3, Q8) ТР2**=**(Q10, Q9) | **78** ТР1= (Q5, Q7) ТР2**=**(Q6, Q9) | **79** ТР1= (Q5, Q7) ТР2**=**(Q10, Q8) | **80** ТР1= (Q6, Q7) ТР2**=**(Q10, Q9) |
| **81** ТР1= (Q1, Q8) ТР2**=**(Q3, Q9) | **82** ТР1= (Q1, Q7) ТР2**=**(Q5, Q9) | **83** ТР1= (Q1, Q8) ТР2**=**(Q6, Q9) | **84** ТР1= (Q1, Q7) ТР2**=**(Q10, Q9) | **85** ТР1= (Q3, Q8) ТР2**=**(Q5, Q9) |
| **86** ТР1= (Q3, Q7) ТР2**=**(Q6, Q8) | **87** ТР1= (Q3, Q7) ТР2**=**(Q10, Q9) | **88** ТР1= (Q5, Q8) ТР2**=**(Q6, Q9) | **89** ТР1= (Q5, Q7) ТР2**=**(Q10, Q9) | **90** ТР1= (Q6, Q8) ТР2**=**(Q10, Q9) |
| **91** ТР1= (Q1, Q2) ТР2**=**(Q3, Q4) | **92** ТР1= (Q1, Q2) ТР2**=**(Q5, Q4) | **93** ТР1= (Q1, Q2) ТР2**=**(Q6, Q4) | **94** ТР1= (Q1, Q2) ТР2**=**(Q10, Q4) | **95** ТР1= (Q3, Q2) ТР2**=**(Q5, Q4) |
| **96** ТР1= (Q3, Q2) ТР2**=**(Q6, Q4) | **97** ТР1= (Q3, Q2) ТР2**=**(Q10, Q4) | **98** ТР1= (Q5, Q2) ТР2**=**(Q6, Q4) | **99** ТР1= (Q5, Q2) ТР2**=**(Q10, Q4) | **100** ТР1= (Q6, Q2) ТР2**=**(Q10, Q4) |

# Приложение 4. Характеристики узлов и сетей

**Характеристики кластера (Cluster)**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| № | Тип узла | Количество процессорных ядер | Мод (Модель)  | ЦПУ, ГГц | ОП(Мбайт/с) | Пропускной канал сети (внутренней шины)Мбит/с | Мбайт/с(диск)/коэффициент p\* | Количество дисков  | Цена кластера( тыс.$) |
| 1 | кластер | 24 | SE | 2,3 | 10600 | 200000 | 100/0.01=10000 | 4 | 36 |
| 2 | кластер | 16 | SE | 2,9 | 12800 | 200000 | 120/0.4=300 | 8 | 45 |
| 3 | кластер | 32 | SE | 2.27 | 10600 | 200000 | 120/0.012=10000 | 8 | 79 |
| 4 | кластер | 120 | SE | 2.8 | 14900 | 200000 | 120/0.01=12000 | 24 | 180 |
| 5 | кластер | 240 | SE | 2.8 | 12800 | 200000 | 120/0.01=12000 | 48 | 338 |
| 6 | кластер | 56  | SE | 2.5 | 25600 | 200000 | 500/0.02=25000 | 7 | 472 |
| 7 | кластер | 512 | SE | 2.0 | 23400 | 200000 | 3200/0.14=23000 | 8 | 310 |

\* - коэффициент p, учитывающий, что считываемого блока нет в ОП

**Характеристики клиента (PC)**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| № | Тип | Количество машин в узле (рабочих станций) | Мод (Модель)  | ЦПУ, ГГц | ОП(Мбайт/с) | Пропускной канал сети (внутренней), Мбит/с | Диск,Мбайт/с | Количество дисков |
| 1 | клиент | 5 | - | - | - | - | - | - |
| 2 | клиент | 10 | - | - | - | - | - | - |
| 3 | клиент | 20 | - | - | - | - | - | - |

**Характеристики сетей**

|  |  |  |
| --- | --- | --- |
| № | Скорость сети (Мбит/с) | Узлы в сети |
| 1 | 10000 | Cluster#№, PC#№ |
| 2 | 1000 | Cluster#№, PC#№ |

**Варианты:**

|  |  |
| --- | --- |
| **1.** Cluster #1, PC#1, Сеть#1**2**. Cluster #2, PC#1, Сеть#1**3**. Cluster #3, PC#1, Сеть#1**4**. Cluster #4, PC#1, Сеть#1**5**. Cluster #5, PC#1, Сеть#1**6**. Cluster #6, PC#1, Сеть#1**7**. Cluster #7, PC#1, Сеть#2**8**. Cluster #1, PC#1, Сеть#2**9**. Cluster #2, PC#1, Сеть#2**10**. Cluster #3, PC#1, Сеть#2**11**. Cluster #4, PC#1, Сеть#2**12**. Cluster #5, PC#1, Сеть#2 | **13**. Cluster #6, PC#2, Сеть#1**14**. Cluster #7, PC#2, Сеть#1**15**. Cluster #1, PC#2, Сеть#1**16**. Cluster #2, PC#2, Сеть#1**17**. Cluster #3, PC#2, Сеть#1**18**. Cluster #4, PC#2, Сеть#1**19**. Cluster #5, PC#2, Сеть#2**20**. Cluster #6, PC#2, Сеть#2**21**. Cluster #7, PC#2, Сеть#2**22**. Cluster #1, PC#2, Сеть#2**23**. Cluster #2, PC#2, Сеть#2**24**. Cluster #3, PC#2, Сеть#2 |

# Приложение 5. Параметры обращений к транзакциям.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| REQUEST№ | Частота (кол/ч) – для online,Число пакетов – для offline | тип request | узел (откуда) | транзакция |
| 1 | 2 | online | PC#№ | ТР1 |
| 2 | 1 | online | PC#№ | ТР2 |
| 3 | 10 | offline | Cluster#№ | ТР1 |
| 4 | 20 | offline | Cluster#№ | ТР2 |

**Варианты обращений к транзакциям:**

**1**. REQUEST №: 1, 4.

**2**. REQUEST №: 2, 3.

# Приложение 6. Варианты выполнения вычислений

**Вариант**

|  |  |
| --- | --- |
| **1.** 1\*, 2\*, 4\***2.** 2, 1, 5**3.** 3, 2, 4**4.** 4, 1, 3**5.** 5, 1, 3**6.** 6, 1, 2**7.** 7, 2, 3**8.** 8, 2, 7**9.** 9, 1, 6**10.** 10, 1, 6**11.** 11, 5, 1**12**. 12, 2, 1 | **13.** 13, 2, 3**14.** 14, 4, 5**15.** 15, 5, 6**16.** 16, 7, 8**17.** 17, 9, 11**18.** 18, 10, 19**19.** 19, 14, 15**20.** 20, 15, 16**21.** 21, 17, 18**22.** 22, 19, 20**23.** 23, 21, 22**24.** 24, 16, 12 |

\* - это номер варианта в Приложении 4.

# Приложение 7. Классификация кластеров

Наиболее распространенной системой классификации параллельных систем баз данных является система, предложенная Майклом Стоунбрейкером. Схематично данная классификация изображена на рисунке П8.1. Здесь P обозначает процессор, M – модуль оперативной памяти, D – дисковое устройство, N – соединительную сеть.



Рисунок П8.1 - Классификация Стоунбрейкера.

В соответствии с классификацией Стоунбрейкера параллельные системы баз данных могут быть разделены на следующие три базовых класса в зависимости от способа распределения аппаратных ресурсов:

a) SE (Shared-Everything) – архитектура с разделяемыми памятью и дисками;

б) SD (Shared-Disks) – архитектура с разделяемыми дисками;

в) SN (Shared-Nothing) – архитектура без совместного использования ресурсов.

# Приложение 8. Расчёт числа записей в соединении нескольких таблиц

1. Число записей в соединении двух подзапросов Q1 и Q2.

,

T(Qi)= pi·T(Ri) – число записей в i-ом подзапросе, T(Ri) – число записей в исходной таблице Ri (см. Приложение 1), pi – вероятность, что запись исходной таблицы удовлетворяет условию поиска по этой таблице (см. курс ТОРА),

 I(Qi,a) = min(T(Qi), I(Ri,a)) – мощность атрибута соединения "а" в таблице Qi ; I(Ri,a) - мощность атрибута соединения "а" в исходной таблице Ri (см. п. 4.2.3 в файле «tpch2.17.1.pdf»)

2. Мощности атрибутов в соединении H1:

а) мощность атрибута соединения ("а") в результирующей таблице

;

б) мощности остальных атрибутов (b):

пусть *bi* – атрибут таблицы *Qi*, *i*= 1, 2, тогда

1) если $T\left(Q\_{1}⊳⊲Q\_{2}\right)\geq $ $T(Q\_{i}),$ то $I\left(Q\_{1}⊳⊲Q\_{2},b\_{i}\right)=$*I(*$Q\_{i},b\_{i}),$

2) если $T\left(Q\_{1}⊳⊲Q\_{2}\right)<$ $T(Q\_{i}),$ то $I\left(Q\_{1}⊳⊲Q\_{2},b\_{i}\right)=T\left(Q\_{1}⊳⊲Q\_{2}\right)\*I\left(Q\_{i},b\_{i}\right)/T(Q\_{i})$.

3. По приведённым выше формулам можно оценить характеристики соединений , , …, .

4. Если выполняется группирование Group by (c1,…ck), число записей в результирующей таблице можно оценить по формуле:

 

**Примечания.**

1. Расчёт числа записей выполнять в соответствии с выданным вариантом в следующей последовательности:

Q21:

P $⊳⊲$ PS$⊳⊲S⊳⊲N⊳⊲R$, group by,

Q41:

L $⊳⊲$ O, group by,

Q71:

N1xN2(декартово произведение)$⊳⊲$S$⊳⊲$L$⊳⊲$O$⊳⊲$C(по двум атрибутам),

Q81:

P$⊳⊲$L$⊳⊲$ S$⊳⊲$ O$⊳⊲$ C$⊳⊲$ N1$⊳⊲$ R$⊳⊲$ N2

Q91:

S $⊳⊲$L$⊳⊲$ P$⊳⊲$ O$⊳⊲$ N$⊳⊲$ PS

2. В КСАМ описание соединения таблиц можно выполнять в этом или в другом порядке.

# Приложение 9. Расшифровка csv-файла

По результатам моделирования выгружается csv-файл, который содержит информацию расчета.

По столбцам:

**Serial** – порядковый номер вычисления. Количество вычислений зависит от варьируемых параметров, указанных при выполнении вычислений (пункт «Параметры расчётов»).

для online**: \_\_changed\_[*имя узла*]\_[*поле*]** - варьируемый параметр для узла типа PC, поле=NodeCount – количество машин.

для offline: **\_\_changed\_[*имя обращения к транзакции*]\_[*поле*]** - варьируемый параметр для обращения к транзакции, поле=Frequency – число пакетов (частота), выполняемых параллельно.

**\_\_db\_[*имя физической БД*]\_[*имя запроса*]\_time** – время выполнения запроса (с.) на одном ядре процессора.

**\_\_[*имя транзакции*]\_disc-p** – загрузка диска кластера.

**\_\_[*имя транзакции*]\_** **proc-p** – загрузка одного процессора кластера.

**\_\_[*имя транзакции*]\_** **net-p** – загрузка сети, связывающей кластер и узел обращения.

**\_\_[*имя транзакции*]\_size** – размер результата выполнения транзакции в байтах (для режима online – это объем данных, который передаётся по сети на одну рабочую станции PC).

Расчеты характеристик транзакции производятся с точки зрения обращения к транзакции, поэтому для обращения типа offline загрузка устройств кластера нулевая.

 Для обращения к транзакции типа online загрузка каждого устройства кластера не должна превышать 1.

**Примечание**: Если вы открыли csv в Microsoft Excel и все столбцы оказались в одном, то выполните следующие действия в меню:

Данные -> Текст по столбцам -> Выбрать с разделителями «запятая» -> Дополнительно десятичный разделитель «.»

# Приложение 10. Интерфейс редактирования модели



# Приложение 11. Титульный лист и задание.

|  |  |
| --- | --- |
|  | **Министерство образования и науки Российской Федерации****Федеральное государственное бюджетное образовательное учреждение** **высшего образования****«Московский государственный технический университет****имени Н.Э. Баумана****(национальный исследовательский университет)»****(МГТУ им. Н.Э. Баумана)** |

ФАКУЛЬТЕТ «Информатика и системы управления» (ИУ)

КАФЕДРА «Системы обработки информации и управления» (ИУ5)

**ОТЧЕТ ПО ПРОЕКТНО-ТЕХНОЛОГИЧЕСКОЙ ПРАКТИКЕ**

Студент: Фамилия Имя Отчество

Группа: ИУ5-ххМ

Тип практики: учебная

Название предприятия: МГТУ им. Н.Э.Баумана

Студент **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ ФИО**

 *подпись, дата*

Руководитель практики **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Плужникова О.Ю.**

 *подпись, дата*

Оценка \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*2021 г.*

|  |  |
| --- | --- |
|  | **Министерство образования и науки Российской Федерации****Федеральное государственное бюджетное образовательное учреждение** **высшего образования****«Московский государственный технический университет****имени Н.Э. Баумана****(национальный исследовательский университет)»****(МГТУ им. Н.Э. Баумана)** |

ФАКУЛЬТЕТ «Информатика и системы управления» (ИУ)

КАФЕДРА «Системы обработки информации и управления» (ИУ5)

**ИНДИВИДУАЛЬНОЕ ЗАДАНИЕ НА ПРОЕКТНО-ТЕХНОЛОГИЧЕСКУЮ ПРАКТИКУ**

Тип практики: учебная

Название предприятия: МГТУ им. Н.Э.Баумана

Продолжительность практики: 72 академических часов

Направление подготовки: 09.04.01 «Информатика и вычислительная техника»

***Задание***

Проанализировать временные характеристики функционирования информационной системы (ИС) на основе базы данных на этапе её проектирования (на примере аналитических запросов TPC-H к хранилищу данных) с помощью пакета КСАМ 2.0 и выбрать оптимальный вариант аппаратного комплекса.

Студент **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ ФИО**

 *подпись, дата*

Руководитель практики **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Плужникова О.Ю.**

 *подпись, дата*