

БАЗЫ ДАННЫХ

часть II

Распределенные и
параллельные системы
управления базами данных

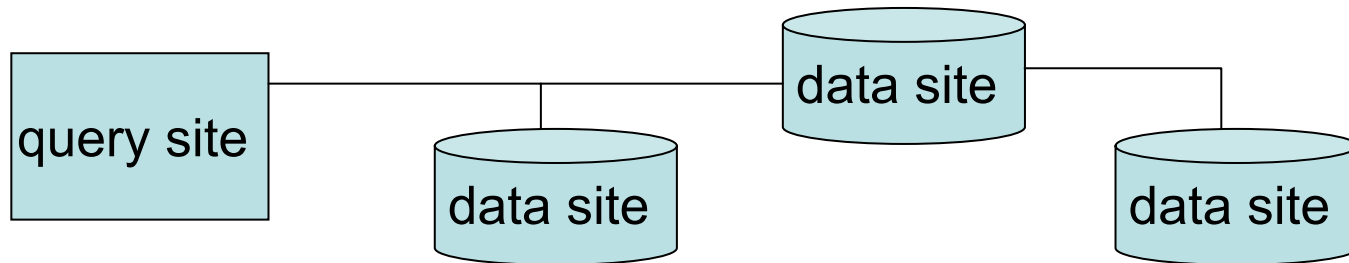
Распределенные и параллельные СУБД

Распределенная база данных (DDB - distributed database) - это совокупность множества взаимосвязанных баз данных, распределенных в компьютерной сети.

Система управления распределенной базой данных определяется как программная система, которая позволяет управлять базой данных таким образом, чтобы ее распределенность была прозрачна для пользователей

Распределенные и параллельные СУБД

Важнейший отличительный признак - слабосвязанный характер среды, где каждый узел имеет собственную операционную систему и функционирует независимо.



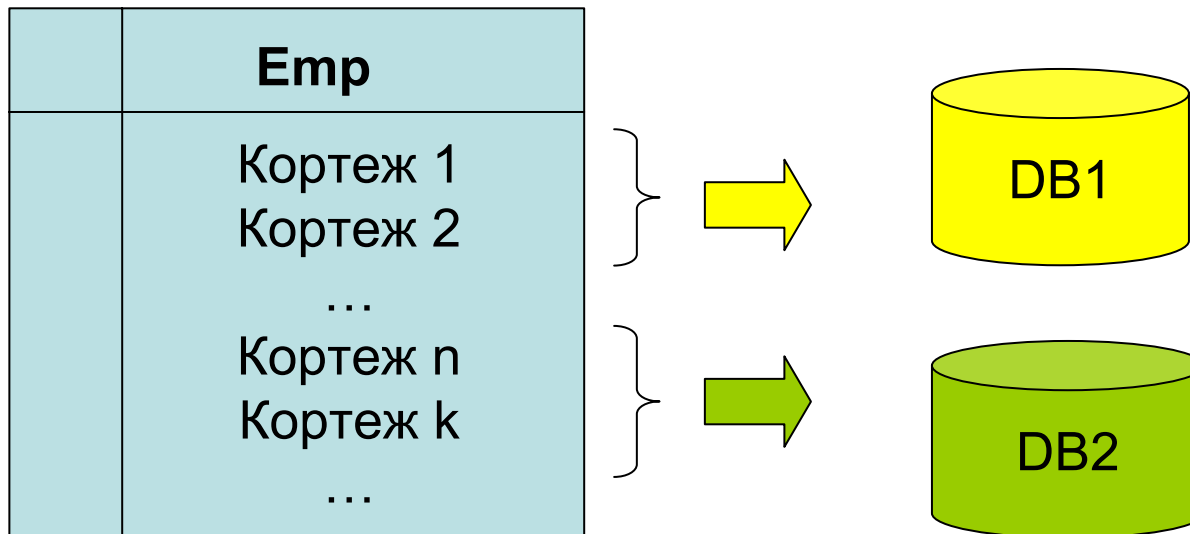
Распределенные и параллельные СУБД

База данных физически распределяется по узлам данных при помощи **фрагментации** и **репликации**, или **тиражирования** данных.

Отношения, принадлежащие реляционной базе данных, могут быть фрагментированы на *горизонтальные* или *вертикальные* разделы.

Распределенные и параллельные СУБД

Горизонтальная фрагментация реализуется при помощи операции селекции, которая направляет каждый кортеж отношения в один из разделов, руководствуясь предикатом фрагментации.



Распределенные и параллельные СУБД

Параллельную СУБД можно определить как реализацию СУБД для многопроцессорного компьютера.

Решение заключается в применении широкомасштабного параллелизма, чтобы усилить мощность отдельных компонентов путем их интеграции в целостную систему на основе соответствующего программного обеспечения параллельных баз данных.

Распределенные и параллельные СУБД

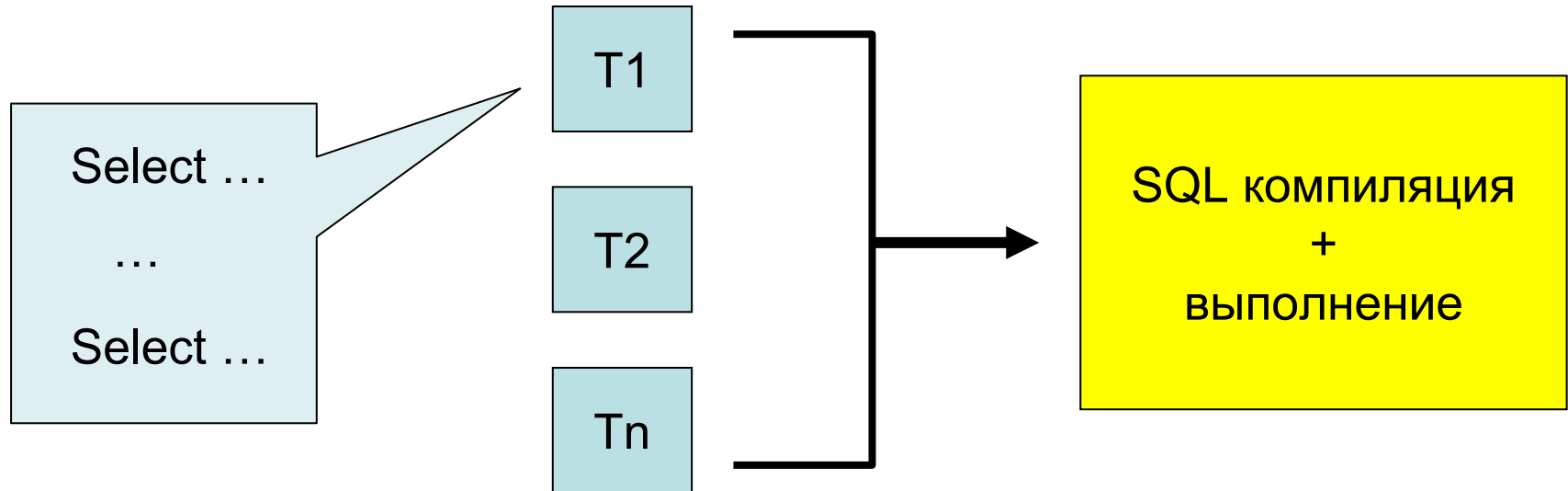
В программном обеспечении базы данных могут быть предусмотрены три вида параллелизма, присущие приложениям интенсивной обработки данных:

- **межзапросный параллелизм**
- **внутризапросный параллелизм**
- **внутриоперационный параллелизм**

Распределенные и параллельные СУБД

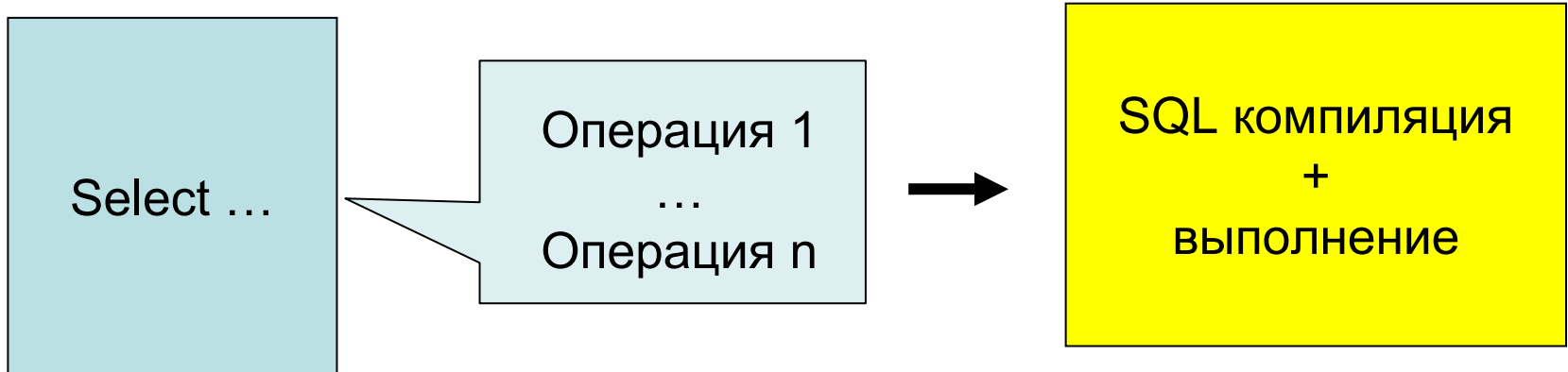
Межзапросный параллелизм

предполагает одновременное выполнение множества запросов, относящихся к разным транзакциям.



Распределенные и параллельные СУБД

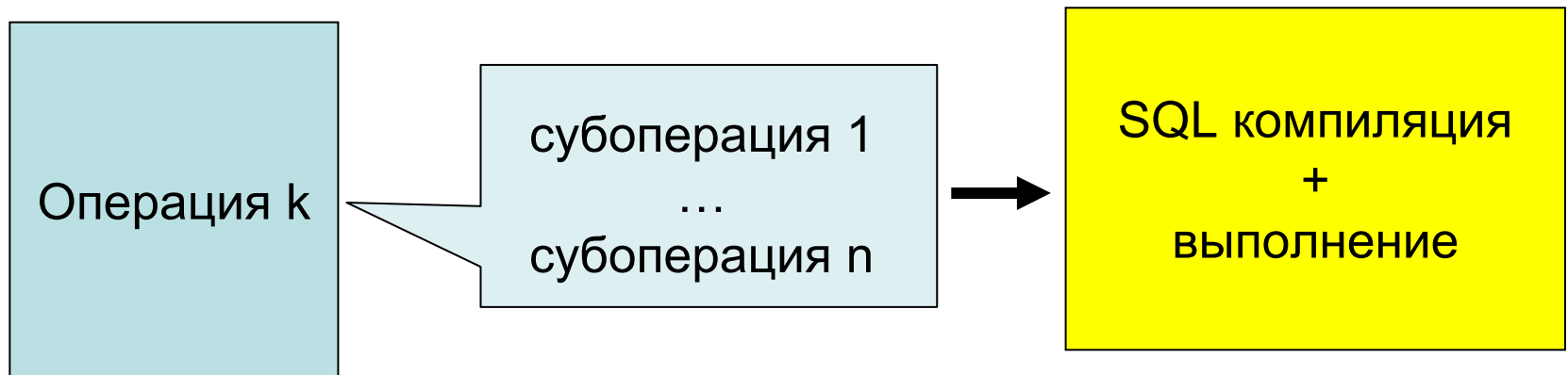
Под **внутризапросным параллелизмом** понимается одновременное выполнение сразу нескольких операций (например операций выборки), относящихся к одному и тому же запросу.



И внутризапросный, и межзапросный параллелизм реализуется на основе *разбиения данных* по типу горизонтальной фрагментации.

Распределенные и параллельные СУБД

Понятие **внутриоперационного параллелизма** означает параллельное выполнение одной операции в виде множества субопераций с применением, в дополнение к фрагментации данных, также *фрагментации функций*.



Распределенные и параллельные СУБД

Идентифицирующие характеристики параллельных и распределенных СУБД:

1. Распределенная/параллельная база данных - это именно база данных, а не "коллекция" файлов, индивидуально хранимых на разных узлах сети.
2. Система обладает полной функциональностью СУБД. Она не сводится по своим возможностям ни к распределенным файловым системам, ни к системам обработки транзакций.
3. Распределение (включая фрагментацию и репликацию) данных по множеству узлов невидимо для пользователей (**прозрачность**).

Распределенные и параллельные СУБД

Технология распределенных/параллельных баз данных распространяет основополагающую для управления базами данных концепцию *независимости данных* на среду, где данные распределены и тиражированы по множеству компьютеров, связанных сетью.

Это обеспечивается за счет нескольких видов прозрачности: *прозрачность сети* (следовательно, *прозрачность распределения*), *прозрачность репликации* и *прозрачность фрагментации*.

Распределенные и параллельные СУБД

Прозрачность доступа означает, что пользователи имеют дело с единым логическим образом базы данных и осуществляют доступ к распределенным данным точно так же, как если бы они хранились централизованно.

В идеале полная прозрачность подразумевает наличие языка запросов к распределенной СУБД, не отличающегося от языка для централизованной СУБД.

Распределенные и параллельные СУБД

Высокая производительность - одна из важнейших целей, на достижение которой направлены технологии параллельных СУБД. Как правило, она обеспечивается за счет сочетания нескольких взаимно дополняющих решений, таких как применение операционных систем, ориентированных на поддержку баз данных, параллелизм, оптимизация, баланс загрузки.

Наличие операционной системы, "осведомленной" о специфических потребностях баз данных, упрощает реализацию функций баз данных нижнего уровня и способствует снижению их стоимости.

Затраты на передачу сообщения могут быть значительно снижены за счет применения специализированного коммуникационного протокола.

Распределенные и параллельные СУБД

Технологии распределенных и параллельных СУБД направлены также на повышение надежности, поскольку благодаря репликациям данных исключаются одиночные точки отказа.

Отказ одного узла или сбой на линии связи не приводит к выходу из строя всей системы. Даже если часть данных становится недоступной, при правильной организации системы пользователи могут иметь доступ к остальной части информации.

Распределенные и параллельные СУБД

В идеале параллельная (и, в меньшей степени, распределенная) СУБД обладает свойством **линейной расширяемости** и **линейного ускорения**.

Распределенные и параллельные СУБД

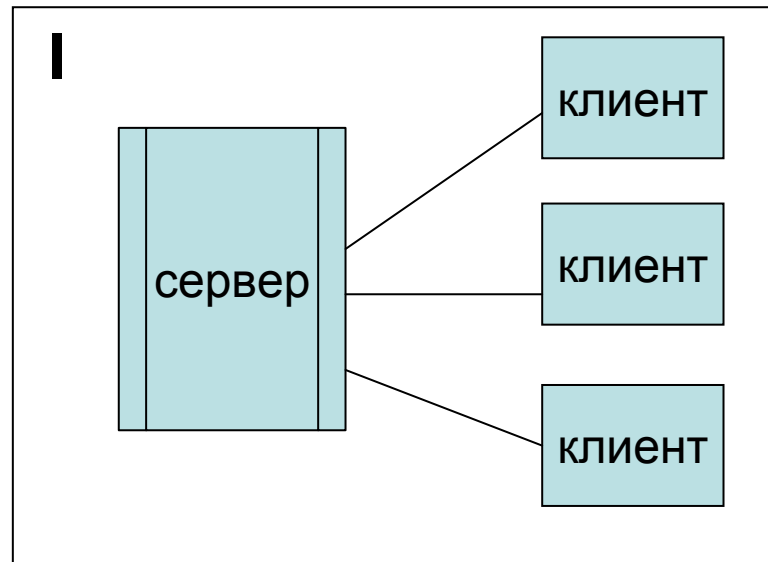
Под **линейной расширяемостью** понимается сохранение того же уровня производительности при увеличении размера базы данных и одновременном пропорциональном увеличении процессорной мощности и объема памяти.

Распределенные и параллельные СУБД

Линейное ускорение означает, что с наращиванием процессорной мощности и объема памяти при сохранении прежнего размера базы данных пропорционально возрастает производительность.

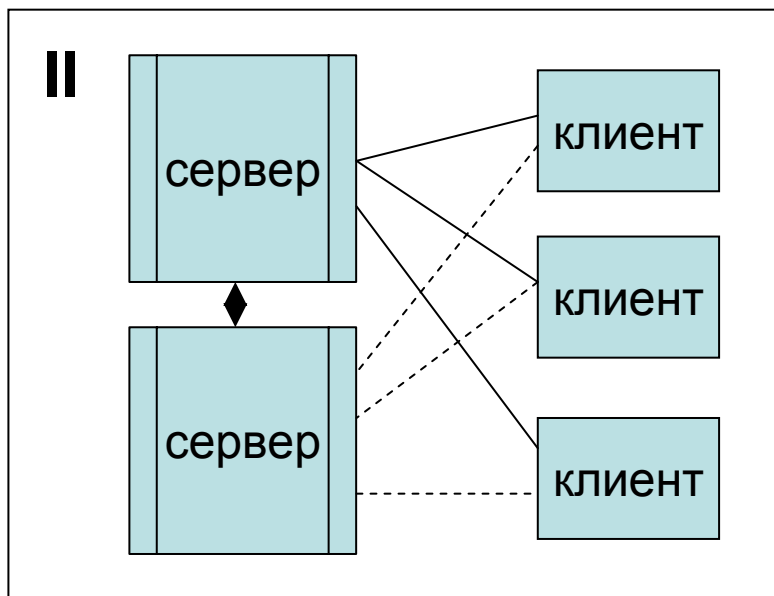
Распределенные и параллельные СУБД

Наиболее популярна в настоящее время архитектура **клиент-сервер**, когда множество машин-клиентов осуществляют доступ к одному серверу баз данных.



Распределенные и параллельные СУБД

Архитектура типа *много-клиентов/много-серверов*, когда база данных размещена на множестве серверов, которым, для того чтобы вычислить результат пользовательского запроса или выполнить транзакцию, необходимо взаимодействовать друг с другом.

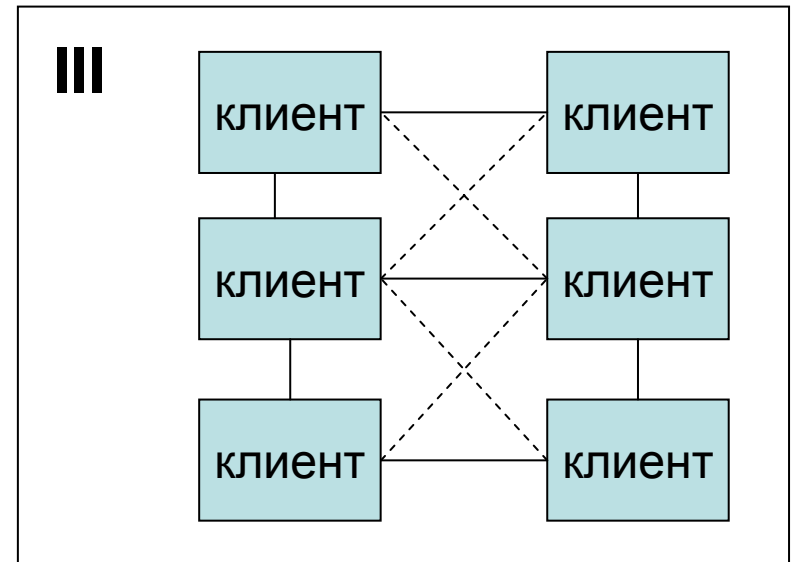


Каждая клиентская машина имеет свой "домашний" сервер; ему она направляет пользовательские запросы. Взаимодействие серверов друг с другом прозрачно для пользователей.

Распределенные и параллельные СУБД

В истинно распределенной СУБД клиентские и серверные машины не различаются. В идеале каждый узел может выступать и как клиент, и как сервер.

Такие архитектуры, тип которых определяют как *равный-к-равному*, требуют сложных протоколов управления данными, распределенными по множеству узлов.



Распределенные и параллельные СУБД

Архитектуры параллельных систем варьируются между двумя крайними точками, называемыми **архитектура без разделяемых ресурсов** и **архитектура с разделяемой памятью**. Промежуточную позицию занимает архитектура с разделяемыми дисками.

Распределенные и параллельные СУБД

В случае **неразделения ресурсов** каждый процессор имеет эксклюзивный доступ к собственной оперативной памяти и к набору дисков.

Архитектуры без разделяемых ресурсов обладают тремя важнейшими **преимуществами**: низкие затраты, расширяемость, высокая доступность

Наиболее существенные **проблемы** - сложность реализации и (потенциальные) трудности соблюдения баланса загрузки.

Распределенные и параллельные СУБД

Подход, основанный на разделении памяти, заключается в том, что каждый процессор посредством быстрых линий связи (высокоскоростных шин или кросс-панельных коммутаторов) соединен со всеми модулями памяти и дисковыми устройствами.

Сильные стороны - простота и хороший баланс загрузки.

Наиболее существенные **проблемы** - стоимость, ограниченная расширяемость, невысокая надежность.

Распределенные и параллельные СУБД

В системах с разделяемыми дисками каждый процессор имеет доступ к любому дисковому устройству посредством специальных соединений и эксклюзивный доступ к своей собственной оперативной памяти.

Преимущества: низкие затраты, расширяемость, хороший баланс загрузки, высокая доступность, простота миграции с однопроцессорных систем.

Трудности: сложность системы, потенциальные проблемы производительности.

ВОПРОСЫ ?

СОВЕТ ДНЯ:

Этапы «большого»
пути:

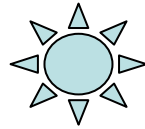
Школа



Человеческое состояние:

Детство...

Поступление в вуз,
первые два курса



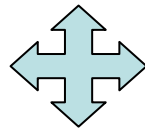
Отрочество...

Третий – пятый курсы



Юность...

Окончание вуза



Скоро пенсия!

Живи – пока молодой!