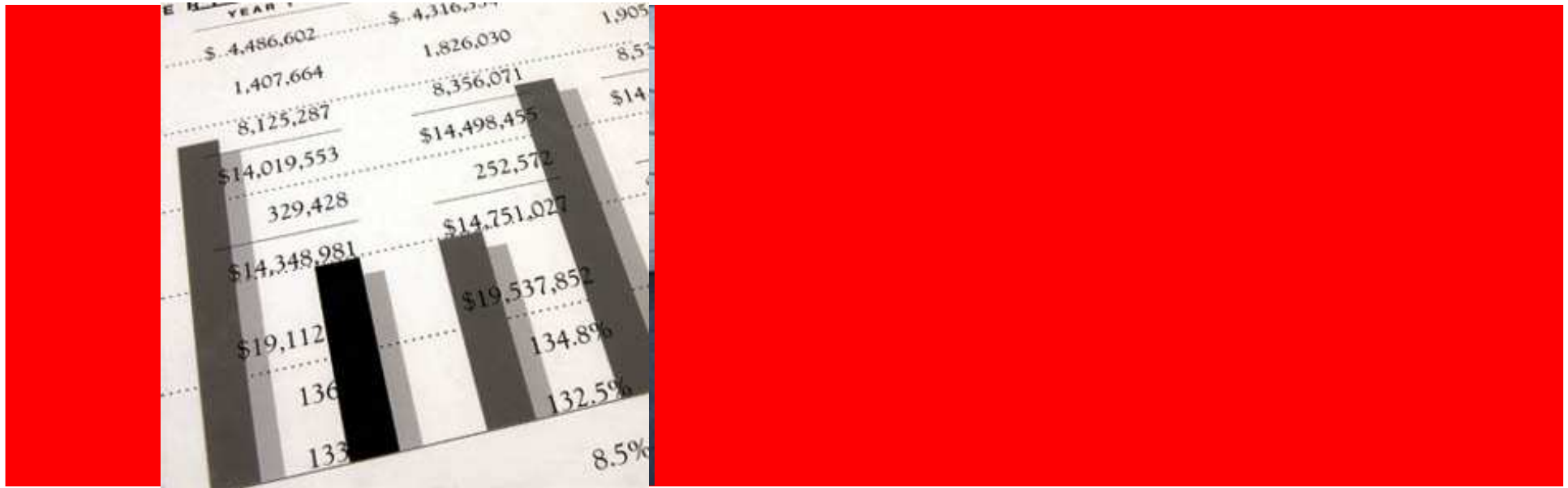


ORACLE®



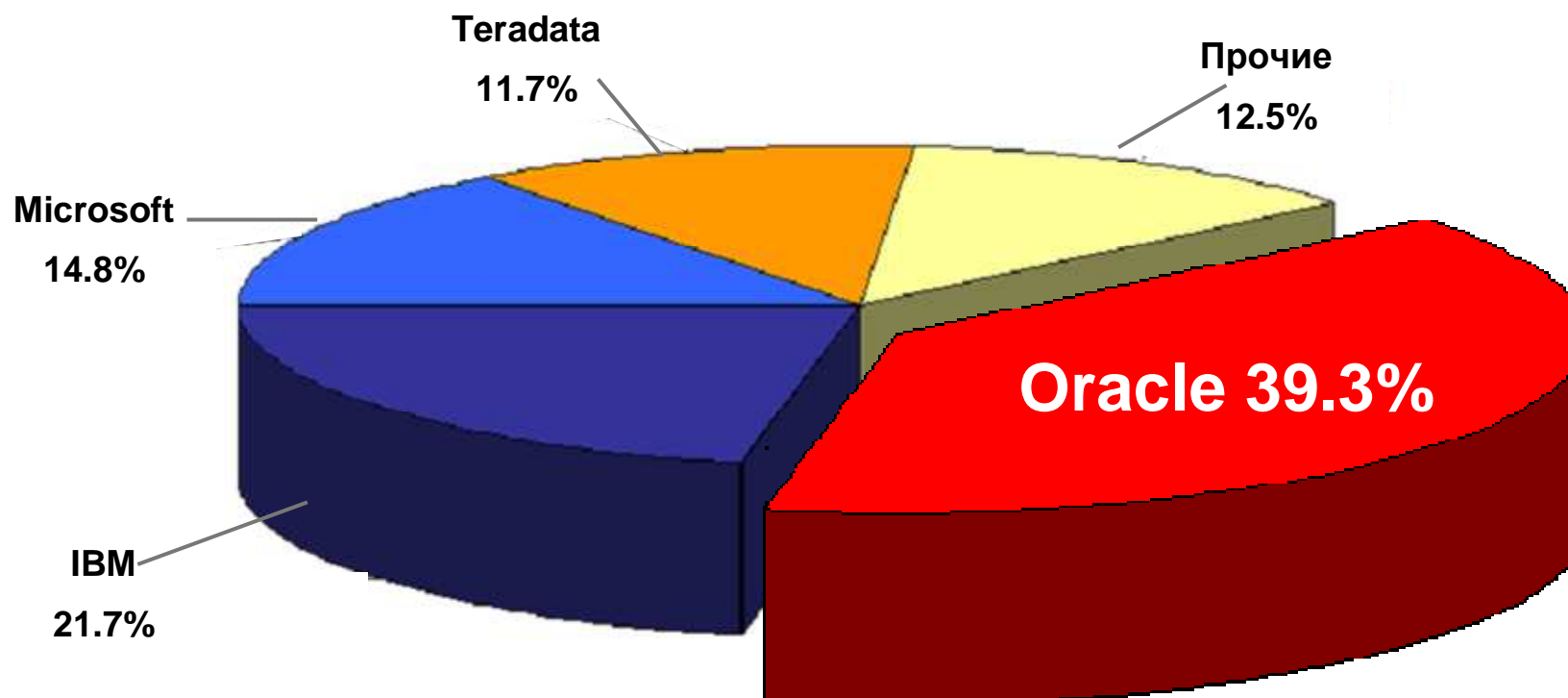
ORACLE®

Oracle Exadata и HP Oracle Database Machine НОВЫЕ ВОЗМОЖНОСТИ для хранилищ данных

Андрей Пивоваров
Менеджер по технологическому консалтингу
Oracle CIS

ORACLE®

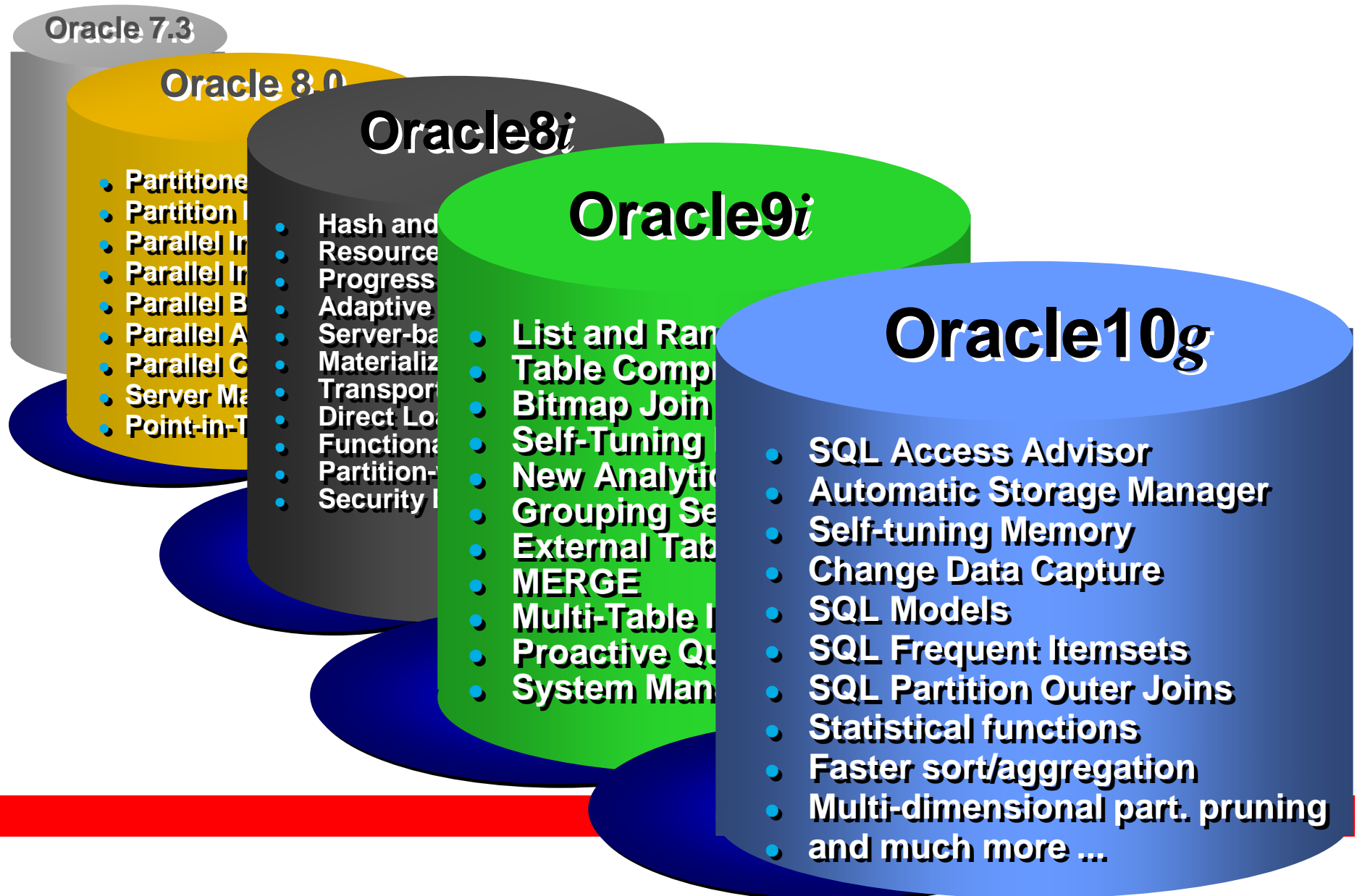
Oracle – лидер в области хранилищ данных



ORACLE

Источник: IDC, Aug 2008 – “Worldwide Data Warehouse Management Tools 2007 Vendor Shares”

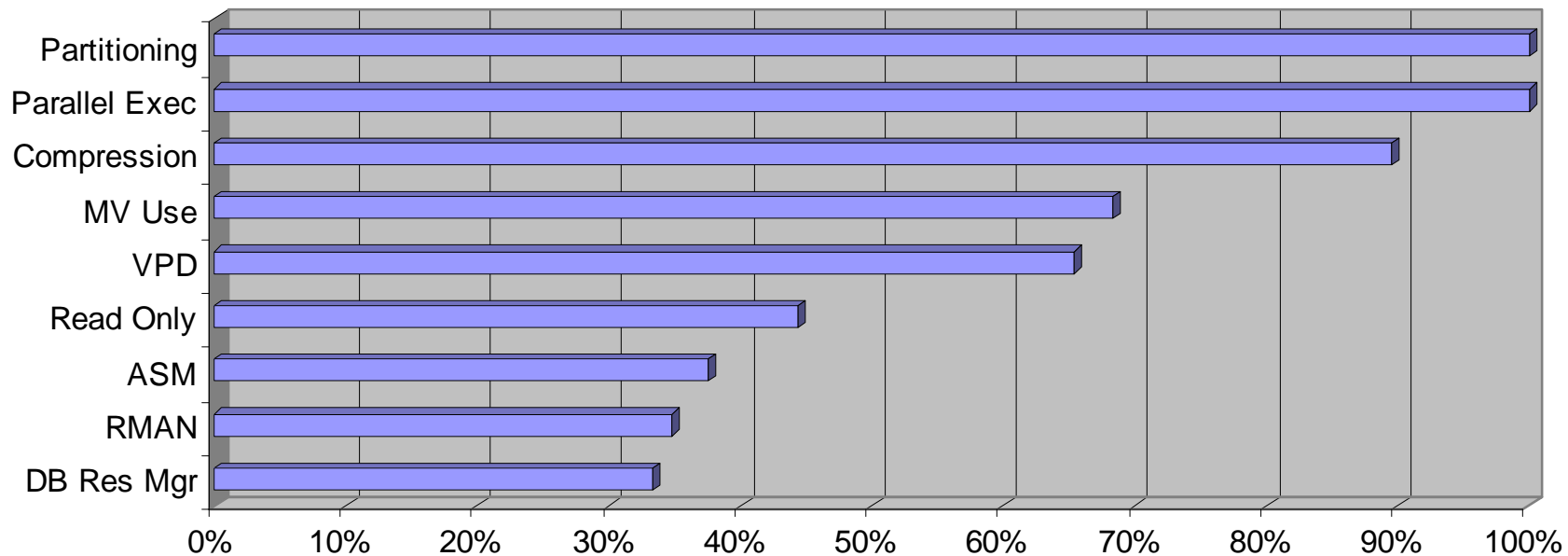
Oracle Database – платформа для DW&BI



НОВЫЕ ВОЗМОЖНОСТИ ДЛЯ ХД В 11g

- VLDB
 - Composite Range-Range
 - Composite List-Range
 - Composite List-List
 - Composite List-Hash
 - REF Partitioning
 - Virtual Column Partitioning
 - Compression enhancements
- Performance
 - Query Result Cache
- Data loading
 - Change data capture enhancements
 - Materialized view refresh enhancements
- Manageability
 - Partition Advisor
 - Interval Partitioning
 - SQL Plan Management
 - Automatic SQL Tuning with Self-Learning Capabilities
 - Enhanced Optimizer Statistics Maintenance
 - Multi-Column Optimizer Statistics
 - ASM Fast Resync, Fast VLDB Startup and other enhancements
- SQL
 - SQL Pivot and Unpivot
 - Continuous Query Notification
- OLAP
 - Materialized view refresh and SQL rewrite
 - Continued database integration
 - Cube metadata in the Data Dictionary
 - Fine-grained data security on cubes
 - Simplified application development
 - Fully declarative cube calculations
 - Cost-Based Aggregation
 - Simpler calculation definitions
- Data Mining
 - Simplified development and deployment of models
 - Supermodels: data preparation combined with mining model
 - Additional packaged predictive analytics
 - Integration in database dictionary
 - New algorithms: “General Linear Models”
 - Encapsulates several widely used analytic methods
 - Multivariate linear regression; logistic regression

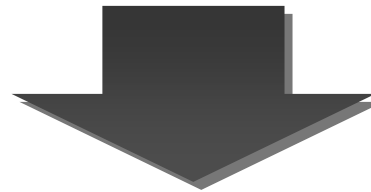
Возможности Oracle, используемые в больших ХД



Query Rewrite (переписывание запросов)

Оптимизатор
автоматически
переписывает
запрос для
обращения к
детальным
данным

```
SELECT p.brand, r.country, t.month,  
       SUM(s.amt) tot_sales  
FROM   sales s, region r,  
       time t, product p  
WHERE  s.region_id = r.region_id  
AND    s.sdate = t.curdate  
AND    s.prod_code = p.prod_code  
GROUP BY p.brand, r.country, t.month  
HAVING SUM(s.amt) > 5000000;
```

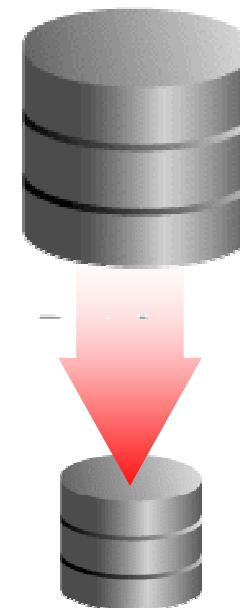


```
SELECT brand, country, month, tot_sales  
FROM   sales_sumry  
WHERE  tot_sales > 5000000;
```

Сжатие (Compression)

- Таблицы и индексы могут быть сжаты
 - Можно управлять даже на уровне партиций
 - Обычное сжатие до 3:1
- Требуют больше процессорного времени при загрузке
 - Распаковка почти не занимает лишнего времени
- Данные занимают меньше места на диске
 - И требуют меньше времени на чтение
- Полная прозрачность

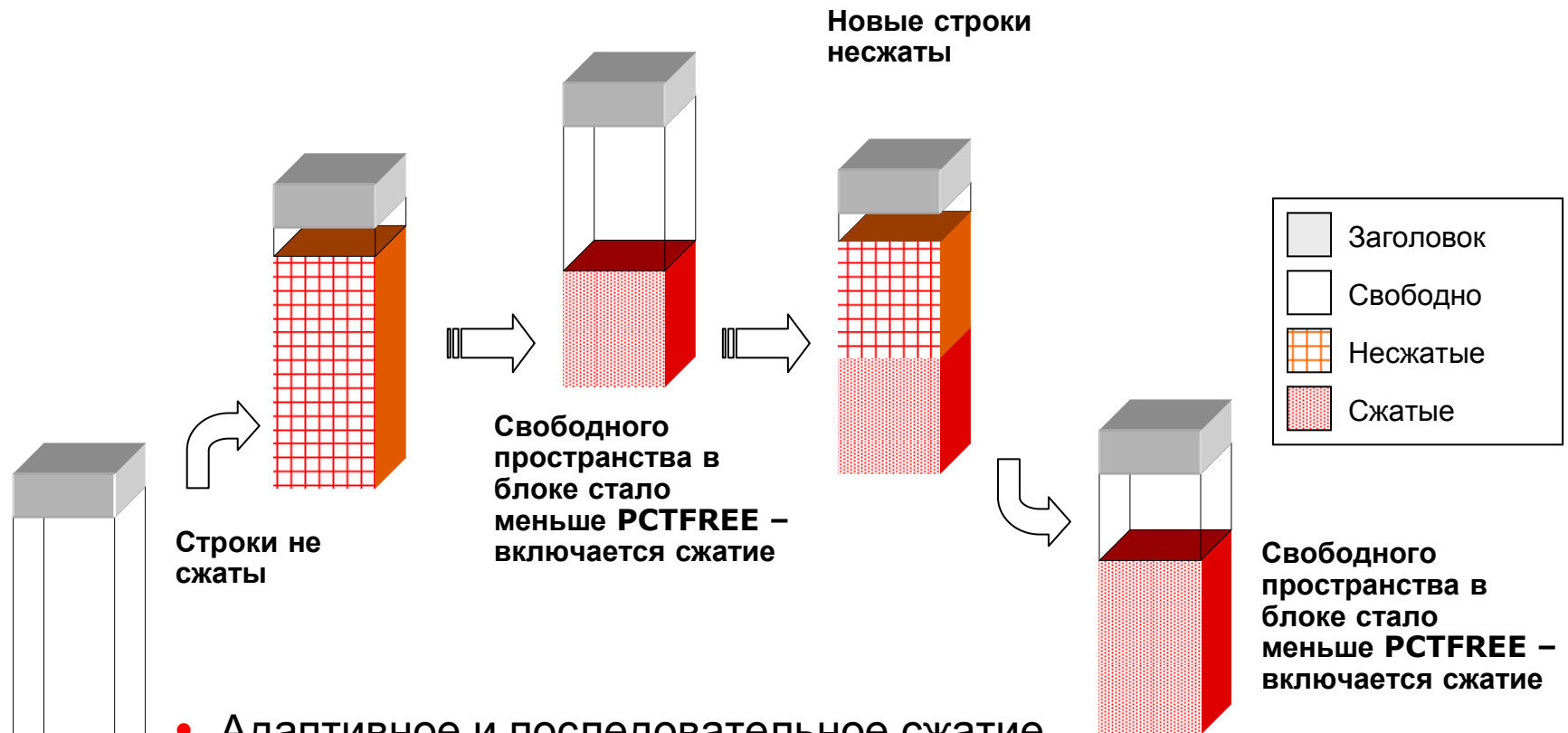
ORACLE[®] 11g
DATABASE



3X

Compression

OLTP Table Compression



- Адаптивное и последовательное сжатие
- Компрессия автоматически включается только когда свободного пространства в блоке станет меньше PCTFREE
- Сжатие устраняет “дырки” после операций delete и максимизируют непрерывность данных в блоке

Кэш результатов SQL запросов

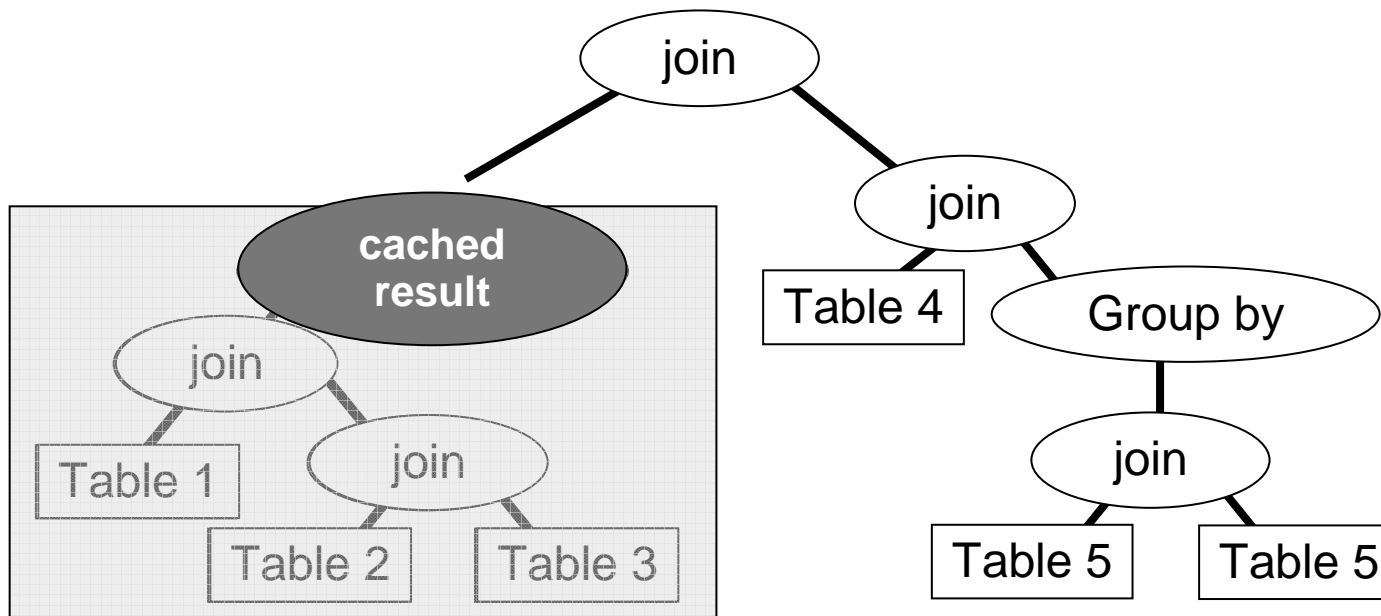
- Хранит результат запросов
 - При повторном запросе результат берется из кэша
- Удобно для ускорения запросов к хранилищу
 - Долгих запросов
 - Вычислительно сложных
 - Возвращающих немного записей

Id	Operation	Name
0	SELECT STATEMENT	
1	RESULT CACHE	fz6cm4jbpcwh48wcyk60m7qypu
2	SORT GROUP BY ROLLUP	
* 3	HASH JOIN	

etc.

SQL Query Result Cache

- Кэшируются результаты queries, query blocks, или pl/sql function calls
- Read consistency поддерживается
 - Результаты сбрасываются из кэша, если участвующие в запросе объекты были изменены с помощью DML/DDDL
- Результат хранится в кэше вместе с bind переменными



11g Cube Fact View

```
DESC units_cube_view;
```

Name	Null	Type		
-----	-----	-----		
TIME		VARCHAR2(100)	}	
CHANNEL		VARCHAR2(100)		Keys
CUSTOMER		VARCHAR2(100)		
PRODUCT		VARCHAR2(100)		
UNITS		NUMBER	}	
SALES		NUMBER		Stored facts
COST		NUMBER		
SALES_PRIOR_PERIOD		NUMBER	}	
SALES_DIFF_PRIOR_PERIOD		NUMBER		Calculated facts
SALES_PCT_DIFF_PRIOR_PERIOD		NUMBER		

11g Cube Fact View

```
SELECT time,  
       product,  
       customer,  
       channel,  
       units  
FROM units_cube_view  
WHERE rownum < 15;
```

TIME	PRODUCT	CUSTOMER	CHANNEL	UNITS
TOTAL	TOTAL	TOTAL	TOTAL	4000968
CY1999	TOTAL	TOTAL	TOTAL	330425
CY2003	TOTAL	TOTAL	TOTAL	534069
CY1998	TOTAL	TOTAL	TOTAL	253816
CY2005	TOTAL	TOTAL	TOTAL	565718
CY2006	TOTAL	TOTAL	TOTAL	584929
CY2004	TOTAL	TOTAL	TOTAL	587419
CY2000	TOTAL	TOTAL	TOTAL	364233
CY2002	TOTAL	TOTAL	TOTAL	364965
CY2001	TOTAL	TOTAL	TOTAL	415394
CY2000.Q1	TOTAL	TOTAL	TOTAL	88484
CY2001.Q2	TOTAL	TOTAL	TOTAL	97346
CY2001.Q3	TOTAL	TOTAL	TOTAL	105704
CY2005.Q3	TOTAL	TOTAL	TOTAL	138953

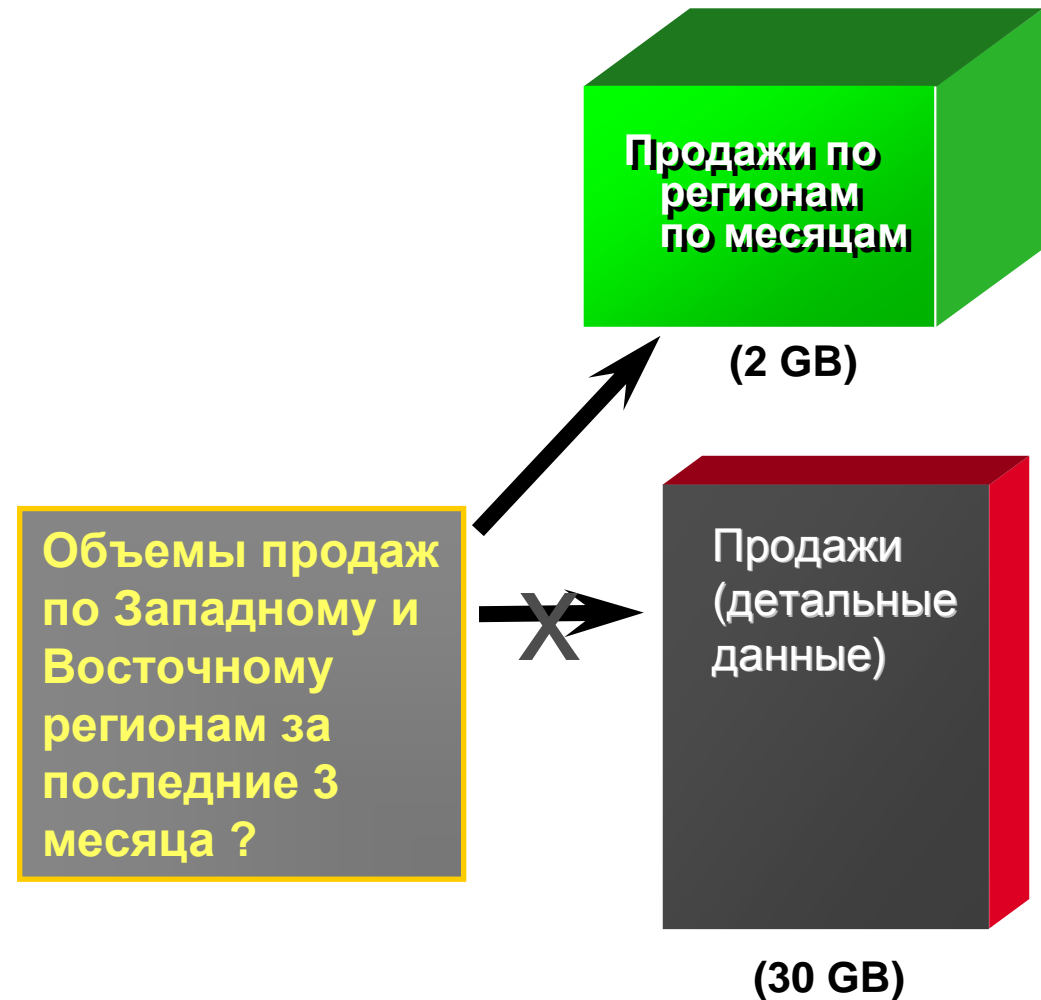
Агрегаты по годам

Агрегаты по кварталам

Итого по продуктам и др.
измерениям

Материализованные представления

- Хранение суммарных данных
- Использование материализованных представлений при автоматической оптимизации запросов
- Средства модификации запросов
- Различные методы обновления (полное, частичное и отложенное)



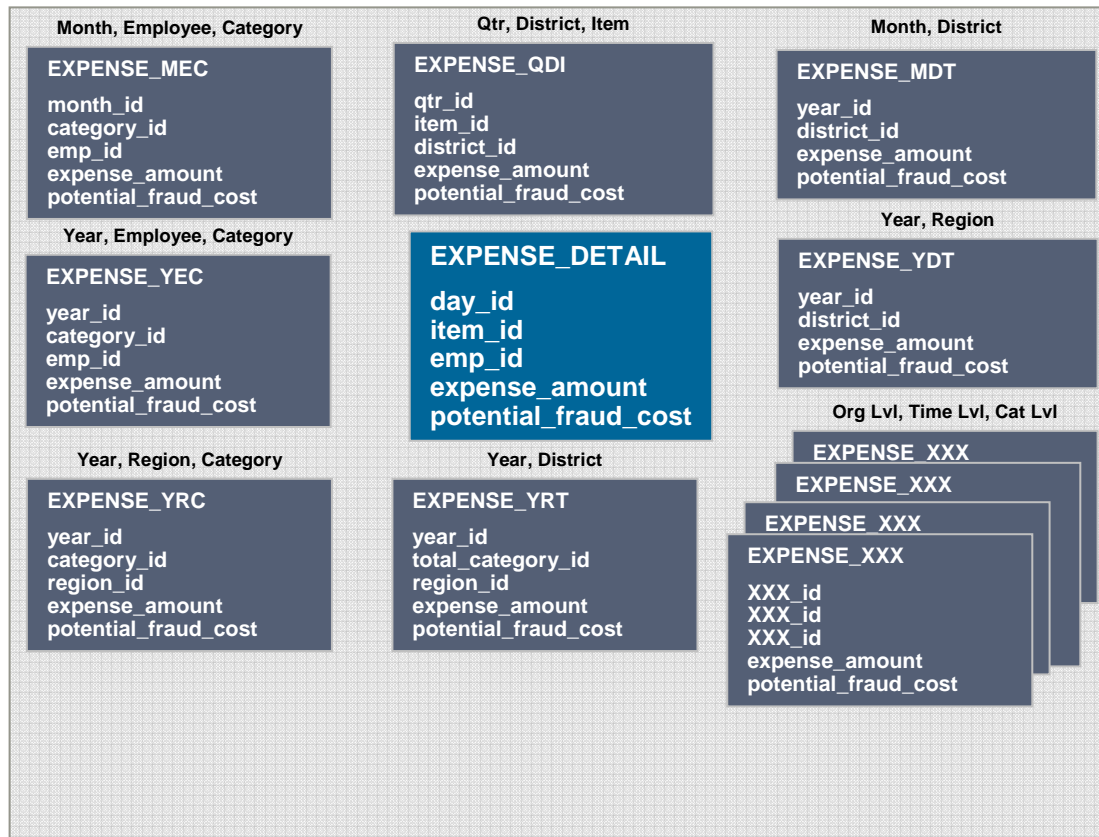
Materialized Views

COGNOS

MicroStrategy

ORACLE

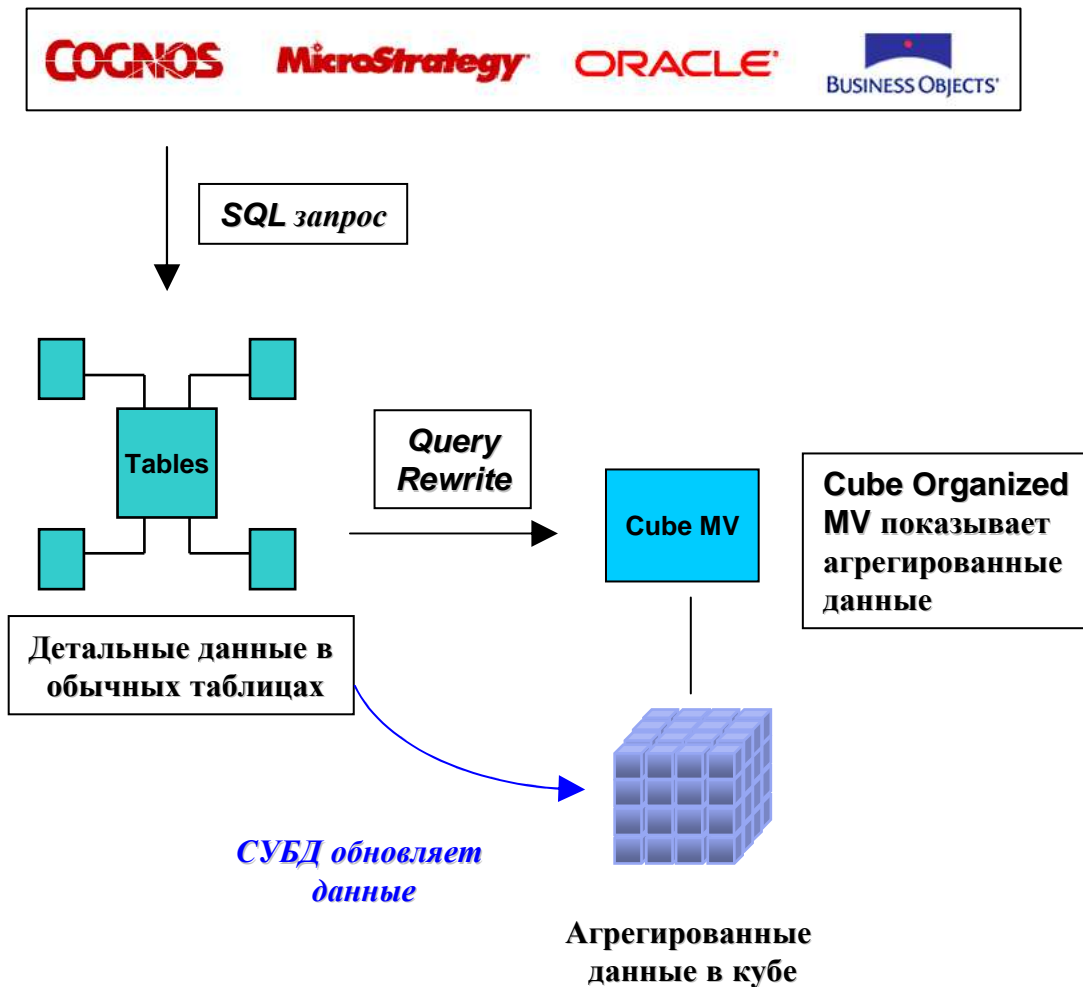
BUSINESS OBJECTS



- Создание MV при нерегламентированных запросах - проблема
- Пользователи хотят ровный отклик
- Появление большого количества MV

ORACLE

11g Cube-Organized Materialized Views



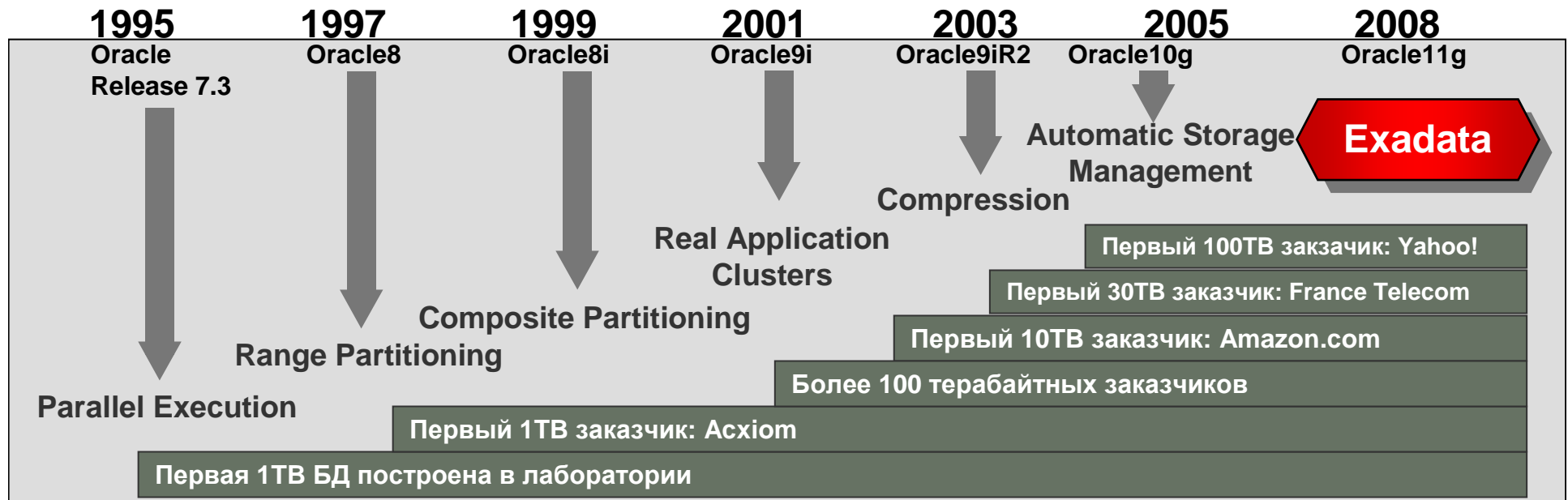
- Детальные данные в реляционных таблицах
- Приложение генерирует SQL запрос
- Агрегированные данные в OLAP кубе
- Доступ к агрегированным данным в кубе – автоматический, через query rewrite
- Приложение не меняется, но запросы работают быстрее

HP Oracle Exadata и Database Machine

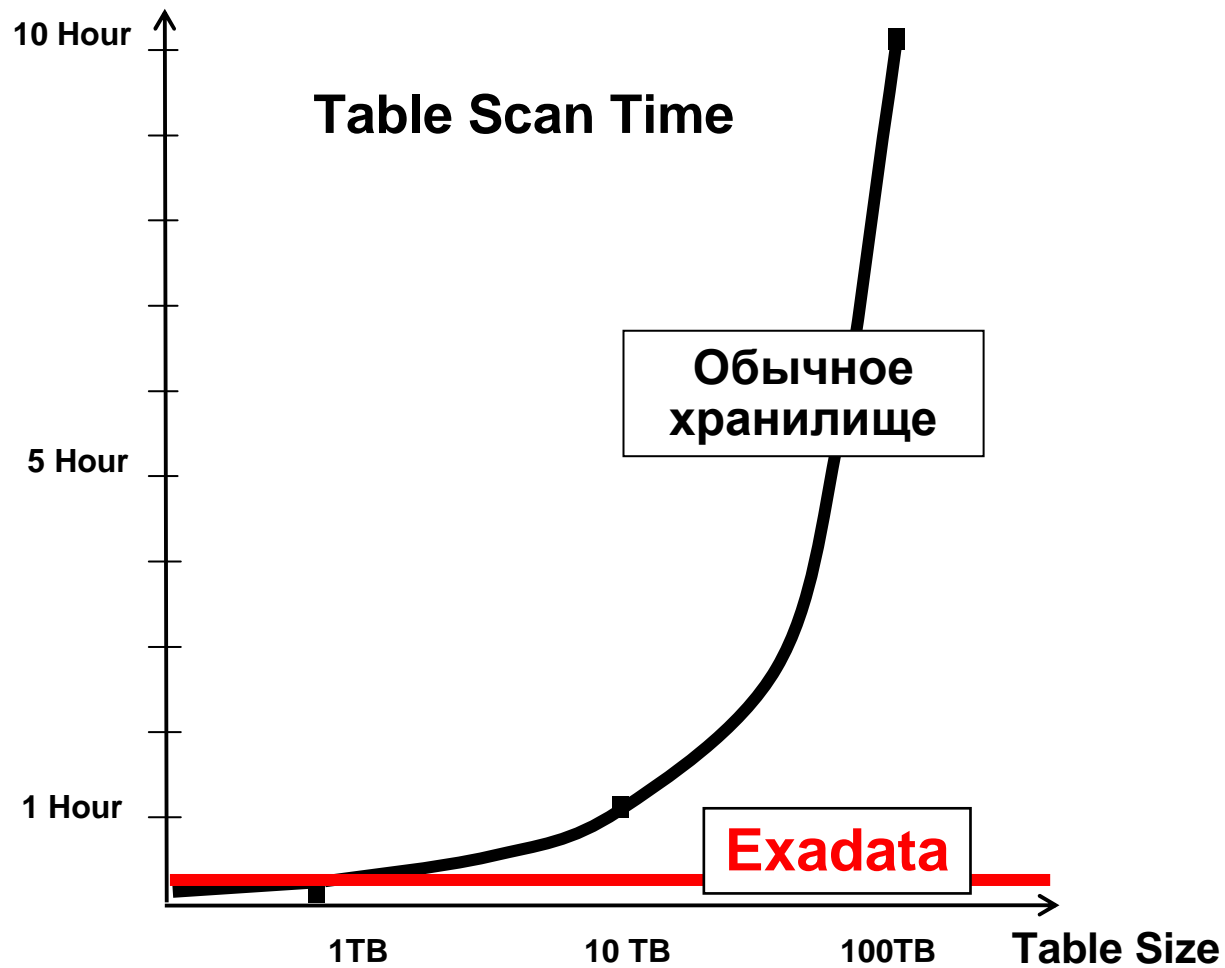


ORACLE

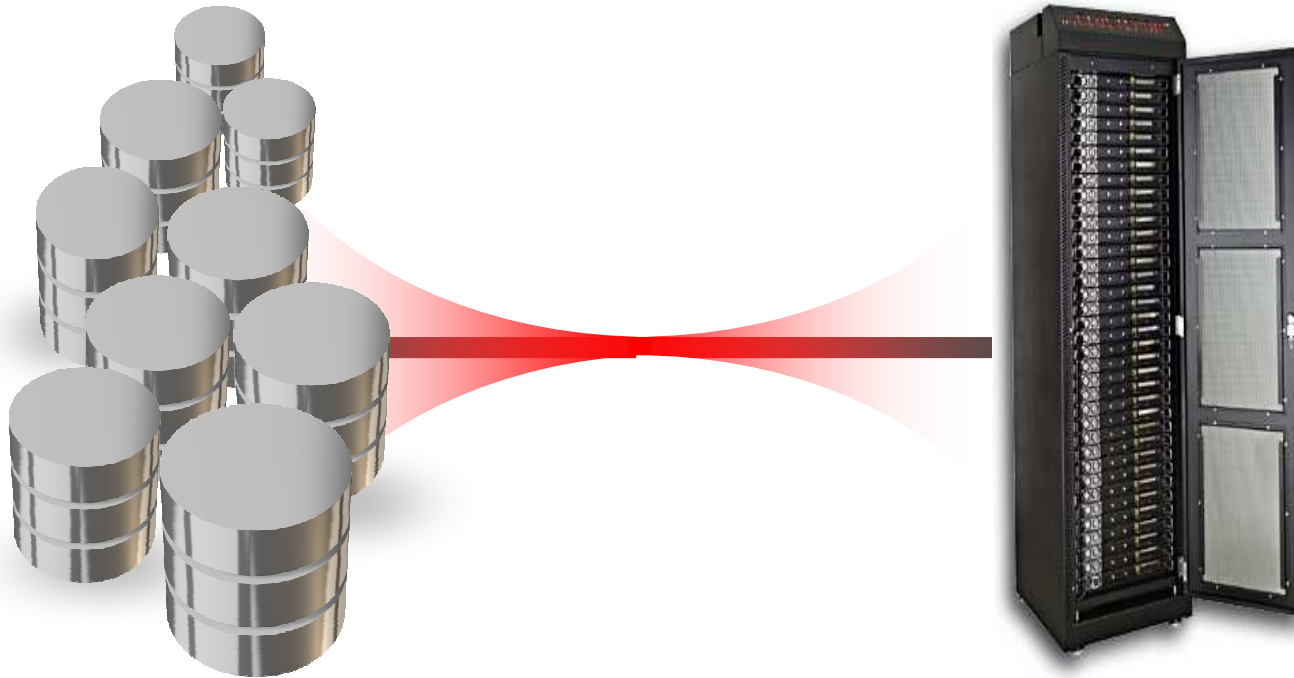
Поддержка сверхбольших хранилищ данных



Время доступа к сверхбольшим ХД

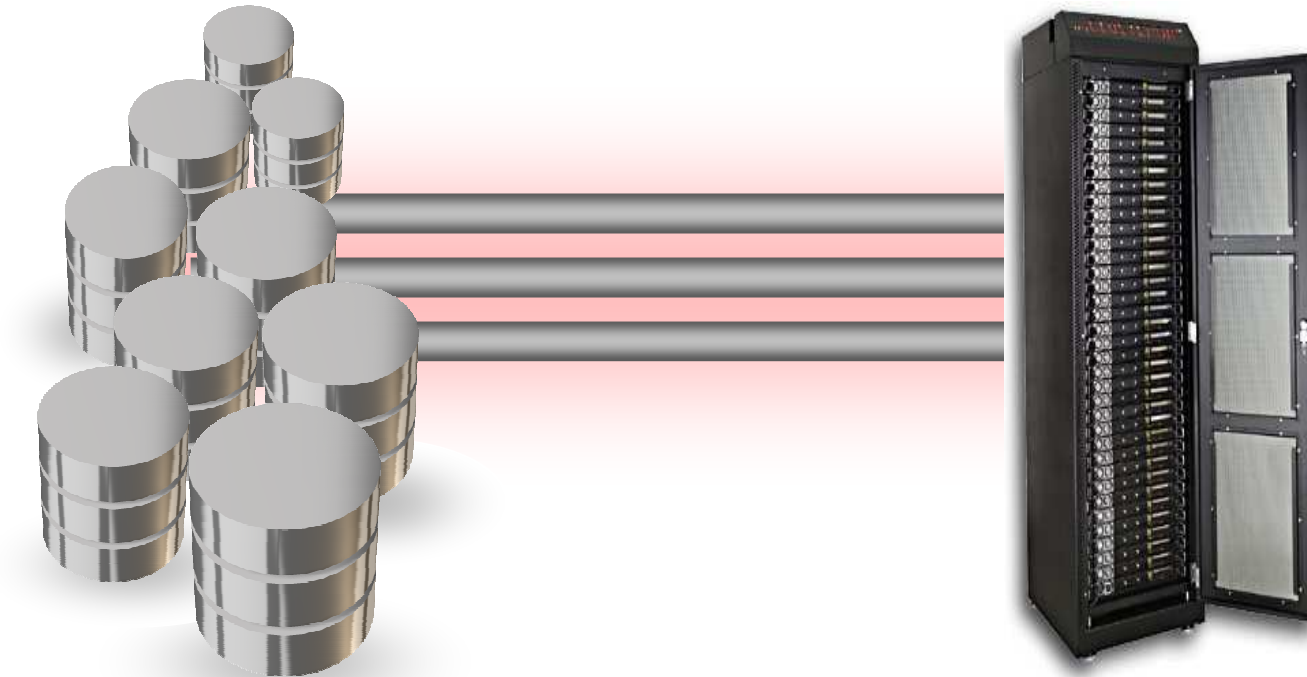


Проблема производительности сверхбольших ХД



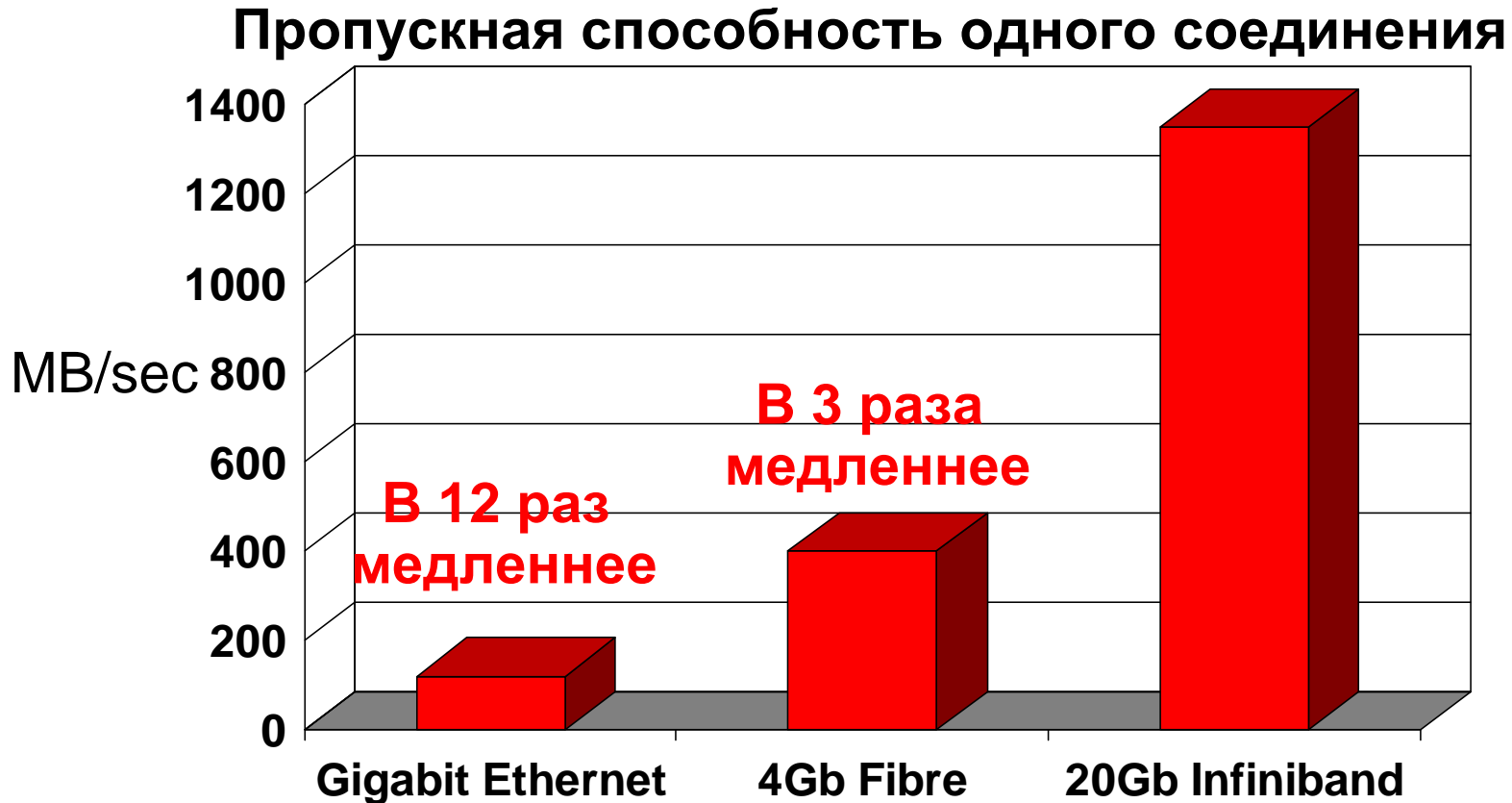
- Большие хранилища сканируют десятки, сотни и тысячи дисков
- Соединения между дисками и серверами ограничивают скорость передачи данных в десятки и сотни раз
- В результате хранилища становятся медленнее по мере роста

Решение проблемы



- Прокачивать меньше данных
- Увеличить количество каналов связи
- Сделать каналы связи шире

Пропускная способность Infiniband

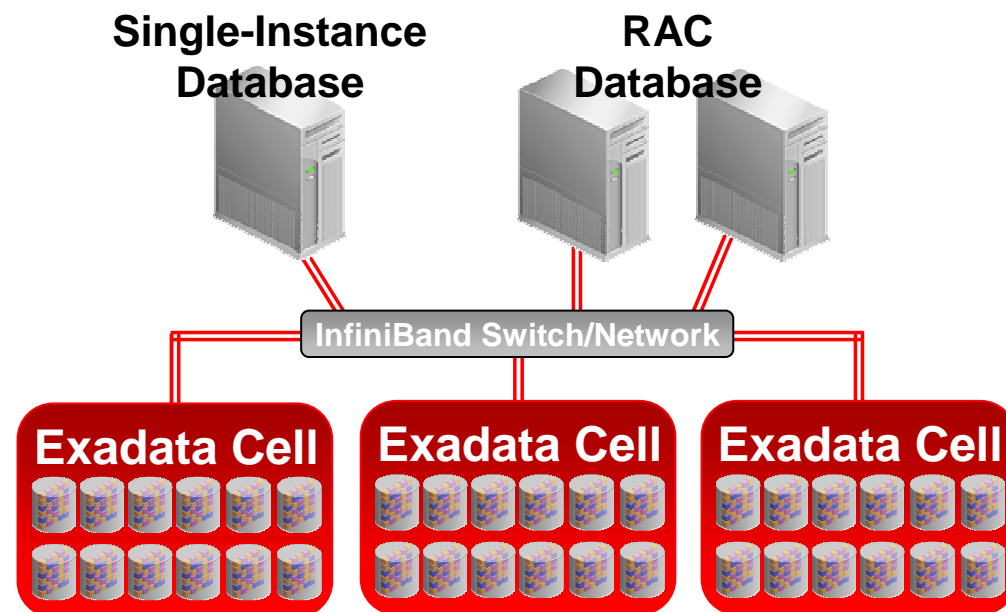


Exadata – новая архитектура

- Exadata передает меньше данных
 - Обработка запросов перенесена на систему хранения для того, чтобы уменьшить количество передаваемых данных

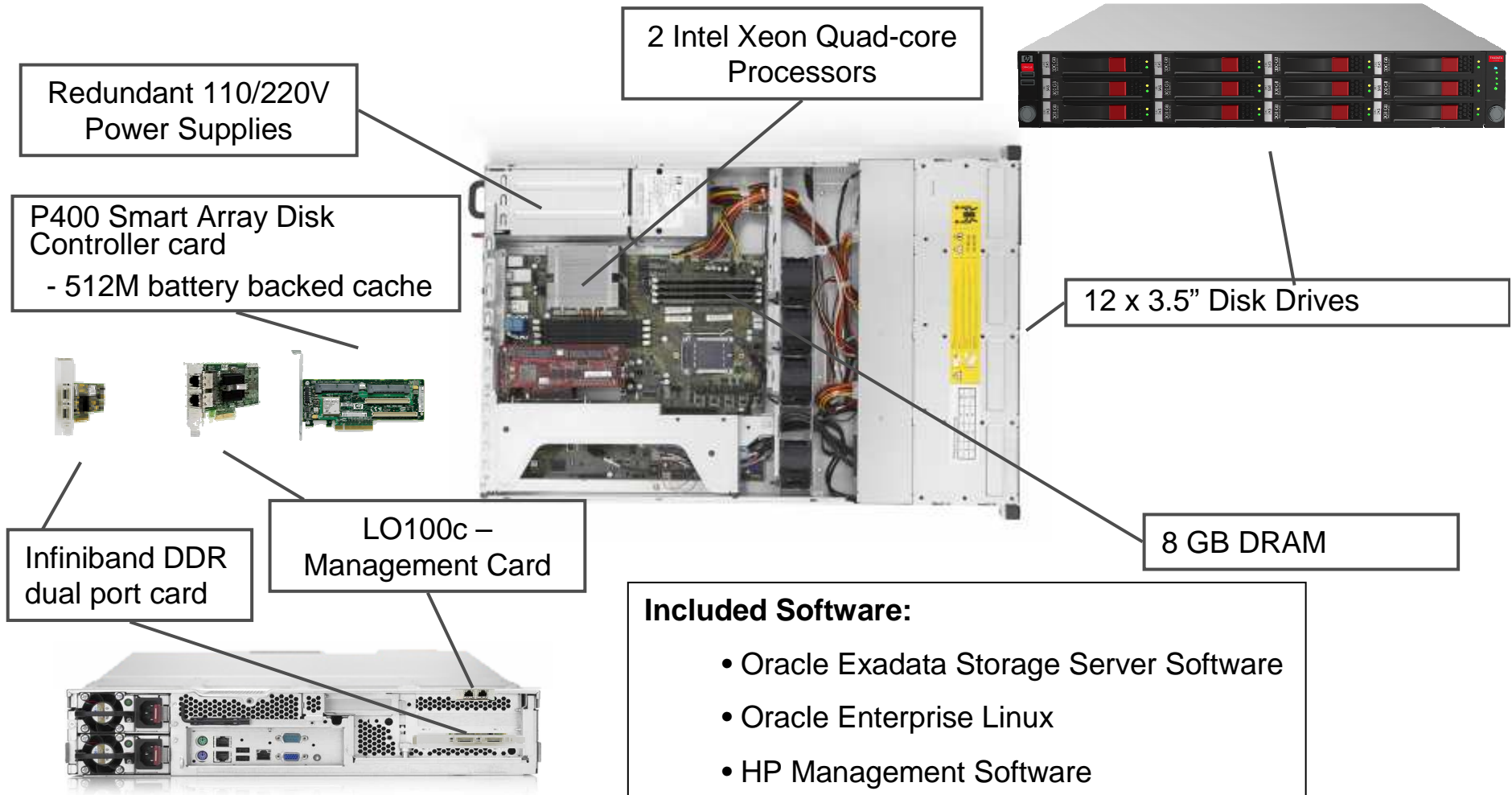


Конфигурация системы с Exadata



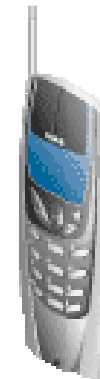
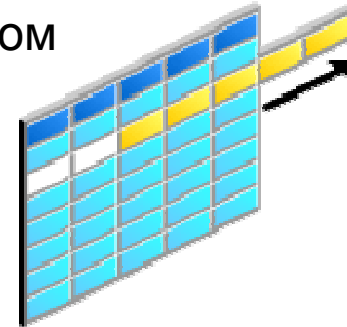
- Каждая ячейка Exadata – самостоятельный сервер с установленными дисками и ПО Exadata
- Данные «размазаны» между многими ячейками Exadata
- Нет ограничения на количество ячеек в системе

HP Exadata Storage Server Hardware Details

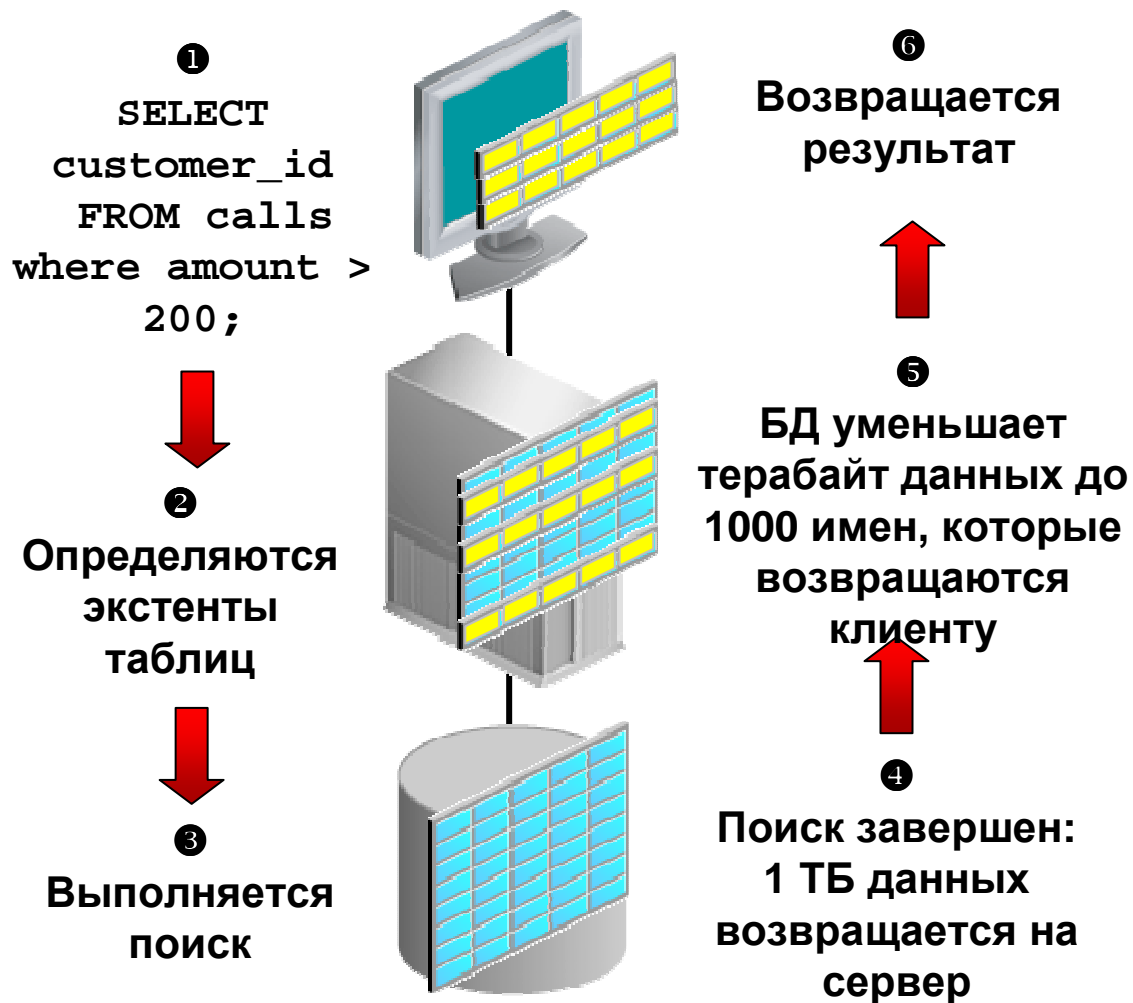


Перенос нагрузки по сканированию на Exadata

- Ячейки Exadata выполняют «умное» сканирование данных, которое уменьшает объем данных, обрабатываемый сервером БД
- Значительное уменьшение объема перекачки данных
 - Обычно в десятки раз
- Запрос выполнится, даже если во время выполнения откажет диск
- Пример «умного» сканирования:
 - Телеком оператор хочет найти клиентов, которые тратят более \$200 за один телефонный звонок.
 - Таблица содержит 1 терабайт данных. Данные о таких VIP клиентах занимают 2 мегабайта. Это в ~500 тысяч раз меньше.

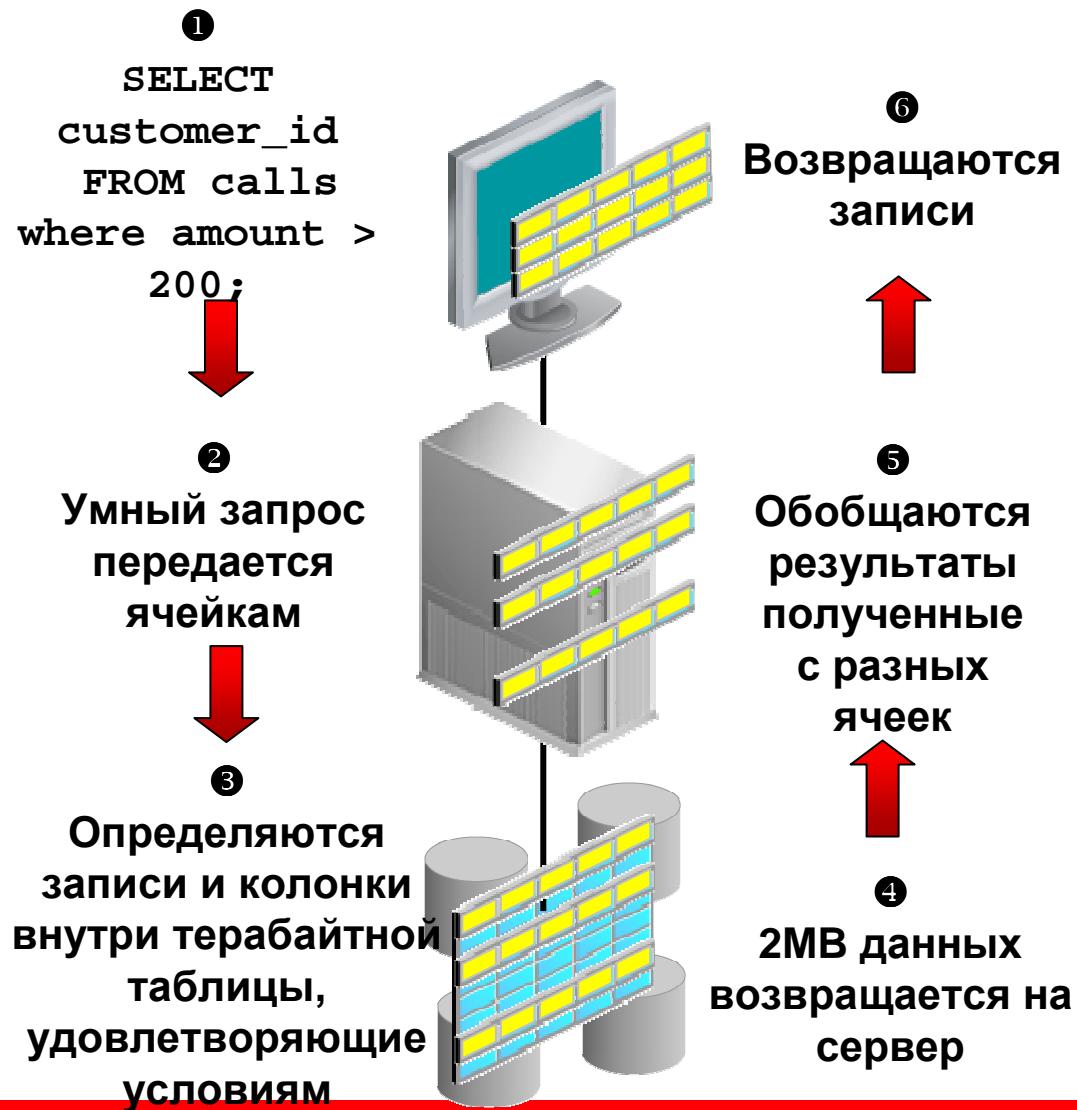


Традиционное выполнение запроса



- Пример:
 - Оператор хочет найти клиентов, которые тратят больше \$200 на один звонок
- С традиционным хранилищем, анализ данных производится сервером БД
- Большая часть данных отсеивается за ненадобностью
- Данные, которые отсеиваются, тем не менее тратят время на их передачу

Выполнение запросов с Exadata



- Только нужные колонки
 - customer_id
 - И нужные записи
 - where amount>200
- Возвращаются на сервер БД
- Разгружается процессор БД
- Не передается лишний объем данных

План запроса с использованием Exadata

- TABLE ACCESS STORAGE FULL
- INDEX ACCESS STORAGE FULL
- storage(SYS_OP_BLOOM_FILTER(...))

Id	Operation	Name	E-Rows
0	SELECT STATEMENT		
1	SORT AGGREGATE		1
2	PX COORDINATOR		
3	PX SEND QC (RANDOM)	:TQ10000	1
4	SORT AGGREGATE		1
5	PX BLOCK ITERATOR		46
* 6	TABLE ACCESS STORAGE FULL	SALES	46

Predicate Information (identified by operation id):

```
6 - storage(:Z>=:Z AND :Z<=:Z)
    filter("PRICE"<25000)
```



Терминология

- Ячейка Exadata – «умное» устройство хранения данных на базе сервера HP DL180-G5
 - На каждой ячейке установлено 12 дисков, Enterprise Linux и ПО Oracle Exadata
- Oracle Exadata Software- специальное ПО, которое понимает структуру хранения данных в блоках СУБД Oracle
 - Может осуществлять фильтрацию данных и другие операции
- HP Oracle Database Machine – система, состоящая из
 - 14 ячеек Exadata
 - 4 Infiniband маршрутизатора
 - 8 серверов HP DL 360 – G5
 - ПО: Enterprise Linux + Oracle 11g + RAC + Partitioning

HP Oracle Database Machine: Key Components



Database Server Grid

8 серверов, каждый состоит из:

- 1 HP DL 360-G5 c
 - 2 Intel Quad-core processors
 - 32 GB RAM
 - 4 146GB SAS disks
 - Dual-port Infinibad Host Channel Adapter (HCA)
- Oracle Enterprise Linux
- Oracle Database 11g Enterprise Edition with Real Application Clusters and Partitioning

4 Infiniband Switches

Каждый с 24 портами

Exadata Storage Server Grid

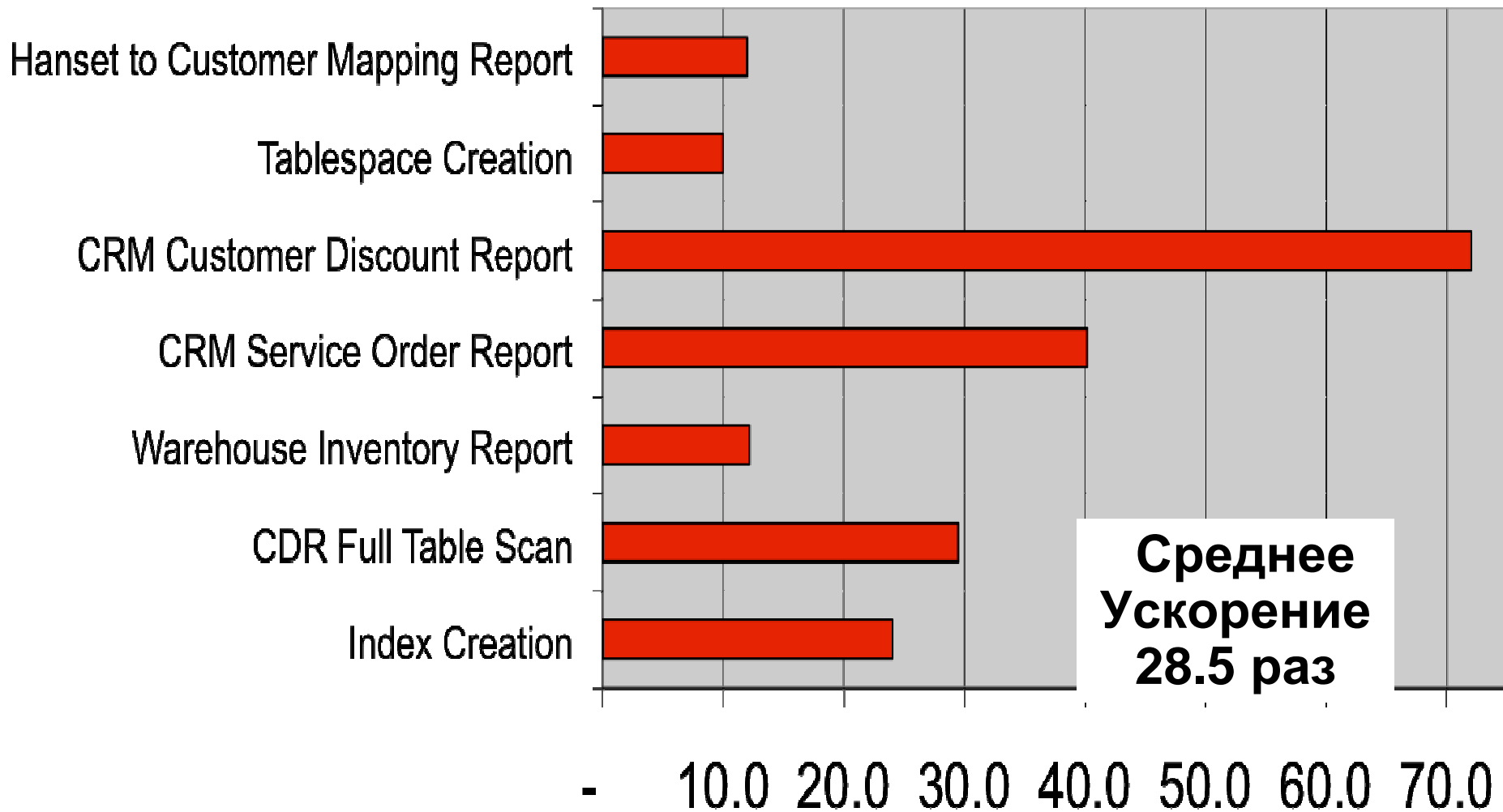
14 серверов, каждый состоит из:

- One HP DL180-G5 with
 - 2 Intel Quad-core processors
 - 8GB RAM
 - 12 300GB SAS or 1TB SATA disks
 - Dual-port Infiniband Host Channel Adapter (HCA)
- Oracle Enterprise Linux
- Oracle Exadata Storage Server Software

Retailer Exadata Speedup – 3x to 50x



Telco Exadata Speedup – 10x to 72x





Преимущества архитектуры Exadata

- Exadata позволяет ускорить выполнение запросов и другие операции в десятки раз
 - Меньше прокачивается данных – разгружается сервер БД
 - Значительный объем обработки запросов переносится на систему хранения
 - Обработка распараллеливается как на уровне массива ячеек Exadata, так и на уровне кластера БД
- Для приложений это совершенно прозрачно, так как они работают с «обычной» СУБД Oracle 11g
 - Не надо переписывать приложения и запросы при миграции на Exadata
 - Возможна миграция данных в архитектуру с Exadata без остановки сервера



Дополнительная информация

- **Oracle.com:**

<http://www.oracle.com/exadata>

- **Oracle Exadata Technology Portal on OTN:**

<http://www.oracle.com/technology/products/bi/db/exadata>

- **Oracle Exadata white papers:**

<http://www.oracle.com/technology/products/bi/db/exadata/pdf/exadata-technical-whitepaper.pdf>

<http://www.oracle.com/technology/products/bi/db/exadata/pdf/migration-to-exadata-whitepaper.pdf>



ORA



ORACLE®

ORACLE