



ТЕХНОЛОГИИ БАЗ ДАННЫХ

Третья практическая конференция

Стратегия и тактика выбора СУБД и инструментов обработки данных

Пленарная часть

10:00

Андрей Николаенко (IBS)

Стратегия и тактика выбора СУБД и инструментов обработки данных

10:25

Олег Бартунов, Александр Коротков (Postgres Professional)

Оптимизация производительности СУБД PostgreSQL

10:50

Дмитрий Пудов (Angara Technologies Group)

Современные технологии защиты баз данных

11:15

Алексей Миловидов (Яндекс)

Секреты эффективного использования СУБД ClickHouse

11:40

Кофе-пауза

12:05

Марк Ривкин (Oracle)

Автономные базы данных. Пора ли администраторам баз искать новую работу?

12:30

Евгений Степанов (Micro Focus)

Какой должна быть современная аналитическая СУБД?

12:55

Сергей Кузнецов (ИСП РАН, МГУ)

Управление данными, энергонезависимая память: проблемы и перспективы

13:20

Обед

Технологии и разработка

14:05

Борис Новиков (СПбГУ)

Декларативные средства: новый виток развития?

14:30

Артём Шитов (GridGain)

Открытая платформа распределенной работы с большими данными

14:55

Юрий Буйлов (CarPrice)

Микросервисы и нейронные сети против больших данных

15:20

Кофе-пауза

15:35

Роман Симаков (Ред Софт)

Технологии обеспечения доступности в СУБД «Ред База Данных»

16:00

Дмитрий Ленёв (Oracle)

Что нового в MySQL 8.0?

16:25

Дмитрий Еманов (FirebirdSQL Project)

СУБД Firebird: мониторинг, трассировка и диагностика

16:50

Дмитрий Павлов (Arenadata)

Greenplum 5.0 – на волне Open Source

Практика

14:05

Олег Морозов (TmaxSoft)

Tibero 6: опыт миграции

14:30

Михаил Сеткин (Райффайзенбанк)

Data Lake в банке

14:55

Олег Орлов («Неофлекс»)

Корпоративное хранилище на открытом ПО: миф или реальность?

15:20

Кофе-пауза

15:35

Кирилл Варламов (OnGrid Systems)

Блокчейн: как обеспечивается безопасность распределённого реестра?

16:00

Кирилл Юхин (Mail.ru, Tarantool.org)

Оптимизация запросов в СУБД In-memory

16:25

Александр Кондиайн (R-Style Softlab)

Практика оптимизации промышленных решений на платформе Oracle

16:50

Павел Труханов (Okmeter)

Мониторинг баз данных

История одного проекта

5НФ + BLOB в Oracle Database

Один инструмент

Подобрались к сотням терабайтов: резервная копия? отказ
всего сразу? расширение СХД?



5НФ в Oracle Database и исходные сообщения в Cassandra

Разгружена транзакционная база

Раздельные SLA для инструментов

Точки эффективности Cassandra (до 2.0) –
не более 32 ГБ ОЗУ на узел, до 1 ТБ на узел



Вместо Cassandra – ScyllaDB

Эффективное использование
современного оборудования

Совместимость с Cassandra CQL

Segmentation fault (core dumped)...

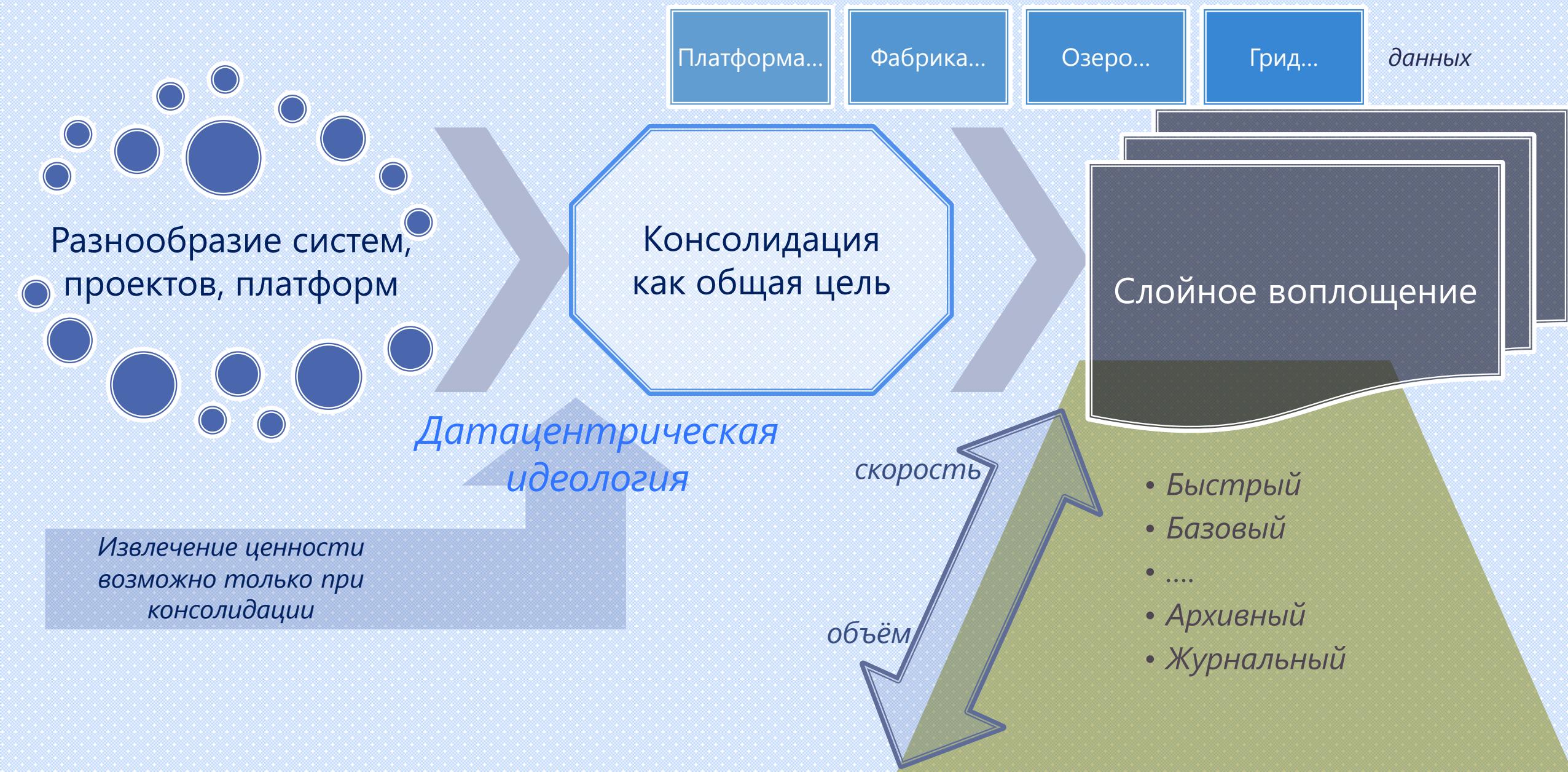


Вместо Oracle Database – PostgreSQL, вместо ScyllaDB – HBase

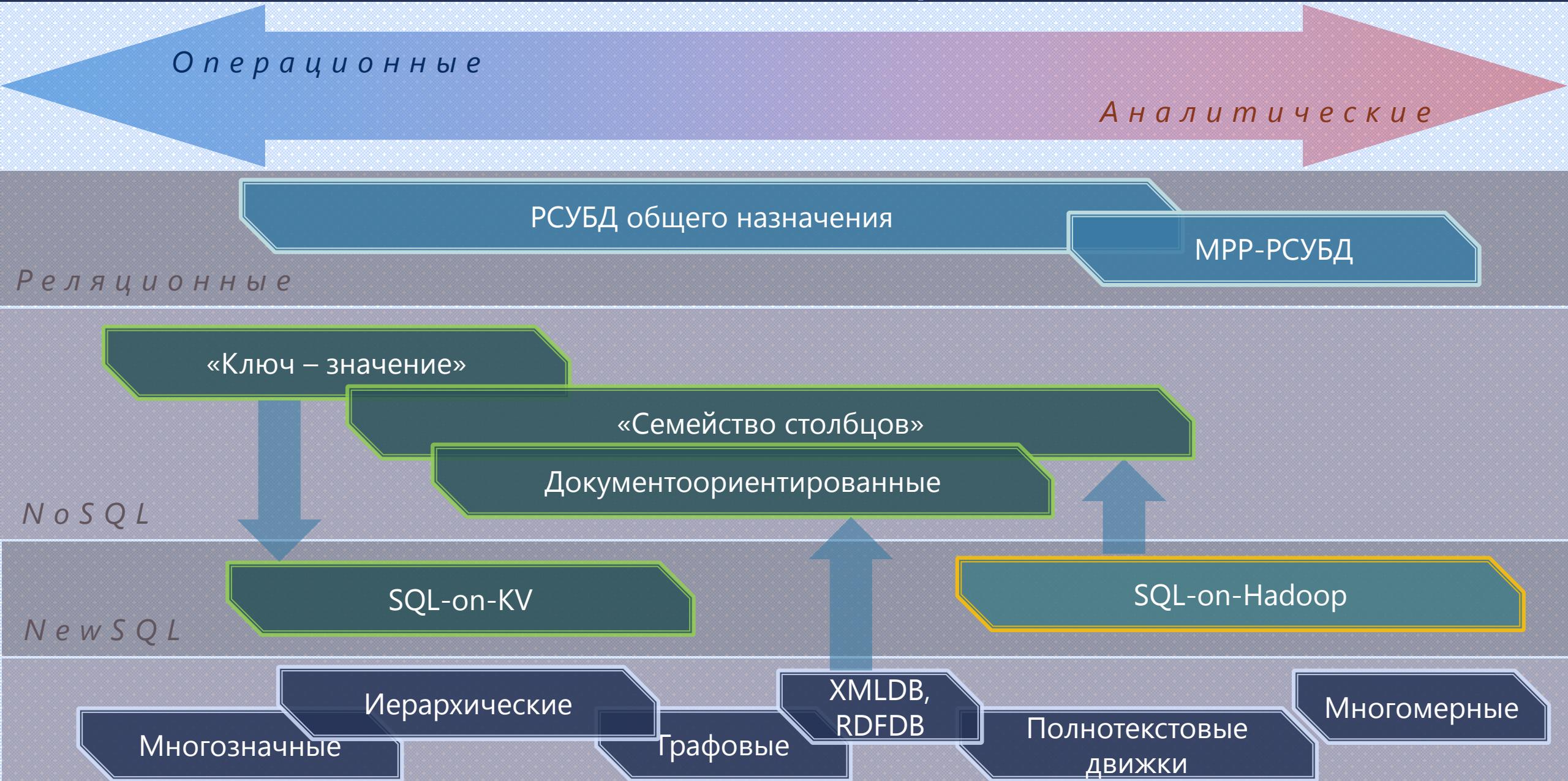
Экономия на лицензиях при сохранении PL/SQL-кода и
показателей производительности

Совместимость с экосистемой Hadoop, стабильность

Снова выборы, и снова «зоопарк»?



Поле выбора



Операторионные СУБД: один сегмент (2016)



Оператори ционные СУБД: один сегмент (2017)



СУБД для аналитики: почти тоже (и Hadoop)



Объём рынка

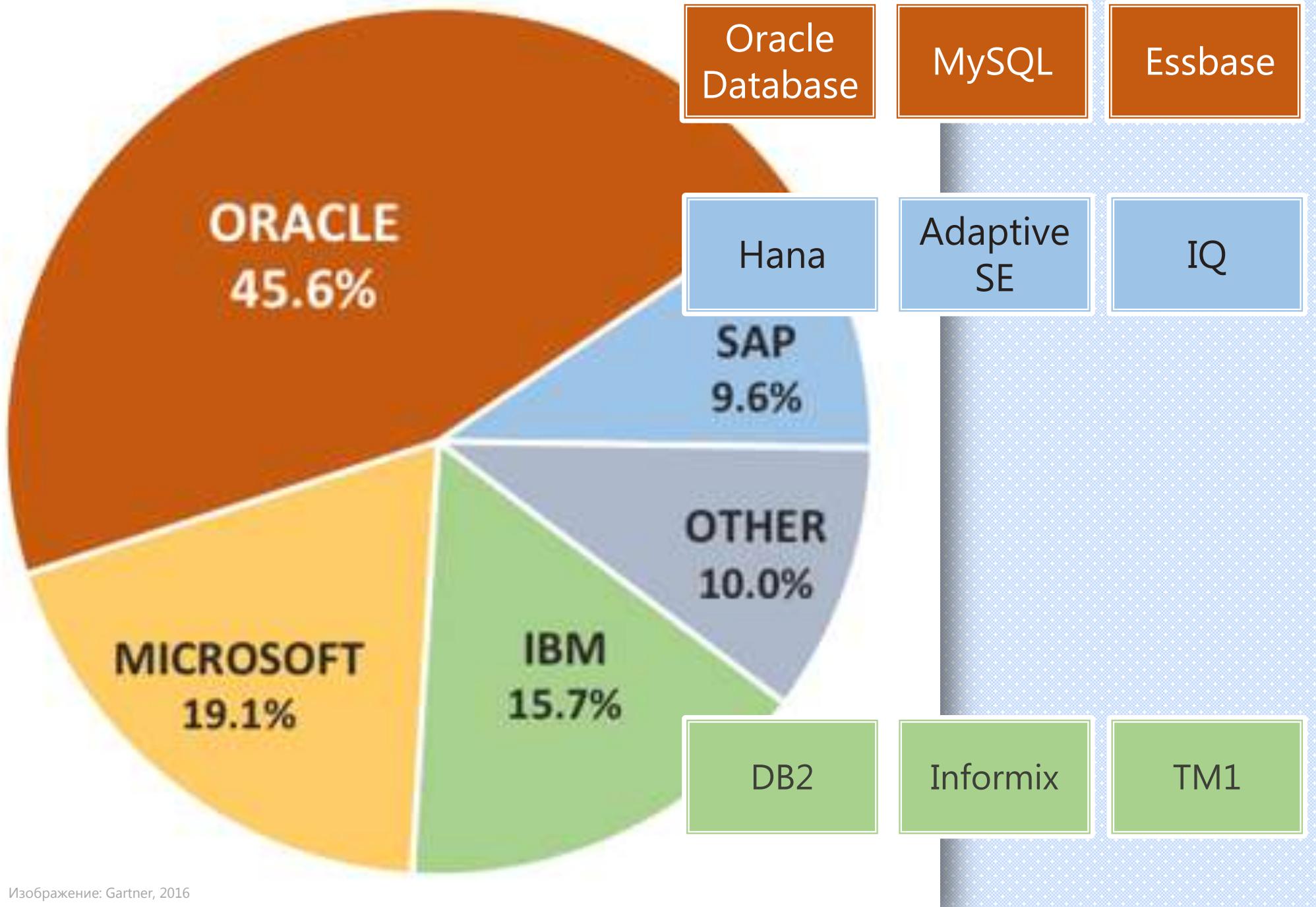
451 Research: к 2016 году
объём рынка NoSQL
превзойдёт \$1 млрд

IDC: рынок больших
данных к 2016 году
– \$16 млрд

Gartner: резидентные
СУБД в 2015 году – \$6 млрд

IDC: рынок СУБД в 2016
году – \$45,7 млрд

РСУБД общего назначения:
основа коммерческого
рынка (2016, Gartner)



Фантастическое будущее Oracle DB 18c

10:0

10:2

10:50

11:15

11:40

12:05

Марк Ривкин (Oracle)

Автономные базы данных. Пора ли администраторам баз искать новую работу?

12:30

12:55

13:20

MySQL: прогресс (уже в Oracle)

14:0

14:3

14:55

15:20

15:35

16:00

Роман

Технологии с

ных»

Дмитрий Ленёв (Oracle)

Что нового в MySQL 8.0?

16:25

16:5

Как Oracle DB, но из Кореи

14:05

Олег Морозов (TmaxSoft)

Tibero 6: опыт миграции

14:30

Михаил Сеткин (Райффайзенбанк)

Data Lake в банке

14:55

Олег Орлов («Неофлекс»)

Корпоративн

тность?

15:20

15:35

16:00

16:15

16:30

Массово-параллельные РСУБД

Teradata

Netezza

- IBM

Greenplum

- Pivotal
→ Dell-EMC

ParAccel

- Actian

DatAllergro

- Microsoft

Aster Data

- Teradata

Dataupia

VectorWise

- Actian

Exasol

InfiniDB

Hadapt

- Teradata

Vertica

- HP → HPE → Micro Focus

Posgres XL

Citus DB

ClickHouse

– начинались как машины баз данных



– на основе PostgreSQL

Greenplum (Arenadata DB)

14:

14:

14:5

15:20

15:35

16:00

16:25

16:50

Роман

Техноло

Дмитрий Лен

Что нового в MySQL 8.0:

Дмитрий Еманов (FirebirdSQL Project)

СУБД Firebird: мониторинг, трассировка и диагностика

Дмитрий Павлов (Arenadata)

Greenplum 5.0 – на волне Open Source

Российский МРР-прорыв и Vertica (уже британская)

10:0

10:25

10:50

11:15

Алексей Миловидов (Яндекс)

Секреты эффективного использования СУБД ClickHouse

11:40

12:00

12:30

Евгений Степанов (Micro Focus)

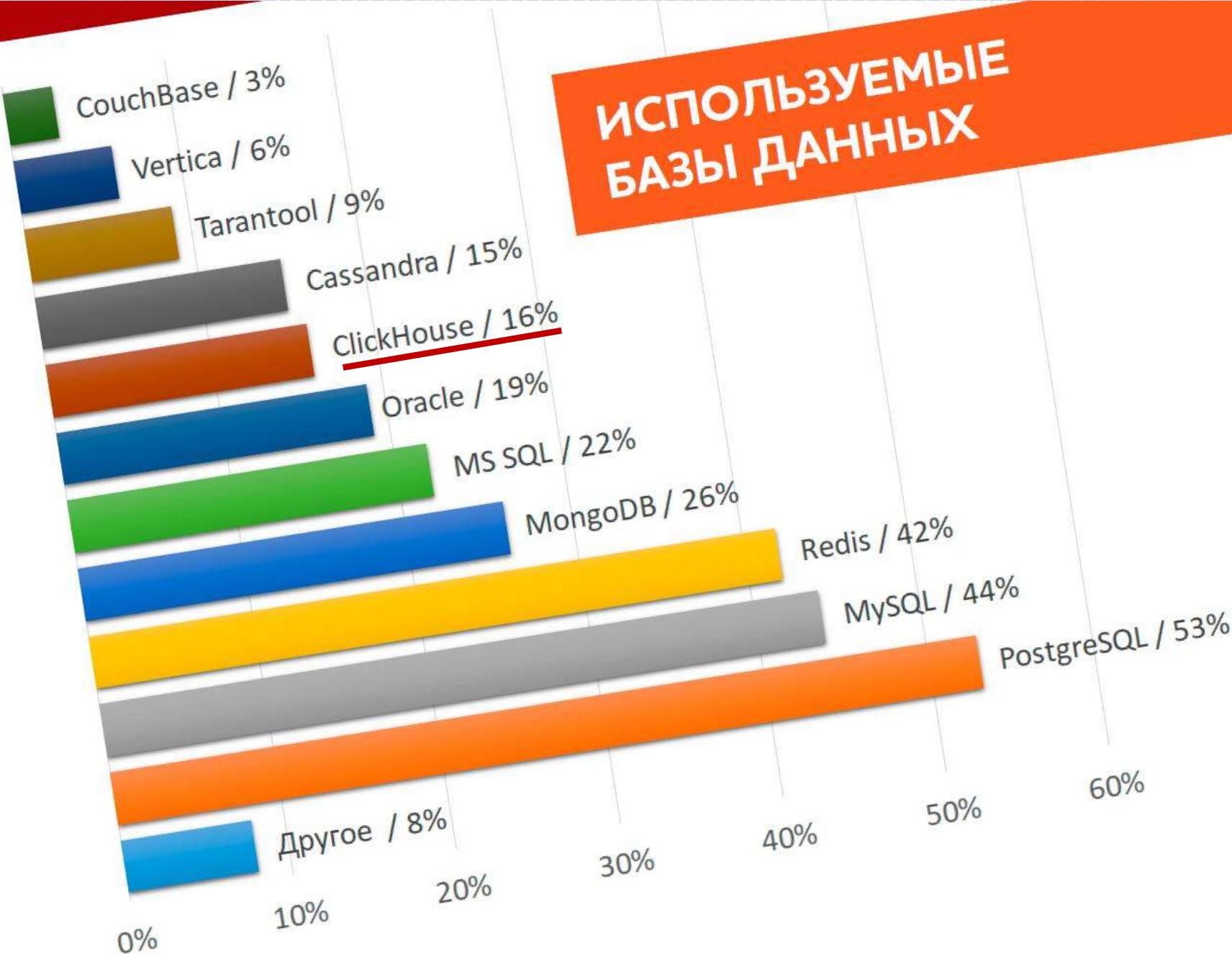
Какой должна быть современная аналитическая СУБД?

12:55

13:20

Яндекс-эффект: результаты опроса на HighLoad++-2017

ИСПОЛЬЗУЕМЫЕ БАЗЫ ДАННЫХ



Изображение: Ontico, 2017

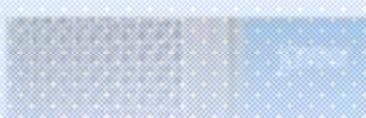
М а ш и н ы б а з д а н н ы х



Цена аналитического МРР-терабайта

... на примере линейки *Teradata* (2014)

Integrated 1. Big Data Platform	Data Warehouse Appliance	Active Enterprise Data Warehouse	Aster Big Analytics Appliance	Appliance for Hadoop
\$3.8K/TB	\$34K/TB	\$34K-\$69K/TB	\$8K/TB	\$2K/TB



Анонс: уже завтра о машинах баз данных

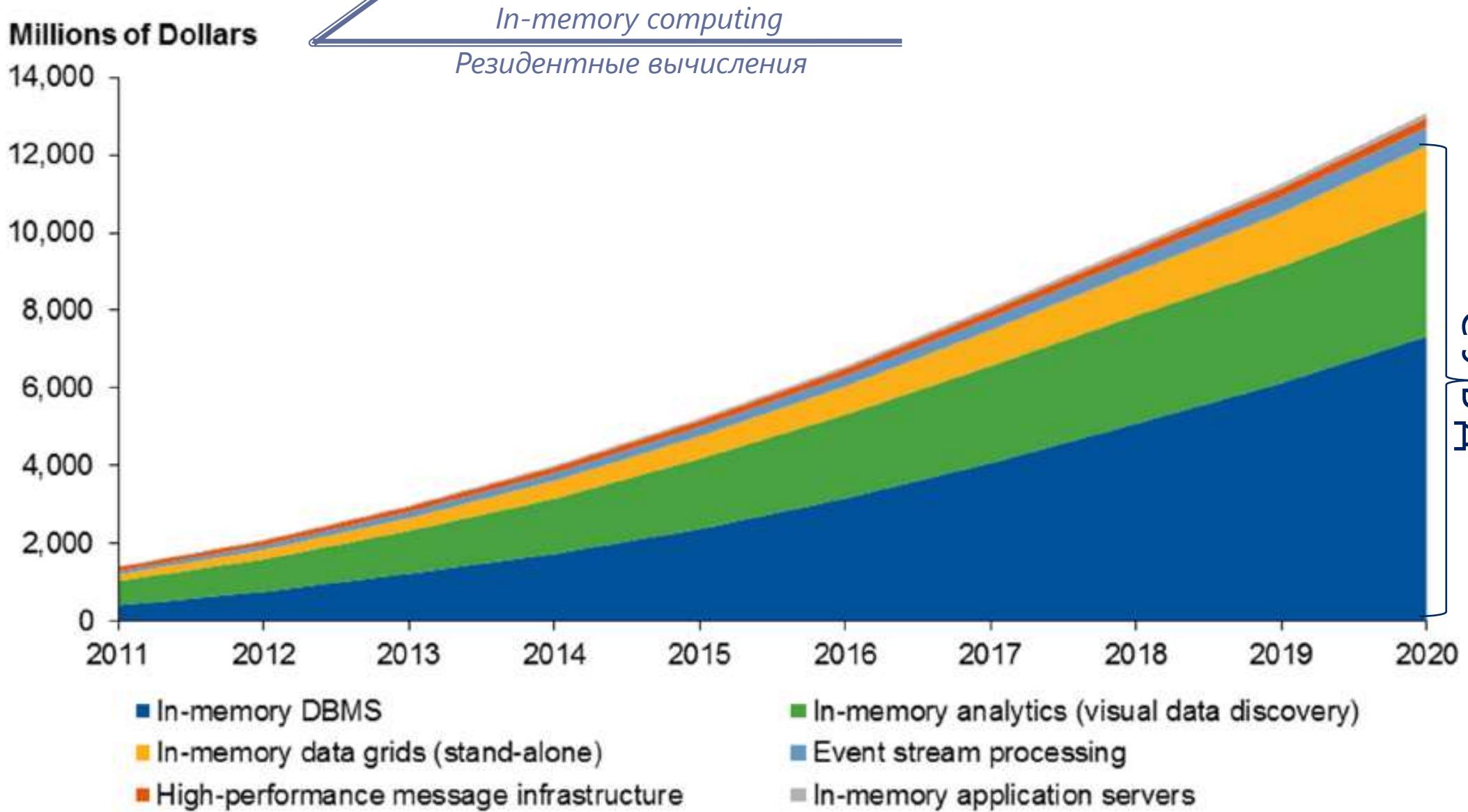
Очередное 197-е заседание семинара Московской секции ACM SIGMOD состоится в четверг, **30 ноября 2017 г.**, в 16:20 во 2-м учебном корпусе МГУ, ауд 524.

На заседании семинара будет заслушан доклад

Машины баз данных: концентрированное обозрение

Другой Фланг: Борьбе за отзывчивость

Figure 3. Market Outlook for IMC Technologies



Source: Gartner (October 2015)

Резидентные СУБД

Реляционные

SAP Hana

Oracle
TimesTen

MemSQL

Реляционные
расширения

Oracle 12g
in-memory
option

IBM DB2
BLU

SQL Server
in-memory
OLTP

MOLAP

IBM TM1

Oracle
Essbase for
Exalytics

Jedox Palo

Ключ –
значение

Redis

Aerospike

Tarantool

Резидентные гриды

Начало: распределённый кэш Java-объектов

На каждый узел J2EE-сервера приложений ставится спец. ПО

Одно из ключевых требований – поддержка JSR-107 (java cache)

Пробрасывает операции в СУБД

Поддерживает основные ORM (Hibernate, Toplink, ...)

Требуется рефакторинг слоя model в MVC-приложении

Oracle
Coherence

WebSphere
Extreme Scale

Tibco
ActiveSpaces

Hazelcast

Gigaspaces
ХАР

Gridgain

Red Hat
Ininispan

Tmax
Ininicache

GridGain: не грид, но фабрика, или просто СУБД?

14:05

14:30

Артём Шитов (GridGain)

Открытая платформа распределенной работы с большими данными

14:55

Юрий Буйлов (CarPrice)

Микросервисы и

15:20

15:35

P

16:00

16:25

16:

Резидентная (и не только) СУБД из Mail.ru

14:0

14:3

14:55

15:20

15:35

Кирилл
Юхин

Блокчейн: как

его реестра?

16:00

Кирилл Юхин (Mail.ru, Tarantool.org)

Оптимизация запросов в СУБД In-memory

16:25

16:5

NoSQL = No[Rel]+SQL

SQL и SQL-подобные

«Ключ – значение»

Redis

Tarantool

Aerospike

GenieDB

RocksDB

IMDG как класс

CockroachDB

MyRocks

Документо-ориентированные

MongoDB

ToroDB

CouchDB

Couchbase

NosDB

RethinkDB

CrateDB

«Семейство столбцов»

HBase+Phoenix

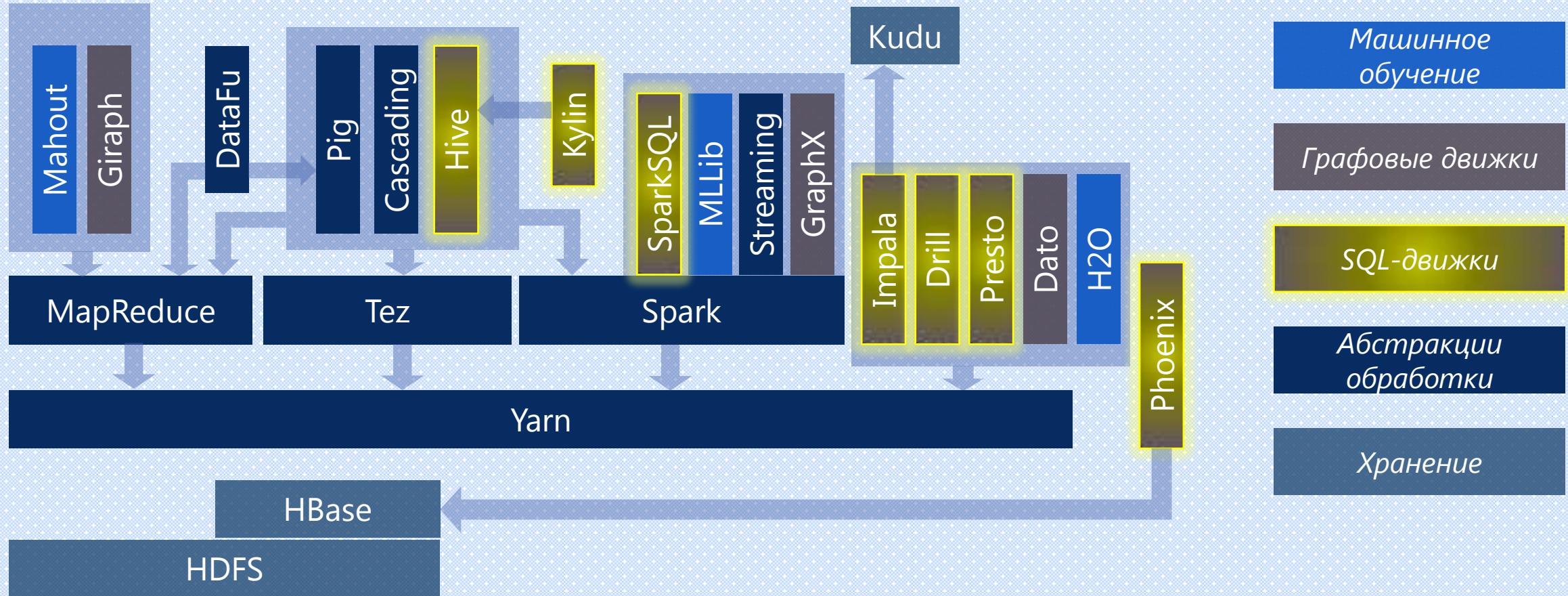
Cassandra

ScyllaDB

RavenDB

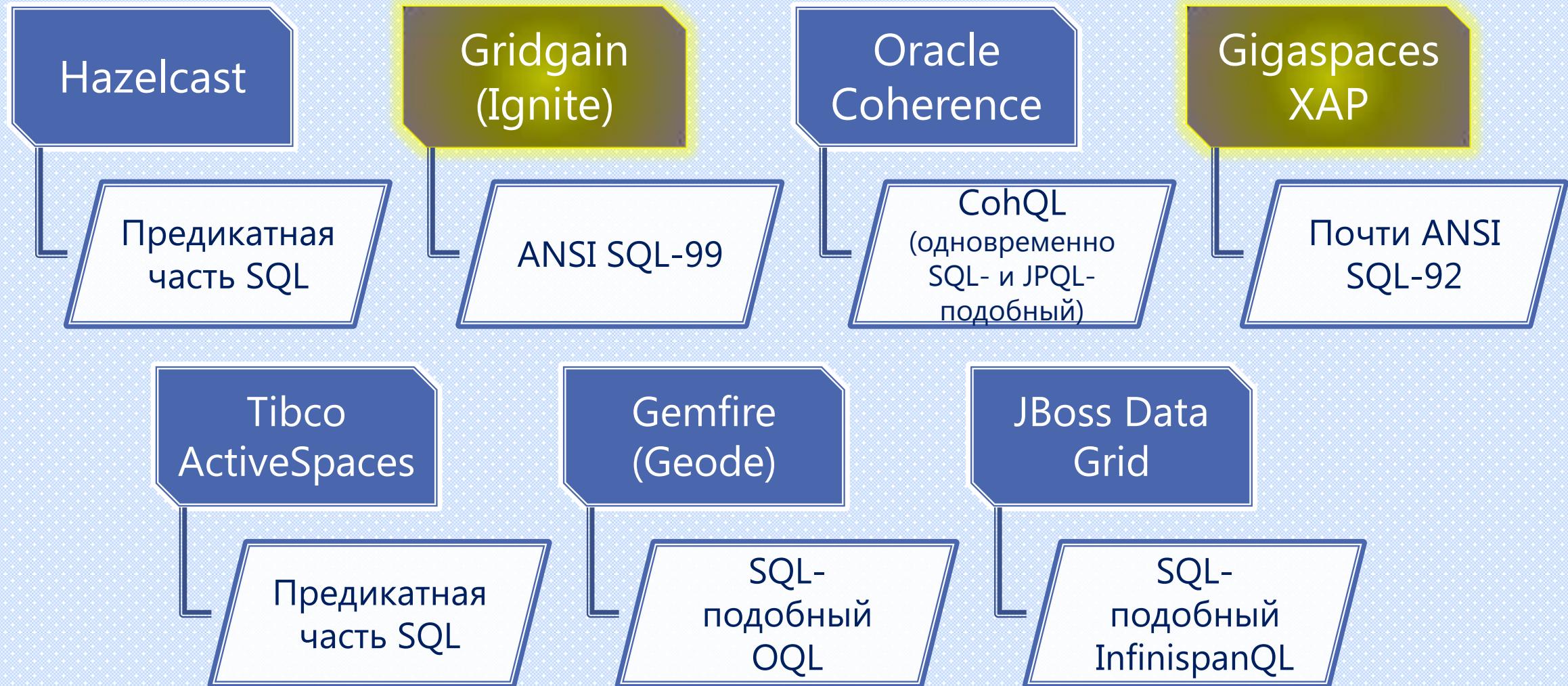
Accumulo

SQL в Hadoop и на Hadoop



SQL в IMDG

Гонка уже не за SQL как таковым, а за глубиной соответствия стандартам



Structured *English* Query Language

Руководство оператора
ЭВМ.
Структурированный язык
запросов SQL

Вытеснен идеями QBE

- (Query-by-example,
Моше Цоф, 1974)

SQL: задумка

Dennis McLeod »
IBM Research Laboratory
San Jose, CA

Keywords and phrases: data base management, data base query languages, data selection, very high level languages, user-oriented languages

Abstract

The problem of the translation and compatibility of high level data base query languages is addressed via a case study of the languages SEQUEL and Query by Example. In particular, an algorithm for translating queries on a relational data base stated in Query by Example into queries expressed in SEQUEL is introduced. This algorithm is used as a vehicle to describe a consistent interpretation of the semantics of SEQUEL and Query by Example. The compatibility of Query by Example and SEQUEL is briefly discussed. Examples of the operation of the algorithm are presented.

« Current address: MIT Laboratory for Computer Science, 545 Technology Square, Cambridge MA 02139.

Introduction

The purpose of this paper is to examine the problem of the translation and compatibility of high level, nonprocedural query languages. In particular, query languages for relational data bases [4,5,6] are considered. The principal goal is to develop techniques useful in designing algorithms for translating "programs" in one such query language (the source language) into "programs" in another (the target language). In this paper, we address this problem by examining two specific data base query languages.

The major reasons for analyzing the translation and compatibility of SEQUEL and Query by Example are:

1. It provides a mechanism by which multiple query languages may be supported in a single data base system, allowing queries to be stated in a language most appropriate to the user and/or the query itself.

2. It requires the actual implementation of a complete language processor for only one query language.

3. It provides some degree of insight into the semantics of the source language, in terms of the semantics of the target language.

To provide a concrete framework for the analysis of translation compatibility, the query language Query by Example [8,9,12] (hereafter referred to as QBE) has been selected as the source language and the query language SEQUEL chosen as the target

on the translation algorithm is that all QBE queries of the type in [8] must be handled. The most recent version of SEQUEL, SEQUEL-2 [3], will be used herein.

A Translation Algorithm

In this section, a specific algorithm for translating a QBE query into a SEQUEL query is outlined; details may be found in the appendix.

The input to the algorithm is a QBE query. For the purposes of this paper, it will be assumed that the algorithm operates on QBE queries in linearized form, similar to that outlined in [10]. (It is a relatively straightforward matter to write a translator which reduces any QBE query into its equivalent linear form.) The output of the algorithm is (the text of) a SEQUEL query.

Algorithm Input and Output. A QBE query consists of a set of t tables, where TABLE- i has m_i columns and n_i rows, and so forth. That is, each query is of the form:

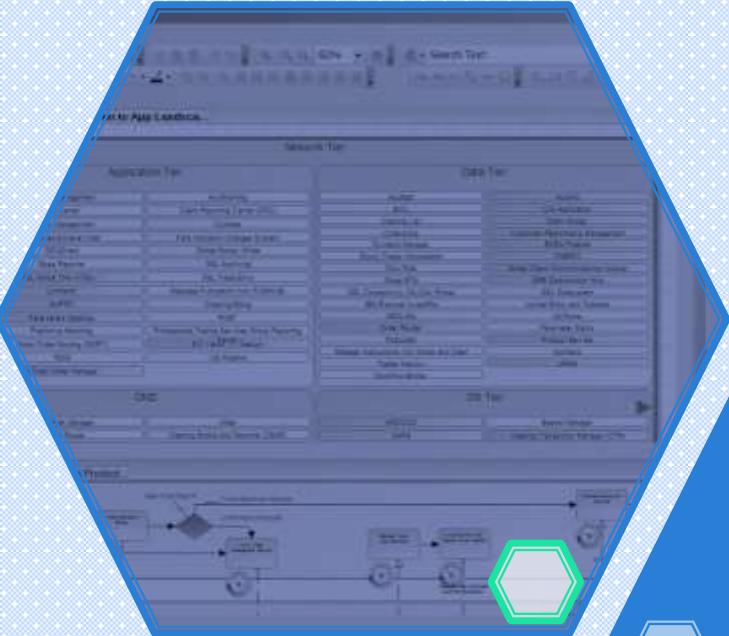
TABLE-1	Column-1	...	Column-m1	
-	-	-	-	<- row-1
-	-	-	-	<- row-n1
⋮	⋮	⋮	⋮	⋮

TABLE-t	Column-1	...	Column-mt	
-	-	-	-	<- row-1
-	-	-	-	<- row-n1
⋮	⋮	⋮	⋮	⋮

Each table actually has an associated name, such as "EMP" or "SALES". Table names are denoted TABLE- i , where i ranges from 1 to the number of tables in the query. Each column of a table has a name, such as "Salary" or "Department". Column names are denoted Column- j , where j ranges from 1 to the number of columns in a table. Note that a column is uniquely identified by a table and column pair. In a QBE query, each table has a number of rows. Rows are denoted (k), where k ranges from 1 to the number of rows in the table.

ICSE '76: Proceedings of the 2nd international conference on Software engineering

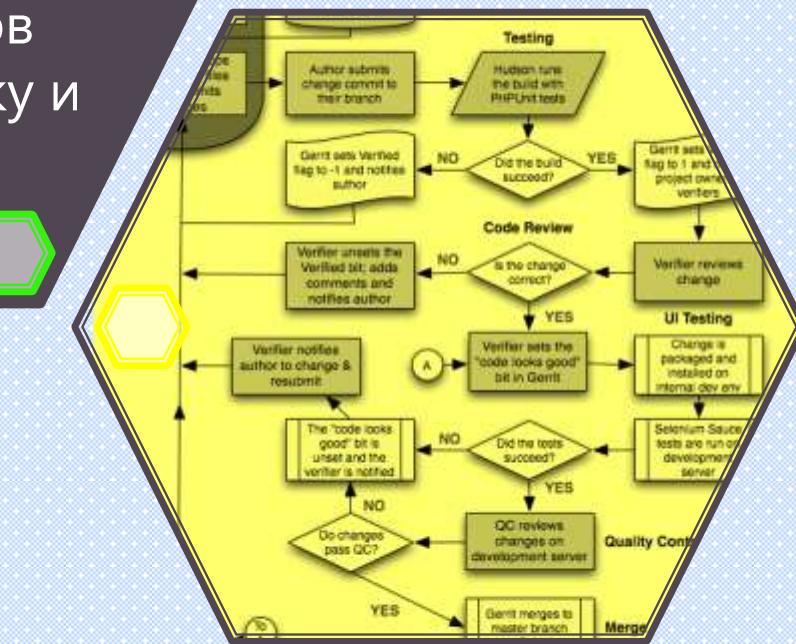
SQL в 2010-е



На SQL спросить
проще и яснее,
чем вникать в
тонкости работы
drag-n-drop-BI

На SQL спросить
быстрее, чем
просить
программистов
сделать витринку и
картинку

```
SELECT PROD, SUM(AMT)  
FROM SALES  
WHERE CAT = 52  
GROUP BY CUBE (SHOP, RG)  
HAVING SUM(AMT) >100
```



На SQL готовы “разговаривать” не только программисты

Декларативный ренессанс (не только SQL)

14:05

Борис Новиков (СПбГУ)

Декларативные средства: новый виток развития?

14:30

Артём Шитов (GridGain)

Открытая платформа распределенной работы с большими данными

14:55

Юрий Буйлов (CarPrice)

Микросервисы

15:20

15:35

16:00

16:15

16:30

Тактика выбора

По объёму

Умещающееся в ОЗУ
одного узла

Умещающееся на
накопители одного
узла

Обрабатываемые в
одном кластере

По интенсивности

Операции ввода-
вывода,
обеспечиваемые
СУБД

Пропускная
способность,
обеспечиваемая
СУБД

Одновременные
подключения,
обеспечиваемые
СУБД

Как проверить?

Эталонные тесты

Воспроизведение
нагрузки

Интеграционные

Стыкуемость с
имеющимися
технологиями

Устанавливаемость
на имеющееся
оборудование

Обрабатываемые объёмы

от
зы
чи
в
ос
тъ

Многомерные

«Ключ – значение» резидентные

РСУБД общего назначения (блочный доступ)

РСУБД общего назначения (POSIX)

«Семейство столбцов»

MPP-РСУБД

Документоориентированные

Hadoop

Полнотекстовые движки

Сотни ГБ

Терабайты

Десятки
ТБ

Сотни ТБ

Петабайты

Объём

“Недозаказ”

UTF-8

Денормализация

PCTFREE

RAID

Факторы репликации

Ландшафты (разработки, тестирования...)

“Перезаказ”

«Храним всё в исходных конвертах в трёх копиях»... (GRC!)

«Через 4 года натравим нейросеть на накопленное»

Стратегия озера: уже в банках

14:00

14:30

Михаил Сеткин (Райффайзенбанк)

Data Lake в банке

14:55

Олег Орлов («Неофлекс»)

Корпоративное

реальность?

15:20

15:35

16:00

16:25

16:

Нужно ли хранить, чтобы потом «натравить» DL?

14:0

14:30

14:55

Юрий Буйлов (CarPrice)

Микросервисы и нейронные сети против больших данных

15:20

15:35

Pc

16:00

16:2

16:5

Эталонные тесты: TPC (enterprise-нагрузки)

	TPC-A/B	TPC-C	TPC-E	TPC-H
Предметная область	Розничный банк	Оптовый склад	Брокерский дом	Дискретный сбыт
Таблиц	4	9	33	8
Чтение-запись	100%	92%	23,10%	8%
Чтение	0%	8%	76,90%	92%
Ограничения целостности	0	0	22	17
Генерация данных	Случайная	Случайная	Псевдо-действительная	Случайная
Метрика	tpsA	tpmC	tpsE	QphH@Size

Эталонные тесты ТРС: трудности

Малое количество опубликованных результатов

Сложность самостоятельного проведения

- Готовые инструменты не воспроизводят спецификацию
- Сравнимость только внутри определённого инструмента

Сложность интерпретации

- Весьма специфические модели, редко подходящие под конкретную нагрузку

Эталонные тесты: YCSB (веб-нагрузки)

Нагрузка	Тип	Пример	% чтений	% записей
A	Атомарные чтения и записи	запись действий в веб-сессии	50%	50%
B	Преимущественное чтение	тегирование фото	95%	5%
C	Атомарные чтения	доступ к профилю пользователя	100%	0%
D	Чтение новых записей	обновление статуса пользователя и его показ	95%	5%
E	Запрос коротких интервалов	запрос потока дискуссии	90%	10%
F	Чтение - обновление - запись	изменение профиля пользователем	50%	50%

Поверхностная спецификация

- Сложность сопоставления результатов – разные гарантии в различных конструкциях

Упрощённые и атомарные нагрузки

- Подходят для узкого класса операционных сценариев

Caution: новый инструмент бенчмаркетинга

- Энтузиасты и разработчики торопятся что-то выполнить без выравнивания условий у конкурентных систем

Воспроизведение нагрузок

Средства на стороне СУБД

Oracle Real Application Testing
(Database Replay)

MS SQL Server Distributed Replay

Средства эмуляции нагрузки

JMeter

LoadRunner

Grinder

Gatling

...

Системы с полным API-доступом (обычно JSON via HTTP)

Журналирование и
«самозахват»

Повтор с интервалами

Разделение типов
нагрузки
(категорий
пользователей)

Масштабирование по
данным?

Тонкости зашнуровки

Как
размножить
базу данных?

«Вегетативное
размножение»

Повторные со
случайным
сдвигом?

Наложение
реальных данных
(из открытых
данных)

Влияние на
аналитику

Предсказуемые
выборки

Пониженная
селективность

Влияние на
OLTP

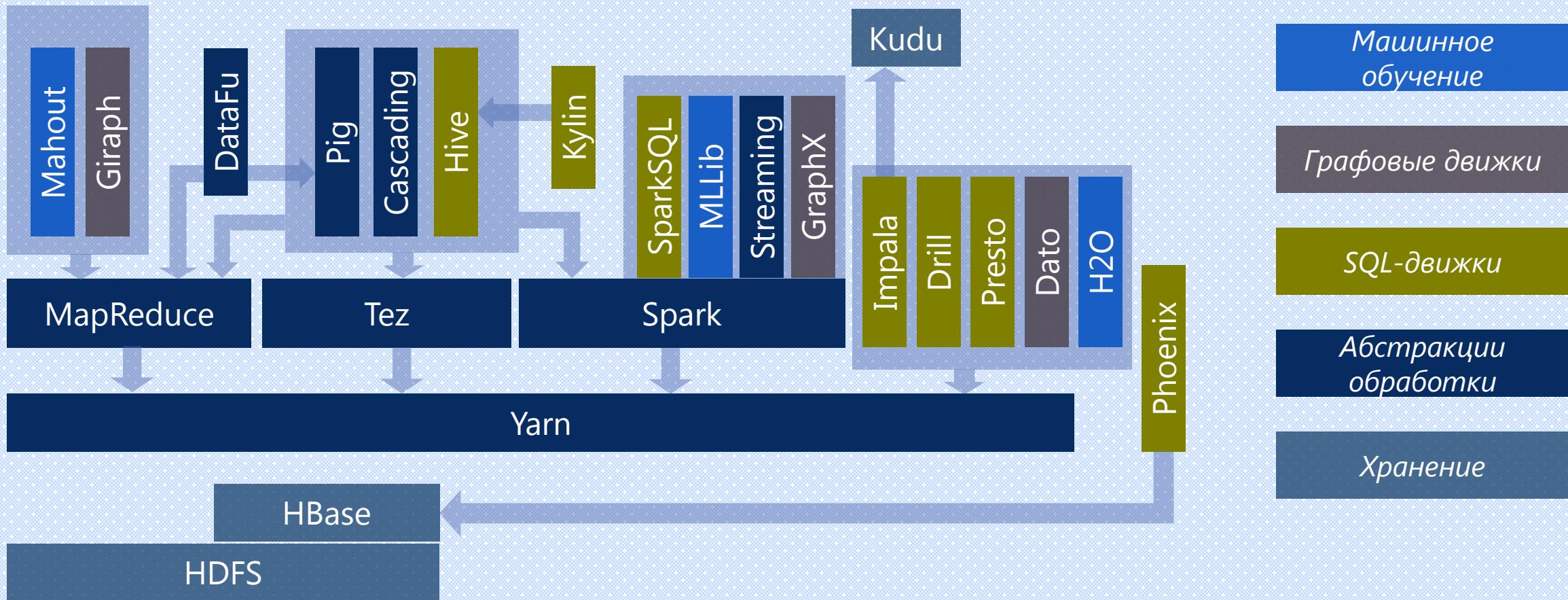
Ключи, индексы...

Размножение
пользовательского
поведения?

Экосистемная классика: LAMP



Многообразие путей в экосистеме Hadoop



SMACK – новая экосистема?



MEAN – НОВЫЙ LAMP



MongoDB is the leading NoSQL database, empowering businesses to be more agile and scalable.

express

Express is a minimal and flexible node.js web application framework, providing a robust set of features for building single and multi-page, and hybrid web applications.



AngularJS lets you extend HTML vocabulary for your application. The resulting environment is extraordinarily expressive, readable, and quick to develop.



Node.js is a platform built on Chrome's JavaScript runtime for easily building fast, scalable network applications.

but **includes** much more:



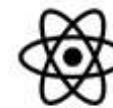
GraphQL



webpack

BABEL

AND EVEN (DRUMROLL...)



React



mocha

LYCE: функционально-акторный мир

L



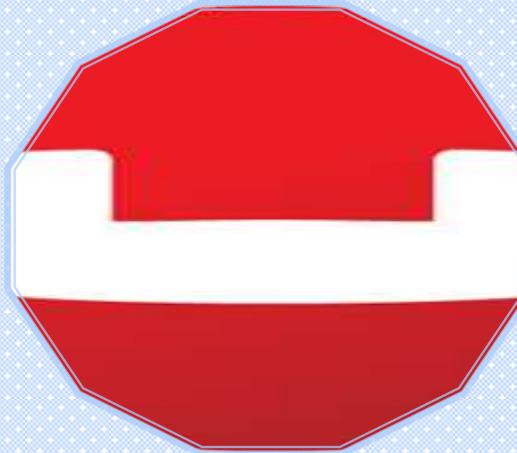
Linux

Y



Yaws

C



Couchbase

E

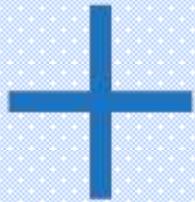


Erlang

ELK – новая классика для журнальных баз



elastic



logstash

kibana

Стеки и экосистемы: за и против

Гарантии

- сочетаемость
- проторенныe пути

Намеренные ограничения

- «пришлось выбрать *Solr* вместо *ElasticSearch*»

Устойчивость, независимость, безопасность

Неочевидные требования могут иметь решающее значение

Репликация в удалённый ЦОД: а как быть с Hadoop?

Интеграция с корпоративным каталогом

Разграничение доступа на уровне столбцов...

Санкционная устойчивость

Сертифицируемость

...

Защитить базу данных

10:00

10:25

10:50

Дмитрий Пудов (Angara Technologies Group)

Современные технологии защиты баз данных

11:15

Алексей Милови

Секреты эф

11:40

12:05

12:30

12:55

13:20

боту?

Техники из блокчейна

14:0

14:30

14:55

15:20

15:35

Кирилл Варламов (OnGrid Systems)

Блокчейн: как обеспечивается безопасность распределённого реестра?

16:00

16:25

16:50

Стоимость владения

Проверенное
и дорогое

Свободное и
прорывное

Поддержка

Лицензии

Лицензии

...

Поддержка

Поддержка

Миграция

Свободное и российское: от PostgreSQL

10:00

10:25

Олег Бартунов, Александр Коротков (Postgres Professional)

Оптимизация производительности СУБД PostgreSQL

10:50

Дмитрий Пудов (Angara Technologies Group)

Современные технологии

11:15

Алексей Миллер

Секреты

11:40

12:05

12:30

12:55

13:

Свободное и российское: от Firebird и Greenplum

14:00

14:30

14:55

15:20

15:35

Роман Симаков (Ред Софт)

Технологии обеспечения доступности в СУБД «Ред База Данных»

16:00

16:25

Дмитрий Еманов (FirebirdSQL Project)

СУБД Firebird: мониторинг, трассировка и диагностика

16:50

Дмитрий Павлов (Arenadata)

Greenplum 5.0 – на волне Open Source

Уходить в СПО или выжать всё из имеющегося?

14:0

14:30

14:55

Олег Орлов («Неофлекс»)

Корпоративное хранилище на открытом ПО: миф или реальность?

15:20

15:35

16:0

16:25

Александр Кондиайн (R-Style Softlab)

Практика оптимизации промышленных решений на платформе Oracle

16:50

Обслуживаемость

14:

14:

14:5

15:20

15:35

16:00

16:25

16:50

Кир

Блокчей.

Кирилл Юхин

Оптимизация запросов в С

Александр Кондиайн (R-Style Softlab)

Практика оптимизации промышленных решений на платформе Oracle

Павел Труханов (Okmeter)

Мониторинг баз данных

Стратегия: что будет через год, три, пять...?

Hadoop 3.0

Помехоустойчивое
кодирование уже в
«бете»

Кластеры станут
вполовину меньше

Вертикальный x86: есть шанс на экономичность

Single system image

От массового
параллелизма к
борьбе с
неравномерным
доступом к памяти

ПЛИС: новый импульс

Intel + Altera

AMD + Xilinx?

Энергонезависимая память

Долой файловую и
блочную логику!

Нужны ли WAL?

Фабрика
персистентной
памяти на RDMA?

Байтovo-адресуемые накопители уже есть, что дальше?

10:

10:2

10:5

11:15

11:40

12:05

12:30

12:55

13:20

Марк Риъ.

Автономные базы

изискать новую работу?

Евгений Степанов (Microsoft)

Какой должна быть современная аналитическая СУБД?

Сергей Кузнецов (ИСП РАН, МГУ)

Управление данными, энергонезависимая память: проблемы и перспективы

<https://t.me/databasetech>

Спасибо за внимание!