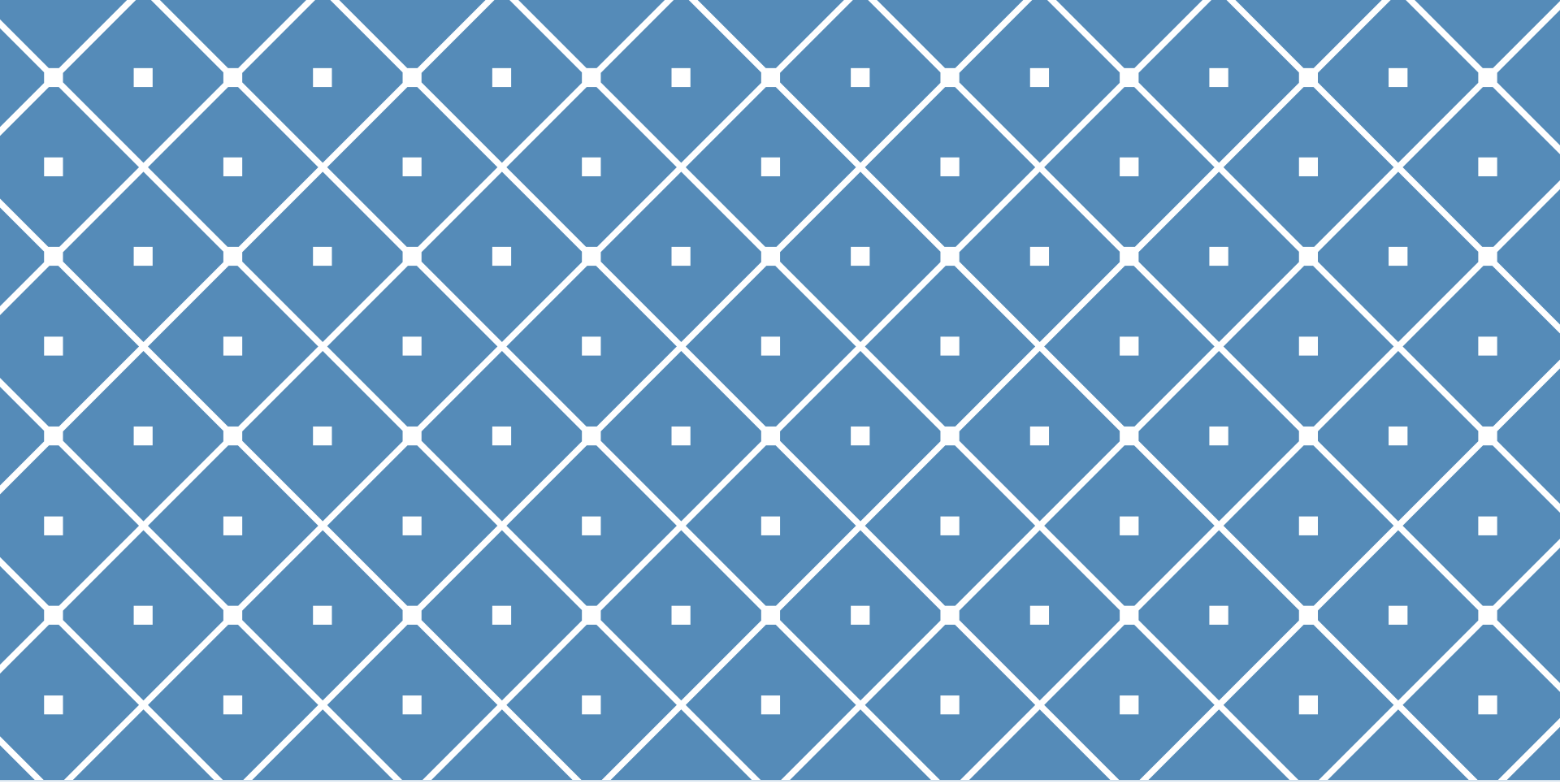


СЕМИНАР МОСКОВСКОЙ СЕКЦИИ ACM SIGMOD, 21 НОЯБРЯ 2013

ОБЛАЧНЫЕ СУБД

Андрей Николаенко,
IBS
anikolaenko@ibs.ru
anikolaenko@acm.org

СОДЕРЖАНИЕ

I ОПРЕДЕЛЕНИЕ

- Что значит «облачность» в контексте СУБД?
- Какие СУБД считать облачными, а какие нет?

ОБЛАЧНЫЕ ВЫЧИСЛЕНИЯ

… модель обеспечения повсеместного и удобного сетевого доступа по требованию к общему пулу конфигурируемых вычислительных ресурсов, […] которые могут быть оперативно предоставлены и освобождены с минимальными эксплуатационными затратами или обращениями к провайдеру

Существенные характеристики:

- Самообслуживание по требованию
- Универсальный доступ по сети
- Пулирование ресурсов
- Эластичность
- Учёт потребления

Модели обслуживания

- SaaS
- **PaaS**
- IaaS

Модели развёртывания

- Частная
- Общественная
- Публичная
- Гибридная

…предоставление возможностей частично или полностью развернуть в облачной инфраструктуре самостоятельно разработанные или приобретённые приложения с использованием языков программирования, библиотек, сервисов, инструментов, поддерживаемых провайдером. Потребитель в этой модели не управляет сетями, серверами, средствами хранения…

DBAAS

Согласно определению стандарта – предоставление потребителям СУБД по облачной модели вычислений – Database-as-a-Service – является относится к модели обслуживания PaaS

- *Не IaaS – "потребитель в этой модели не управляет сетями, серверами, средствами хранения"*
- *Не SaaS – это не приложения, а только средство для них*

Сервис предоставления СУБД внутри только одной организации также может быть облачным

- …в *частной* модели развёртывания, если ему присущи все существенные характеристики

СУБД совместного доступа через не всегда является DBaaS

- … если не выполняются существенные характеристики

ЧТО НЕ РАССМАТРИВАЕТСЯ КАК DBAAS

Облачные хранилища (cloud storage, такие как iCloud, Dropbox и т. п.)

Хостинги небольших веб-приложений, где СУБД предоставляется с серьёзными ограничениями без возможности масштабирования

PaaS-платформы, в которых предоставляется сервис СУБД неуправляем и представляется как неотъемлемая часть более общего сервиса

СУБД, разделяемые несколькими приложениями, если не реализовано специализированных средств масштабирования, балансировки нагрузки, живой миграции

ГЛОССАРИЙ

Провайдер (*provider*)

- Поставщик облачных услуг – организация или подразделение в организации, предоставляющие сервис

Аренда (*tenant*)

- Потребитель облачного сервиса СУБД – пользователь, группа или организация, которой выделена СУБД

Мультиарендность (*multitenancy*)

- Способность программного обеспечения работать с несколькими арендами

Зона (*zone*, *availability zone*, *service location*)

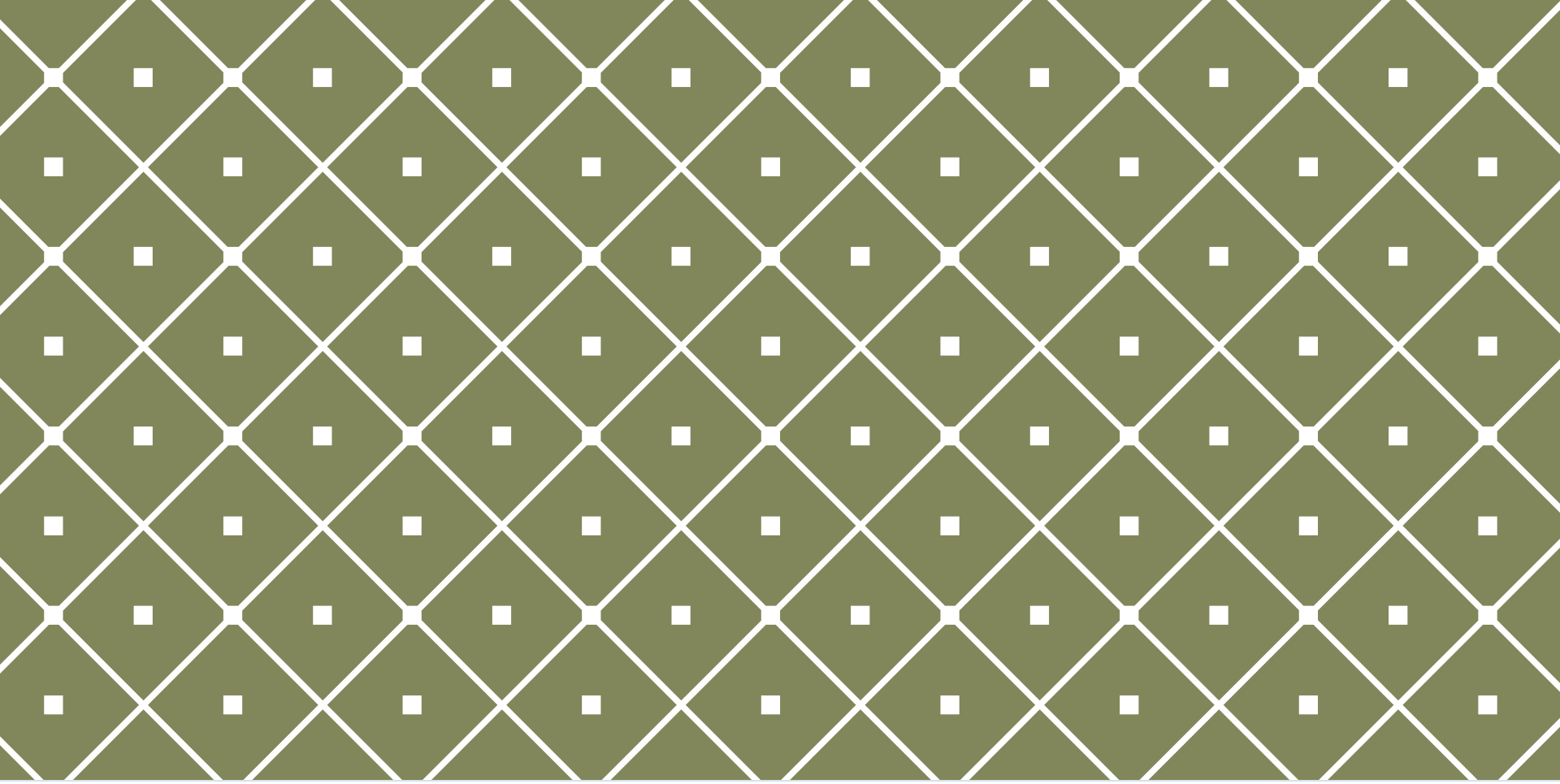
- Географически обособленная зона на стороне провайдера – дата-центр на определённой территории

Геокластер (*geocluster*)

- Кластер, узлы которого находятся в разных *зонах*

Секционирование (*), распределение, шардинг (*partitioning*, *distribution*, *sharding*)

- Распределение базы данных по множеству узлов
- (*) традиционно – разнесение данных одной таблицы



II АРХИТЕКТУРНЫЕ РЕШЕНИЯ

- Виртуализация
- Shared nothing
- *“Shared something”*
- …

ВИРТУАЛИЗАЦИЯ

Использование виртуальных серверов, виртуализации сетей хранения данных и возможностей, предоставляемых инфраструктурной виртуализацией

- Динамическое (автоматическое или автоматизированное) выделение ресурсов (процессорных, оперативного и постоянного хранения)
- Мгновенное или быстрое создание, клонирование виртуальных машин
- Зрелые техники виртуализации, отработанные конфигурации, готовые программно-аппаратные комплексы

АРХИТЕКТУРА SHARED NOTHING

Естественный способ обеспечения масштабирования по вычислительным узлам – *архитектура без разделения ресурсов* с [автоматическим] распределением [базы] данных по узлам обработки

С 1980-х годов широко используется в масштабируемых комплексах для хранилищ данных

- Teradata, Britton-Lee, Nonstop SQL, позднее – Netezza, Greenplum и др.)

Изначально подразумевается в NoSQL-системах (*sharding*)

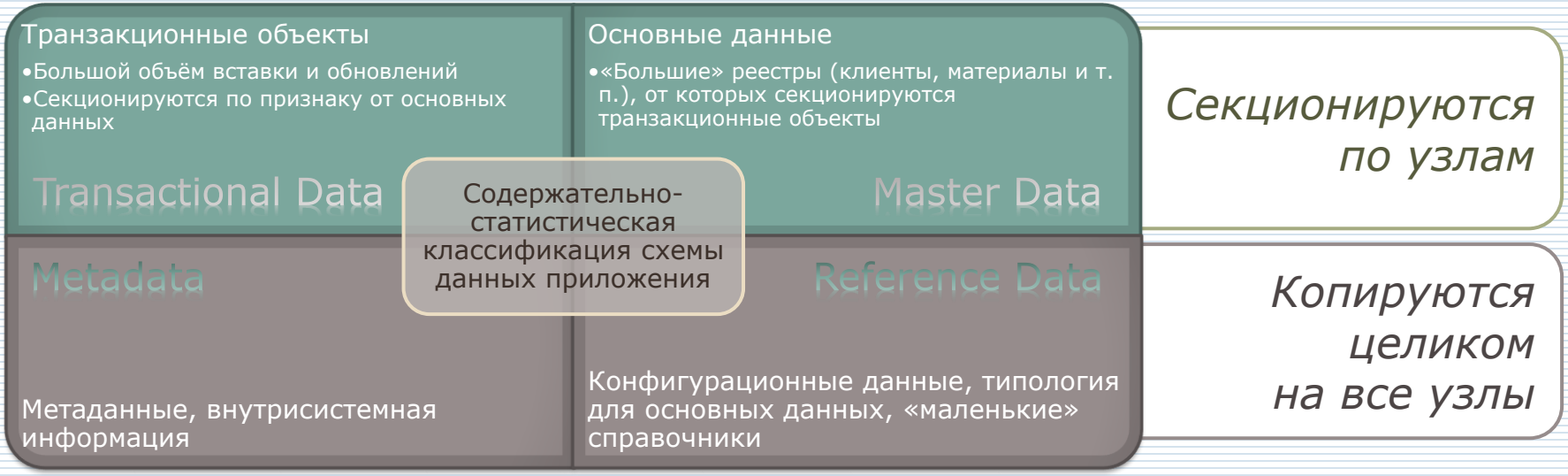
Подразумевает прямое подключение устройств хранения (DAS)

	Зависит от качества выбора ключа секционирования	
	Требует сложного решения по координации узлов	
	Осложнённая реализация OLTP с транзакционным контролем	

“SHARED SOMETHING”

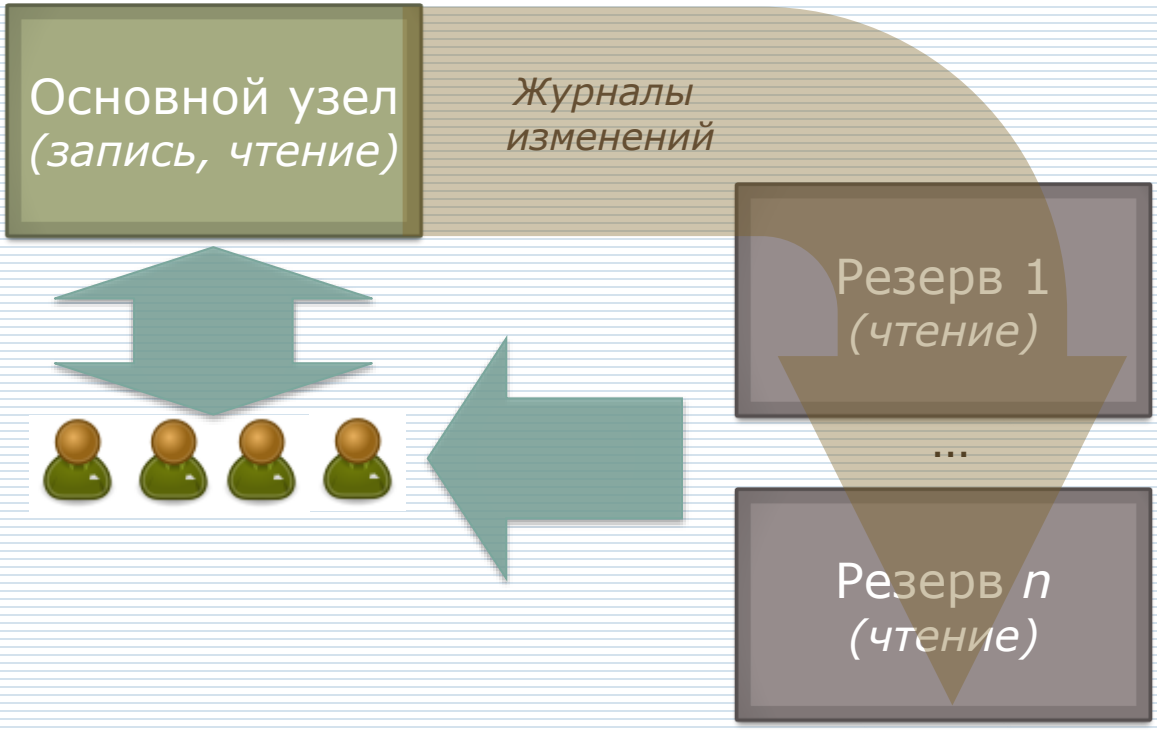
Основная проблема в реляционной модели для Shared Nothing – необходимость частого доступа к относительно небольшому набору данных

David Taniar, **Clement H. C. Leung**, **Wenny Rahayu**, and **Sushant Goel** High Performance Parallel Database Processing and Grid Databases. Wiley Publishing, 2008, isbn: 9780470107621



РЕЗЕРВЫ, ДОСТУПНЫЕ НА ЧТЕНИЕ

Зачастую достаточно одного узла, обрабатывающего операции вставки, обновления, удаления, при этом активность на чтение можно динамически балансировать по резервам



ПАКЕТНАЯ ВАЛИДАЦИЯ

Вместо моментальной проверки ссылочной целостности при пользовательской операции, выполнения всех каскадных обновлений – создаётся заявка на операцию, затем пакетный обработчик последовательно реализует заявки

Успешно работает многопользовательских веб-приложениях (интернет-магазинах, системах бронирования)

- в большинстве случаев пользователи даже не догадываются, что их запрос обработан последовательным «пакетником»
- асинхронный JavaScript обновляет «статус заказа» у клиента, если всё хорошо – это происходит через 2-3 с.

Реализуется только для транзакционных данных

В том или ином виде, использованием MapReduce-бэкенда хранения требует «пакетника», проводящего операции, видимые пользователям как синхронные

Практический вариант двухфазной фиксации изменений

КОНТЕЙНИРОВАНИЕ

Контейнер балансирует общий пул выделенных ему ресурсов среди подключённых к нему баз

- Потребитель владеет целостным экземпляром СУБД с виртуализированными ресурсами

Подключаемая база данных может переподключена из одного контейнера в другой

- Притом мгновенно, в случае thin provisioning

Подключаемая база данных «выбирает» контейнер в зависимости от своей нагрузки и состояния контейнер

Разработчик-поставщик СУБД использует стандартные механизмы, библиотеки, разборщики, интерпретаторы

МАНИПУЛЯЦИИ СО СХЕМОЙ ДАННЫХ

Общая таблица

- «Первый антипарттерн проектирования баз данных» (*Том Кайт*)
- Реализован в Database.com
- Полностью переписан интерпретатор SQL

Разделение по идентификатору аренды с автоматической инъекцией

- `WHERE … TENANT = :tenant_id`
- Замена схемы данных представлениями с инъекциями и триггерами на `INSTEAD OF`

Соглашение об именовании схемы

- Автоматическое разыменование схем, например
  - `… :schema_name + “$$” + :tenant_id`

III ОБЕСПЕЧЕНИЕ ОБЛАЧНЫХ ХАРАКТЕРИСТИК

- Автоматизация развёртывания
- Эластичность
- Балансировка
- Изоляция аренд
- Живая миграция

АВТОМАТИЗАЦИЯ РАЗВЁРТЫВАНИЯ

Средства автоматизации развёртывания виртуальных и облачных инфраструктур	• *Шаблоны виртуальных машин* • *Стандартное или специфическое API для развёртывания*
Для схемных мультиаренд	• *всего лишь создание записи об аренде* • *… или автоподгоподготовка схемы*
Средства контейнеризации	• `CREATE PLUGGABLE DATABASE [AS CLONE]` • `ALTER PLUGGABLE DATABASE UNPLUG INTO …`
Автоматизация останова, парковки	• … в том числе, на медленных средствах хранения

ЭЛАСТИЧНОСТЬ

Подключение новых экземпляров СУБД к той же базе данных при росте нагрузки и отключение при спаде

- Oracle Database Real Application Cluster, IBM DB2 PureScale, SAP ASE Clustering
- Максимум 10-12 узлов
- Неэффективно для мнопользовательских и сильноконкурентных тразнакционных нагрузок

Запуск новых узлов базы данных

- В архитектурах Shared-Nothing, в реализациях готовых к массовому параллелизму
- «Сужение» может быть проблемой, но может быть перекрыто мультиарендностью

Манипуляции с местом хранения

- [Статистически] малоиспользуемые данные переносятся на медленные носители
- «Горячий кэш» в оперативной памяти, в том числе, распределённый

Переклассификация аренды

- Разделение аренд на одноузловые и распределённые (Elastras)
- Как только нагрузка на аренду превышает определённую метрику, она распределяется
- Как только нагрузка спадает – аренда возвращается на один узел

УЧЁТ ПОТРЕБЛЕНИЯ И УРОВЕНЬ СЕРВИСА

Возможные биллинговые метрики:

- Время работы
- Использование ЦПУ
- Использование ОЗУ
- Объём хранения
- Трафик
- Количество записей
- Количество схем
- Количество пользователей
- Максимальное количество одновременно работающих пользователей
- Среднее количество одновременно работающих пользователей
- Количество узлов обработки
- Количество процессов обработки
- Количество операций
- Количество транзакций
- Количество операций ввода-вывода
- Стоимость учтённых в базе данных операций

SLA-метрики:

- Процент доступности
- Среднее время отклика запроса
- IOPS
- Время запуска
- Время восстановления

БАЛАНСИРОВКА

В рамках одного физического узла

Гипервизором	Средствами контейнеризации	Сценариями администрирования СУБД
(в виртуальных средах)	*(в контейнерных СУБД)*	*(искусством разработчиков провайдера)*

В распределённом ландшафте

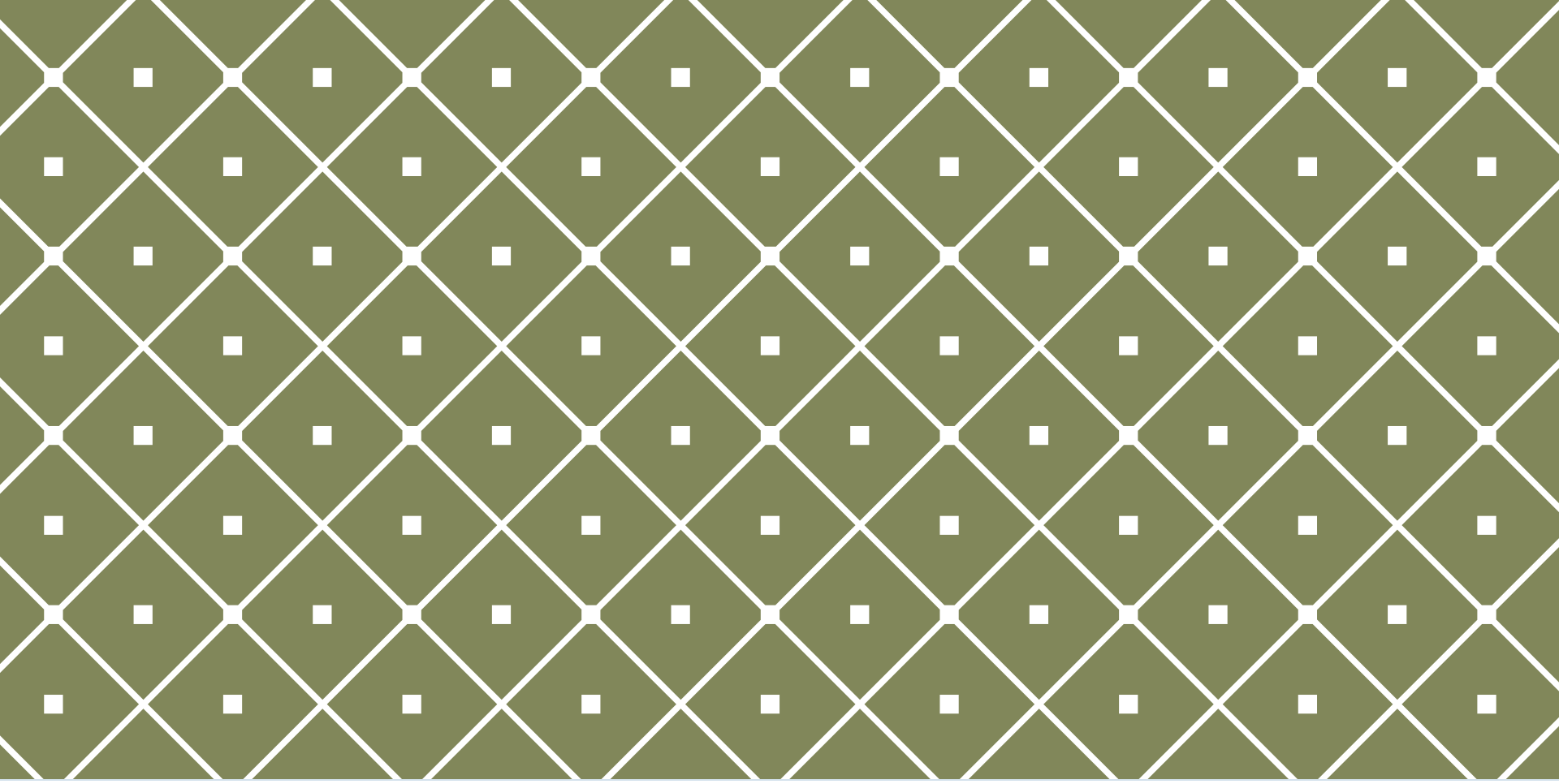
Распределение пользователей по узлам		Распределение операций по узлам
При использовании резервов доступных на чтение – операции чтения балансируются по резервным узлам	*В решениях «активный – активный»*	*В зависимости от решения по организации распределения*

ИЗОЛЯЦИЯ АРЕНД

ЖИВАЯ МИГРАЦИЯ

При сбое сервиса на каком-либо из узлов или в какой-либо из зон система целиком должна быстро («вживую») переместиться на другие узлы или в другое расположение

Холодные и горячие резервы	• *В различных зонах* • *Применение быстрых средств резервирования (журналов изменений)*
Средства контейнеризации	• *Переезд подключаемой БД из одного контейнера в другой*
Особые алгоритмы	• *В Elastras – алгоритм Albatros, работает для HFDS*



IV ДИСКУССИИ

- ACID, теорема CAP, согласованность в конечном счёте
- Языки доступа к данным

ACID, ТЕОРЕМА CAP, СОГЛАСОВАННОСТЬ В КОНЕЧНОМ СЧЁТЕ

Трактовка “consistency”

- ACID *целиком*? «*мгновенная согласованность*»?

«Согласованность в конечном счёте»

Практическая актуальность проблемы распада на секции

- Насколько часто встречается распад на секции в сравнении с отключением одной зоны целиком для всех потребителей?

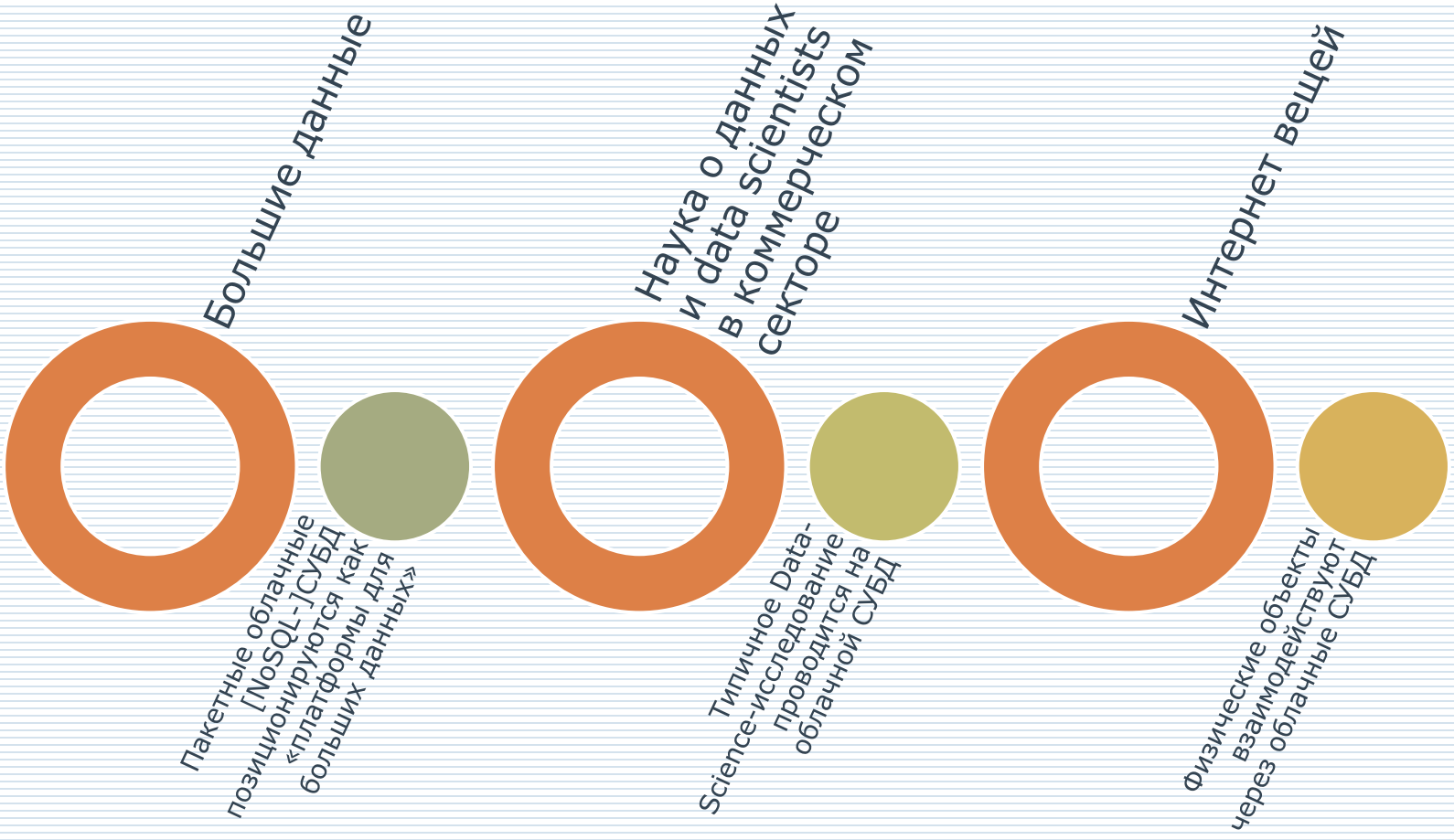
ЯЗЫКИ ДОСТУПА К ДАННЫМ

Дискуссии о необходимости поддержки SQL, о стандартах SQL	• Нужен ли SQL-доступ к NoSQL-базам? • Is SQL really kludge?
Необходимость процедурных расширений	• Нужен ли «императивный код» на стороне СУБД или он должен обрабатываться в связующем программном обеспечении?
Веб-API (XML, REST)	• Большинство облачных служб СУБД создали JSON API • Каково будущее и статус SPARQL, MQL? • Перспективы стандартизации новых языков доступа
Поддержка статистических языков в базе данных, на узлах обработки	• Oracle Database, Netezza поддерживают R на стороне базы данных • ScaleR (Revolution Analytics) реализует массово-параллельную R-обработку в сети Hadoop

V ВЛИЯЮЩИЕ ТЕХНОЛОГИИ

- Тренды
- Облачное ПО

ТРЕНДЫ НАЧАЛА 2010-Х



ОБЛАЧНОЕ ПЛАТФОРМЕННОЕ ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ

Облачные платформы разработки

Связующее программное обеспечение

- Серверы приложений
- Средства сообщений, ESB
- Средства выполнения бизнес-процессов
- Средства управления содержанием
- …

Управление данными

- Средства массово-параллельной пакетной обработки
- ETL
- Управление основными данными
- Управление жизненным циклом информации

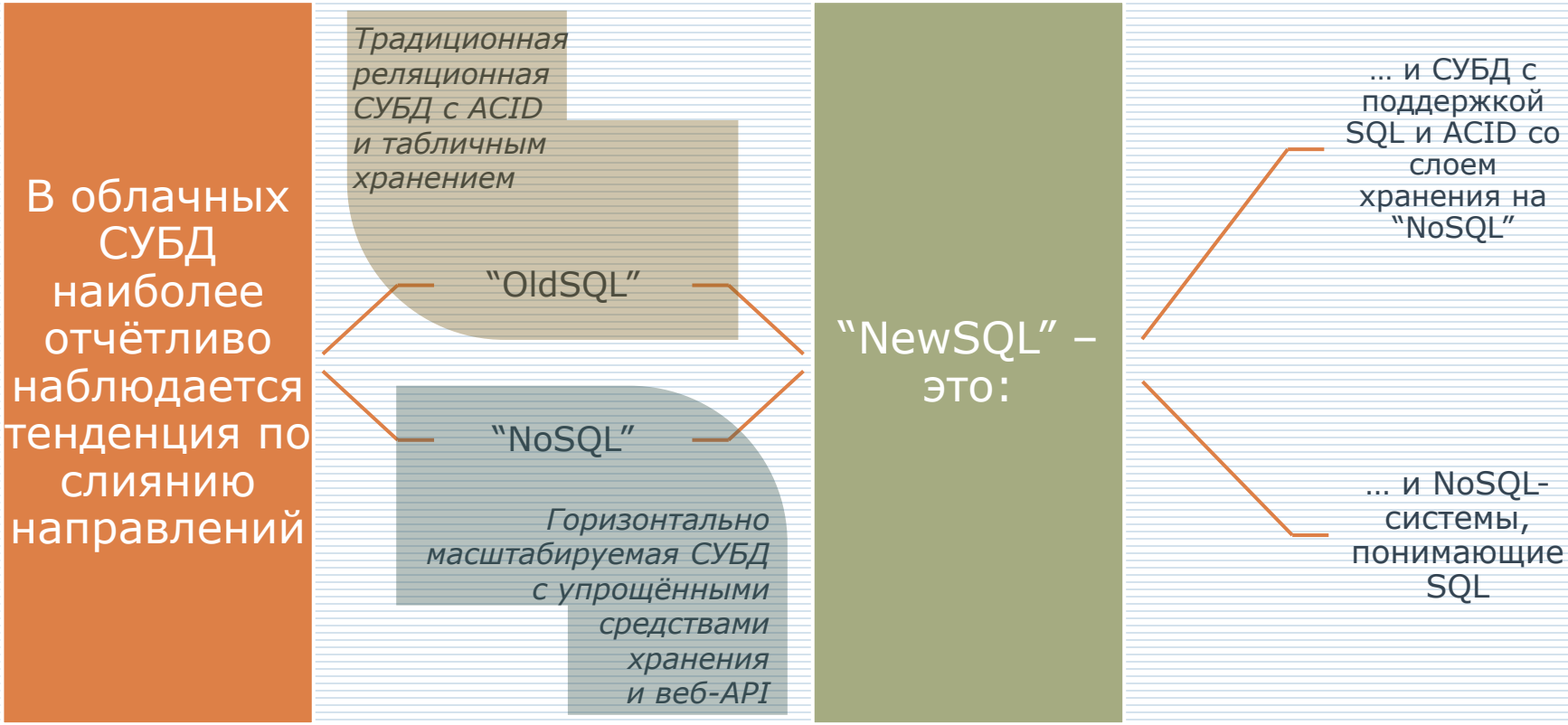
Облачная аналитика

- Cloud Business Intelligence
- Облачные визуализаторы
- Облачные статпакеты

VI ТЕНДЕНЦИИ

- Конвергенция SQL и NoSQL
- Интеграция…
- Аппаратные реализации
- …

КОНВЕРГЕНЦИЯ SQL И NOSQL



ИНТЕГРАЦИЯ

Интеграция

- … с другими PaaS-технологиями
- … в среды разработки

Перекрёстная поддержка облачными СУБД друг друга как источников данных

Переносы

- приватизация (перенос в публичное облако)
- публикация (перенос в частное облако)
- гибридизация (формирование гибридного облака)

АППАРАТНЫЕ РЕАЛИЗАЦИИ

Использование в некоторых облачных СУБД предконфигурированных аппаратных решений

SAP Hana Cloud (работает только на конкретном оборудовании)	Clustrix (аппаратно-программная облачная СУБД)	Oracle Exadata (продвигается для частного облака)

РЕАЛИЗАЦИЯ СЛОЯ ХРАНЕНИЯ

- Колоночное хранение
- Хранение в оперативной памяти
- Автоматический кэш с учётом производительности устройств хранения (SDRAM → NAND SSD → 15k HDD → 7.2k HDD → 4.2k HDD → LTO)
- «Распределённые файловые системы» с MapReduce-API как слой хранения
- Специализированные аппаратные реализации слоя хранения (программируемые вентильные матрицы, закрытые аппаратно-программные комплексы)
- Поддержка сверхбольших систем хранения

ФУНКЦИОНАЛЬНОЕ И ЛОГИЧЕСКОЕ ПРОГРАММИРОВАНИЕ

Shan Shan Huang, **Todd Jeffrey Green**, and **Boon Thau Loo**. Datalog and emerging applications: an interactive tutorial / Proceedings of the 2011 ACM SIGMOD International Conference on Management of data (SIGMOD '11), p. 1213-1216, DOI=10.1145/1989323.1989456

Всё чаще используются в контексте СУБД (для разработки и управления СУБД, как внутренние языки СУБД):

R	Erlang	Haskell	O'Caml

VII ПЕРСПЕКТИВЫ

- Стандартизация
- Интерес к Data Intensive Computing

СТАНДАРТИЗАЦИЯ

Отмечены попытки стандартизации PaaS в целом:

Cloud Application Management for Platforms (CAPM) – проект стандарта для PaaS (OASIS)

Пока эти предложения ничего существенного для стандартизации DBaaS не представляют, лишь являются свидетельствами стремления к стандартизации

Перспективный стандарт для DBaaS должен охватить следующие аспекты:

- Терминология, классификация
- Стандарт API обслуживания (развёртывание, запрос ресурсов, регистрация пользователей … - в терминах СУБД)
- Минимальный интероперабельный REST API
- Принципы для языковых API

ИНТЕРЕС К DATA INTENSIVE COMPUTING

Отмечаются первые признаки интереса правительственных и академических организаций к Data Intensive Computing как отдельной, специфической отрасли суперкомпьютинга (высокопроизводительных вычислений, high performance computing)

Отработка технологий Data Intensive Computing будет проходить в средах облачных СУБД

Технологии для больших самоуправляемых DIC-кластеров будут актуальны для облачных СУБД

VIII ИСТОРИЯ

- Что было до…?
- Когда появились?
- На чём закрепились?

ДООБЛАЧНАЯ ЭПОХА

Консолидация баз данных

- Один большой сервер для организации или группы организаций, в разных схемах (или экземплярах) работают разные приложения

Хостинг-провайдеры:

- Типичная услуга: выделение базы mySQL

xAMPP

- *x* =(Linux
  | FreeBSD
  | Windows)
- **A**pache
- **M**ySQL
- **P**HP

Автоматическое выделение экземпляров БД

Выставляются лимиты на объёмы и процессы

Учёт потребления (объёмы)

2006: КОНЦЕПТУАЛИЗАЦИЯ ОБЛАЧНЫХ ВЫЧИСЛЕНИЙ

Август 2006

Amazon запускает *Elastic Computing Cloud*

Август 2006

Эрик Шмидт (президент *Google*) говорит об облачной модели вычислений

Осень 2006

Термин начали использовать ASP:

- *Salesforce.com*
- *Zoho*

… и хостинг-провайдеры

- *Rackspace*
- *Joyent*

Говорят обычно о двух характеристических признаках:

Обработка данных производится на мощностях провайдера

Потребитель не заботится о конкретных выделенных для него ресурсах, ресурсы масштабируются по мере необходимости

Чётких определений нет, некоторыми считается «красным словцом»:

- *«…компьютерная индустрия – единственная отрасль, движимая модой в большей степени, чем индустрия женской моды»*

2008: КЛЕРМОНТСКИЙ ОТЧЁТ

Раз в несколько лет ведущие исследователи баз данных собираются на пару дней в уютном месте и формулируют отчёт о состоянии дел и перспективах в отрасли:

1989 Лагуна-бич

1990 Пало-Альто

1995 Пало-Альто

1998 Асиломар

2003 Лоуэлл

2008 Клермонт

В разделе о перспективных направлениях исследований:

1	2	3	4	5
Пересмотр архитектуры СУБД	Декларативное программирование для новых платформ	Взаимосвязь структурированных и неструктурированных данных	**Облачные службы данных** *(cloud data services)*	Мобильные приложения и «виртуальные миры»

«Весь этот подпункт производит впечатление "застолбления участка", на котором, вполне вероятно, могут находиться золотые жилы»,
С. Д. Кузнецов, 2008

2009: ФОРМИРОВАНИЕ РЫНКА

Marcus Collins. *Cloud Databases: Structure in a Nebulous World* // Burton research report, Gartner, 20 Nov 2009, ID: G00203887

Выделены облачные характеристики, выделены 2 модели обслуживания

- PaaS-СУБД
- SIaaS-СУБД

Дано сравнение трёх участников

- Amazon SimpleDB
- Google AppEngine Data store
- Microsoft Windows Azure Table

Парадигма MapReduce отмечена как ключевая перспективная технология реализации облачных СУБД

2009: CLOUDDB'09

В ноябре 2009 года в рамках ACM CIKM (*Conference on Information and Knowledge Management*) прошёл первый научный семинар "Cloud Data Management"

(несмотря на аббревиатуру, тематические рамки семинара несколько шире облачных СУБД)

Проводится ежегодно:

- 2009 Гонконг
- 2010 Торонто
- 2011 Глазго
- 2012 Мауи
- 2013 Сан-Франциско

- *Персонализация как услуга*
- *Обработка запросов к массовым маршрутным данным на базе MapReduce*
- *Эффективный многомерный индекс для облачного управления данными*
- *Как управление репликами в облаке может использовать Data Grid Protocol*
- *Использование масштабируемого колоночного хранилища для построения распределённого текстового индекса*

2014 *Чикаго* пройдёт 14 апреля в рамках CIKM, CfP до 13 ноября 2014

2010: DATABASE.COM

Декабрь 2010: слой обработки данных для force.com – платформы разработки облачных приложений – выведен в отдельный сервис в домене database.com

- 3 пользователя
- 100 тыс. записей
- 50 тыс. транзакций в месяц

Бесплатно

$10

- Каждые 100 тыс. записей

- Каждые 150 тыс. транзакций в месяц

$10

Внутри – Oracle Database

SOAP, REST API

ODCB, JDBC

2011: NIST'11

Peter Mell and **Timothy Grance**. The NIST Definition of Cloud Computing / Recommendations of the National Institute of Standards and Technology, Sep 2011, NIST SP 800-145, MD 20899-8930

… модель обеспечения повсеместного и удобного сетевого доступа по требованию к общему пулу конфигурируемых вычислительных ресурсов, […] которые могут быть оперативно предоставлены и освобождены с минимальными эксплуатационными затратами или обращениями к провайдеру

Существенные характеристики:

- Самообслуживание по требованию
- Универсальный доступ по сети
- Пулирование ресурсов
- Эластичность
- Учёт потребления

Модели обслуживания

- SaaS
- PaaS
- IaaS

Модели развёртывания

- Частная
- Общественная
- Публичная
- Гибридная

2013: ТЕКУЩИЙ МОМЕНТ

David Linthicum. Why 2013 will be the year of the cloud database // Inforworld, Oct 12 2011

1 марта 2013

- Rackspace покупает ObjectRocket (MongoDB DBaaS)

1 мая 2013

- Закрыта Xeround Database

26 июня

- Выпущена Oracle Database 12*c*

9 октября

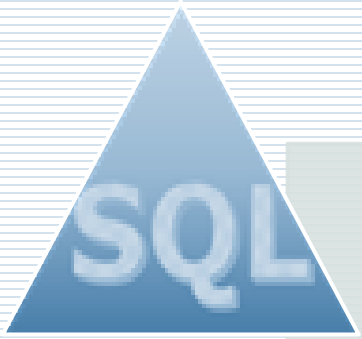
- TransLattice поглощает StormDB

Рост интересна к «большим данным» и “data intensive computing”

IX ТАКСОНОМИЯ

- По реализации
- По модели предоставления
- По модели обработки
- По сфере применения

ПО МОДЕЛИ РЕАЛИЗАЦИИ



В современном контексте относится не столько к языку **SQL**, сколько к *реляционной модели*

"*The idea is that SQL itself is a kludge*" – по мнению веб-разработчиков второй половины 2000-х

NoSQL (2009): СУБД с нетрадиционной моделью хранения для обеспечения горизонтального масштабирования

Модели:

- «Ключ – значение»
- «Документо-ориентированные» (XML, RDF, JSON, …)
- Графовые

NewSQL – доступ к NoSQL-системам на языке SQL
или реляционные системы с поддержкой SQL и shared-nothing

ПО МОДЕЛИ ПРЕДОСТАВЛЕНИЯ

Уникальные

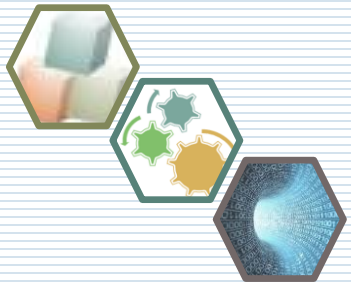
- Развёрнутые в публичном облаке
- Разработанные для частного облака одной организации

Тиражируемые

- Для создания собственного облака

Единый релиз	Поддержка на стороне провайдера
Сокрытие деталей реализации	Только эксплуатационная документация

Релизы, версии, патчи, разные платформы	Поддержка инсталляций
Физический доступ к компонентам	Документация на развёртывание



ПО МОДЕЛИ ОБРАБОТКИ

OLTP

Интенсивная многопользовательская вставка и обновление

Небольшие операции

Стабильная модель

ERP, CRM, биллинг, системы бронирования, core banking…

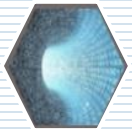
OLAP

Многопользовательская выборка

Большие объёмы выборок

Относительно стабильная модель, известные показатели, измерения

Business Intelligence, мониторинг, DSS …



Batch

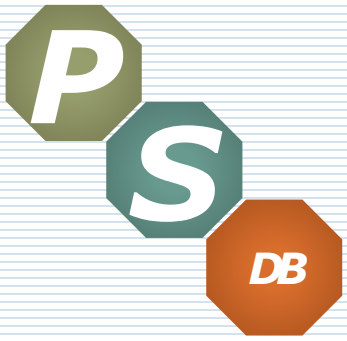
Небольшое количество пользователей, выборка, частая пересборка

Сверхбольшие объёмы [пакетных] обработок

Нестабильная модель

Data Mining, ad-hoc-аналитика, статистический анализ, …

ПО ПРИМЕНЕНИЮ



P — Для использования в рамках облачной платформы

Как правило –
интегрированная часть облачной среды разработки

S — Для использования в облачных приложениях

Особые средства для работы с облачными приложениями
(основные данные, модели доступа)

DB — Общего применения: как классическая СУБД

Должна предоставлять все возможности традиционной СУБД:
язык доступа, средства администрирования, резервирования…

X ПОСТАВЩИКИ И СЕРВИСЫ

- Google
- Amazon
- Salesforce
- Microsoft
- Oracle
- …

GOOGLE

СУБД	Запуск	Реализация	Предоставление	Обработка	Применение
BigTable	2006				
BigQuery	2010				
Megastore	2011				
Cloud Datastore	2012				
CloudSQL	2012				
Spanner	2012				
F1	2012				

GOOGLE (2)

BigTable (2006)

«Разреженная, распределённая, хранимая `map`»:

- `(row:string, column:string, time:int64) → string`

Данные хранятся отсортированными по **row**

Хранилище секционировано по лексикографическим диапазонам

Секция (*tablet*) – единица распределения

Колонка (`column`) – в формате:

- `“Family:Qualifier”`

Используется в большом количестве продуктов Google:

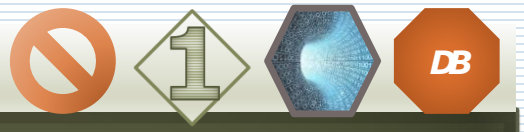
Google Reader	Google Maps	Google Book Search	My Search History	Google Earth
Blogger.com	Google Code hosting	Orkut	YouTube	Gmail

Прототип для Apache HBase

Google

GOOGLE (3)

BigQuery (2010)




```
$ bq query "SELECT word, COUNT(word) as count FROM publicdata:samples.shakespeare >
  WHERE word CONTAINS 'raisin' GROUP BY word"
Waiting on job_dcda37c0bbed4c669b04dfd567859b90 ... (0s) Current status: DONE
+---------------+-------+
|     word      | count |
+---------------+-------+
| Praising      |   4   |
| raising       |   5   |
| raisins       |   1   |
| praising      |   7   |
| dispraising   |   2   |
| dispraisingly |   1   |
+---------------+-------+
```

datasets	jobs	tabledata	tables
delete	get	list	delete
get	getQuery Results		get
insert	insert		insert
list	list	insertAll	update
update	query		list
			path

Google

GOOGLE (4)

Megastore (2011)

NoSQL-система с синхронной репликацией и целостными (ACID) транзакциями внутри секций

Используется как СУБД для интерактивных приложений Google и для сервиса Cloud Datastore

Cloud Datastore (2012)

Публичный сервис, работающий на Megastore

API для node.js, Python (Protobuf), Java (Protobuf), Ruby (JSON)

По состоянию на конец 2013 года находится в статусе “Preview”

Google

GOOGLE (5)

Cloud SQL (2012)

MySQL в из дата-центров Google, предоставляемая заказчикам

Режим репликации

- синхронный
- асинхронный

Ценообразование

- почасовое
  - $0,025 – $3,08 в час
    - $0,24 за ГБ в месяц
    - $0,10 за 1 млн IO
- пакеты ($0,36 – $46,84 в день)

Google

GOOGLE (6)

Spanner (2012)

James C. Corbett, **Jeffrey Dean**, **Michael Epstein**, **Andrew Fikes**, **Christopher Frost**, **J. J. Furman**, **Sanjay Ghemawat**, **Andrey Gubarev**, **Christopher Heiser**, **Peter Hochschild**, **Wilson Hsieh**, **Sebastian Kanthak**, **Eugene Kogan**, **Hongyi Li**, **Alexander Lloyd**, **Sergey Melnik**, **David Mwaura**, **David Nagle**, **Sean Quinlan**, **Rajesh Rao**, **Lindsay Rolig**, **Yasushi Saito**, **Michal Szymaniak**, **Christopher Taylor**, **Ruth Wang**, and **Dale Woodford**. Spanner: Google's Globally Distributed Database // ACM Trans. Comput. Syst. 31, 3, Article 8 (August 2013), 22 p. DOI=10.1145/2491245

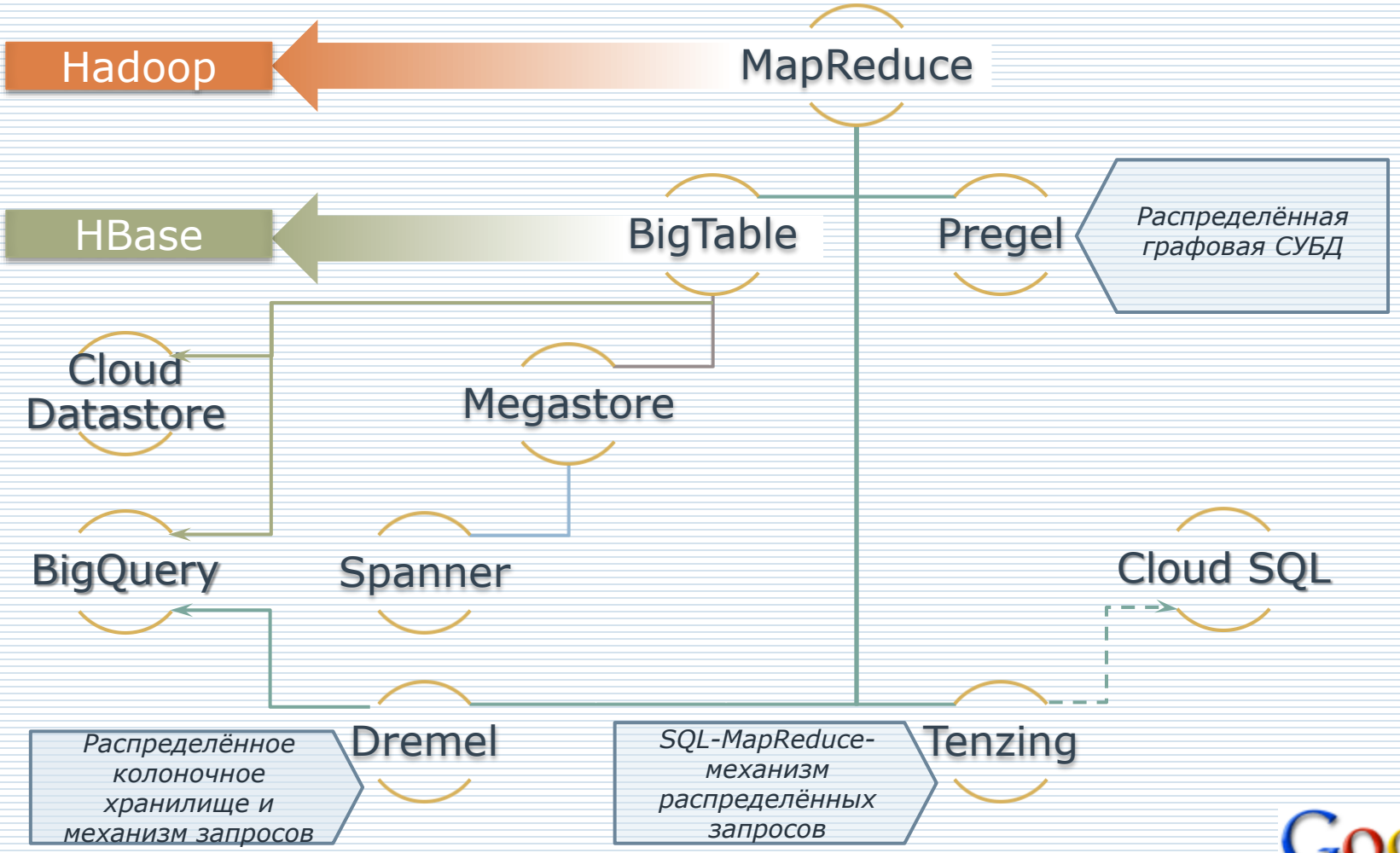
Облачная СУБД – «автоматически распределяющая данные по сети Paxos-узлов», «созданная для работы триллионов записей на миллионах узлов в сотнях дата-центров»

F1 (2012)

Сервис СУБД для рекламных приложений Google, реализованный на Spanner

Google

GOOGLE (5)



Google

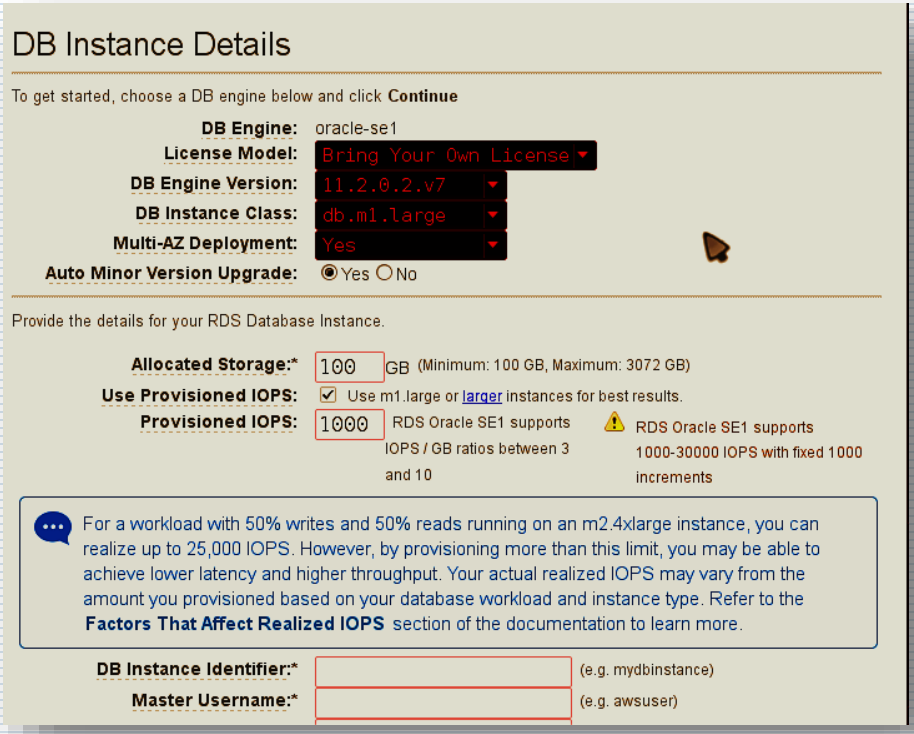
AMAZON

СУБД	Запуск	Реализация	Предоставление	Обработка	Применение
Relational Database Service	2009				
Elastic MapReduce	2009				
Dynamo DB	2012				
ElastiCache	2013				
Redshift	2013				

AMAZON (2)

Relational Database Service (2009)

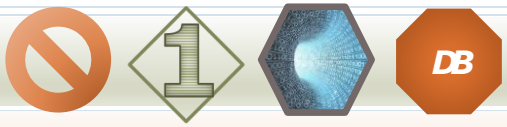
- MySQL
- Oracle Database
  - Standard Edition One
  - Standard Edition
  - Enterprise Edition
- Microsoft SQL Server
  - Express
  - Web
  - Standard
  - Enterprise



amazon

AMAZON (3)

Elastic MapReduce (2009)



Экосистема Hadoop (собственная сборка или MapR, включая HBase, Hive, Pig), с биржей котировок на ресурсы (spot instances)

DynamoDB (2012)

Собственная высокодоступная автомасштабируемая NoSQL-СУБД («ключ-значение»), «согласованная в конечном счёте» через векторные часы, с хранением на SSD, ставшая прототипом для Riak

ElastiCache (2013)

Коммерческая реализация Memcached – распределённого хэш-табличного кэша в оперативной памяти с многоязыковым API

amazon

AMAZON (4)

Redshift (2013)

«Реляционное хранилище данных петабайтных масштабов» – колоночная СУБД без разделяемых ресурсов (shared-nothing) на базе решения ParAccel Analytic Platform (принадлежит с 2013 года компании Actian)

Узел-XL

- 2 CPU / 4,4 ECU
- 2 ТБ DAS
- 15 ГБ ОЗУ
- До 32 узлов
- $0,85 в час
- k$3 за 3 года резерва

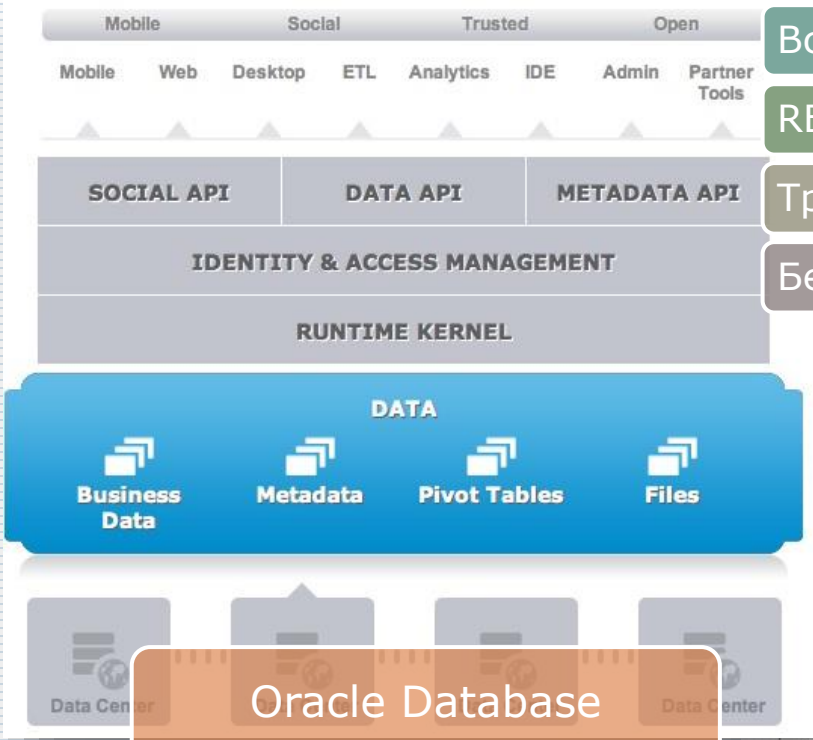
Узел-8XL

- 16 CPU / 35 ECU
- 120 ГБ ОЗУ
- 16 ТБ DAS
- До 100 узлов
- $6,8 в час
- k$24 за 3 года резерва

amazon

SALESFORCE.COM

database.com (2010)



- Все аренды в общей таблице
- REST/SOAP API, SQL
- Триггеры, хранимые процедуры
- Безопасность на уровне записи

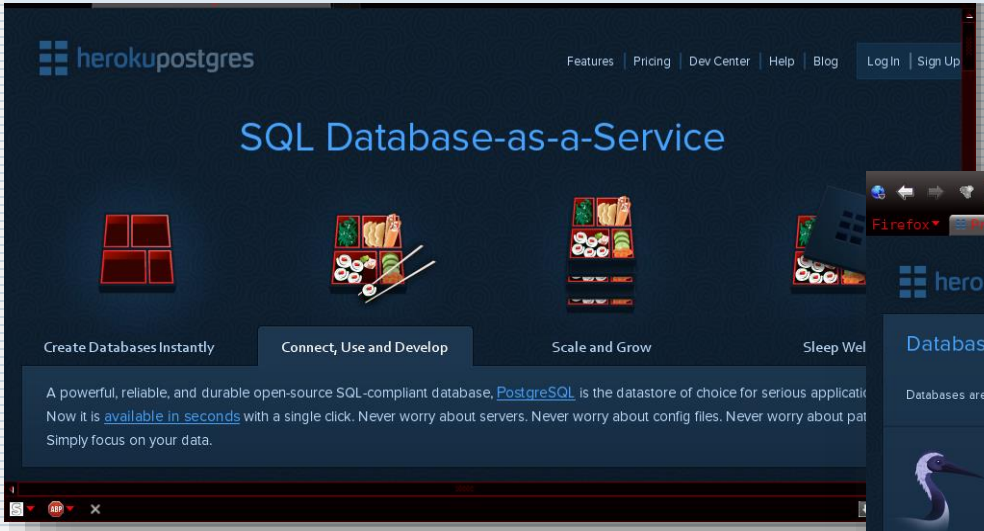
Oracle Database

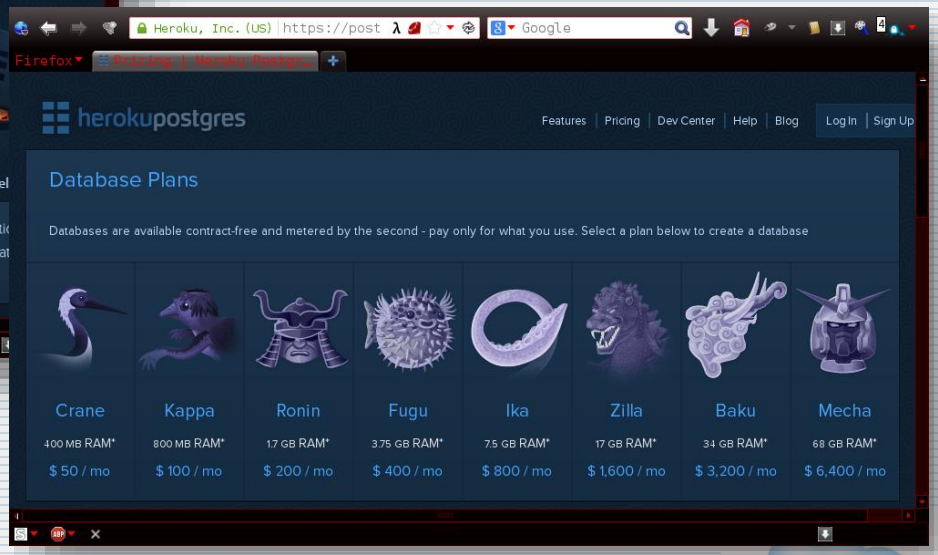
Oracle Fusion Middleware

SALESFORCE.COM (2)

postgres.heroku.com (2013)

Просто хостинг PostgreSQL, но используется в основном для платформы Heroku (но может быть использоваться самостоятельно)





(В платформу Heroku входят также Cloudant, Membase, Redis, MongoDB)

MICROSOFT

SQL Azure (2009)

Специальная редакция Microsoft SQL Server, поддерживается две реплики, SQL, T-SQL, REST/SOAP-API

«Таблицы»

- объекты «ключ-значения»
- распределяемые по узлам
- доступные через REST API
- до 200 ТБ

BLOB

- распределяемые двоичные данные
- монтируются как NTFS-тома
- до 200 ТБ

- Есть проект Cloud SQL Server от Microsoft Research
- Есть материалы для развёртывания SQL Server в частном облаке

ORACLE

Public Cloud Database (2012)

Oracle Database Enterprise Edition 11*g*/12*c*

S5	S20	S50	Полный экземпляр
Пространство: 5 ГБ	Пространство: 20 ГБ	Пространство: 50 ГБ	…
Трафик: 20 ГБ	Трафик: 120 ГБ	Трафик: 300 ГБ	
$175 в месяц	$900 в месяц	2000 в месяц	

ORACLE (2)

Database 12*c* (2013)

В рамках сети хранения с функцией thin provisioning – мгновенно

Живая миграция
в другую контейнерную БД

Подключаемая база данных (PDB)

Подключаемая база данных (PDB)

...

Контейнерная база данных (**CDB**)

XEROUND

Xeround Database (2010–2013)

Технология		
Распределённое хранилище в оперативной памяти	Реализовано в виде виртуальных машин	Выглядит из вне – как одна база MySQL

Было реализовано в IaaS- и PaaS-сервисах		
Rackspace	Amazon EC2	Heroku

2005

- Основана с концепцией «виртуализации данных»

2006

- Названа наиболее многообещающим израильским стартапом

2010

- Запущена Xeround Database Beta

2011

- Запущена Xeround Database

2013

- Сервис закрыт, «*мы работаем над распределённым средством управления данными нового поколения*»

CLUSTRIX

Clustrix Database (2012)

x86-64-узлы 1U с 48 ГБ ОЗУ и 7 SSD (минимум 3 сервера)

Может быть предоставлена на Amazon EC2, Rackspace (устанавливаются аппаратные комплексы)

Горизонтально масштабируемая

Общего назначения (OLTP/OLAP)

Полностью совместимая с MySQL

CLOUDANT

Cloudant

Основные идеи и схема REST API заимствованы у Amazon Dynamo DB

Массачусетская компания, разрабатывает ответвление Couch DB – BigCouch (документо-ориентированная СУБД, написанная на Erlang) и предоставляет её как сервис

BigCouch

BigCouch выложен на GitHub, и можно собрать свой DBaaS

TRANSLATTICE

StormDB (2010)

StormDB – географически распределённый PostgreSQL в режиме «активный – активный», восстанавливающийся быстрым применением журналов изменений

Созданы кластеры на Amazon EC2, есть поставки на x86-64-серверах Dell PowerEdge

Авторы StormDB считают, что DBaaS – это IaaS (не PaaS)

NUODB

NuoDB (2013)

Массачусетская компания Джеймса Старки (ключевого разработчика Interbase), основана в 2007 году)

За Старки закреплён патент на «эластичную облачную СУБД»

В проекте называлась NimbusDB, начиная с релиза-кандидата (2012) переименована в NuoDB

Первый выпуск – для Solaris x86-64, потом выпущены версии для Linux, OS X, Windows

В качестве хранилища поддерживаются файловая система, HDFS, Amazon S3

ДРУГИЕ (1)

Профессиональные хостинги с сервисами обслуживания, холодными резервами, эластичностью (в рамках одного продуктивного узла)

Caspio.com (2009)

Хостинг MS SQL Server, предлагается также совместно с платформой разработки на базе .Net

Enterprise DB (2009)

Хостинг PostgreSQL, с географически удалённым холодным резервом, можно построить репличный кластер (компания также оказывает услуги по техподдержке PostgreSQL)

ДРУГИЕ (2)

MySQL-хостинги

SkySQL (2009)

Хостинг MariaDB (форк MySQL); разработчик MariaDB, оказывает также услуги по платному сопровождению, CTO – Микаэль Видениус

ClearDB (2009)

Геокластеризованный MySQL

Genie DB (2012)

Согласованный в конечном счёте [часами Лэмпорта] геокластер на MySQL

ДРУГИЕ (3)

SAP Hana Cloud (2013)

Колоночная СУБД в оперативной памяти Hana с поддержкой SQL и MDX: 1 ГБ бесплатно, 10 ГБ – €400 в месяц, 750 ГБ - k€16 в месяц

Teradata Agile Analytic Cloud (2013)

Публичный хостинг аппаратно-программных комплексов Teradata

IBM SmartCloud DB (2013)

Публичный хостинг Informix и DB2 для любой платформы

HP Cloud RDB (2013)

Публичный хостинг MySQL от Hewlett-Packard, beta

ДРУГИЕ (4)

Yahoo! Sherpa (2008)

NoSQL-СУБД «ключ-значение» с «настраиваемой целостностью», по идеям сходна с Google Big Table. Используется только внутри Yahoo! Несколько исследовательских статей, находимых по аббревиатуре PNUTS (Platform for Nimble Universal Table Storage)

Garantia Data

Публичный хостинг Memcached и публичный хостинг Redis

Cloudbase.io (2012)

Документо-ориентированная СУБД с гео-API, ориентированная для мобильных приложений

XI ИССЛЕДОВАТЕЛИ И ИССЛЕДОВАНИЯ

- Судипто Дас
- Филип Бернштейн
- Стив Бобровский

- Обзор публикаций

СУДИПТО ДАС

Целиком исследования посвящает облачным СУБД, предложил множество проектов

В основном публикуется совместно (Амр Эль-Аббади, Филип Бернштейн и др.)

Соавтор книги “Data Management in the Cloud” (MC, 2013)

Основной проект – ***Elastras***, облачная СУБД с поддержкой ACID и SQL со слоем хранения на HDFS

Лауреат Премии ACM имени Джима Грея за лучшую диссертацию (2013):

«…в его диссертации заложен фундамент для построения масштабируемых транзакционных баз данных в мультиарендных облачных инфраструктурах. В ней разработаны техники обеспечения масштабируемости как на основе статического, так и динамического секционирования по вычислительным узлам и обеспечения эластичности посредством живой миграции. Комбинация этих техник обеспечивает эластичную масштабируемость без ущерба транзакционной согласованности. Работа также примечательна широтой охвата и полноценной реализацией, подтверждённой тщательной оценкой производительности»

ФИЛИП БЕРНШТЕЙН

Ветеран исследований СУБД, соавтор 5 из 6 отчётов ведущих исследователей о состоянии дел в отрасли СУБД

Ведёт в Microsoft Research проект **Cloud SQL Server**
С 2012 года в Microsoft Research работает и **Судипто Дас**

СТИВ БОБРОВСКИЙ

Известен книгами по архитектуре Oracle Database (середина 1990-х)

С середины 2000-х ведает миру об устройстве слоя хранения CRM Salesforce.com и облачной СУБД Database.com

ОБЗОР ПУБЛИКАЦИЙ

В «Цифровой библиотеке» ACM зафиксировано более **200** публикаций, содержащих в названии или реферате в каком-либо виде упоминание облачных СУБД, мультиарендных СУБД, СУБД в модели PaaS

Наиболее примечательные публикации (**60**) собраны в подшивку (пользователям ACM DL может быть предоставлена по запросу на mailto:anikolaenko@acm.org)

Типичные темы публикаций:

- Проект *XXX* облачной СУБД
- SQL vs NoSQL в облачной среде
- Новый язык *xxSQLxx* для облачных сред
- Учёт потребления, экономические аспекты
- Безопасность, разграничение доступа в условиях мультиарендности
- Научные БД в облаках

ОБЗОРНЫЕ ПУБЛИКАЦИИ

2009

- **Edward P. Holden**, **Jai W. Kang**, **Dianne P. Bills**, and **Mukhtar Ilyassov.** Databases in the cloud: a work in progress / In Proceedings of the 10th ACM conference on SIG-information technology education (SIGITE'09). ACM, New York, NY, USA, 138-143. DOI=10.1145/1631728.1631765

2010

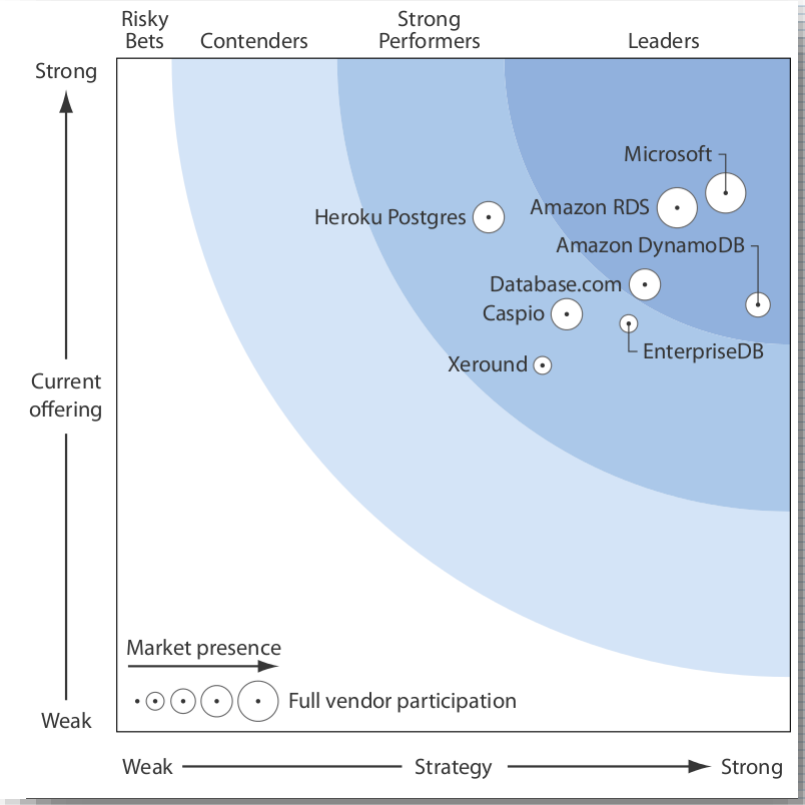
- **Daniel J. Abadi**, **Michael Carey**, **Surajit Chaudhuri**, **Hector Garcia-Molina**, **Jignesh M. Patel**, and **Raghu Ramakrishnan**. Cloud databases: what's new? Proc. VLDB Endowment, **3**, 1-2 (September 2010), 1657-1657

2011

- **Edward P. Holden**, **Jai W. Kang**, **Geoffrey R. Anderson**, and **Dianne P. Bills**. Databases in the cloud: a status report / Proceedings of the 2011 conference on Information technology education (SIGITE'11). ACM, New York, NY, USA, 171-176. DOI=10.1145/2047594.2047642

АНАЛИТИКА

Noel Yuhanna. The Forrester Wave: Enterprise Cloud Databases, Q4 2012 // Forrester



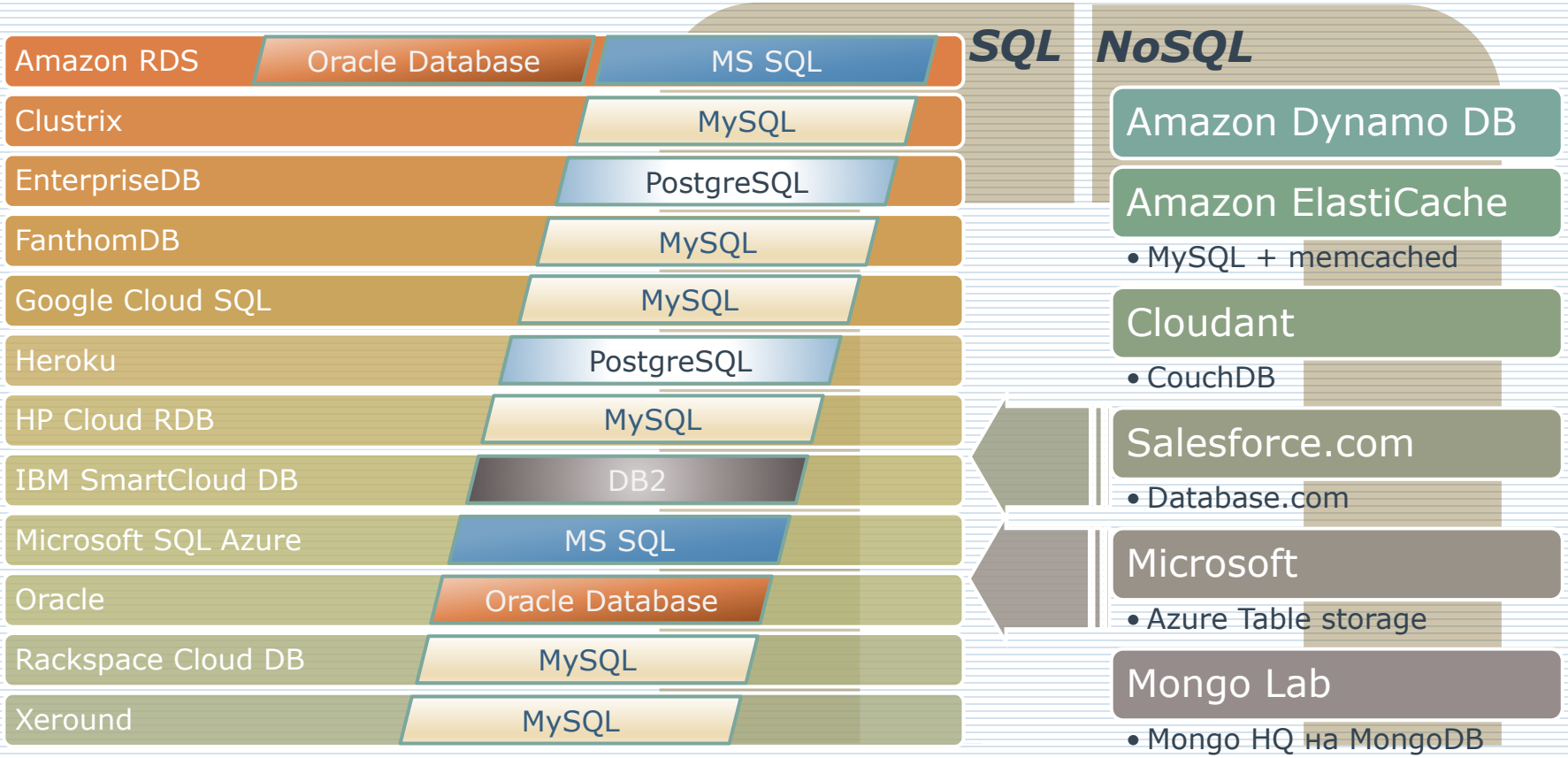
Активное предложение на рынке до апреля 2012 года

Более 50 заказчиков (средних и малых организаций), использующих СУБД в продуктивном режиме

10 упоминаний респондентами Forrester за последний год

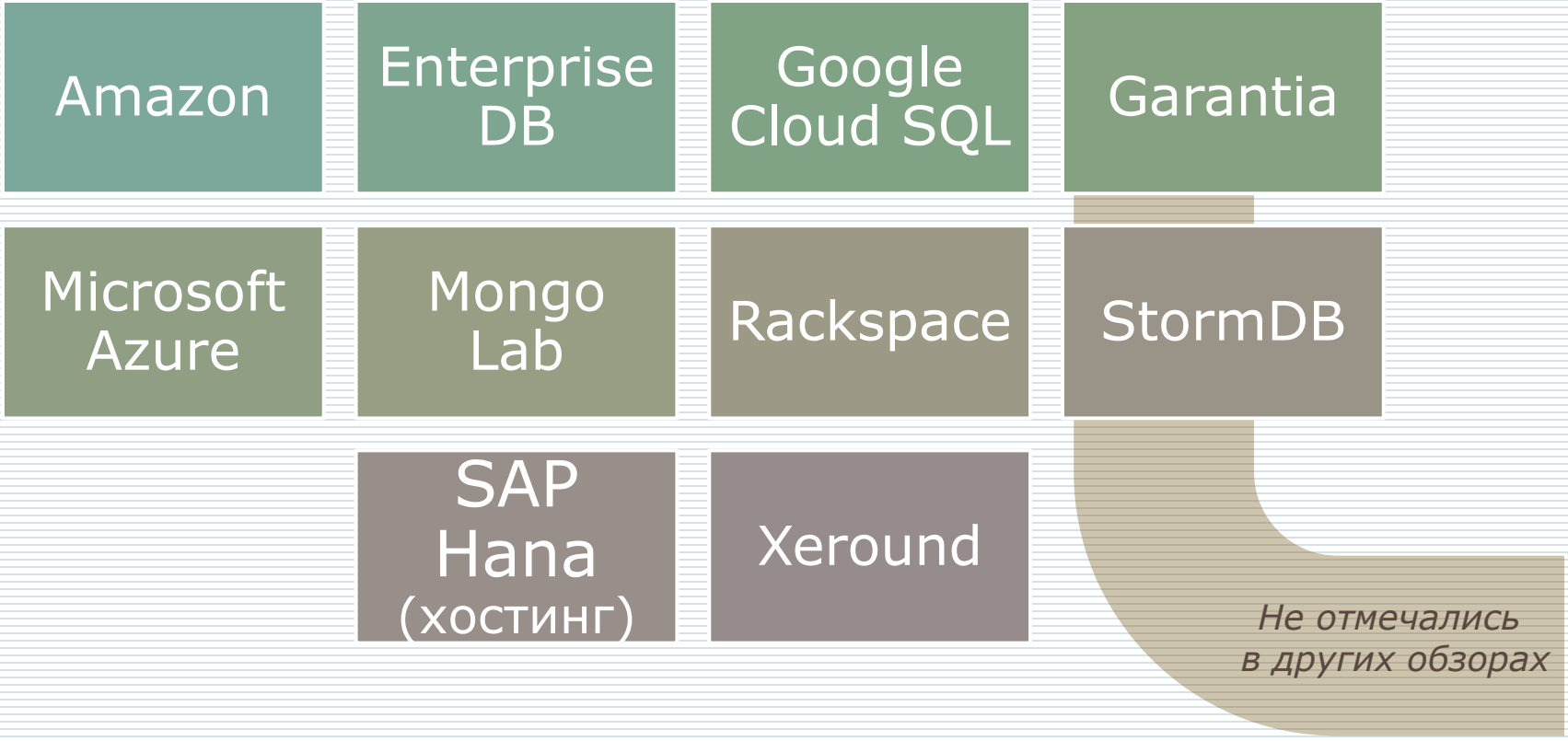
ОБЗОРЫ В ПЕРИОДИКЕ (1)

Derrick Harris. Cloud databases 101: Who builds 'em and what they do // Gigaom, Jul 20 2012



ОБЗОРЫ В ПЕРИОДИКЕ (2)

Brandon Butler. 10 of the most useful cloud databases // Network World, Dec 19 2012



ОБЗОРЫ В ПЕРИОДИКЕ (3)

Klint Finley. 7 Cloud-Based Database Services // Read Write Web, Jan 12 2011

Написан по следам объявления о запуске Database.com, по-видимому, первый обзор облачных СУБД в периодике

Database.com

Xeround

Microsoft Azure SQL

Amazon SimpleDB

Google DataStore

ClearDB

CouchOne

Не отмечена в других обзорах

ОСВЕЩЕНИЕ В ПЕРИОДИКЕ

The Register – британская онлайн-газета, всеохватно следящая за всеми новинками в ИТ, с особым интересом к стартапам и крупным ИТ-проектам

Dbms2.com Курта Монаша – каждодневные наблюдения за рынком и исследованиями в области СУБД (с уклоном к коммерческим аспектам)

Журнал «**Открытые системы. СУБД**» – целенаправленно следит за новостями на рынке облачных СУБД (Леонид Черняк)

Citforum.ru – публикует (публиковал?) статьи и переводы с комментариями С. Д. Кузнецова

На основе обзорной части данного доклада подготовлена и опубликована статья:
- **А. Николаенко.** Год облачных СУБД // Открытые системы. СУБД, №9, 2013, сс. 42-47

МОСКОВСКАЯ СЕКЦИЯ ACM SIGMOD 21 НОЯБРЯ 2013

mailto:
anikolaenko@acm.org