



PERCONA
Performance Consulting Experts

Системы хранения данных в MySQL

Oct 6-7, 2008

HighLoad++

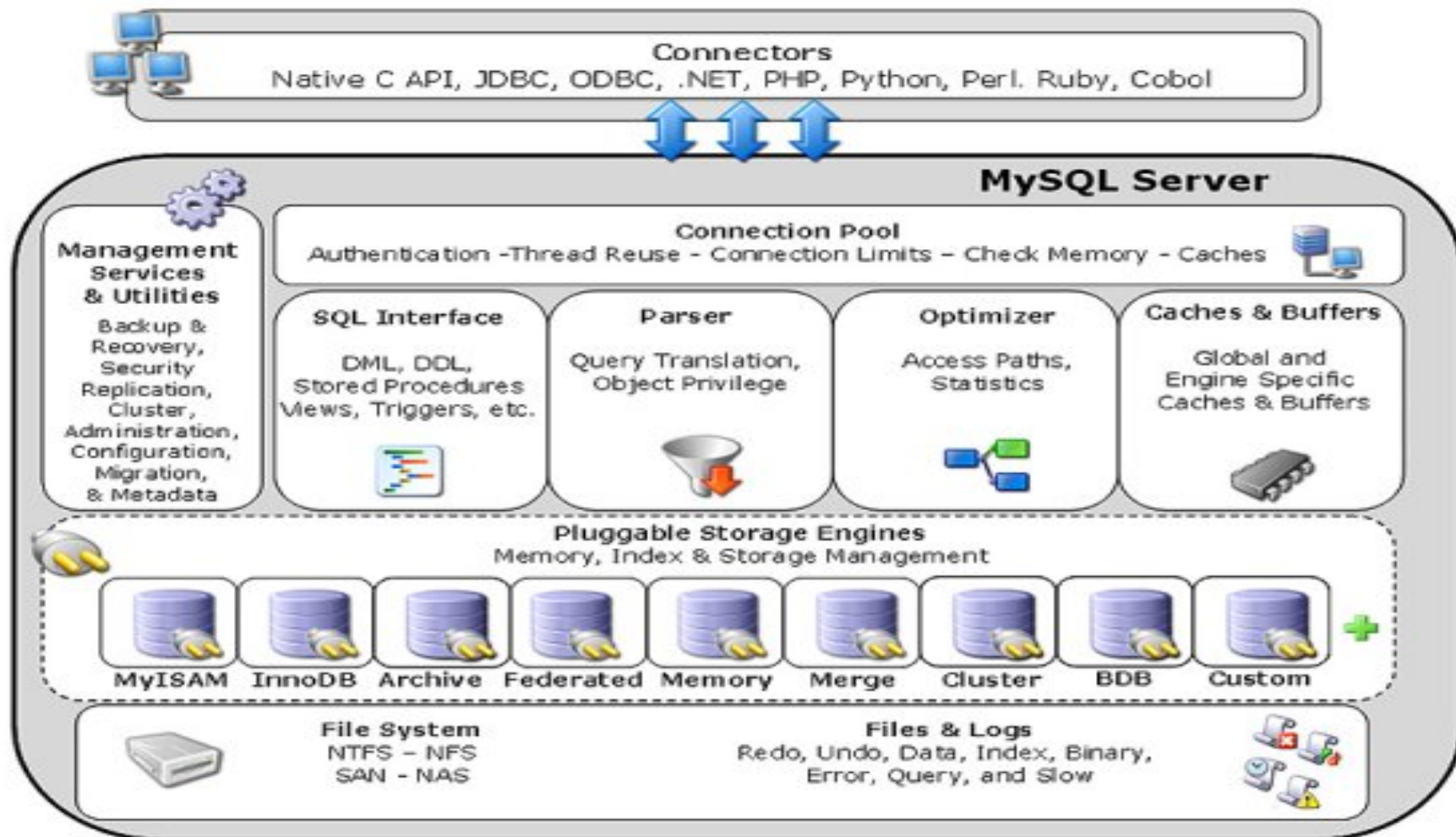
Moscow, Russia

by Peter Zaitsev, Percona Inc

О системах Хранения Данных

- MySQL Сервер состоит из 2х уровней
 - “SQL ” - Функции верхнего уровня
 - “Системы Хранения (Storage Engines)” - Хранение транзакции блокировки итд
- Начиная с MySQL 5.1
 - Системы хранения могут быть модульными
 - Можно компилировать и распространять отдельно
 - MySQL активно содействовал их созданию как внутри так и снаружи
 - Так что их развелось много
 - Партнеры MySQL выпускают свои системы (часто с закрытым кодом)

Архитектура MySQL



Что такое системы хранения

- Системы хранения отвечают за **хранение**
 - Могут реализовать разные концепции хранения и доступа – компрессию, удаленное хранение итд
- Не могут реализовывать ф-ии верхнего уровня – сортировка группировка JOIN
 - В будущем планируется снять эти ограничения
- Некоторые компании (например Kickfire) перехватывают полностью выполнение запроса что позволяет реализовать другие методы выполнения.

Путь систем хранения

- Подход MySQL к расширяемости достаточно необычен
 - Практически нет модульного расширения языка индексов итд
 - Однако широкий выбор систем хранения данных
- Достоинства
 - Разным приложения имеют разные требования к данным
 - перманентность, транзакции, блокировки, компрессия итд
- Недостатки
 - Потеря производительности
 - 2х вазный протокол фиксации транзакций (локально) два лога
 - Сложность
 - Разработка и тестирование (все эти комбинации)
 - Выбор систем хранения для приложения
 - Обслуживание – бакап, балансировка настроек сервера итд

Типы систем Хранения

- Системы Хранения Общего Назначения
 - Транзакционные
 - InnoDB, Falcon, PBXT, Maria (в будущем)
 - Не транзакционные
 - MyISAM, ISAM (устарел), Maria(сейчас)
- Кластерные
 - NDB, ScaleDB (Закрытый Код)
- Системы Хранения Специального Назначения
 - Memory, Federated, Archive, Blackhole, CSV, NitroDB(CS), InfoBright, Queue, Graph(CS), SphinxSE ...

Как использовать на практике

- Выберите основную SE для приложения
 - InnoDB наиболее стандартный выбор сейчас
- Используйте другие для того для чего они хороши
 - InnoDB - Доступ с высокой конкуренции, надежность
 - MyISAM – компактная, нет транзакций быстрое обновление – не критичные и временные данные
 - MEMORY – временные таблицы
 - Federated – удаленный доступ к данным (редко и простые запросы)

Системы Хранения

Обзор систем хранения общего назначения

MyISAM

- Традиционная система хранения MySQL
- Впервые появилась в MySQL 3.23.x
 - Кто из вас работал с этой версией ?
- Основана на ISAM которая основана на UNIREG создано более 15 лет назад
- Табличные блокировки, Не Транзакционная, Не устойчива к сбоям
- Компактная (типично 2-3 раза меньше InnoDB), Быстрые обновления, Может быть использована в режиме только для чтения

Когда использовать MyISAM

- В случае только чтения или в основном чтения когда важен размер данных
 - Замечание: MyISAM не всегда быстрее чем InnoDB для чтения
- Когда нужна быстрая запись
 - Но не чтение и запись для одной таблицы одновр.
 - Журналирование, Временные Таблицы, Обработка данных
- Когда время восстановления данных не критично
 - Большие MyISAM таблицы могут требовать часов для восстановления после сбоя.

InnoDB

- Была создана Heikki Tuuri
- Теперь принадлежит Oracle Corp
- Несколько лет практически не получал развития однако на MySQL UC вышло обновление
 - Плагин с компрессией и быстрым созданием индексов
- Провинутая система с поддержкой транзакций
 - MVCC, блокировки на уровне строк, кластеринг данных
- Автоматическое восстановление при сбое
- Поддержка внешних ключей (Foreign Keys)

Когда использовать InnoDB

- Во многих случаях хороший выбор по умолчанию
- Когда нужны транзакции или внешние ключи
- Когда нужна высокая конкурентность
 - Так что читатели не блокируют писателей
- Если не хочется терять данные при потере питания
- Таблицы (особенно индексы) больше чем MyISAM
 - Часто 2-5 раз больше
- Сложное восстановление при повреждении
- Загрузка данных и создание индексов медленно

Falcon

- Дизайн Jim Starkey – создатель FireBird
- Основная планируемая транзакционная система MySQL
 - Но официально не называется конкурентом Innodb
- Фокус дизайна – работа с большим объемом памяти и многими процессорами
- Много инновационных (непроверенных?) дизайнерских идей
- Нет поддержки кластерных ключей, не может использовать Covering Indexes
- По прежнему не стабильна.

Когда использовать Falcon ?

- Слишком мало реального использования чтобы сказать
 - Хотя ряд наших клиентов уже использует Falcon на Slave серверах для тестов
 - А вы ?
- Будет сложно заменить InnoDB для многих приложений
- Может быть более масштабируем чем InnoDB если проблемы не будут исправлены
- Более быстрый для маленьких транзакций (в теории)

Maria

- Разработка ведется Michael Widenious
 - С небольшой группой старожилов MySQL
- Гибрид MyISAM и InnoDB по идеологии
 - MyISAM – компактное хранение, структура
 - InnoDB - MVCC, транзакции, восстанавливаемость
- Может использовать страничную организацию
- Поддержка транзакций будет опциональна
- На данном этапе доступны “Crash Safe” версия
- Пока что не оптимизированна
- На данном этапе Альфа

Для чего будет хороша Maria ?

- Еще менее используется чем Falcon
- Хорошая замена для MyISAM
 - Системные таблицы, кэшируемые большие временные таблицы
- Будет хорошая замена InnoDB для ряда приложения
 - Когда кластеринг по первичному ключу не критичен
- Легкость использования MyISAM
 - Восстановление таблиц, Перемещение таблиц между серверами на уровне файлов.

PVHT

- Разработана Paul McCullagh from PrimeBase
- Специально для MySQL (не адаптация)
- Была “переписана” 3 или 4 раза
 - Последняя версия ACID по умолчанию
- Много инновационных идей
 - Запись данных один раз, лог на транзакцию
- Фокус на эффективной работе с блобами
- Пока что весьма не стабильна

Для чего использовать РВХТ ?

- Использовать вместе с MyBS – хранение Блобов (например файлов) в базе данных
- Логгинг данных (особенно большие записи)
- Может хорошо работать с SSD дисками
- Будет видно лучше когда будет более готова к реальному использованию
- В некоторых бенчмаках показывала очень хорошие результаты.

What is about SolidDB

- Last year we looked into SolidDB
 - And this year we removed it from consideration
- SolidDB for MySQL project stopped after Solid Technologies was bought by IBM
 - So it never become real alternative
- The code is available on SourceForge
 - But does not have user community to move it forward.

Системы Хранения

Использование специальных систем хранения данных в MySQL

MEMORY

- Хранит данные в памяти
 - При перезапуске сервера теряется содержимое
- Используется неявно сервером
 - Для разрешения многих сложных запросов
- Сохранение временных данных
- Осторожно – только строки фиксированного размера. Есть патч от Ebay решающий это
- Можно пре-загружать данные для быстрого доступа
- Аккуратно использовать с репликацией.

ARCHIVE

- Быстрое и компактное сжатое хранение данных
- Нет поддержки индексов (только сканирование)
- Не блокирующие вставки
- Может занимать существенно меньше MyISAM
- Хорошо работает для хранения логов
 - Одна таблица на день или Partitions в MySQL 5.1 для эффективности

FEDERATED

- Доступ к данным хранящемуся на удаленном сервере
- Хорошо подходит когда нужно несколько строк с удаленного сервера
- Плохо работает с JOIN – требуется много обращений к удаленному серверу
- Опасайтесь больших результатов запросов к удаленному серверу
 - Результат полностью материализуется в памяти сервера инициатора запросов.

BLACKHOLE

- Изначально была создана как пример и для тестов производительности
- Затем оказалась полезной для организации фильтрованной репликации
 - Но не без проблем
 - **ALTER TABLE TBL ... ENGINE=MYISAM**
 - Сконвертирует таблицу в MyISAM

Тесты производительности

**Интересно какова
производительность ?**

DBT2 Размер данных

- Загружаем 200 “складов” от DBT2
- РВХТ упала во время загрузки
- Maria удивляет большим объемом данных

DBT2 Время загрузки

- 200W Время загрузки (минуты)
- 2*Dual Core Xeon 5148, 16G Ram
 - 8GB выделенно на буфера
- Быстрое создание индексов **не** использовано для InnoDB

DBT2 Results

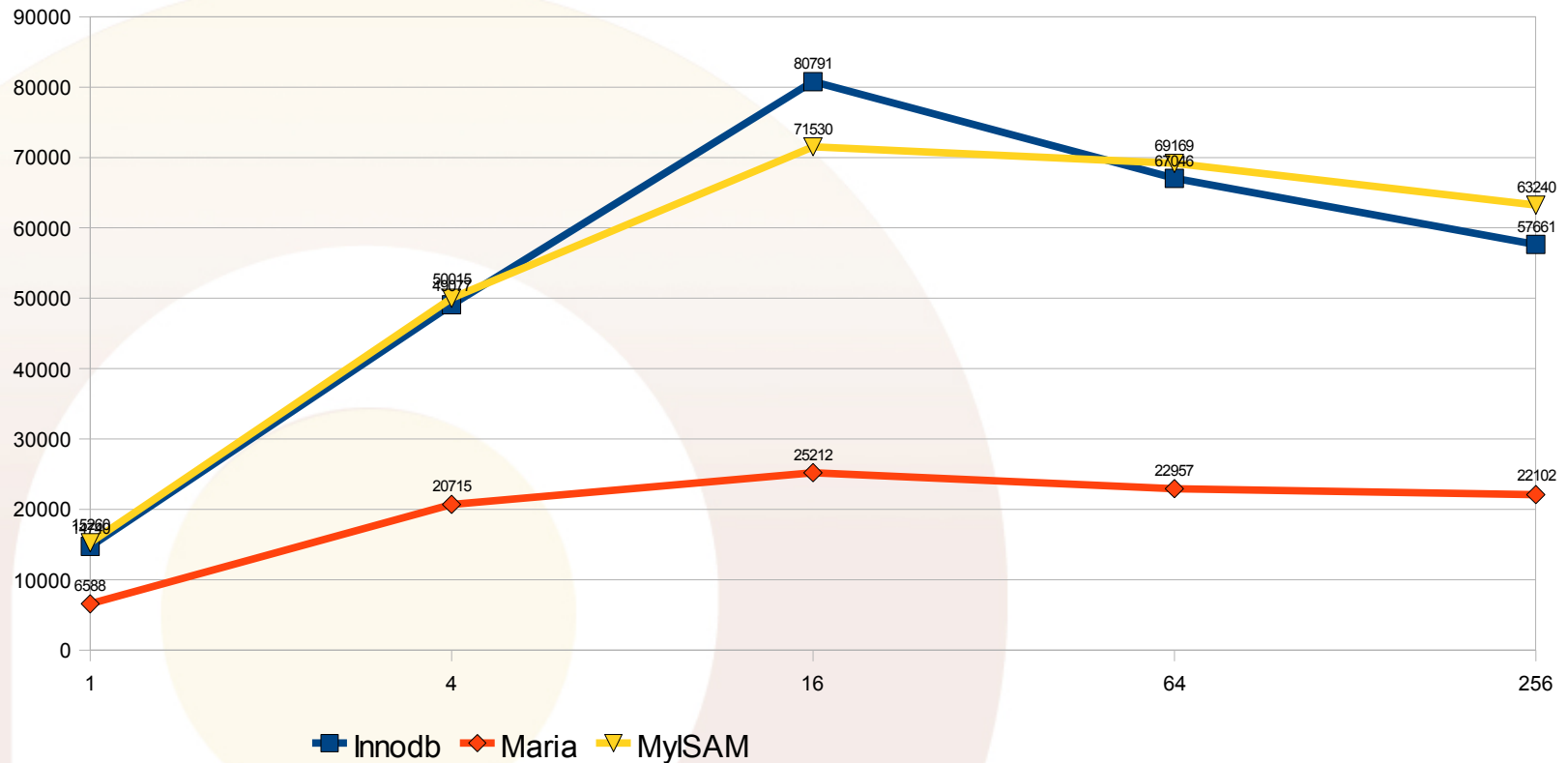
- DBT2 10W Результаты (CPU интенсивные)
- Falcon значительно улучшился с прошлого года
- 5.1 немного медленнее чем 5.0 с Innodb
- Maria показала весьма не стабильные результаты

Микро тесты пр-ости

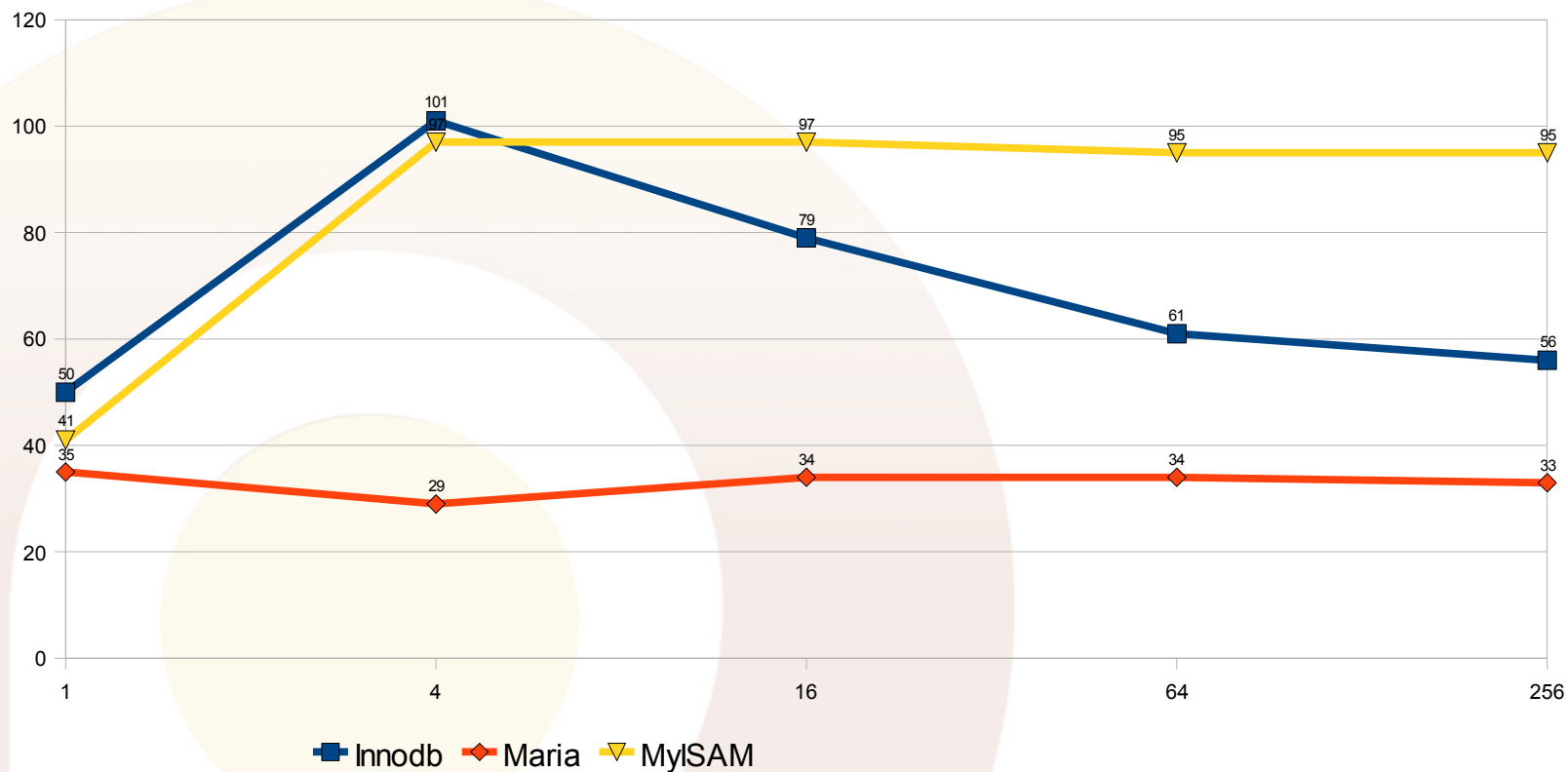
- CPU интенсивные, Запросы/Сек
- 1.000.000 строк в таблицы
- Dual Quad Core Xeon
- Смотрим типичные типы доступа к данным
- Детали по схеме и запросов доступны на
 - <http://www.mysqlperformanceblog.com/files/benchmarks/ph>
- Falcon не включен так как он падал на ряде тестов
- PBXT и вовсе висло при загрузке данных

Сканирование таблицы

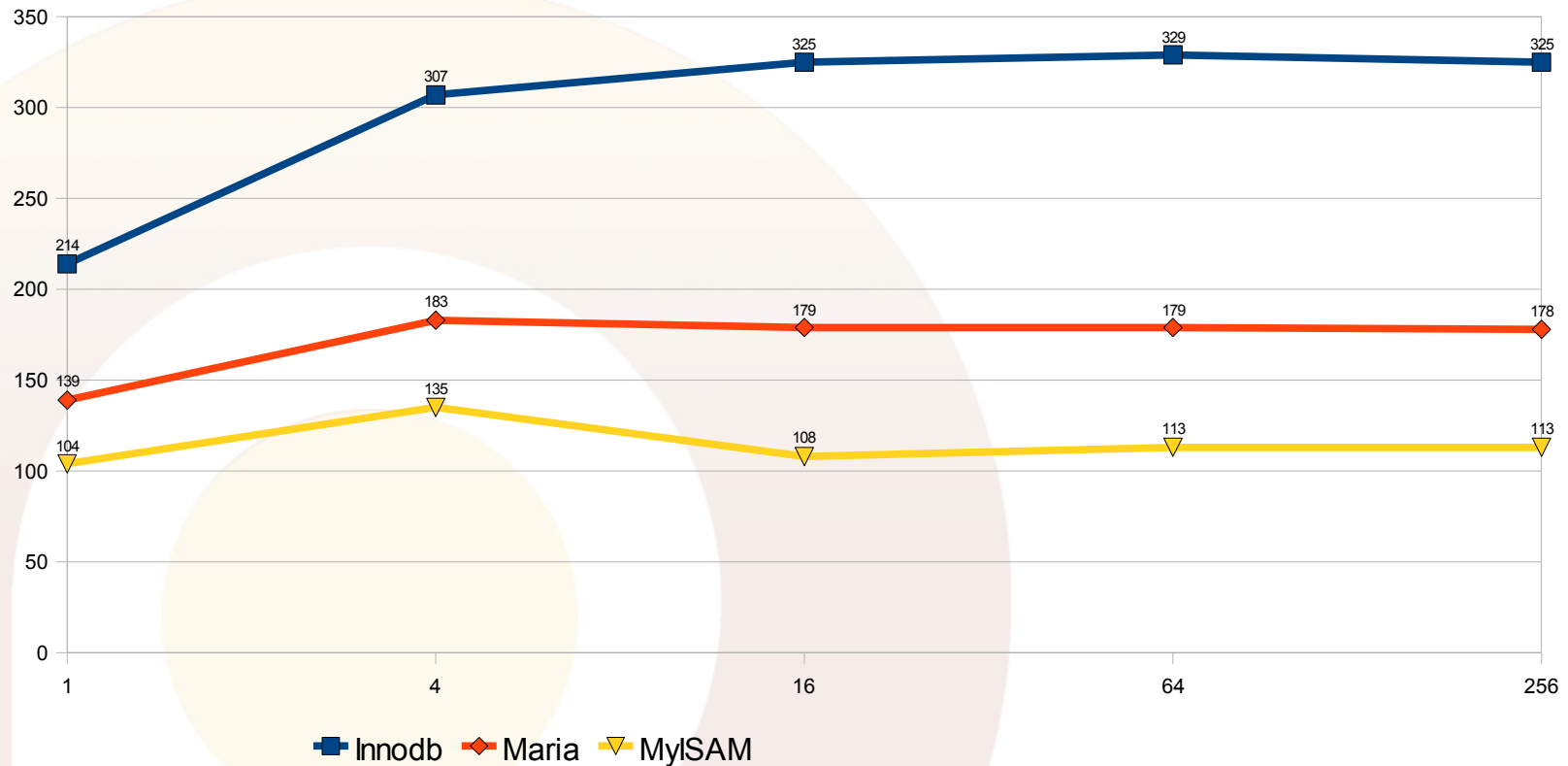
Доступ по первичному ключу



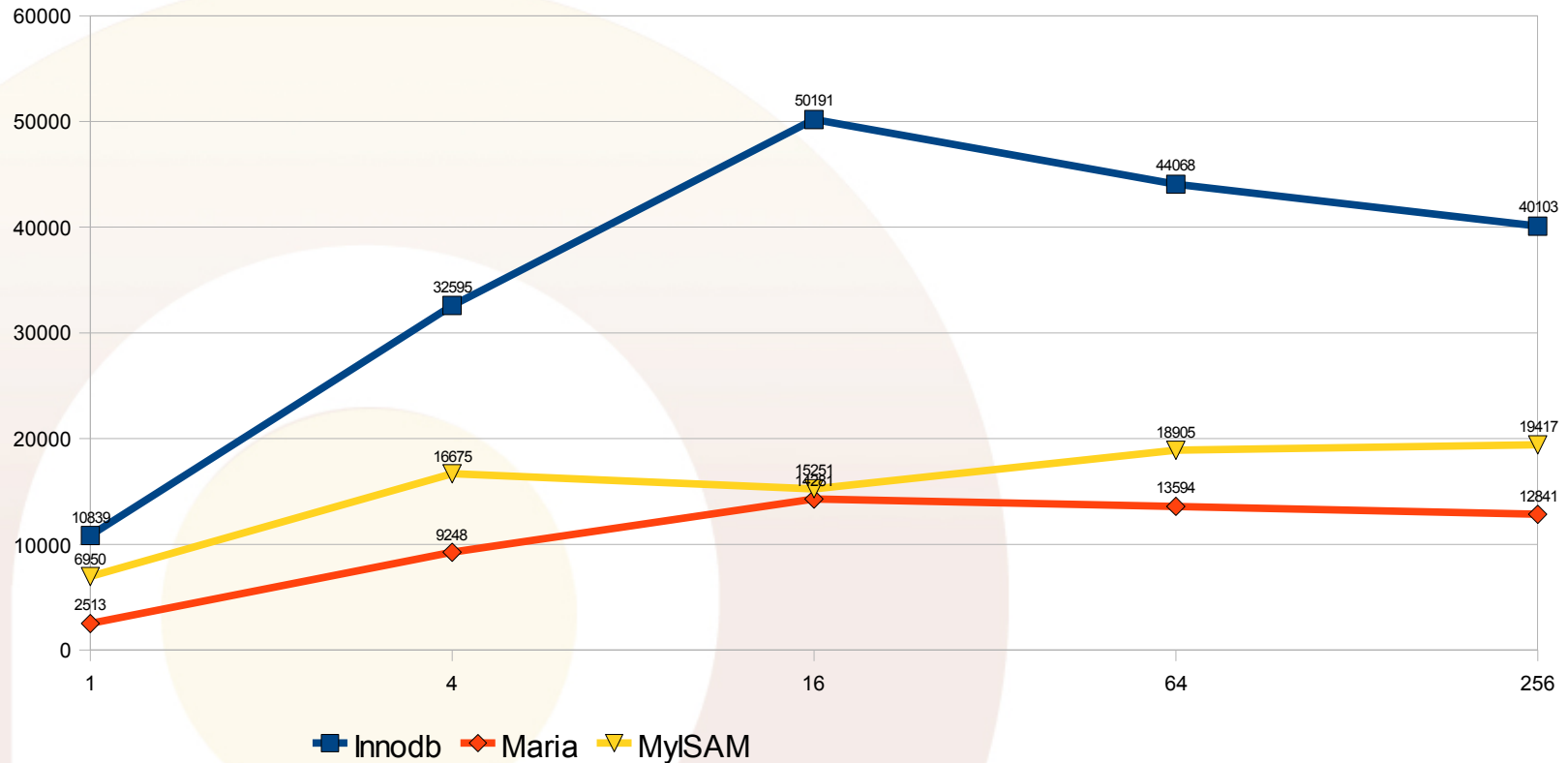
Доступ по вторичному индексу



Доступ по covering index



Досту по ключу с LIMIT



Спасибо что пришли

- Вопросы ? Предложения ?
 - pz@percona.com
- Да мы занимаемся консалтингом в области производительности
 - <http://www.percona.com>
- Посмотрите нашу книгу
 - Полностью переписанное издание

