

Дрессируем Greenplum

Полезные советы для тех, кто завёл свой первый Greenplum



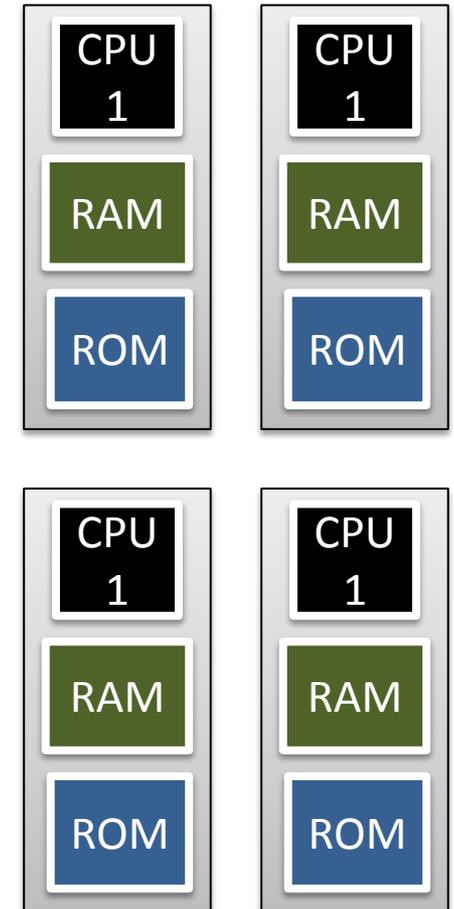
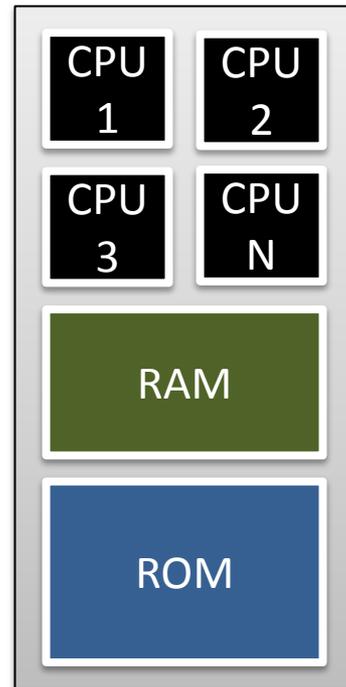
Кто такой Greenplum?

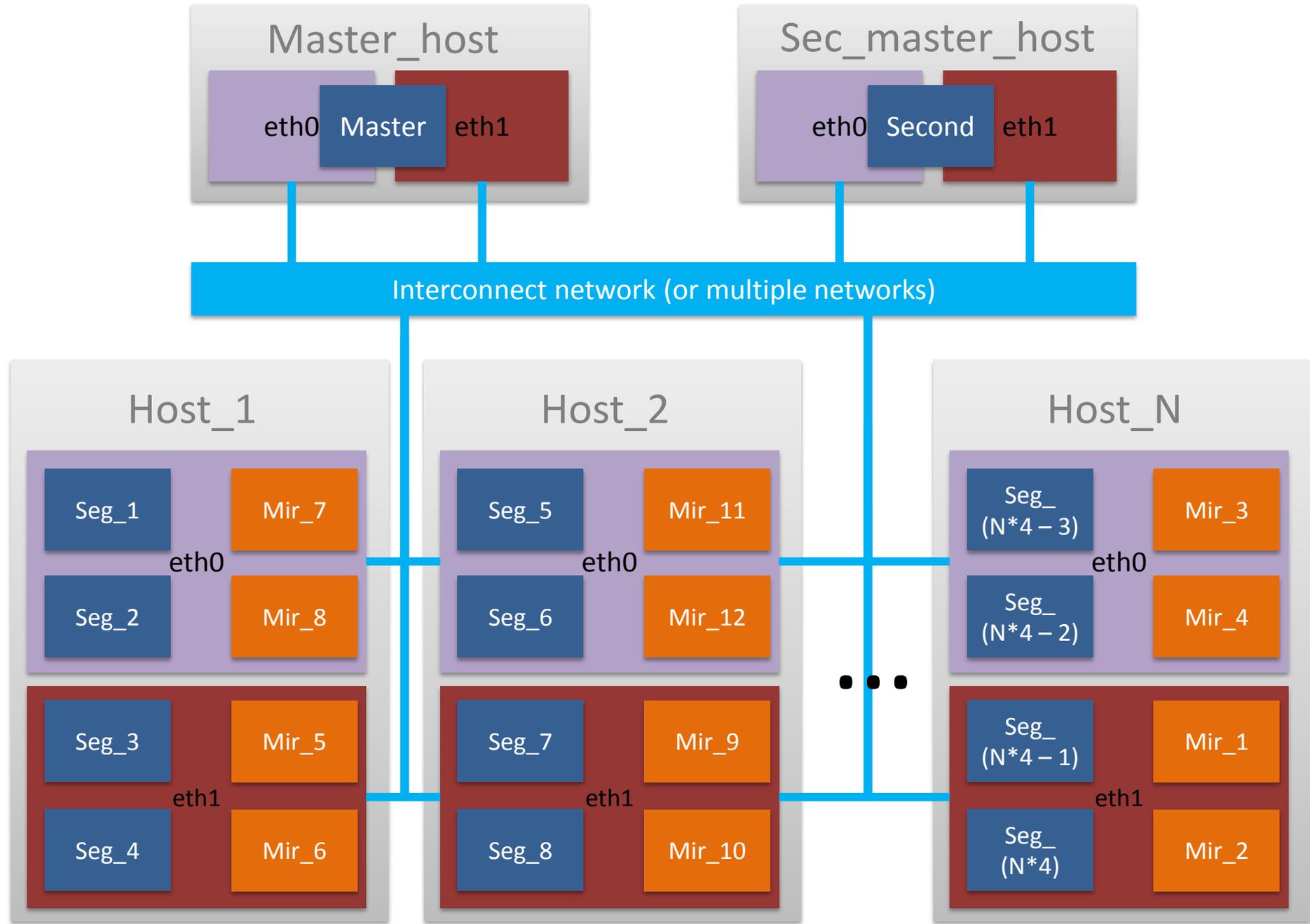
- Кластерная СУБД на базе PostgreSQL
- Форк Postgres 8.3 с большим числом бэк-портов из более новых версий
- Колоночное и строковое хранение
- Есть полное, гибкое резервирование
- Open source – C, Python (обязка)

Shared-nothing MPP on SMP

SMP (symmetric multiprocessing) MPP (massive parallel processing)

- Горизонтальная масштабируемость
- Линейный рост производительности
- Отказоустойчивость





Почему люди заводят Greenplum?

Greenplum создан для аналитики:

- Сложные запросы, обрабатывающие большие объёмы
- Небольшое число QPS

Задачи:

- Клиентская аналитика
- Формирование отчётности
- Анализ оттока клиентов
- Предиктивный анализ
- Формирование маркетинговых кампаний и т.д.



at&t



Bakrie Telecom



Bloomberg



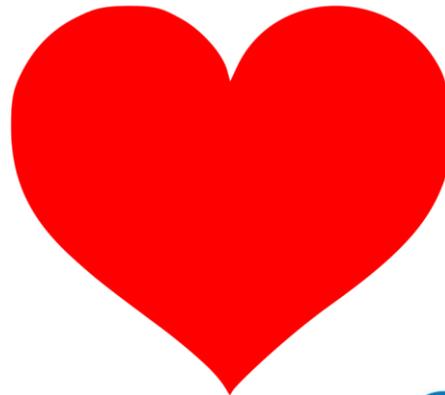


Tinkoff



Интеграция

Сегменты и мастер Greenplum – PostgreSQL-инстансы. Это обеспечивает интеграцию с любым ПО, поддерживающим PostgreSQL.



Когда Greenplum использовать не стоит

- НЕ является шардированной заменой для OLTP СУБД
- НЕ справится с большим QPS
- НЕ справится с очень большим числом (> 1000) одновременно работающих запросов



Где он будет жить?

- Любые x86
- CPU/SPACE - исходя из задач, можно подправить компрессией
- RAID10, RAID1 – хорошо
- RAID5, RAID6, JBOD – плохо
- Виртуализация – хуже



Поздравляем, у вас Greenplum!

- Mirror strategy
- Ваши данные:
 - Вынести каталоги на отдельные устройства
 - Отдельные filesystems на SSD, Hadoop, лентах
 - Ключ распределения в таблицах – очень важно!

Наблюдаем за питомцем

- Мониторинг событий СУБД - статус, упавшие сегменты
- Мониторинг состояния – длительные сессии, блокировки
- Мониторинг железа – собирайте и храните все метрики:
 - CPU
 - Net I/O
 - Disk Latency, I/O%, IOPS, по возможности SMART
- Обеспечьте себе возможность углублённого анализа всех собираемых метрик базы, железа, запросов и логов

Он растёт



- Экспанд – это не страшно
- Даунтайм – до 30 минут
- Редистрибуция таблиц - в фоне

Ваши данные

- Собирайте исторические размеры всех объектов БД на ФС
- Skew – очень плохо:
 - Выбран плохой ключ распределения – дата, тип клиента – всё
 - Аномалии в данных
 - Некорректный рестор из бекапа
- Ищите skew по:
 - Размерам дата-файлов на сегментах (см п.1)
 - Числу строк на сегментах (служебное поле `gp_segment_id`)
 - Аномалиям в нагрузке на железо
- Автоматизируйте поиск skew!

Хороший джойн – локальный джойн

- Джойны по ключам распределения работают локально
- Нелокальные джойны:
 - Redistribute одной или обеих таблиц
 - Broadcast одной из таблиц (меньшей)
- Хинтов нет, но есть немного хаков
- Хорошая практика – автоматизированный анализ самых долгих запросов и алерты на broadcast/redistribute



Где данные?



- Heap tables: много маленьких update, delete
- Append-optimized tables:
 - Batch insert
 - Column store
 - Compression
- Используйте партиции и разные filesystems

Ласка & смазка

- MVCC: следите за bloat'ом
- Частые операции с метаданными = частый VACUUM каталогов
- Собирайте статистику после больших изменений
- Настройте ILM: репартиционирование, архивирование
- Настройте управление spill-файлами
- Отнимите у пользователей ~~клавиатуру~~ ресурсы – настройте ресурсные очереди/группы

PXF – 5.0's new feature

- PXF – загрузка в параллель в/из:
 - Hive
 - HDFS
 - Hbase
 - JDBC
 - ...
 - Whatever (можно писать свои плагины)

Nifi

DataNode

DataNode

DataNode

DataNode

Hadoop

GP seg

GP seg

GP seg

GP seg

GP seg

ADB

GridNode

GridNode

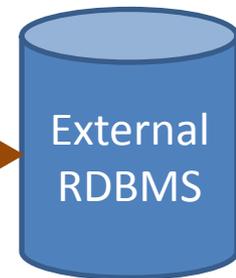
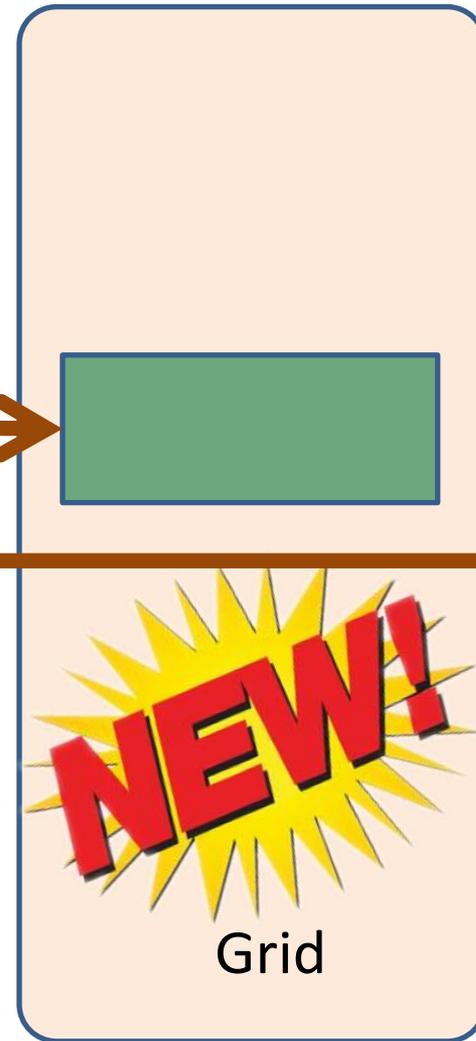
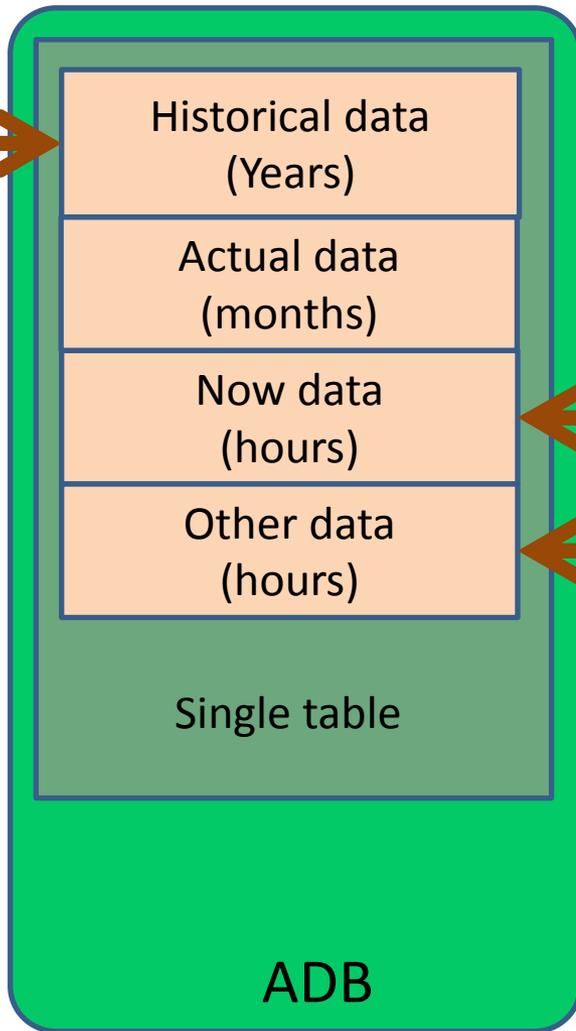
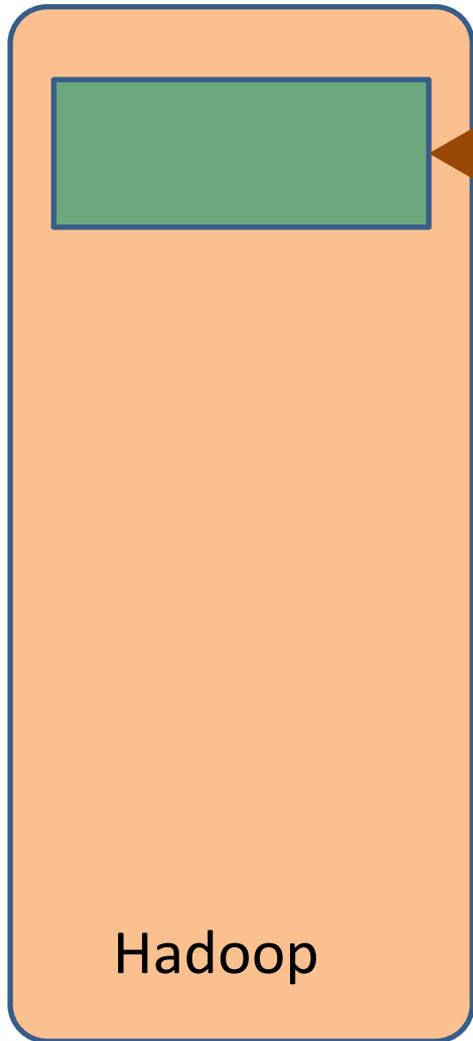
GridNode

NEW!

Grid

Zeppelin

Nifi



Zeppelin

Он всё равно не слушается?



Обратитесь к профессионалам:

- Свой дистрибутив на основе GPDB 5.0
 - Мониторинг
 - Автоматизация установки
 - Автоматизация администрирования
 - Оперативное устранение багов GPDB, критичных именно для вас
- Большая практическая экспертиза GPDB
- Архитектурный консалтинг
- Поддержка
- Мы в России :)

pds@arenadata.io

ARENADATA

ADB 5.0 – состав и развитие дистрибутива

Monitoring

Hardware metrics

Database metrics

Data metrics

Alerting

Sprint #1

Sprint #2

Sprint #3

Installation

Requirements check

Configuration planning

Capacity calculation

GPDB 5.0



Performance

Zstd compression

Configuration

Resource management

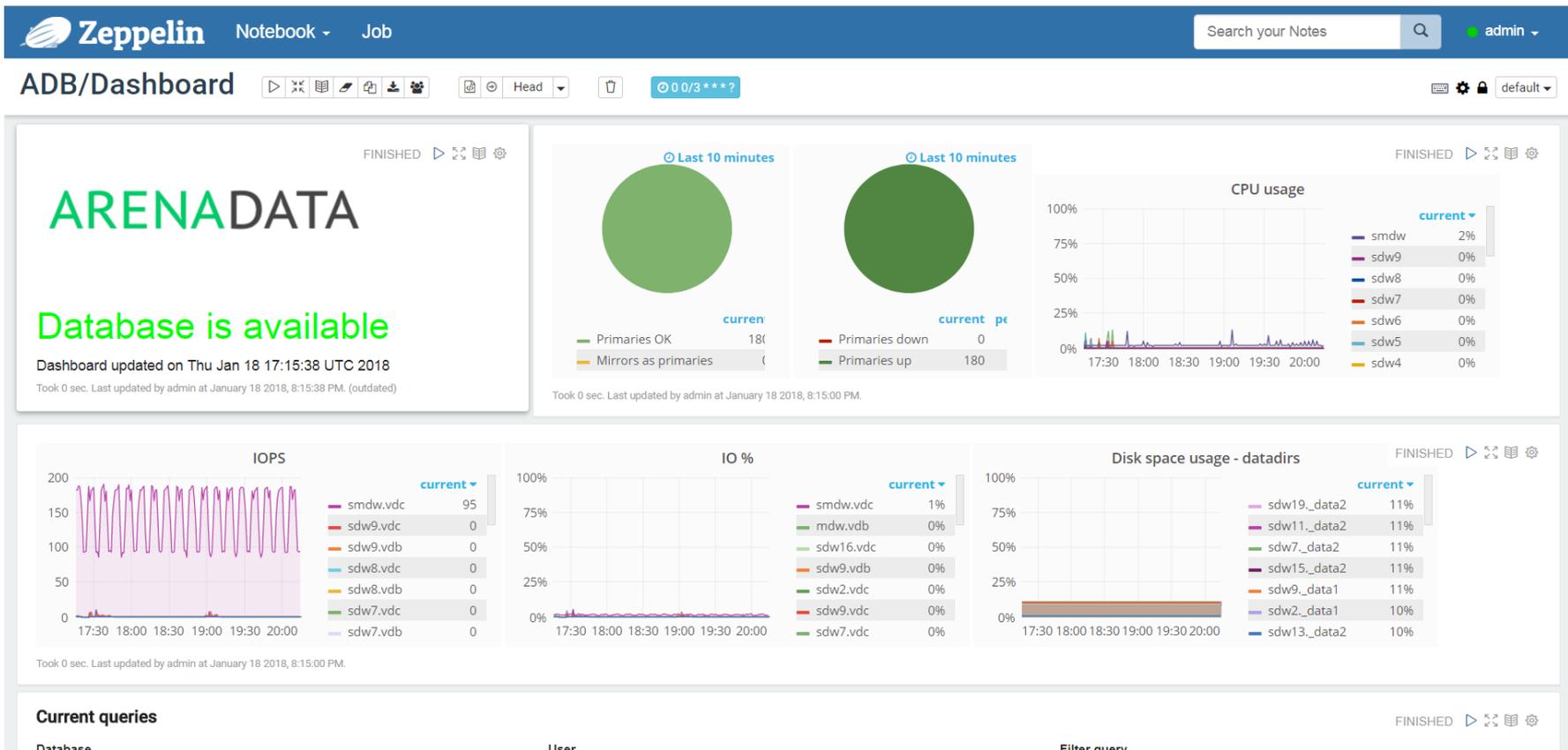
System configuration

Security configuration

Hadoop integration

Cluster expand

Мониторинг и управление из коробки



- Graphite
- Grafana
- Zeppelin

Мы умеем выгуливать Greenplum в облаке



- Совместный проект с провайдером облачных услуг Digital Energy
- Различные варианты сотрудничества
- Возможность работы с очень мощными VM (до 40 ядер, до 256 Гб памяти)

Questions?



Dmitriy Pavlov
d@arenadata.io
2017

ARENADATA