

Единый mapreduce

Или о нашем родстве с
Сусаниным

Зачем?

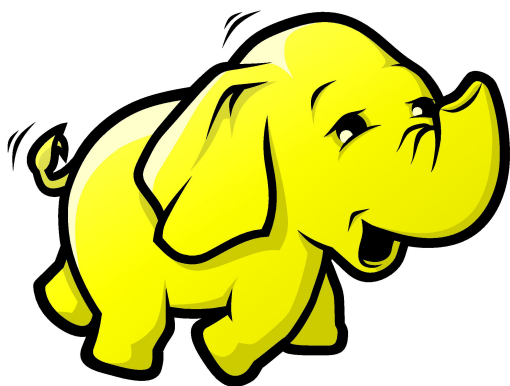
- Во всем мире всего несколько развитых систем:
 - в Google и в Microsoft свои,
 - Hadoop у почти всех остальных крупных интернет-компаний (Yahoo, Facebook, Amazon, etc).
- У нас в компании сейчас 3 системы MapReduce на кластерах из 100+ машин: YaMR, YT, Hadoop
- И мы продолжаем с разной скоростью их развивать

Почему сейчас?

- Кажется, мы и так уже с этим затащили.
- Проектов все больше и не понятно на базе чего делать новые сложные проекты.
- Дорого.



И что вы предлагаете?



Проведем публичный тендер

«Какая из этих систем станет основной системой для Яндекса в будущем (2016+)»



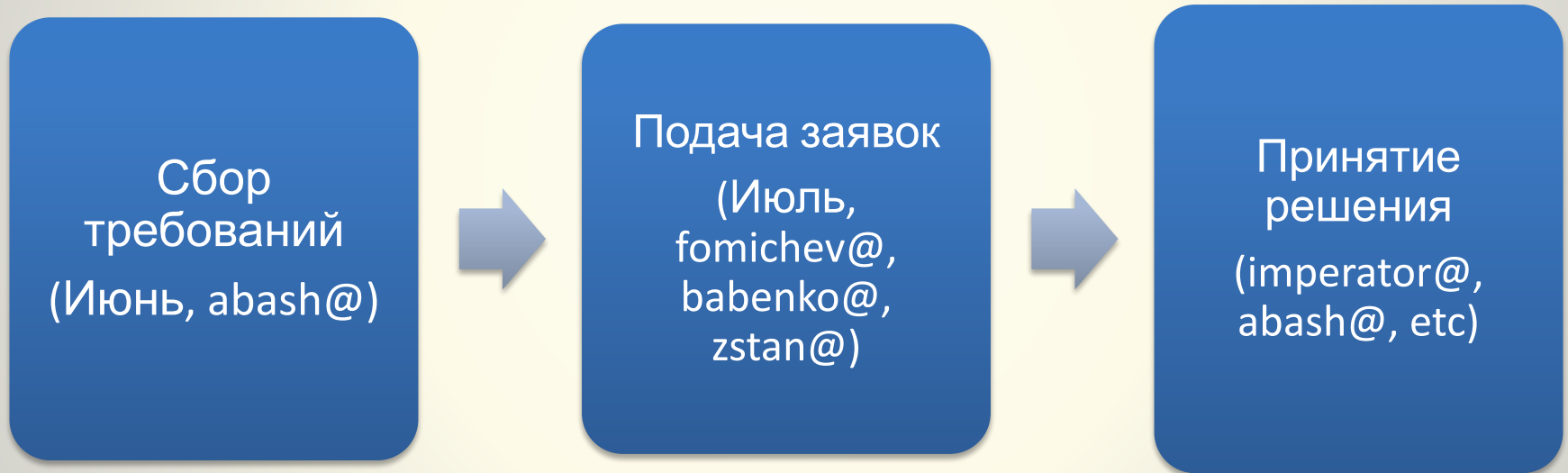
Участвуют

- YT
- YaMR/KiKiMR
- Hadoop

Выбранная технология:

- Будет рекомендована для всех новых проектов.
- В её развитие будем вкладывать основные усилия.
- Для других технологий в империи мы будем придумывать планы миграции на неё.

План



Сбор требований

- Покроют не только классический MapReduce но и прилегающие территории (random access, realtime-processing, etc).
- Будут собраны с команд («что умеем»)
- Будут доформулированы по результатам встреч с ключевыми пользователями («что реально нужно»):
 - UserData processing
 - Crawling
 - Index generation
 - Direct
 - Statistics
 - ...
 - <https://wiki.yandex-team.ru/mrtender> - записывайтесь на поговорить! 😊
- По завершению будут опубликованы на всеобщее обозрение и дообсуждены.

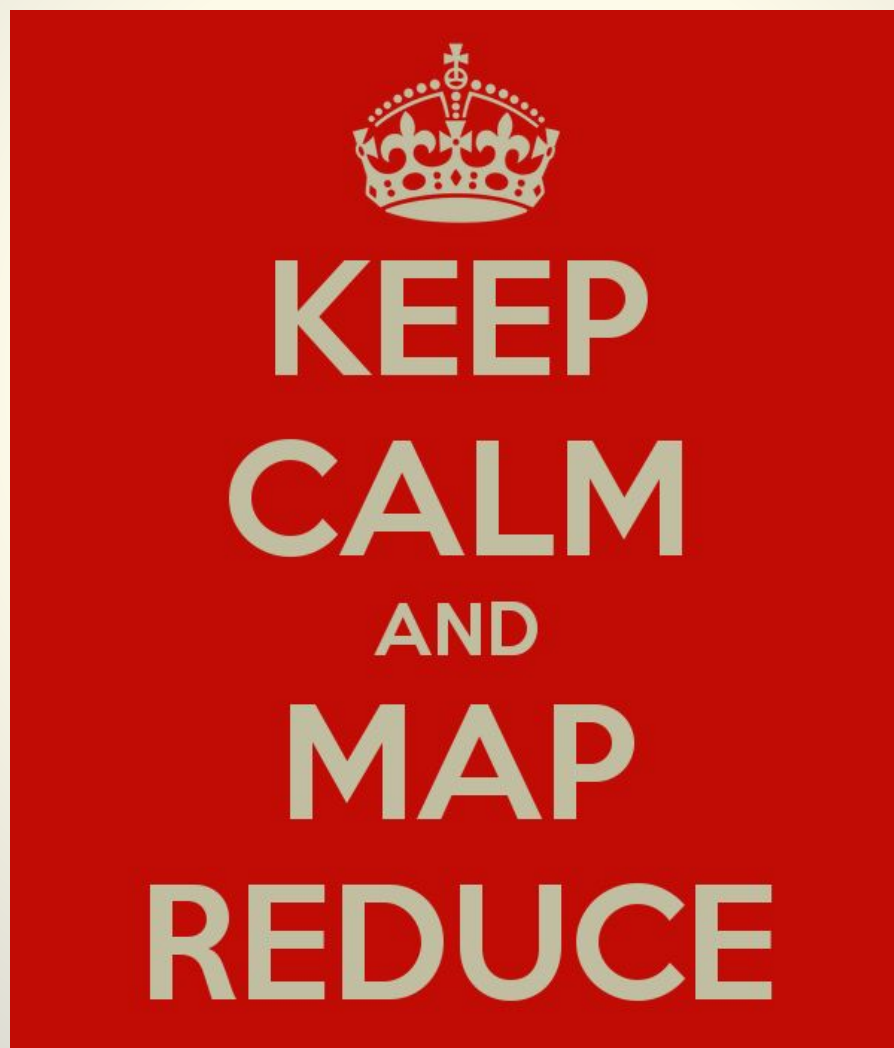
Подача заявок

- Все участники должны рассказать:
 - Как они могут удовлетворить всех пользователей (местами поговорить о high-level design'e)
 - Предоставить план по доработкам своих систем под их нужды
 - Рассказать про уникальные возможности их систем
 - ...
 - (полный список чего нужно – в июне)

Выбор победителя

- Будет собрана комиссия из ключевых пользователей. В нее войдут abash@, imperator@.
- Состав будет опубликован
- Людей сильно лоббирующих ту или иную технологию будут удалены из комиссии
- Все команды будут заслушаны, предложения изучены, вопросы заданы, etc
- По итогам комиссия выберет итоговое решение примерно в конце лета

Что это меняет?



Для пользователей любого MR

- Проекты на Q2/Q3 – не меняются.
- Если можно подождать до конца лета – подождите.
- До конца лета не начинайте долгих проектов на одной из технологий и не мигрируйте никуда (а то может случиться, что делать придется 2 раза).
- Если у вас есть уникальные сценарии – запишитесь тут:
<https://wiki.yandex-team.ru/mrtender>

FAQ

- Все куплено, победитель известен.
- Вы напишите такие требования, что выиграет ХХХ.
- Был у нас уже один «тендер» пару лет назад...

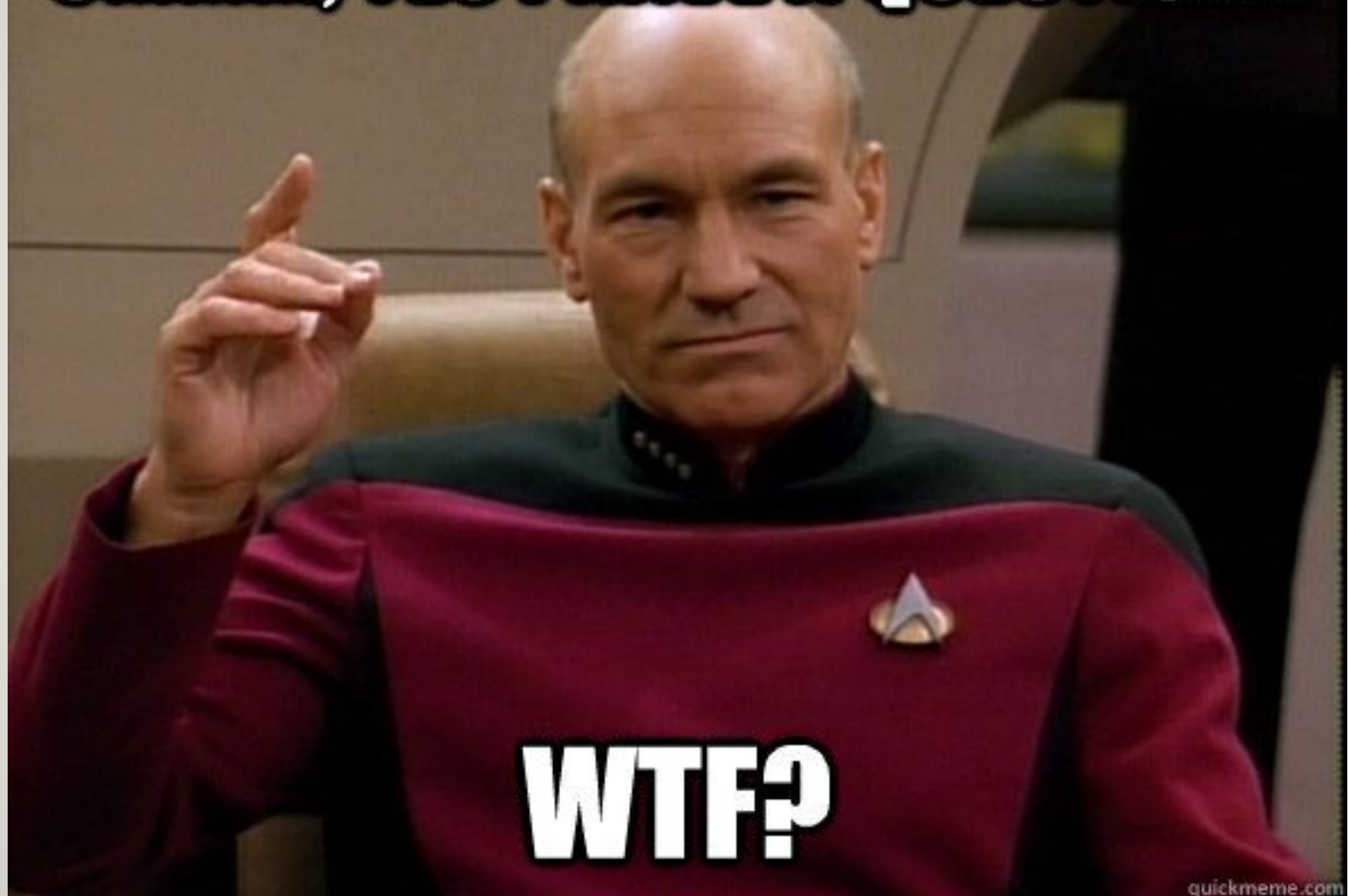
FAQ

- Все равно выиграет ХХХ.
- Почему ХХХ вообще участвует?
- Что с нашими текущими планами и нашими пользователями?
- Что с верхнеуровневым языком (aka SQL)?

FAQ

- Тех кого не выберут, уволят?
- Кластера других технологий осенью выключат?

UMMM, YES I HAVE A QUESTION...



WTF?

И это еще не все новости



KiWi будет заморожено

- Дальнейшее развитие KiWi как технологии будет заморожено, мы уходим от систем с Eventual Consistency.
- Планы по переходу и возможному сохранению совместимости появятся после проведения тендера
- Не зависимо от предыдущего пункта будут выполнены обязательства перед Oxygen/Antispam.
- В июне будет выпущен стейбл с квотированием CPU, reliable ExtCall, и прочими ожидаемыми фичами
- Поддержка и багфиксы будут осуществляться для всех продакшн-процессов (Oxygen, etc)
- Команда KiWi будет переориентирована на задачи в KiKiMR/YAMR/YQL

Если у вас есть хотелки/вопросы – chubinskiy@