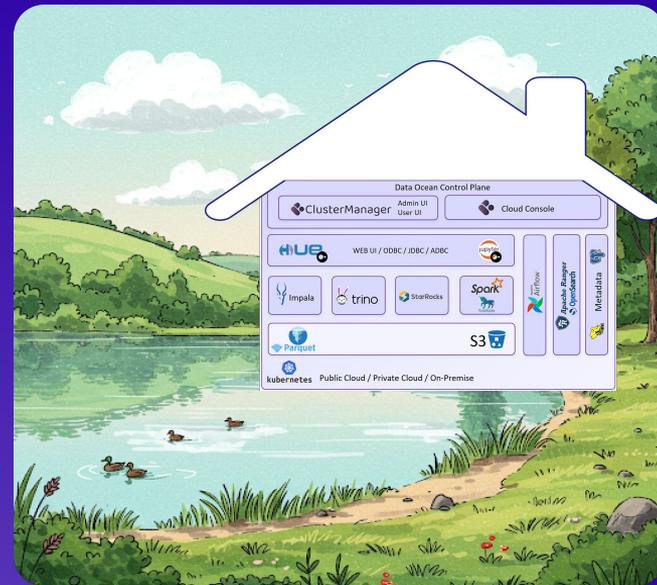




Nova

data
ocean

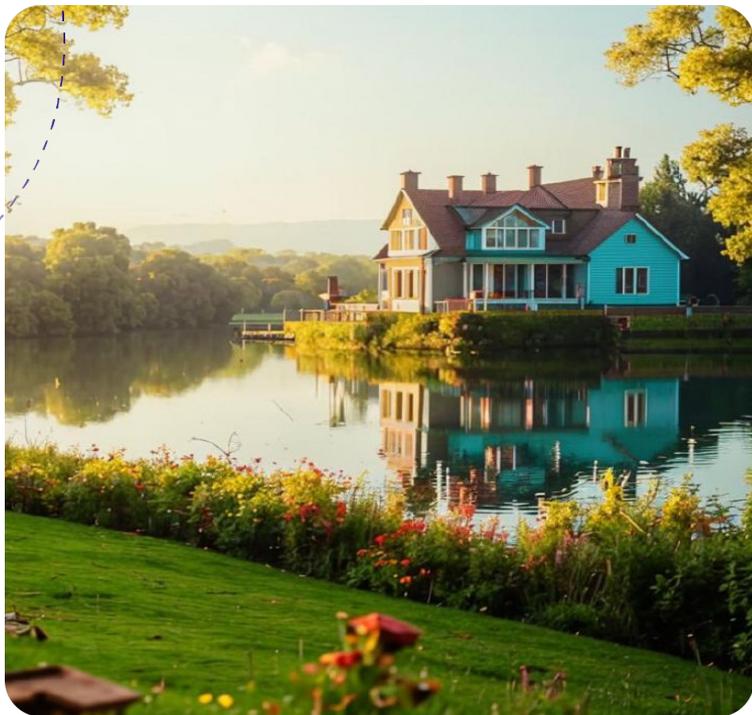
Дом у озера без DIY-доработок



Евгений Морозов,
Data Sapience

Архитектор, руководитель команды разработки
Lakehouse-платформы данных Data Ocean Nova

Даешь Data Lakehouse в каждый доклад!



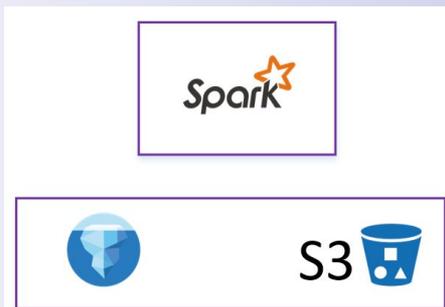
Ожидание

- Все только и говорят, что о **Lakehouse**
 - У всех он уже есть или вот-вот будет
 - Так ли это на самом деле? Давайте попробуем разобраться
- A dashed blue arrow points from the bottom right of the purple box towards the bottom right of the slide.

Строим дом у озера. С чего все начинается

У нас Spark, и он работает с Iceberg над S3. Это Lakehouse!

Seen ✓



Seen ✓

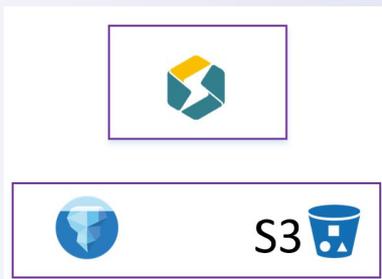
Постойте, а как же SQL-доступ к данным?
Уж не ClickHouse ли вы собрались где-то рядом
пристроить и данные там дублировать?!

Seen ✓

Строим дом у озера. С чего все начинается

Хорошо, мы возьмем StarRocks или Trino. Все задачи решены!
Это Lakehouse!

Seen ✓



Зайдите в отдел Data Science
и покажите им картинку.

Seen ✓

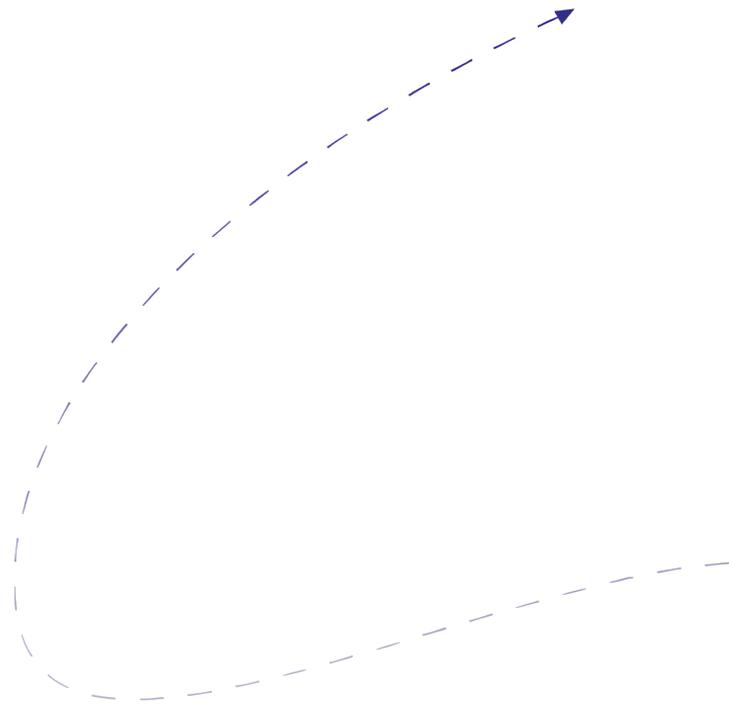


Seen ✓

Строим дом у озера. С чего все начинается

Ладно, берем StarRocks или Trino и еще Spark.
Это Lakehouse!

Seen ✓



А вы уверены, что оно все вместе будет
согласованно работать?

Seen ✓

Строим дом у озера. Стройка только началась

Вот теперь у нас есть общий метакаталог. Подружили! Это Lakehouse!

Seen ✓



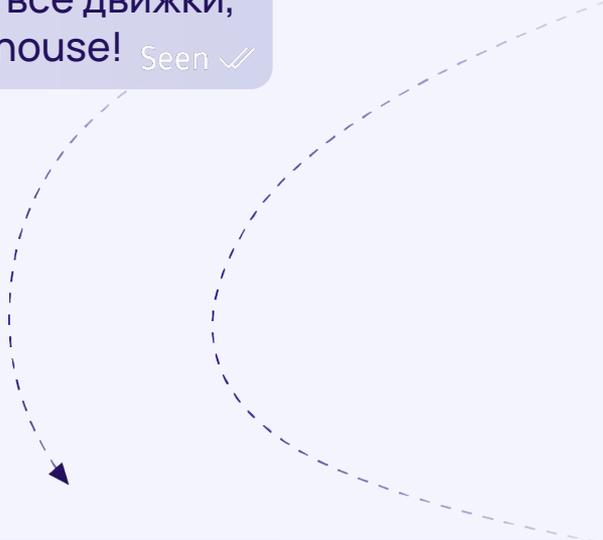
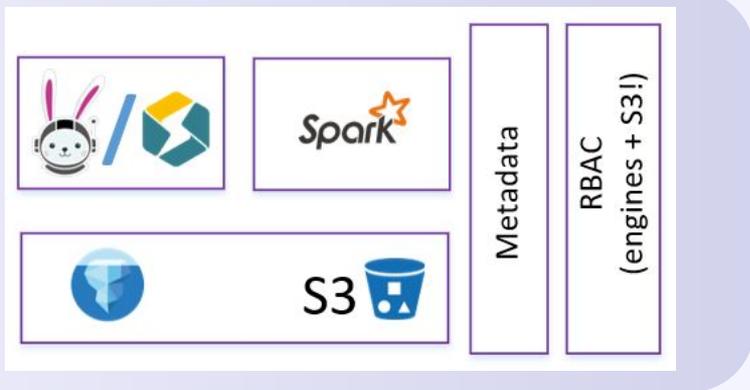
А как насчет общей ролевой модели на все компоненты?

Она ведь наверняка и на S3 может права раздавать, правда?

Seen ✓

Строим дом у озера. Стройка только началась

Получи! Раздаем ролевой доступ из одного сервиса на все движки, фреймворки и даже на S3 (и даже облачный!). Это Lakehouse! Seen ✓



Скажите, а где тут у вас журналы работы посмотреть в одном месте и формате?

Seen ✓

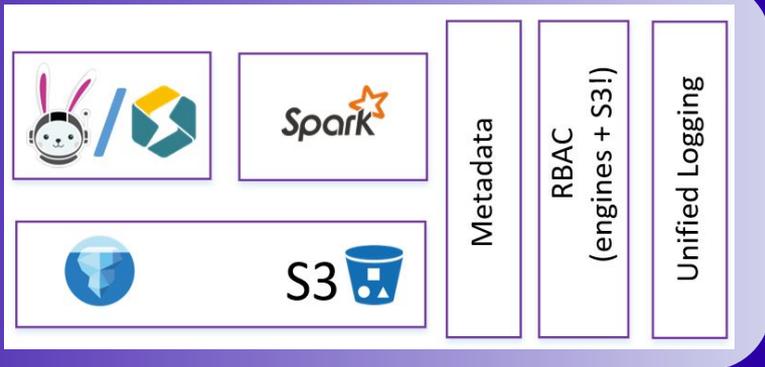


Seen ✓

Строим дом у озера. Черновые отделочные работы

Доволен? Это Lakehouse!

Seen ✓



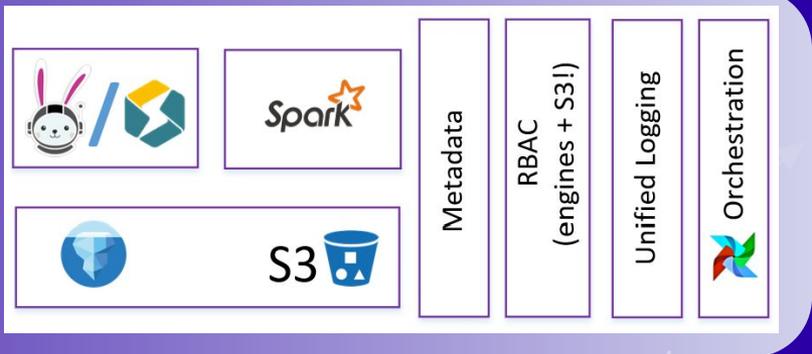
Да, но вот еще бы оркестратор сразу, чтобы без
напильника...

Seen ✓

Строим дом у озера. Черновые отделочные работы

Все? Теперь Lakehouse?

Seen ✓



Да, но хочется графический интерфейс управления и мониторинга.

Seen ✓

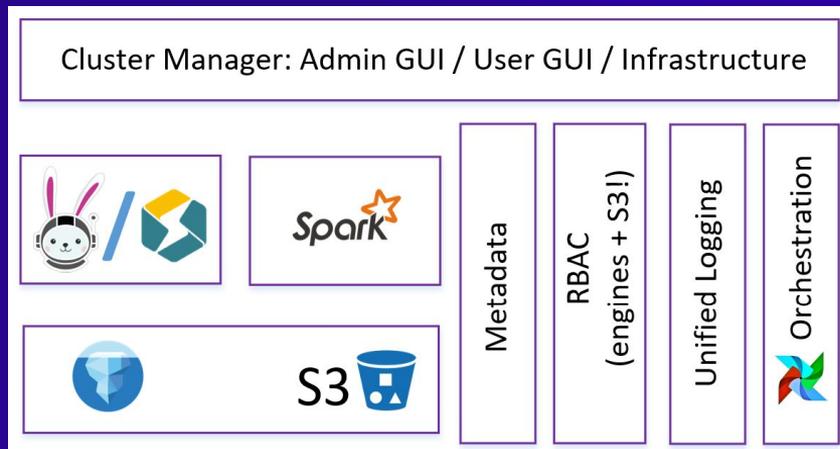
Можем на графану вывести.

Seen ✓

Все могут, а вот чтобы все сразу — в одном приложении...

Seen ✓

Строим дом у озера. Чистовая отделка



Все? Теперь Lakehouse?

Seen ✓

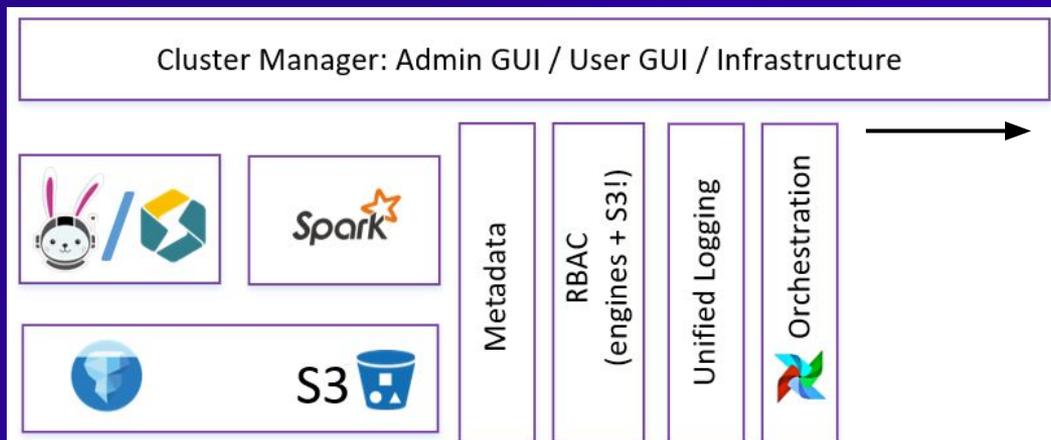
Да, но крыша маловата.

Seen ✓

??

Seen ✓

Строим дом у озера. Чистовая отделка



У нас деградирует чтение/запись S3? И таблица сломалась. 200 млн файлов в S3. Попробовали сами руками написать лекарство на Spark, но он просит 2 Tb+ RAM и работает 6 часов над одним объектом. Помогите!

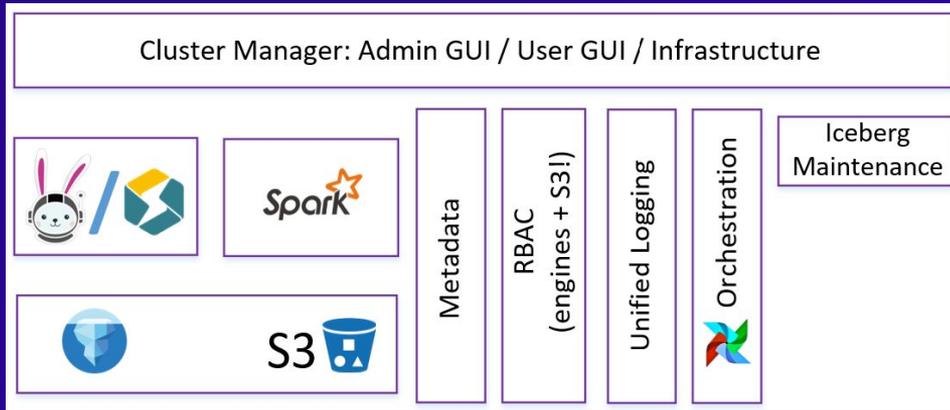


???

Seen ✓

Seen ✓

Строим дом у озера. Чистовая отделка



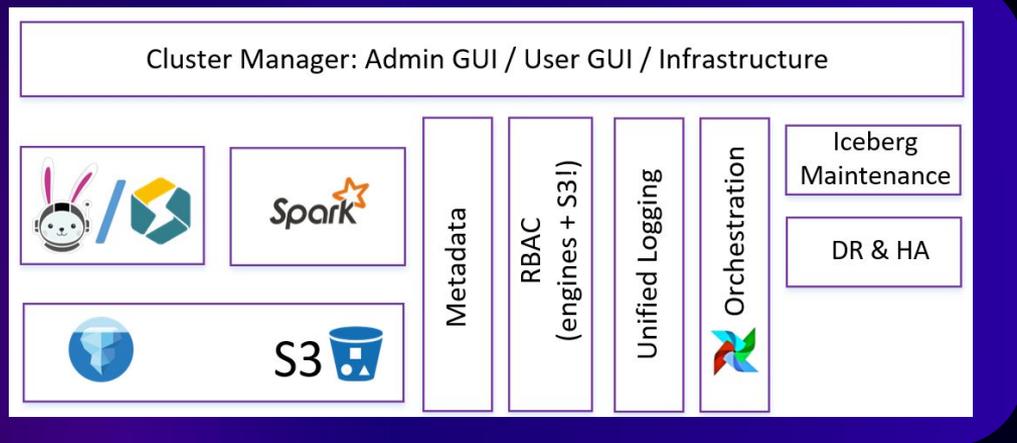
Теперь у нас есть свой уникальный сервис обслуживания. Наконец-то Lakehouse!

Seen ✓

Да, но.. Нам бы HA и DR, да еще на разные ЦОДы, да под наши RTO/RPO.

Seen ✓

Строим дом у озера. Обустраиваем

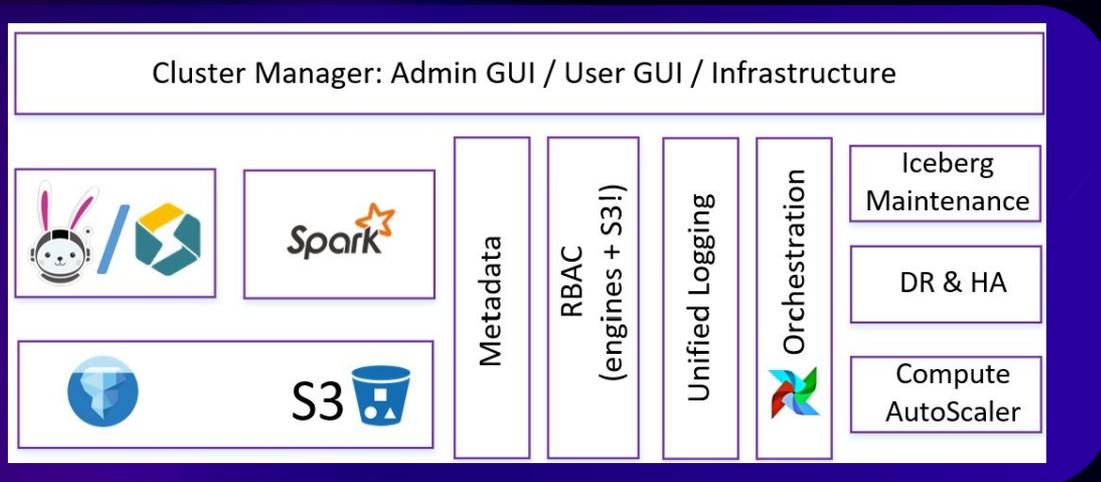


Довольны?
Seen ✓

Да, но вот бы еще нам сервис автоматического масштабирования и перераспределения ресурсов кластера.

Seen ✓

Строим дом у озера. Обустраиваем



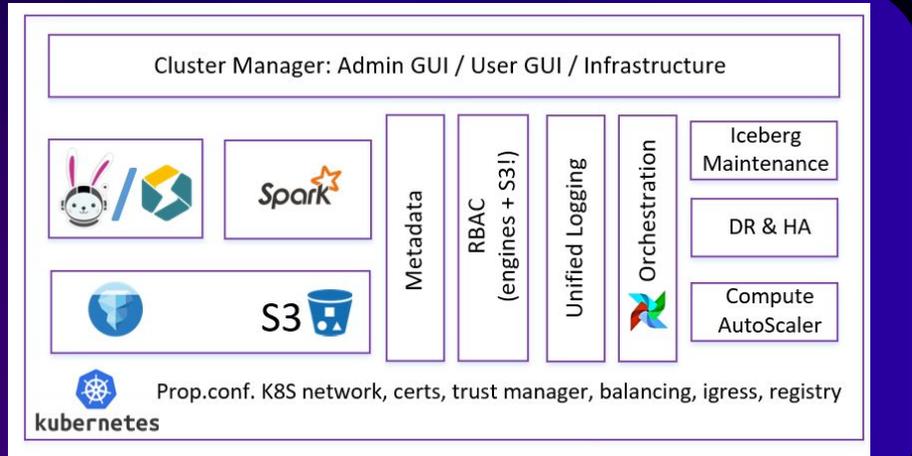
Lakehouse «под ключ», так?

Seen ✓

А кубер кто поставит и правильно настроит? А S3 вы тоже ставите, с поддержкой? Ведь ставите..., да?

Seen ✓

Строим дом у озера. Обустраиваем



То, что надо! Вы же проверяли performance на 500Tb в S3 с сотнями тысяч запросов в сутки и 500 работающими сессиями?

Seen ✓

Проверяли, ну... почти.



Seen ✓

Строим дом у озера. Регистрируем в Росреестре

Отлично! Теперь только осталось сходить в департамент информационной безопасности нашего банка, пройти проверку инфраструктуры и всего решения и можно выходить в прод...

Seen ✓



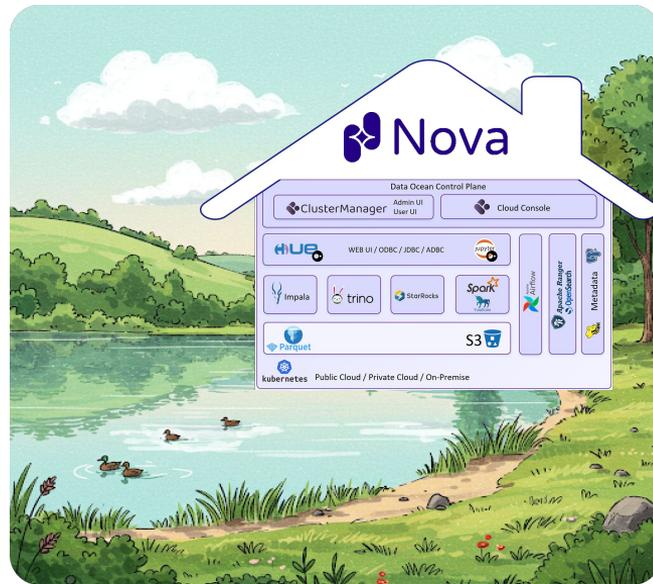
Seen ✓

⊘ Пользователь удалил свой профиль.



✓

Строим дом у озера. Почувствуйте разницу



И вот только теперь у вас есть настоящий Lakehouse



**Благодарим
за внимание**

Евгений Морозов

**data
ocean**

