

Что такое GreenPlum и зачем нам единое аналитическое хранилище?

- Greenplum — MPP СУБД, мощный и гибкий инструмент для аналитической обработки больших объемов данных
- Зачем мы построили новое аналитическое хранилище:
 - Сбёрмаркету нужен прозрачный контролируемый и прогнозируемый сбор ключевых витрин - новое хд будет его обеспечивать
 - Нам всем нужна историчность данных, нужно иметь возможность отследить историю изменения тех или иных показателей - GP ее обеспечивает
 - Данные так же не должны дублироваться - источник правды в отчетах всегда должен быть только один - единый детальный слой для всех сущностей должен это обеспечить
 - На следующем этапе развития хранилища мы сможем сделать проверки качества данных - только корректно построенное и правильно организованное хд сможет сделать эти проверки эффективными
 - И в итоге придем к долгожданному управлению на основе данных 😊

Схема DWH TOBE:

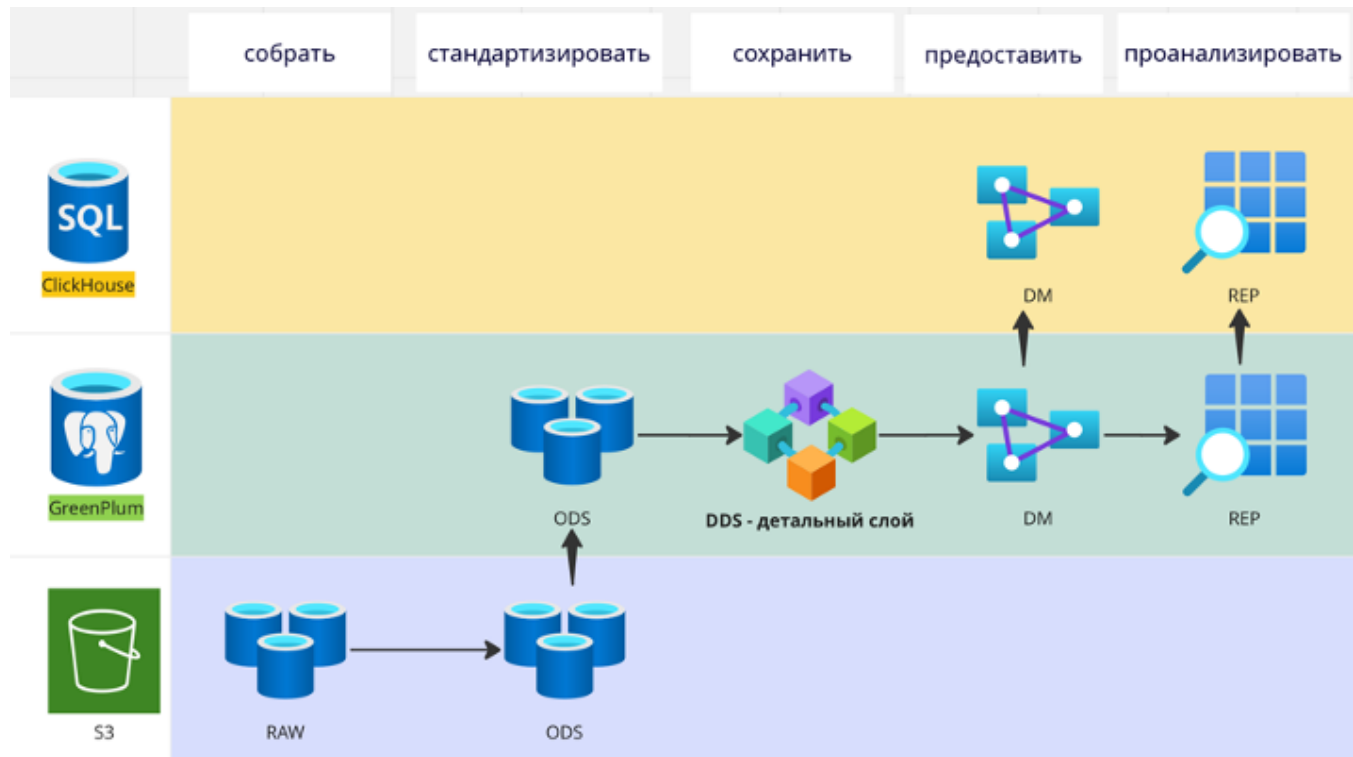


Схема слоев внутри GP:

