Вчера SAP за 10 млн $, сегодня Java микросервисы за 10 млн. руб.

Алексей Топчий

X5 Tech

@topchy_alexey

JPoint X5 Tech

СОДЕРЖАНИЕ

JPoint X5 Tech

1 → 2 → 3 →

Проблема

Вендор лок
для важного сервиса,
обслуживать дорого,
разработаем своё

Задача

- ✓ Вынести алгоритм ценообразования в микросервисы
- ✓ Собственная разработка/ поддержка
- ✓ Масштабируемость/ отказоустойчивость/ изоляция

Что делали

- ✓ Проектирование - распилить монолит на микросервисы, поправки на BigData
- ✓ Поиск оптимальной архитектуры, состава сервисов и выбор БД
- ✓ Выбор способа и критериев параллелизации, проблемы параллельной обработки больших данных
- ✓ Оптимизация кода сервисов
- ✓ Раскатка, масштабирование, баланс между «оптимизировать»/ «уместить в ресурсы VM» / «залить железом»
- ✓ Оптимизация доступа к БД и её структуры

СОДЕРЖАНИЕ

JPoint X5 Tech

4

Результат

- ✓ Как мы раскатываем пилот
- ✓ Итоговая работающая архитектура
- ✓ Оставшиеся вопросы

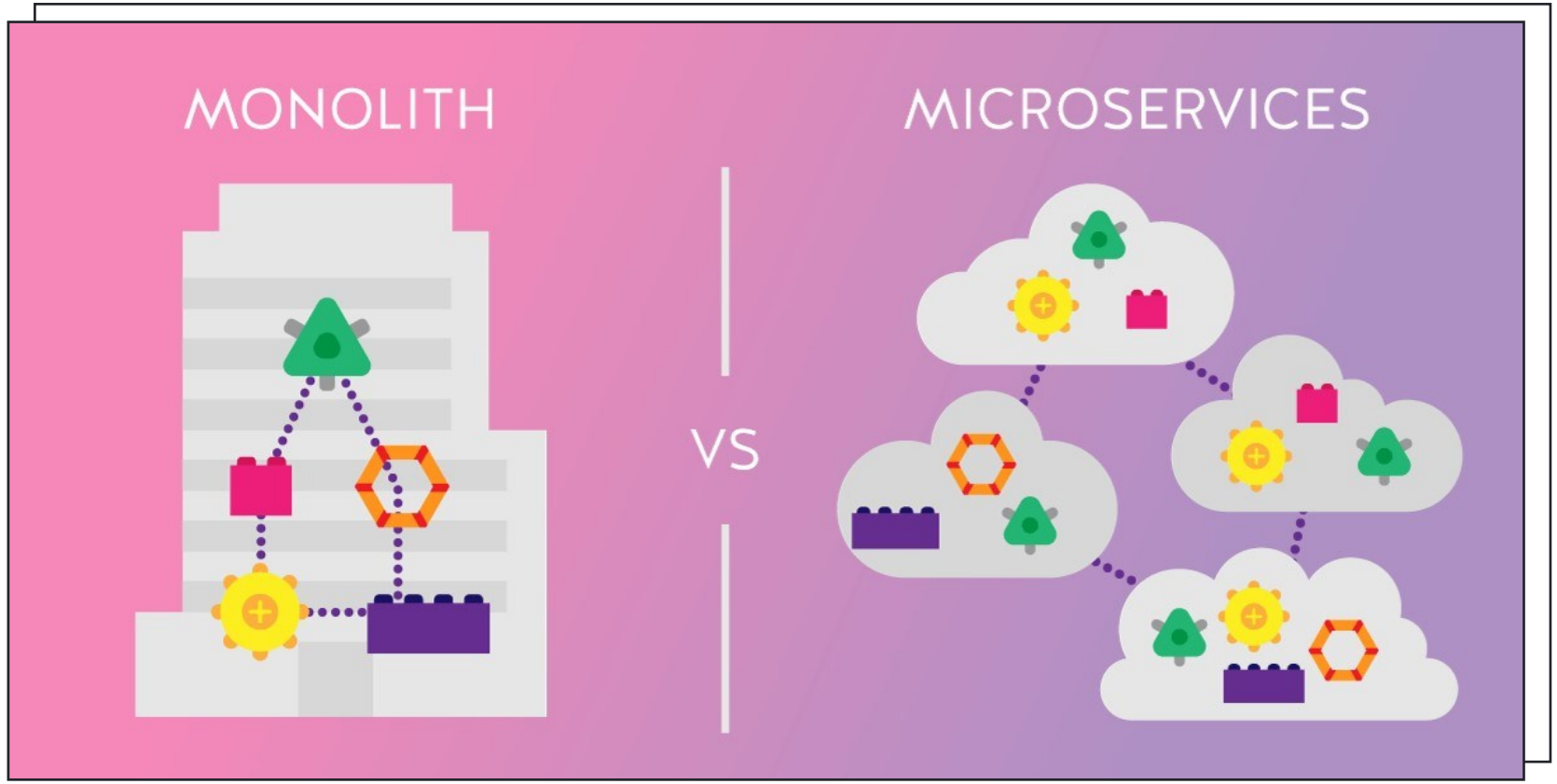
ПРИЧИНЫ

Зачем было менять работающее решение?

ОЖИДАНИЯ

JPoint X5 Tech

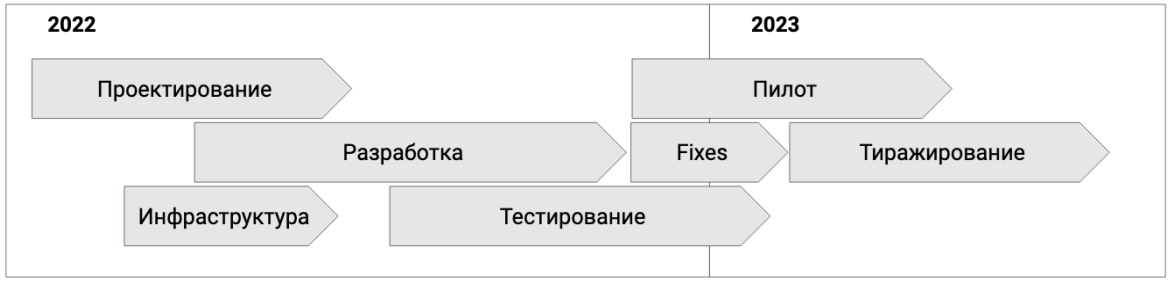
Какие преимущества мы рассчитывали получить



КОМАНДА

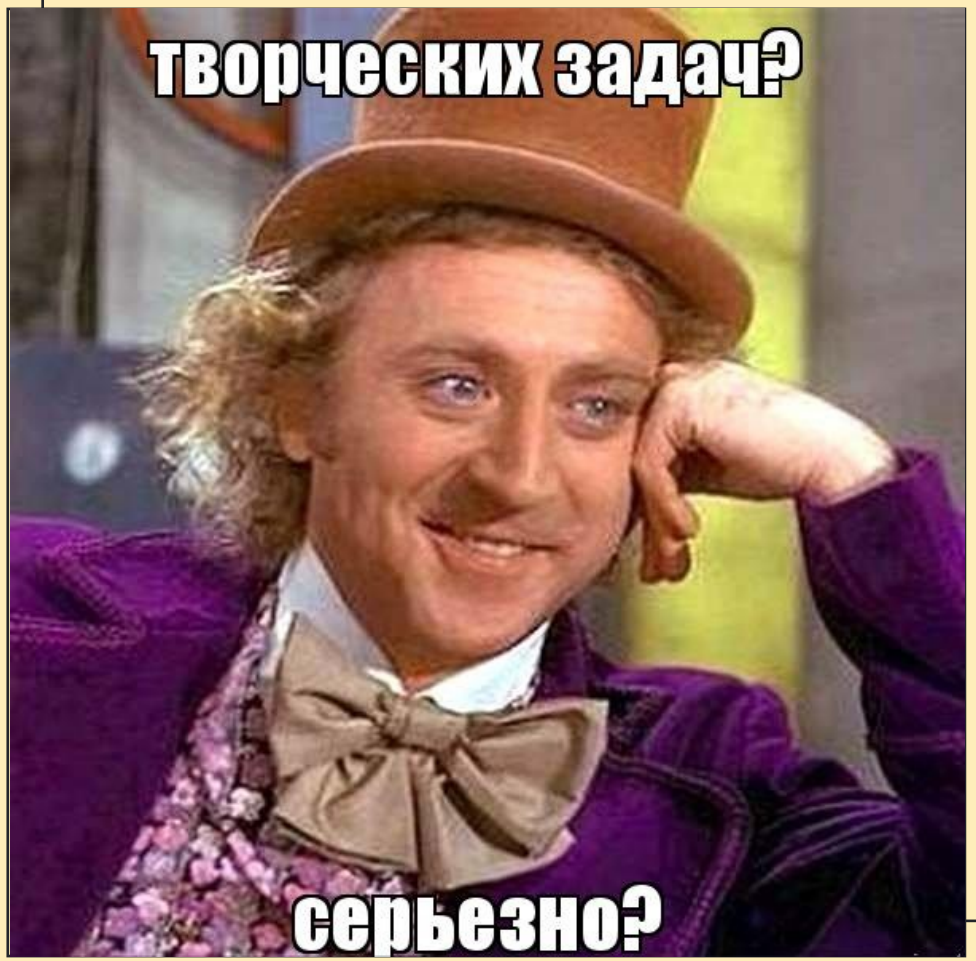
У нас было

- Примерное понимание алгоритма
- Кроссфункциональная команда
- Много энтузиазма
- Бюджет



МОТИВАЦИЯ

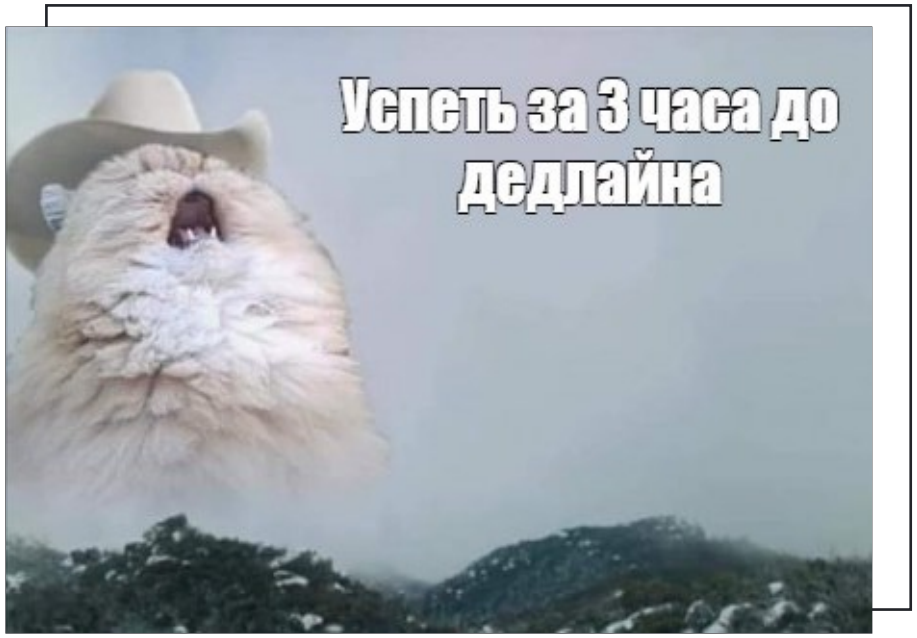
Зачем это бизнесу – понятно,
но зачем это разработчикам?



3 ЭТАПА ЗА 3 ЧАСА

- ✓ Загрузка данных
- ✓ Расчёт
- ✓ Верификация и отправка

Загрузка данных → Расчёт → Сверка и отправка

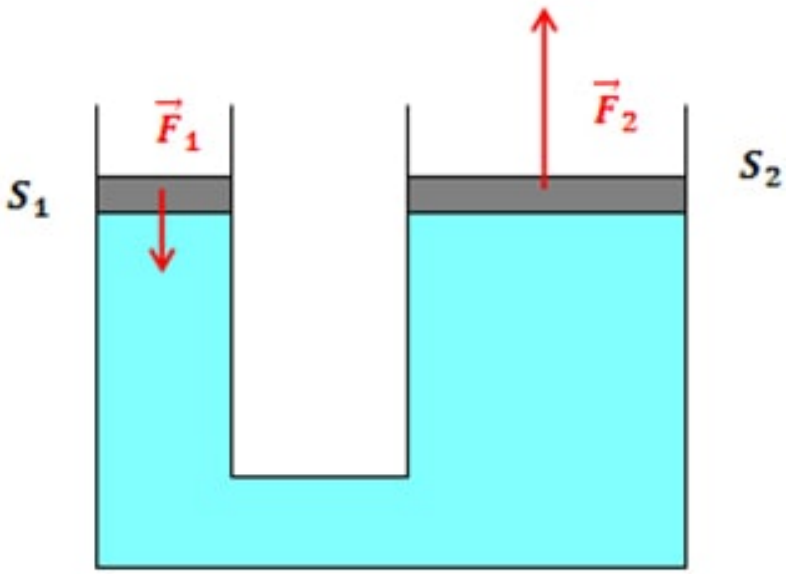


Загрузка **данных**

JPoint X5 Tech

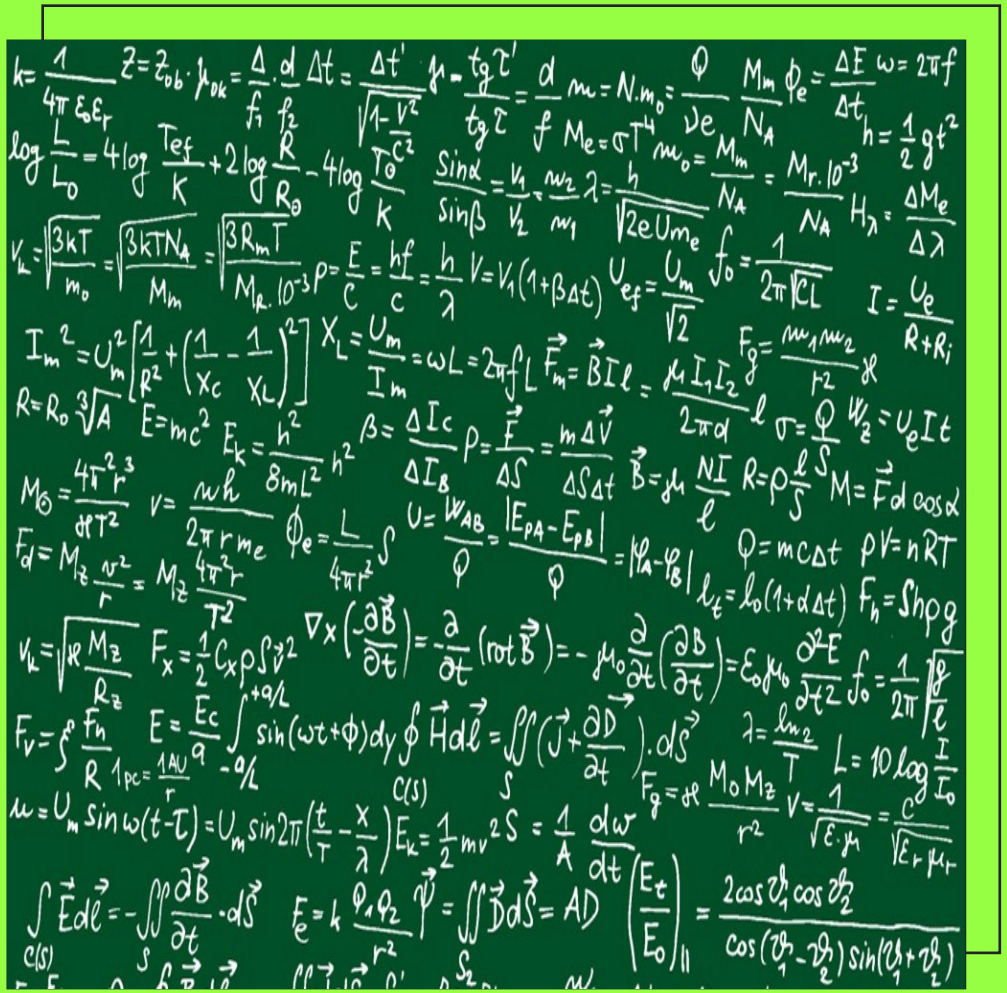
Структура, количество и формат данных

SZ12	D1	56.98	500	KG	GREEN
SZ25	D1	128.00	25	ST	GREEN
DS25	D2	41.23	120	ST	YELLOW



Посчитать
200 млн. цен

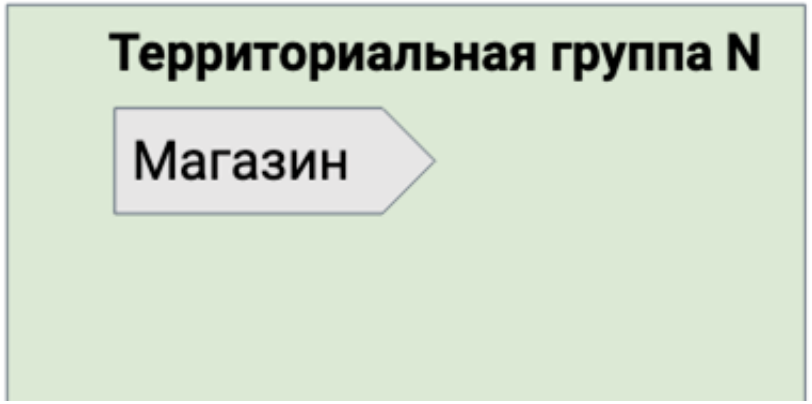
- 20 тыс. магазинов
- Почему цен так много?



Проверить, **отправить в магазины**

JPoint X5 Tech

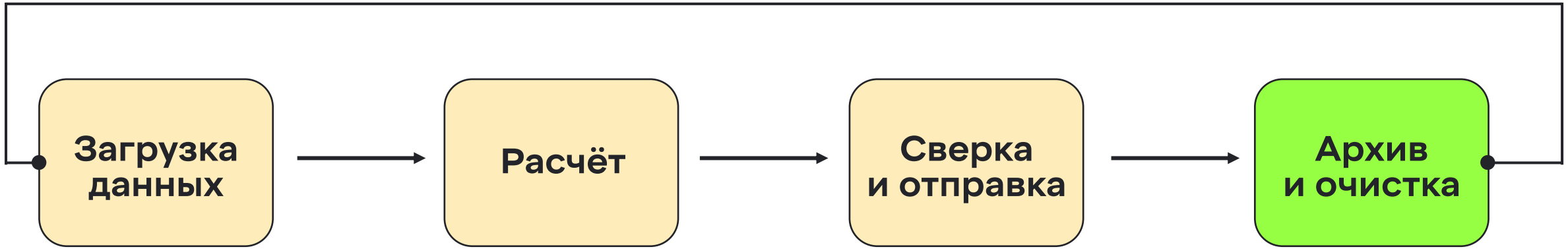
- 500 тыс. пакетов по 300 PLU
- Магазины делятся на 400 территориальных групп, в каждой от 1 до 1000+ магазинов



Данные обновляются ежедневно

JPoint X5 Tech

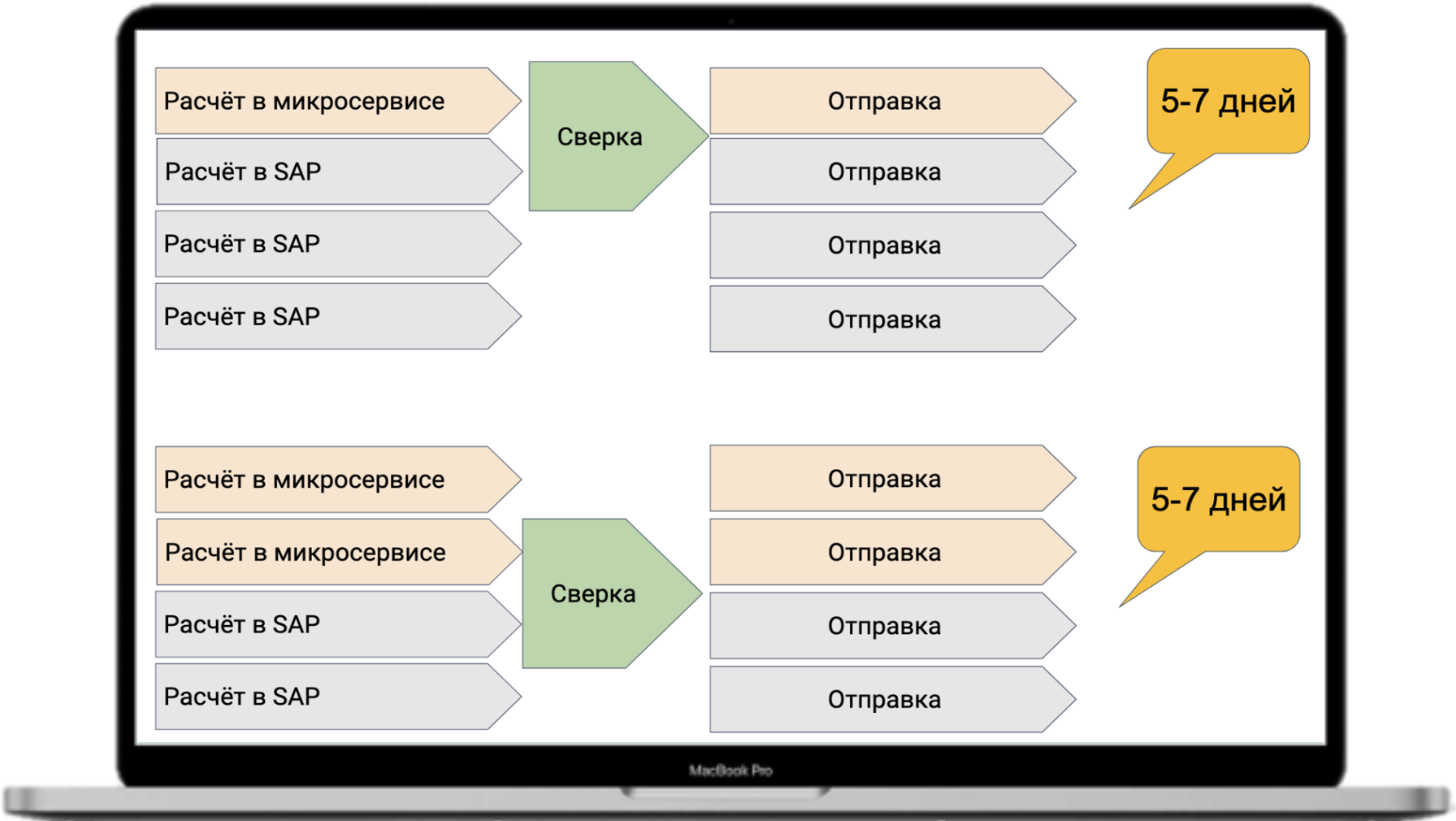
Зачем?



Пилот и тиражирование

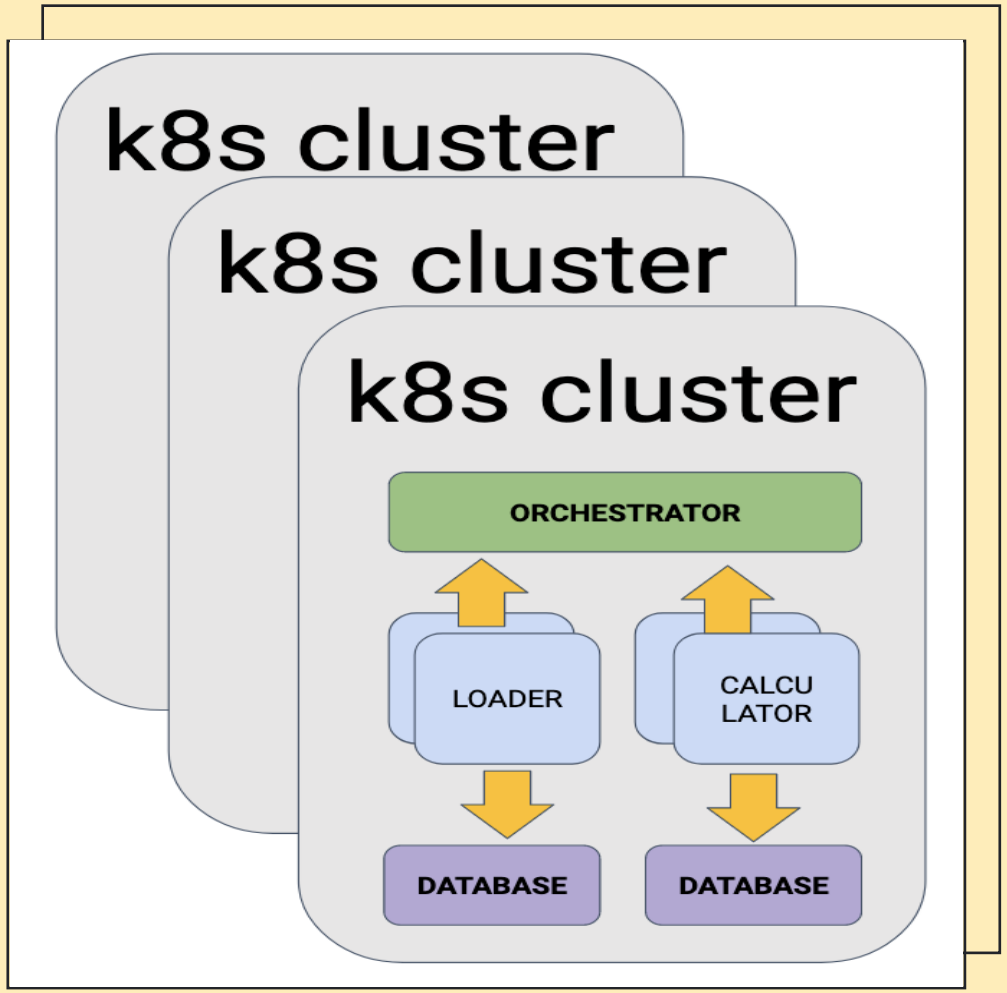
JPoint X5 Tech

Есть ограничения



РЕШЕНИЕ

- Микросервисы
- Изолированные БД
- Распределённая архитектура
- Отказоустойчивость



Перенос **алгоритма**

JPoint X5 Tech

Зачем воспроизводить точную копию алгоритма? Костыли тоже переносить?

Шаг 1 — 15 пунктов

Шаг 2 — 13 пунктов

Шаг 3 — 6 пунктов

Пункт 17 - Получить номера материалов

Определим переменную date_count=current_date + integer '1'

Проверяем выборку, полученную в пункте 16 на наличие распределительных цепочек по алкоголю.

```
SELECT vkorg, vtweg, pltyp FROM calc_message_scope_stage_2
WHERE calc_message_id=calc_message.id AND data_id='step_16_result'
AND (vkorg='ZALC' OR pltyp='20')
```

Вариант 1 - Если распределительные цепочки по алкоголю найдены

```
WITH var AS (SELECT current_date + integer '1' AS date_count)

SELECT DISTINCT artnr FROM wlk1
JOIN wrs1 ON wlk1.filia=wrs1.asort
CROSS JOIN var
WHERE DATE(wrs1.datab)<=date_count AND DATE(wrs1.datbi)>=date_count AND
AND  DATE(wlk1.datab)<=date_count AND DATE(wlk1.datbi)>=date_count AND
```

Записываем результат пункта в таблицу calc_message_scope_stage_2:

Атрибут	Значение
id	NEWID()
calc_message_id	calc_message.id
data_id	step_17_result
artnr	wlk1.artnr

Перенос **алгоритма**

JPoint X5 Tech

Шаг 1

100+ SQL

Шаг 2

140+ SQL

Шаг 3

80+ SQL

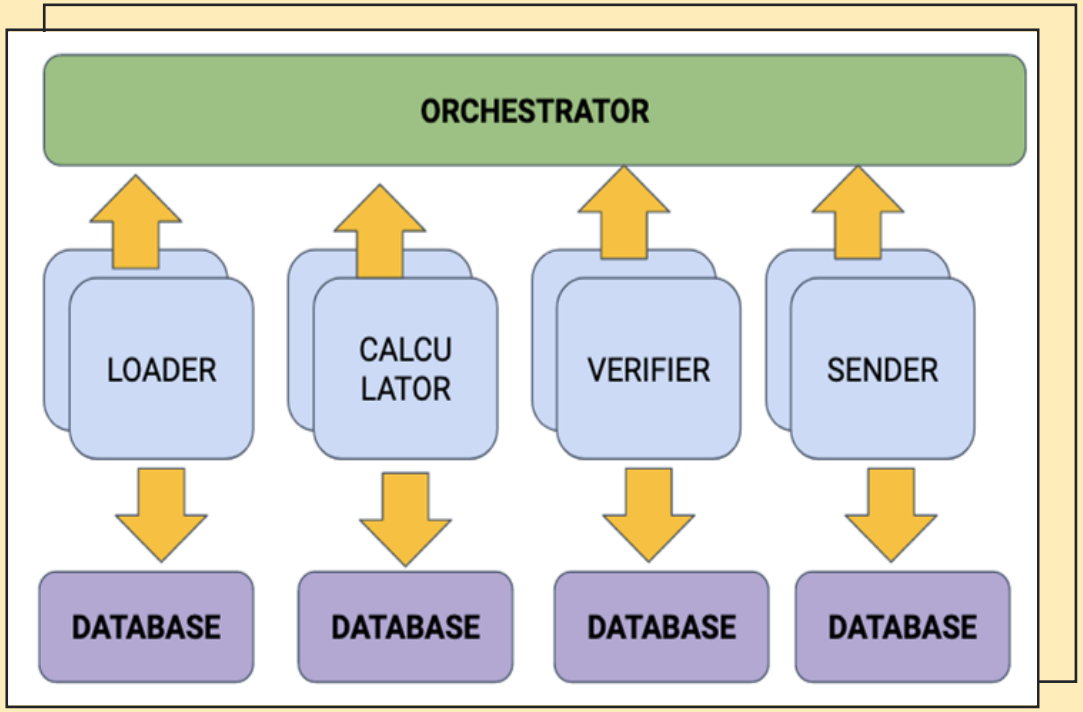
```
WITH var AS (SELECT current_date + integer '1' AS date_count)

SELECT DISTINCT artnr FROM wlk1
JOIN wrs1 ON wlk1.filia=wrs1.asort
CROSS JOIN var
WHERE DATE(wrs1.datab)<=date_count AND DATE(wrs1.datbi)>=date_count AND
AND  DATE(wlk1.datab)<=date_count AND DATE(wlk1.datbi)>=date_count AND
```

Правильная **архитектура**

- Лёгкий оркестратор
- Сервис хранилища
- Обмен через REST или KAFKA

Кластер 1



РЕАЛИЗАЦИЯ

JPoint X5 Tech

Узкое место - загрузка данных в хранилище

Сериализация/десериализация

ORM (Кеширование/мэппинг DTO)

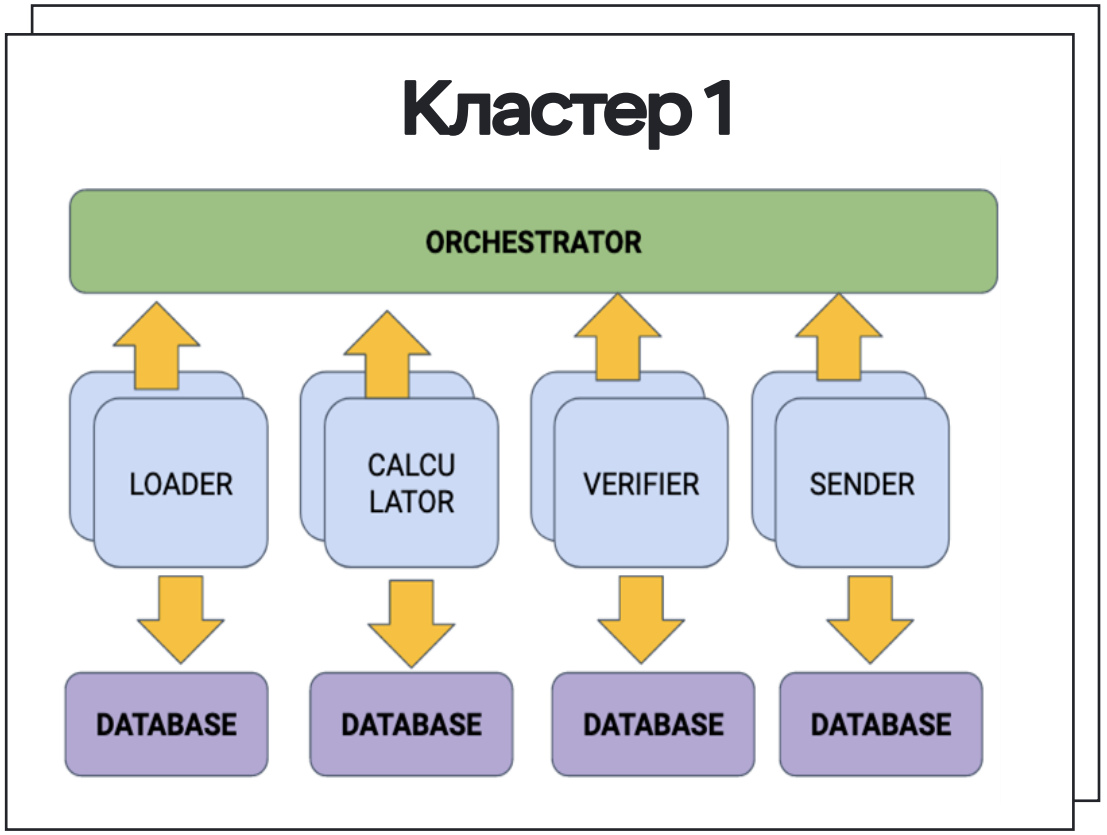
Запись в БД

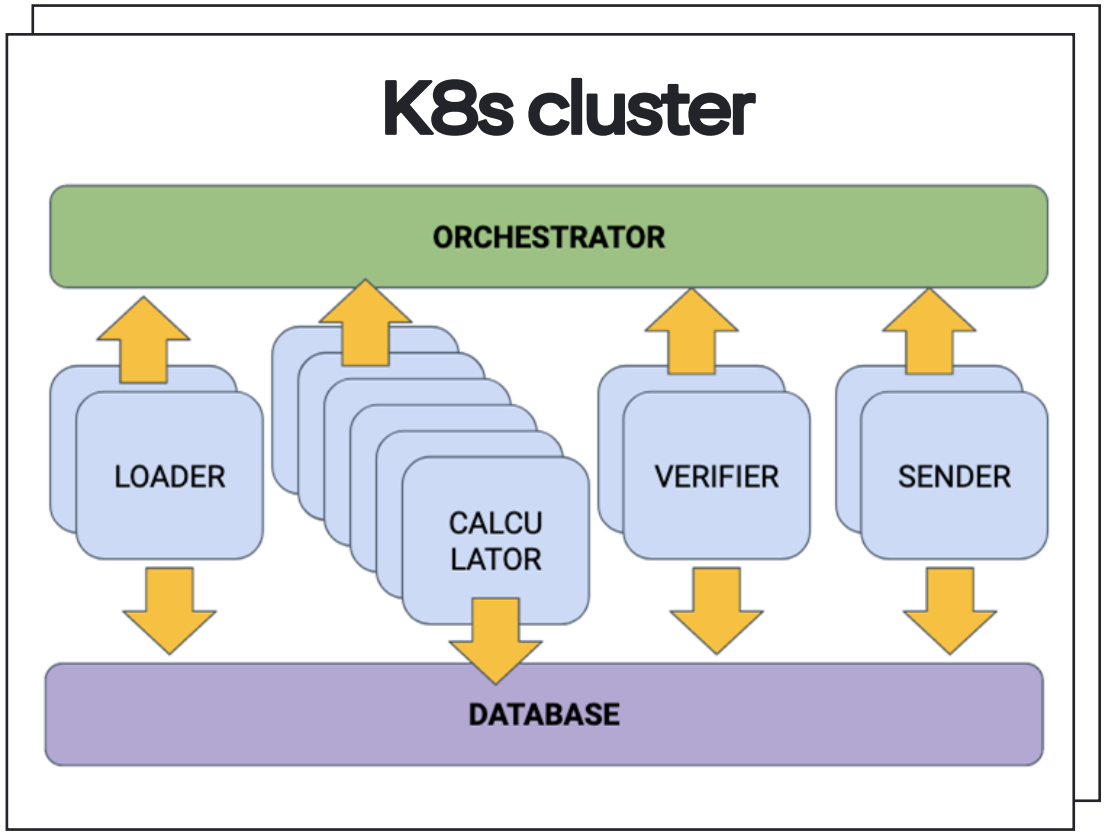
Прочее

РЕАЛИЗАЦИЯ

JPoint X5 Tech

Переосмысление паттернов

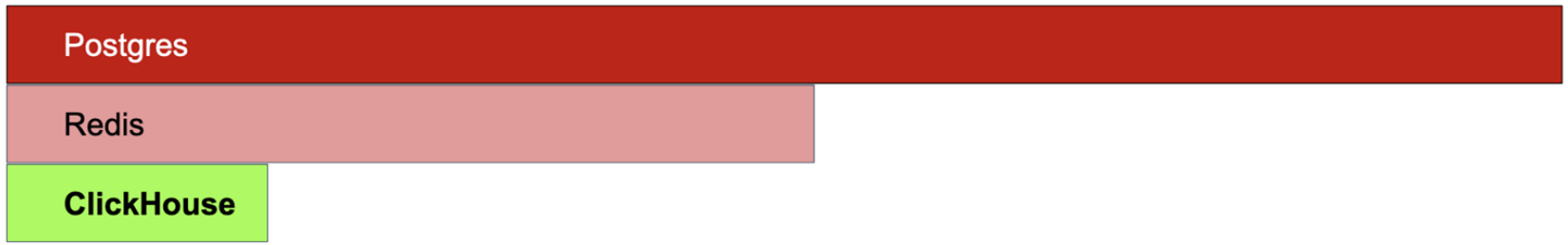




ВЫБОР БД

JPoint X5 Tech

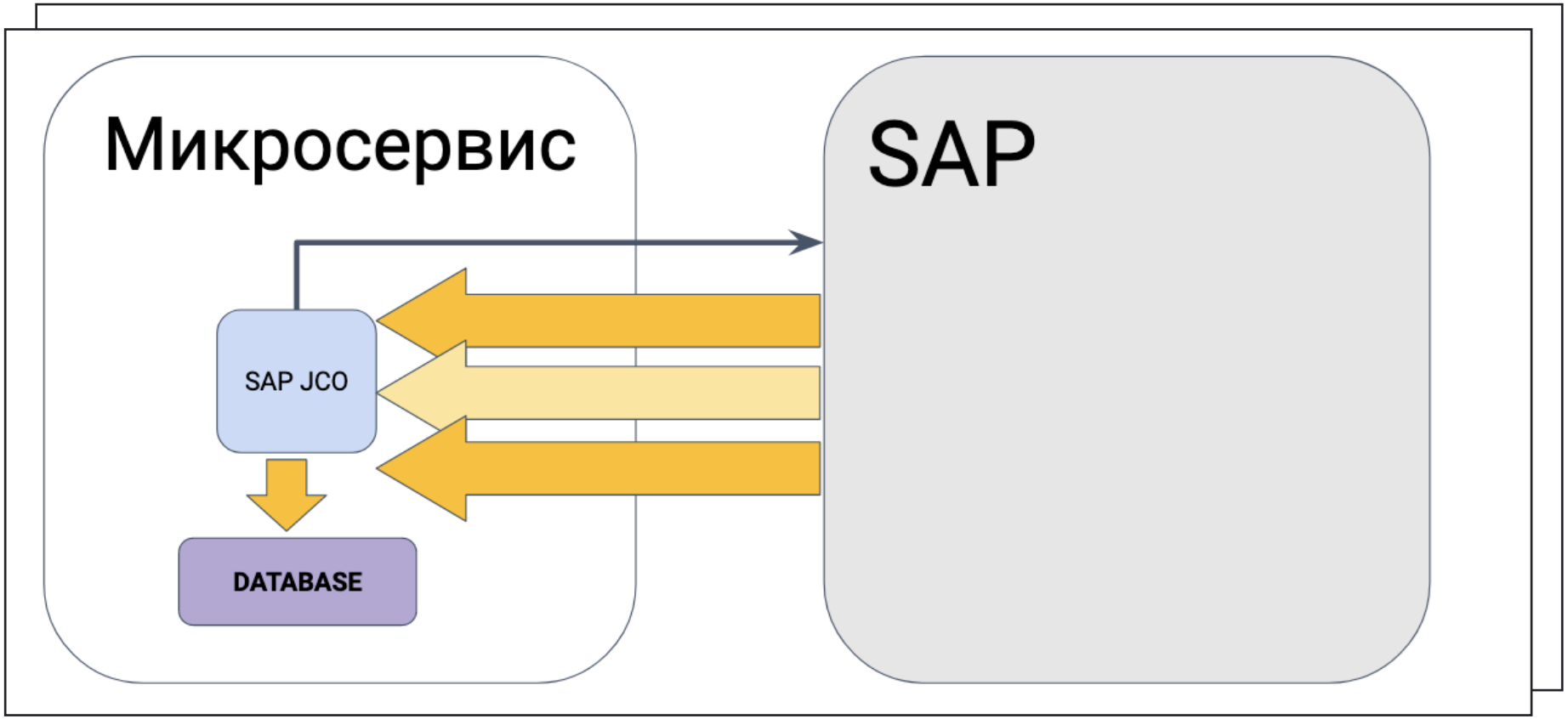
- ✓ Postgres
- ✓ ClickHouse memory engine
- ✓ Redis in memory



Первый этап: **загрузка**

JPoint X5 Tech

SAP PI vs SAP JCO



Первый этап: **загрузка**

JPoint X5 Tech

Как убедиться в том, что все данные корректно загружены?

name	`formatReadableQuantity(total_rows)`
marc	1.45 billion
prop	1.43 billion
calc_history_position	678.07 million
calc_data_assort_for_werks	560.98 million
konp	440.86 million
konh	403.59 million
calc_history_position_to_send	379.12 million
wlk1	309.79 million
calc_data_matnr_for_werks	292.99 million
a071	282.49 million

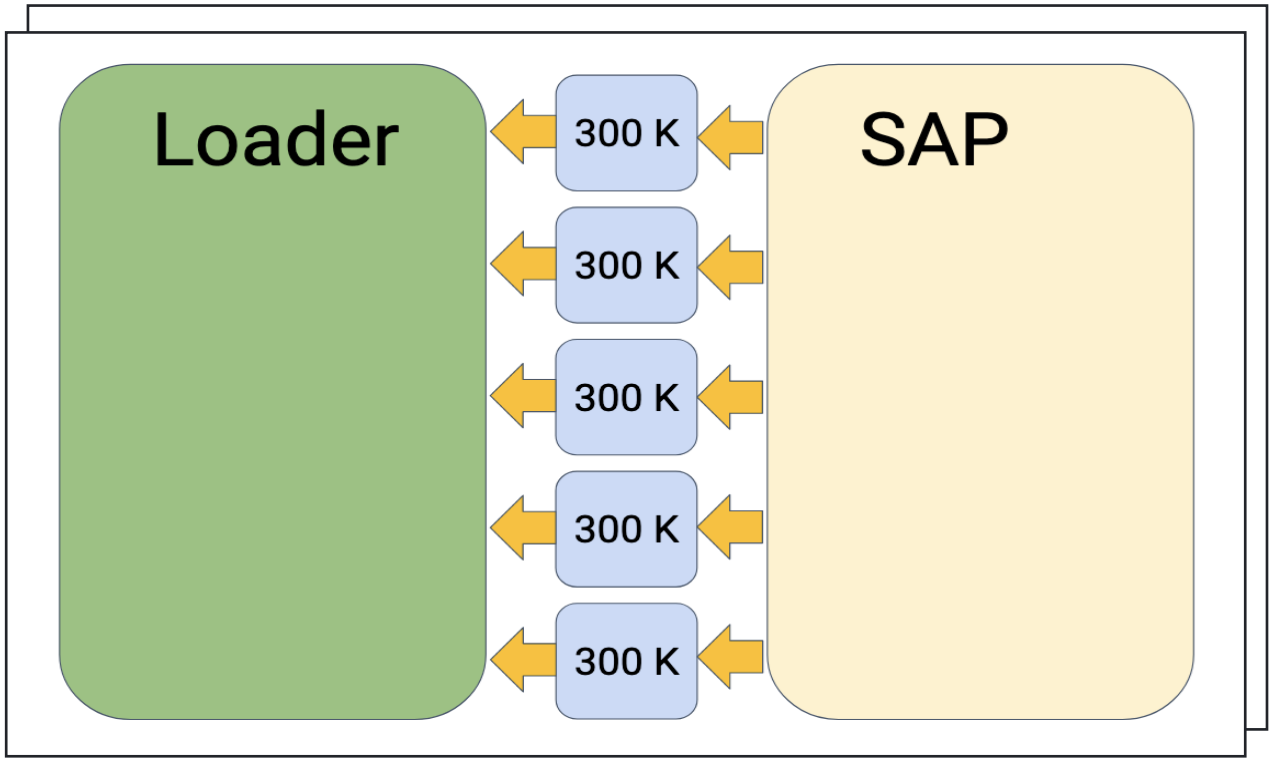
```
with
    batchCte as (
        select lower(table) table, sum(batchSize) as batchRows, sum(filteredRows) filtered,
               sum(batchSize) - sum(filteredRows) rows
        from log_batches where batchSize > 0 group by table
    ),
    sysTablesCte as ( select * from system.tables where database = currentDatabase()),
    totalRowsCte as (
        select b.table as TABLE,
               coalesce(t.total_rows, 0)  as ROWS_DB,
               b.batchRows as ROWS_BATCH,
               b.filtered as FILTERED,
               coalesce(t.total_rows, 0) + b.filtered - b.batchRows as DIFF
        from batchCte b left join sysTablesCte t on upper(t.name) = upper(b.table)
    )
select TABLE, ROWS_DB, DIFF, ROWS_BATCH as ROWS_SAP, FILTERED
from totalRowsCte
```

TABLE	ROWS_DB	DIFF	ROWS_SAP	FILTERED
a550	2294114	0	2294114	0
a521	1	0	1	0
a623	3182	0	3182	0
tvko	1118	0	1118	0
a609	74	0	74	0
wakh	2749824	0	2749824	0
marm	3204614	0	3204614	0
tvarvc	24	0	24	0
lfa1	1144693	0	1251225	106532
a540	298658	0	298658	0
a657	49	0	49	0

Первый этап: **загрузка**

JPoint X5 Tech

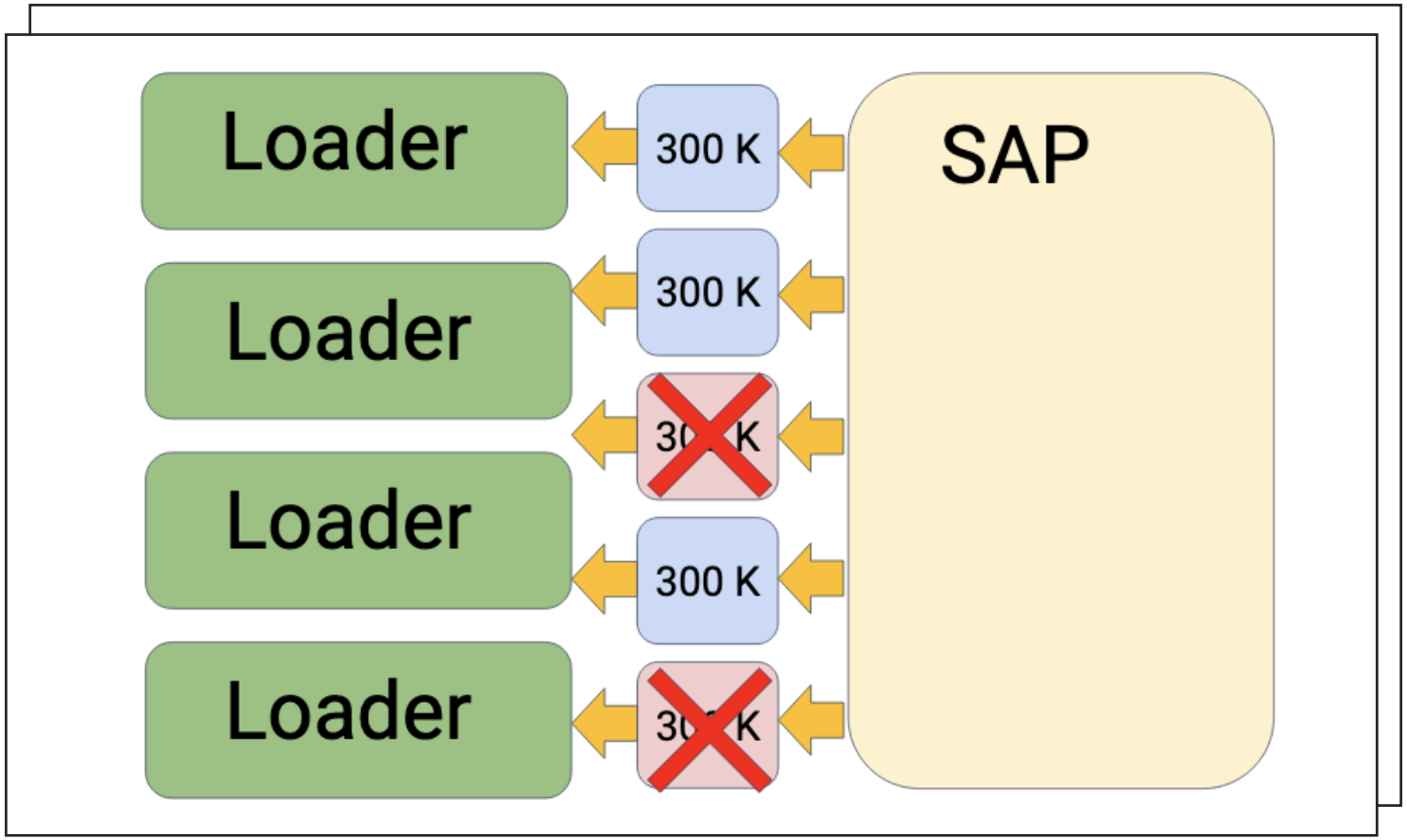
- Нужно быстрее, значительно быстрее
- Эффективное количество потоков обработки соответствует количеству процессоров?



Первый этап: **загрузка**

JPoint X5 Tech

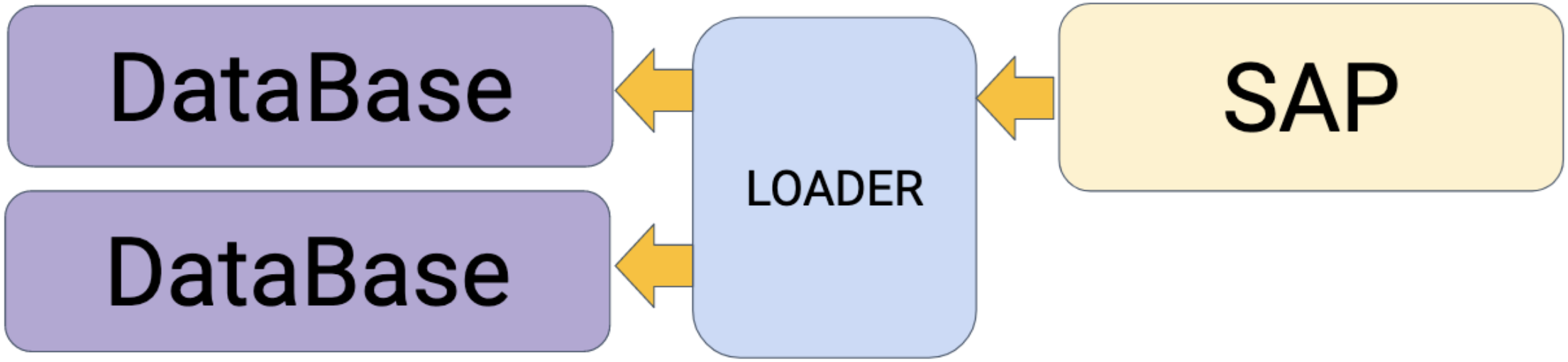
Теряются пакеты



Первый этап: **загрузка**

JPoint X5 Tech

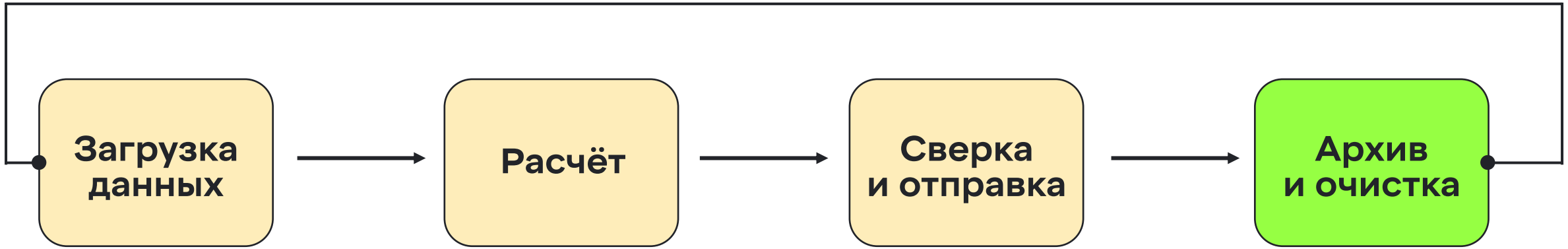
...и отказоустойчивее



Данные обновляются ежедневно

JPoint X5 Tech

Что это значит применительно к данным БД (в памяти)



Первый этап: расчёт

JPoint X5 Tech

**Много запросов,
несколько сотен разных запросов**

	Total Time		Total Time (CPU)	
	1,053,035 ms	(100%)	1,016,592 ms	(100%)
pareCalculation (Object, String, String	664,297 ms	(63.1%)	656,300 ms	(64.6%)
calcImplementation (java.util.Map, Str	664,297 ms	(63.1%)	656,300 ms	(64.6%)
SpringCGLIB$$bcbaa12a.getVrkme (St	357,888 ms	(34%)	354,897 ms	(34.9%)
SpringCGLIB$$bcbaa12a.getKnvi ()	222,393 ms	(21.1%)	220,639 ms	(21.7%)
cData (String, String, java.time.LocalDa	35,731 ms	(3.4%)	34,792 ms	(3.4%)
cerBySpringCGLIB$$2449cec2.load (jav	14,296 ms	(1.4%)	13,746 ms	(1.4%)
ce.lambda$calcImplementation$7 (j	8,809 ms	(0.8%)	8,545 ms	(0.8%)
ng, String, java.time.LocalDate, String)	8,516 ms	(0.8%)	8,340 ms	(0.8%)
SpringCGLIB$$bcbaa12a.getMlan (java	5,389 ms	(0.5%)	5,035 ms	(0.5%)

VisualVM 2.1.2

ru.x5.prices.conditions.Application (pid 66879)

Overview | Monitor | Threads | Sampler | Profiler

ru.x5.prices.conditions.Application (pid 66879)

Profiler | Settings

Profile: CPU | Memory | JDBC | Locks | Stop

Status: profiling running (21,871 methods instrumented)

Profiling results

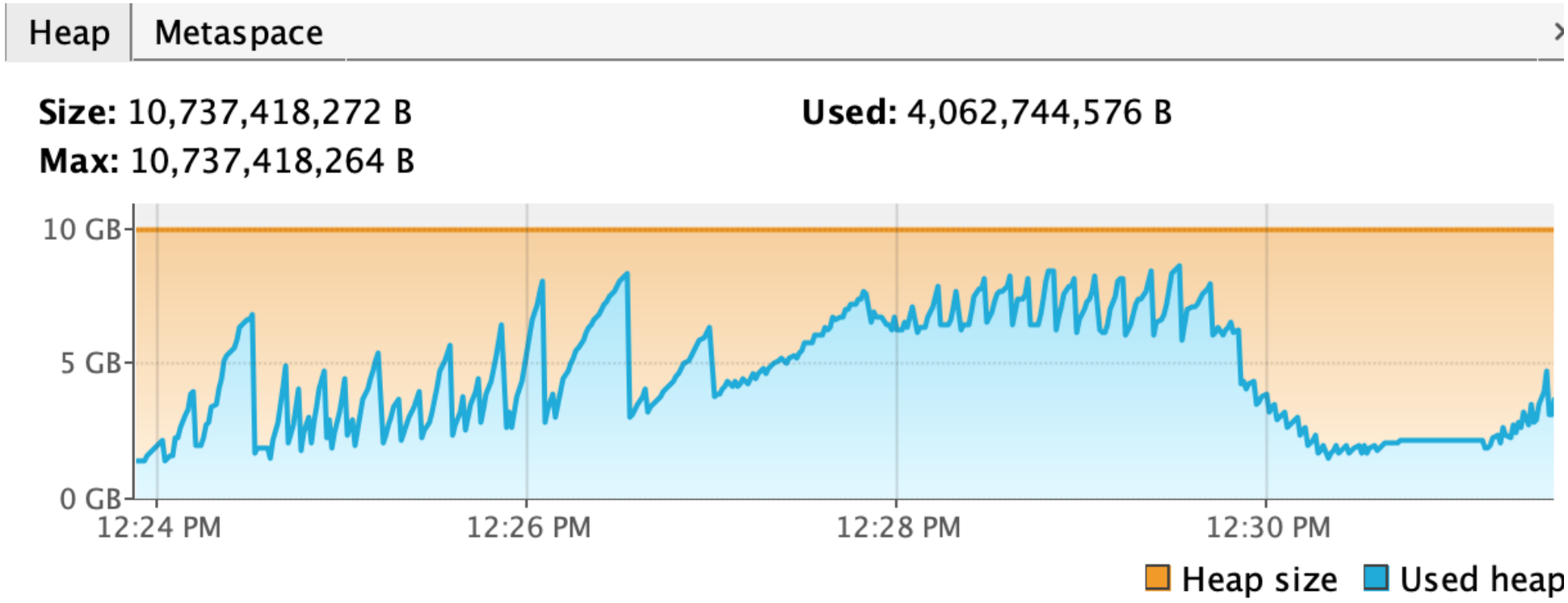
Results: View: Collected data: Snapshot

Name	Total Time		Total Time (CPU)		Invocations
http-nio-8087-exec-2	1,053,035 ms	(100%)	1,016,592 ms	(100%)	153,043
ru.x5.prices.conditions.service.processing.stage.impl.step1.Stage5Service.prepareCalculation (Object, String, String	664,297 ms	(63.1%)	656,300 ms	(64.6%)	1
ru.x5.prices.conditions.service.processing.stage.impl.step1.Stage5Service.calcImplementation (java.util.Map, Str	664,297 ms	(63.1%)	656,300 ms	(64.6%)	1
ru.x5.prices.conditions.storage.repository.MasterDataDao$$EnhancerBySpringCGLIB$$bcbaa12a.getVrkme (St	357,888 ms	(34%)	354,897 ms	(34.9%)	1
ru.x5.prices.conditions.storage.repository.MasterDataDao$$EnhancerBySpringCGLIB$$bcbaa12a.getKnvi ()	222,393 ms	(21.1%)	220,639 ms	(21.7%)	1
ru.x5.prices.conditions.storage.repository.CalcDataDao.setResultsSCalcData (String, String, java.time.LocalDa	35,731 ms	(3.4%)	34,792 ms	(3.4%)	1
ru.x5.prices.conditions.storage.repository.tables.MaraDataDao$$EnhancerBySpringCGLIB$$2449cec2.load (jav	14,296 ms	(1.4%)	13,746 ms	(1.4%)	1
ru.x5.prices.conditions.service.processing.stage.impl.step1.Stage5Service.lambda$calcImplementation$7 (j	8,809 ms	(0.8%)	8,545 ms	(0.8%)	4,845

Первый этап: расчёт

JPoint X5 Tech

Закешировать данные было отличным решением

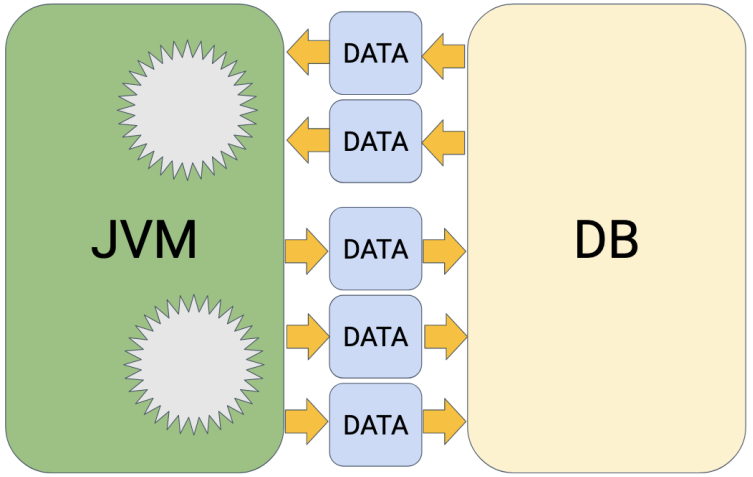


Первый этап: расчёт

JPoint X5 Tech

Выносим запросы за цикл, стало быстрее.
Но растёт потребление памяти.
Замкнутый круг

```
with werksCte as (select value from all_werks), // 10 728 334 records
     excludeWerksCte as (SELECT DISTINCT REFSITE
                           FROM mmsiteref where REFSITE != 'HR01'
                          UNION ALL
                         SELECT DISTINCT VLGFI
                           FROM tvkov where VLGFI != 'HR01'),
     t001wCte as (select DISTINCT WERKS, KUNNR from t001w
                  WHERE WERKS IN werksCte and WERKS not in excludeWerks
     ),
     assortCte as (select distinct WERKS, KUNNR, ASORT,SOTYP, ASSORTYP
           from t001wCte
                    INNER JOIN wrsz on wrsz.LOCNR = t001wCte.KUNNR
                    INNER JOIN wrs1 on wrs1.ASORT = wrsz.ASORT
         )
select  WERKS, KUNNR, FILIA, SOTYP, ASSORTYP, ARTNR, QUELL, NEGAT, STRLI
from wlk inner join assort on wlk1.FILIA = assort.ASORT and STRLI != 'X'
```



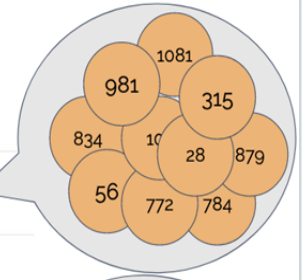
Прочитать и передать по сети 100 М строк

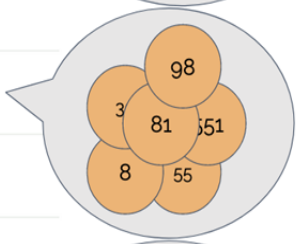
Обработать

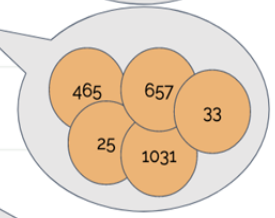
Передать по сети и записать 50-300 М строк

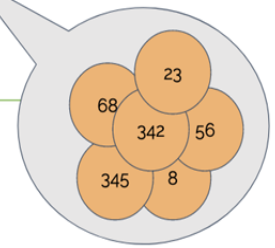
Второй этап: **потоки расчёта**

Федеральный округ	Пятёрочка	Перекрёсток	Чижик	Красный Яр и Слата	Всего
Центральный	6 797	538	302	–	7 703
Приволжский	4 993	135	127	–	5 256
Северо-Западный	1 913	161	–	–	2 081
Уральский	1 729	67	88	–	1 885
Южный	2 149	61	–	–	2 211
Сибирский	1 159	–	–	537	1 696
Северо-Кавказский	424	9	–	–	433
Дальневосточный	–	–	–	58	58
Всего	19 164	971	517	595	21 323





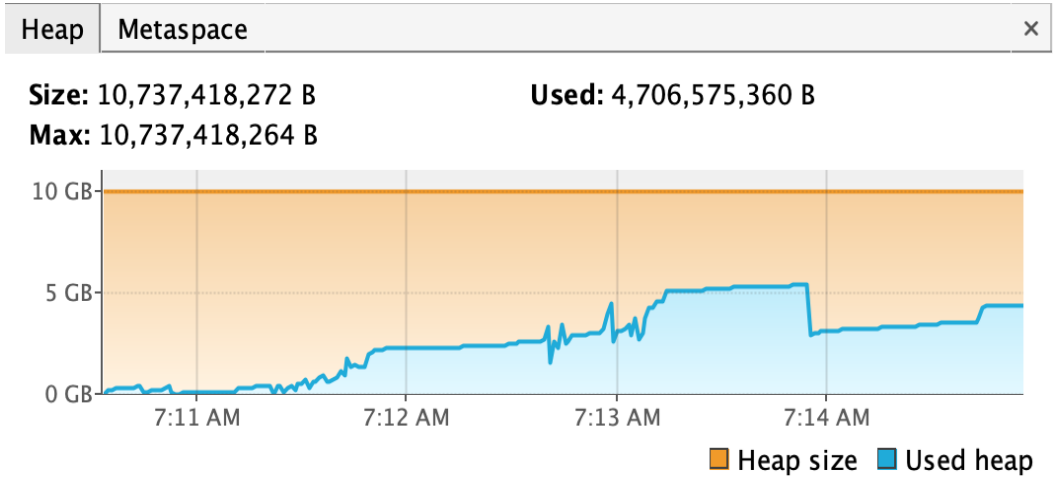




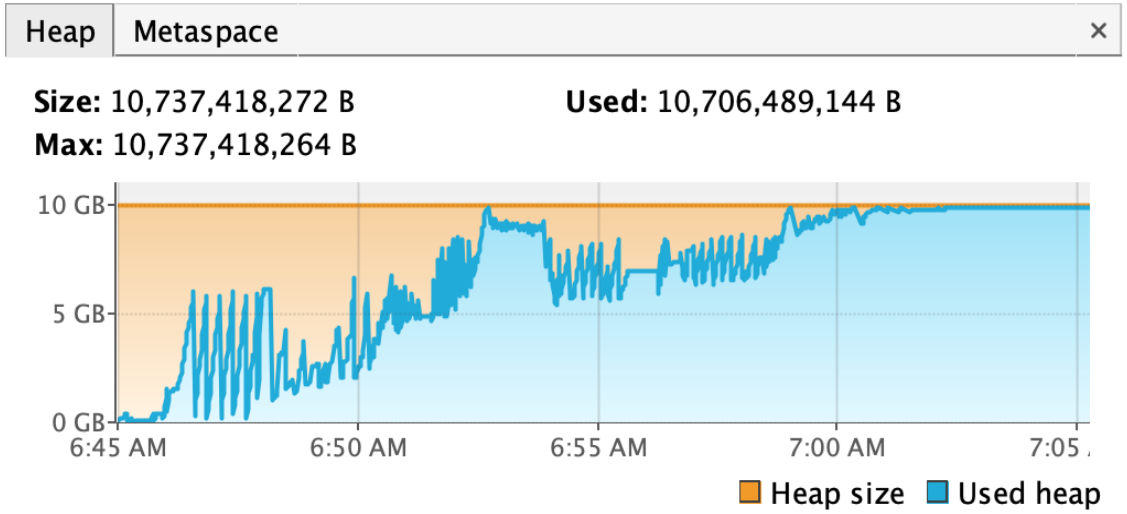
Второй этап: **потоки по группам**

JPoint X5 Tech

- **Большие группы – мало ресурсов**



- **Маленькие группы – много ресурсов**

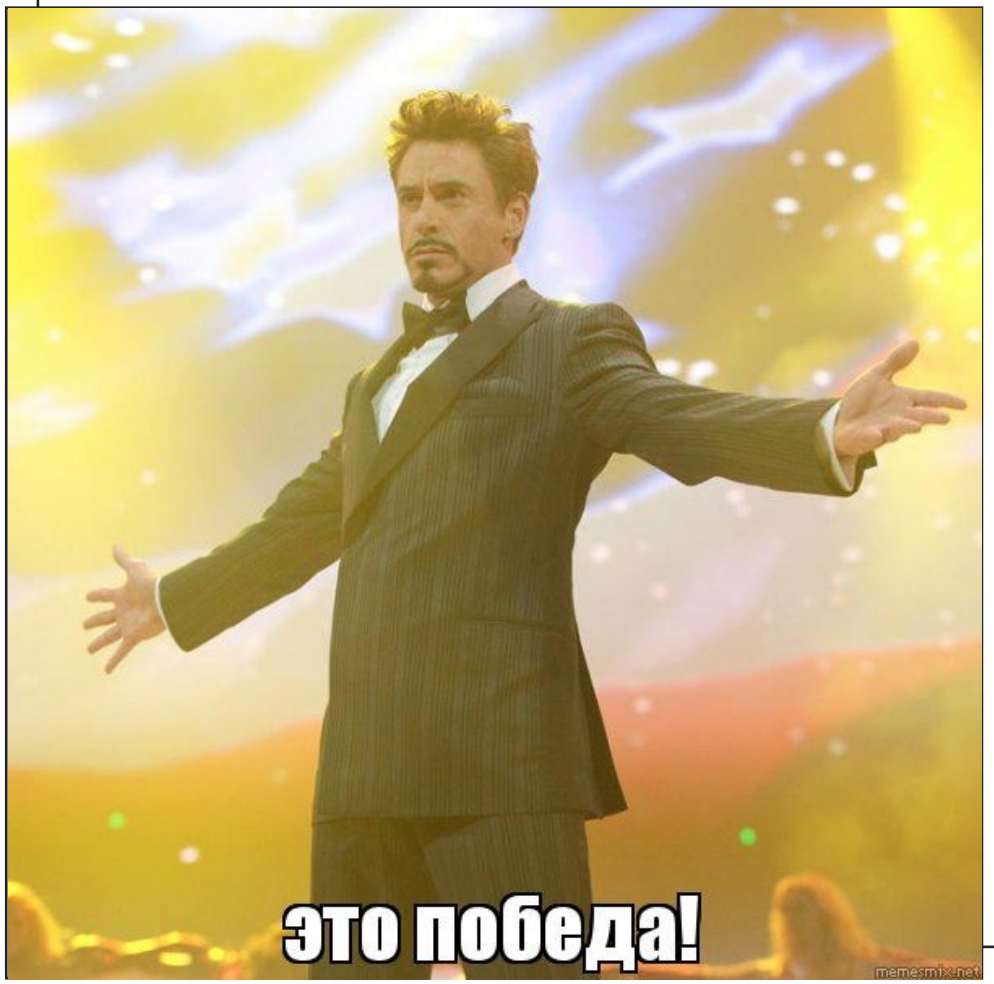


Второй этап: **потоки по группам**

JPoint X5 Tech

Оптимизировали до нужной длительности расчёта

- **1000** магазинов за 40 минут
- **20** инстансов сервиса расчёта

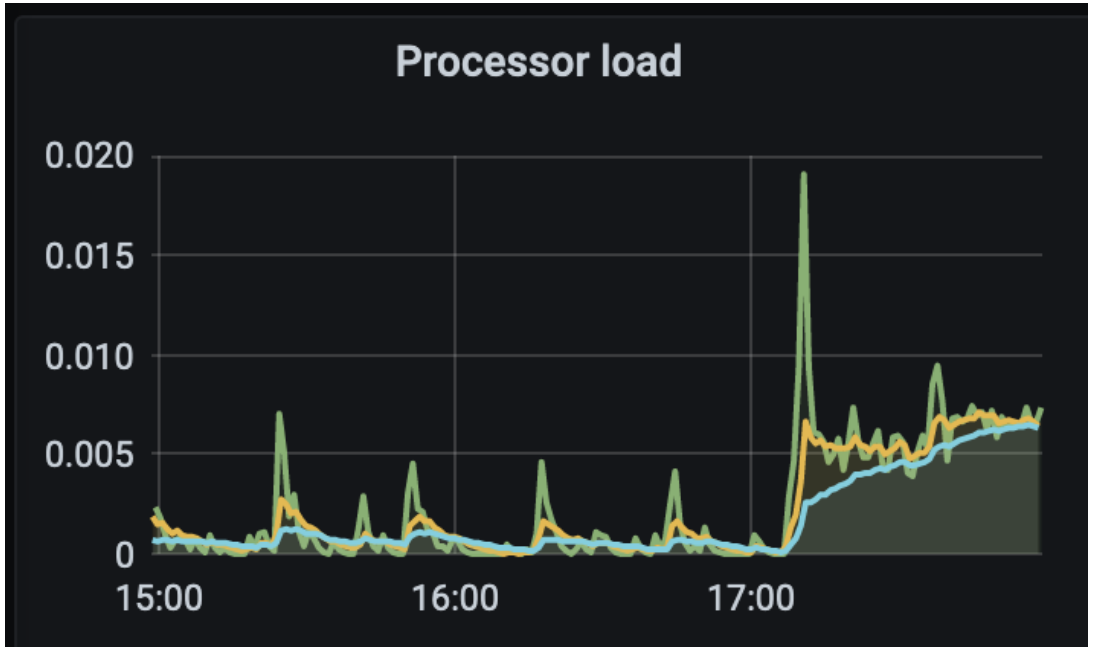


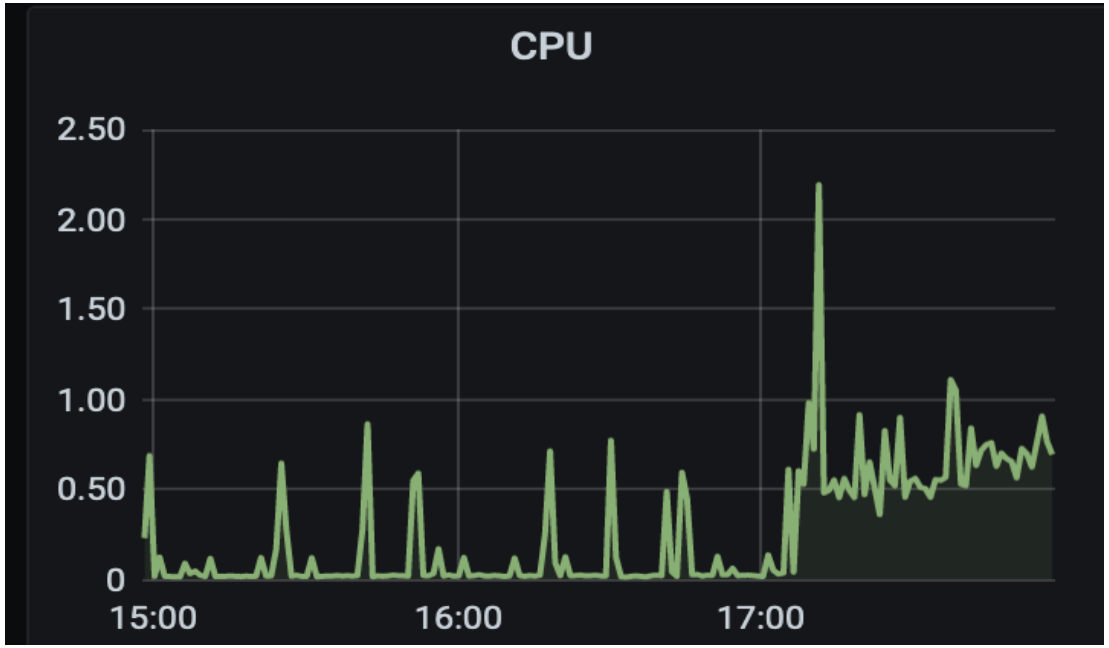
Второй этап: **потоки по группам**

JPoint X5 Tech

Оптимизировали до нужной длительности расчёта

Получилось!!! Или нет?





Второй этап: **потоки по группам**

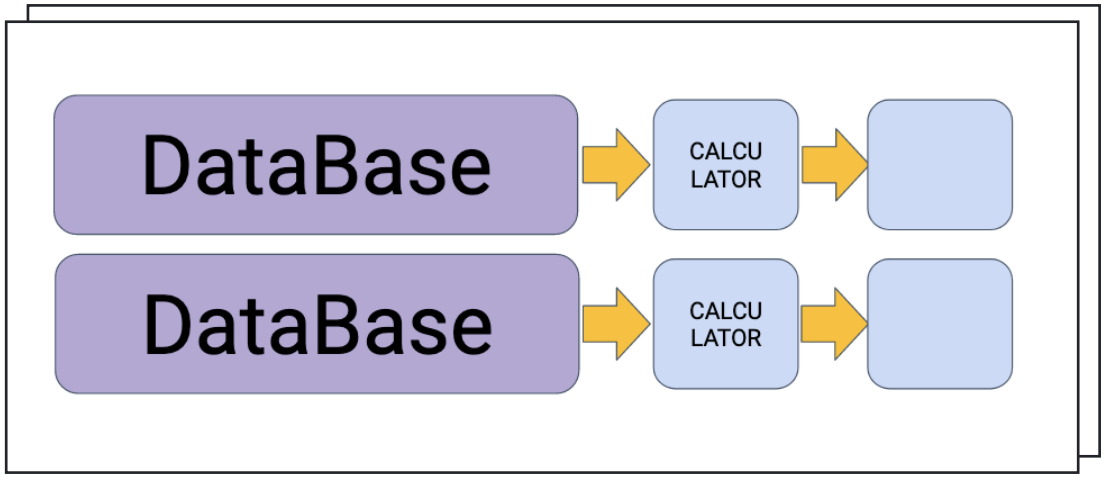
JPoint X5 Tech

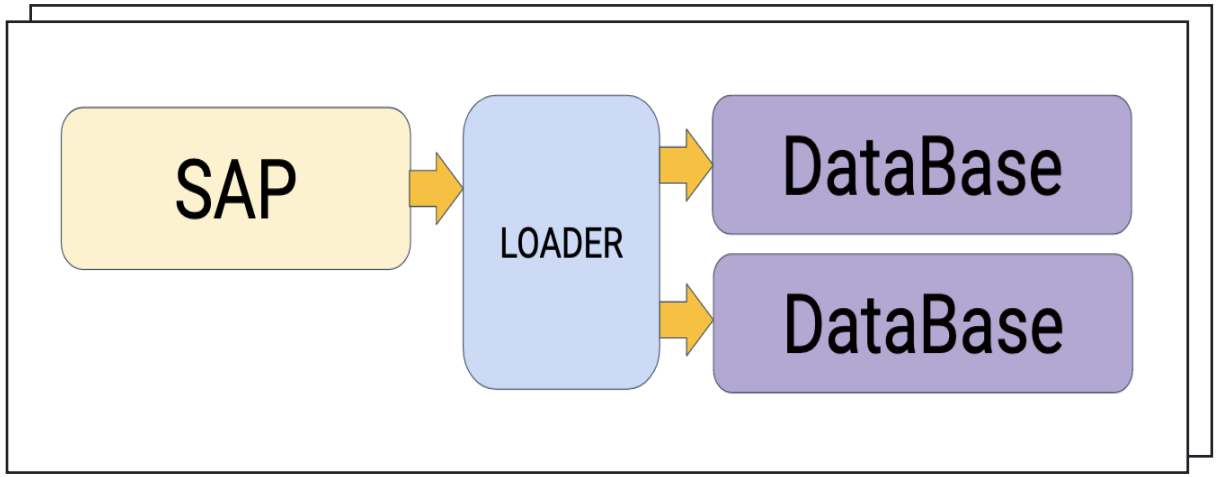
Узкое место – хранилище в памяти

Второй этап: **потоки по группам**

JPoint X5 Tech

Как устранить узкие места, если они под капотом хранилища?





Первый этап: **загрузка**

JPoint X5 Tech

**Как убедиться в том,
что все данные корректно загружены?**

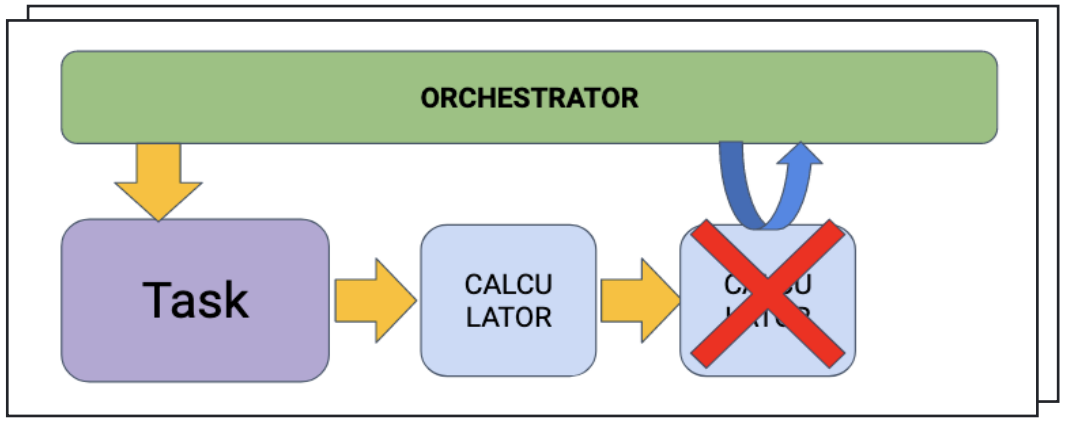
```
<mt_fr_umf>
  <logParameter name="BUILD_DATE">20230308193236
  <logParameter name="SENDER">SRV</logParameter>
  <logParameter name="SEND_TYPE">F</logParameter
  <logParameter name="SEND_ID">SRV_100500</logPa
  <logParameter name="RECORD_NUMBER">95000</logP
  <logParameter name="TABLE_NAME">JDA_IN_INVENTO
  <logParameter name="MSG_ID">7a580864-b8b4-0725
  <logParameter name="MSG_ROW_COUNT">0</logParam
  <rows/>
</mt_fr_umf>
```

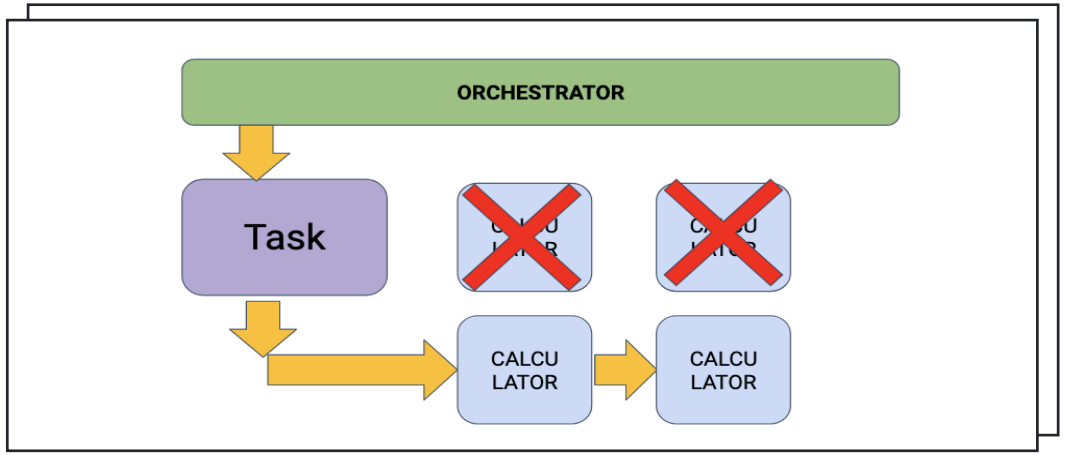
Ограничения

Сотни тысяч пакетов
по 300+ PLU = сотни миллионов цен

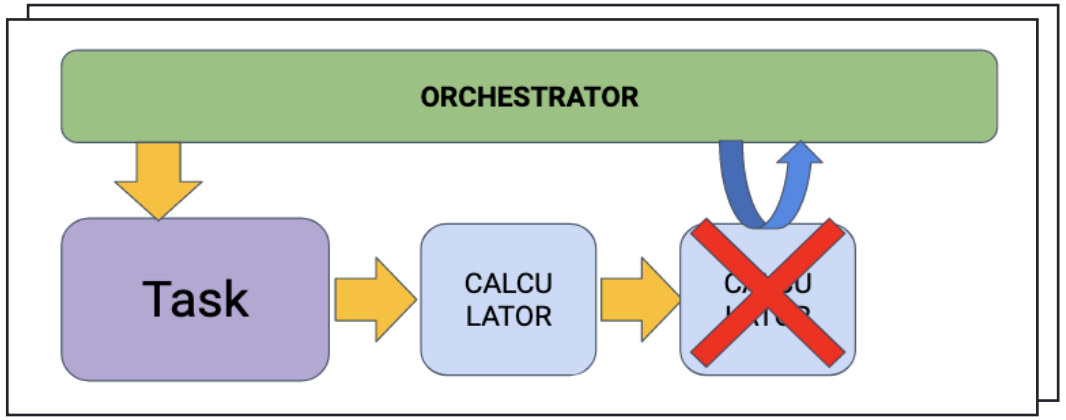
Оркестратор

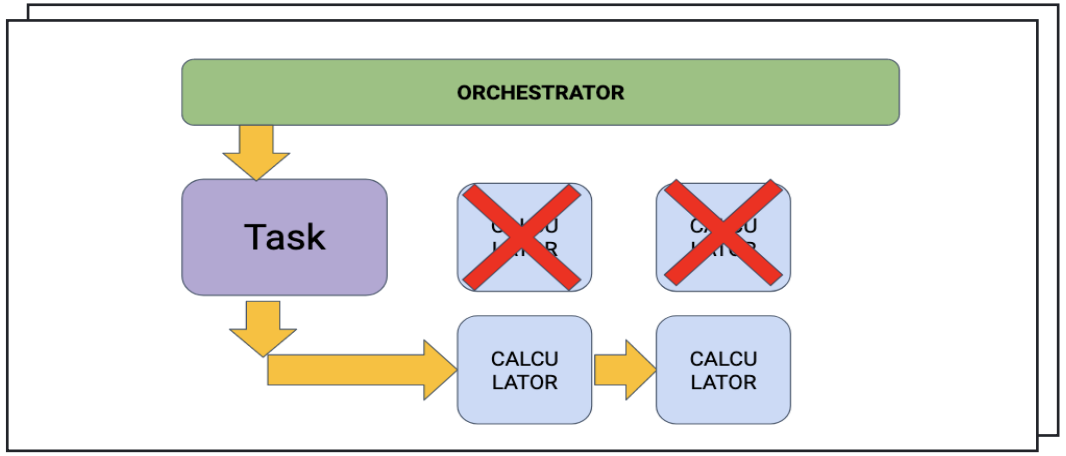
Мультивселенная безумия





Оркестратор

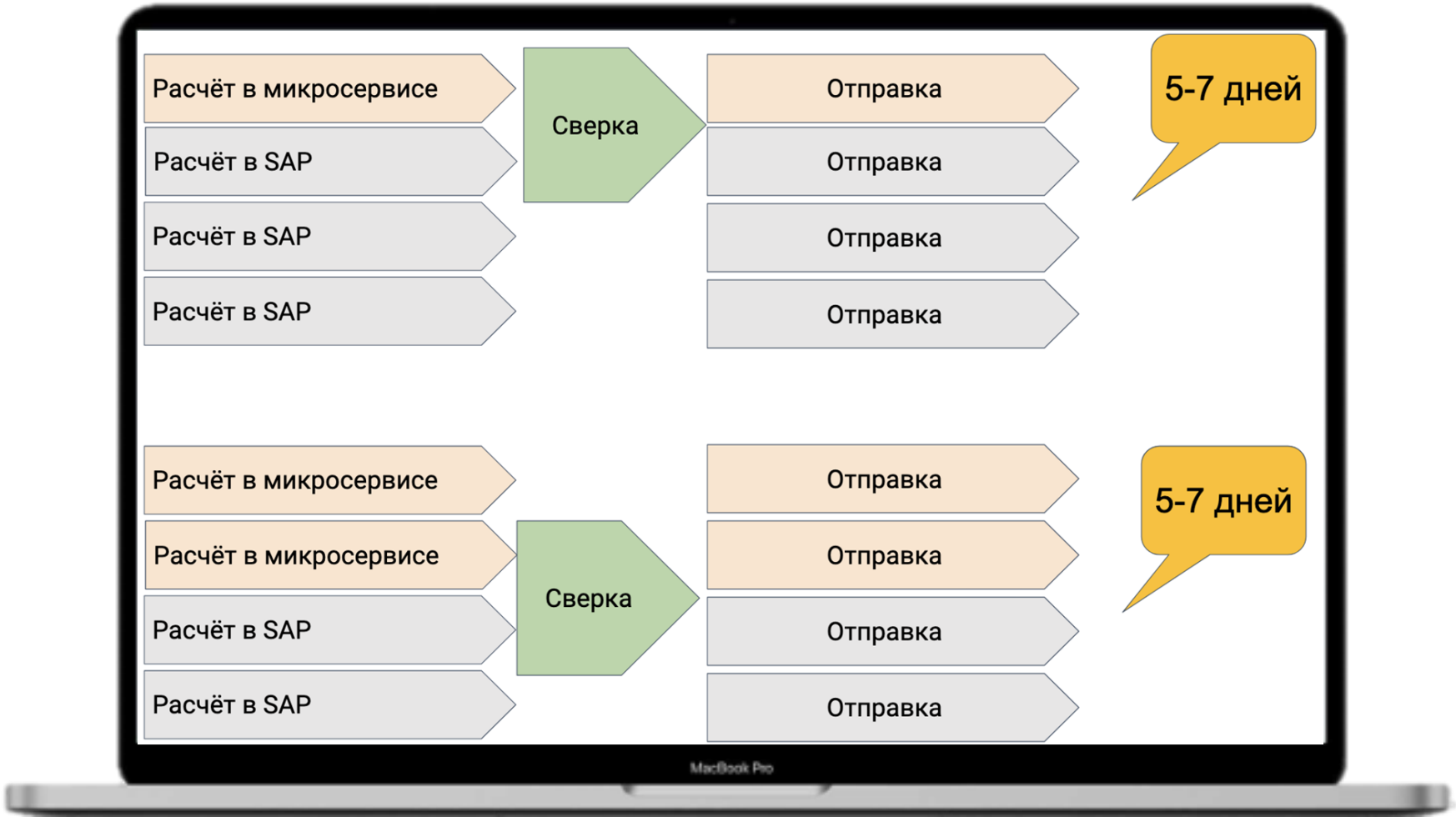




- Это хелсчеки сервисов или алгоритма?
- Что делать с обилием хелсчеков?

Внедрение

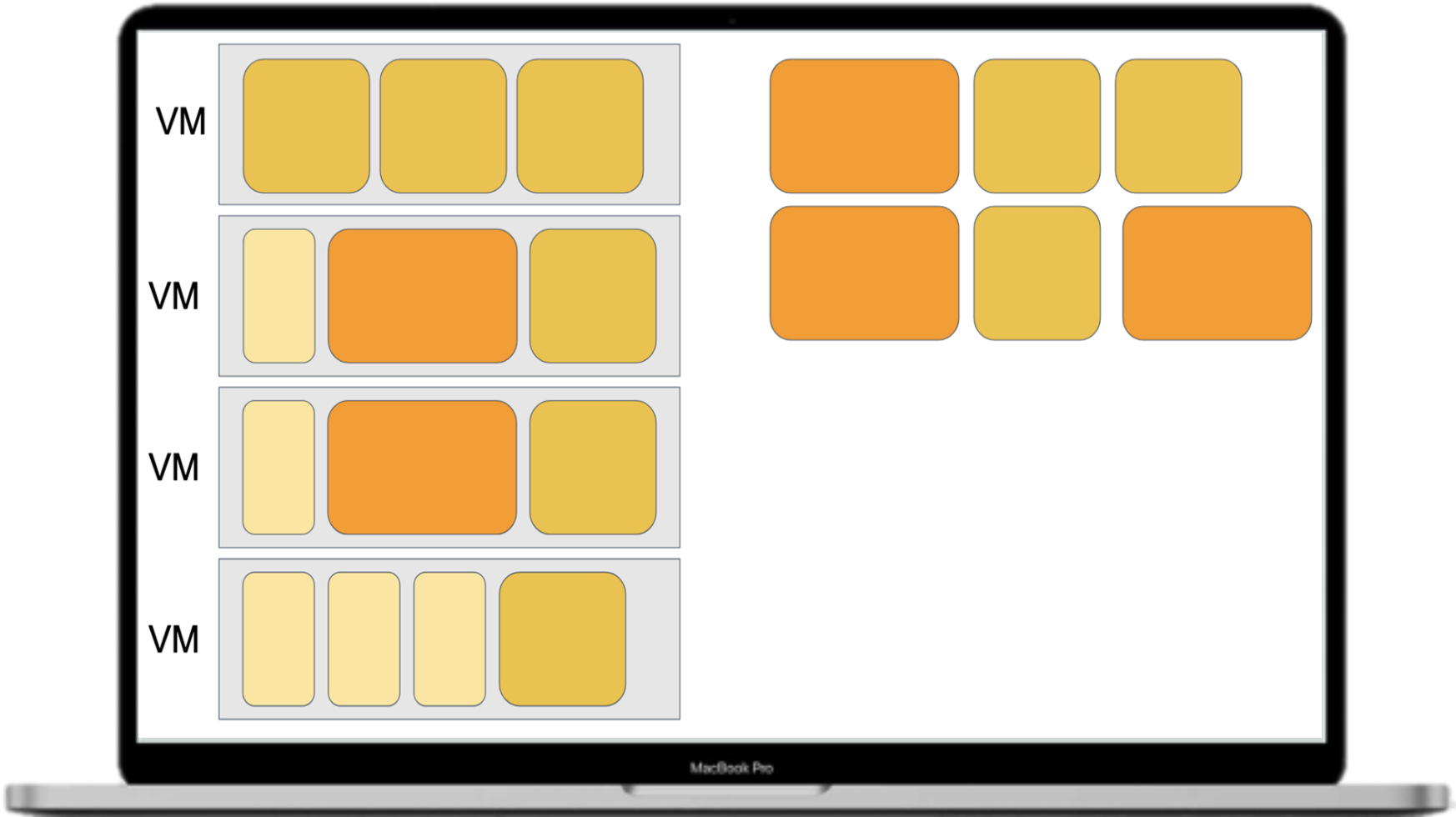
Бесшовный переход с САП на микросервисы



Утилизация ресурсов

JPoint X5 Tech

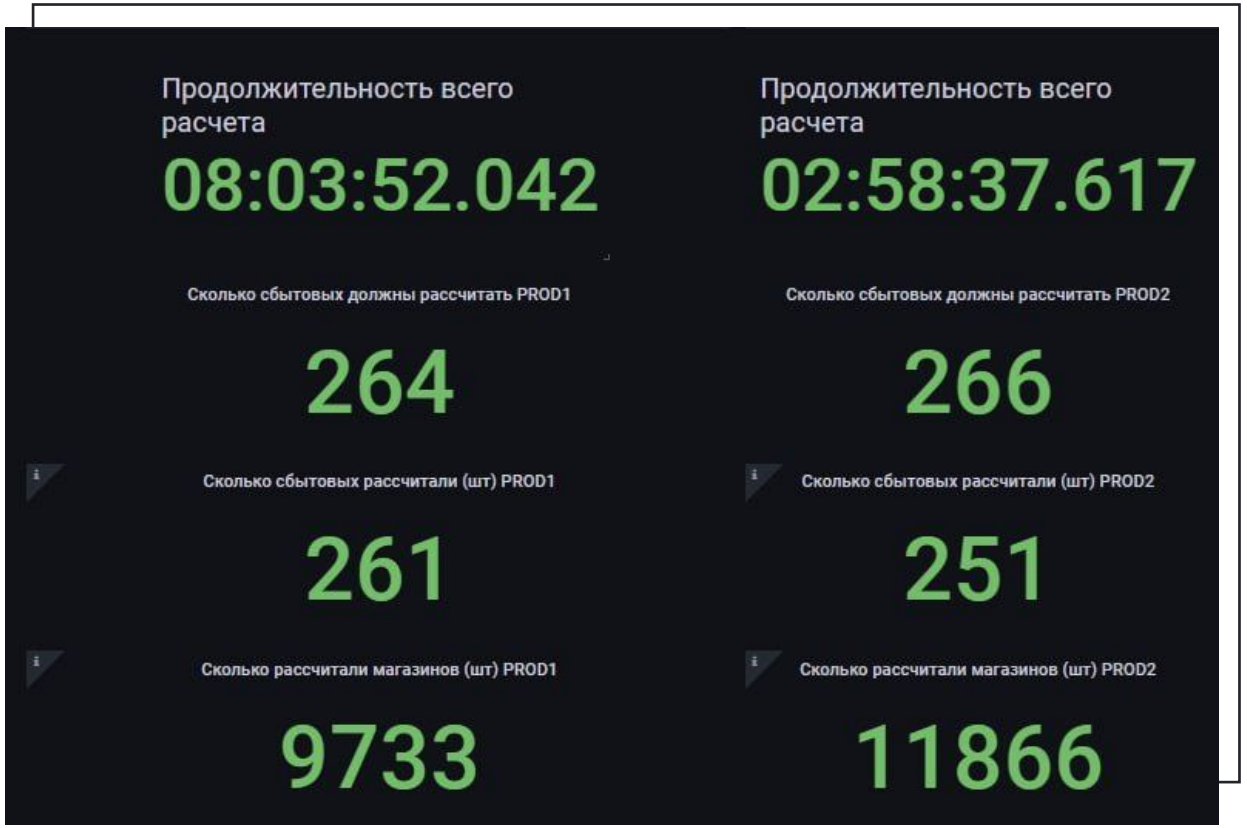
Олимпиадная задача



Перенос **алгоритма**

JPoint X5 Tech

Лучше тысячи слов



SAP Загрузка

НЕТ

SAP Расчёт

2,5 часа

SAP Отправка

1 час

ВЫВОДЫ

JPoint X5 Tech

- Только профилировщик

ВЫВОДЫ

JPoint X5 Tech

- Только профилировщик
- Оптимизируем самый критичный сервис

ВЫВОДЫ

JPoint X5 Tech

- Только профилировщик
- Оптимизируем самый критичный сервис
- Когда он доведён до оптимума - нагрузка на все сервисы параллельно

ВЫВОДЫ

JPoint X5 Tech

- Только профилировщик
- Оптимизируем самый критичный сервис
- Когда он доведён до оптимума - нагрузка на все сервисы параллельно
- Узкое место может мигрировать в другие сервисы/архитектурные юниты

ВЫВОДЫ

JPoint X5 Tech

- Только профилировщик
- Оптимизируем самый критичный сервис
- Когда он доведён до оптимума - нагрузка на все сервисы параллельно
- Узкое место может мигрировать в другие сервисы/архитектурные юниты
- Соотношение 80% результата даёт оптимизация 20% критичных участков

X5 Tech

Обо мне

- **Более 20 лет опыта** во всех областях разработки
- Предпочитаю баланс результата и усилий
- Fake it until make it everywhere
- Ценность не в технологиях – ценность в продукте

Алексей Топчий

@topchy_alexey