

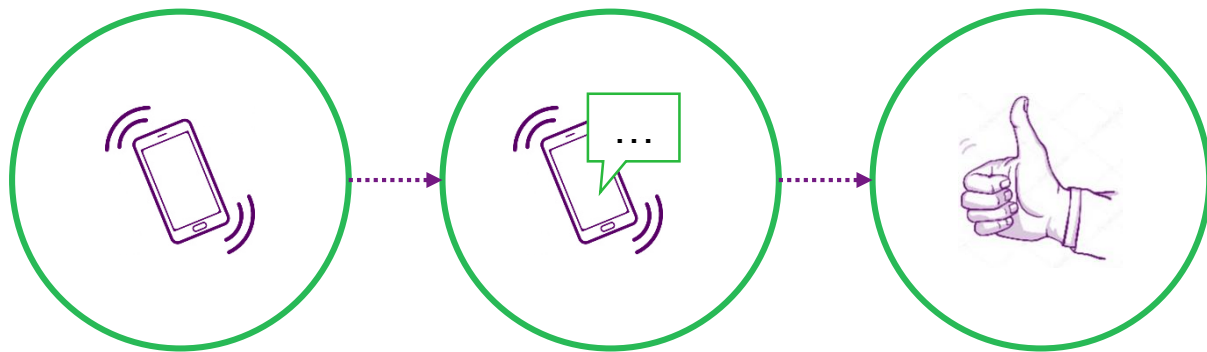
Держим тонус моделей в продуктиве

Селезнев Артем

Data Scientist @ MegaFon



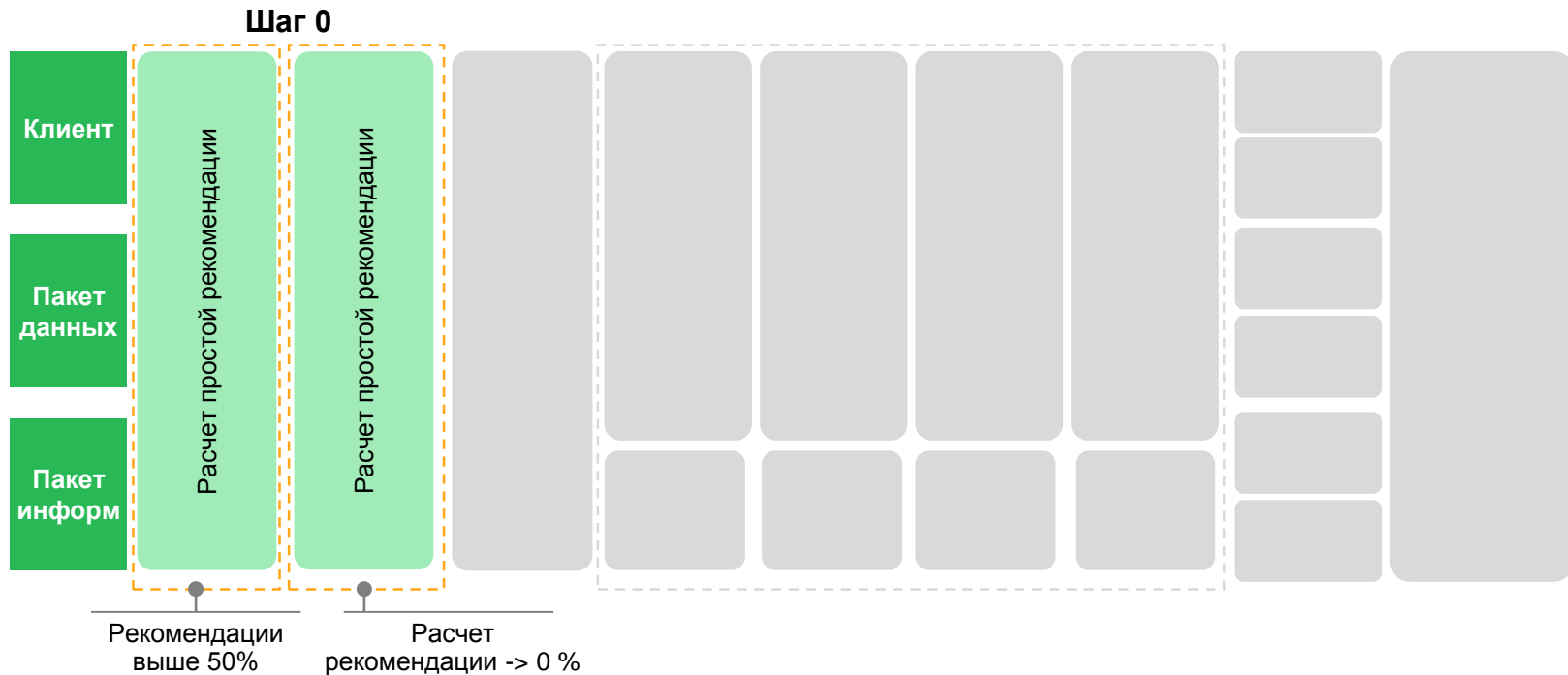
РЕКОМЕНДУЕМ В УСЛОВИЯХ ТРЯСКИ



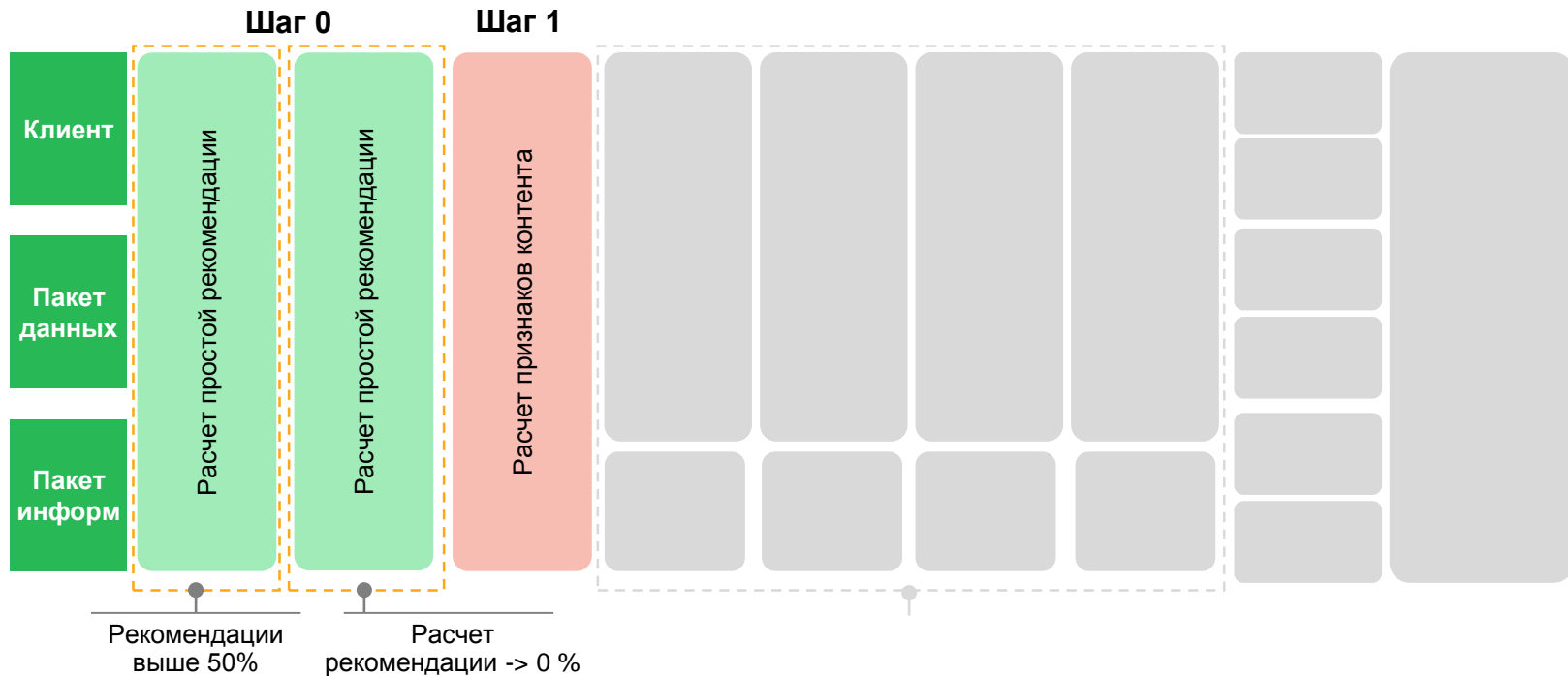
A woman with her hair in a bun, wearing a blue jumpsuit and white gloves, is leaning over the open hood of a car. She is working on the engine compartment. The image has a purple tint. The text "Что там «под капотом»?" is overlaid in white.

Что там «под капотом»?

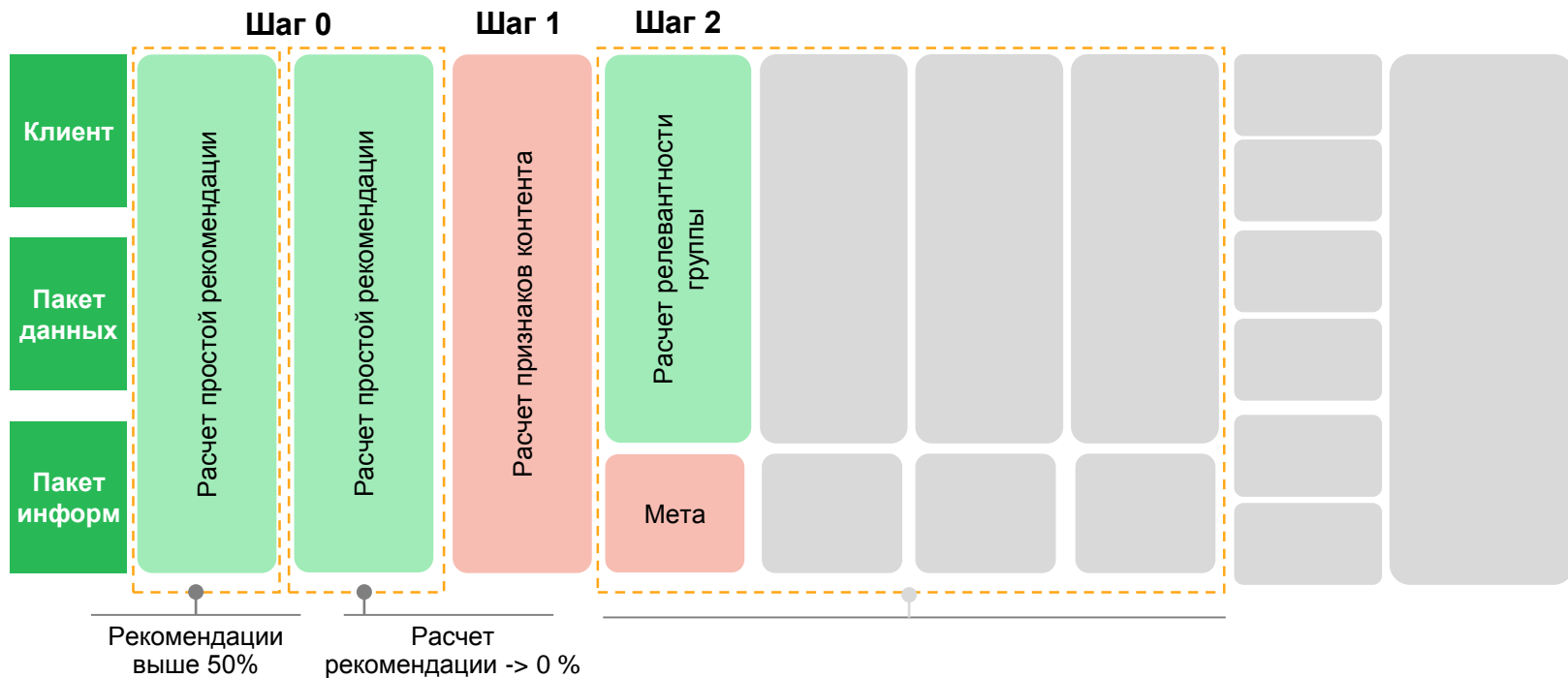
КОГДА ДАННЫХ МНОГО



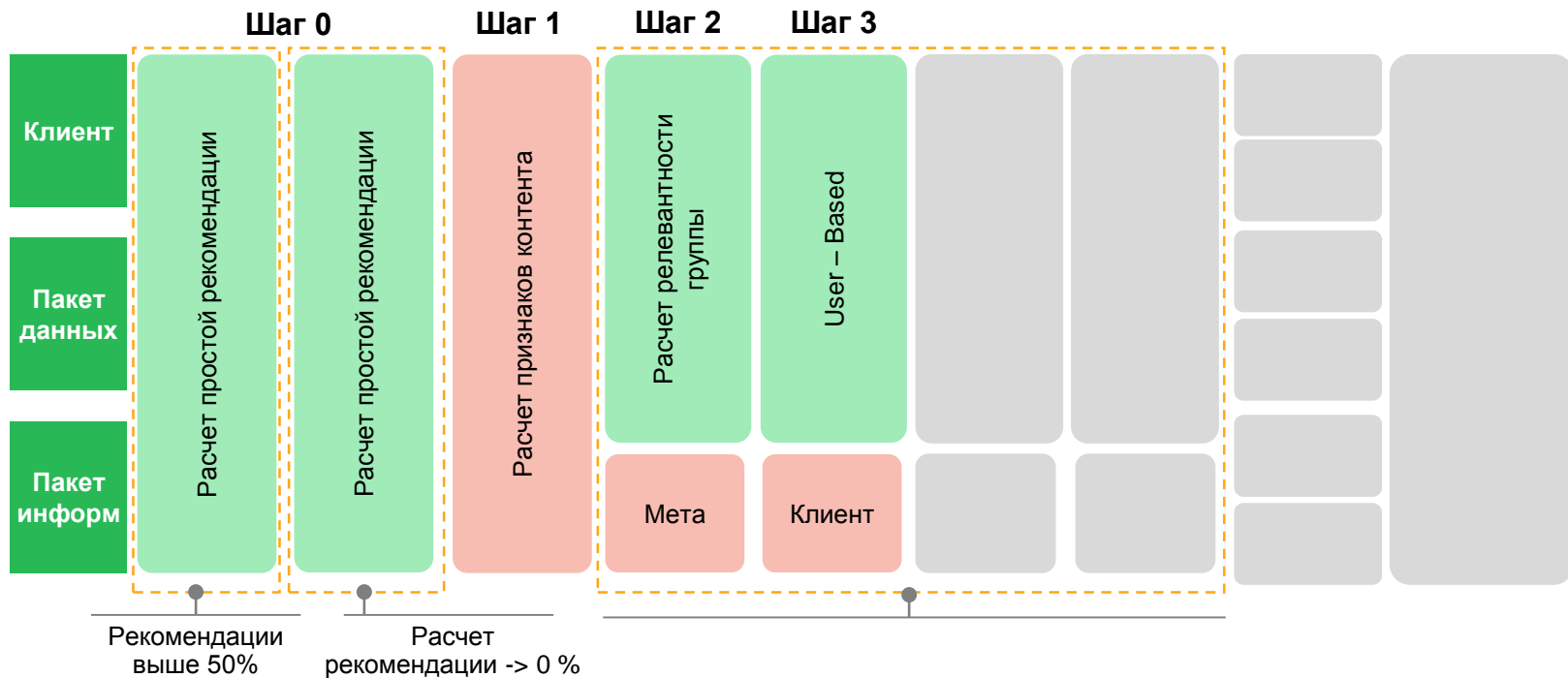
КОГДА ДАННЫХ МНОГО



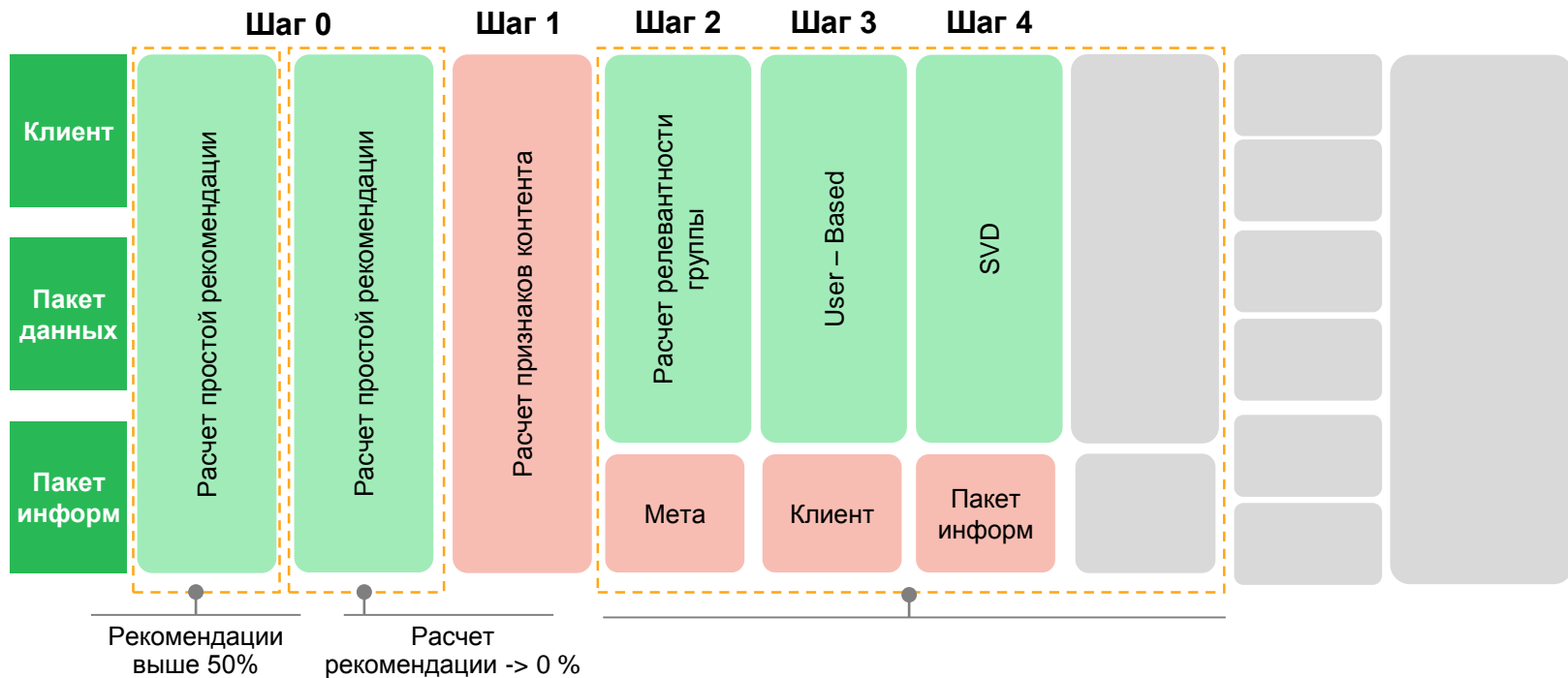
КОГДА ДАННЫХ МНОГО



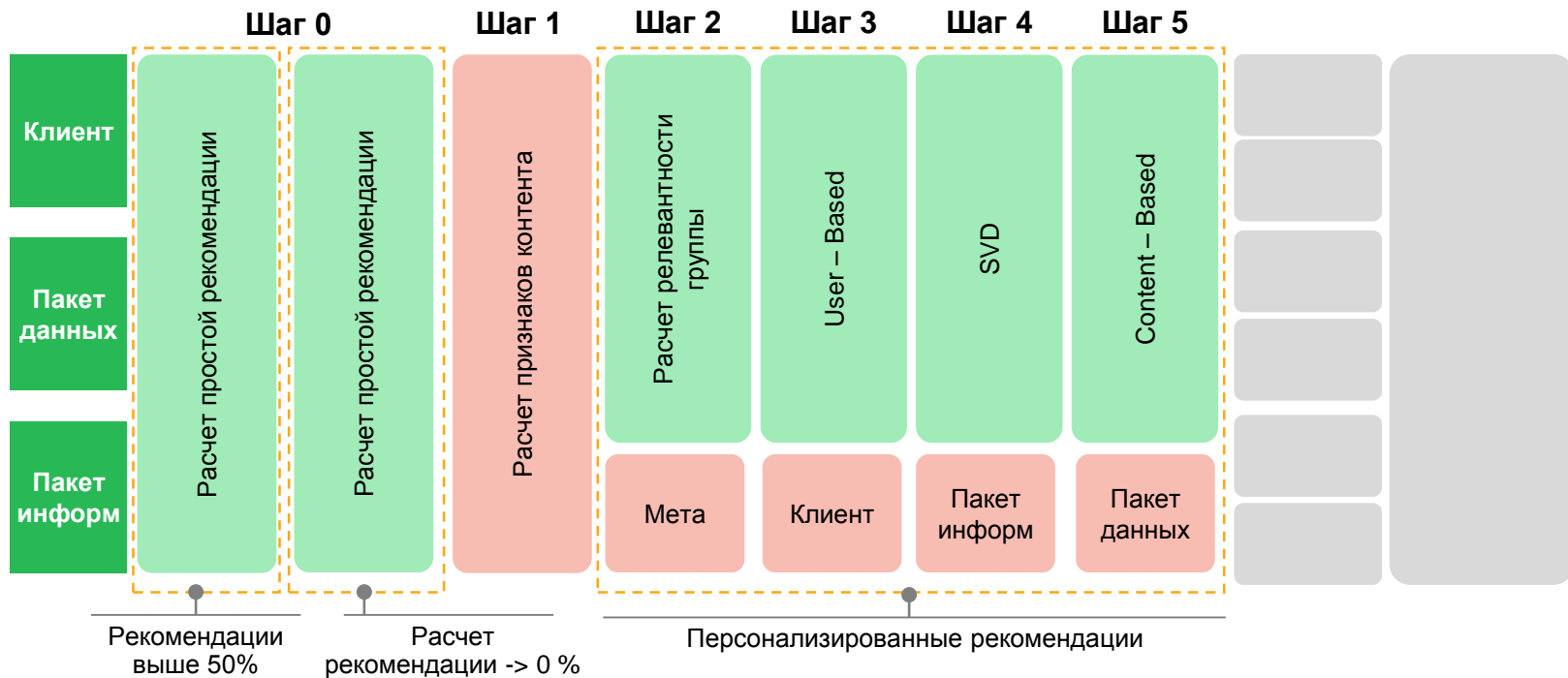
КОГДА ДАННЫХ МНОГО



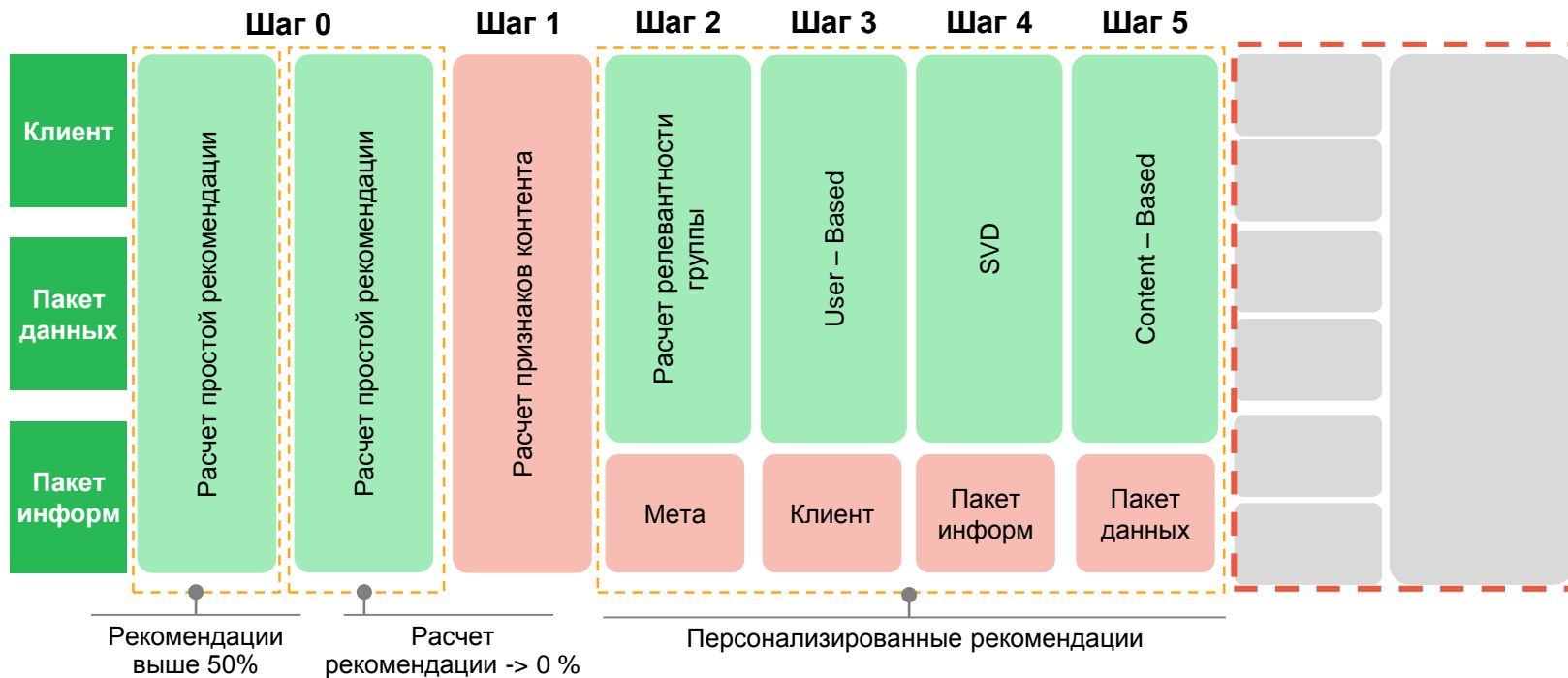
КОГДА ДАННЫХ МНОГО



КОГДА ДАННЫХ МНОГО



КОГДА ДАННЫХ МНОГО (и ещё не хватает)



ПРОБЛЕМЫ КОНТЕНТА

1. Голодание (специализированного) контента

ПРОБЛЕМЫ КОНТЕНТА

1. Голодание (специализированного) контента
2. Каннибализация популярным контентом

ПРОБЛЕМЫ КОНТЕНТА

1. Голодание (специализированного) контента
2. Каннибализация популярным контентом
3. Постоянство контента (не постоянство пользователя)

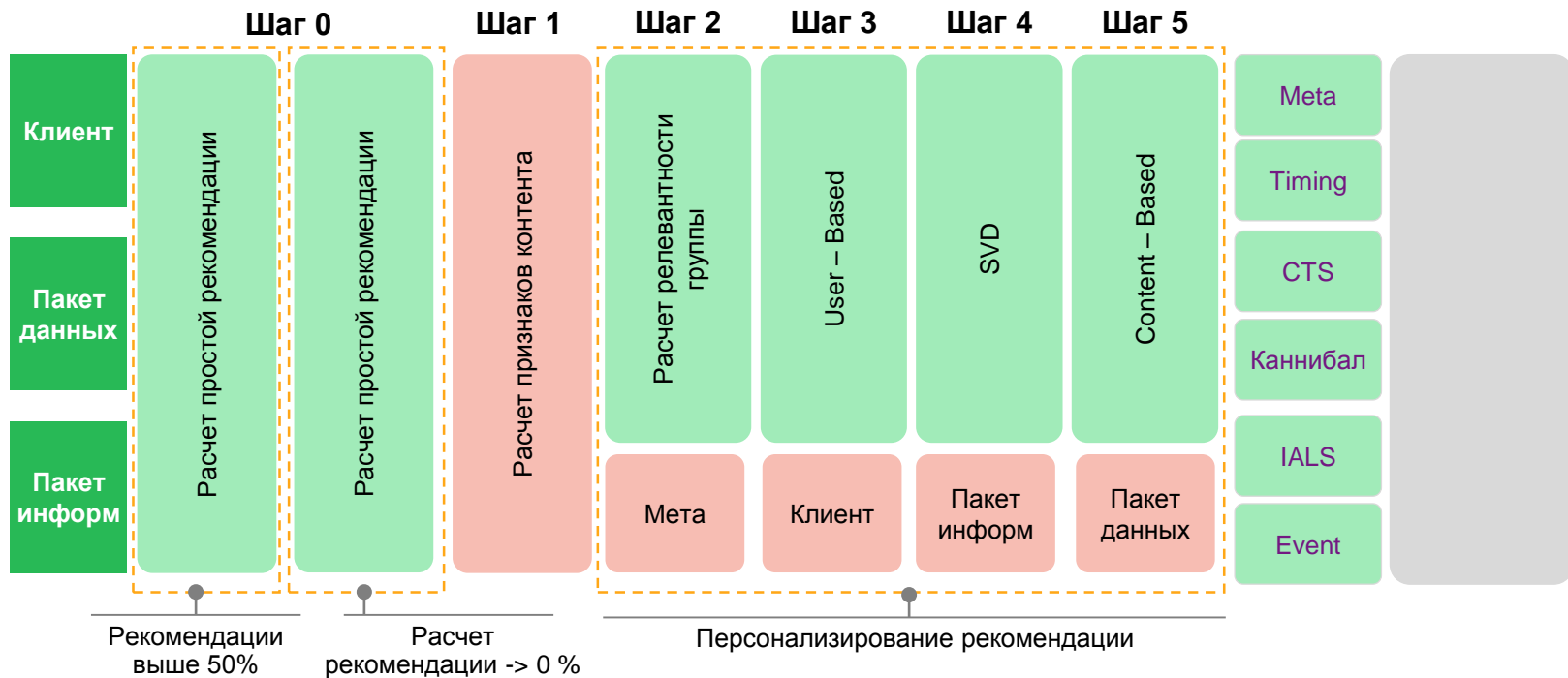
ПРОБЛЕМЫ КОНТЕНТА

1. Голодание (специализированного) контента
2. Канныбализация популярным контентом
3. Постоянство контента (не постоянство пользователя)
4. Новые форматы и содержание предложений

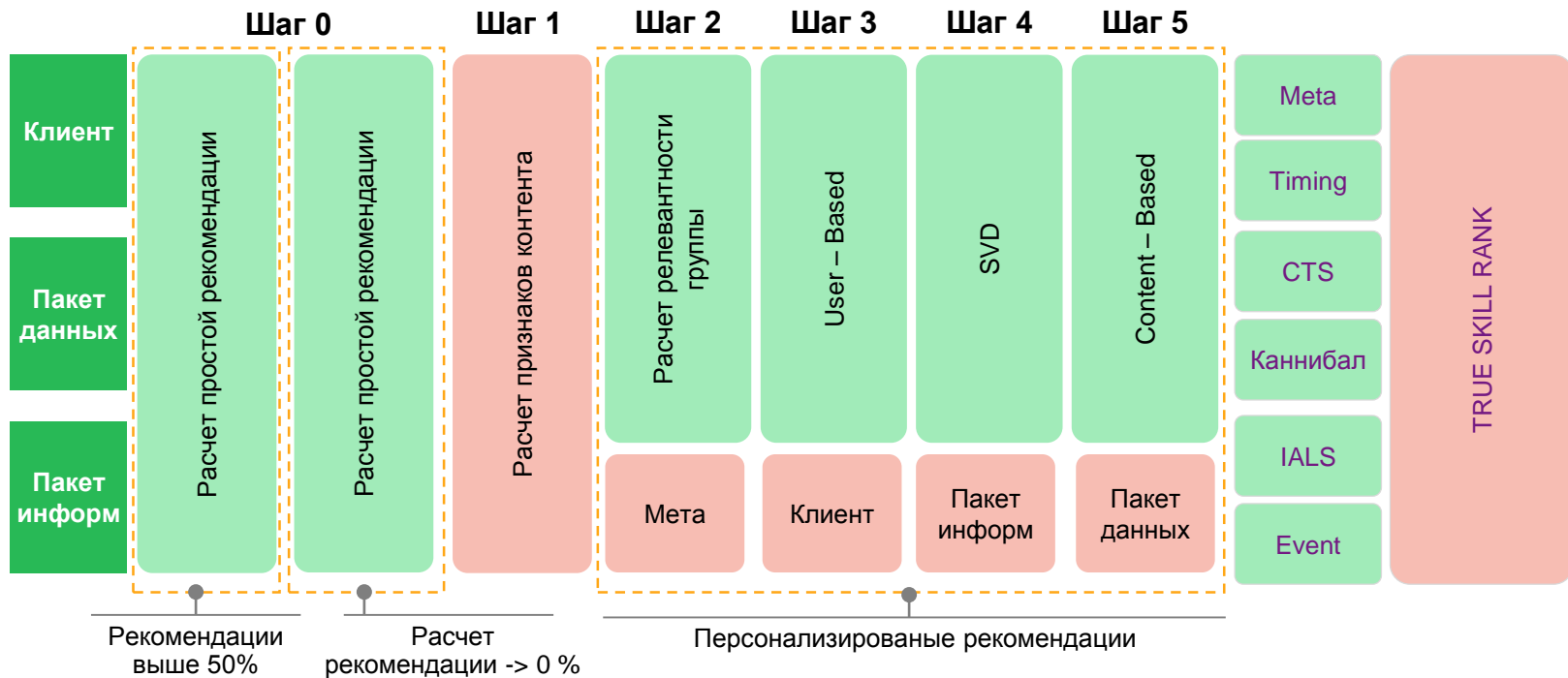
ПРОБЛЕМЫ КОНТЕНТА

1. Голодание (специализированного) контента
2. Канныализация популярным контентом
3. Постоянство контента (не постоянство пользователя)
4. Новые форматы и содержание предложений
5. События по контенту

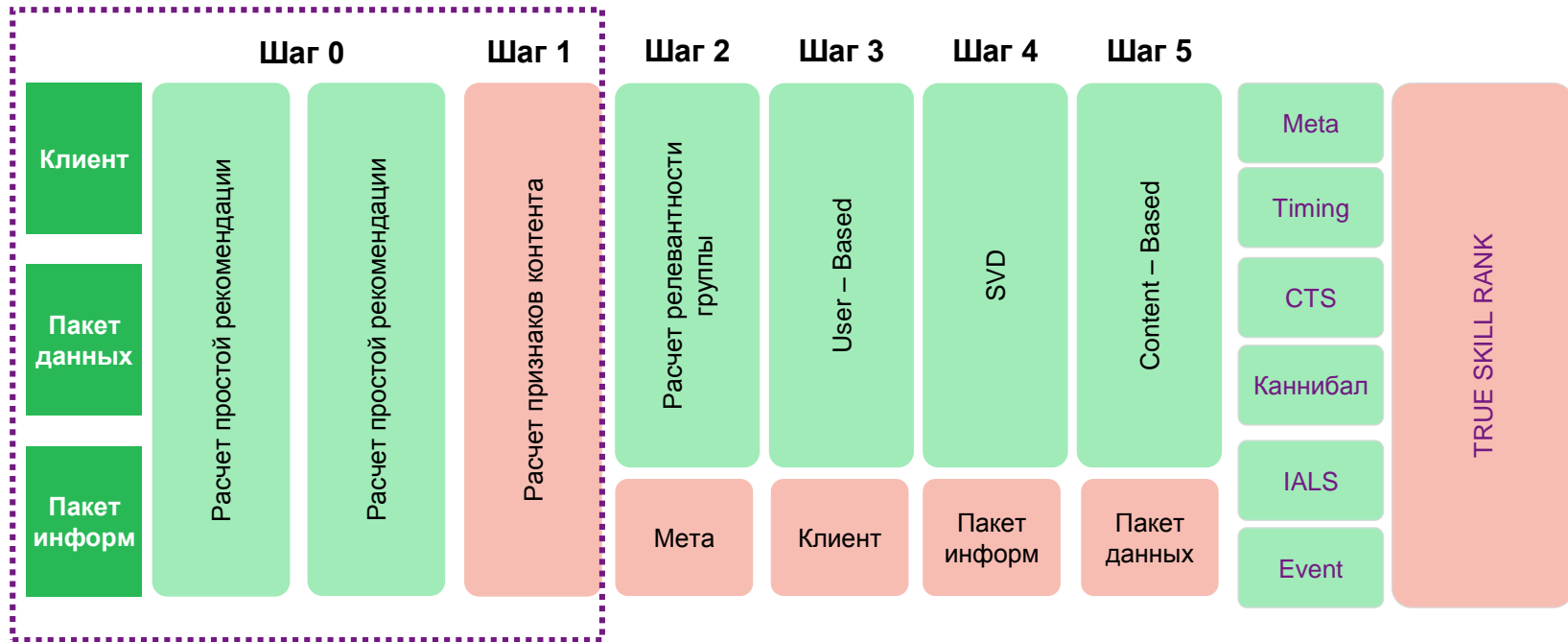
КОГДА ДАННЫХ МНОГО



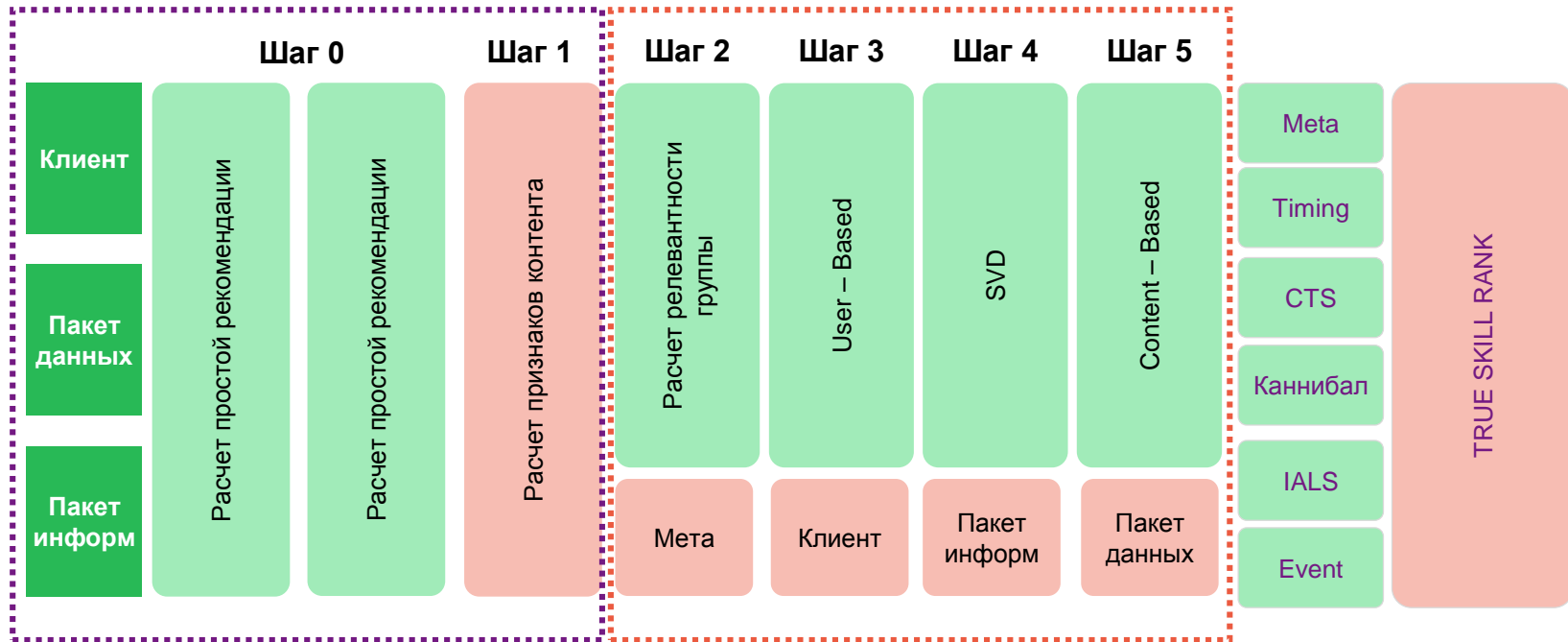
КОГДА ДАННЫХ МНОГО



КОГДА ДАННЫХ МНОГО



КОГДА ДАННЫХ МНОГО



КОГДА ОБУЧАЕМСЯ

После отклика пользователя?

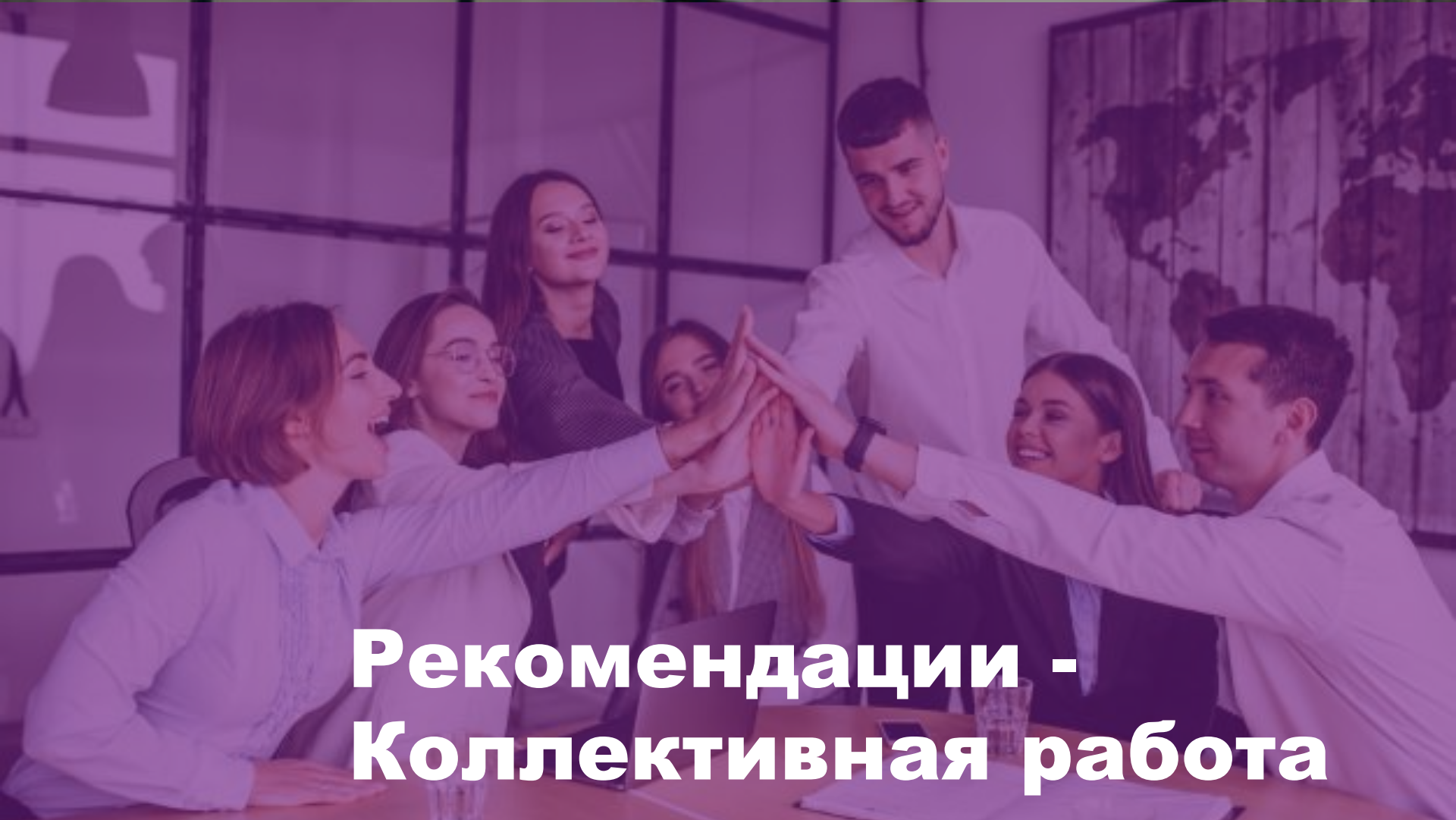
1 раз в месяц

КОГДА ОБУЧАЕМСЯ

~~После отклика пользователя?~~

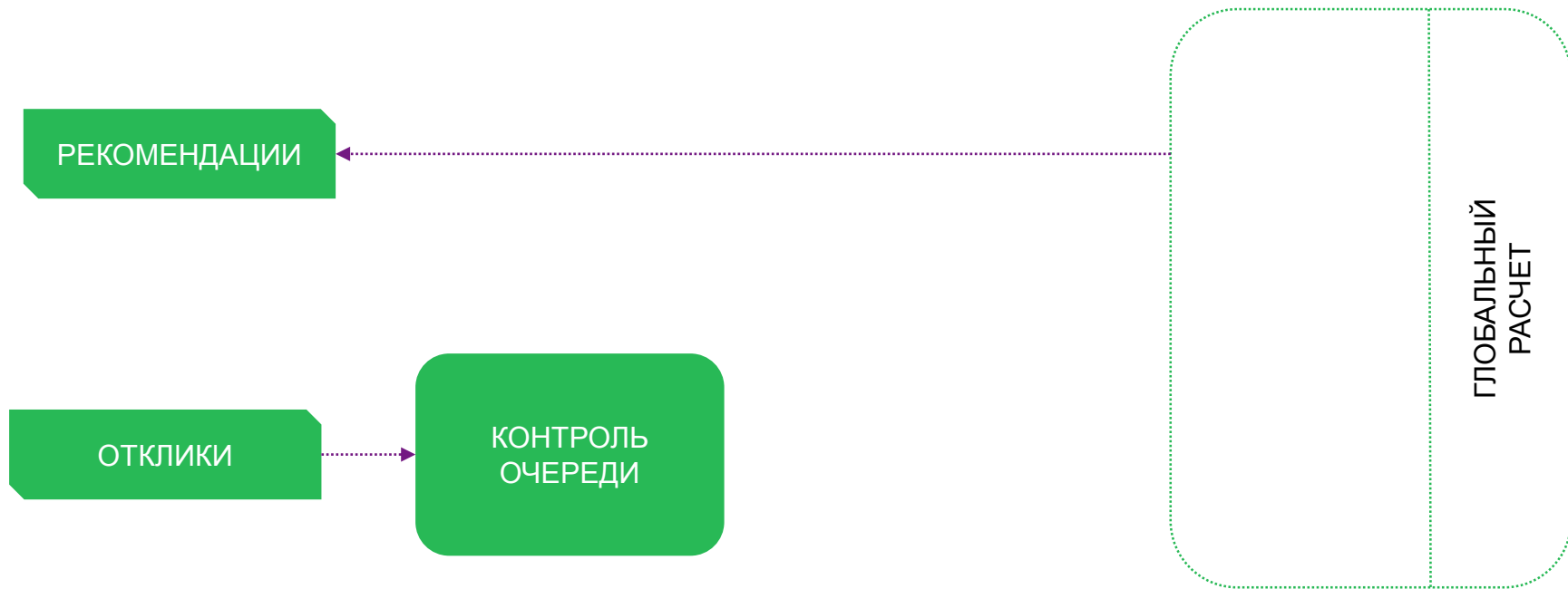
После отклика группы

1 раз в месяц

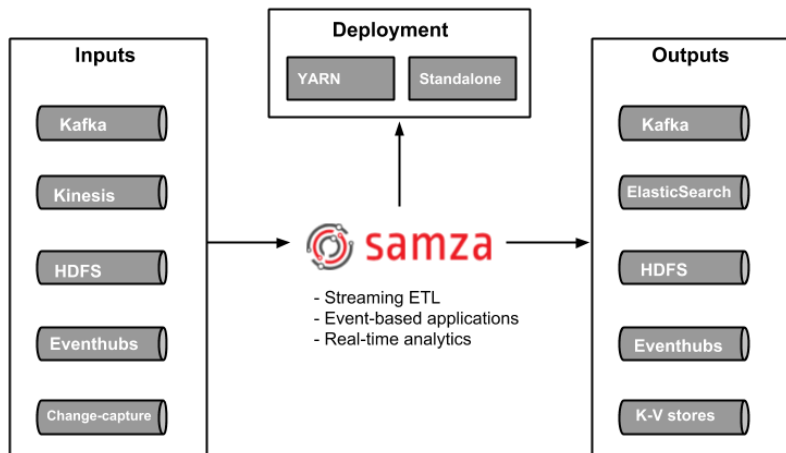


Рекомендации - Коллективная работа

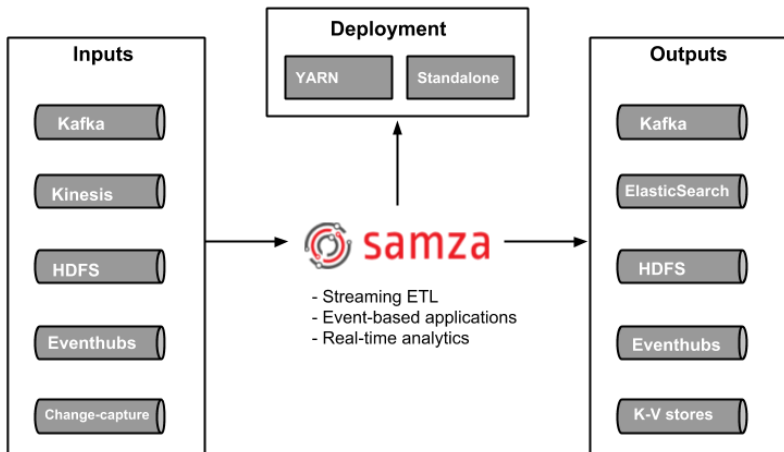
ПРОЦЕССЫ В СХЕМЕ



SAMZA

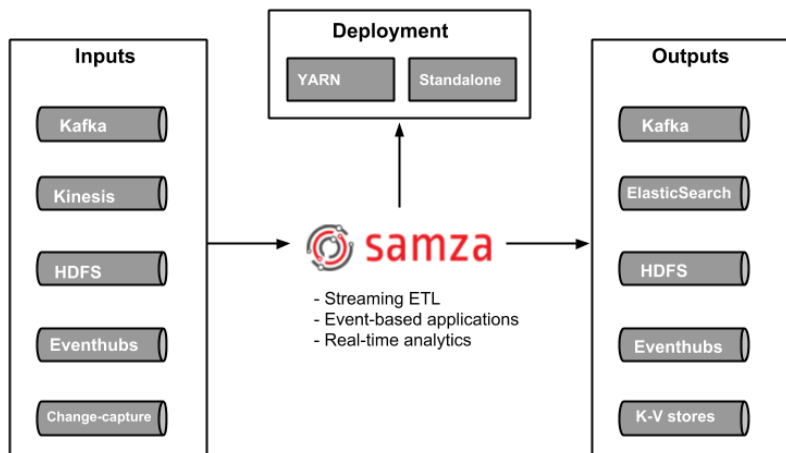


SAMZA



1. Появление объект: ?

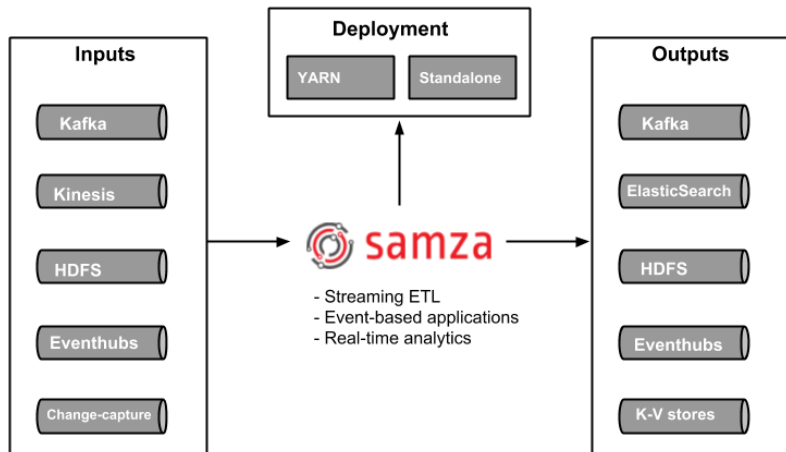
SAMZA



1. Появление объект: 0.1

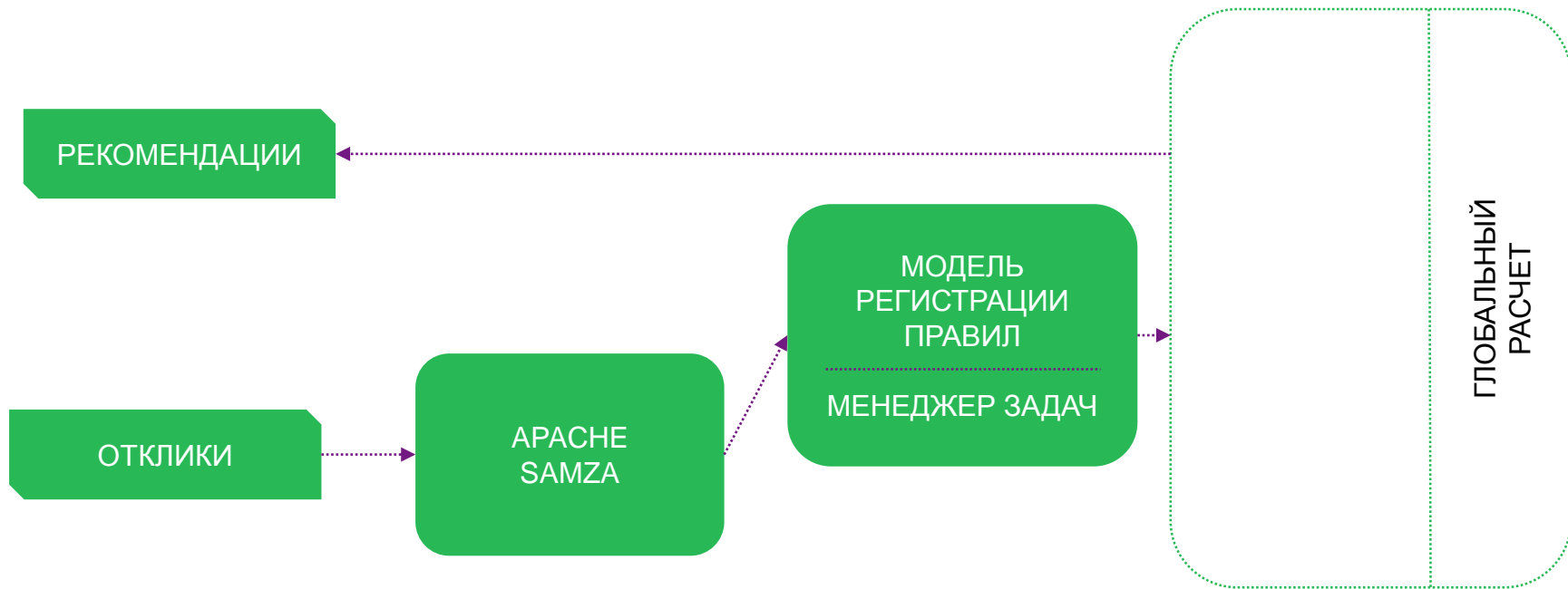
2. Оценка момента
появления: $(m_i)^k$

SAMZA

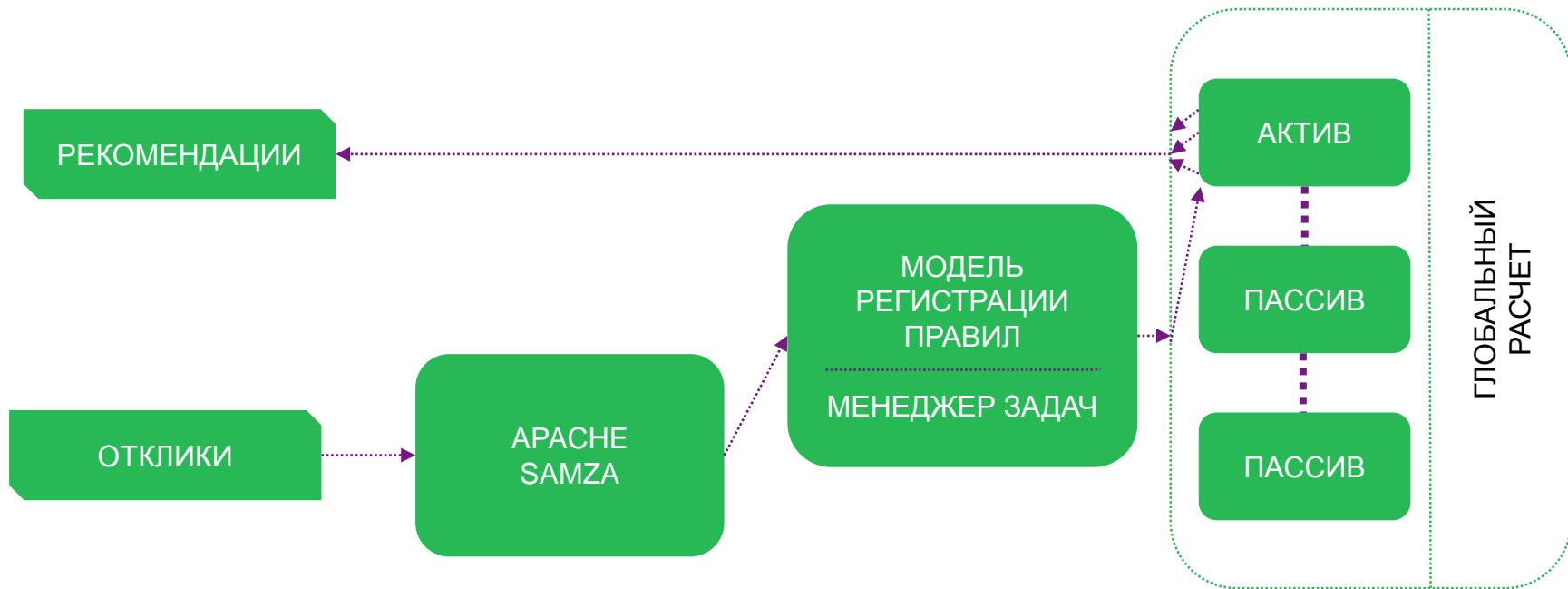


1. Появление объект: 0.1
2. Оценка момента появления: $(m_i)^k$
3. $n(2X.value - 1)$

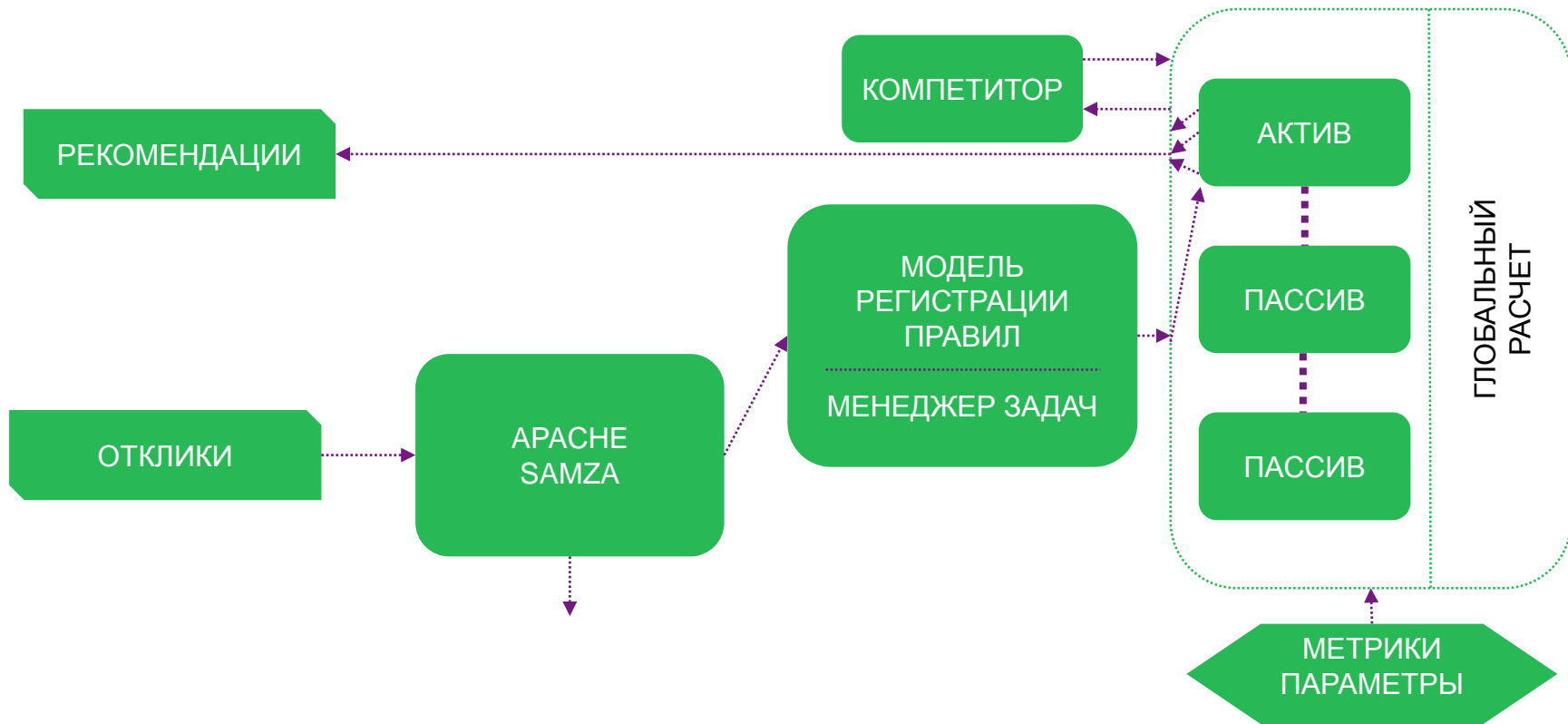
ПРОЦЕССЫ В СХЕМЕ



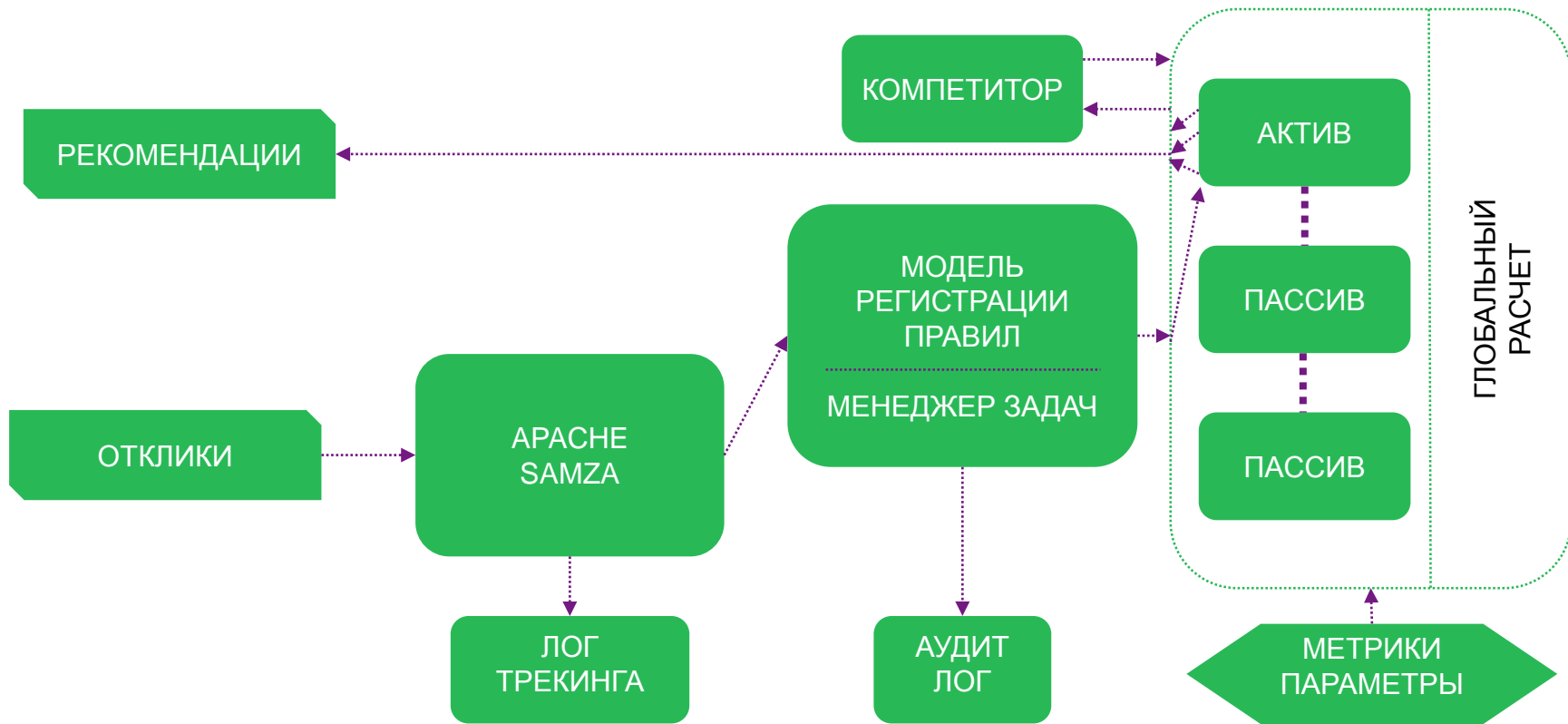
ПРОЦЕССЫ В СХЕМЕ



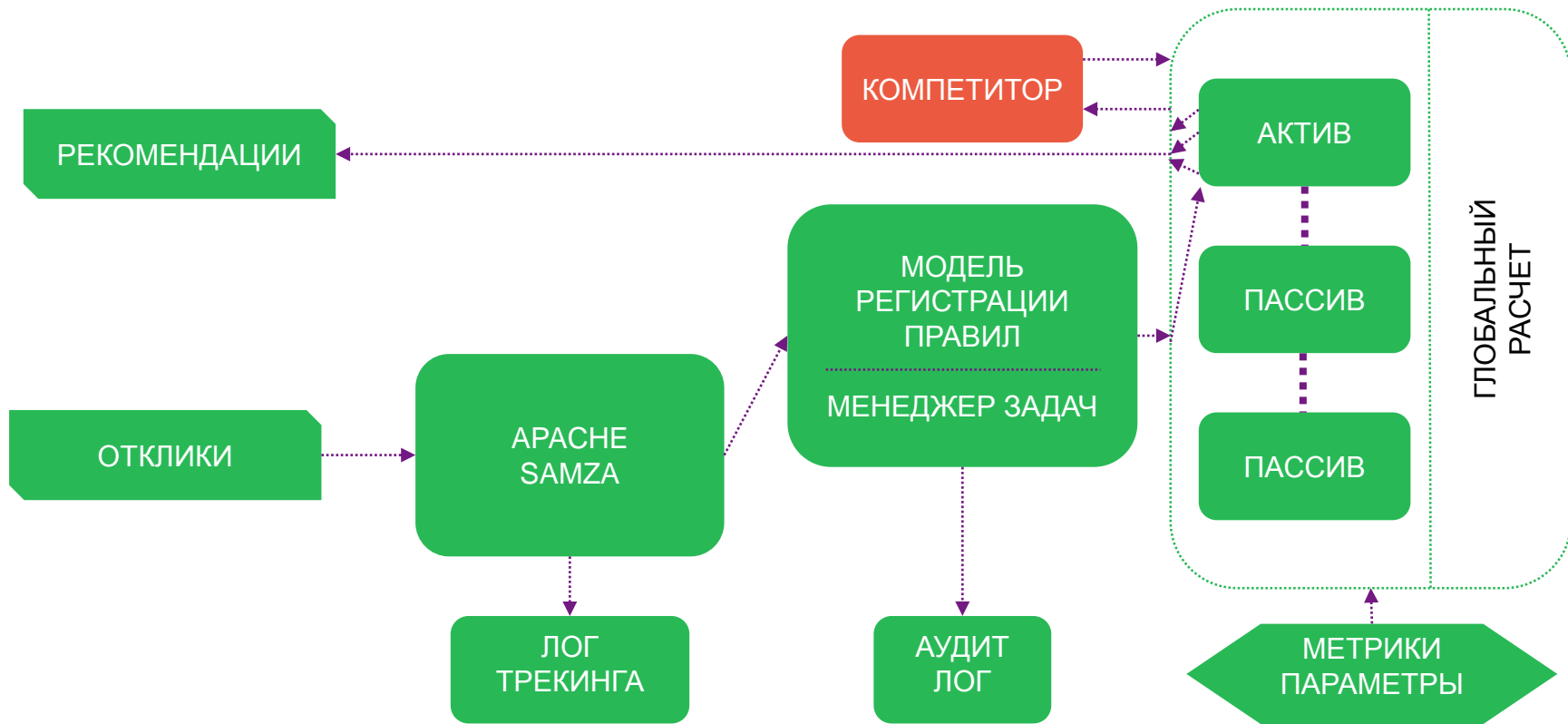
ПРОЦЕССЫ В СХЕМЕ



ПРОЦЕССЫ В СХЕМЕ



ПРОЦЕССЫ В СХЕМЕ



**Сравнить
результат**



Как сравнить модели и найти лучшую

1. Как сравниваются модели?

1. Построение таблицы по каждой модели

bind_type	model 1	...	model №
Bind	Bad % / Good %	:-:	Bad % / Good %

Как сравнить модели и найти лучшую

1. Как сравниваются модели?

1. Построение таблицы по каждой модели

bind_type	model 1	...	model №
до 30	32%	:-:	34%
35 – 40	16%	:-:	19%

Как сравнить модели и найти лучшую

1. Как сравниваются модели?

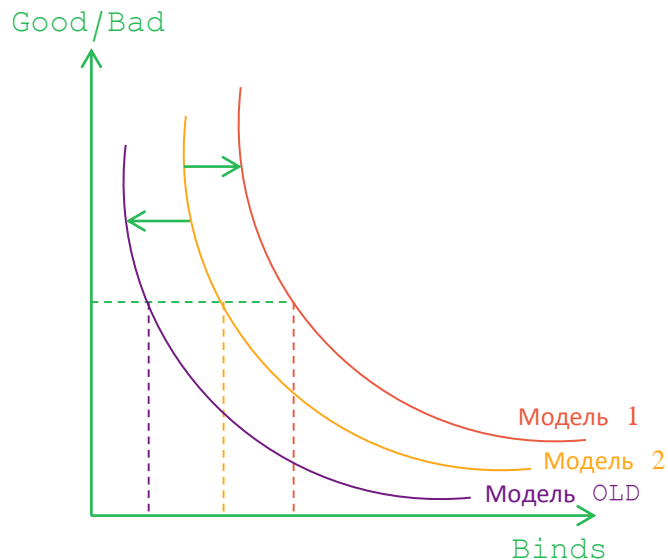
1. Построение таблицы по каждой модели

bind_type	model 1	...	model №
до 30	32%	⋮	34%
35 – 40	16%	⋮	19%

Как сравнить модели и найти лучшую

1. Как сравниваются модели?

1. Построение таблицы по каждой модели
2. График



Как сравнить модели и найти лучшую

1. Как сравниваются модели?

1. Построение таблицы по каждой модели
2. График
3. Дивергенция

$$DIV = \frac{(\overline{good} - \overline{bad})^2}{[0.5 * (\sigma_{good}^2 + \sigma_{bad}^2)]}$$

Как сравнить модели и найти лучшую

1. Как сравниваются модели?

1. Построение таблицы по каждой модели
2. График
3. Дивергенция

$$DIV = \frac{(\overline{good} - \overline{bad})^2}{[0.5 * (\sigma_{good}^2 + \sigma_{bad}^2)]}$$

0 -> UNLIM

Как сравнить модели и найти лучшую

1. Как сравниваются модели?

1. Построение таблицы по каждой модели
2. График
3. Дивергенция
4. Таблица выигрышей

	score group	количество в группе	факт.good	факт.bad	model good	model bad
Model Old						
Model 1						
.....						

Как сравнить модели и найти лучшую

1. Как сравниваются модели?

1. Построение таблицы по каждой модели
2. График
3. Дивергенция
4. Таблица выигрышей

	score group	количество в группе	факт.good	факт.bad	model good	model bad
Model Old	до 30	1280	950	330	870	410
Model 1	до 30	1280	950	330	844	350
.....						

Как сравнить модели и найти лучшую

1. Как сравниваются модели?

1. Построение таблицы по каждой модели
2. График
3. Дивергенция
4. Таблица выигрышей
5. Стабильность

	bind_type	actual %	expected %	actual - expected	actual / expected	ln(actual / expected)	index
Model Old							
Model 1							
.....							

index рассчитывается по формуле:

$$index = \frac{\sum(actual\% - expected\%)}{\ln(actual\%/expected\%)}$$

Как сравнить модели и найти лучшую

1. Как сравниваются модели?

1. Построение таблицы по каждой модели
2. График
3. Дивергенция
4. Таблица выигрышей
5. Стабильность



Как сравнить модели и найти лучшую

1. Как сравниваются модели?

1. Построение таблицы по каждой модели

score	expected good % (model old)	% model 1 / expected good % (model old)	..

2. График

3. Дивергенция

4. Таблица выигрышей

5. Стабильность

6. Групповой отчет

Как сравнить модели и найти лучшую

1. Как сравниваются модели?
2. Цена ошибки!

НАШЛИ ИЛИ ПОТЕРЯЛИ?

1. Уменьшение времени реакции
2. Активность рекомендаций
3. Шаг в сторону универсальной платформы рекомендаций

НАШЛИ ИЛИ ПОТЕРЯЛИ?

1. Уменьшение времени реакции
2. Активность рекомендаций
3. Шаг в сторону универсальной платформы рекомендаций

1. Нет Python в процессе работы с данными
(Apache Samza работает с Java | Scala)

Рекомендации начинаются с тебя



Артем Селезнев, Data Scientist

@SeleznevArtem

