

Data driven'ность — как оцифровать пользу аналитики для бизнеса

СберМаркет

🛒 tech



Леонтьев Иван

И.О. Директора по аналитике СберМаркета



@Ivan_Leontev



in/ivan-leontev-serg

ИЗ ЭТОГО ДОКЛАДА ВЫ УЗНАЕТЕ ...

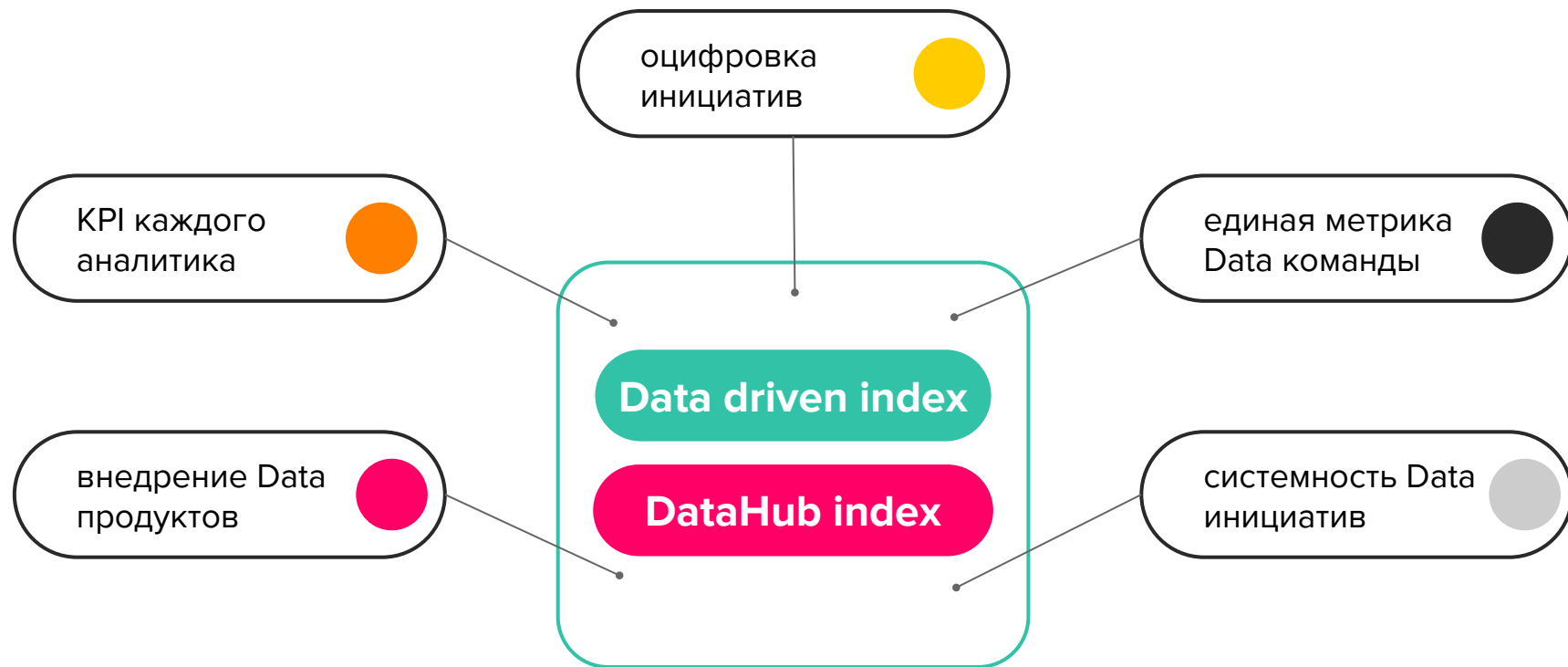
- Ценность data-driven подхода к data культуре

- Как мы внедряли data-driven подход?

- Стоило ли оно того?

- Как применить такое у себя?

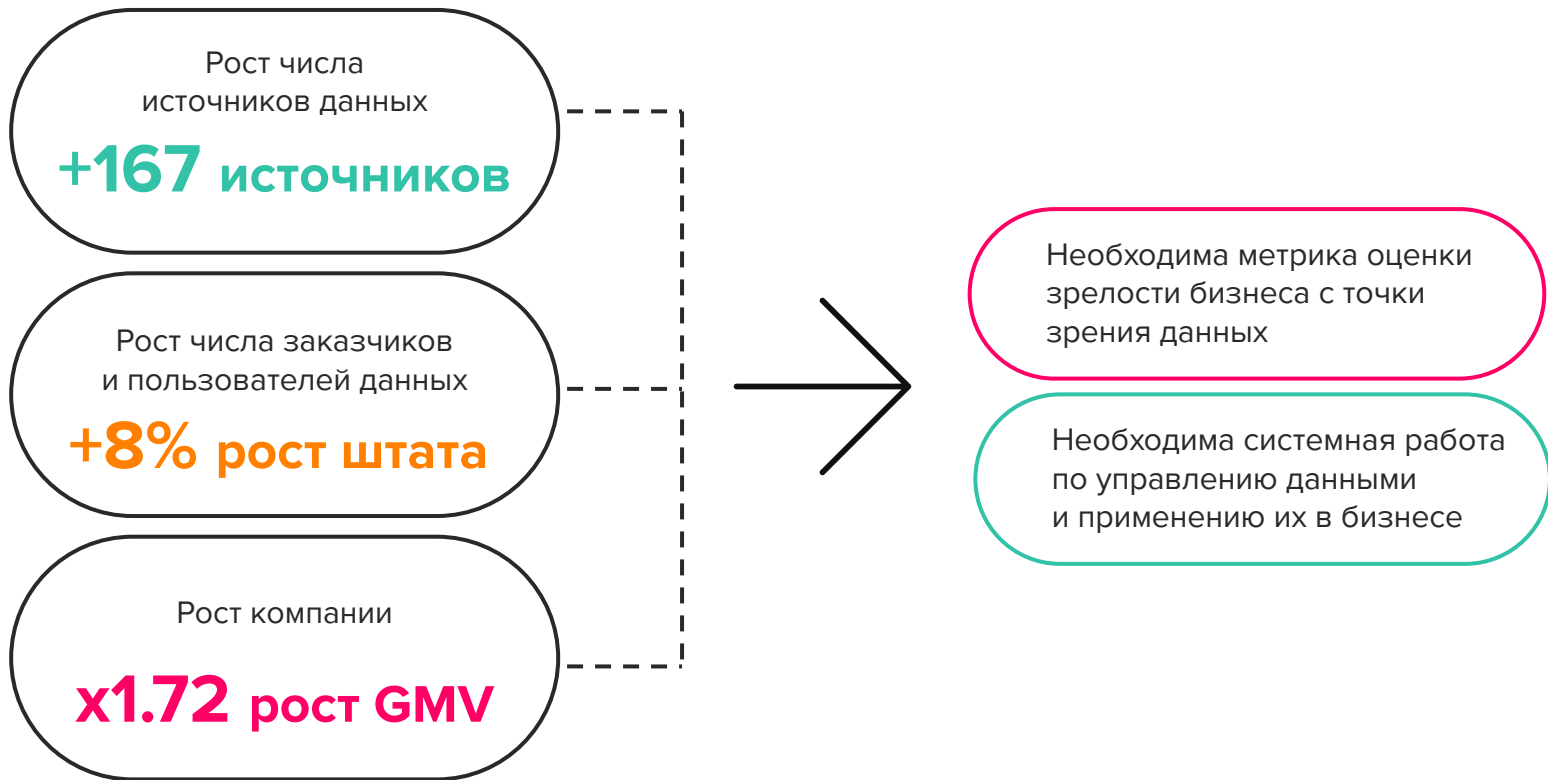
ценность data driven подхода к дата культуре



почему бы не использовать например ФОТ / GMV?

- Аналитики не влияют напрямую на ФОТ и тем более на весь GMV
- ФОТ / GMV не дает оценки зрелости дата культуры
- Метрика не actionable
- Команды без прямых инкрементов по такой метрике не имеют добавленной ценности

как возник запрос на data driven'ность?



как мы оцифровали data driven'ность?

Data driven index **72%**

Индекс
использования
данных **62%**



- Учет инструментов (BI системы, Grafana)
- Разный вес для разных ролей, учет количества дашбордов
- Разный вес для дашбордов (сертификация)

Индекс
качества
данных **58%**



- Количество использования ключевых витрин в дашбордах
- Качество расчета ключевых витрин и их источников
- Доля багов / количество событий в событийных данных

Индекс
стабильности
данных **96%**



- Стабильность ключевых витрин и их расчета
- SLA обновления витрин 8 утра
- Разный вес витрин в зависимости от их важности для бизнеса (70% vs 30% события)

как посчитать использование данных?

1

Индекс использования
данных

2

Индекс качества
данных

3

Индекс стабильности
данных

Сотрудники

Различный вес
в зависимости от должности

- CEO = **16**
- CEO-1 = **8**
- CEO-2 = **4**
- CEO-3 = **2**
- Специалист = **1**

Инструменты

Весы инструментов
зависят от их качества

- Сертифицированный
дашборд Tableau или Grafana
для ДИТ = **1**
- Tableau без сертификации
или Grafana не для ДИТ = **0.7**
- Дашборд в Metabase = **0.5**
- Карточка в Metabase = **0.3**

Частота использования

Повышенная активность
пользователя влияет
на его вес

- Сертифицированный
дашборд в Tableau **+100%**
- Дашборд Tableau
без сертификации **+50%**
- Дашборд в Metabase **+25%**
- Карточка в Metabase **+10%**

как посчитать качество данных?

1

Индекс использования
данных

2

Индекс качества
данных

3

Индекс стабильности
данных

Качество
бизнес
данных

70%

Качество
событийных
данных

30%

- За основу взят golden set витрин которые планируются к переезду в GreenPlum
- Витринам проставляются баллы
 - **1** — реплицирована в CH и есть select
 - **0,9** — реплицирована в CH, не используется
 - **0,7** — есть в CH.prod, расчет через DWH airflow
 - **0,4** — есть в CH.prod, расчет через analytics airflow
 - **0** — sandbox
- Итоговый результат делится на кол-во витрин

- Исходные данные — заведенные в Jira баги
- Итоговый результат = $1 - \frac{\text{кол-во багов}}{\text{кол-во уникальных пар событие+платформа}}$

как посчитать стабильность дашбордов?

- 1 Индекс использования данных
- 2 Индекс качества данных
- 3 **Индекс стабильности данных**

- За основу берется сет ключевых витрин для обновления Tableau
- Вес каждой витрины - кол-во дашбордов зависящих от нее
- Если витрина обновилась до SLA - 1 балл, опоздание ≤ 10 мин - 0,5 баллов
- $\text{Стабильность} = \text{сумма вес } \times \text{балл} / \text{общий вес витрин}$

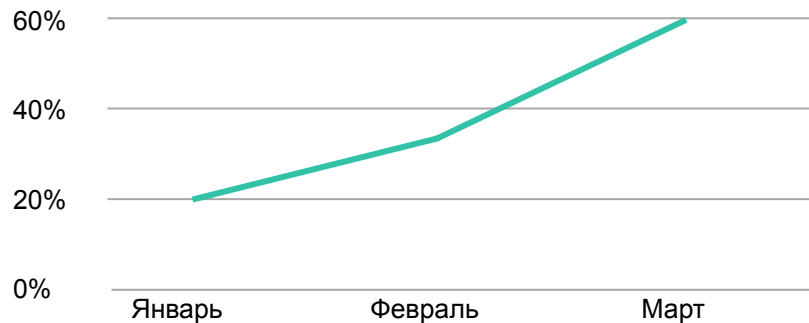
в чем польза data driven'ности для бизнеса?



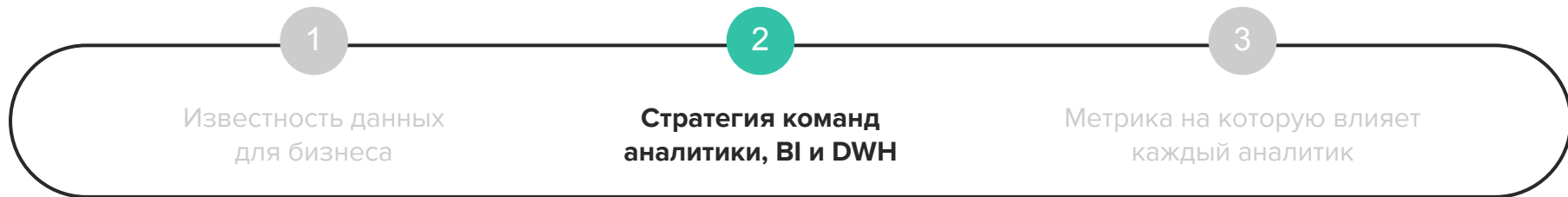
x1.8

возросло использование данных за квартал за счет обучения по BI системе и ознакомления с ключевыми витринами и кубами данных

это снизило долю Adhoc задач на 15%



в чем польза data driven'ности для бизнеса?



Использование данных

- Популяризация дашбордов и обучение сотрудников
- Погружение в потребности бизнеса. Например, мониторы и дашборды для команд и self-service

Стабильность данных

- Выстраивание системы алертингов на всех этапах ELT
- Внедрение стандартов кода

Качество данных

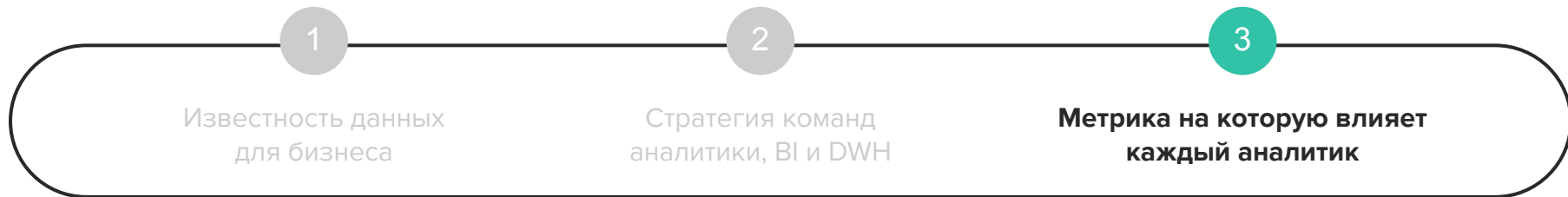
- Перенос расчёта ключевых витрин в GreenPlum
- Сверка данных в процессе преобразований

в чем польза data driven'ности для бизнеса?

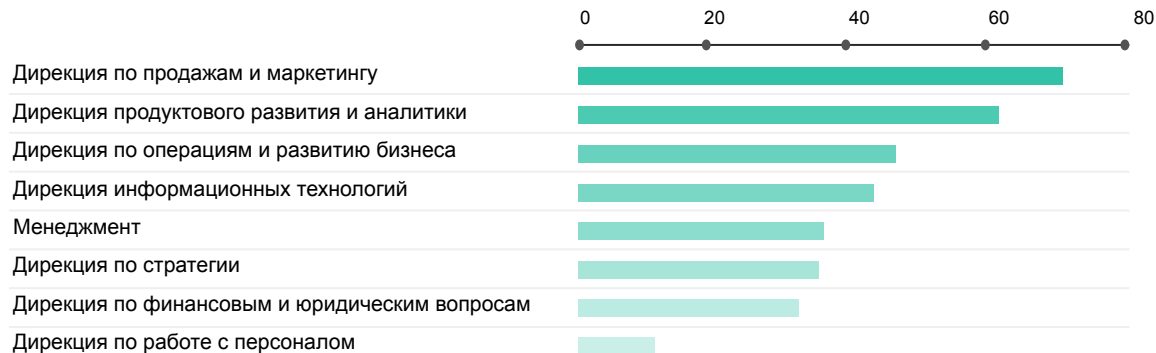


На эту метрику влияет каждый аналитик. Она может быть одним из OKR, который напрямую зависит от сделанных дашбордов/витрин

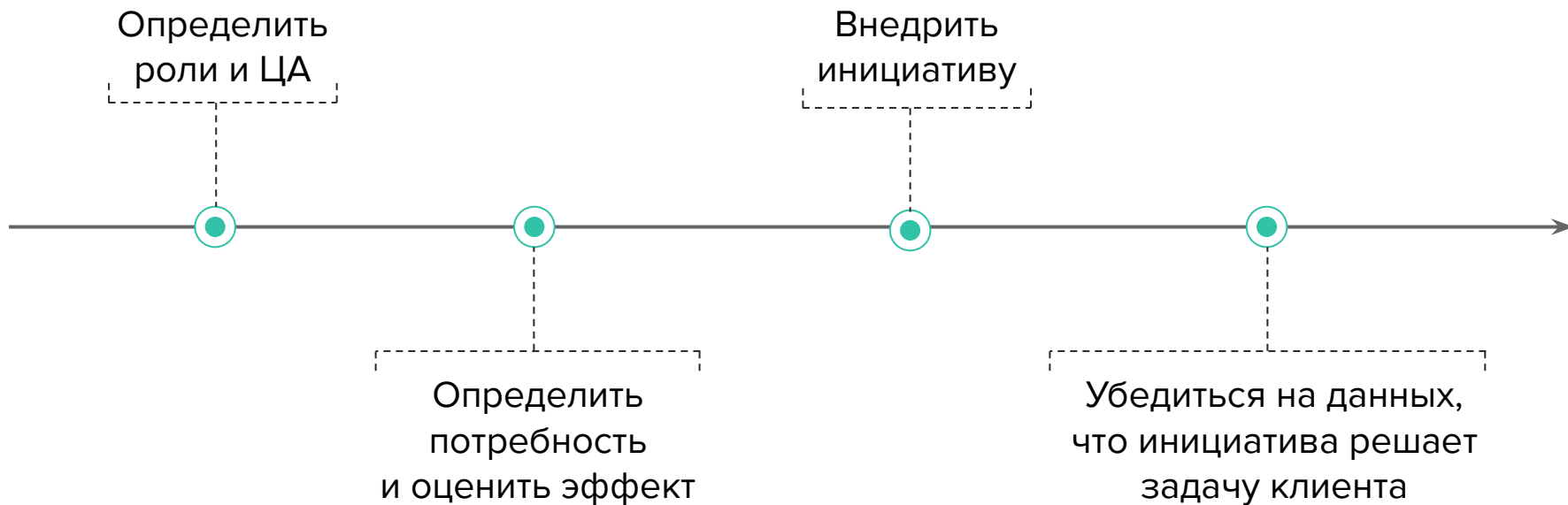
в чем польза data driven'ности для бизнеса?



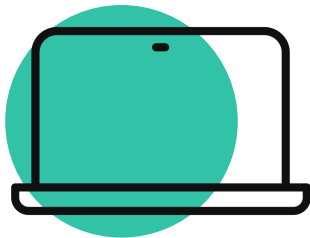
На эту метрику влияет каждый аналитик. Она может быть одним из OKR, который напрямую зависит от сделанных дашбордов/витрин



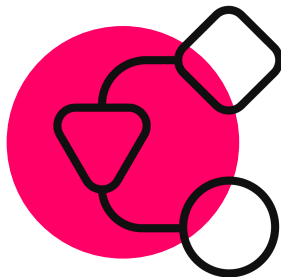
как повысить data driven'ность конкретного отдела?



когда стоит применять такой подход?



Есть BI инструмент
с возможностью
мониторинга
посещаемости

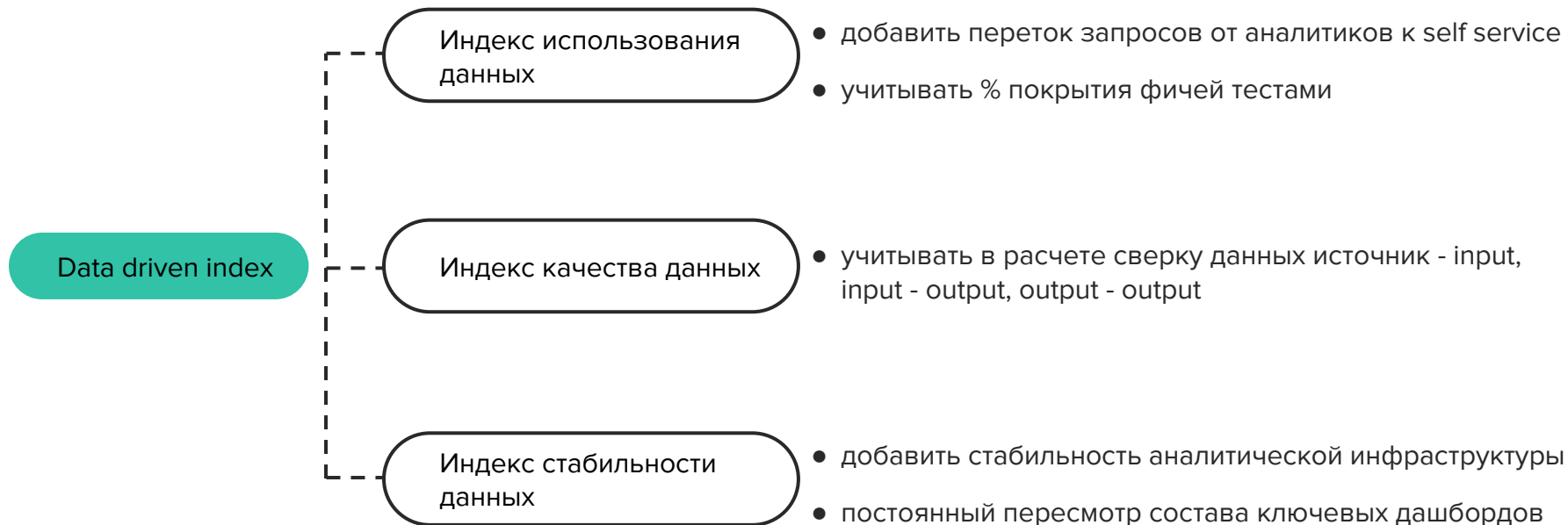


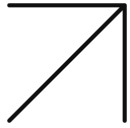
Есть регулярные
дашборды / важные
процессы, основанные
на данных



В компании большое
число сотрудников
на одного аналитика
(например,
50 в СберМаркете)

вектор развития индекса data driven'ности





все это о пользе данных
для бизнеса, а что о пользе
данных для аналитики?

что такое метаданные?

Метаданные — данные о данных,
свойства и признаки данных

Дата-каталог



Метаданные

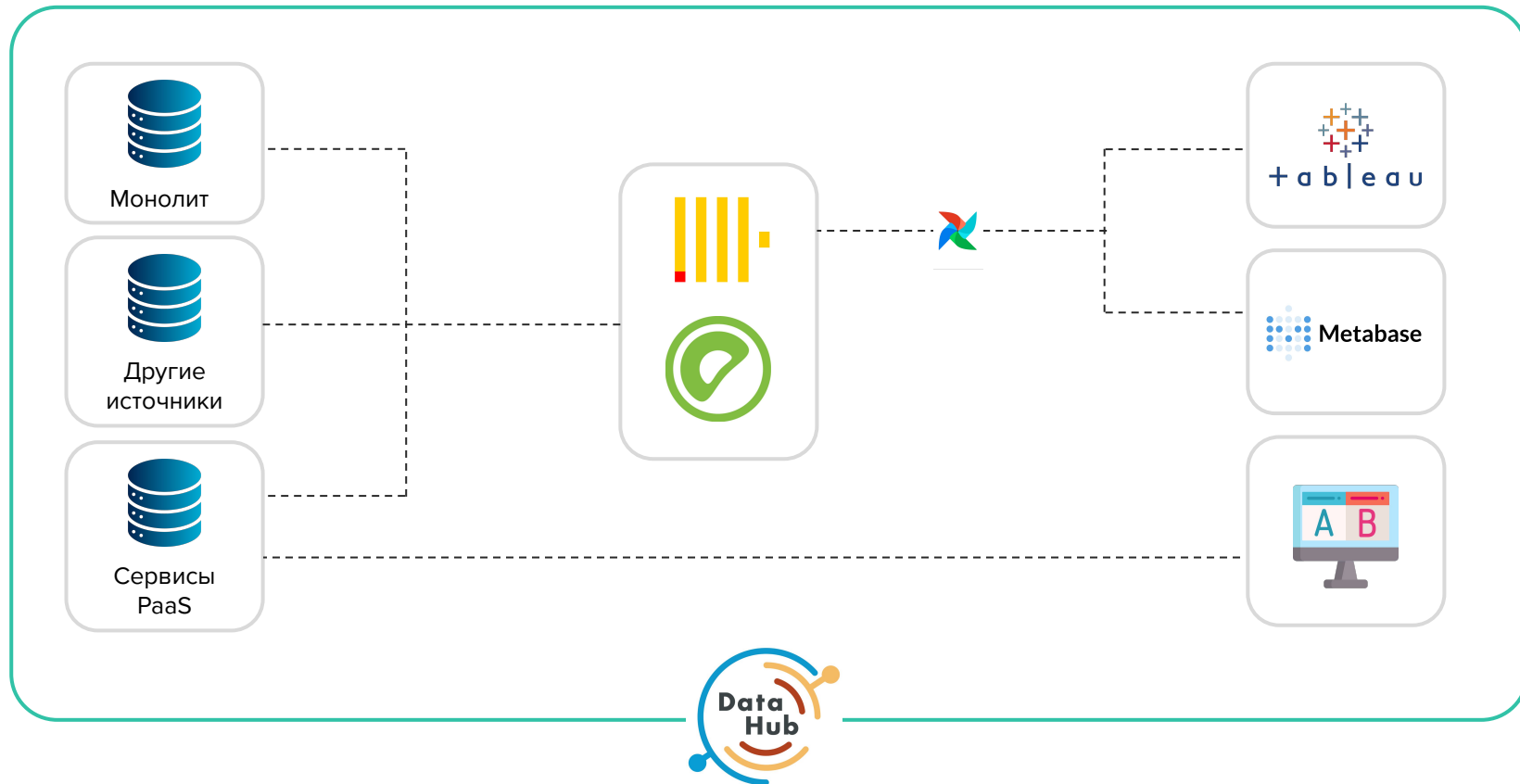


Поиск и разметка



Data lineage

архитектура данных и дата каталог в ней



логика расчета индекса дата-каталога

DataHub index **33%**

Индекс
заполняемости **11%**

Индекс
персональных
данных **14%**

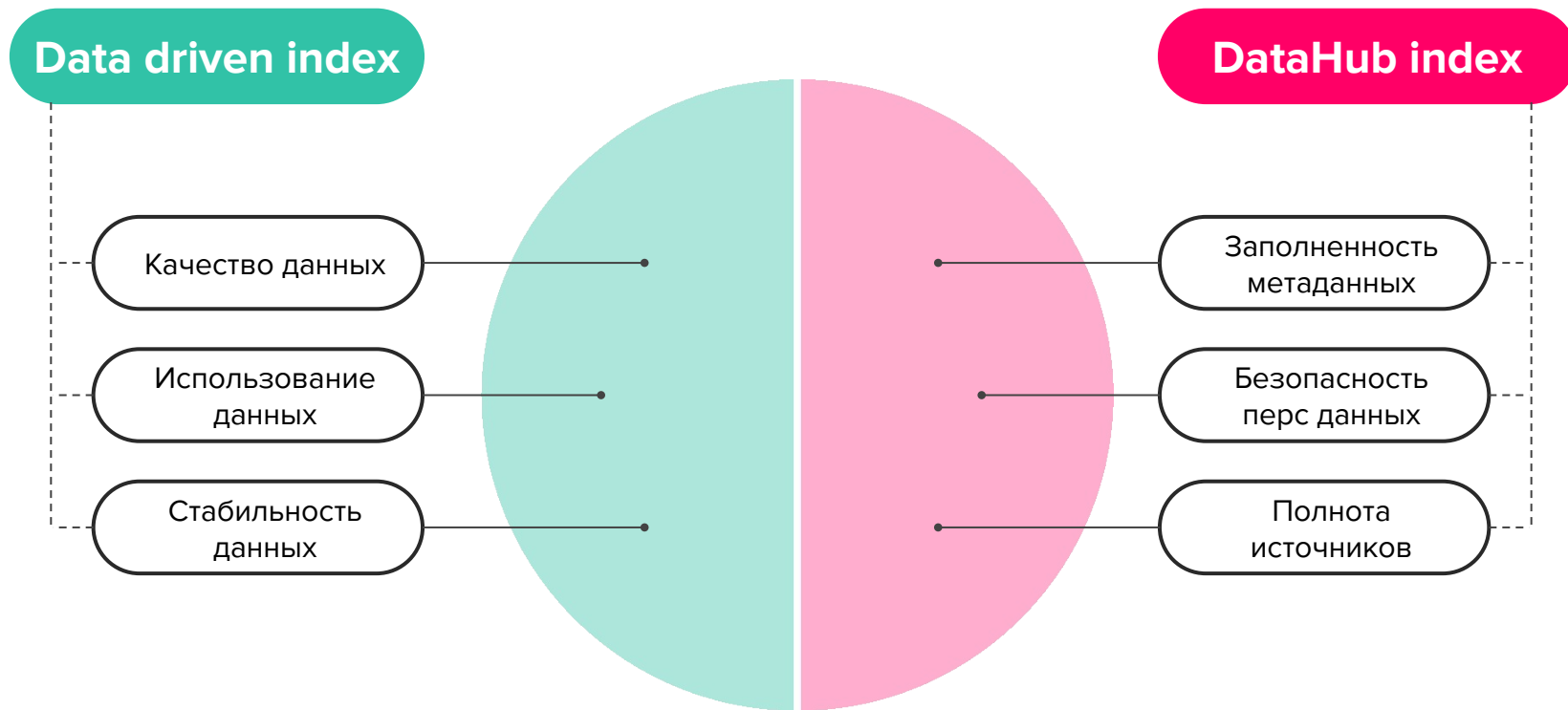
Доля
добавленных
DB **75%**

	Tier 1	Tier 2	Tier 3	Tier 4
Проставлен Tier	1	1	1	1
Домен	1	1	1	1
Владелец	1	1	1	1
Документация	1	1	1	0
Описание полей	1	1	0	0
Ссылка на gitlab	1	1	0	0
Запросы (пример запроса к таблице)	1	0	0	0
Lineage	1	0	0	0
Веса	5	4	3	2

- % просмотренных и протегированных ИБ таблиц на наличие персональных данных

- Доля подключенных источников к DataHub от планового числа

so what



как это влияет на бизнес?



- Усиление контроля за стабильностью в команде DWH
- Снижение доли Adhoc задач на 15%
- Рост доли положительных отзывов от бизнес заказчиков



- Снижение времени аналитиков на выяснение данных
- Повышение безопасности данных

с чего начать внедрение такого подхода?



Определить цели внедрения метрики в команде



Сделать первую итерацию расчета и начать применять



Кастомизировать метрику под потребности компании

keep in touch



Леонтьев Иван

И.О. Директора по аналитике
СберМаркет

