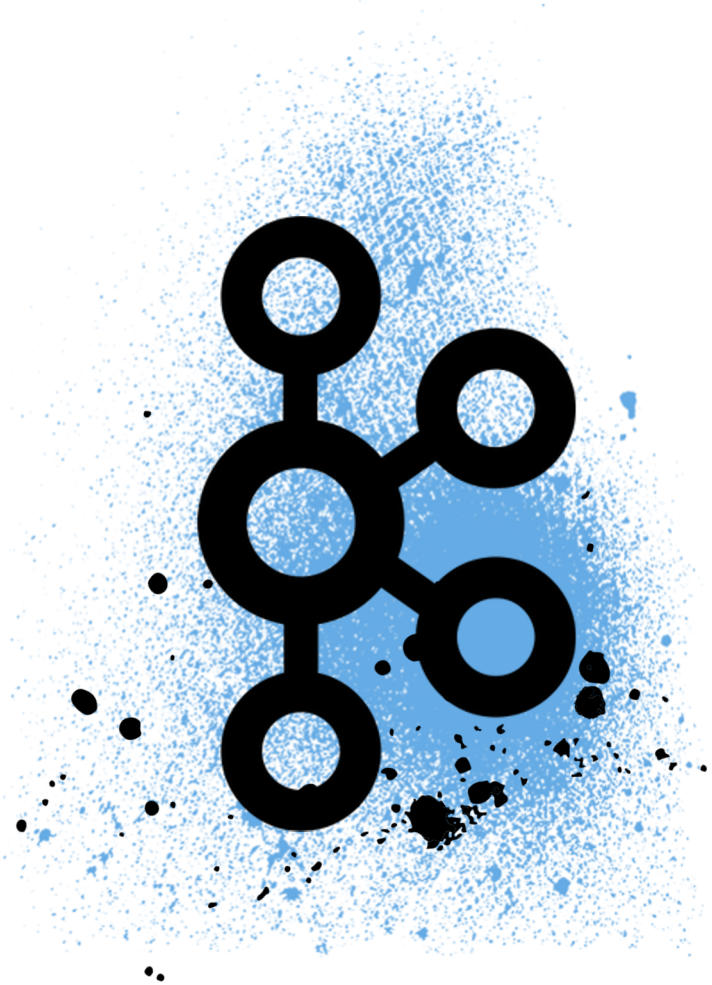
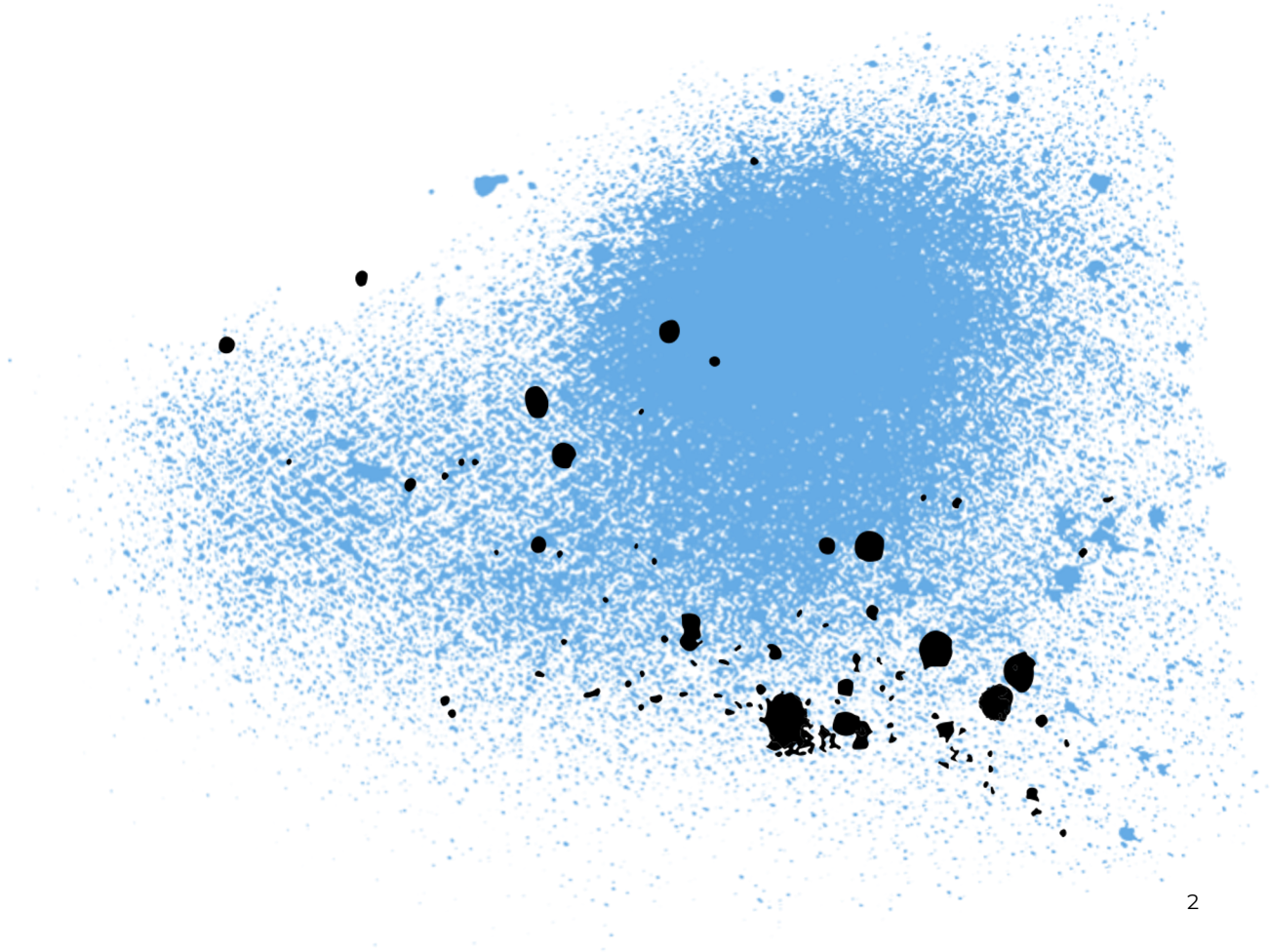
avito.tech CONFLUENT

Apache Kafka и инфраструктурные сателлиты



DevOops Conf Online
2020

Привет!



Анатолий Солдатов

Engineer, Avito.ru

- https://t.me/MrAnatoly
- https://avito.tech/

Виктор Гамов

Developer Advocate, Confluent

- https://t.me/gAmUssA
- https://t.me/proKafka

Как все начиналось

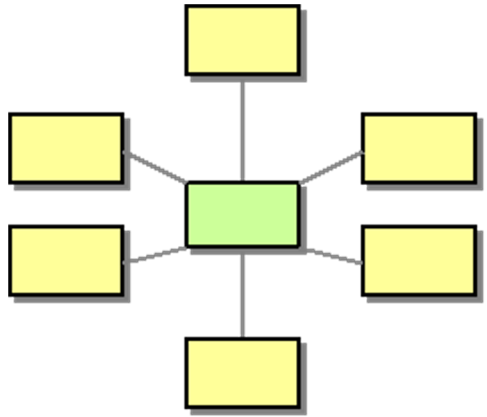
Как все начиналось

Проблема

- Сотни сервисов
- Синхронное взаимодействие
- Ненадежно

Решение

- Асинхронное взаимодействие
- Message broker pattern
- Apache Kafka



Подробнее про message broker в Авито – https://habr.com/ru/company/avito/blog/465315/

И как сейчас

100+
серверов

25
Gbit/sec

800k
events/sec

А как там дела у буржуев?

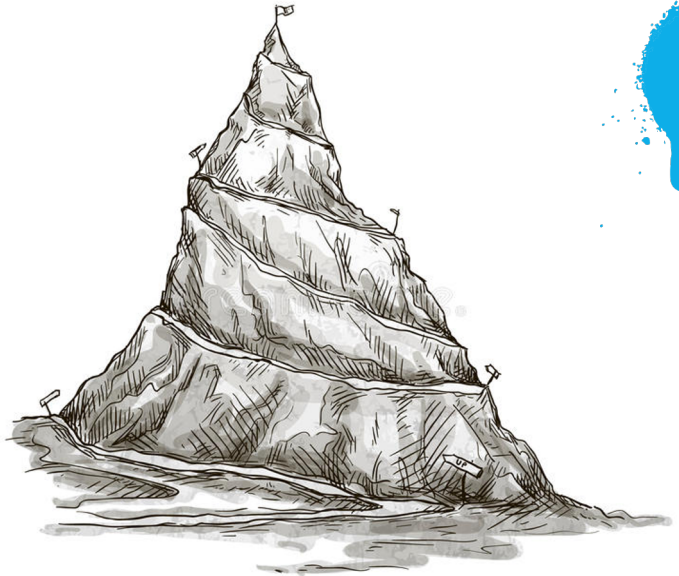
4500+

Apache Kafka® кластеров

450%

рост прибыли от клауда

План

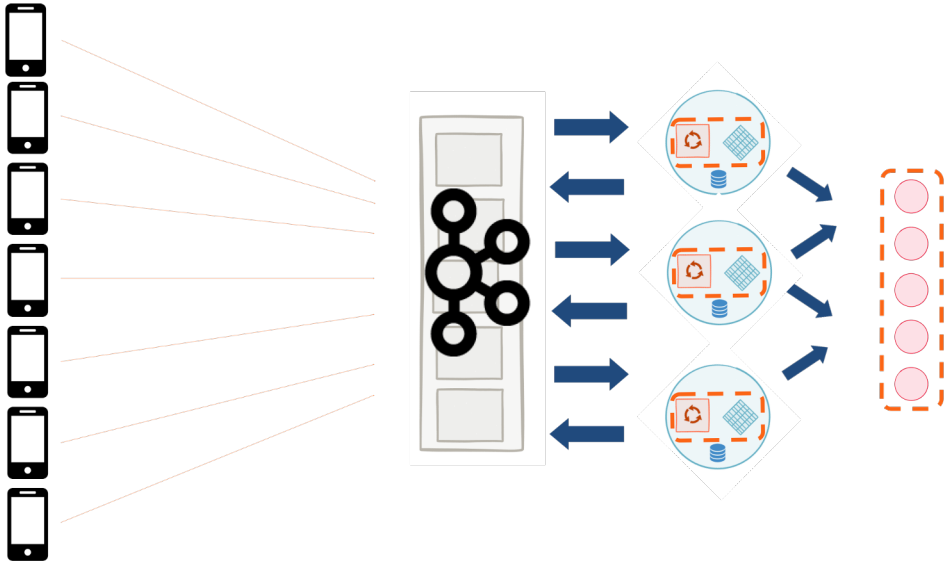


План

- Kafka 101
- Установка
- Мониторинг
- Бэкапы
- Schema management
- Connectors
- Proxy
- Multi-DC
- Автоматизация

Apache Kafka 101

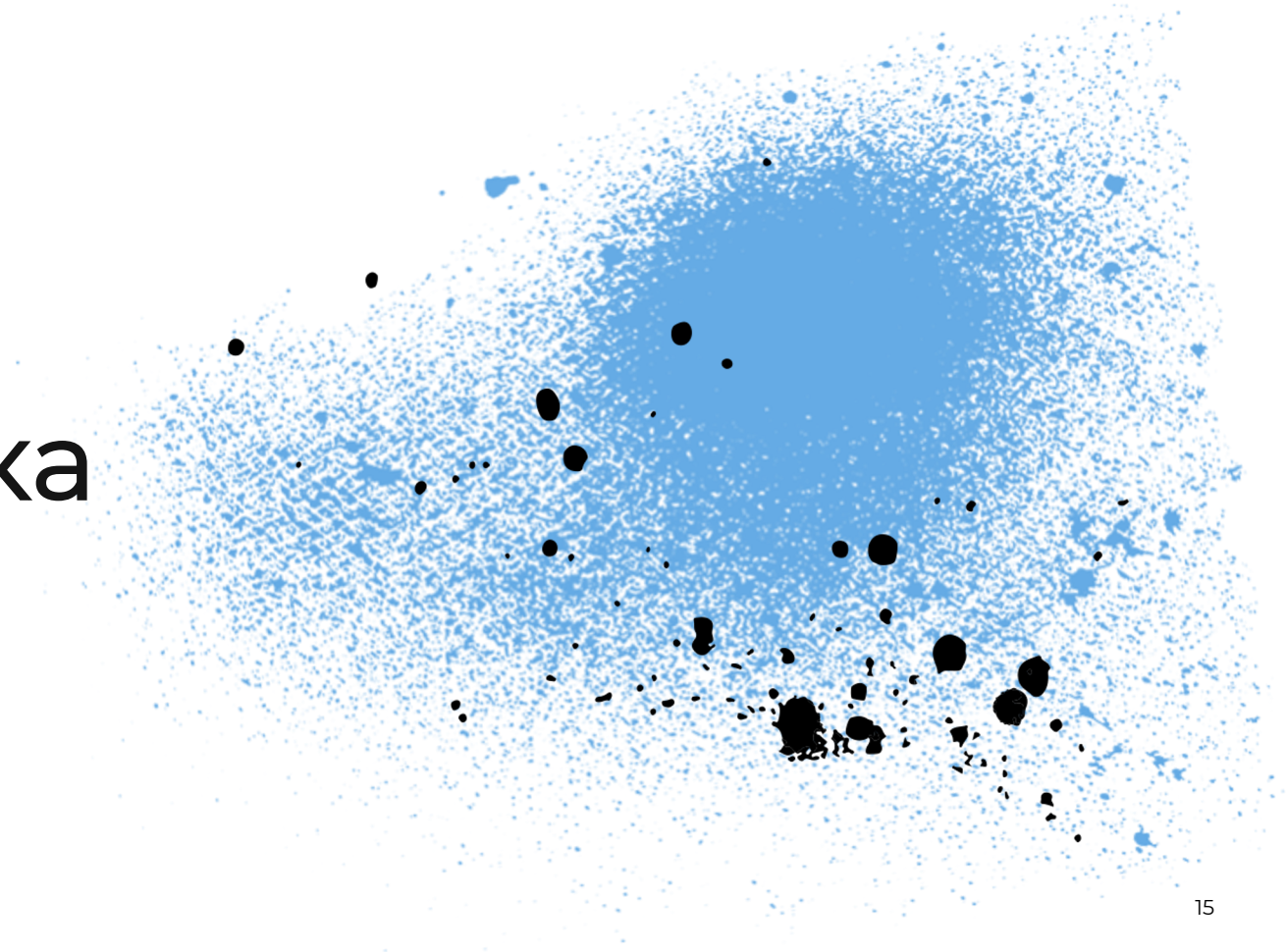
Apache Kafka 101



Вы находитесь здесь

- Kafka 101
- **Установка**
- Мониторинг
- Бэкапы
- Schema management
- Connectors
- Proxy
- Multi-DC
- Автоматизация

Установка



Установка

Community

- Apache Kafka
- Confluent Platform Community
- Cloudera Free

Enterprise

- Confluent Enterprise
- Cloudera Platform Enterprise
- Arenadata
- И другие

Установка

Облака

- TODO
- TODO
- TODO

Железо

- TODO
- TODO
- TODO

Установка

Заранее продумать

- Требования к железу (http://kafka.new)
- Файловая система
- Версия Kafka
- Конфигурация кластера
- Топология кластера

Точка входа – https://docs.confluent.io/platform/current/kafka/deployment.html

Установка

Коммунальный ZK

- Один для всех

Выделенный ZK

- Один под кластер Kafka

Установка

А точно ли все работает?

Вы находитесь здесь

- Kafka 101
- Установка
- **Мониторинг**
- Бэкапы
- Schema management
- Connectors
- Proxy
- Multi-DC
- Автоматизация

Мониторинг

Мониторинг

Сложности

- Много частей
- Мониторинг нужен и для инструментов вокруг (например, Connectors)
- DDoS метриками

Мониторинг

Метрики, требующие особого внимания

- Min ISR
- Under-replicated partitions
- Метрики сети, железа
- Consumer lag

Мониторинг

Уровни

- SLO мониторинг
- Bird’s-eye view
- Cluster view
- Broker view

Мониторинг

SLO мониторинг

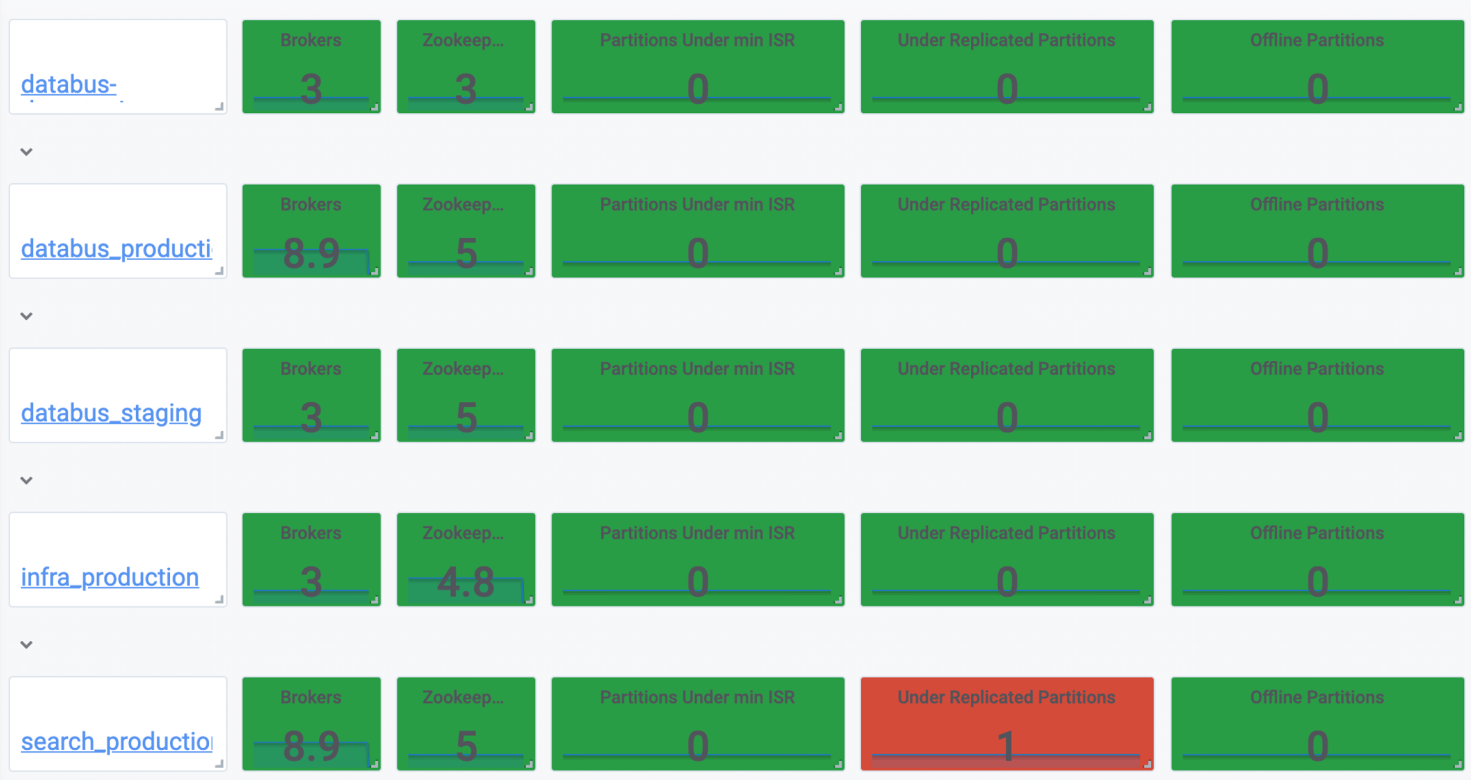
- Доступность на запись и чтение
- End to End latency < 100ms
- Error budgets

Мониторинг

Bird’s-eye view

- Число доступных брокеров и zk к общему числу брокеров
- min ISR
- Under-replicated partitions
- Offline partitions

Мониторинг bird’s-eye view

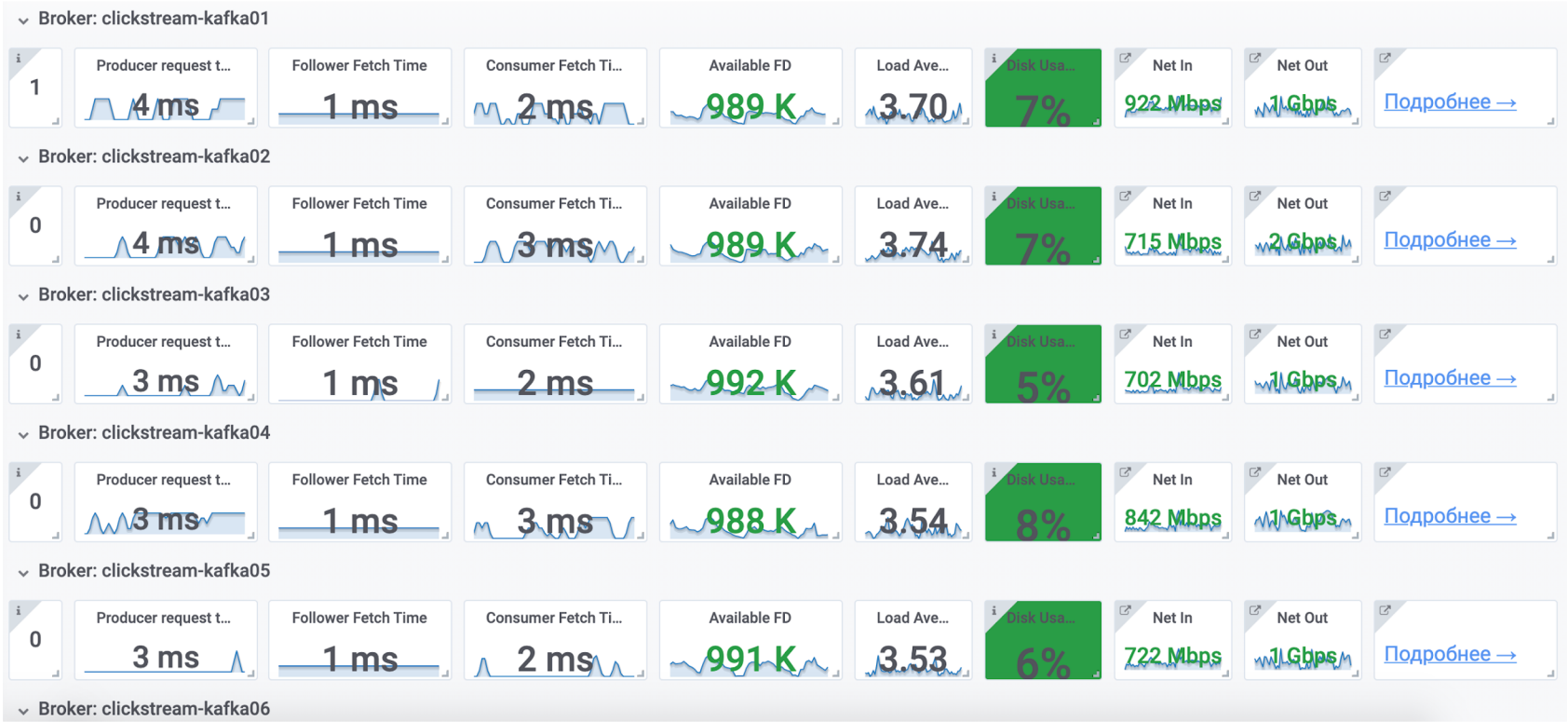


Мониторинг

Cluster view

- Bytes In/Out
- Controller
- Network, Disk, CPU
- Consumer lag

Мониторинг/cluster view

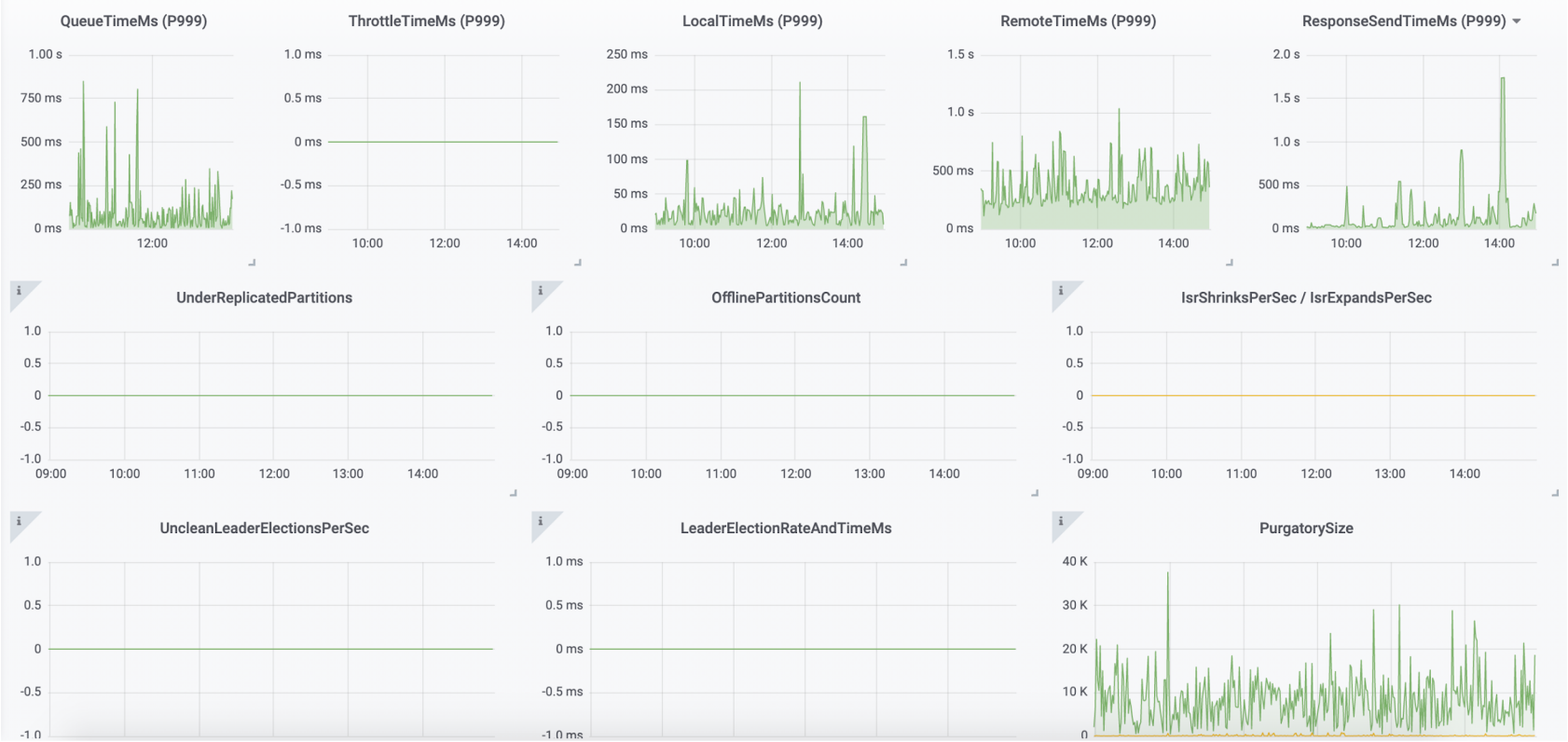


Мониторинг

Broker view

- Total time
- Network processor avg idle time
- Queue time, Remote time
- Purgatory size
- Leader election time

Мониторинг/broker view

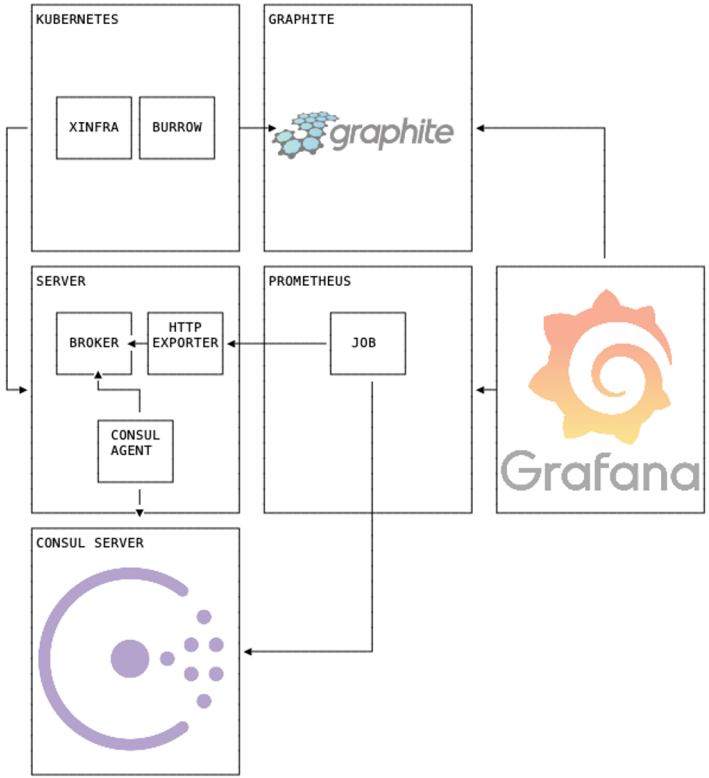


Мониторинг

Инструменты

- JMX to Prometheus exporter
- Prometheus, Graphite
- Consul
- Burrow/Consumer Freshness Tracker
- Xinfra-monitor
- Grafana

Мониторинг


KUBERNETES
XINFRA
BURROW
GRAPHITE
graphite
SERVER
BROKER
HTTP EXPORTER
CONSUL AGENT
PROMETHEUS
JOB
Grafana
CONSUL SERVER

Мониторинг

Альтернативные инструменты

- Confluent Control Center (Enterprise)
- Datadog Kafka Dashboard (Enterprise)
- Lenses (Enterprise)
- yahoo/CMAK
- И другие

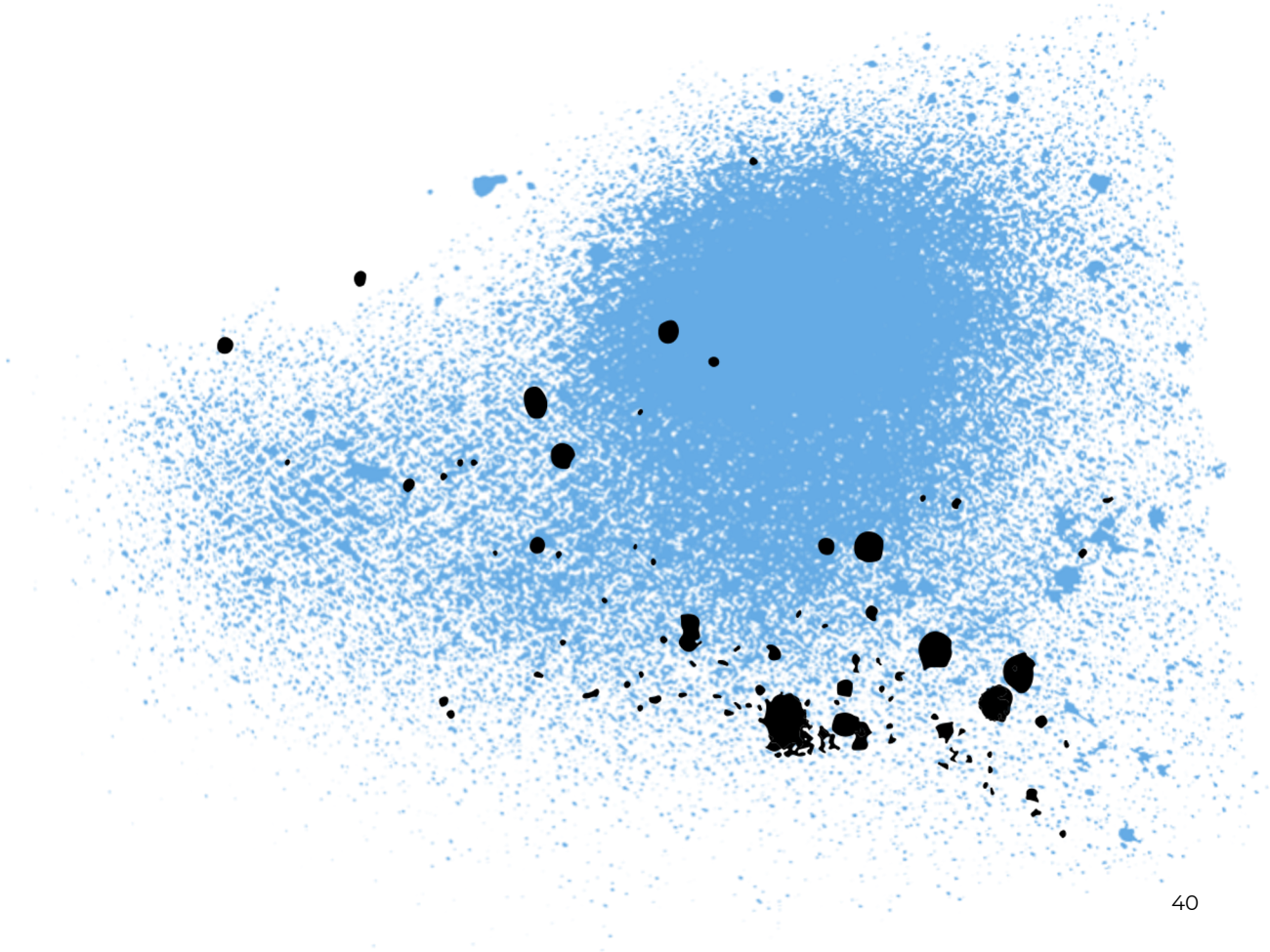
Мониторинг

Когда все идет не по плану

Вы находитесь здесь

- Kafka 101
- Установка
- Мониторинг
- **Бэкапы**
- Schema management
- Connectors
- Proxy
- Multi-DC
- Автоматизация

Бэкапы



Бэкапы

Kafka

- Мы не делаем бэкапы
- Держим сервера в разных ДЦ и стойках
- Используем фактор репликации 3
- Используем rack awareness
- Имеем команду 24x7

Бэкапы

ZooKeeper

- Бэкапы не нужны, KIP-500

KIP-500: Replace ZooKeeper with a Self-Managed Metadata Quorum – https://cwiki.apache.org/confluence/display/KAFKA/KIP-500%3A+Replace+ZooKeeper+with+a+Self-Managed+Metadata+Quorum

Бэкапы

ZooKeeper

- Бэкапы нужны
- Бэкапы дешевые
- Данные в ZK очень важны
- Снимать бэкапы, например, с помощью burry.sh
- Обязательно валидировать бэкапы

Бэкапы

Анархия – мать порядка

Вы находитесь здесь

- Kafka 101
- Установка
- Мониторинг
- Бэкапы
- **Schema management**
- Connectors
- Proxy
- Multi-DC
- Автоматизация

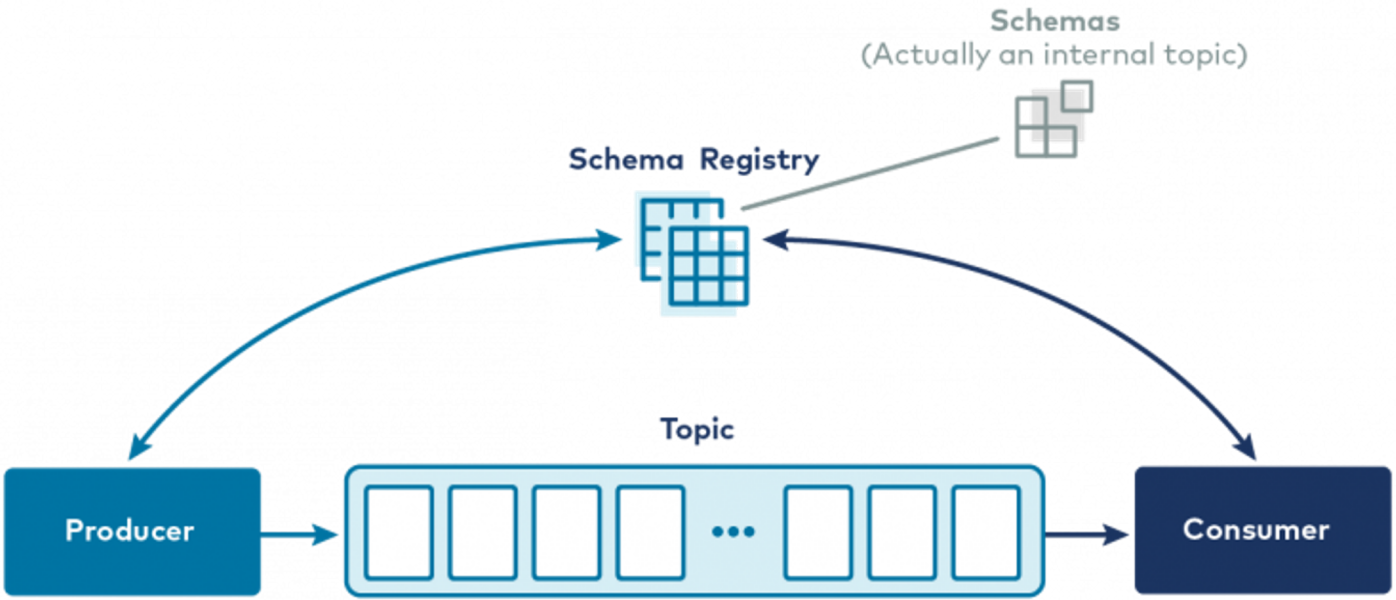
Schema management

Schema management

Schema registry 101

- TODO
- TODO
- TODO

Schema management



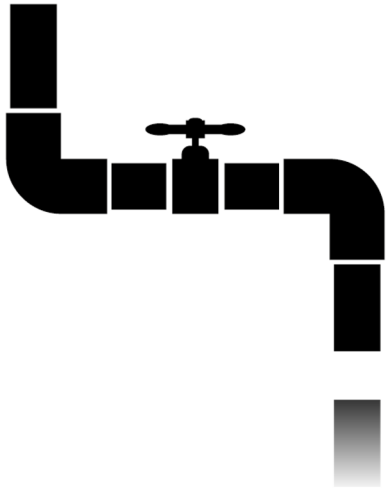
Schema management

Альтернативные технологии

- Red Hat Integration
- Azure Schema Registry
- Самодельные решения

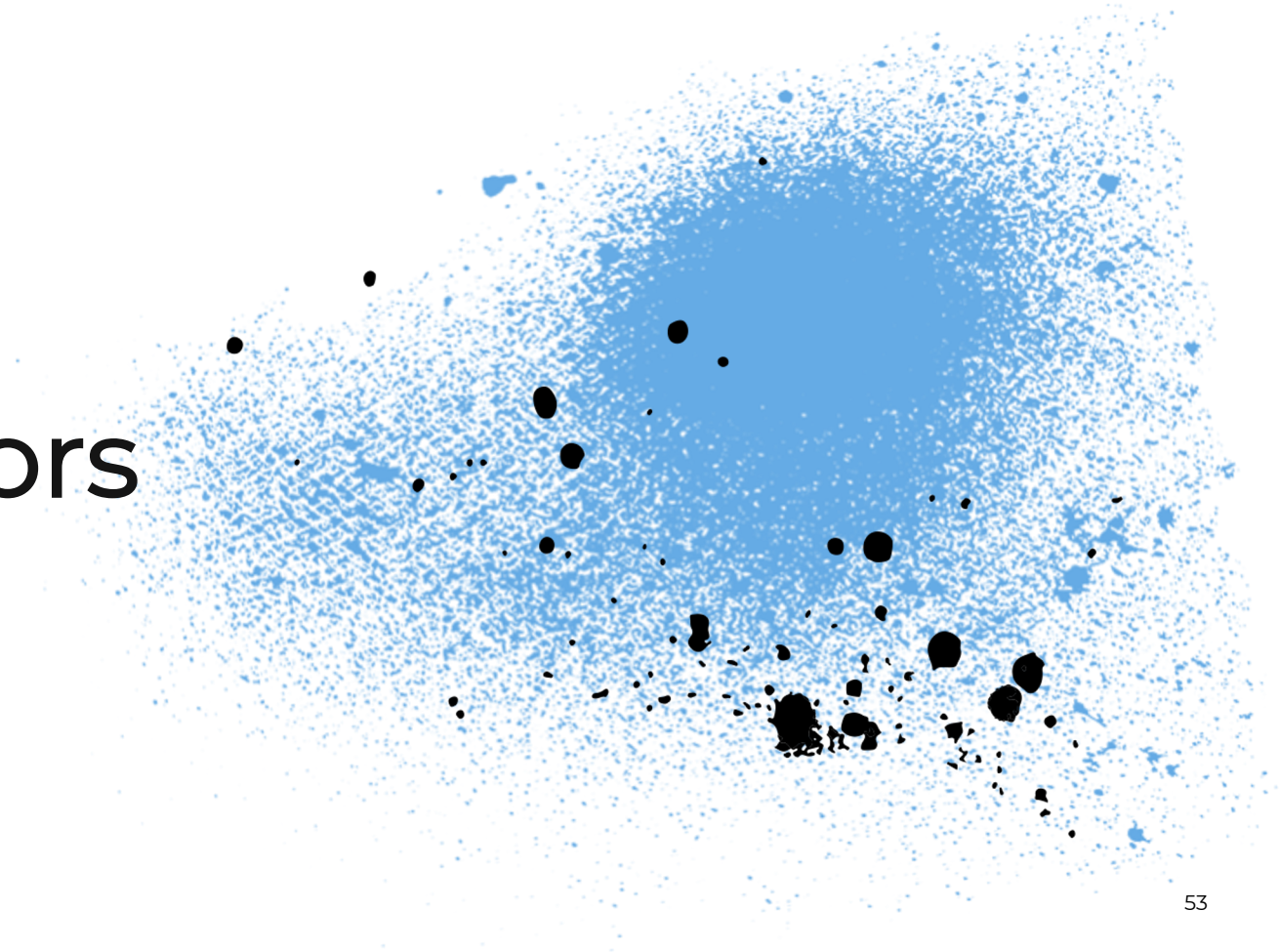
Подробнее про brief-схемы в Авито –
https://www.youtube.com/watch?v=VjMloZzEq2A&ab_channel=AvitoTech

Schema management



Из кранов с шумом потекла вода (Булгаков)

Connectors



Connectors


Data Source
Kafka Connect
Kafka
Cluster
Kafka Connect
Data Sink

Connectors

Connectors 101

- Source / Sink
- SMT
- Converters

Connectors

Connector hub

- confluent.io/hub
- Lenses connectors

Connectors

Connector deployment & management

- TODO
- TODO

Connectors

Самодельный PostgreSQL -> Kafka source

- Демон на go, паттерн transactional outbox
- Т.к. Debezium для PostgreSQL имеет нюансы

Подробнее про Debezium vs самописный коннектор в Авито – https://www.youtube.com/watch?v=w4w7J4acNo0&ab_channel=AvitoTech

Connectors

Толку от вашей Кафки (разработчик на Perl)

Proxy & Gateway

Proxy

Зачем нужен proxy

- Абстрагирование от технологии
- Централизованное управление конфигурацией
- Интеграция языков со слабой поддержкой Kafka

Proxy

Confluent REST Proxy

- Community licensed
- Admin features
- Produce / consume messages

Proxy

Liiklus

- Schema - cloudevents.io
- Offset management

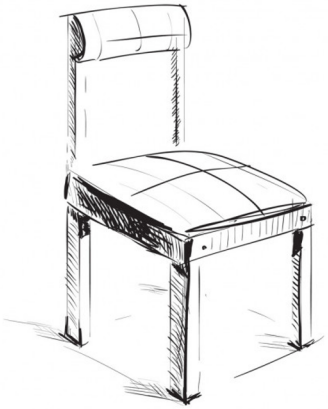
Proxy

Самодельный data-bus

- Поддержка DLQ/Defer
- Поддержка роутинга в разные кластеры

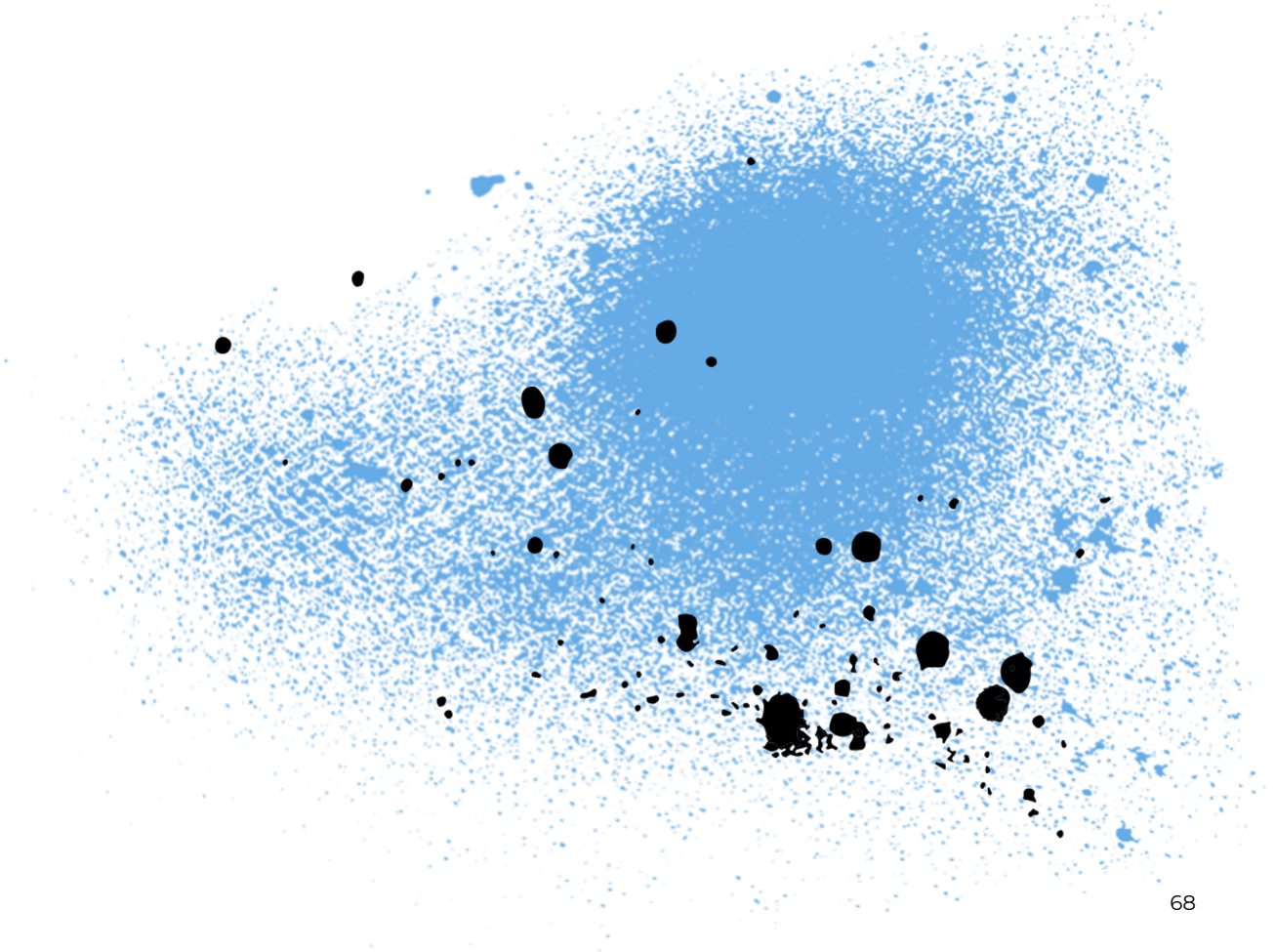
Подробнее про data-bus в Авито – https://www.youtube.com/watch?v=MoqG6iFERPw&ab_channel=AvitoTech

Proxy & Gateway



Почем у вас огурцы соленые? Ну, хорошо, дайте два!
(Двенадцать стульев)

Multi-DC



Multi-DC

Топология

- Идеальный мир – три дата-центра
- Реальный мир – два дата-центра
- Правильный реальный мир – два с половиной дата-центра

Multi-DC

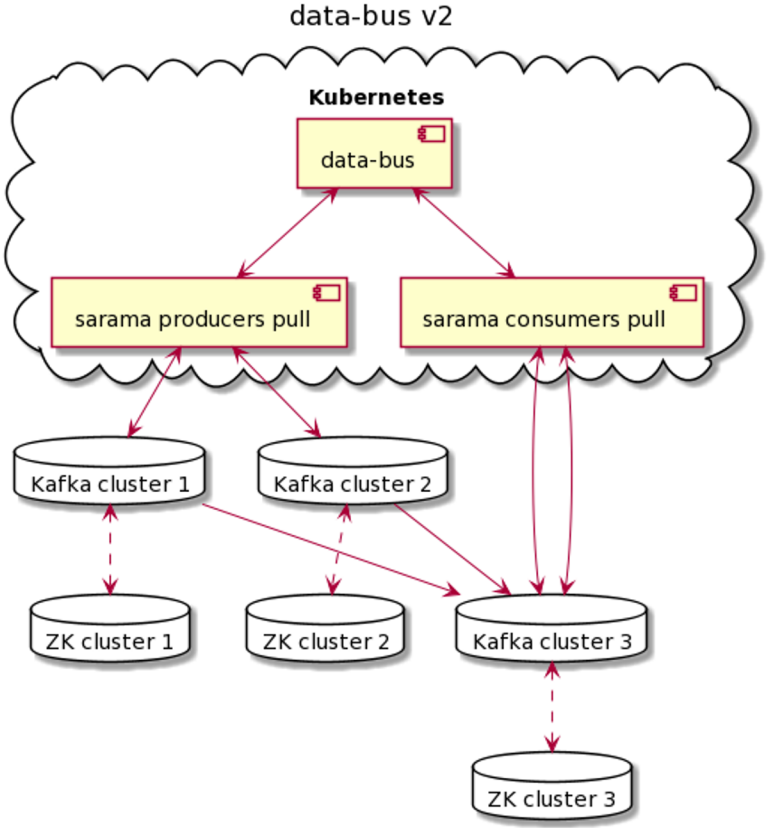
https://gamov.dev/devoops-moscow-2020

Multi-DC

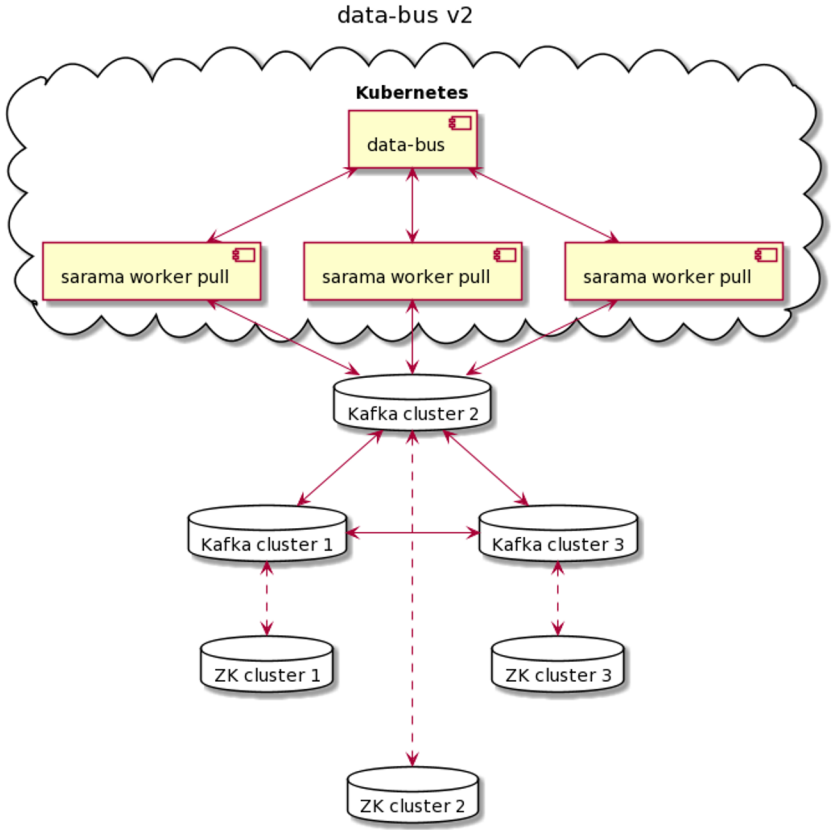
А еще

- Stretched cluster это нормально (если latency около 10ms и ниже)
- Follower Fetching в Kafka 2.4+
- OSS репликаторы полны сюрпризов :)
- Сложные топологии могут быть оправданы и в рамках одного DC (Federation)

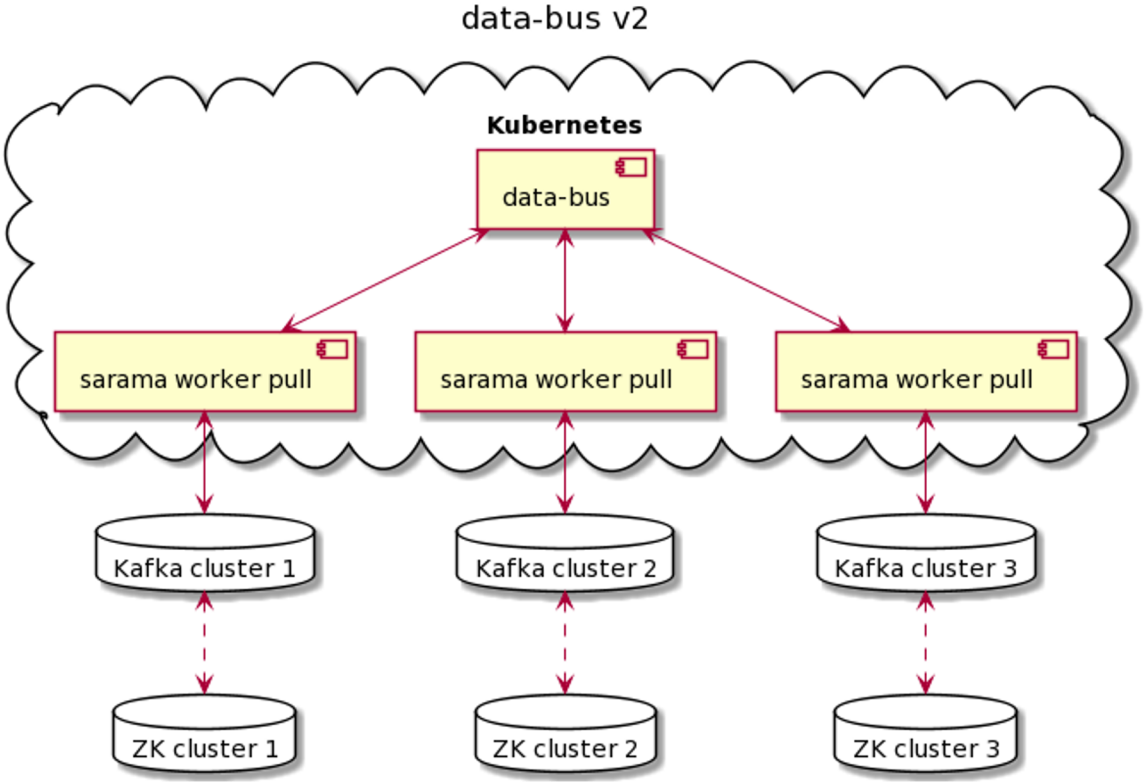
Multi-DC wierd write-active



Multi-DC failover



Multi-DC independent

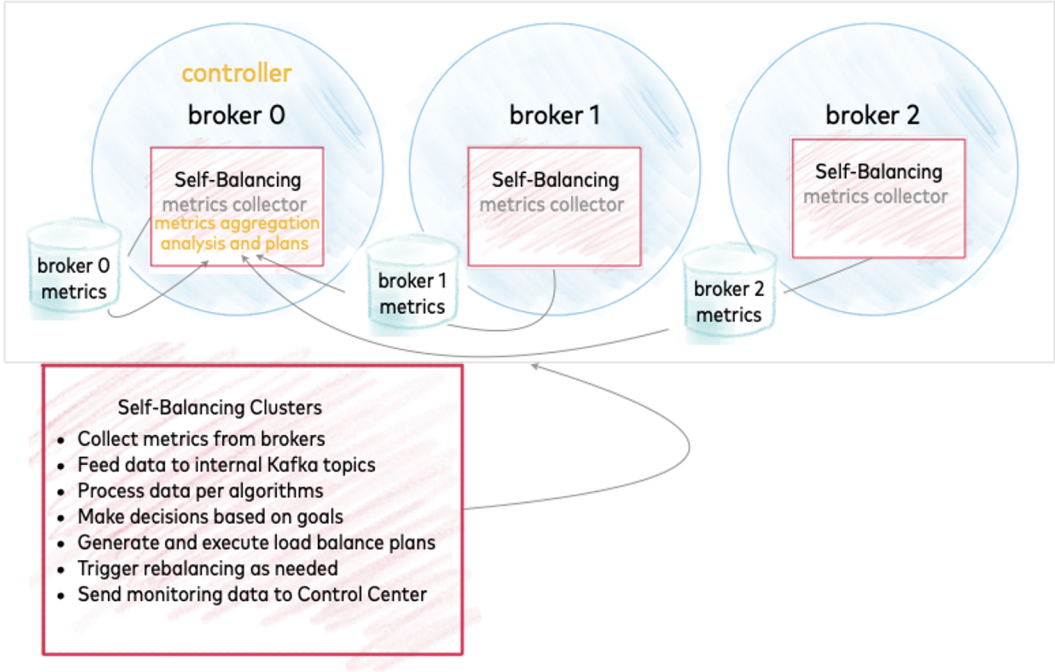


Multi-DC

Обычно те, кто лучше других умеют работать, лучше других умеют не работать

Автоматизация...

Автоматизация

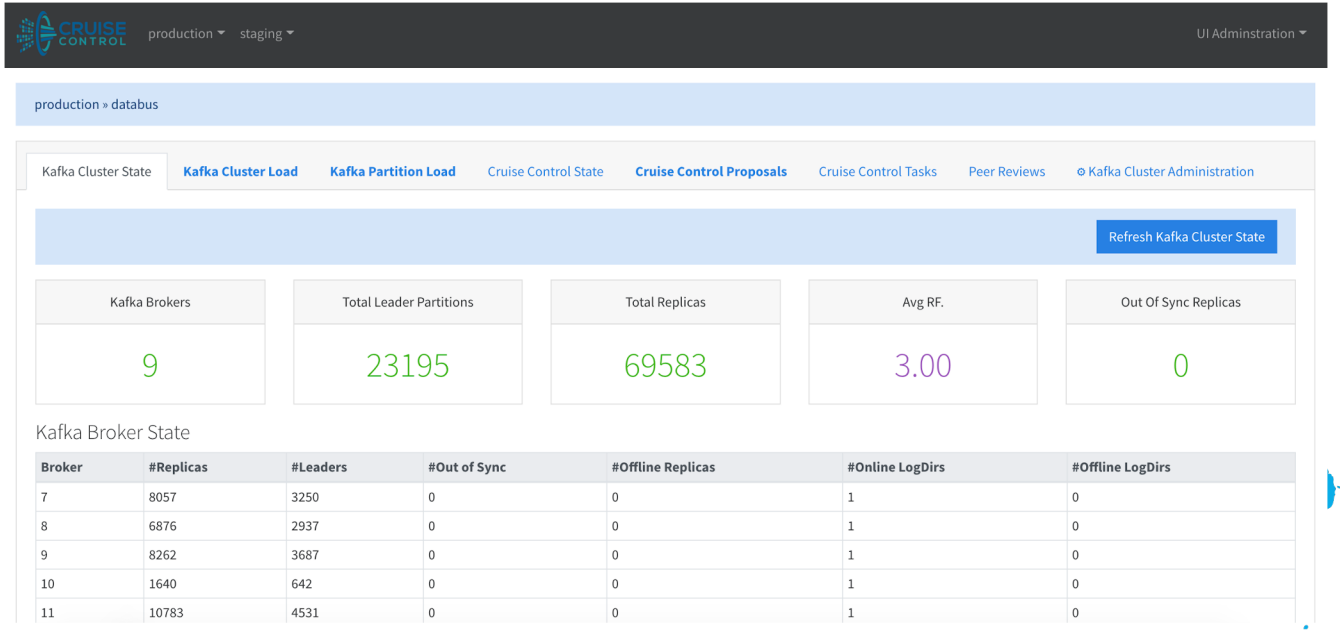


Автоматизация Cruise Control

На что обратить внимание

- Cruise Control из мастер-ветки плохо работает с версиями Kafka 2.4+
- Начинать стоит с ручного управления (UI или REST)
- Затем можно использовать автоматический self-healing

Автоматизация Cruise Control



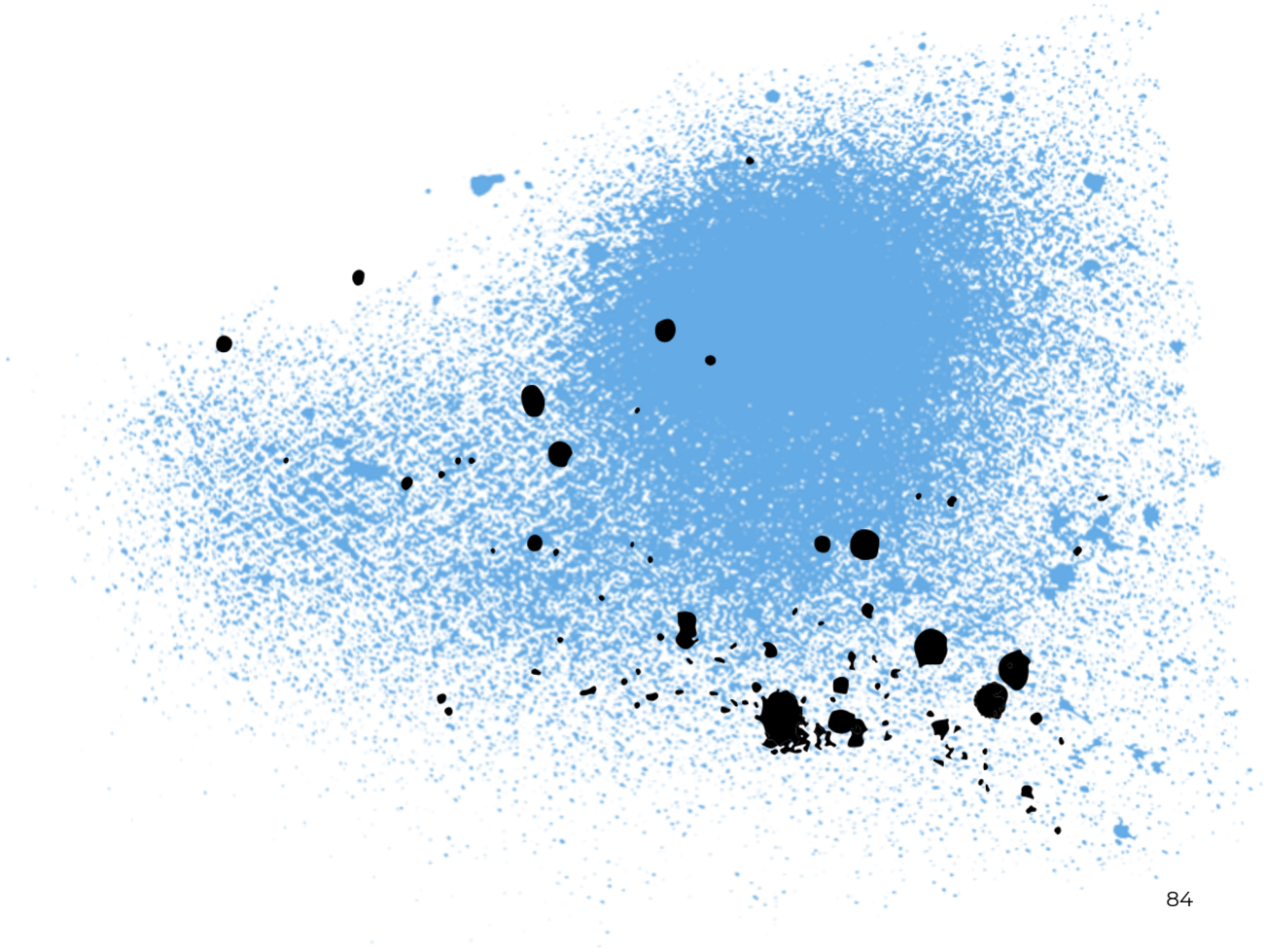
Автоматизация Cruise Control

Goal	Status	Include All Topics	Min Monitored Partition %	Required Snapshots
RackAwareGoal	ready	true	0	1
ReplicaCapacityGoal	ready	true	0	1
DiskCapacityGoal	ready	true	0.9	1
NetworkInboundCapacityGoal	ready	true	0.9	1
NetworkOutboundCapacityGoal	ready	true	0.9	1
CpuCapacityGoal	ready	true	0.9	1
ReplicaDistributionGoal	ready	true	0	1
PotentialNwOutGoal	ready		0.9	2
DiskUsageDistributionGoal	ready	true	0.9	1
NetworkInboundUsageDistributionGoal	ready		0.9	2
NetworkOutboundUsageDistributionGoal	ready		0.9	2
CpuUsageDistributionGoal	ready		0.9	2
TopicReplicaDistributionGoal	ready	true	0	1
LeaderReplicaDistributionGoal	ready	true	0	1
LeaderBytesInDistributionGoal	ready		0.9	2

Автоматизация

You know, I learned something today... (Kyle)

Выводы



Выводы

01. Убедитесь, что у вас есть мониторинг, бэкапы и схемы

02. Определитесь с подходом к эксплуатации (build vs buy)

03. Используйте те технологии, которые работают для вас!

04. Экосистема Kafka очень богата, всегда есть альтернативы


интернет
avito.tech
интернет
дизайн
мнения аудитории
команда
Kafka
DevOops
математика
изучение
мнение
Kafka
Спасибо за внимание!
avito.tech
мнения аудитории
Kafka
платформа
анализ
Confluent
дизайн
Kafka
изучение
дизайн
avito.tech
ui/ux
интеллект
анализ
it-специалист
машинное обучение
avito.tech
интернет
Kafka
Kafka
мнения аудитории
команда
Kafka
Kafka
анализ
мнения аудитории