

Сервис хостинга электронных каталогов библиотек: обзор, основные характеристики

Колобов О.С.

Институт сильноточной электроники СО РАН

Князева А.А., Турчановский И.Ю.

Институт вычислительных технологий СО РАН

Управление ресурсами в библиотеке

Управление ресурсами

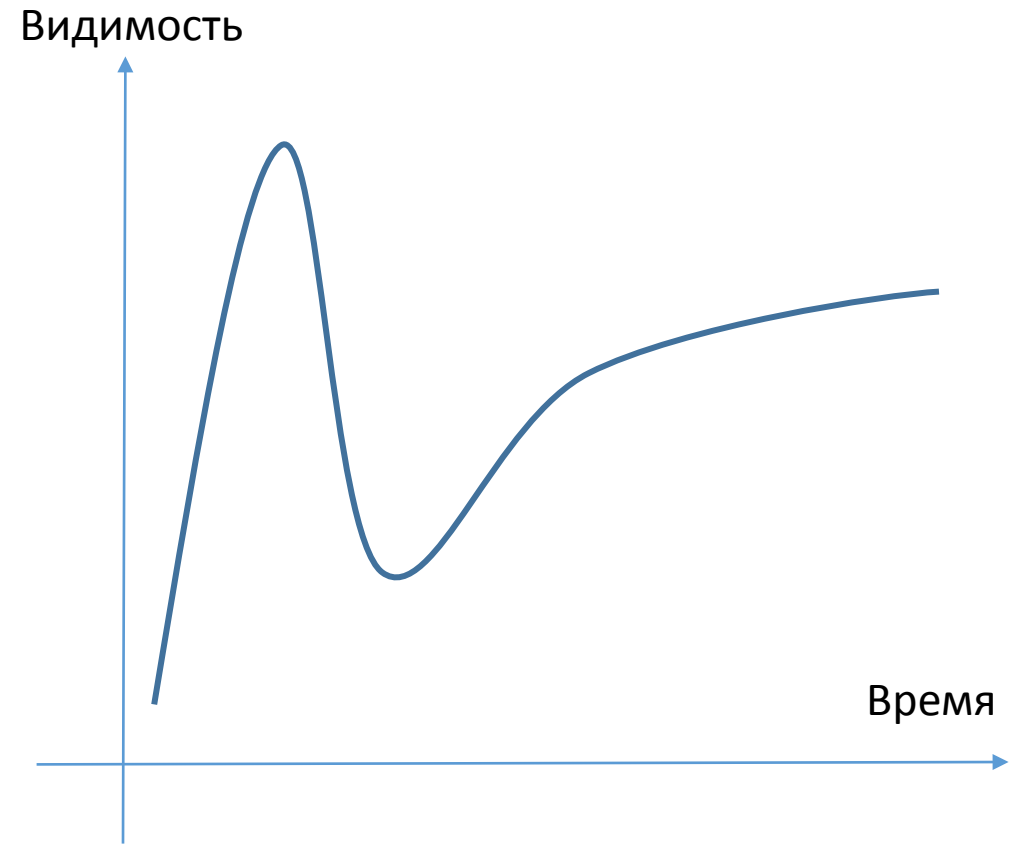
- Поддержка различных форматов для записей метаданных
- Предоставление специфических полей записей
- Индексирование специфических полей записей
- Добавление записей метаданных из различных источников (включая внешние)

Управление сервисами

- Воссоздание связей между записями метаданных (*authority control*)
- Дискавери сервис (*discovery service*)
- Рекомендательный сервис (*recommender service*)

Наша мотивация

- Крупные библиотеки обращаются к услугам сторонних организаций (сервисов) для поиска в своих данных.
- Стагнация систем распределенных электронных каталогов библиотек, которые были созданы ранее.
- Исследовательская работа без необходимых инструментов.



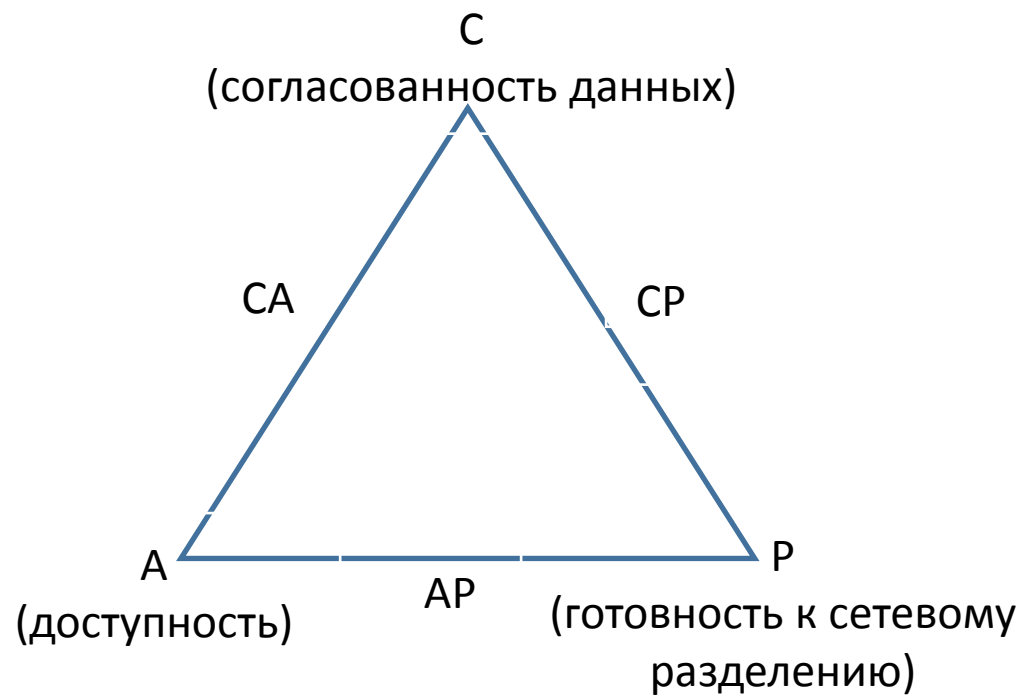
Наш путь

- Применять простой **инструмент управления ресурсами**
- **Создать платформу** для новых библиотечных сервисов
- **Интегрировать платформу в среду PaaS** в пределах действия юрисдикции [отечественного законодательства]

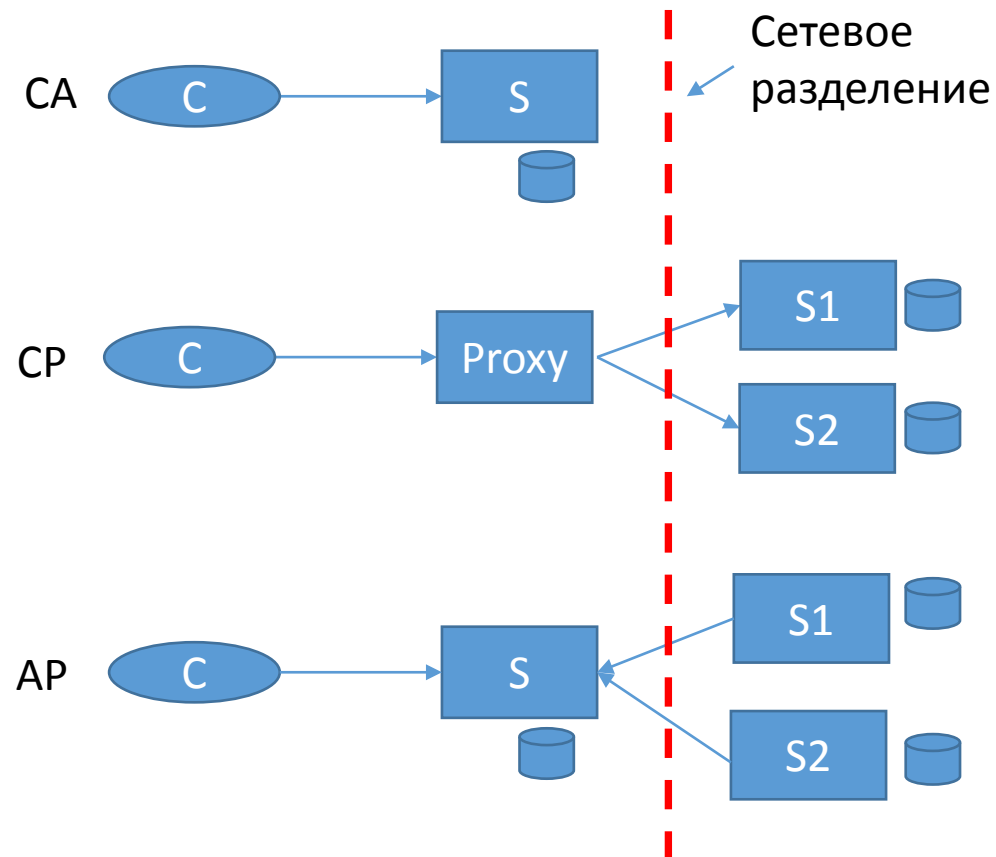
MARC-системы сегодня

- Монолитная организация
- Очень дорогое масштабирование
- Длительный жизненный цикл изменений (1-2 релиза в год)
- Большой барьер входа в тему
- Проприетарное программное обеспечение

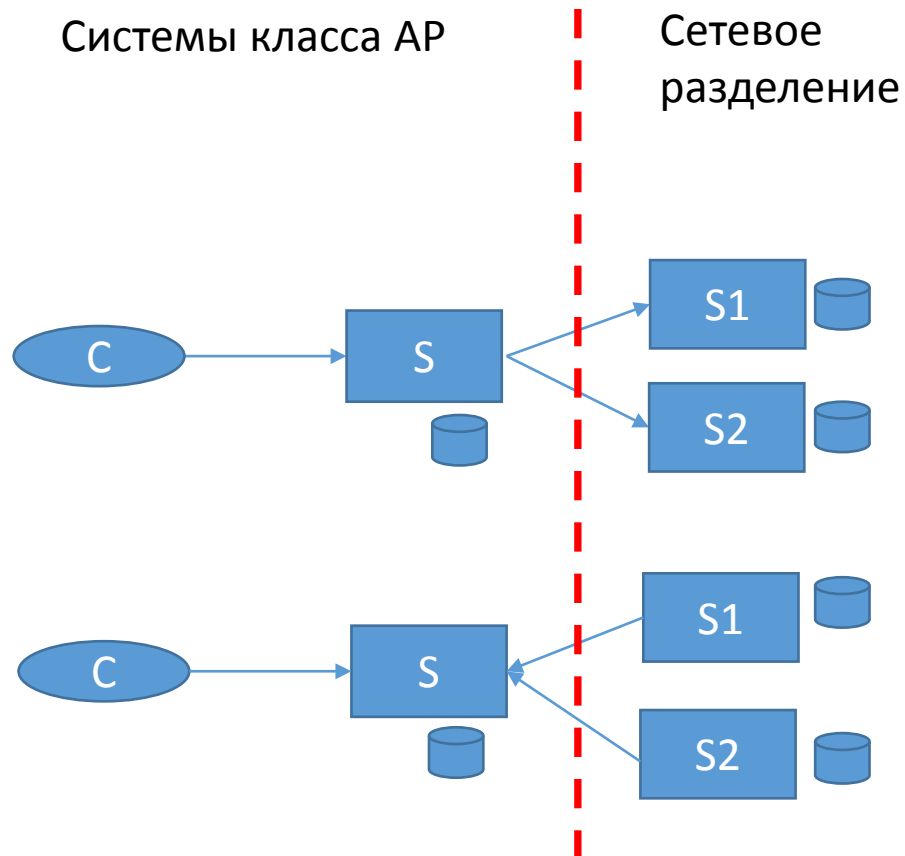
Теорема CAP



CAP theorem (Brewer, E. 2000)



Согласования данных. Тянуть или толкать?



$$T = T_{\text{period}} + T_{\text{sync}}$$

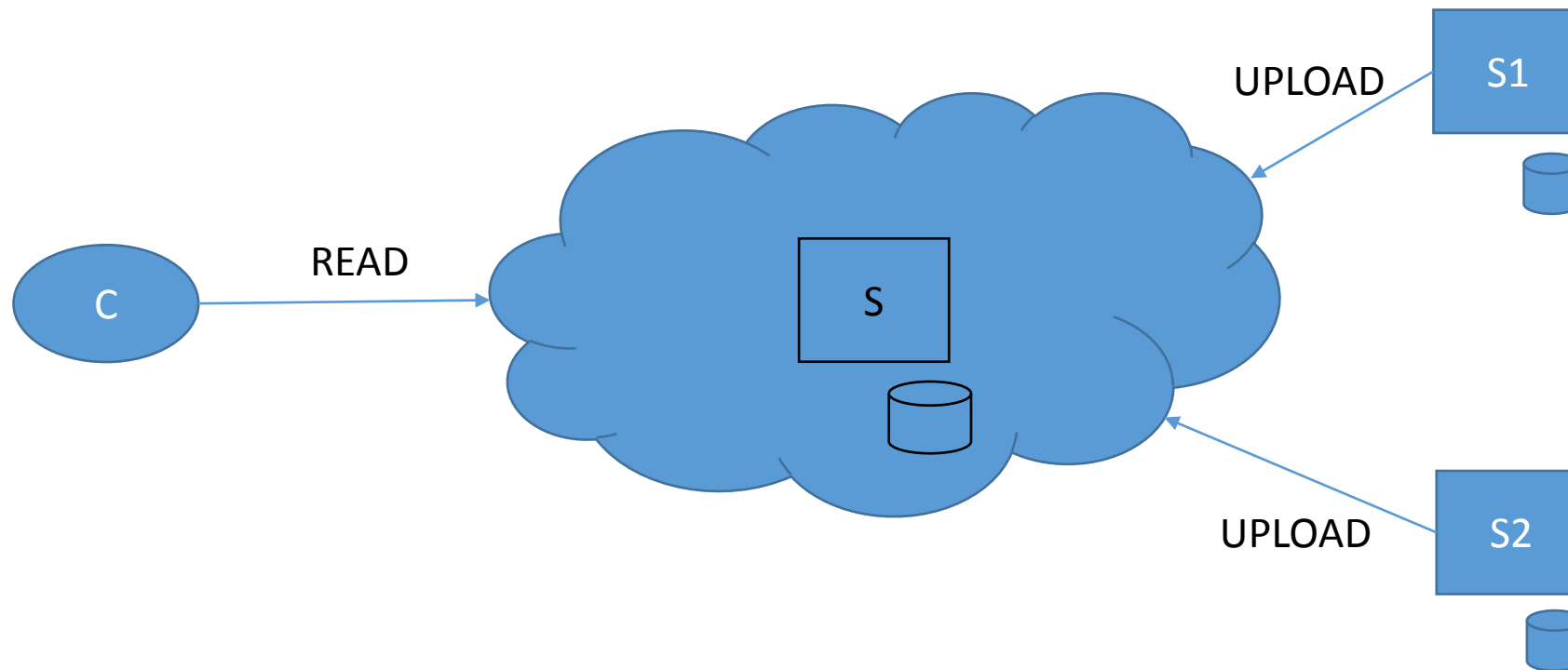
$$T = T_{\text{sync}}$$

T – время согласования данных

T_{period} – длительность интервала

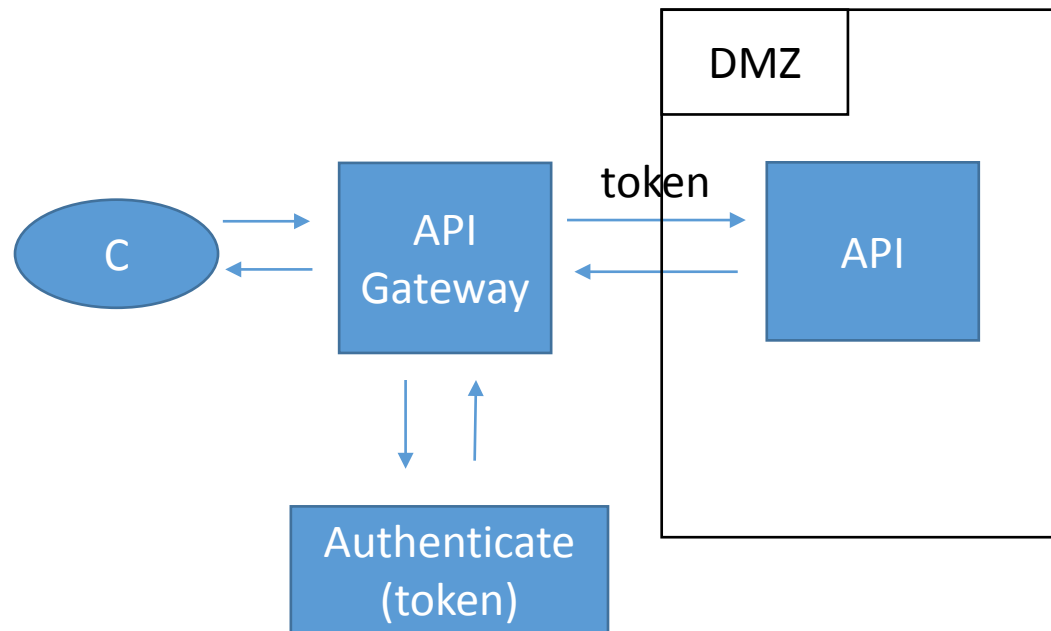
T_{sync} – время синхронизации

Модель



Платформа

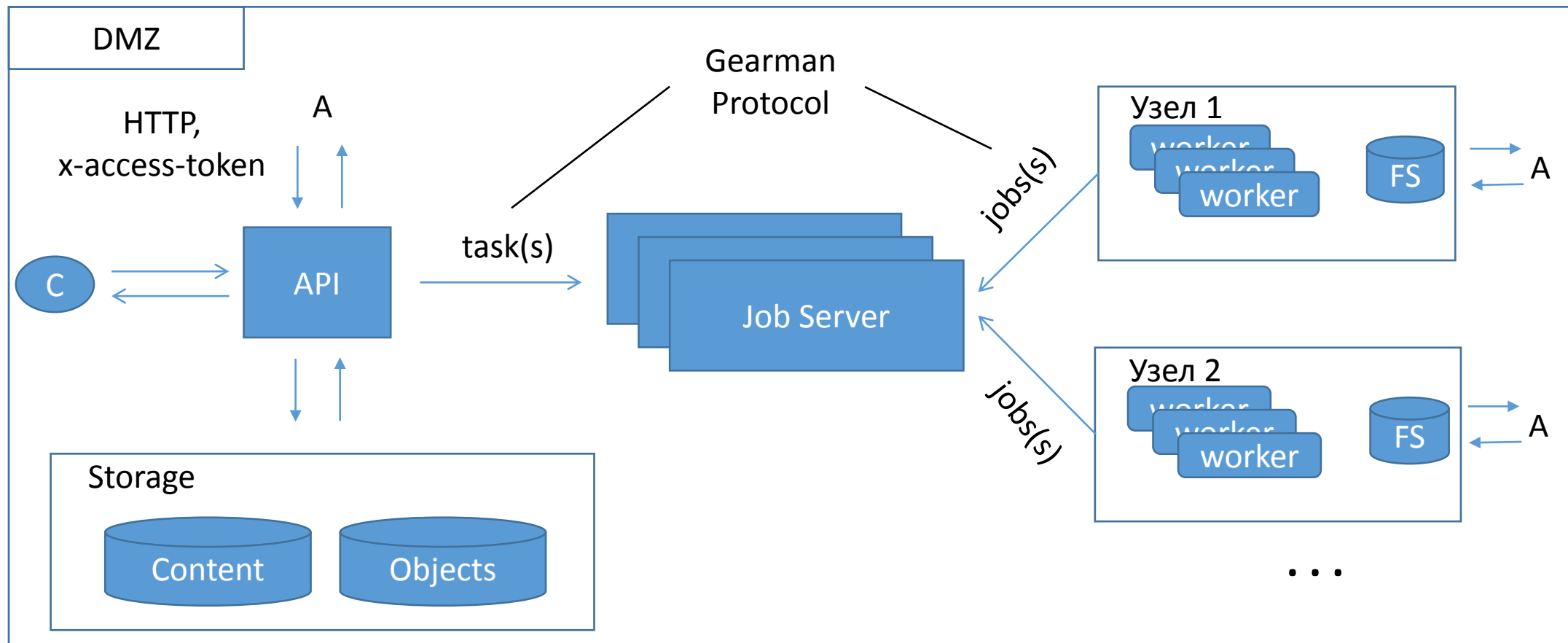
Платформа



Шаблоны

- Microservices
- API Gateway
- Access Token
- Database per Service
- Server-side fragment Composition
- Client-side UI Composition

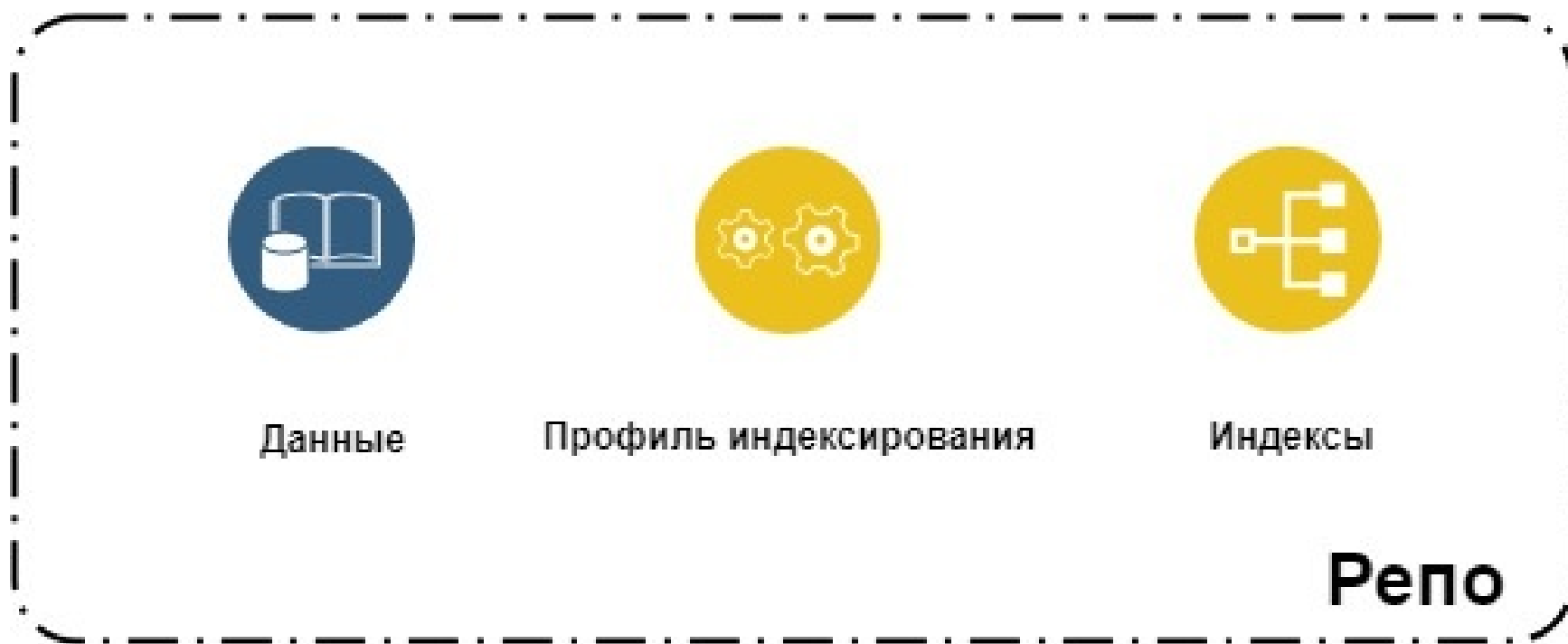
Архитектура



Основные характеристики

- Динамический репозиторий
- Асинхронная обработка запроса

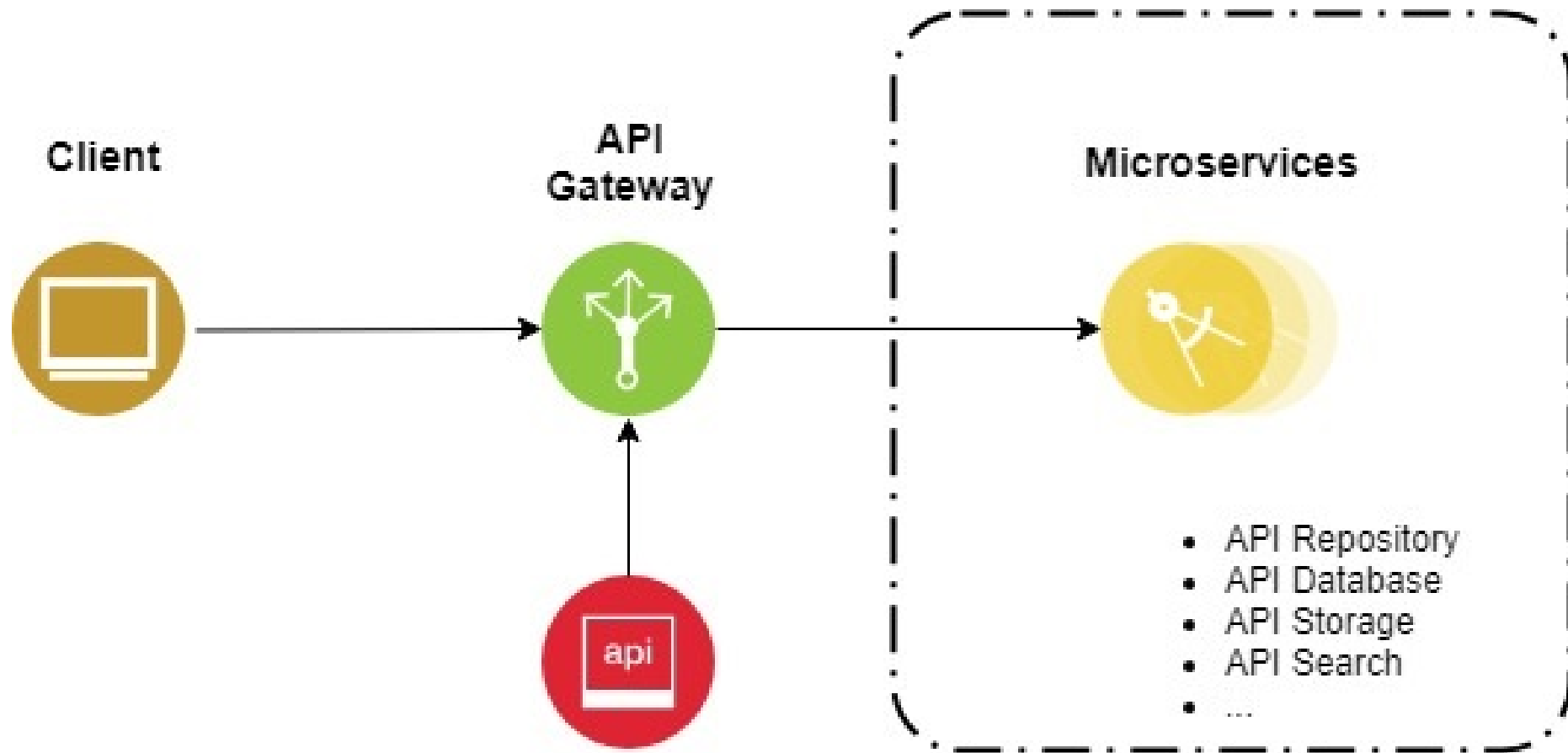
Репозиторий



Кратко о Zebra

- Сервер предназначен для поиска и извлечения информации в структурированных текстовых данных [XML, MARC]
- Создан на основе нескольких моделей для обмена сообщениями (протоколы Z39.50, SRU)
- Поддерживает различные модели поиска: булева модель, векторная модель (алгоритм TF-IDF)
- Высокая производительность при работе с большими базами данных (~100Gb данных, ~40-60M записей)
- GPL License (используете этот сервер, открываете свой код)

API Gateway

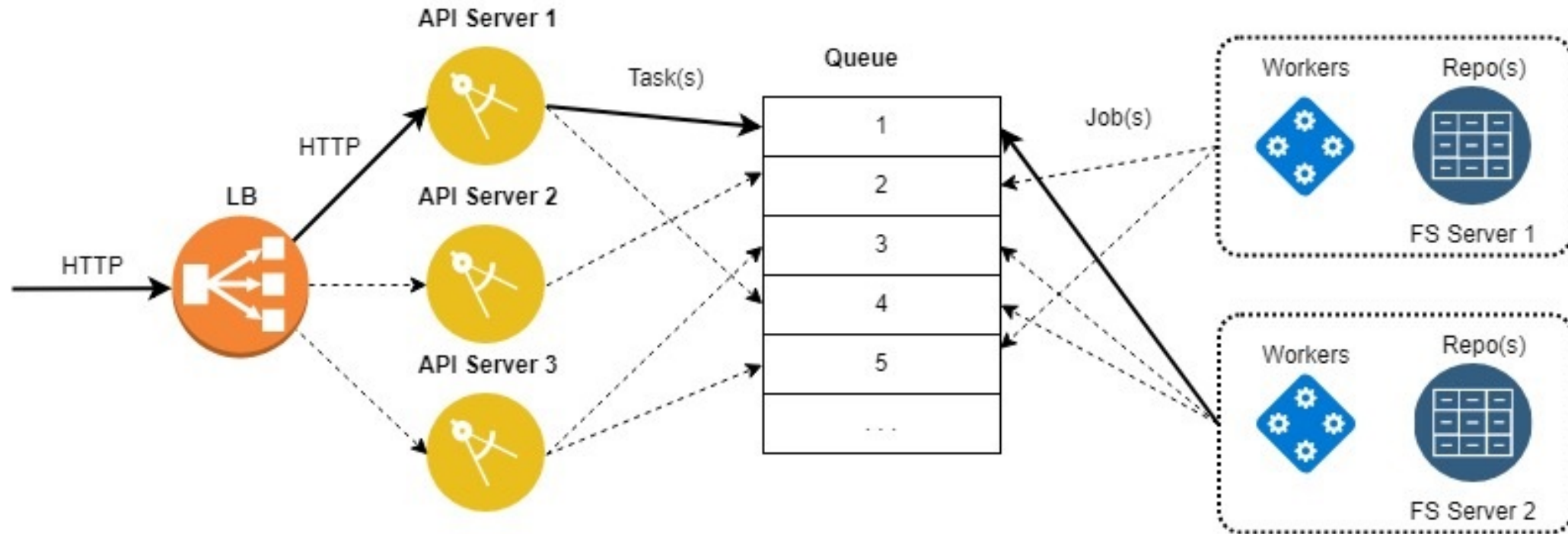


**Authenticate
(token)** АРБИКОН 2018

APIs

- Namespace API
- Repository API
- Database API
- Storage API

Асинхронная обработка запроса



- LB -- балансировка нагрузки
- API Server -- API-сервер
- Queue -- очередь сервера очередей
- Worker -- процесс для работы с репо
- Task -- задача для worker
- Job -- работа для worker
- FS Server -- Файловый сервер

Методы работы с запросами

- Background jobs
- Map/reduce
- Pipeline processing

Консольный клиент платформы `cuba`

- Скрипт для `sh`
- Доступ к API
 - `cuba ns [opts\*]`
 - `cuba repo [opts\*]`
 - `cuba db [opts\*]`
 - `cuba st [opt\*]`
- Зависимость от ПО
 - `curl`
 - `jq`

Пример

```
$cuba ns -c myns | cuba repo -c myrepo -t grs.marc.rusmarc | cuba -c  
db mydb
```

```
$ uuid=`cuba -q '{"repositoryName":"myrepo", "namespaceName":  
"myns"}' | jq -r 'data.uuid`
```

```
$cuba st -c myfile -f file:///tmp/data/simple-rusmarc-utf8 | cuba -x  
update -u ${uuid}
```

Полезные свойства

- Сервис и API – платформа
- Раздельный поиск
- Раздельное индексирование
- Раздельное хранение
- Простая миграция данных
- Горизонтальное масштабирование

Заключение

- Для каждой распределенной системы необходимо соблюдать разумный баланс доступности и согласованности данных.
- Применение шаблонов проектирования позволяет избежать грубых ошибок на этапе проектирования системы.
- Асинхронная обработка запросов – основа для решения задачи масштабирования распределенной системы.
- Программная платформа распределенной системы – основа для новых сервисов сети.

Спасибо за внимание!

Сервис хостинга электронных каталогов библиотек: обзор, основные характеристики

Колобов О.С.

okolobov@gmail.com

Институт сильноточной электроники СО РАН