

openBD API

日本国内で最も速い、自由な書誌情報

カ-リル™

コンセプト

- 超高速（キャッシュしなくてもよい）
- 使いやすい
- 落ちない（大量アクセス可能）

ウェブサービスにおける書誌データのニーズ

- 書誌情報の充実（ウェブに表示）
- 書影の表示
- 検索インデックスのリッチ化／精度向上
- データ管理からの開放

（間違いの修正がすぐ反映されること）

高速化のための工夫

- 70万件のデータ

XML形式（約5GB）→JSON形式（3.2GB） 36%削減

- KVS時代のAPI設計

Redis + Golang すべてのデータをメモリ上に展開

キャッシュレス化のための工夫

- 一括取得
 - 検索結果にJOINしたい
 - フロントエンドのみで完結可能（CORS）
- マルチデータセンター展開



使いやすさのための工夫

- JPOのONIX運用に適合したJSON変換

XMLの課題…データ取得のためのプログラムが煩雑

仕様書の確認を必須

JSON変換時に複数項目はリストに、単一項目はオブジェクトに変換

実データを見ながらプログラムを書くだけで高い互換性を実現できる

- ONIXと版元独自書誌を統合したまとめ書誌を提供

書名・著者表記のレンダリング済みバージョン

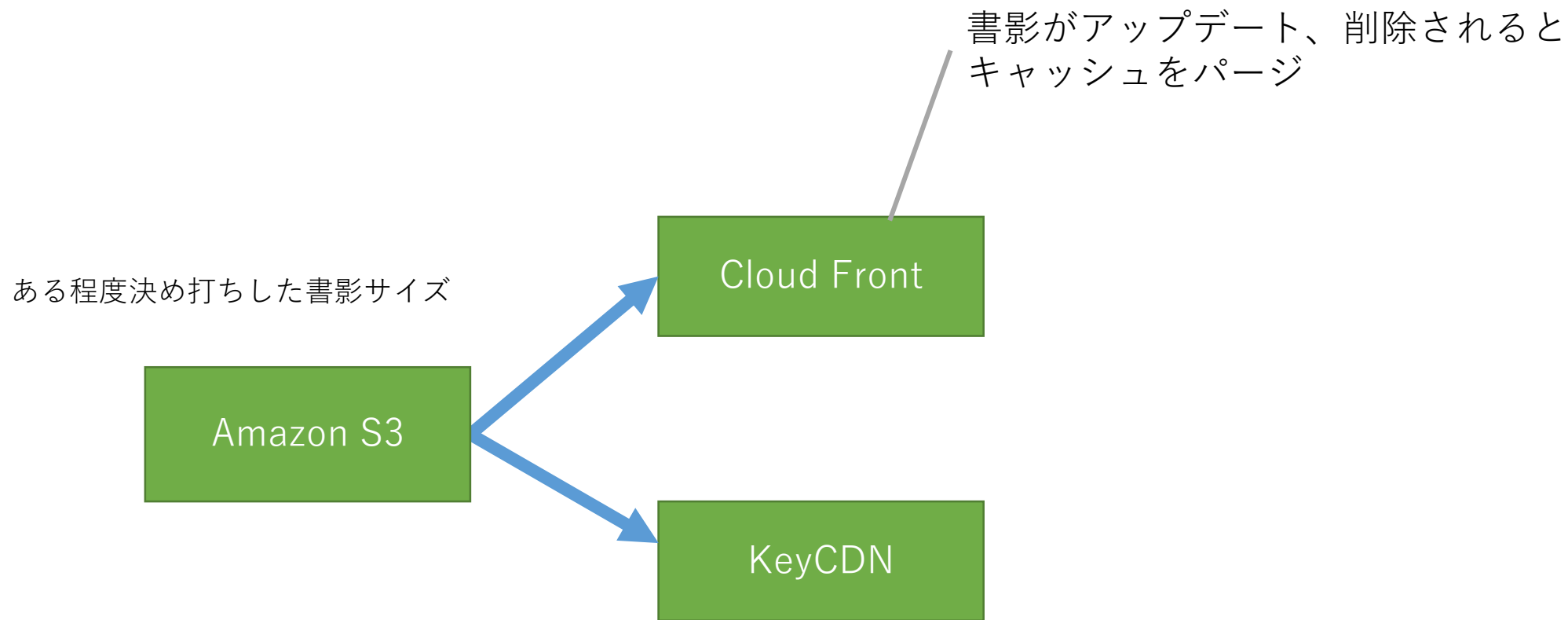
API構成

- 書誌情報の取得 (**get**) ※—— 10000件まで一括指定ができる
- 収録範囲の取得 (**coverage**)
- JSONスキーマの取得 (**schema**)
- 書影→URL規則で対応
- 見てみましょう

性能目標・実測値

- 20000リクエスト / 秒の実測値
(1件単位で取得)
- 10000リクエスト単位での取得→1秒で応答

書影情報の考え方



openBD API v1（最終リリース候補版 2017/1/22）

基本エンドポイント：<http://api.openbd.jp/v1/>

応答はすべてJSON

- 収録範囲の取得 (/coverage) [GET]
- 少量の書誌の取得 (/get) [GET]
 - 1,000件まで複数指定
 - isbn…ISBNを指定。複数指定の場合はカンマ[,]で区切る
 - pretty…JSONを整形表示する
- 大量の書誌を取得 (/get) [POST]
 - 10,000件まで複数指定
 - isbn…ISBNを指定。複数指定の場合はカンマ[,]で区切る

レイテンシに関する考え方

- 現在のデプロイ先は「Google Cloud Platform」
- 同じデータセンター内では1ms～3ms以内で応答
- 今後、「AWS」「さくらインターネット」に順次展開予定

実測の応答速度（フレッツ回線で計測・・・レイテンシ25ms程度）

- 10件の取得・・・30ms
- 1000件の取得・・・500ms
- 10000件の取得・・・5500ms
- 全件の取得・・・50秒 / 4接続で並行取得・・・30秒

データ形式についての追記事項

- 基本的なデータは版元ドットコムから
- Summary – 使いやすい基本事項
 - title,author,publisher,series,pubdate,isbn

実装上の要点

- なるべくまとめて取得する（HTTPのペイロードを減らす）
- Keep-Aliveを推奨（セッションプール）
- HTTP/2（SSLの使用）を推奨
- 同時接続数が100接続程度であれば、
ウェイトなどの配慮は不要
 - 5000req/seq以上の性能は常時確保
- 全件取得は、coverageで取得したリストを
10000件ごとに分割してgetする

書影の実装

- CDNによる提供
- 大規模利用は独自CDNを推奨
 - （キャッシュパージ連携に対応 – CloudFront/KeyCDN）