Openlink Virtuoso for LinkedMusic

Hong Van Pham
Hanwen Zhang

McGill

LinkedMusic

DDMAL DISTRIBUTED DIGITAL MUSIC ARCHIVES & LIBRARIES LAB

Presentation Goals

01

Introduce

What is Virtuoso? Who's Openlink? What are the (relevant) components of Virtuoso?

02

Implement

How do we use Virtuoso for LinkedMusic? What choices were made? How to intergrate other software?

03

Demonstrate

How does that look like in action? How can I test this myself?

04

Expand

Which direction are we taking? What problems do we have with the current implementation? What is our next mile stone?

01.
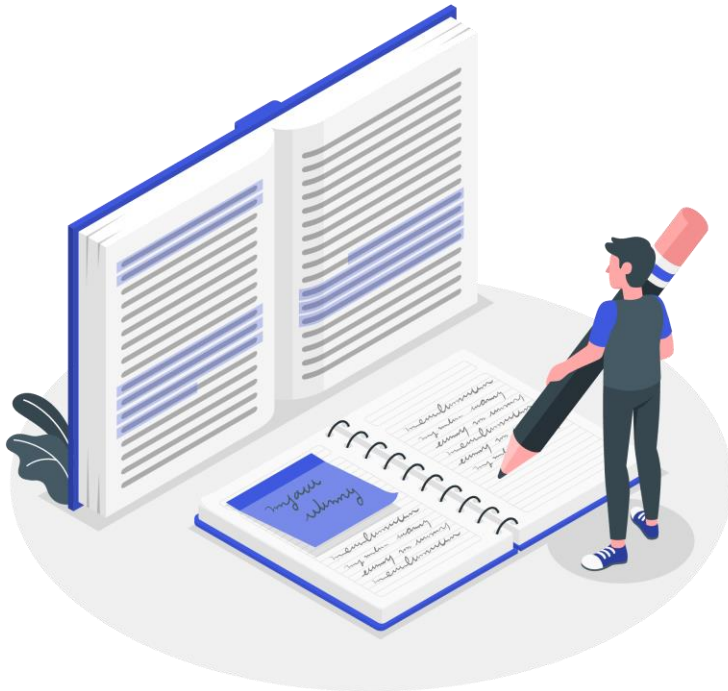
Introduce

Introduction

OPENLINK
SOFTWARE

Openlink

OpenLink Software is headquartered in Burlington, Massachusetts, USA. They specialize in data connectivity solutions, semantic web technologies, middleware, integration, and security solutions for effective data management.

Virtuoso

Combined multi-model DBMS and Data Virtualization platform. It serves as a hybrid data integration platform, providing support for various data models, including relational, graph, and document-oriented data.
It allows organizations to manage, query, and integrate data from different formats and locations.

(ChatGPT)

Virtuoso

DBMS

Multi-model platform, manages relational, RDF, and document-oriented data. Suitable for handling large datasets and providing a unified solution for diverse data types.

Linked Data

Support RDF, enable robust representation and querying of data through RDF triples. Integrates SPARQL, facilitates deployment of interconnected linked data sets.

Search Engine

Full-text search across various data types. Integrate semantic search into applications with SPARQL-Full Text Extensions.

Component: Sponger Middleware

The Virtuoso Sponger is a middleware component designed generate Linked Data from a variety of data sources, and supports a wide variety of data representation and serialization formats.

- Extract non-RDF data and convert them to Linked Data format

- Virtuoso Content Crawler

Even though most of the data we consider are RDF data, the concept of Sponging (transform and extract structured data) is still very relevant for the demo.

Sponging: extracts structured data from various web sources and converts it into RDF format
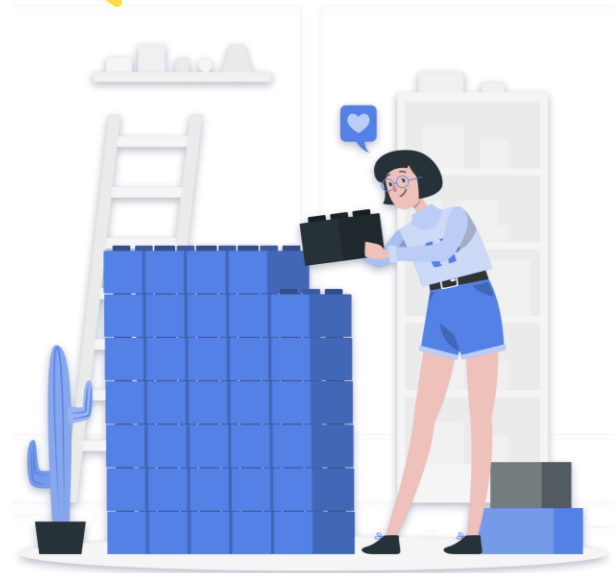
Compared to Apache Jena

Data model support

Virtuoso supports both SPARQL for RDF data and SQL for relational data, while **Jena** is made specifically and only for RDF data.

DBMS

Virtuoso offers full-fledged DBMS capabilities, while **Jena** only provides some storage capabilities.

In-built search-engine

Jena lacks built-in full-text search capabilities while **Virtuoso** offers full-text search and indexing across different data types.
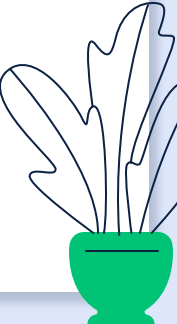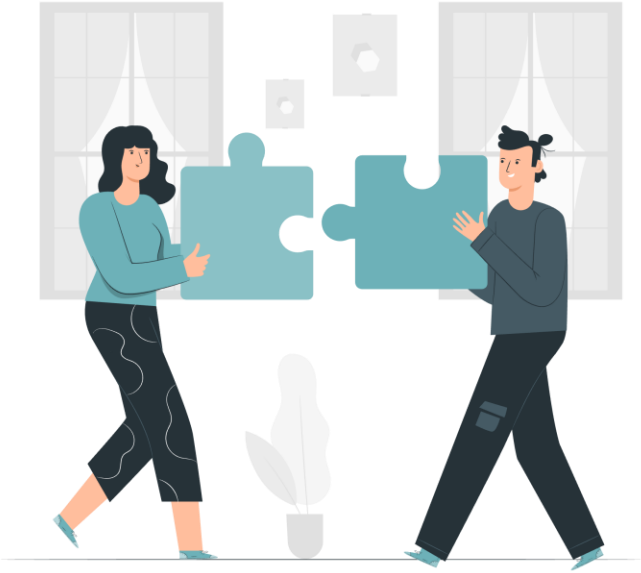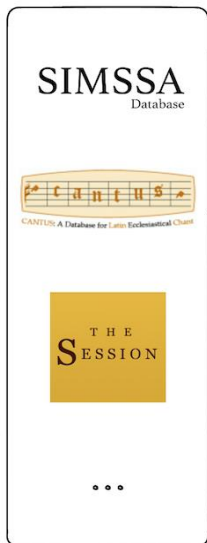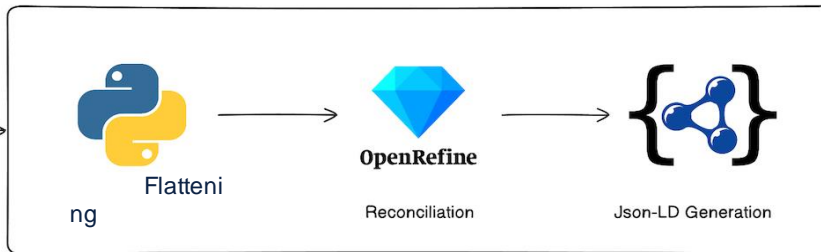
Virtuoso is a more comprehensive solution

02.

Implement

Previous Pipeline (Oct 2023)

Sources

SIMSSA
Database

cantus
CANTUS: A Database for Latin Ecclesiastical Chant

THE
SESSION

. . .

Pipeline

Flatteni
ng

OpenRefine
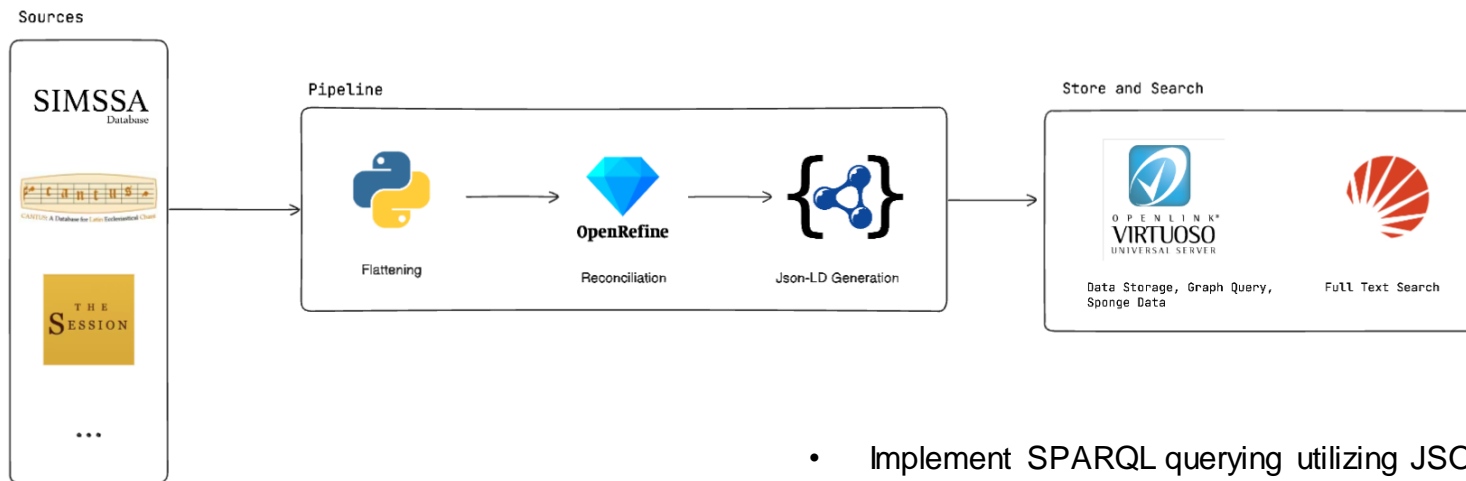Reconciliation

Json-LD Generation

Index and Search

Solr

elasticsearch

- No graph database implementation

- Didn't utilize the JSON-LD format

- Manually pulling wikidata information
  (alternative names, different languages, etc.)
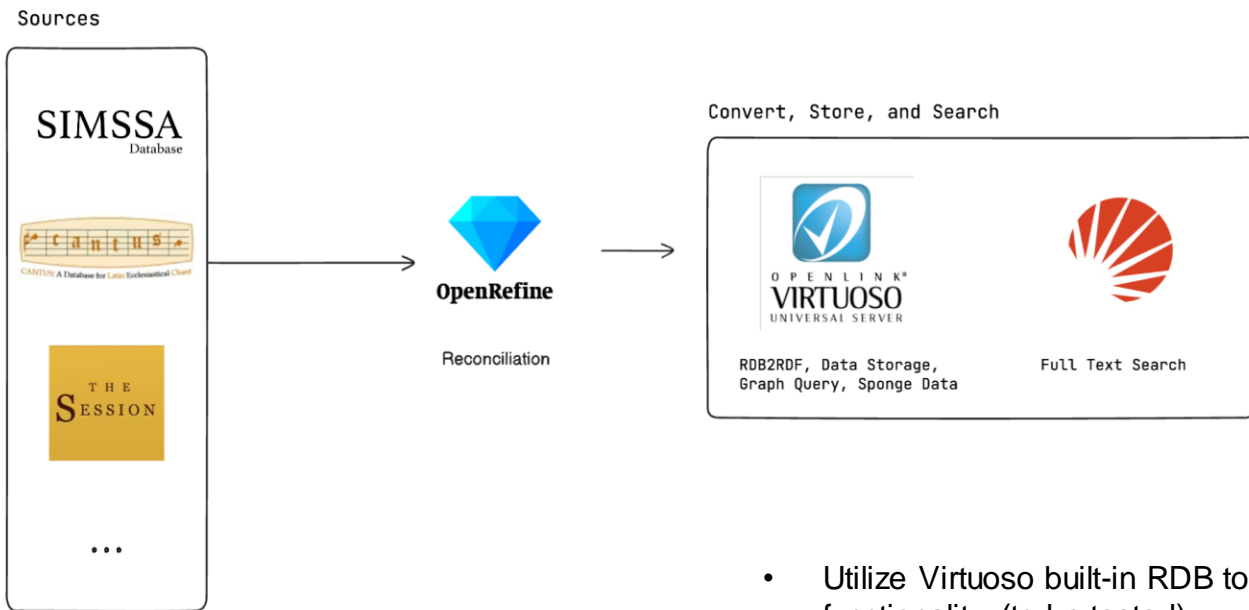
Current Pipeline (not final)

- Implement SPARQL querying utilizing JSON-LD format

- Automatically Sponge external information (Wikidata, schema.org, other databases, etc.)

Possible Future Implementation

- Utilize Virtuoso built-in RDB to RDF mapping functionality (to be tested)

RDB to RDF Mapping

Our use case: import from SIMSSA (PostgreSQL) to Virtuoso

Approach 1
Link PostgreSQL into Virtuoso using Conductor web interface (enterprise version)

Approach 2
Manually preprocess and load into Virtuoso

Option 2.1
Convert into CSV and upload into Virtuoso Conductor
--> lose schema information and complex data structure

Option 2.2
Convert into RDF and load into Virtuoso
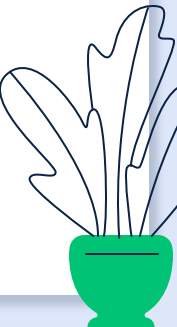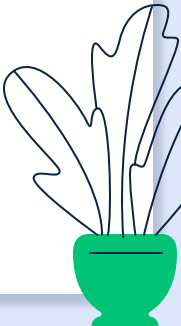--> use R2RML (RDB to RDF Mapping Language)

03.

Demo

https://virtuoso.staging.simssa.ca

(McGill VPN/network necessary)

04.

Expand

OPENLINK®
VIRTUOSO
UNIVERSAL SERVER

What next?

1. **Re-design pipeline?**
   - **Reconcile before flattening?**
   - **Might not need json-ld anymore? (rdb2rdf and/or utilize CSVs)**

2. **Combine Virtuoso with other software**
   - **Solr for more full-text search power?**
   - **UI?**

3. **Get more data and databases!**

Thank you!

van.pham2@mail.mcgill.ca
hanwen.zhang4@mail.mcgill.ca

McGill

LinkedMusic

DDMAL DISTRIBUTED DIGITAL MUSIC ARCHIVES & LIBRARIES LAB