**TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI**

**VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG**

---------\*\*\*---------



**Báo cáo môn học**

**LẬP TRÌNH HƯỚNG ĐỐI TƯỢNG**

**Đề tài:Tìm hiểu công nghệ lưu trữ và truy vấn ArangoDB**

**Giảng viên hướng dẫn: TS. Trịnh Tuấn Đạt**

**Sinh viên thực hiện: Nhóm 8**

1.Trần Minh Hiếu MSSV: 20170075

2.Trần Văn Đạo MSSV: 20170051

3.Trần Hoàng Chuẩn MSSV: 20170043

4.Đinh Văn Hiếu MSSV: 20170072

5.Đỗ Đình Đắc MSSV: 20170049

|  |  |  |  |
| --- | --- | --- | --- |
| Thành viên | MSSV | Công việc | Phần trăm đóng góp |
| Trần Minh Hiếu | 20170075 | Package DAO, models | 60% |
| Trần Hoàng Chuẩn. | 20170043 | Viết báo cáo, vẽ diagram, tổng hợp các package | 10% |
| Trần Văn Đạo | 20170051 | Package query, Rawdata, view, kiểm thử hiệu năng. | 10% |
| Đỗ Đình Đắc | 20170049 | 10% |
| Đinh Văn Hiếu | 20170072 | Creat Data, | 10% |

**MỤC LỤC**

[**I.TÌM HIỂU VỀ CÔNG NGHỆ ARANGODB** 4](file:///C:\Users\MAMXAN~1\AppData\Local\Temp\Rar$DIa5188.14182\BaoCao.docx#_Toc26250002)

[**1.Giới thiệu** 4](file:///C:\Users\MAMXAN~1\AppData\Local\Temp\Rar$DIa5188.14182\BaoCao.docx#_Toc26250003)

[**2. Cài đặt chi tiết** 5](file:///C:\Users\MAMXAN~1\AppData\Local\Temp\Rar$DIa5188.14182\BaoCao.docx#_Toc26250004)

[**3. Các kỹ thuật sử dụng** 7](file:///C:\Users\MAMXAN~1\AppData\Local\Temp\Rar$DIa5188.14182\BaoCao.docx#_Toc26250005)

[**4. Cách truy vấn** 7](file:///C:\Users\MAMXAN~1\AppData\Local\Temp\Rar$DIa5188.14182\BaoCao.docx#_Toc26250006)

[**II. Mô hình dữ liệu** 14](file:///C:\Users\MAMXAN~1\AppData\Local\Temp\Rar$DIa5188.14182\BaoCao.docx#_Toc26250007)

[**III. Chi tiết về chương trình……………………………………………………………………………..**15](file:///C:\Users\MAMXAN~1\AppData\Local\Temp\Rar$DIa5188.14182\BaoCao.docx#_Toc26250008)

[**IV. Kết quả kiểm thử hiệu năng** ..26](file:///C:\Users\MAMXAN~1\AppData\Local\Temp\Rar$DIa5188.14182\BaoCao.docx#_Toc26250009)

[**6.1: Các câu truy vấn cơ bản, nâng cao** 27](file:///C:\Users\MAMXAN~1\AppData\Local\Temp\Rar$DIa5188.14182\BaoCao.docx#_Toc26250010)

[**6.2:Ưu nhược điểm của công nghệ** 27](file:///C:\Users\MAMXAN~1\AppData\Local\Temp\Rar$DIa5188.14182\BaoCao.docx#_Toc26250012)

**NỘI DUNG**

# **I.TÌM HIỂU VỀ CÔNG NGHỆ ARANGODB**

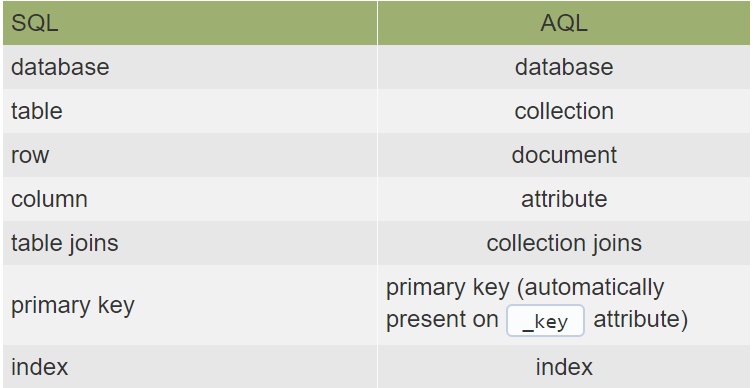
## **1.Giới thiệu**

ArangoDB là một hệ thống [cơ sở dữ liệu đa mô hình](https://en.wikipedia.org/wiki/Multi-model_database" \o "Multi-model database) [miễn phí mã nguồn mở,](https://en.wikipedia.org/wiki/Free_and_open-source" \o "Nguồn mở và miễn phí)được phát triển bởi ArangoDB GmbH. Hệ thống cơ sở dữ liệu hỗ trợ ba mô hình dữ liệu (key /value, document, graph data) với một lõi cơ sở dữ liệu và [ngôn ngữ truy vấn](https://en.wikipedia.org/wiki/Query_language" \o "Query language) thống nhất AQL (Ngôn ngữ truy vấn ArangoDB).

 Ngôn ngữ truy vấn là khai báo và cho phép kết hợp các mẫu truy cập dữ liệu khác nhau trong một truy vấn duy nhất. ArangoDB là một [NoSQL](https://en.wikipedia.org/wiki/NoSQL) hệ thống cơ sở dữ liệu nhưng AQL cũng tương tự như trong nhiều cách hướng [SQL](https://en.wikipedia.org/wiki/SQL)

Các thuật ngữ cơ bản:

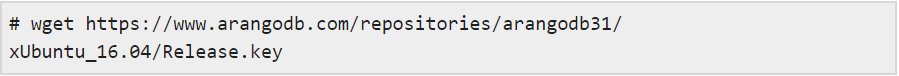
* Key/Value: mẫu dữ liệu lưu trong các bảng băm gồm cặp khóa/giá trị
* Column-Oriented: dữ liệu được lưu trữ trong các phần của các cột, cho phép tính mềm dẻo và dễ dàng trong kết tập
* Document: dữ liệu bao gồm tập tài liệu, mỗi tài liệu gồm nhiều trường mà không nhất thiết phải định nghĩa lược đồ
* Graph: bao gồm các đỉnh nối với nhau bằng các cạnh



## **2. Cài đặt chi tiết**

Với hệ điều hành Linux

* Bước đầu tiên là tải xuống khóa công khai cho kho lưu trữ của nó



* Cài đặt khóa đã lưu bằng dòng mã sau



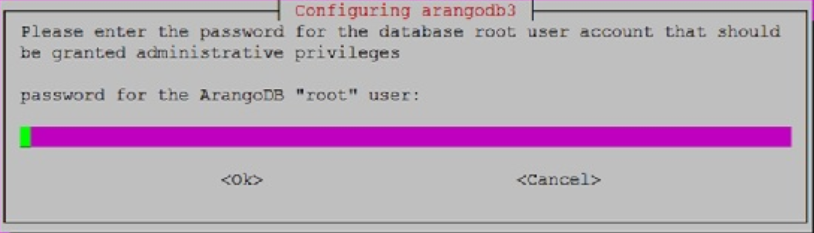
* Chạy các lệnh sau để thêm kho apt và cập nhật chỉ mục

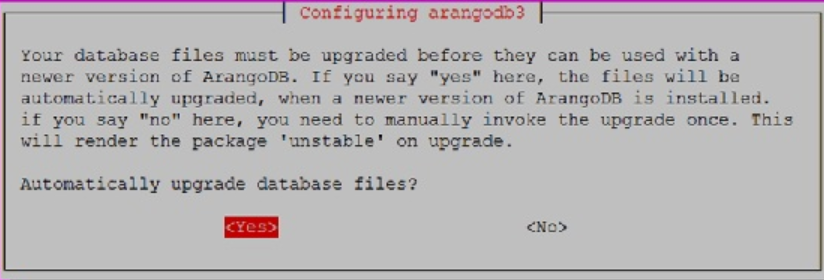


* Cuối cùng có thể cài đặt ArangoDB



* Sau khi cài xong Arango, ta thiết lập cấu hình quản trị







## **3. Các kỹ thuật sử dụng**

* Lưu trữ
* Insert dữ liệu
* Sử dụng java để kết nối dữ liệu với Arango

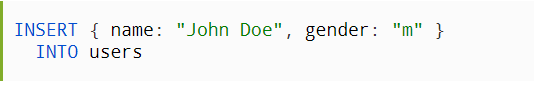
## **4. Cách truy vấn**

Các **Query Language ArangoDB (AQL)** cũng tương tự như **Query Language Structured (SQL)**. Cả hai đều hỗ trợ đọc và sửa đổi dữ liệu thu thập, tuy nhiên AQL không hỗ trợ các hoạt động định nghĩa dữ liệu, chẳng hạn như tạo và xóa cơ sở dữ liệu, bộ sưu tập và chỉ mục

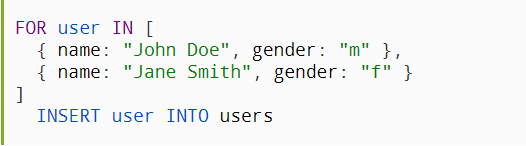
**CHÈN (INSERT)**



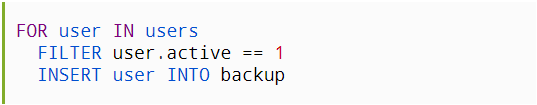
Chèn một hàng / tài liệu



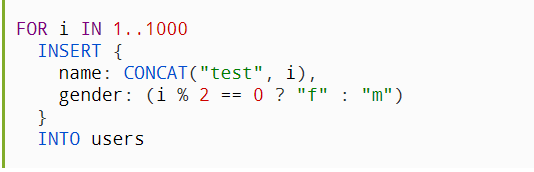
Chèn nhiều hàng / tài liệu



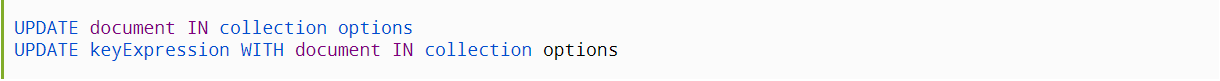
Chèn hàng / tài liệu từ bảng / bộ sưu tập



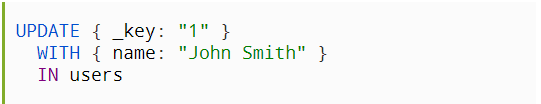
Tạo các hàng / tài liệu kiểm tra



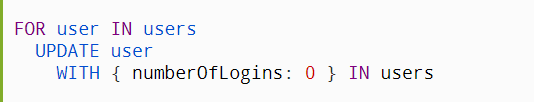
**CẬP NHẬT**



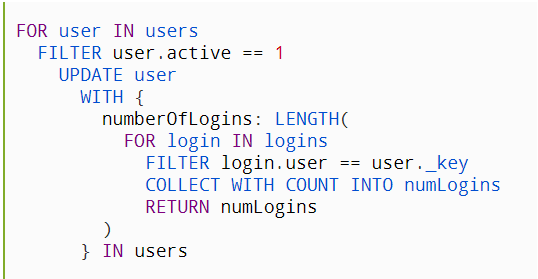
Cập nhật một hàng / tài liệu



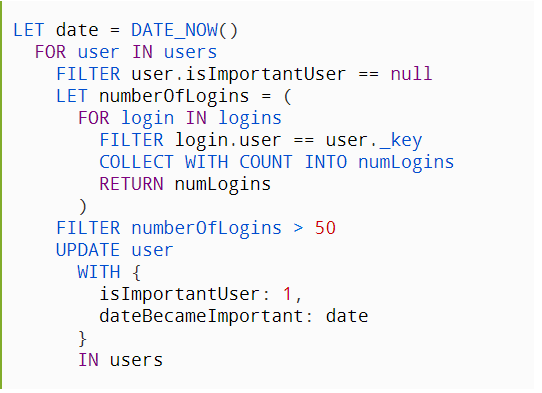
Thêm một cột / thuộc tính mới với giá trị mặc định



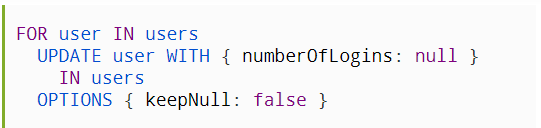
Thêm một cột / thuộc tính mới với giá trị được tính



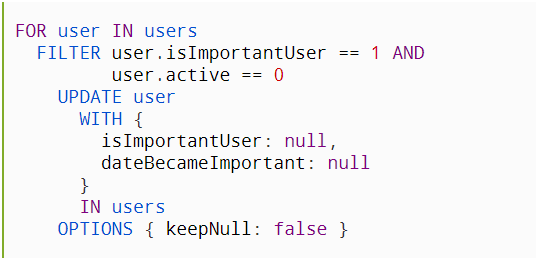
Thêm các cột / thuộc tính tùy chọn



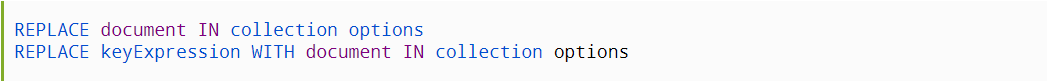
Xóa một cột / thuộc tính



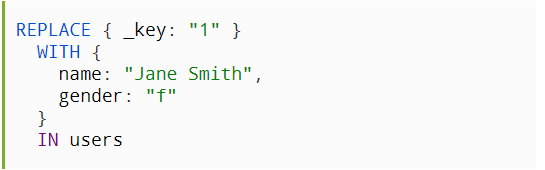
Chỉ xóa một cột / thuộc tính cho một số hàng / tài liệu



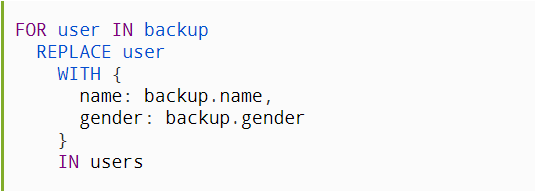
**THAY THẾ**



Thay thế một hàng / tài liệu



Thay thế nhiều hàng / tài liệu trong một bảng / bộ sưu tập

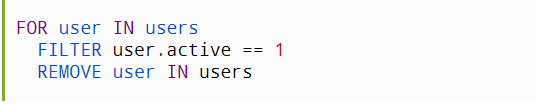


XÓA

Xóa một hàng / tài liệu

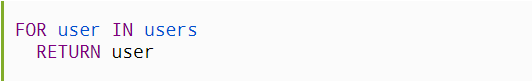


Xóa nhiều hàng / tài liệu

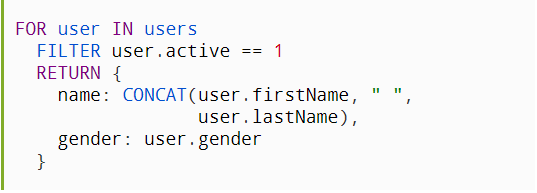


Truy vấn

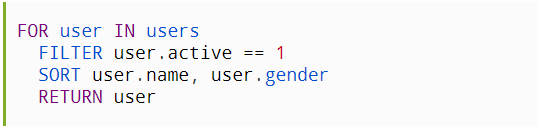
Chọn tất cả các hàng / tài liệu từ một bảng / bộ sưu tập, với tất cả các cột / thuộc tính



Lọc các hàng / tài liệu từ một bảng / bộ sưu tập, với phép chiếu

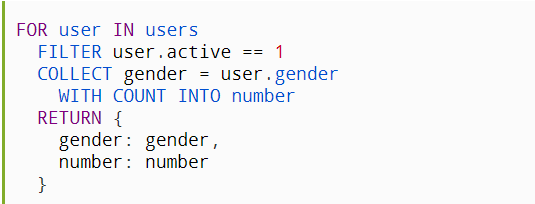


Sắp xếp các hàng / tài liệu từ một bảng / bộ sưu tập

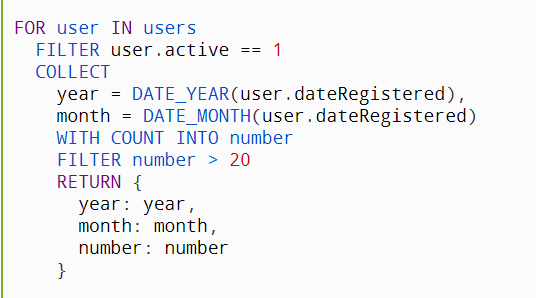


AGGREGATION

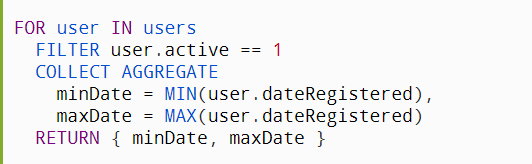
Đếm hàng / tài liệu trong bảng



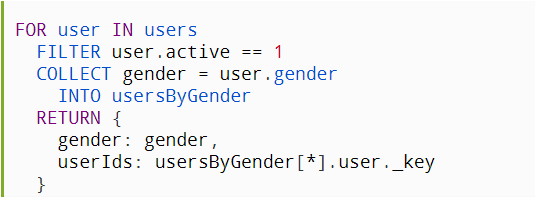
Nhóm các hàng / tài liệu trong một bảng / bộ sưu tập



Tính toán tối thiểu, tối đa các hàng / tài liệu trong một bảng / bộ sưu tập

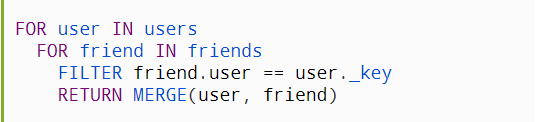


Xây dựng danh sách ngang

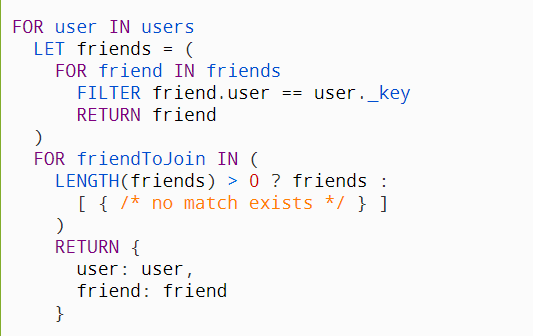


THAM GIA

Inner join



Outer join



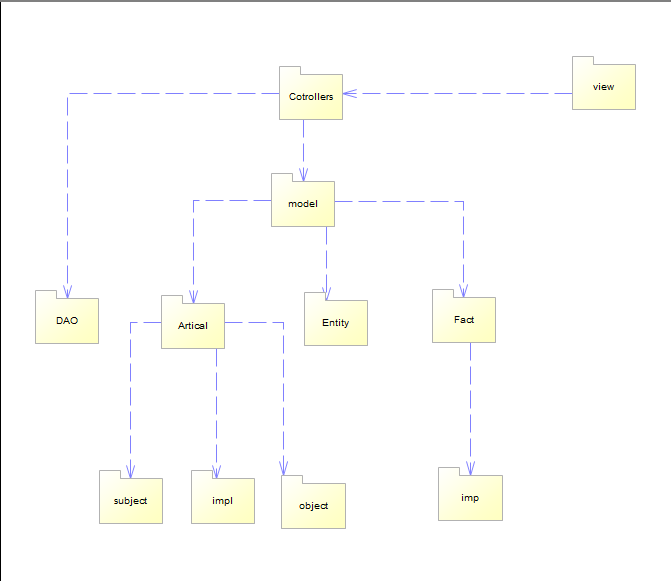
# **II. Mô hình dữ liệu**

Trong bài tập lớn của chúng em, chúng em lựa chọn giải pháp lưu trữ các dữ liệu dưới dạng document. Data về các Article, Fact, và Entities được lưu trữ vào 3 collection riêng biệt.

Mô hình dữ liệu dạng document là một ưu điểm của của NoSQL, thay vì sử dụng nhiều bảng cho mỗi entity khác nhau, người dùng chỉ cần 1 bảng để lưu trữ tất cả. Bù lại việc insert dữ liệu mất khá nhiều thời gian, giảm hiệu năng của hệ thống.

**III.Chi tiết về chương trình**

Biểu đồ sự phụ thuộc giữa các lớp:



Chương trình gồm 4 package chính:

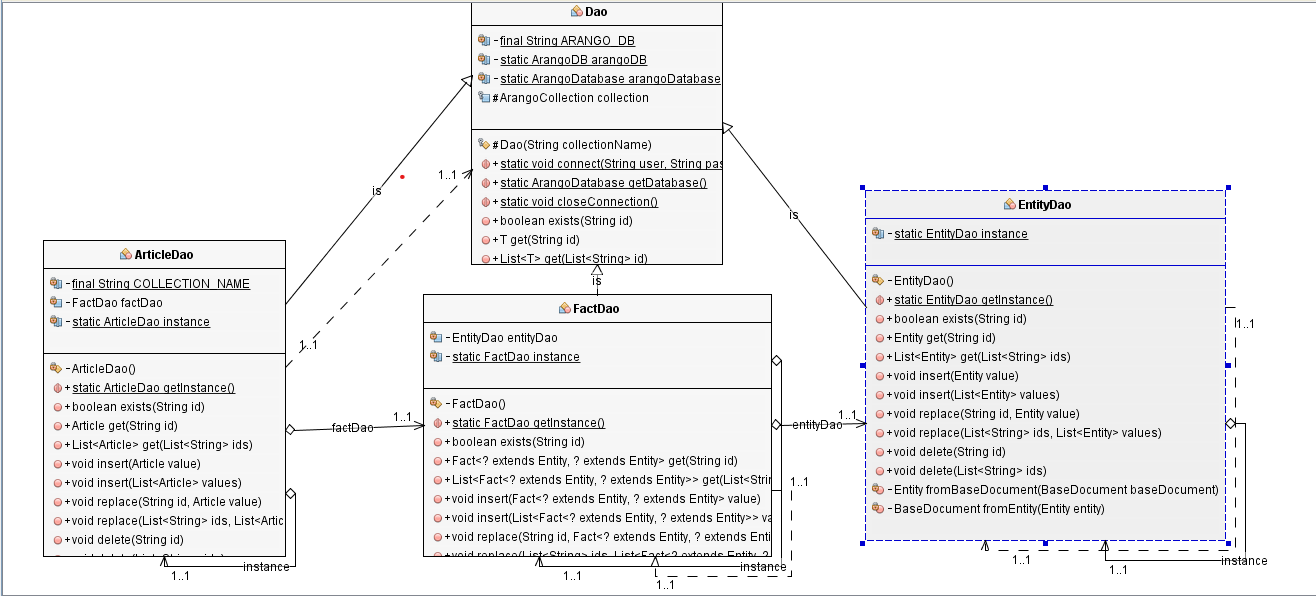
* Dao
* Entities
* Fact
* Model

**II.1.Package DAO:**

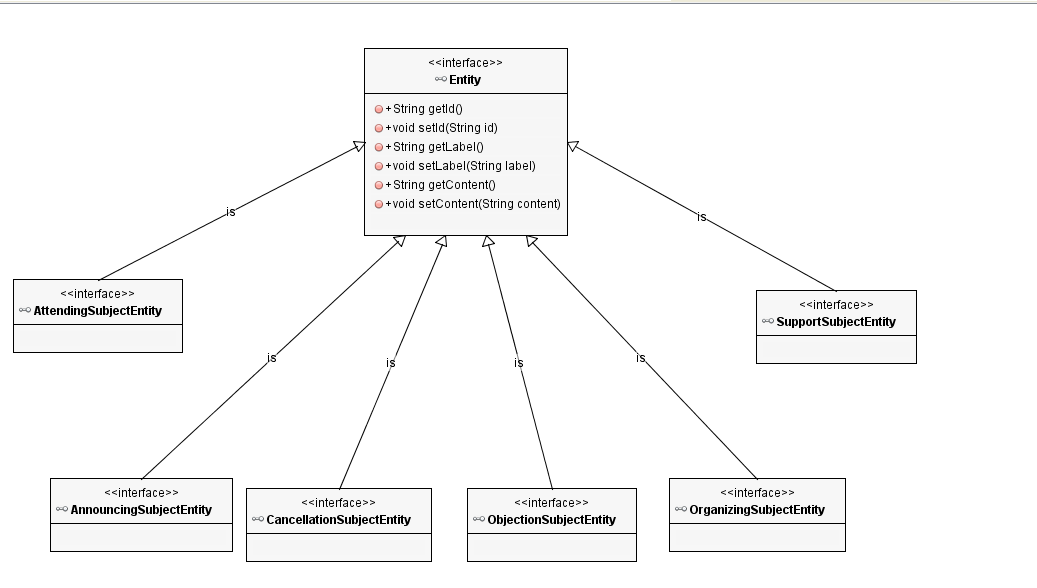
-Tạo conection đến database.

-Các lớp thuộc package dao đều kế thừa từ class dao.

-Tương tác giữa ứng dụng và database bao gồm đọc và ghi dữ liệu.



**II.2.Package Entities:**



Package lưu các lớp thực thể, gồm các lớp thực thể Person, Location, Event, Organization, Time, Country, AEntity. Các lớp đều được kế thừa từ lớp AEntity. Đối với mỗi lớp thực thể sẽ có các phương thức get/set.

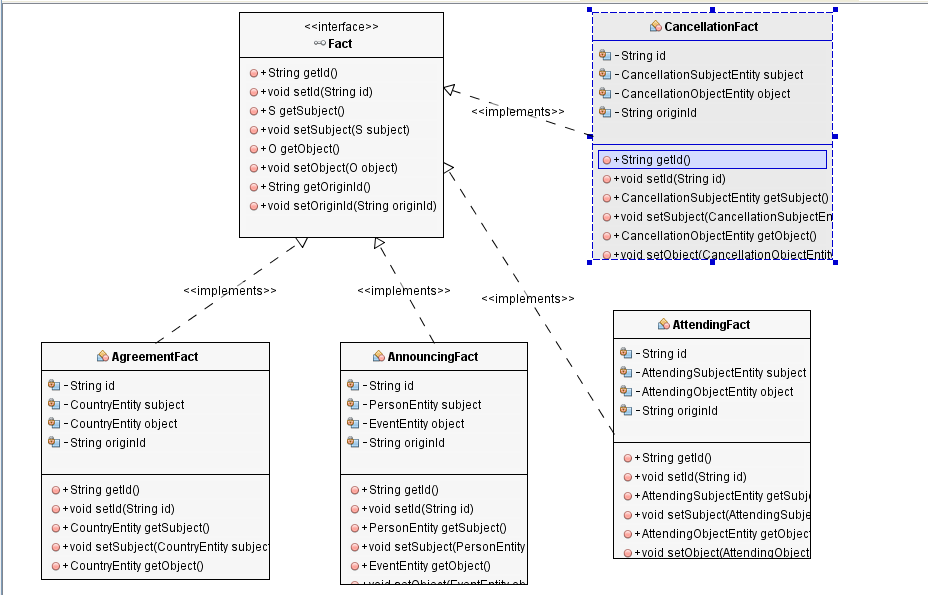
|  |  |  |
| --- | --- | --- |
| class | Thuộc tính | Phương thức |
| CountryEntity | population, gdp, area | 1. Contructor Country  2. (Override)HashMap<Object, Object> getListProperties(): thêm các thuộc tính population, gdp, area  3.các phương thức get/set |
| EventEntity | Kế thừa các thuộc tính của AEntity | 1.Contructor Event |
| LocationEntity | coordinates | 1. Contructor Location  2. (Override)HashMap<Object, Object> getListProperties(): thêm các thuộc tính coordinates  3.các phương thức get/set |
| OrganizationEntity | foundingDay | 1. Contructor Organization  2. (Override)HashMap<Object, Object> getListProperties(): thêm các thuộc tính foundingDay  3.các phương thức get/set |
| PersonEntity | Age, birthday, sex | 1. Contructor person  2. (Override)HashMap<Object, Object> getListProperties(): thêm các thuộc tính age, birthday, sex  3.các phương thức get/set |
| TimeEntity | dateTime | 1. Contructor Time  2. (Override)HashMap<Object, Object> getListProperties(): thêm các thuộc tính dateTime  3.các phương thức get/set |

**II.3.Package Fact:**

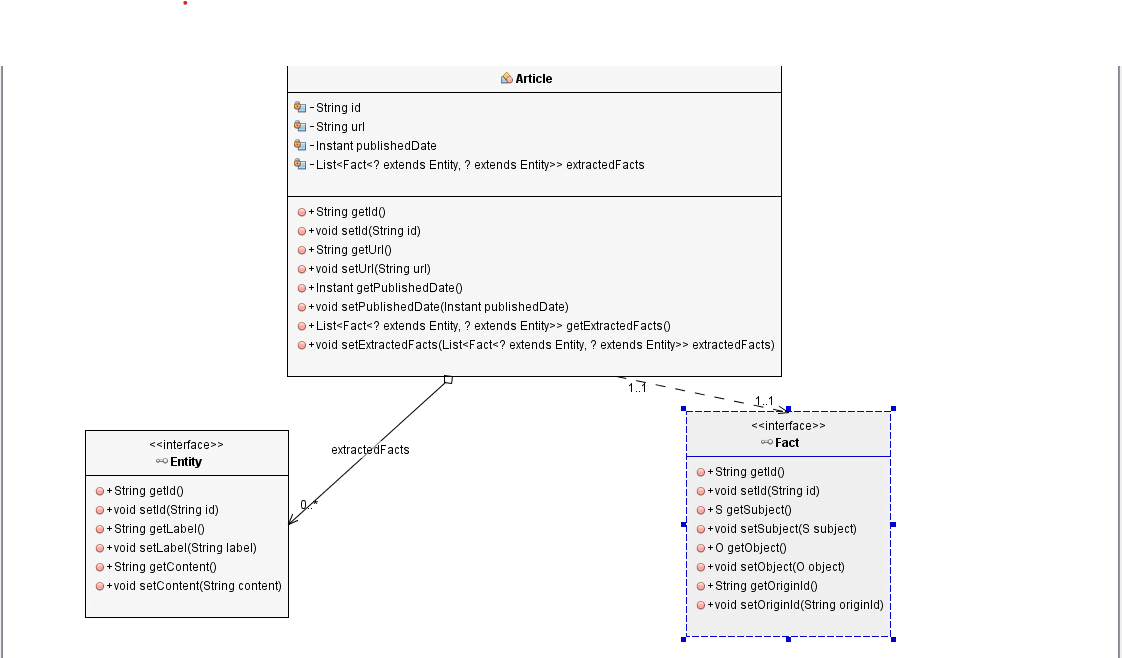
-Các lớp đều được kế thừa từ class Fact.

-Mỗi lớp đều bao gồm các phương thức Id, Subject, Object.

-Các phương thức của lớp: gồm các get và set.



**II.4.Package model:**



**IV.Kết quả kiểm thử hiệu năng:**

**IV.1.Truy vấn cơ bản, nâng cao:**

-Truy vấn cơ bản: Trong file NormalQuery.txt

-Truy vấn nâng cao: Trong file AdvanceQuery.txt

**IV.2.Ưu điểm nhược điểm công nghệ:**

**- Ưu điểm:**

ArangoDB là một cơ sở dữ liệu đa mô hình nguyên gốc. Bạn có thể lưu trữ dữ liệu của mình dưới dạng cặp khóa / giá trị, biểu đồ hoặc tài liệu và truy cập bất kỳ hoặc tất cả dữ liệu của bạn bằng một ngôn ngữ truy vấn khai báo. Bạn có thể kết hợp các mô hình khác nhau trong một truy vấn. Và, do cách tiếp cận đa mô hình nguyên gốc của nó, bạn có thể xây dựng các ứng dụng hiệu suất cao và mở rộng theo chiều ngang với cả ba mô hình dữ liệu.

+Hợp nhất

+Giảm độ phức tạp hoạt động

+**Tính nhất quán dữ liệu mạnh**

**+Tổng chi phí sở hữu thấp hơn**

**+Giao dịch**

**+Đa mô hình**

-Nhược điểm:

+Vẫn không phổ biên lắm, nên thực tế không thể tìm thấy các nhà phát triển kinh nghiệm trong ArrangoDB

+Viết tương đối chậm