

FragRDF: Um Fragmentador de dados RDF baseado em Esquemas

Vinicius Gasparini

viniciuszeiko@gmail.com

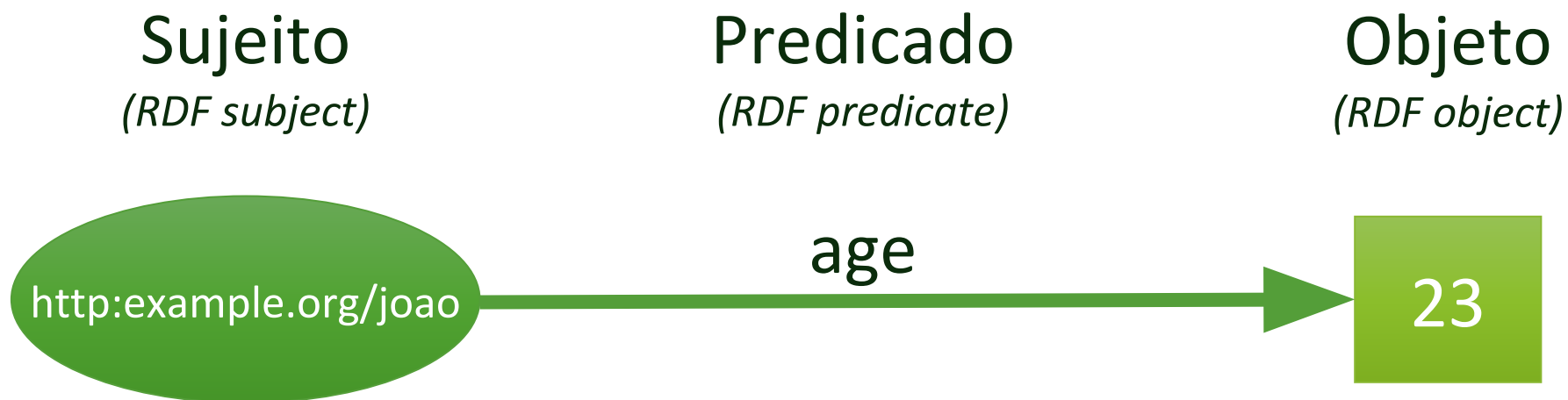
Rebeca Schroeder, PhD

rebeca.schroeder@udesc.br

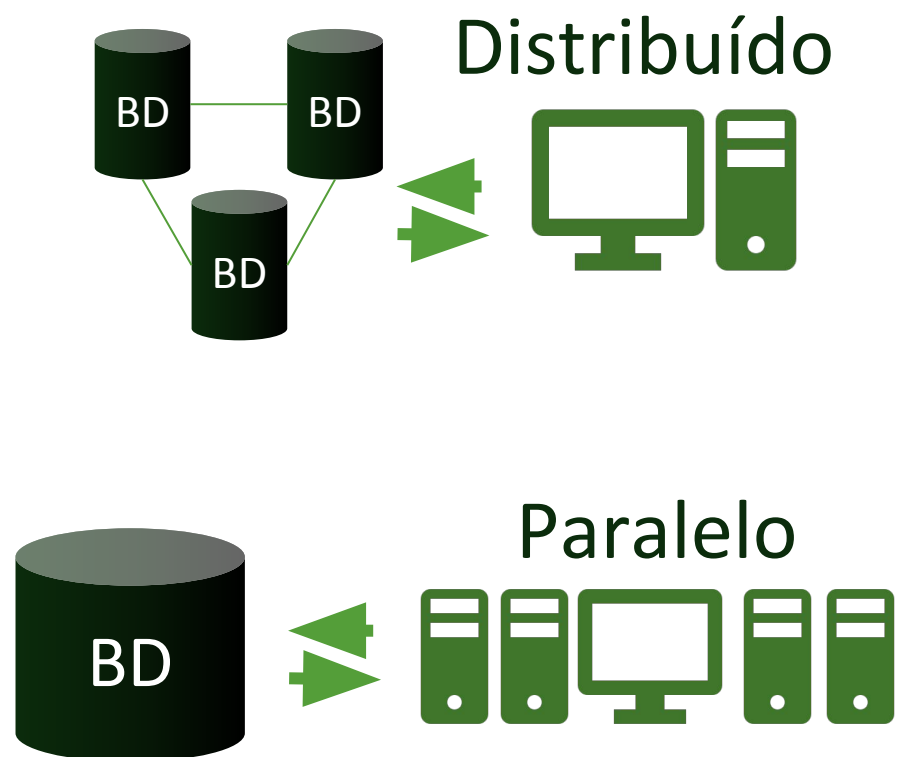
Grupo de Banco de Dados e Engenharia de Software
Departamento de Ciência da Computação
Centro de Ciências Tecnológicas Universidade do Estado de
Santa Catarina

2018

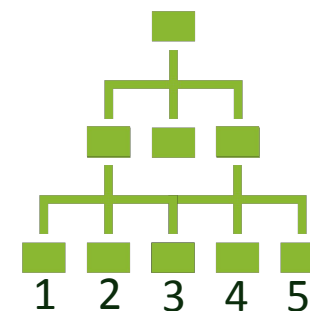
Introdução



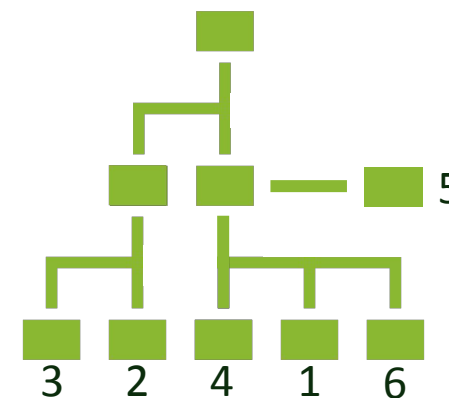
Introdução



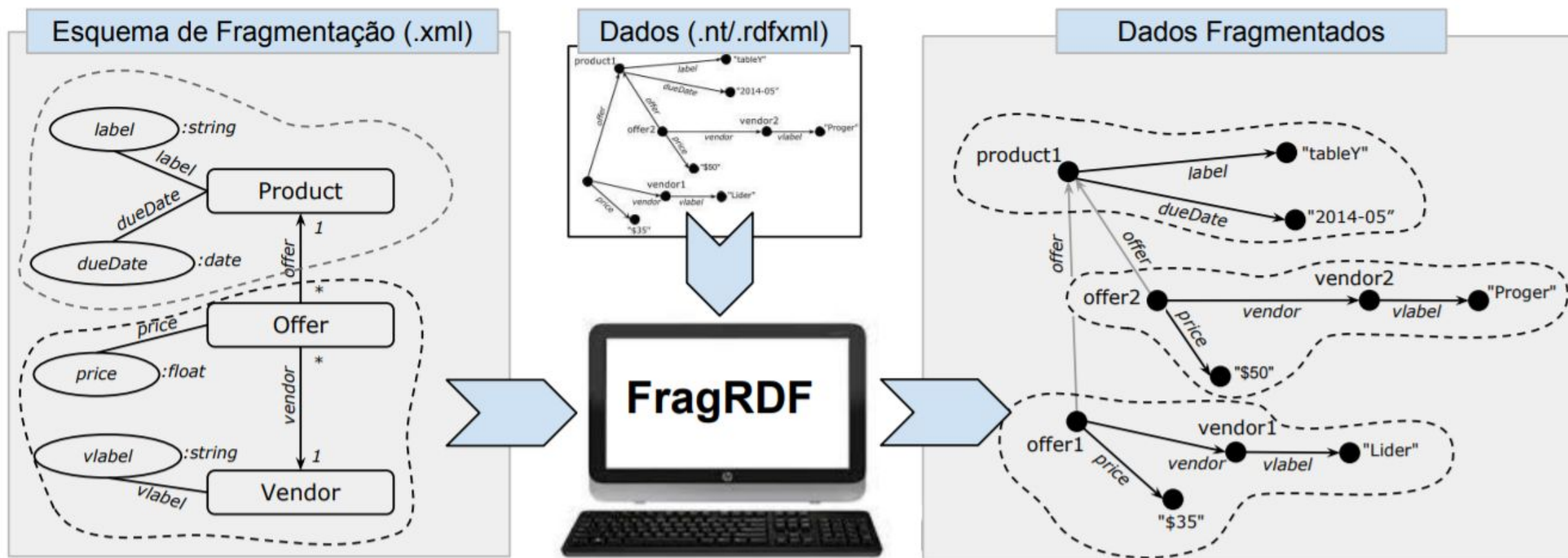
Estrutura de dados



Carga de trabalho



Introdução



Fonte: Gasparini e Schroeder, 2018

Sumário

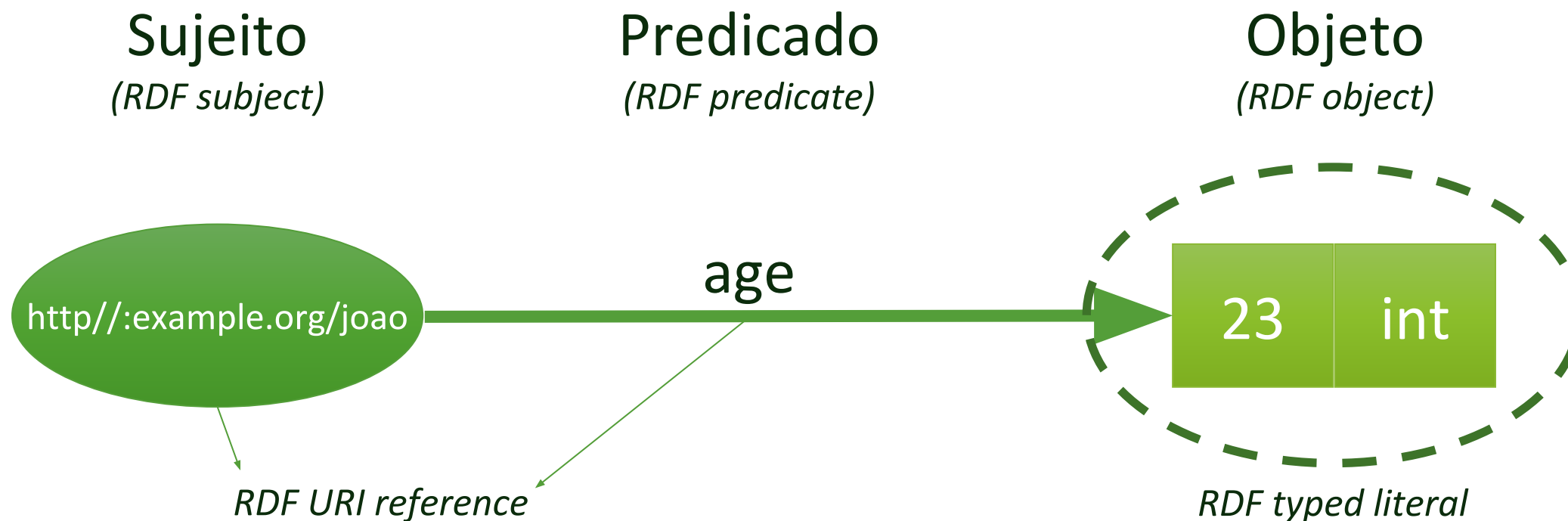
RDF e Métodos de Fragmentação

FragRDF

Experimentos

Conclusões

RDF - *Resource Description Framework*



RDF - *Resource Description Framework*

Sujeito (<i>RDF subject</i>)	Predicado (<i>RDF predicate</i>)	Objeto (<i>RDF object</i>)
<http://example.org/joao>	<http://example.org/ns#age>	"23" <http://example.org/int>

RDF URI reference

http://example.org

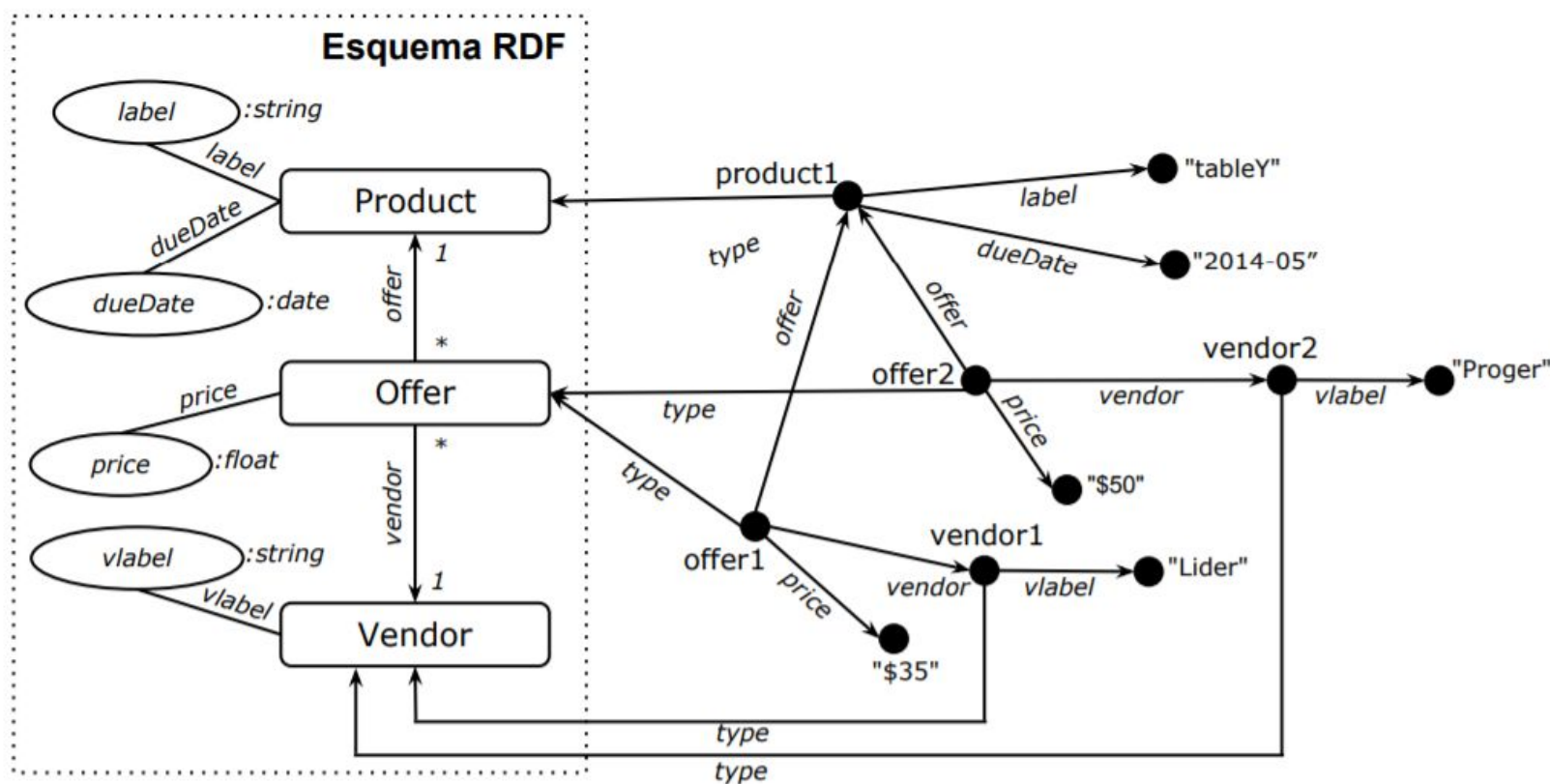
RDF typed literal

"23"

http://example.org/int

RDF - *Resource Description Framework*

Esquema RDF no formato de Grafo



Métodos de Fragmentação

(a) *Triple Store* RDF

Sujeito	Propriedade	Objeto
ID1	type	BookType
ID1	title	"XYZ"
ID1	author	"Fox,Joe"
ID1	copyright	"2001"
ID2	type	CDType
ID2	title	"ABC"
ID2	artist	"Orr,Tim"
ID2	copyright	"1985"
ID2	language	"French"
ID3	type	BookType
ID3	title	"MNO"
ID3	language	"English"
ID4	type	DVDType
ID4	title	"DEF"

(b) Fragmentação Vertical

Sujeito	type	title	copyright
ID1	BookType	"XYZ"	"2001"
ID2	CDType	"ABC"	"1985"
ID3	BookType	"MNP"	NULL
ID4	DVDType	"DEF"	NULL

Sujeito	Propriedade	Objeto
ID1	author	"Fox,Joe"
ID2	artist	"Orr,Tim"
ID2	language	"French"
ID3	language	"English"

(c) Fragmentação Híbrida

Classe: BookType

Sujeito	Title	Author	copyright
ID1	"XYZ"	"Fox,Joe"	"2001"
ID3	"MNP"	NULL	NULL

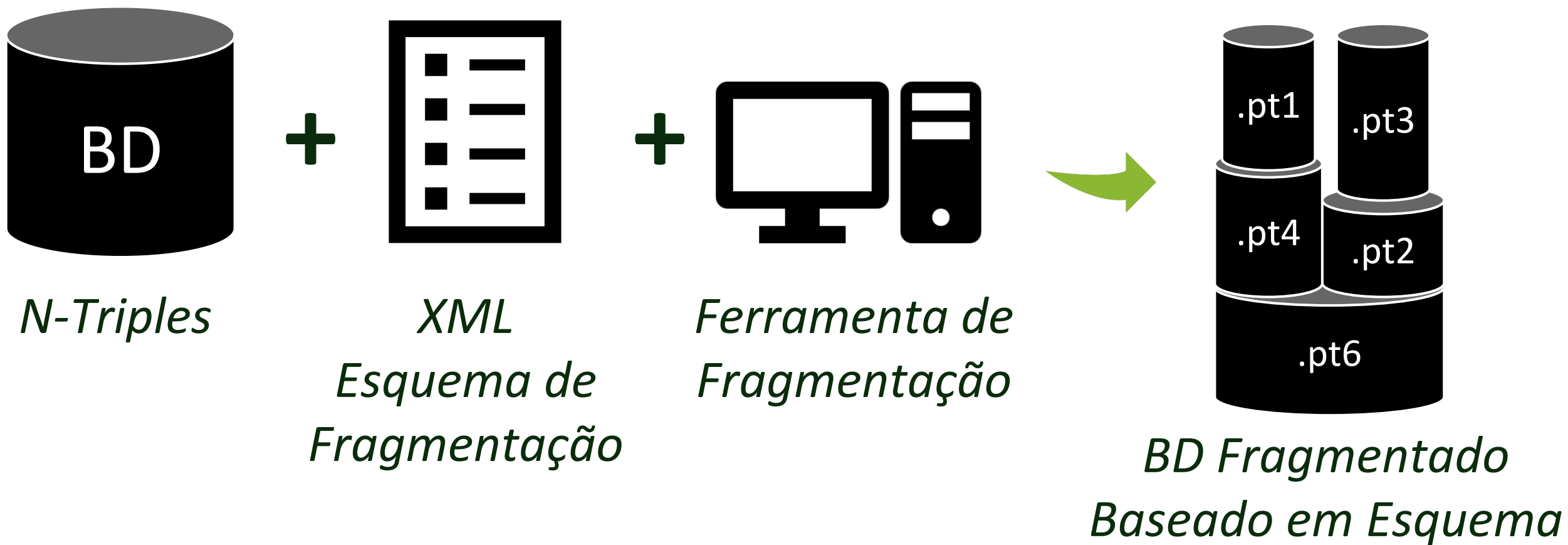
Classe: CDType

Sujeito	Title	Author	copyright
ID2	"ABC"	"Orr,Tim"	"1985"

Sujeito	Propriedade	Objeto
ID2	language	"French"
ID3	language	"English"
ID4	type	DVDType
ID4	title	"DEF"

Fonte: Adaptado de [Abadi et al. 2009]

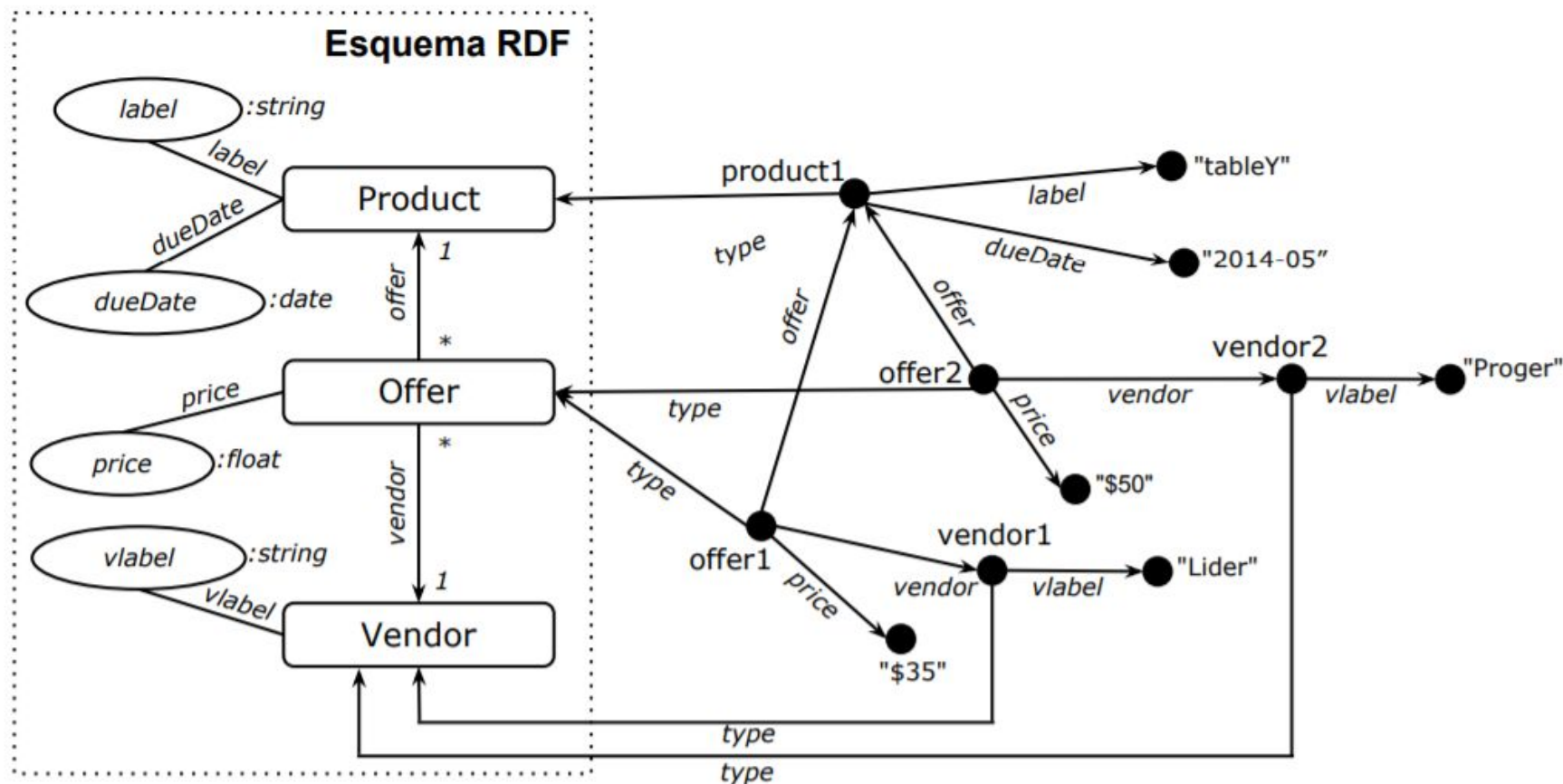
Métodos de Fragmentação



FragRDF

“*FragRDF* é uma ferramenta desenvolvida para fragmentar um conjunto de dados RDF de acordo com um esquema de fragmentação.”
(Gasparini e Schroeder, 2018)

RDF no formato de Grafo



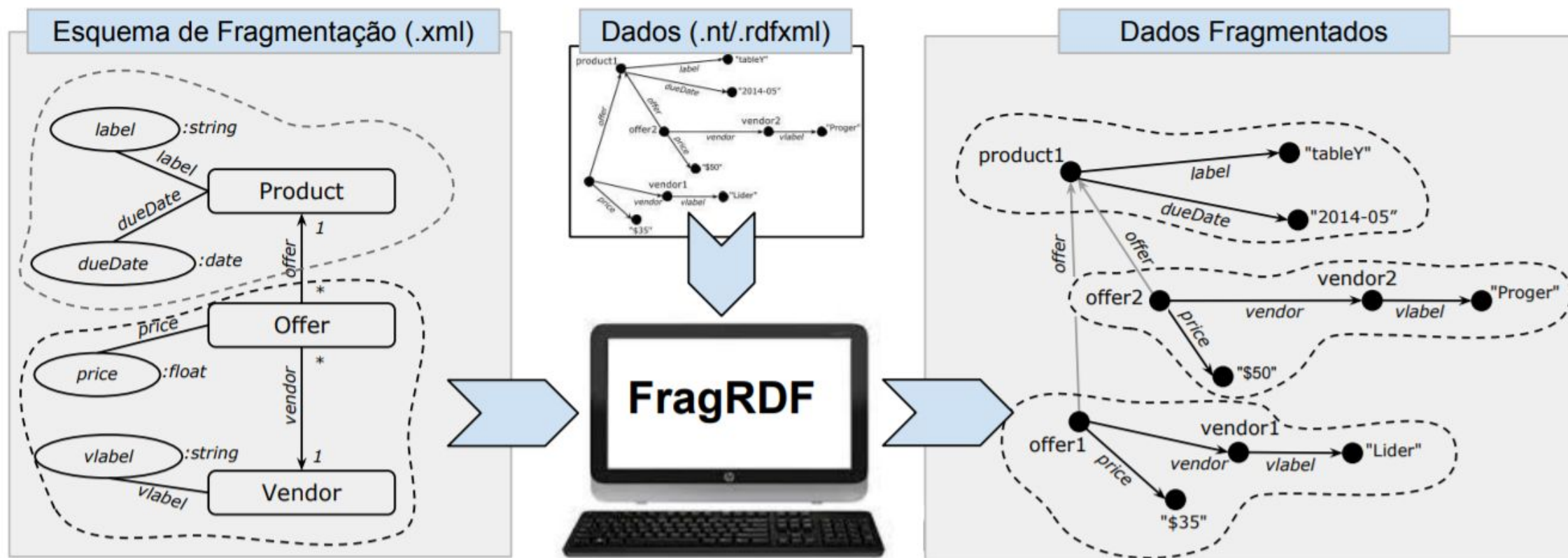
RDF no formato *N-Triples*

```
<http://www4.wiwiss.de/Product1> <http://www4.wiwiss.de/#type> <http://www4.wiwiss.de/Product> .  
<http://www4.wiwiss.de/Product1> <http://www4.wiwiss.de/#label> "tableY" .  
<http://www4.wiwiss.de/Product1> <http://www4.wiwiss.de/#dueDate> "2014-05" <http://www4.wiwiss.de/#date> .  
<http://www4.wiwiss.de/Offer1> <http://www4.wiwiss.de/#type> <http://www4.wiwiss.de/Offer> .  
<http://www4.wiwiss.de/Offer1> <http://www4.wiwiss.de/product> <http://www4.wiwiss.de/Product1> .  
<http://www4.wiwiss.de/Offer1> <http://www4.wiwiss.de/vendor> <http://www4.wiwiss.de/Vendor1> .  
<http://www4.wiwiss.de/Offer1> <http://www4.wiwiss.de/#price> "35.00" <http://www4.wiwiss.de/USD> .  
<http://www4.wiwiss.de/Vendor1> <http://www4.wiwiss.de/#type> <http://www4.wiwiss.de/Vendor> .  
<http://www4.wiwiss.de/Vendor1> <http://www4.wiwiss.de/#label> "Lider" .  
<http://www4.wiwiss.de/Offer2> <http://www4.wiwiss.de/#type> <http://www4.wiwiss.de/Offer> .  
<http://www4.wiwiss.de/Offer2> <http://www4.wiwiss.de/product> <http://www4.wiwiss.de/Product1> .  
<http://www4.wiwiss.de/Offer2> <http://www4.wiwiss.de/vendor> <http://www4.wiwiss.de/Vendor2> .  
<http://www4.wiwiss.de/Offer2> <http://www4.wiwiss.de/#price> "50.00" <http://www4.wiwiss.de/USD> .  
<http://www4.wiwiss.de/Vendor2> <http://www4.wiwiss.de/#type> <http://www4.wiwiss.de/Vendor> .  
<http://www4.wiwiss.de/Vendor2> <http://www4.wiwiss.de/#label> "Proger" .
```

Esquema de Fragmentação no formato XML

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <schemafrag> <!-- definitions for fragmentation schema -->
3   <items> <!-- map elements/attributes to fragments -->
4     <item name="label" entity="Product" frag="1"/>
5     <item name="dueDate" entity="Product" frag="1"/>
6     <item name="offer" entity="Offer" frag="2"/>
7     <item name="price" entity="Offer" frag="2"/>
8     <item name="vendor" entity="Offer" frag="2"/>
9     <item name="vlabel" entity="Vendor" frag="2"/>
10   </items>
11 </schemafrag>
```


Funcionamento do FragRDF



Introdução

O estudo experimental a ser apresentado consistiu em fragmentar em uma única máquina diversos bancos de dados.

Num primeiro momento, variou-se o tamanho de um único arquivo de entrada. Em seguida foi realizado testes variando a quantidade de arquivos de entrada.

Berlin SPARQL Benchmark

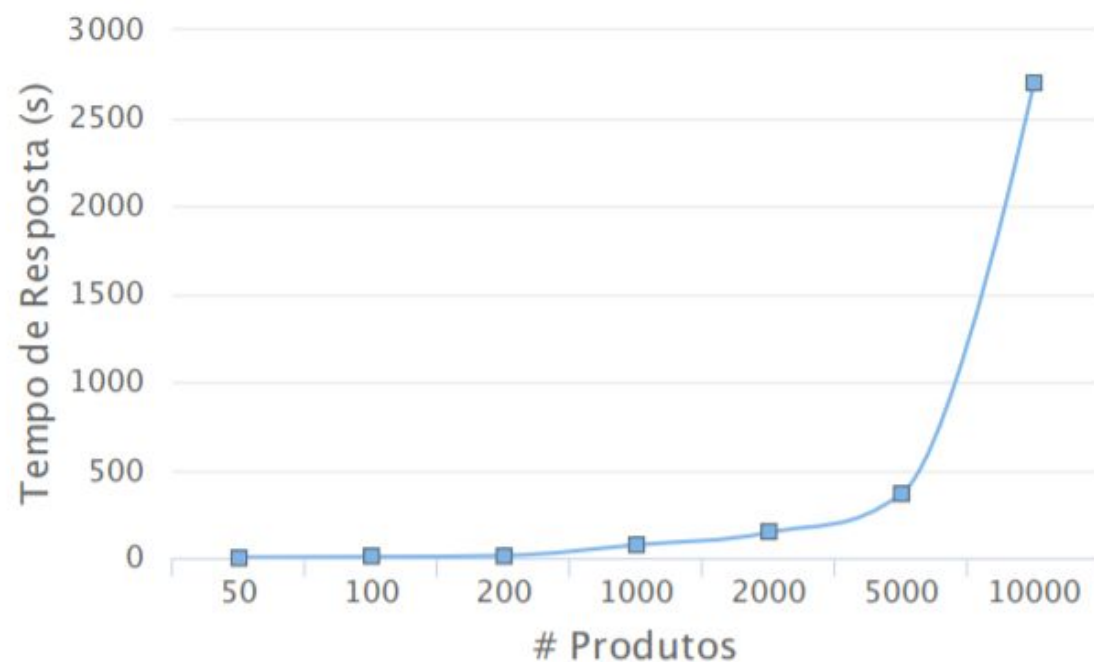
O BSBM é baseado em um caso de uso de um sistema de *e-commerce*. Para a geração das bases, o *benchmark* utiliza um fator de escala baseado no número de produtos a gerar.

Estatísticas das Bases Avaliadas

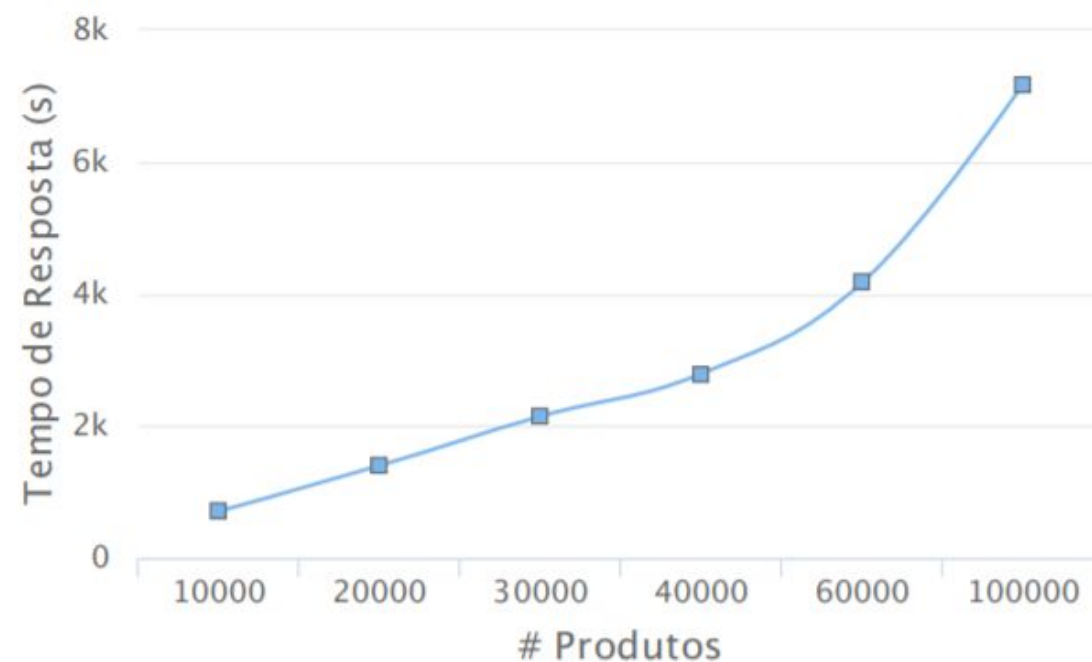
#Produtos	#Triplas	Tamanho (MB)	#Arquivos
50	22.729	5,71	1
100	40.377	10,22	1
200	75.550	19,21	1
1.000	374.911	95,72	1
2.000	725.305	186,04	1
5.000	1.809.874	465,19	1
10.000	3.564.773	919,09	1
10.000	3.564.773	897,62	5
20.000	7.073.571	1.790,02	10
30.000	10.628.484	2.693,76	15
40.000	14.172.772	3.594,61	20
60.000	21.198.599	5.384,19	30
100.000	35.272.182	8.973,50	50

Estatísticas

Experimento 1



Experimento 2



Resultados

Os resultados deste estudo experimental mostram que FragRDF consegue gerar bases de até 35.272.182 triplas em uma única máquina. Não foi possível o teste com bases de tamanhos superiores em virtude do limite de memória da máquina utilizada.

Conclusão

Foi apresentada a ferramenta FragRDF que é capaz de fragmentar conjuntos de dados RDF a partir de um esquema de fragmentação pré-estabelecido. O estudo experimental demonstrou que a ferramenta é capaz de gerar bases fragmentadas de até 35.272.182 triplas em uma única máquina.

Trabalhos Futuros

Pretende-se prosseguir com uma versão paralela da ferramenta, de forma que sua capacidade de processamento se torne escalável. Além disto, disponibilizar mecanismos para controlar a redundância de dados nos fragmentos.

Dúvidas

Vinicius Gasparini

viniciuszeiko@gmail.com

Rebeca Schroeder, PhD

rebeca.schroeder@udesc.br