

Capítulo 7: SPARQL Processamento de consulta em nuvem



01/11/2016

Acadêmico: Maicou Moor Cândido de Lima
Professora: Carmem Satie Hara



Introdução

Introdução



Plataformas
RDF



AMAZON



Armazenamento
e Consulta



Implementação



Evolução
Experimental



Resumo e
Perspectiva

Neste capítulo concentra-se em recentes propostas de técnicas de processamento de **consulta distribuídas e paralelas** que são adequadas para infraestruturas da nuvem .



Introdução

Introdução



Plataformas
RDF



AMAZON



Armazenamento
e Consulta



Implementação



Evolução
Experimental



Resumo e
Perspectiva

Plataformas baseadas em nuvem liberam o desenvolvimento a aplicação para administrar o hardware e fornecer resistência a falhas, bem como dimensionamento elástico de alto e baixo recurso conforme a demanda.



Introdução

Introdução



Plataformas
RDF



AMAZON



Armazenamento
e Consulta



Implementação



Evolução
Experimental



Resumo e
Perspectiva

Há uma grande quantidade de dados RDF publicado na Web hoje em dia e constantemente **aumentam os dados abertos**, por exemplo: DBpedia, BBC, Open Government Data.



Introdução

Introdução



Plataformas
RDF



AMAZON



Armazenamento
e Consulta

Sistemas eficientes foram criados para lidar com **grandes volumes** de dados RDF em um **ambiente centralizado**, estando entre os mais conhecidos o RDF-3X.



Implementação



Evolução
Experimental



Resumo e
Perspectiva

Armazenamento de RDF Baseado em Nuvem



As primeiras classificações de plataformas existentes podem ser feitas de acordo com suas facilidades por baixo do armazenamento. Os sistemas existentes podem ser divididos nas **três categorias** a seguir:

Armazenamento de RDF Baseado em Nuvem



- Sistemas que utilizam "NoSQL" registro de **Key-value**;
- Sistemas baseados em um sistema de **arquivos distribuídos**;
- Sistemas baseados em **outras instalações** de armazenamento.

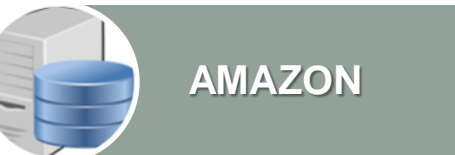
Armazenamento de RDF Baseado em Nuvem



Um segundo ângulo relevante da análise das plataformas RDF baseados em nuvem vem de sua estratégia para o processamento de consultas SPARQL. Pode-se identificar duas classe:

- sistemas baseados no paradigma de **programação paralela** do MapReduce;
 - sistema de tentativa de **reduzir** ou **evitar** completamente os passos do MapReduce.
-

RDF Raciocínio na Nuvem



As regras do RDF(S) de inferência são utilizadas para toda **vinculação** RDFS usando os trabalhos de MapReduce.

Repositório de dados em nuvem (Amazon)



Introdução



Plataformas RDF



AMAZON



Armazenamento e Consulta



Implementação



Evolução Experimental



Resumo e Perspectiva

Hospedagem de dados RDF em uma plataforma (distribuída e paralela), tal **arquitetura** pertencente Software as a Service (SaaS).

Amazon: Serviço de Armazenamento Simples (S3)



É um serviço web de armazenamento ideal para armazenar grandes **objetos** ou **arquivos**. O S3 armazena os dados em blocos chamados **buckets**.

Estrutura Base de Dados DynamoDB (Amazon)

 **Introdução**

 **Plataformas
RDF**

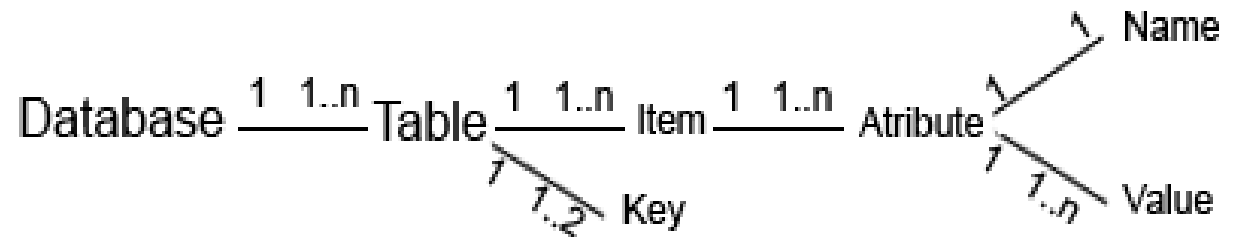
 **AMAZON**

 **Armazenamento
e Consulta**

 **Implementação**

 **Evolução
Experimental**

 **Resumo e
Perspectiva**



 **Introdução**

 **Plataformas RDF**

 **AMAZON**

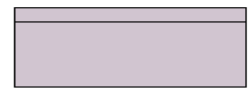
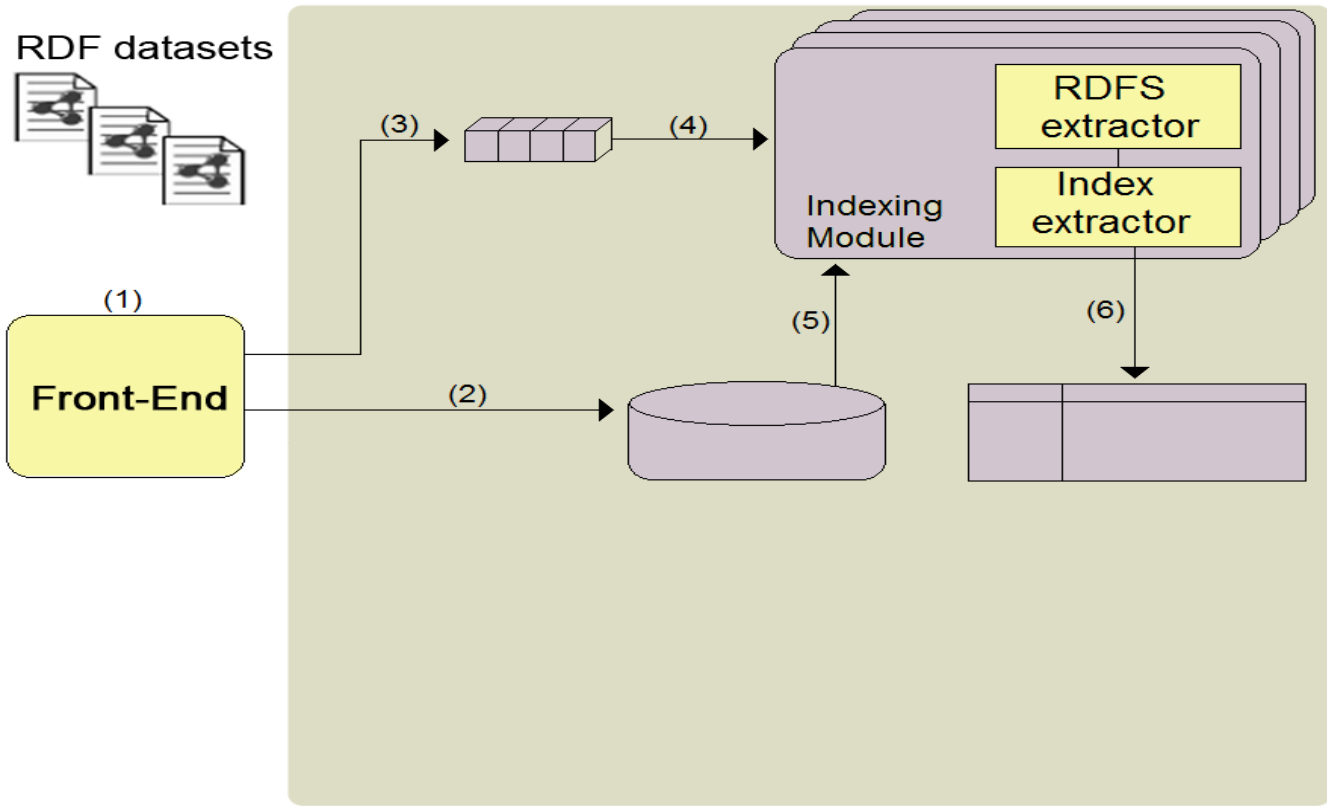
 **Armazenamento e Consulta**

 **Implementação**

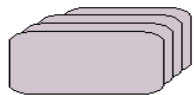
 **Evolução Experimental**

 **Resumo e Perspectiva**

Arquitetura Geral (AMAZON)



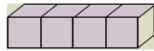
DynamoDB



Instâncias de nuvem de cálculo elástico



Serviço de Armazenamento Simples (S3)



Serviço de fila simples

Introdução

Plataformas RDF

AMAZON

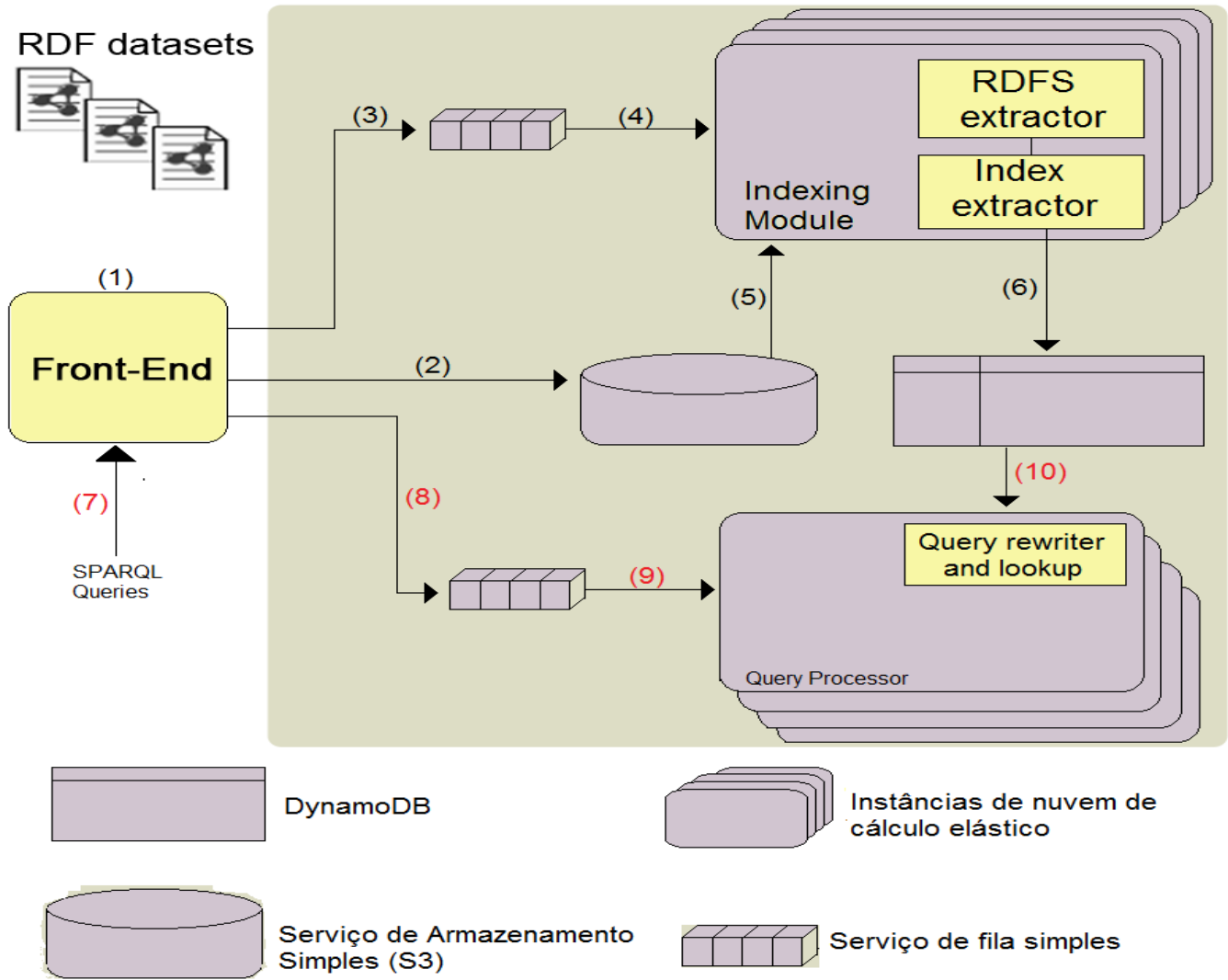
Armazenamento e Consulta

Implementação

Evolução Experimental

Resumo e Perspectiva

Arquitetura Geral (AMAZON)



Introdução

Plataformas RDF

AMAZON

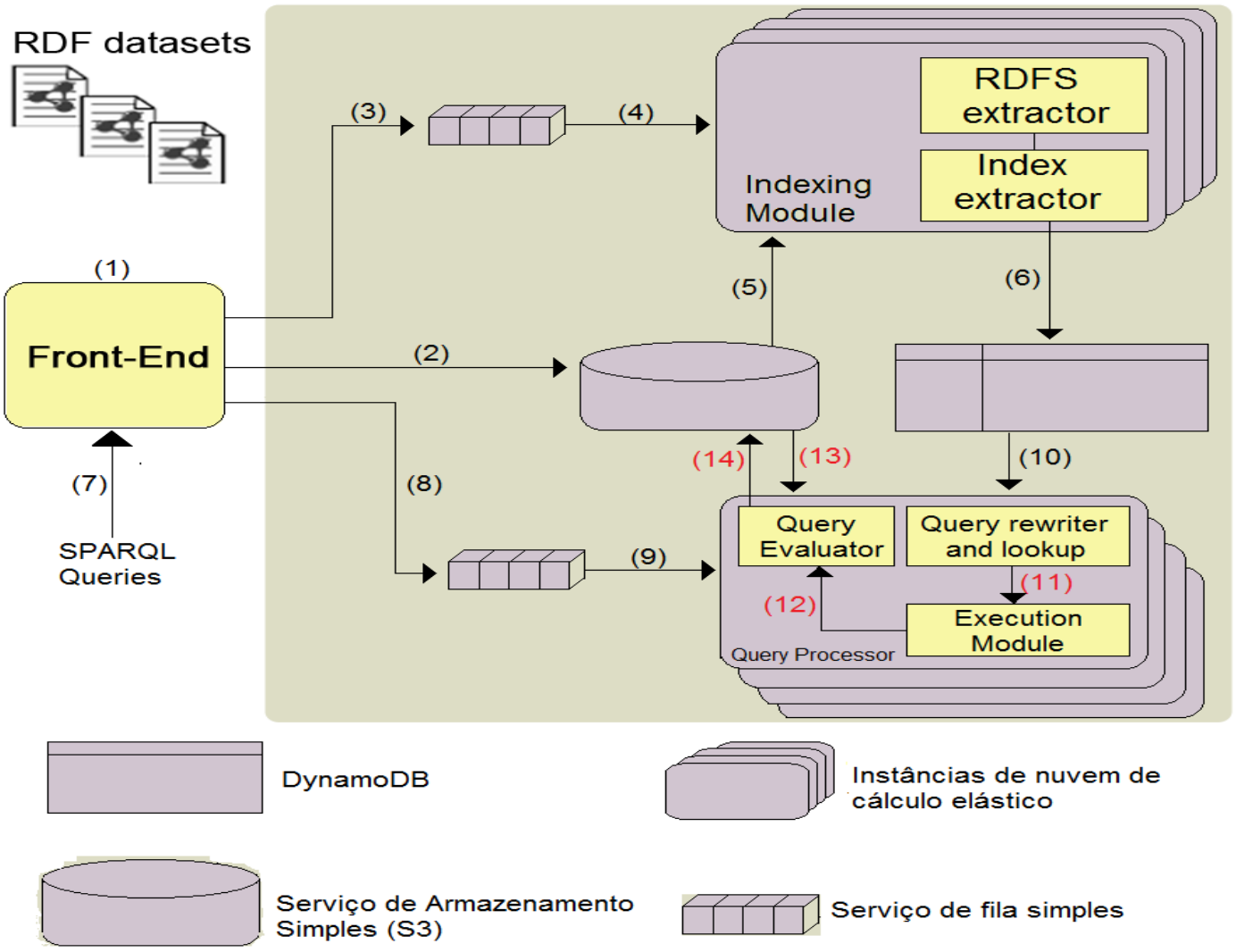
Armazenamento e Consulta

Implementação

Evolução Experimental

Resumo e Perspectiva

Arquitetura Geral (AMAZON)



Introdução

Plataformas RDF

AMAZON

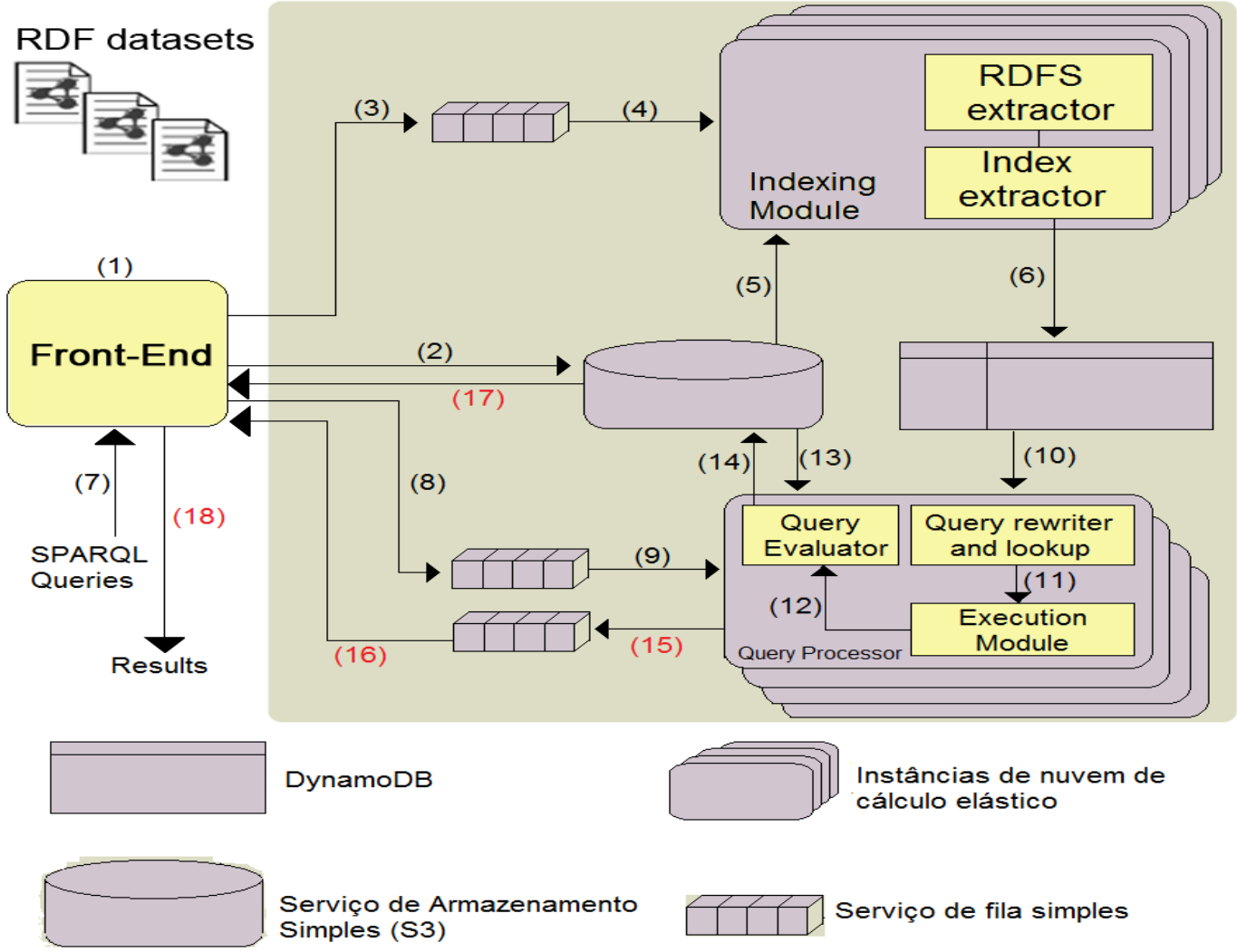
Armazenamento e Consulta

Implementação

Evolução Experimental

Resumo e Perspectiva

Arquitetura Geral (AMAZON)



SPARQL: Processamento de consulta em nuvem (AMADA).



Introdução



Plataformas
RDF



AMAZON



Armazenamento
e Consulta



Implementação



Evolução
Experimental



Resumo e
Perspectiva

Índice RDF

S	Subject	<u>S</u>	Token string for subject
P	Predicate	<u>P</u>	Token string for predicate
O	Object	<u>O</u>	Token string for object
U	URI	<u>N</u>	Constant table name
T	RDF Term		

SPARQL: Processamento de consulta em nuvem (AMADA).



Introdução



Plataformas
RDF



AMAZON



Armazenamento
e Consulta



Implementação



Evolução
Experimental



Resumo e
Perspectiva

Datasets	
D	Set of RDF datasets
ds	a dataset in D
$ ds $	#triples in ds
q	A SPARQL query
$ q $	#triple patterns in q

SPARQL: Processamento de consulta em nuvem (AMADA).



Introdução



Plataformas
RDF



AMAZON



Armazenamento
e Consulta



Implementação



Evolução
Experimental



Resumo e
Perspectiva


DynamoDB	
N	table name
K	item key
A	attr. name
V	attr. value

 **Introdução**

 **Plataformas RDF**

 **AMAZON**

 **Armazenamento e Consulta**

 **Implementação**

 **Evolução Experimental**

 **Resumo e Perspectiva**

Característica de consulta

Query	Struct	#tp	#c	#results	#d by RTS	#d by ATT	#d by ATS
Q1	simple	1	2	1	2	2	1
Q2	simple	1	2	433	3	3	3
Q3	simple	1	1	72829		2	2
Q4	star	2	4	1	19	19	19
Q5	star	3	4	2895	26	26	25
Q6	star	3	3	50686	34	34	34
Q7	star	4	4	42785	39	39	39
Q8	mix	5	6	2	9	9	9
Q9	mix	5	5	12	5	5	5

Introdução

Plataformas RDF

AMAZON

Armazenamento e Consulta

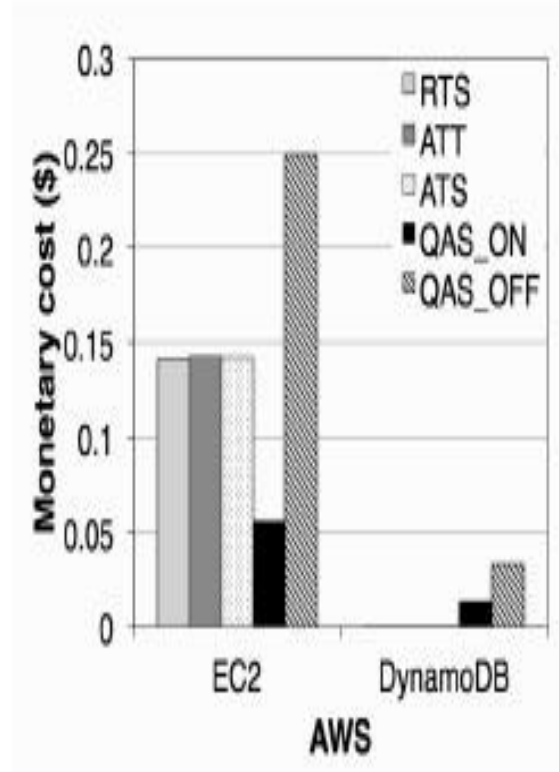
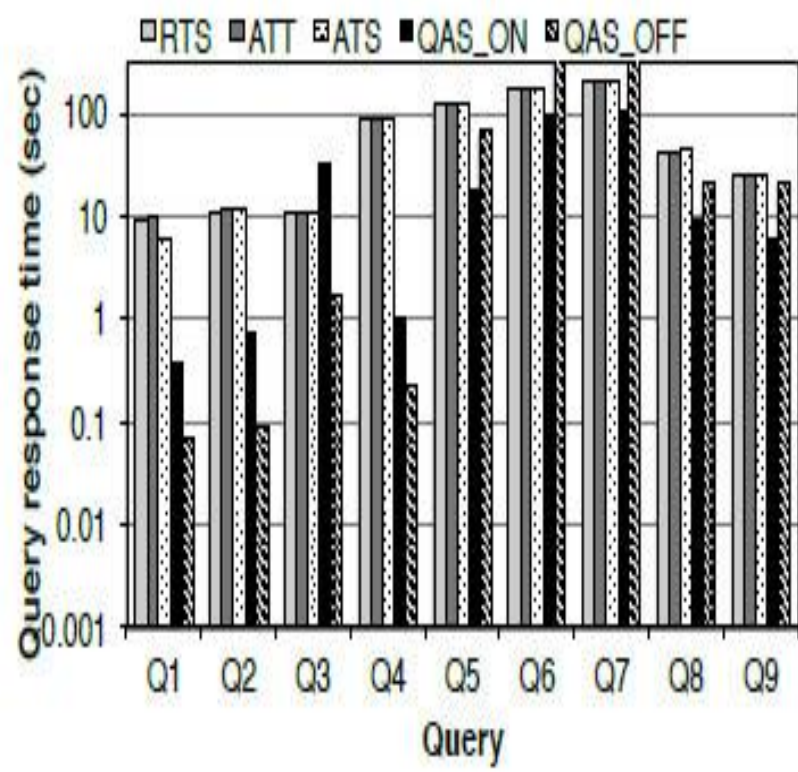
Implementação

Evolução Experimental

Resumo e Perspectiva

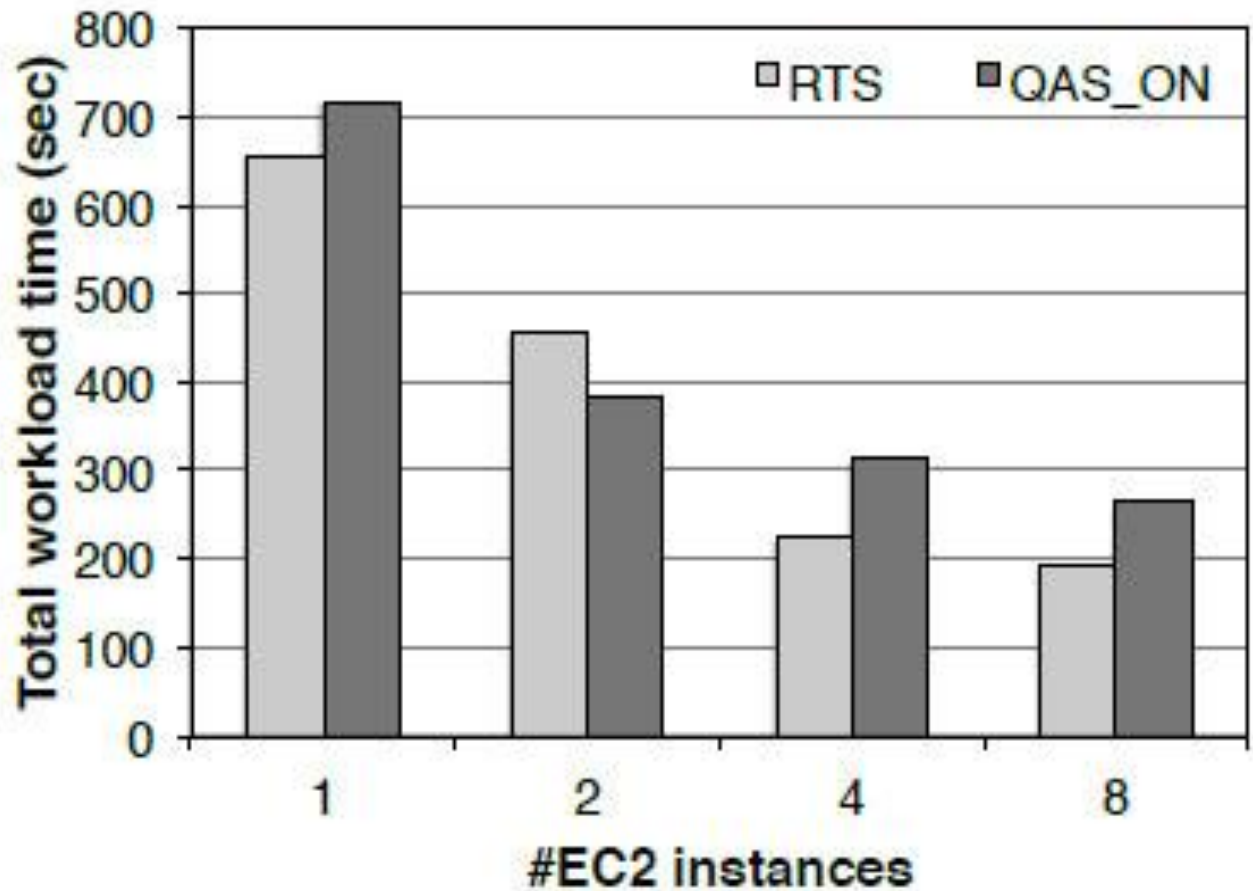
Tempo e Custo

SPARQL Query Processing in the Cloud



Escalabilidade

SPARQL Query Processing in the Cloud



Introdução

Plataformas RDF

AMAZON

Armazenamento e Consulta

Implementação

Evolução Experimental


Resumo e Perspectiva

 **Introdução**

 **Plataformas
RDF**

 **AMAZON**

 **Armazenamento
e Consulta**

 **Implementação**

 **Evolução
Experimental**

 **Resumo e
Perspectiva**

Resumo e Perspectiva

Este capítulo descreve uma arquitetura para armazenar e consultar dados RDF usando serviços em nuvem off-the-shelf, em particular a plataforma AMADA.

Pergunta

1) Sistemas eficientes foram criados para lidar com grandes volumes de dados RDF em um ambiente centralizado, estando entre os mais conhecidos:

A) RDF-3X

B) RDF-2X

C) RDF-1X



Pergunta

2) Qual é a representação correta das letras (**N**, **K**, **A** e **V**) dos quatros níveis de informações do DynamoDB:

A) **N** = item key, **K** = table name, **A** = attr. value, **V** = attr. name.

B) **N** = table name, **K** = item key, **A** = attr. name, **V** = attr. Value.

C) **N** = table name, **K** = attr. name, **A** = item key, **V** = attr. Value.

DynamoDB	
N	
K	
A	
V	

