Linked Data Fragments

Eduardo Vital Alencar Cunha

Agenda

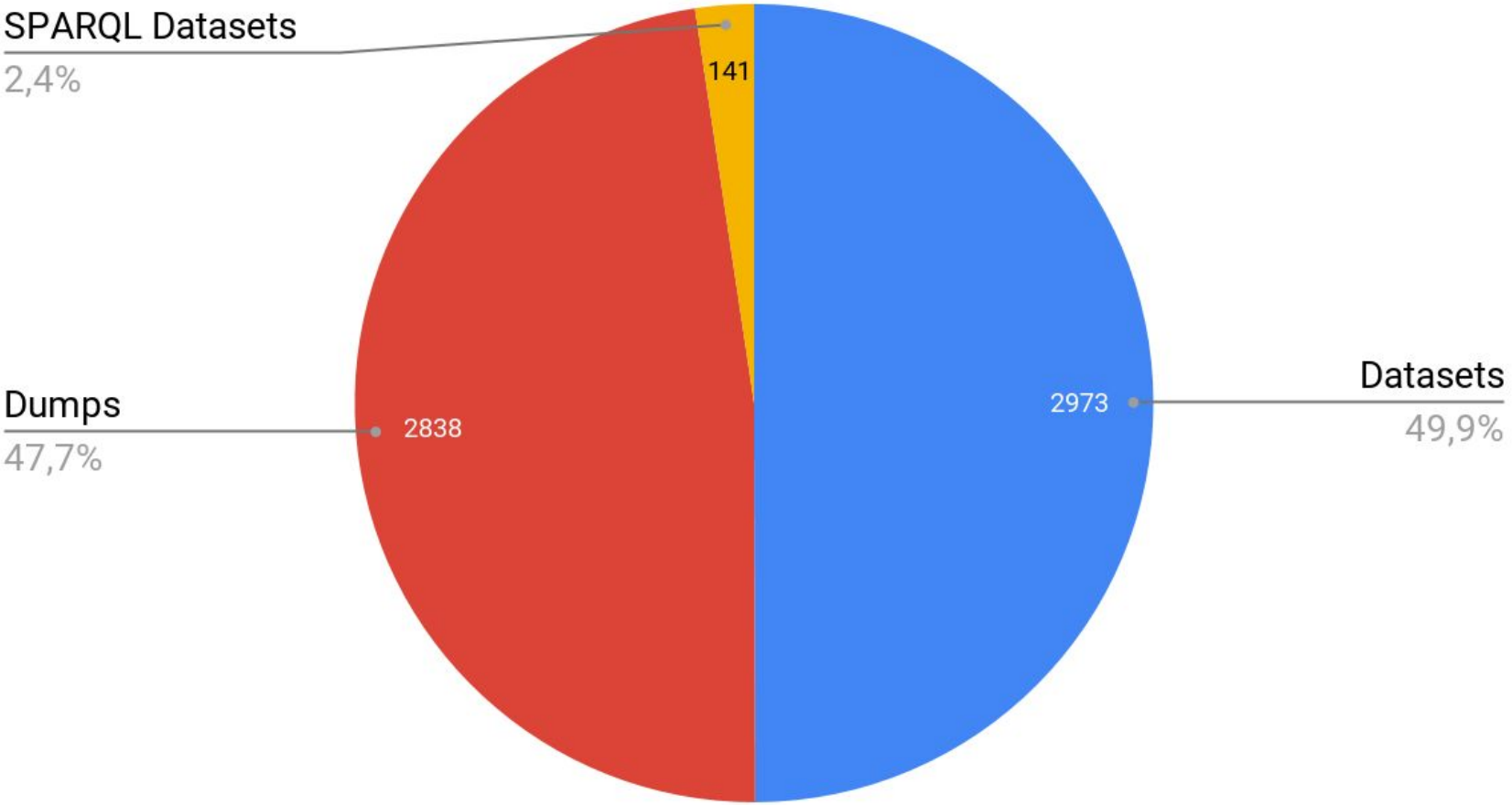
1. O acesso às bases de dados da Web Semântica
2. Os trade-offs da Web Semântica
3. Triple Pattern Fragments
4. Performance
5. Considerações

O acesso às bases de dados da Web Semântica

Acesso às bases de dados

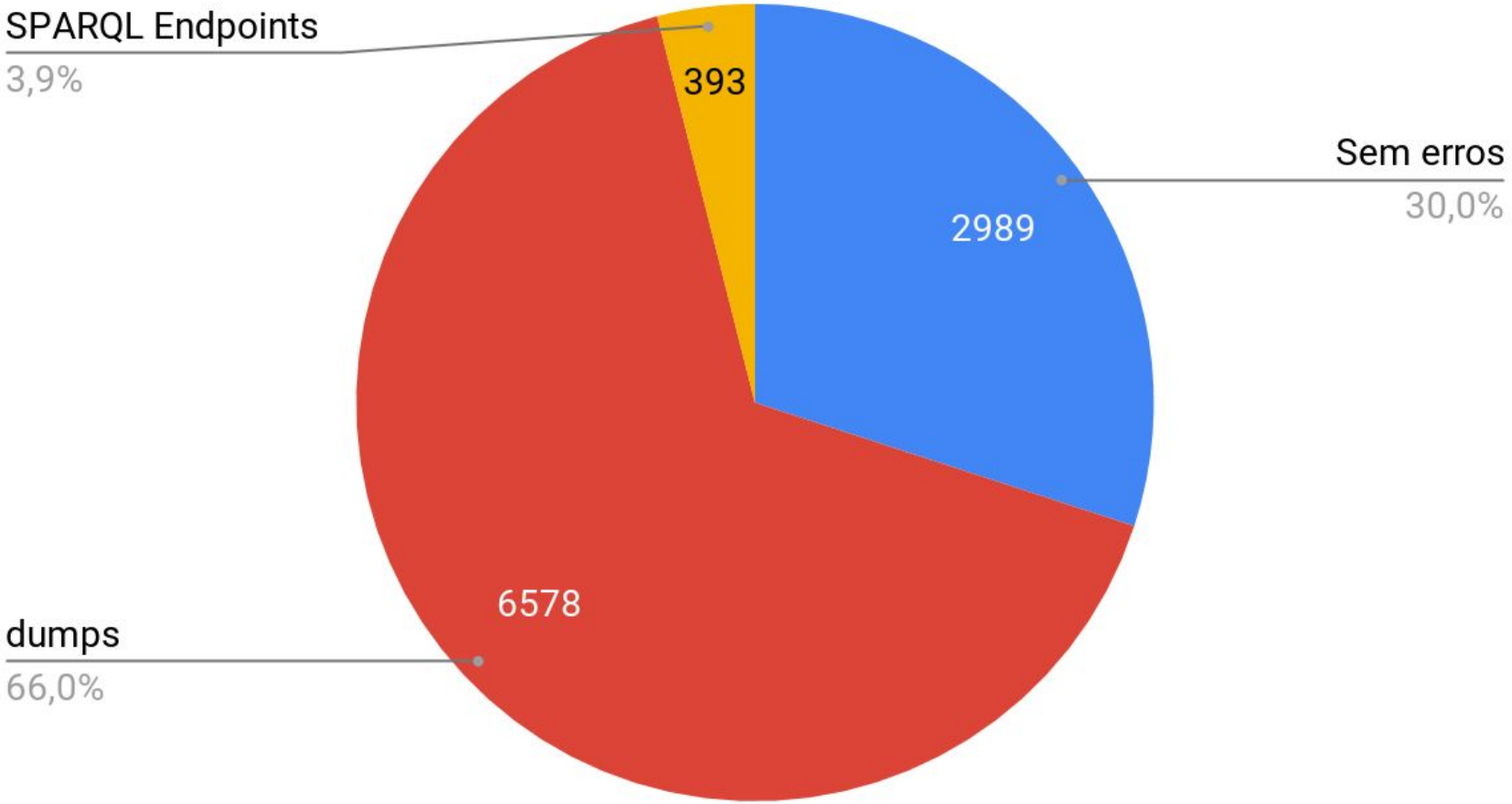
- Linked Data Documents
- Data Dumps
- SPARQL Endpoints

Linked Open Data [1]



Os *trade-offs* da Web Semântica

Linked Open Data Datasets com Erros [1]



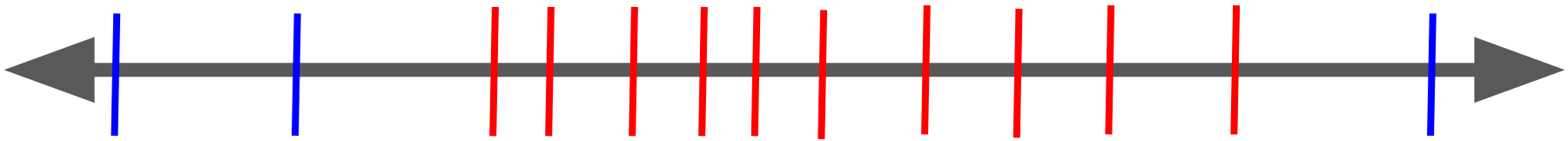
Principais Problemas da Web Semântica Atual [2]

- A maioria das bases de dados da Web Semântica não estão em formatos passíveis de consultas (*queryable*)
- As bases de dados disponíveis via SPARQL sofrem frequentemente com falta de disponibilidade

Trade-offs [3]

Dados desatualizados
Alto volume de tráfego de dados
Alto custo para o cliente
Baixo custo paro o servidor
Alta disponibilidade

Baixa disponibilidade
Alto custo para o servidor
Baixo custo paro o cliente
Baixo volume de tráfego de dados
Dados atualizados



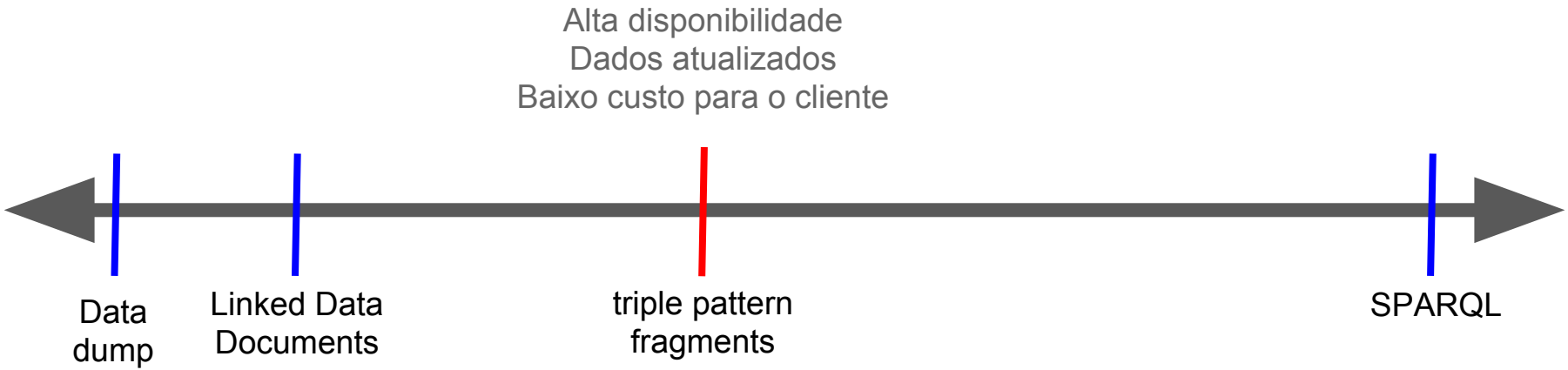
Data dump

Linked Data Documents

SPARQL

Triple Pattern Fragments

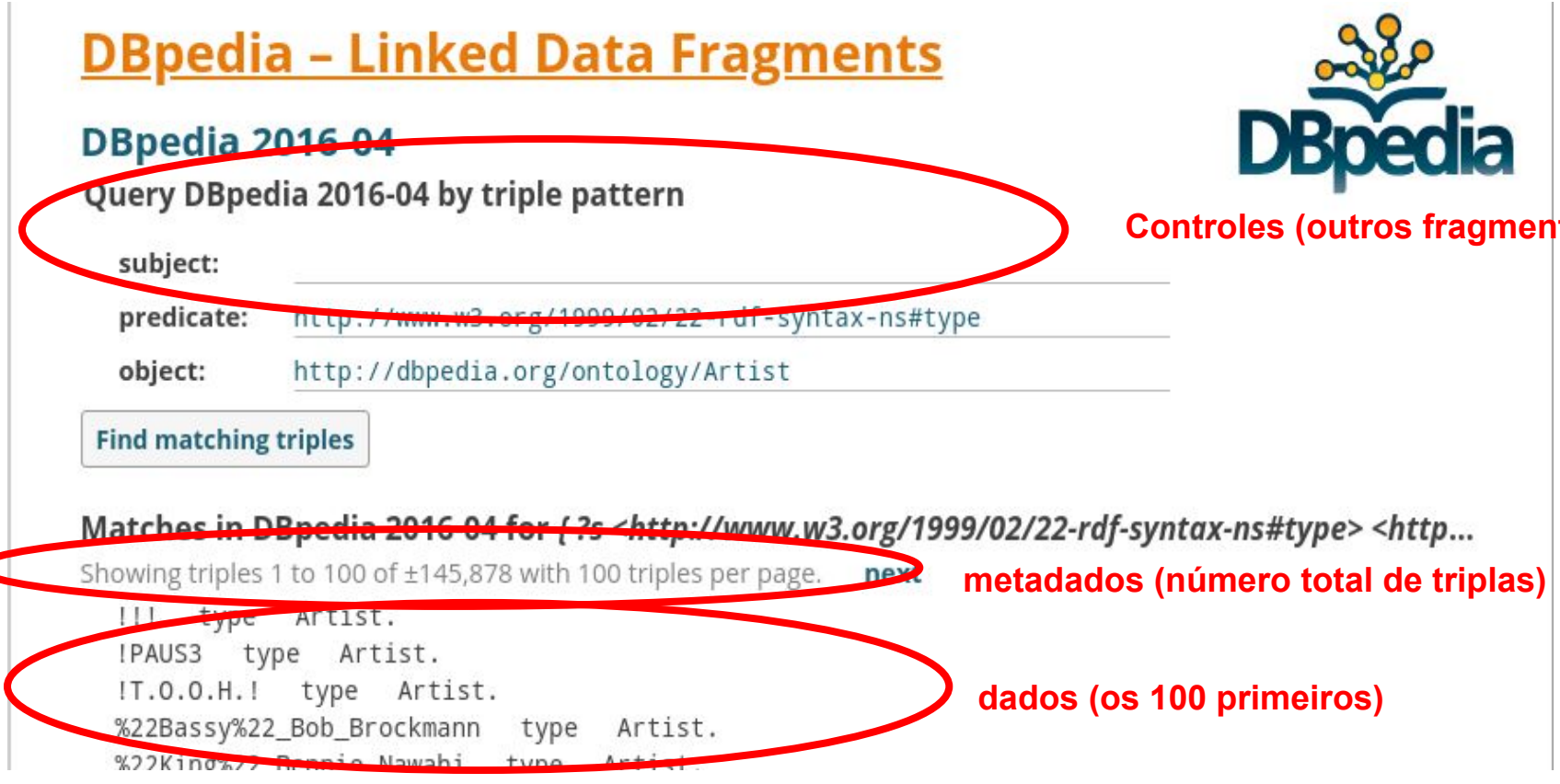
Trade-offs


Alta disponibilidade
Dados atualizados
Baixo custo para o cliente
Data dump
Linked Data Documents
triple pattern fragments
SPARQL

O Triple Pattern Fragments

- **dados**: os resultados da busca
- **metadados**: número de dados da resposta
- **controles**: acesso a outros fragmentos de dados

O Triple Pattern Fragments

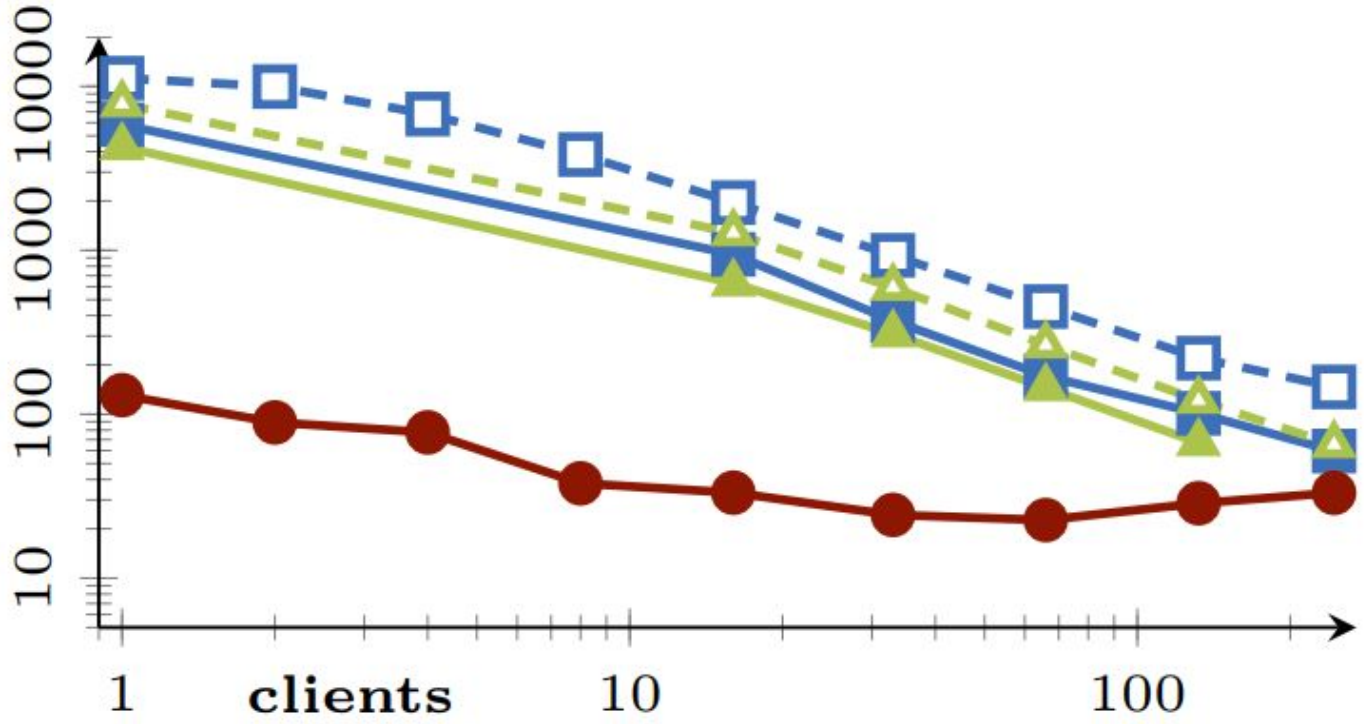


O Triple Pattern Fragments

```
<http://fragments.dbpedia.org/2014/en#dataset> hydra:search [
  hydra:template "http://fragments.dbpedia.org/2014/en
                  {?subject,predicate,object}";
  hydra:mapping
    [ hydra:variable "subject";   hydra:property rdf:subject ],
    [ hydra:variable "predicate"; hydra:property rdf:predicate ],
    [ hydra:variable "object";    hydra:property rdf:object ]
].
```

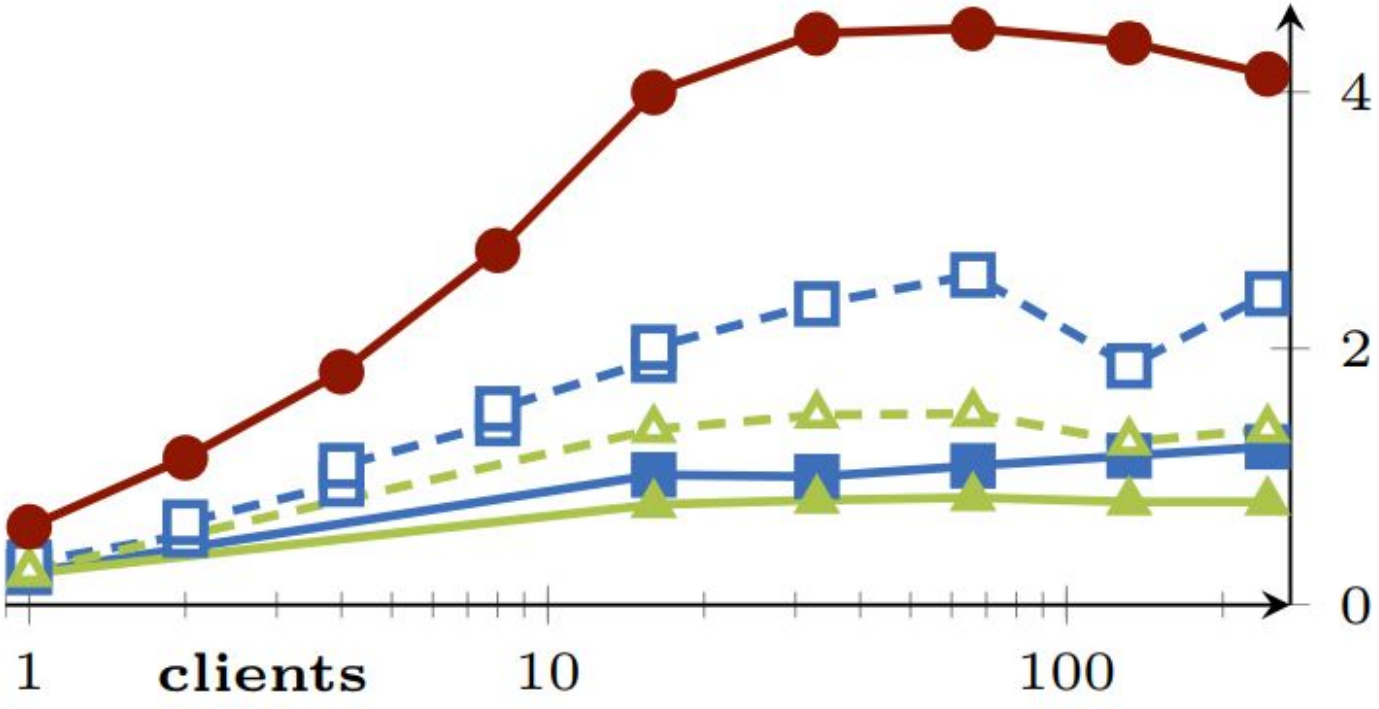
Performance

Resiliente ao aumento de acesso



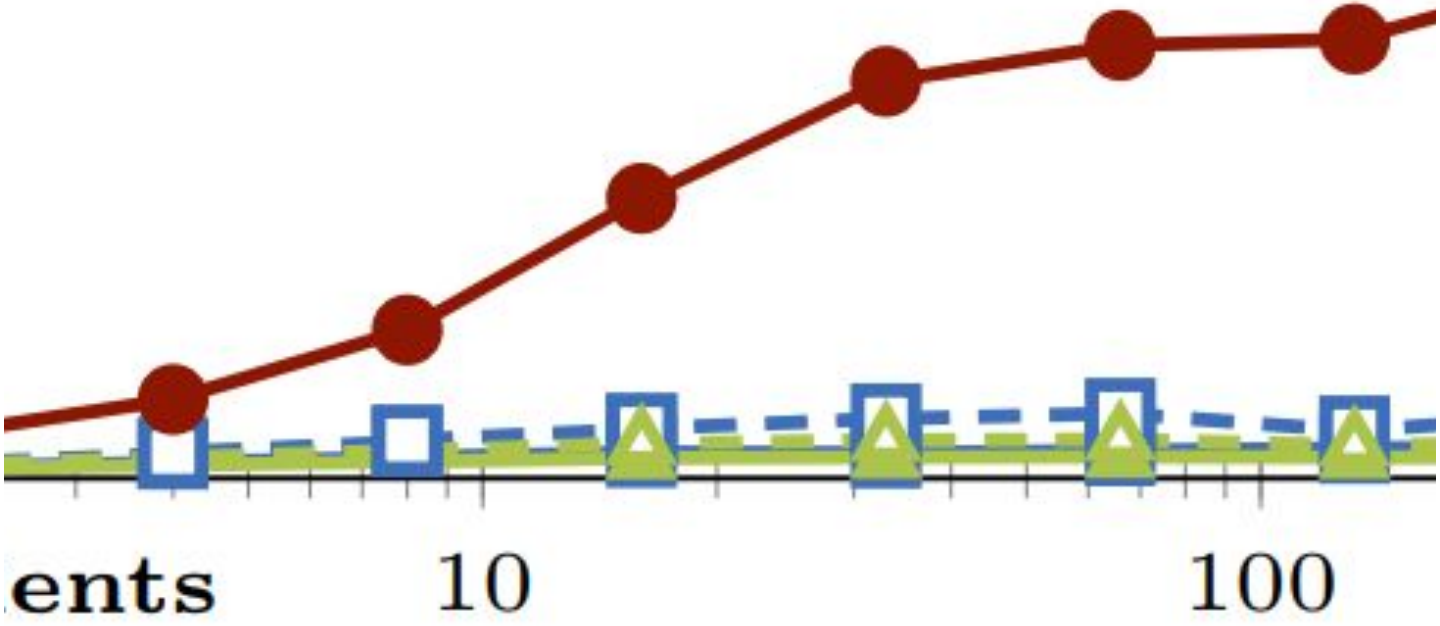
Queries SPARQL executadas por hora [3]

Tráfico maior, requisições mais leves



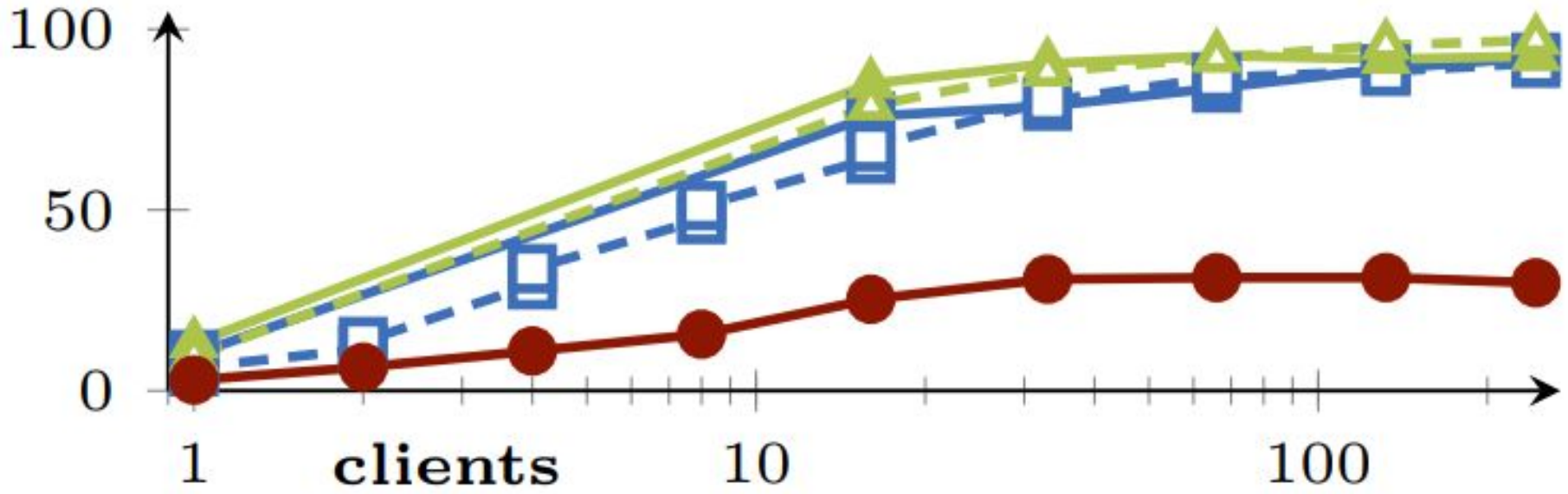
Dados enviados pelo servidor em MB [3]

Cacheamento mais efetivo



Dados enviados pelo servidor em MB [3]

Servidores mais leves



Uso de CPU do servidor por núcleo [3]

Considerações

Bibliografia

1. LODStats. http://stats.lod2.eu/
2. VERBORGH, Ruben et al. **Triple Pattern Fragments: a low-cost knowledge graph interface for the Web. Web Semantics: Science, Services and Agents on the World Wide Web**, v. 37, p. 184-206, 2016.
3. VERBORGH, Ruben et al. **Querying datasets on the Web with high availability**
4. VERBORGH, Ruben et al. **Web-Scale Querying through Linked Data Fragments**. In: LDOW. 2014.