A FINE-GRAINED EVALUATION OF SPARQL ENDPOINT FEDERATION SYSTEMS

Muhammad Saleem, Yasar Khan, Ali Hasnain, Ivan Ermilov, Axel-Cyrille Ngonga Ngomo

ISWC 2016, Kobe, Japan, 20/10/2016

Agile Knowledge Engineering and Semantic Web (AKSW), University of Leipzig, Germany
AGENDA

- SPARQL query federation
- Survey of federated SPARQL query processing systems
- Performance variables
- Evaluation of SPARQL endpoint federation systems
SPARQL QUERY FEDERATION APPROACHES

- SPARQL Endpoint Federation (SEF)
- Linked Data Federation (LDF)
- Hybrid of SEF + LDF
SPARQL Endpoint Federation
SPARQL Endpoint Federation

Rewrite query and get Individual Triple Patterns
SPARQL Endpoint Federation

- Parsing/Rewriting
- Source Selection
- Federator
- Optimizer
- Integrator

Rewrite query and get Individual Triple Patterns
Identify capable source against Individual Triple Patterns
SPARQL Endpoint Federation

1. Parsing/Rewriting
2. Source Selection
   - Federator
   - Optimizer
3. Integrator

Rewrite query and get Individual Triple Patterns
Identify capable source against Individual Triple Patterns
Generate optimized sub-query Exe. Plan

S1 RDF
S2 RDF
S3 RDF
S4 RDF
SPARQL Endpoint Federation

1. Parsing/Rewriting
2. Source Selection
3. Federator
4. Optimzer

Rewrite query and get Individual Triple Patterns
Identify capable source against Individual Triple Patterns
Generate optimized sub-query Exe. Plan
Execute sub-queries

S1 RDF
S2 RDF
S3 RDF
S4 RDF
SPARQL Endpoint Federation

- Parsing/Rewriting
- Source Selection
- Federator
- Optimizer
- Integrator

**Rewrite query and get Individual Triple Patterns**

**Identify capable source against Individual Triple Patterns**

**Generate optimized sub-query Exec. Plan**

**Execute sub-queries**
SPARQL Endpoint Federation

1. Parsing/Rewriting
2. Source Selection
3. Federator
4. Optimizer
5. Integrator

- Rewrite query and get Individual Triple Patterns
- Identify capable source against Individual Triple Patterns
- Generate optimized sub-query Exe. Plan
- Execute sub-queries
- Integrate sub-queries results
SPARQL Endpoint Federation

- Parsing/Rewriting
- Source Selection
- Federator
- Optimizer
- Integrator

Rewrite query and get Individual Triple Patterns
Identify capable source against Individual Triple Patterns
Generate optimized sub-query Exe. Plan
Execute sub-queries
Integrate sub-queries results
TYPES OF SOURCE SELECTION

- Index-free
  - Using SPARQL ASK queries
  - Potentially ensures result set completeness
  - SPARQL ASK queries can be expensive
  - SPARQL ASK cache is beneficial
  - E.g., FedX

- Index-only
  - Only make use of Index/data summaries
  - Less efficient but fast source selection
  - Result set completeness is not ensured
  - E.g., DARQ, LHD
TYPES OF SOURCE SELECTION

- Hybrid
  - Make use of index + SPARQL ASK
  - Most efficient
  - Result set completeness is not ensured
  - SPARQL ASK cache is beneficial
  - E.g., HiBISCuS, ANAPSID, SPLENDID
SURVEY: FEDERATED SPARQL ENGINES

- System Information
  - Title and/or URL of federation engine
  - Code available?
  - Implementation and license
  - Type of source selection
  - Type of join(s) used for data integration
  - Use of cache?
  - Support for catalog/index update
**SURVEY: FEDERATED SPARQL ENGINES**

**C.A.** = Code Availability, **S.S.T.** = Source Selection Type, **I.U.** = Index/catalog Update,
**AGJ** = Adaptive Group Join, **ADJ** = Adaptive Dependent Join

<table>
<thead>
<tr>
<th>Systems</th>
<th>Category</th>
<th>C.A</th>
<th>Implementation</th>
<th>Licencing</th>
<th>S.S.T</th>
<th>Join Type</th>
<th>Cache</th>
<th>I.U</th>
</tr>
</thead>
<tbody>
<tr>
<td>FedX</td>
<td>SEF</td>
<td>✓</td>
<td>Java</td>
<td>GNU A.G.P.L</td>
<td>index-free</td>
<td>bind (VENL)</td>
<td>✓</td>
<td>NA</td>
</tr>
<tr>
<td>LHD</td>
<td>SEF</td>
<td>✓</td>
<td>Java</td>
<td>MIT</td>
<td>hybrid (A+I)</td>
<td>hash/bind</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>SPLENDID</td>
<td>SEF</td>
<td>✓</td>
<td>Java</td>
<td>L.G.P.L</td>
<td>hybrid (A+I)</td>
<td>hash/bind</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>FedSearch</td>
<td>SEF</td>
<td>×</td>
<td>Java</td>
<td>GNU A.G.P.L</td>
<td>hybrid (A+I)</td>
<td>bind, pull based rank join (RMJH)</td>
<td>✓</td>
<td>NA</td>
</tr>
<tr>
<td>GRANATUM</td>
<td>SEF</td>
<td>×</td>
<td>Java</td>
<td>yet to decide</td>
<td>index only</td>
<td>nested loop</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Avalanche</td>
<td>SEF</td>
<td>×</td>
<td>Python, C, C++</td>
<td>yet to decide</td>
<td>index only</td>
<td>distributed, merge</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>ANAPSID</td>
<td>SEF</td>
<td>✓</td>
<td>Python</td>
<td>GNU G.P.L</td>
<td>hybrid (A+I)</td>
<td>AGJ, ADI</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>ADERIS</td>
<td>SEF</td>
<td>✓</td>
<td>Java</td>
<td>Apache</td>
<td>Index only</td>
<td>index-based nested loop</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>DARQ</td>
<td>SEF</td>
<td>✓</td>
<td>Java</td>
<td>GPL</td>
<td>Index only</td>
<td>nested loop, bound</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>LDQPS</td>
<td>LDF</td>
<td>×</td>
<td>Java</td>
<td>Scala</td>
<td>hybrid (C+L)</td>
<td>symmetric hash</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>SIHJoin</td>
<td>LDF</td>
<td>×</td>
<td>Java</td>
<td>Scala</td>
<td>hybrid (C+L)</td>
<td>symmetric hash</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>WoDQA</td>
<td>LDF</td>
<td>✓</td>
<td>Java</td>
<td>GPL</td>
<td>hybrid (A+I)</td>
<td>nested loop, bound</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Atlas</td>
<td>DHTF</td>
<td>✓</td>
<td>Java</td>
<td>GNU L.G.P.L</td>
<td>Index only</td>
<td>SQLite</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>DAW</td>
<td>SEF</td>
<td>×</td>
<td>Java</td>
<td>GNU G.P.L</td>
<td>hybrid (A+I)</td>
<td>based on underlying system</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>SAFE</td>
<td>SEF</td>
<td>✓</td>
<td>Java</td>
<td>GNU G.P.L</td>
<td>hybrid (A+I)</td>
<td>bind (VENL)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>QUETSAL</td>
<td>SEF</td>
<td>×</td>
<td>Java</td>
<td>GNU G.P.L</td>
<td>hybrid (A+I)</td>
<td>based on Sesame</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>HiBISCuS</td>
<td>SEF</td>
<td>✓</td>
<td>Java</td>
<td>GNU G.P.L</td>
<td>hybrid (A+I)</td>
<td>based on underlying system</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
SURVEY: FEDERATED SPARQL ENGINES
SURVEY: FEDERATED SPARQL ENGINES

Code available?

- Yes: 59%
- No: 12%
- Not yet: 29%
SURVEY: FEDERATED SPARQL ENGINES

Code available?
- Not yet: 29%
- Yes: 59%
- No: 12%

Type of source selection
- Hybrid: 65%
- Index-only: 29%
- Index-free: 6%
SURVEY: FEDERATED SPARQL ENGINES

Code available?
- Yes: 59%
- No: 12%
- Not yet: 29%

Type of source selection
- Index-free: 6%
- Index-only: 29%
- Hybrid: 65%

Use of cache?
- Yes: 47%
- No: 53%
SURVEY: FEDERATED SPARQL ENGINES

Code available?
- Not yet: 29%
- Yes: 59%
- No: 12%

Type of source selection
- Index-free: 6%
- Index-only: 29%
- Hybrid: 65%

Use of cache?
- Yes: 47%
- No: 53%

Index update support?
- NA: 12%
- Yes: 29%
- No: 59%
SURVEY: FEDERATED SPARQL ENGINES

- Requirements
  - Result completeness
  - Policy-based query planning
  - Support for partial results retrieval
  - Support for no-blocking operator/ adaptive query processing
  - Support for provenance information
  - Query runtime estimation
  - Duplicate Detection
  - Top-K query processing
  - Supported SPARQL types/ clauses

11/3/2016
SURVEY: FEDERATED SPARQL ENGINES


<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>FedEx</td>
<td>✓</td>
<td>x</td>
<td></td>
<td>✓</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LHD</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPLENDID</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FedSearch</td>
<td>✓</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRANATUM</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avalanche</td>
<td>x</td>
<td>✓</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANAPSID</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADERIS</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DARQ</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDQPS</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIHJoin</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WoDQA</td>
<td>✓</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atlas</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAW</td>
<td>x</td>
<td>✓</td>
<td>based on underlying system</td>
<td>✓</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAFE</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QUETSAL</td>
<td>x</td>
<td>✓</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HiBISCuS</td>
<td>x</td>
<td>✓</td>
<td>based on underlying system</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
SURVEY: FEDERATED SPARQL ENGINES
SURVEY: FEDERATED SPARQL ENGINES

Result completeness?

Yes 18%
No 82%
SURVEY: FEDERATED SPARQL ENGINES

Result completeness?
- Yes: 18%
- No: 82%

Policy-based planning?
- Partial: 6%
- Yes: 6%
- No: 88%
SURVEY: FEDERATED SPARQL ENGINES

Result completeness?

- Yes 18%
- No 82%

Policy-based planning?

- Yes 6%
- Partial 6%
- No 88%

Partial results retrieval?

- Yes 24%
- No 76%
SURVEY: FEDERATED SPARQL ENGINES

Result completeness?
- Yes: 18%
- No: 82%

Policy-based planning?
- Yes: 6%
- No: 88%

Partial results retrieval?
- Yes: 24%
- No: 76%

Adaptive processing?
- Yes: 59%
- No: 41%
SURVEY: FEDERATED SPARQL ENGINES
SURVEY: FEDERATED SPARQL ENGINES

Provenance info?
- Yes: 0%
- No: 94%
- Partial: 6%
SURVEY: FEDERATED SPARQL ENGINES

Provenance info?
- Partial: 6%
- Yes: 0%
- No: 94%

Runtime estimation?
- Yes: 0%
- No: 100%
SURVEY: FEDERATED SPARQL ENGINES

Provenance info?
- Partial: 6%
- Yes: 0%
- No: 94%

Runtime estimation?
- Yes: 0%
- No: 100%

Duplicate detection?
- Partial: 12%
- Yes: 12%
- No: 76%
SURVEY: FEDERATED SPARQL ENGINES

**Provenance info?**
- Partial: 6%
- Yes: 0%
- No: 94%

**Runtime estimation?**
- Yes: 0%
- No: 100%

**Duplicate detection?**
- Partial: 12%
- Yes: 12%
- No: 76%

**Top-k querying?**
- Yes: 100%
- No: 100%
SURVEY: FEDERATED SPARQL ENGINES

<table>
<thead>
<tr>
<th>SPARQL Clause</th>
<th>FedX</th>
<th>Atlas</th>
<th>LHD</th>
<th>SPL</th>
<th>FedS</th>
<th>GRA</th>
<th>Ava</th>
<th>DAW</th>
<th>LDQPS</th>
<th>SIHJoin</th>
<th>ANA</th>
<th>ADE</th>
<th>QWIDVD</th>
<th>DARQ</th>
<th>SAFE</th>
<th>QUETSAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>SERVICE</td>
<td>✓</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FILTER</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unbound QP</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unbound QS</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OPTIONAL</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DISTINCT</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORDER BY</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UNION</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEGATION</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>REGEX</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIMIT</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CONSTRUCT</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DESCRIBE</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASK</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independent Variables</td>
<td>Query</td>
<td>Data</td>
<td>Platform</td>
<td>Endpoints</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>-------</td>
<td>------</td>
<td>----------</td>
<td>-----------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>query plan shape</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>basic triple patterns</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>instantiations and their position</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>join selectivity</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>intermediate results</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>answer size</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>usage of query language expressivity</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>general predicates</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dataset size</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>data frequency distribution</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>type of partitioning</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>data endpoint distribution</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cache on/off</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAM available</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#processors</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#endpoints</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>endpoint type</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>relation graph/endpoint/instance</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>network latency</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>initial delay</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>message size</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>transfer distribution</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>answer size limit</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>timeout</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
EVALUATION

- Benchmarks
  - FedBench
  - Sliced FedBench
  - SP2Bench
- SPARQL endpoint federation engines
  - FedEx
  - SPLENDID
  - LHD
  - DAW
  - ADERIS
  - ANAPSID
## NUMBER OF ASK REQUESTS

<table>
<thead>
<tr>
<th>FedBench</th>
<th>FedX</th>
<th>SPLENDID</th>
<th>LHD</th>
<th>DARQ</th>
<th>ANAPSID</th>
<th>ADERIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD</td>
<td>252</td>
<td>44</td>
<td>0</td>
<td>0</td>
<td>43</td>
<td>0</td>
</tr>
<tr>
<td>LS</td>
<td>297</td>
<td>33</td>
<td>0</td>
<td>0</td>
<td>63</td>
<td>0</td>
</tr>
<tr>
<td>LD</td>
<td>369</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>37</td>
<td>0</td>
</tr>
<tr>
<td>Net</td>
<td>918</td>
<td>96</td>
<td>0</td>
<td>0</td>
<td>143</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sliced FedBench</th>
<th>FedX</th>
<th>SPLENDID</th>
<th>LHD</th>
<th>DARQ</th>
<th>ANAPSID</th>
<th>ADERIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD</td>
<td>280</td>
<td>110</td>
<td>0</td>
<td>0</td>
<td>228</td>
<td>0</td>
</tr>
<tr>
<td>LS</td>
<td>330</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>143</td>
<td>0</td>
</tr>
<tr>
<td>LD</td>
<td>410</td>
<td>140</td>
<td>0</td>
<td>0</td>
<td>270</td>
<td>0</td>
</tr>
<tr>
<td>SP2Bench</td>
<td>660</td>
<td>180</td>
<td>0</td>
<td>0</td>
<td>265</td>
<td>0</td>
</tr>
<tr>
<td>Net</td>
<td>1680</td>
<td>530</td>
<td>0</td>
<td>0</td>
<td>906</td>
<td>0</td>
</tr>
</tbody>
</table>
# Triple Pattern-Wise Sources Selected

<table>
<thead>
<tr>
<th>Query Category</th>
<th>FedBench</th>
<th>Sliced FedBench</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FedX</td>
<td>SPLENDID</td>
</tr>
<tr>
<td>CD</td>
<td>78</td>
<td>78</td>
</tr>
<tr>
<td>LS</td>
<td>56</td>
<td>56</td>
</tr>
<tr>
<td>LD</td>
<td>108</td>
<td>108</td>
</tr>
<tr>
<td>Net</td>
<td><strong>242</strong></td>
<td><strong>242</strong></td>
</tr>
<tr>
<td>SP2Bench</td>
<td>521</td>
<td>521</td>
</tr>
<tr>
<td>Net</td>
<td><strong>1071</strong></td>
<td><strong>1071</strong></td>
</tr>
<tr>
<td>Query Category FedX(cold)</td>
<td>FedX(warm)</td>
<td>SPLENDID</td>
</tr>
<tr>
<td>--------------------------</td>
<td>------------</td>
<td>----------</td>
</tr>
<tr>
<td>CD</td>
<td>151</td>
<td>7</td>
</tr>
<tr>
<td>LS</td>
<td>147</td>
<td>8</td>
</tr>
<tr>
<td>LD</td>
<td>139</td>
<td>8</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>146</strong></td>
<td><strong>8</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query Category P2Bench</th>
<th>Sliced FedBench</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD</td>
<td>284</td>
</tr>
<tr>
<td>LS</td>
<td>207</td>
</tr>
<tr>
<td>LD</td>
<td>323</td>
</tr>
<tr>
<td>SP2Bench</td>
<td>212</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>256</strong></td>
</tr>
</tbody>
</table>
# Result Completeness

<table>
<thead>
<tr>
<th></th>
<th>CD1(90)</th>
<th>CD7(1)</th>
<th>LS1(1159)</th>
<th>LS2(333)</th>
<th>LS3(9054)</th>
<th>LS5(393)</th>
<th>LD1(309)</th>
<th>LD3(162)</th>
<th>LD9(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPLENDID</td>
<td>90</td>
<td>1</td>
<td>1159</td>
<td>333</td>
<td>9054</td>
<td>393</td>
<td>308</td>
<td>159</td>
<td>1</td>
</tr>
<tr>
<td>LHD</td>
<td>77</td>
<td>1</td>
<td>0</td>
<td>322</td>
<td>0</td>
<td>0</td>
<td>309</td>
<td>162</td>
<td>1</td>
</tr>
<tr>
<td>ANAPSID</td>
<td>90</td>
<td>0</td>
<td>1159</td>
<td>333</td>
<td>9054</td>
<td>393</td>
<td>309</td>
<td>162</td>
<td>1</td>
</tr>
<tr>
<td>ADERIS</td>
<td>77</td>
<td>1</td>
<td>1159</td>
<td>333</td>
<td>9054</td>
<td>393</td>
<td>309</td>
<td>162</td>
<td>1</td>
</tr>
<tr>
<td>DARQ</td>
<td>90</td>
<td>1</td>
<td>1159</td>
<td>333</td>
<td>9054</td>
<td>393</td>
<td>309</td>
<td>162</td>
<td>0</td>
</tr>
<tr>
<td>FedEx</td>
<td>90</td>
<td>1</td>
<td>1159</td>
<td>333</td>
<td>9054</td>
<td>393</td>
<td>309</td>
<td>162</td>
<td>1</td>
</tr>
</tbody>
</table>
QUERY RUNTIME
FedBench: FedX(warm) $\Rightarrow$ FedX(cold) $\Rightarrow$ LHD $\Rightarrow$ SPLENDID $\Rightarrow$ ANAPSID $\Rightarrow$ DARQ
FedBench: FedX(warm) → FedX(cold) → LHD → SPLENDID → ANAPSID → DARQ
Sliced FedBench: FedX(warm) → FedX(cold) → LHD → SPLENDID → ANAPSID → DARQ
EFFECT OF DATA PARTITIONING

FedX (first run) FedBench
FedX (first run) SlicedBench

FedX (cached) FedBench
FedX (cached) SlicedBench

SPLENDID FedBench
SPLENDID SlicedBench

ANAPSID FedBench
ANAPSID SlicedBench

DARQ FedBench
DARQ SlicedBench

LHD FedBench
LHD SlicedBench
EFFECT OF DATA PARTITIONING

FedX (first run) FedBench
FedX (first run) SlicedBench

Decreased by 214%

FedX (cached) FedBench
FedX (cached) SlicedBench

Decreased by 199%

SPLENDID FedBench
SPLENDID SlicedBench

Decreased by 227%

ANAPSID FedBench
ANAPSID SlicedBench

Decreased by 392%

DARQ FedBench
DARQ SlicedBench

Increased by 36%

LHD FedBench
LHD SlicedBench

Decreased by 293%
THANKS