



SPARQL ENDPOINT STATUS > DATAHUB.IO

“Monitoring the Status of SPARQL Endpoints”

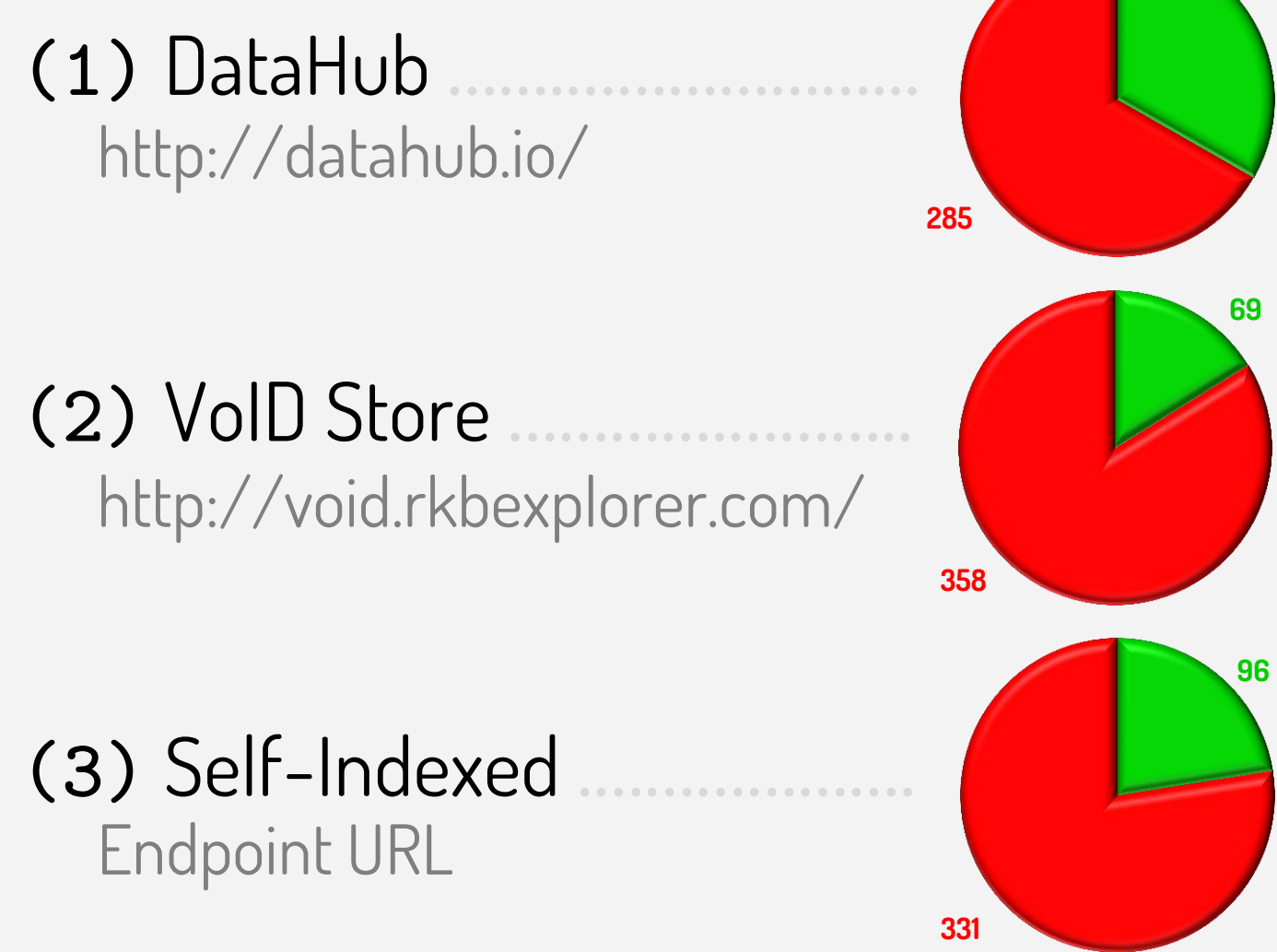
▼ Description:

- ▷ Extract list of 427+ public SPARQL endpoints from Datahub.io
- ▷ Monitor **DISCOVERABILITY, INTEROPERABILITY, PERFORMANCE, AVAILABILITY,**

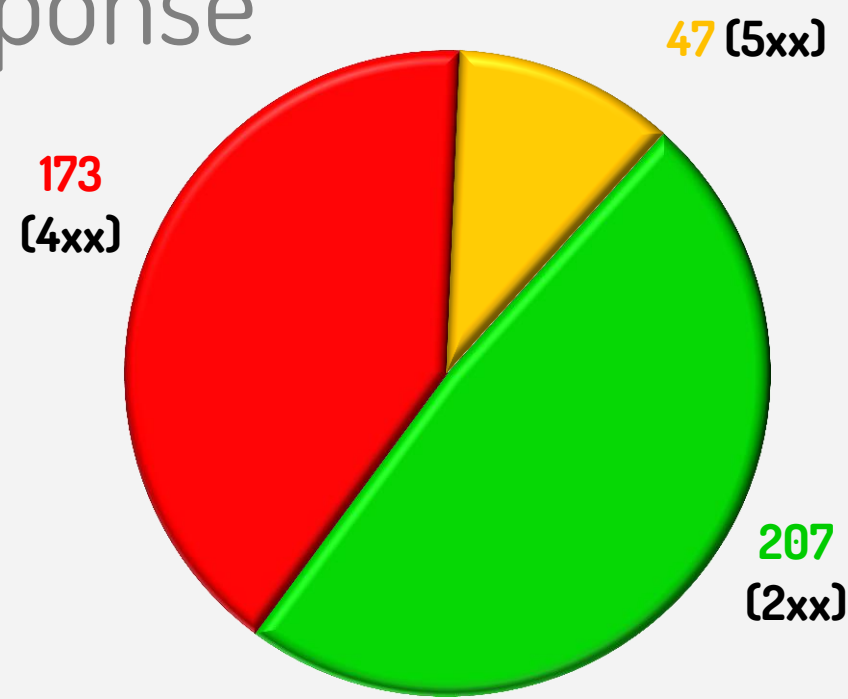
▼ DISCOVERABILITY:

VoID: Discovery

```
PREFIX void: <http://rdfs.org/ns/void#>
SELECT DISTINCT ?ds
WHERE { ?ds a void:Dataset ; void:sparqlEndpoint %%ep . }
```



Dereferencing: HTTP Response Codes



Dereferencing: SPARQL Service Descriptions

Predicate	Nº	Predicate	Nº
sd:resultFormat	38	void:sparqlEndpoint	5
sd:feature	36	void:classes	2
sd:supportedLanguage	36	void:datadump	2
sd:url	34	void:triples	2
sd:endpoint	33	void:vocabulary	2

Dereferencing: HTTP Server Fields

Server-field Prefix	Nº
Apache	157
Virtuoso	71
Jetty	25
nginx	23
Fuseki	6
4s-httpd	3

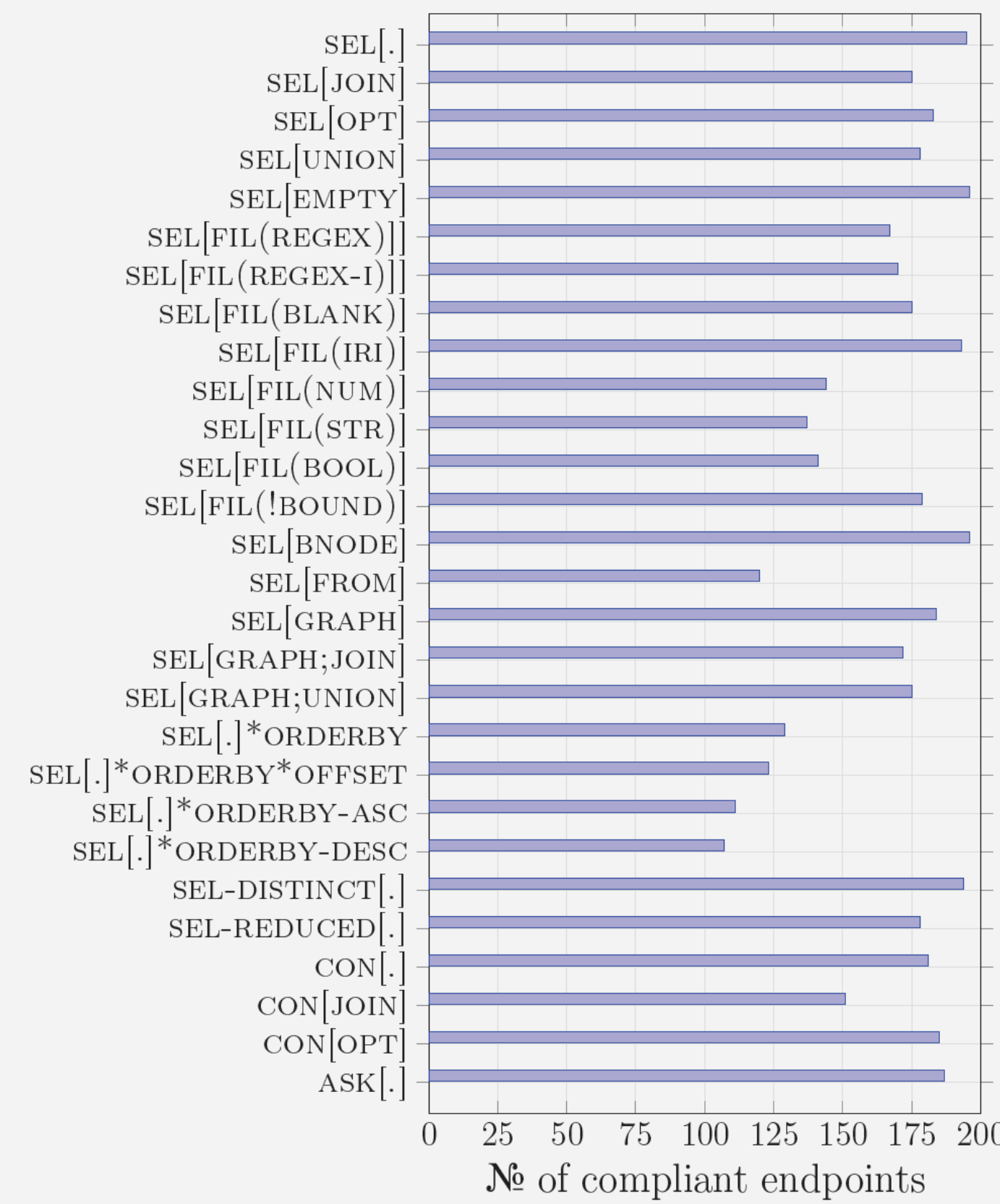
VoID: Filename Conventions

Path	Nº	File-name	Nº
/	52	void.ttl	69
/dataset	20	lds	11
/datasets	6	void.rdf	8
/models	5	GeologicTimeScale.rdf	4
/doc	4	GeologicUnit.rdf	4

▼ INTEROPERABILITY:

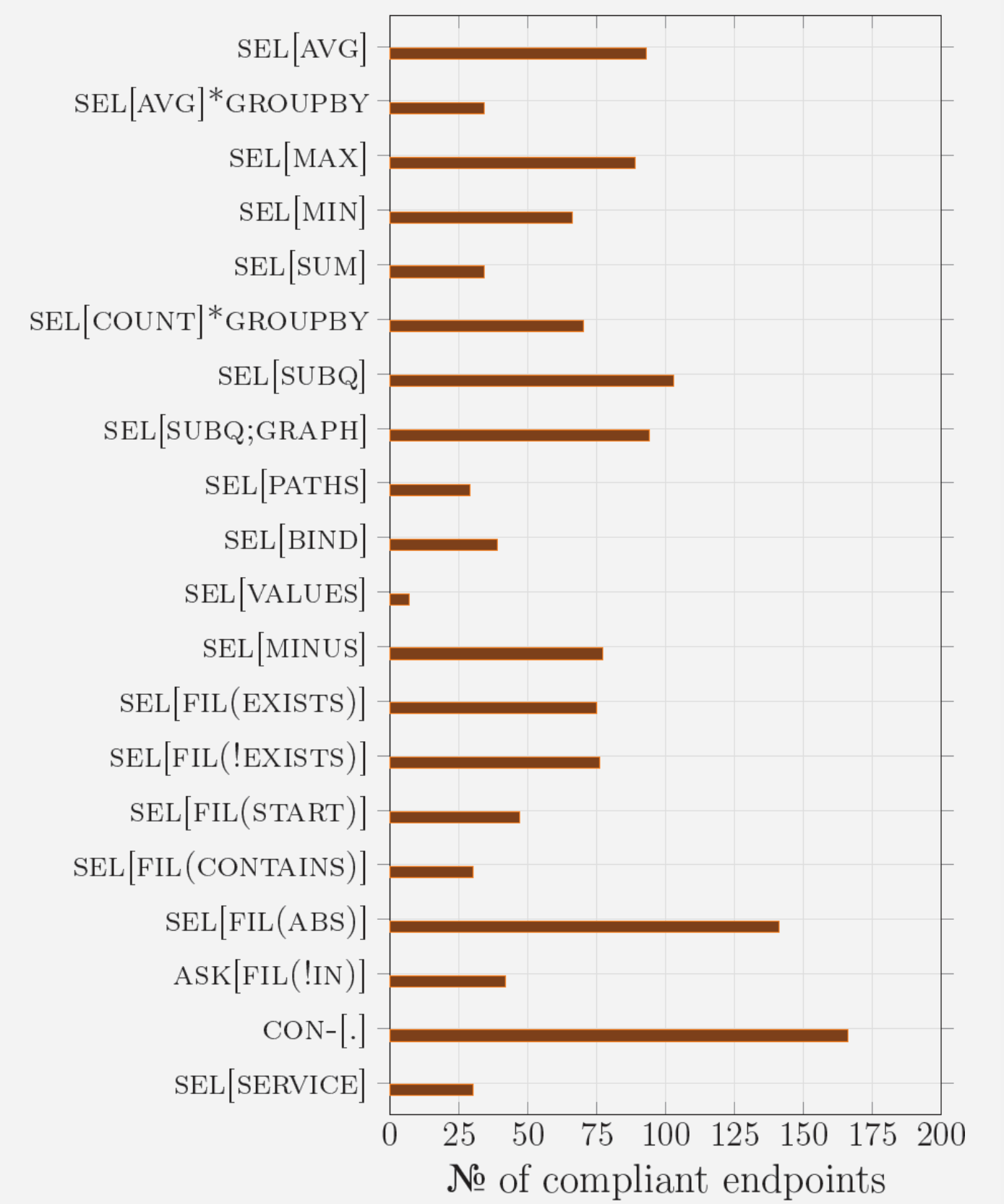
SPARQL 1.0:

Test-case Queries Passed



SPARQL 1.1:

Test-case Queries Passed



Compliance:

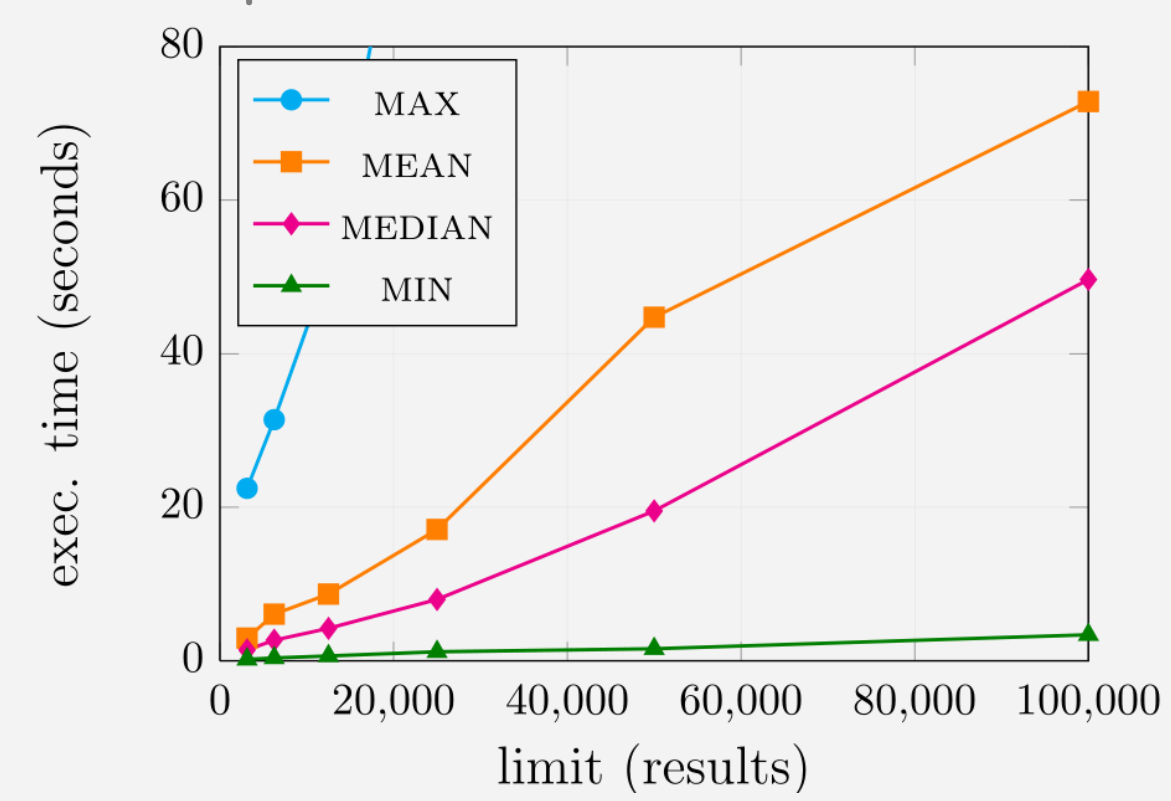
Thresholds on the number of results

```
SELECT * WHERE { ?s ?p ?o } LIMIT 100002
```

Nº of Results	Nº of Endpoints
500	1
1,000	3
1,500	1
5,000	1
10,000	49
20,000	2
40,000	1
50,000	3
100,000	7
TOTAL:	68

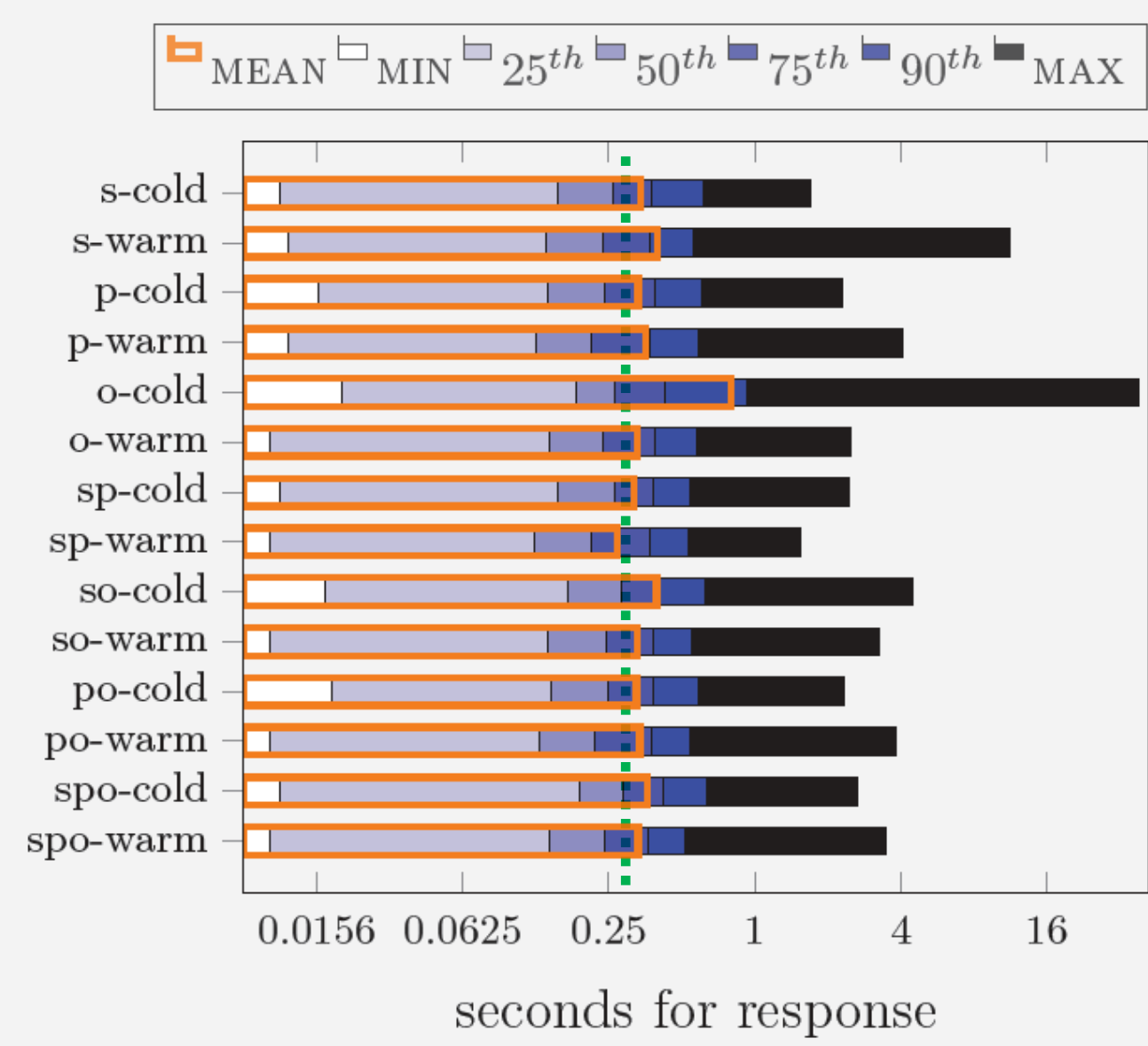
▼ PERFORMANCE:

Streaming Performance: LIMIT Queries



```
SELECT * WHERE { ?s ?p ?o } LIMIT 100002
```

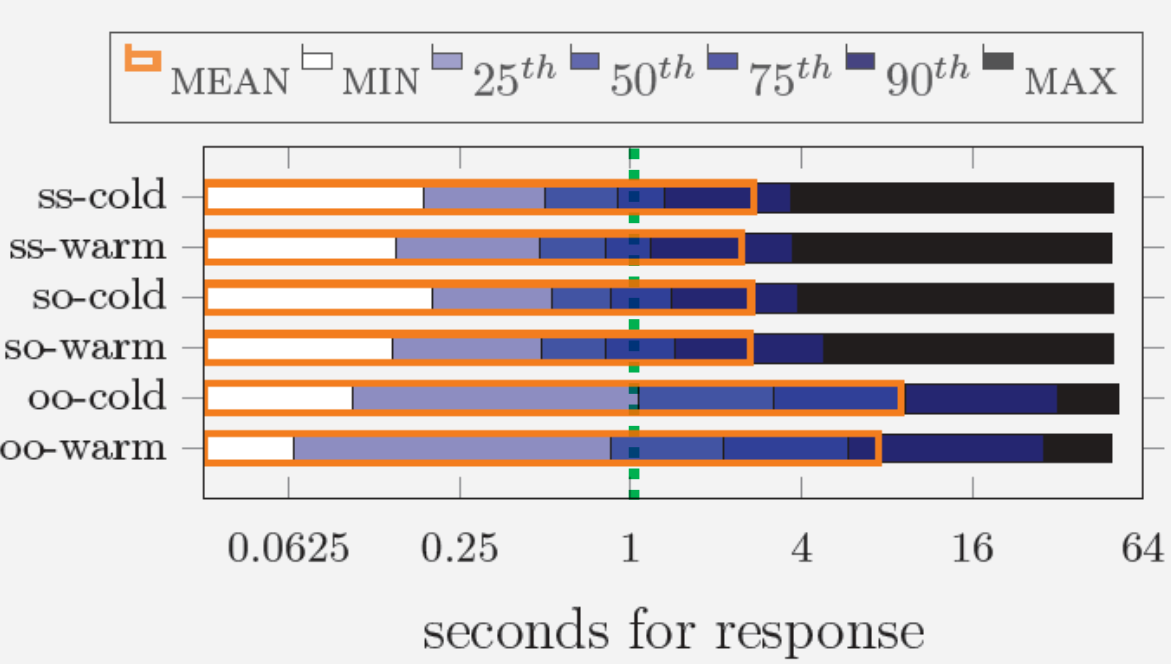
Lookup Performance: Atomic ASK Queries



```
ASK WHERE { ?s <y> <z> }
```

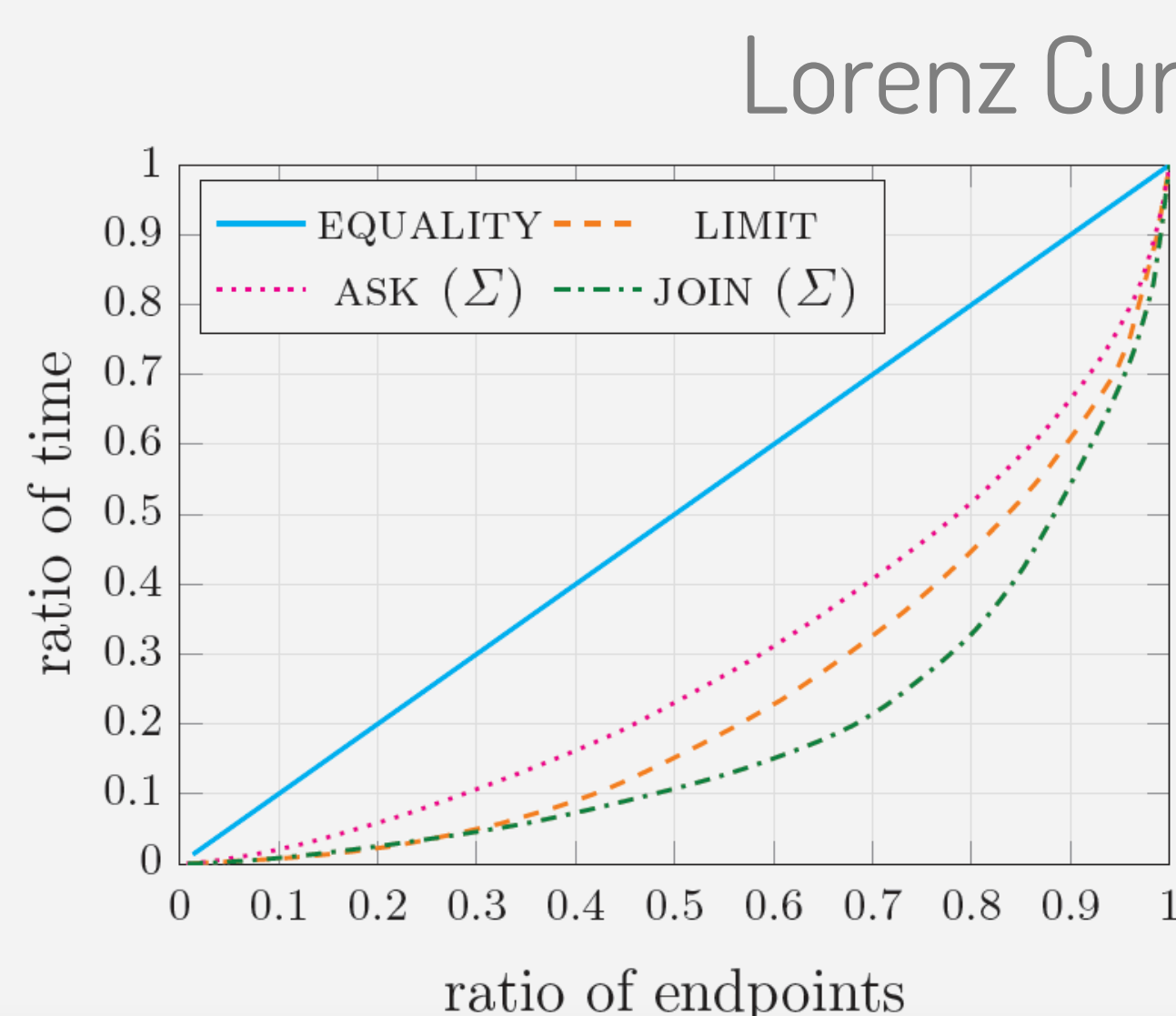
Join Performance:

OPTIONAL with no results



```
SELECT DISTINCT ?s ?p2
WHERE { ?s ?p ?o .
OPTIONAL { ?s ?p2 <x> . }
LIMIT 1000
```

Skewed Performance: Lorenz Curve

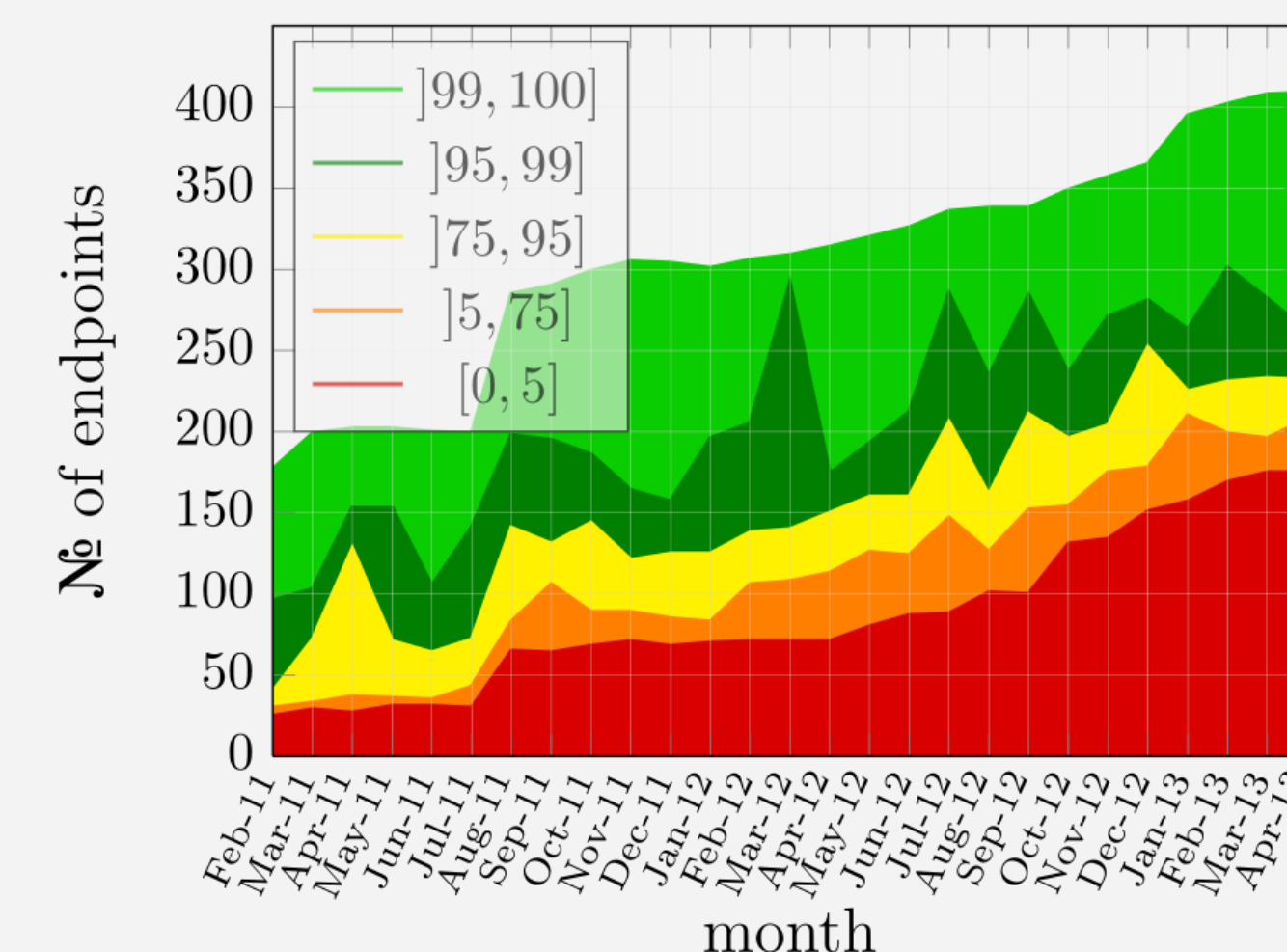


▼ AVAILABILITY:

Uptimes: Hourly Pings over Two Years



Availability: Count



Availability: Ratio

