SOROCA
Highly Scalable Open Source
Earth Observation Catalogue Server

IGUASSU SOFTWARE SYSTEMS

Continuously growing volume of remote sensed images generates growing demand to store these images in a repository, to store EO geospatial metadata in spatial database (catalogue) and to provide standardized catalogue server to access them. This increase of demand for catalogue server and its storage can be effectively satisfied by **scaling.**
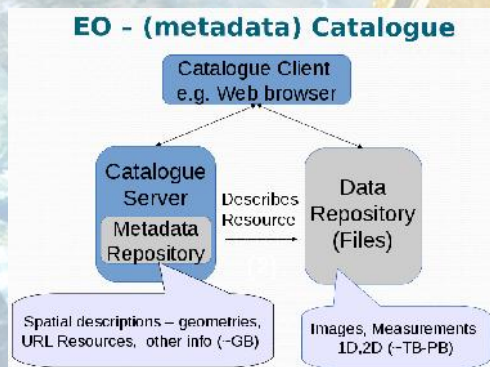

Figure 1.

SOROCA is highly horizontally scalable open source EO Catalogue server developed under European Space Agency (ESA) contract, consisting of:

• **OpenSearch** based server (Conforms to 10-032r3 standard)
• **MongoDB** database (i.e. not relational Database), where the data is distributed to several nodes / shards (Mongod) and, transparently to the user, balanced between them.
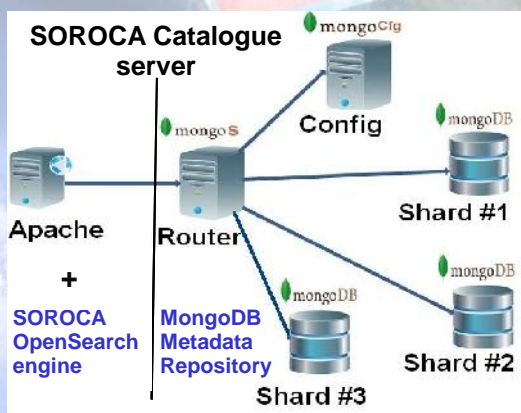

SOROCA Catalogue server

Figure 2.

Commonly used R-tree spatial indexing uses bounding box filtering (Figure 3) to avoid unnecessary CPU utilization. Bounding boxes of some footprints cover the entire earth which decreases performance of bounding box pre - filtering . One option to improve searching is by scaling.

Figure 3.

SOROCA catalogue has the following characteristics:

Basic functions:
• Full text search
• Bounding box search
• Arbitrary geometry search (using WKT)
• Spatial search using a point and a radius
• Get record by id
• Search by name
• Temporal search

Relational Geo function
• Contains, Overlaps, Disjoint, Intersects

Stored Geometries:
• Point, Linestring, Polygon, Multipoint, Multilinestring, Multipolygon,

Output Formats:
• generic XML / RDF / KML / Atom / WKT (well known text)

WKT

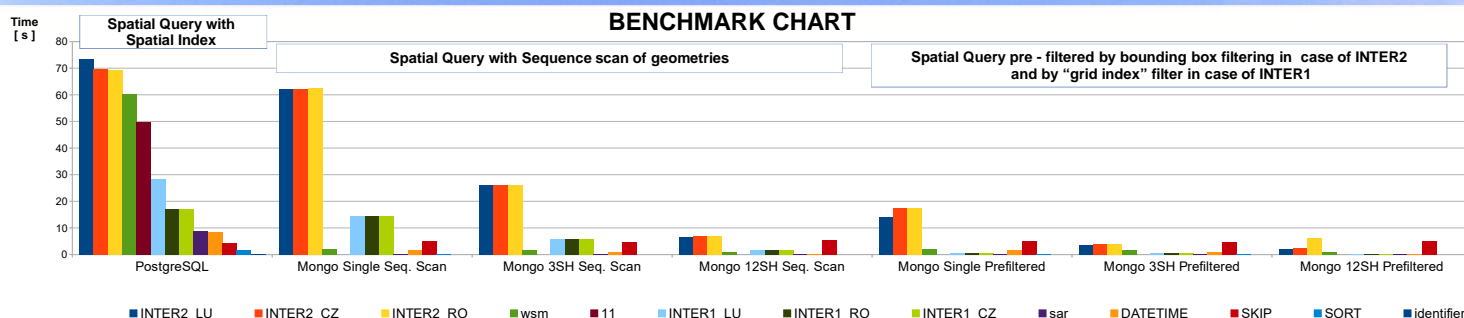**Scalability + failover + loadbalancing**
• provided by MongoDB itself in Metadata repository
• provided by 3th party SW e.g. Keepalived for web server

CONCLUSION

We compared PostgreSQL with Mongo database running on a single machine/node and then with Mongo database running on 3 machines scaled with 3SHards, and 12SHards.
The notable improvement of speed is visible from the Benchmark chart (Figure 4.)
At the same the SOROCA system can easily be incrementally expanded by adding additional machines.

BENCHMARK CHART

Time [s]

Spatial Query with Spatial Index	Spatial Query with Sequence scan of geometries	Spatial Query pre - filtered by bounding box filtering in case of INTER2 and by "grid index" filter in case of INTER1

Legend: ■ INTER2_LU ■ INTER2_CZ ■ INTER2_RO ■ wsm ■ 11 ■ INTER1_LU ■ INTER1_RO ■ INTER1_CZ ■ sar ■ DATETIME ■ SKIP ■ SORT ■ identifier

	INTER2 LU	INTER2 CZ	INTER2 RO	wsm	11	INTER1 LU	INTER1 RO	INTER1 CZ	sar	DATETIME	SKIP	SORT	Identifier
Total record	1422303	1422303	1422303	6650273	6650273	214834	214834	214834	6650273	6650273	6650273	3307	6650273
Results Count	8974	17946	8971	147499	416	153	1380	909	12842	247316	10	3307	1
Description	Spatial search intersects Luxemburg 14 coordinates	Spatial search intersects –Czech Republic 74 coordinates	Spatial search intersects Romania – 102 coordinates	text search "wsm"	Text search "11"	Spatial search intersects Luxemburg 14 coordinates	Spatial search intersects Romania 102 coordinates	Spatial search intersects –Czech Republic –74 coordinates	Text search "sar"	temporal search	Skip (300k) & limit 10	Sort	Search by unique identifier
Area of interest									-	-	-	-	-

Roman Benčík, roman.bencik@iguassu.eu , Iguassu Software Systems a.s., Czech Republic
Tel.: +420 2 3535 1000, http://www.iguassu.eu