

A Novel Approach to GNSS Data Storage Using TileDB in the Context of EPOS

Fernando Geraldes⁽¹⁾

Duarte Arribas⁽¹⁾

Paul Crocker⁽¹⁾

Luís Carvalho⁽¹⁾

Rui Fernandes⁽¹⁾

⁽¹⁾ SEGAL-UBI, Covilhã, Portugal

Summary

- **TileDB**
- **TileDB Implementations**
- **Architecture**
- **Testing - Granularities**
- **Results - Import**
- **Results - Query**
- **Conclusion**

TileDB

main characteristics

- Array database (Array multidimensional)
- Supports large amounts of data
- Supports Dense and Sparse data
- ARCO (Analysis-Ready and Cloud-Optimized)
- Integrates with multiple languages/cloud storage
- Asynchronous fragmented storage
- Applications: Big data processing, real-time monitoring
- Ready for new forms of queries such as machine learning and artificial intelligence

cols

4	2	3	9		
3		9	8	4	1
9	1	2	5	1	2
8	8	9		0	8
	8		2		3
5		1	4	9	

row

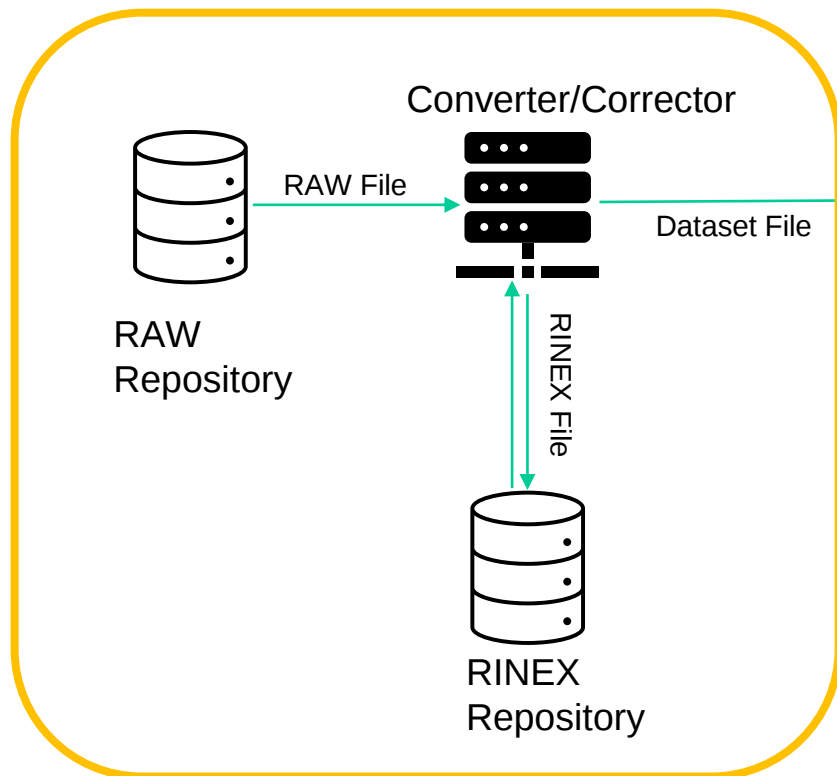
TileDB Implementations

-
Earthscope & GFZ

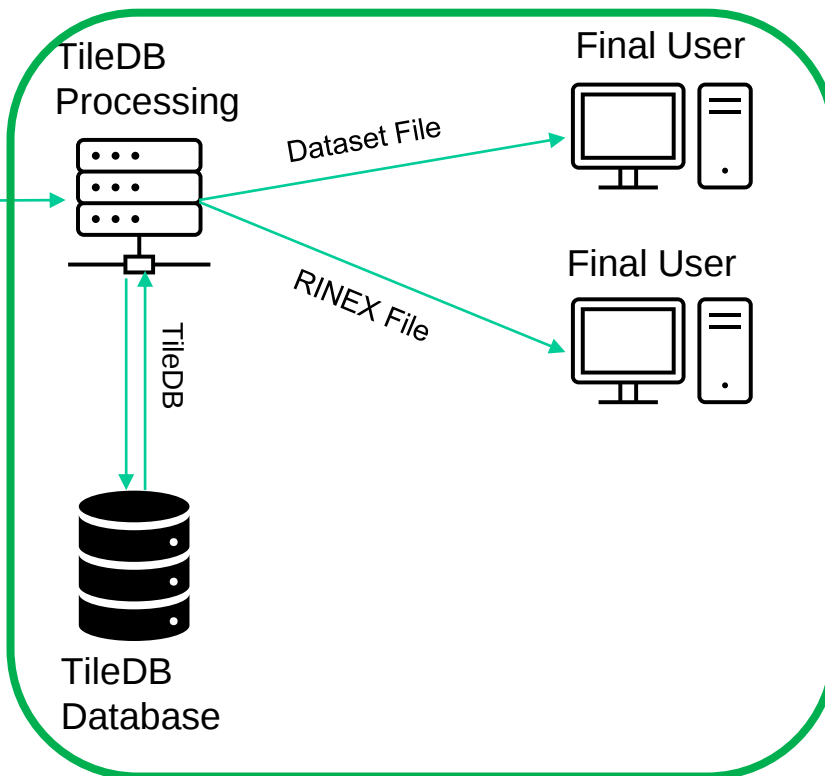
	Storage Granularity	Dimensions	Common Attributes	Organization Specific Attributes
 EarthScope Consortium	One TileDB per Station	1.Time; 2.System; 3.Satellite; 4.Observation codes.	1.Range; 2.Phase; 3.Doppler; 4.Signal-to-noise; 5.Flags; 6.Frequency channel number (FCN).	1.Slip.
 GFZ Helmholtz Centre for Geosciences				1.LLI; 2.SSI.

Architecture

Previous Presentation



Focus of this presentation



Testing - – Granularities

Station-Only

Station-Year

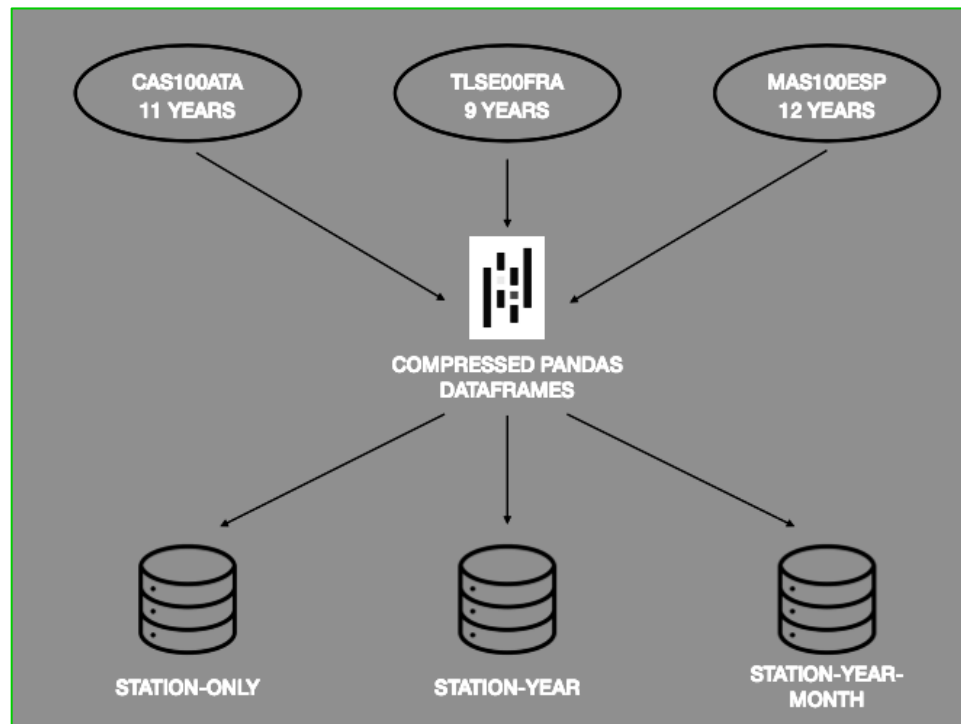
Station-Year-Month

➤ Three random stations:

CAS100ATA	- 11 Years
TLSE00FRA	- 9 Years
MAS100ESP	- 12 Years

➤ Three different granularities:

Station-only
Station-year
Station-year-month



Testing

Queries – Import Metrics – Query Metrics

➤ Executed queries:

24h queries data:

- daily solutions
- special case: crossing December 31st – January 1st

➤ Import Metrics:

Number of Tiles

Time

Size

➤ Query Metrics:

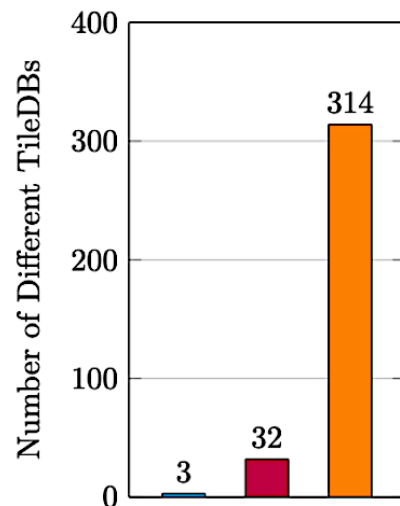
Time

Memory

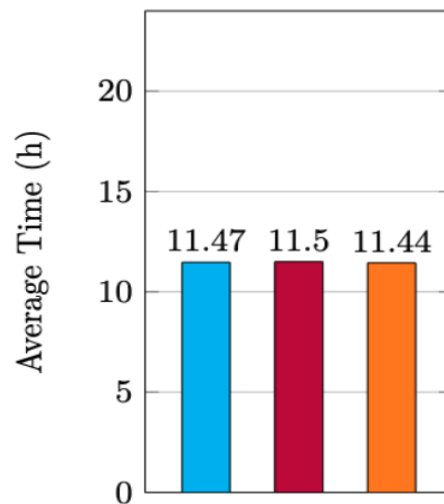
CPU %.

Results - Import

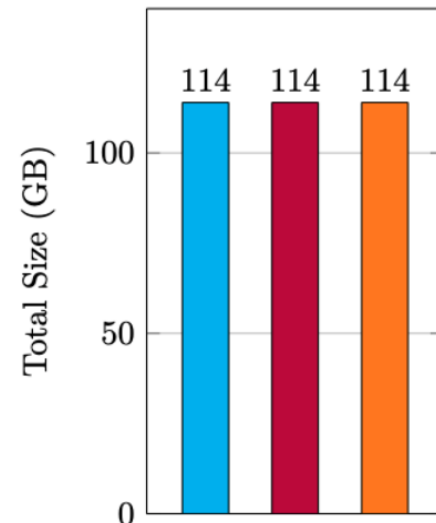
Number of Different TileDBs



Average Time (h)



Total Size (GB)



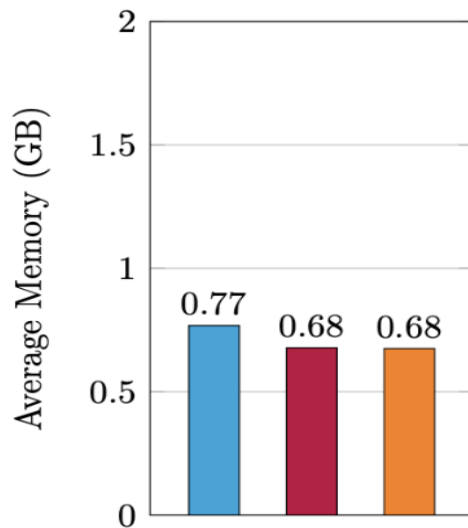
■ TileDB Station Only

■ TileDB Station Year

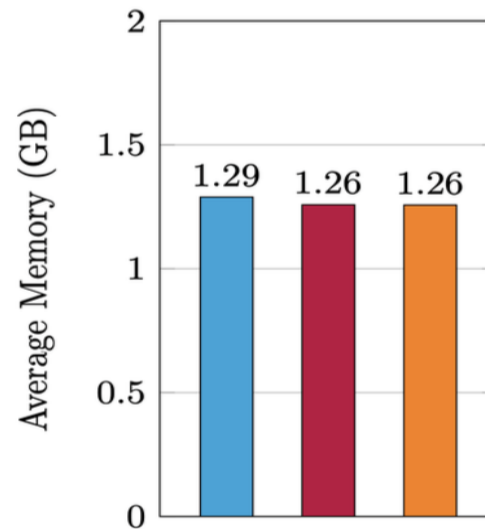
■ TileDB Station Year Month

Results - Query

Average Memory (GB)



Queries for 1 day



Queries for 2 day

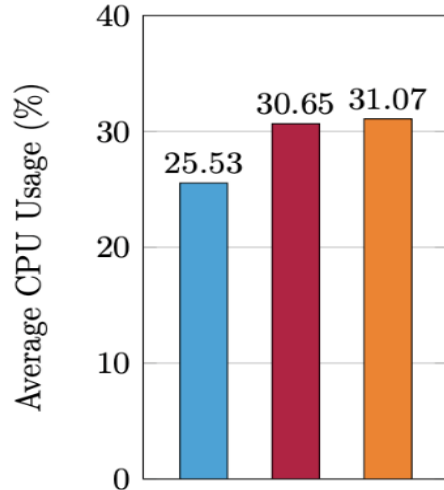
■ TileDB Station Only

■ TileDB Station Year

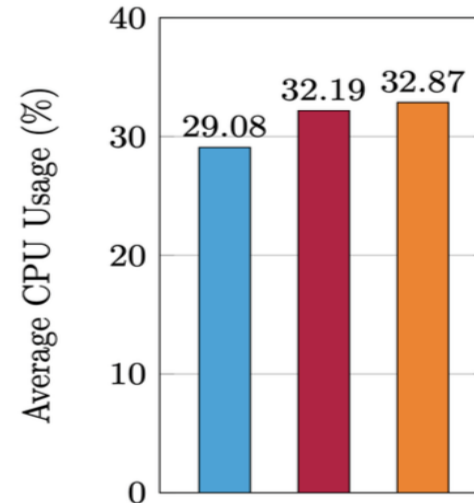
■ TileDB Station Year Month

Results - Query

Average CPU (%)



Queries for 1 day



Queries for 2 day

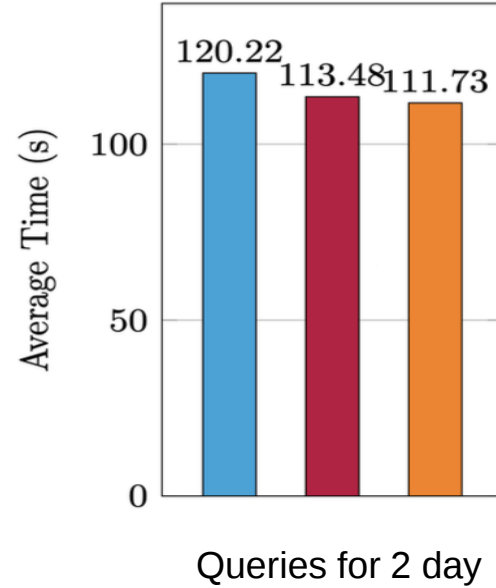
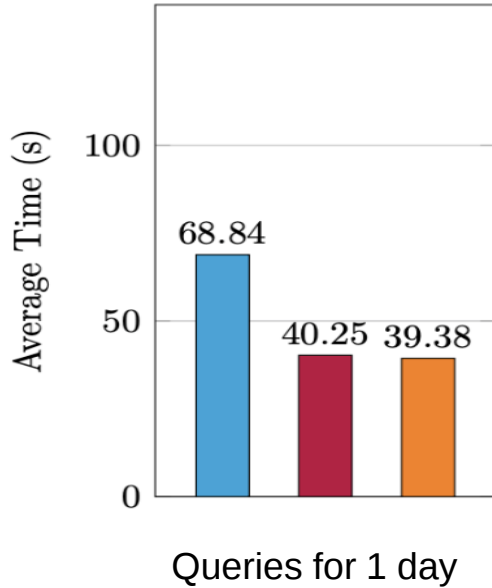
■ TileDB Station Only

■ TileDB Station Year

■ TileDB Station Year Month

Results - Query

Average Time (s)



TileDB Station Only

TileDB Station Year

TileDB Station Year Month

Conclusion

- **TileDB looks as a promising solution to replace RINEX files for storing (not disseminating) GNSS observations**
- **Several implementations are already in use worldwide, particularly at EarthScope**
- **However, further research is still needed to improve its operational use.**
- **We focus here on the granularity of the TileDB to store data for a station.**
- **The results show the Station-Year is the best compromise for the granularity since it improves the retrieval of the data without the need to create too many TileDBs**

Thank you for your attention!

fernando.geraldes@segal.ubi.pt