



Performance Evaluation of TimescaleDB for Storage of Historical Data from WinCC OA SCADA Systems

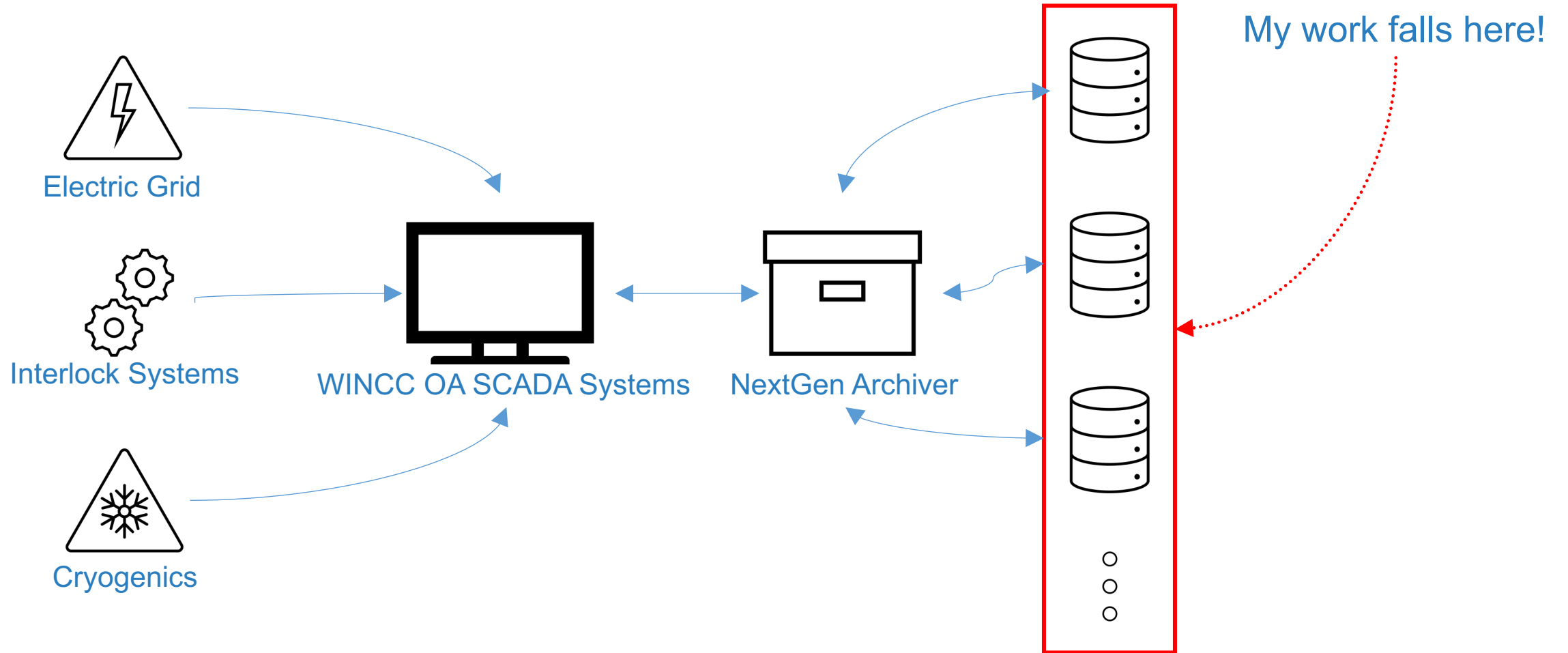
CERN openlab summer student lightning talk session

Mehant Kammakomati

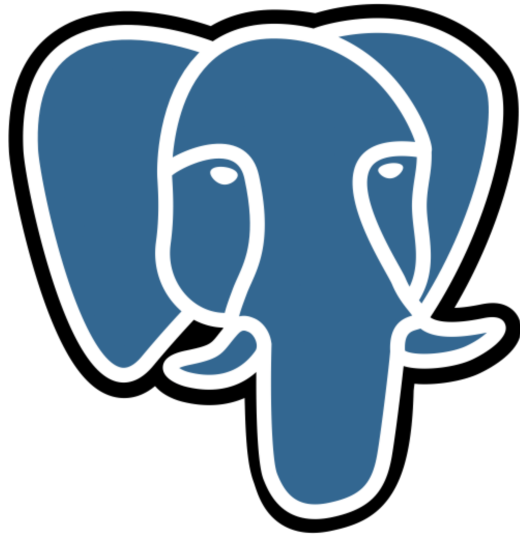
Supervisors: Rafal Kulaga, Anthony Hennessey

07 / 09 / 2021

What This Project Is About?



TimescaleDB Database



+

Extends PostgreSQL to allow performant storage of time-series data

Time-series oriented features

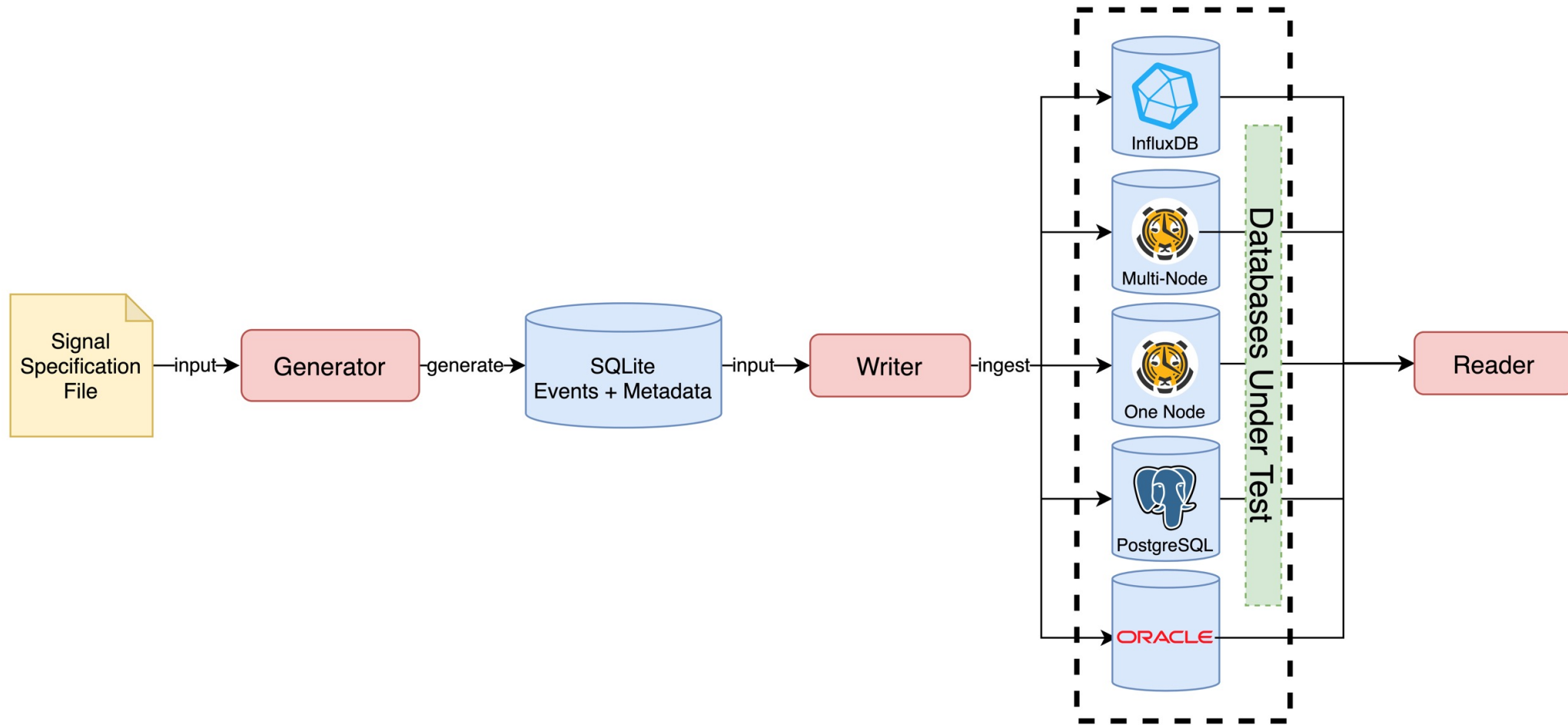
Multi-node variant

Time- and space-based partitioning

=



Benchmarking Methodology



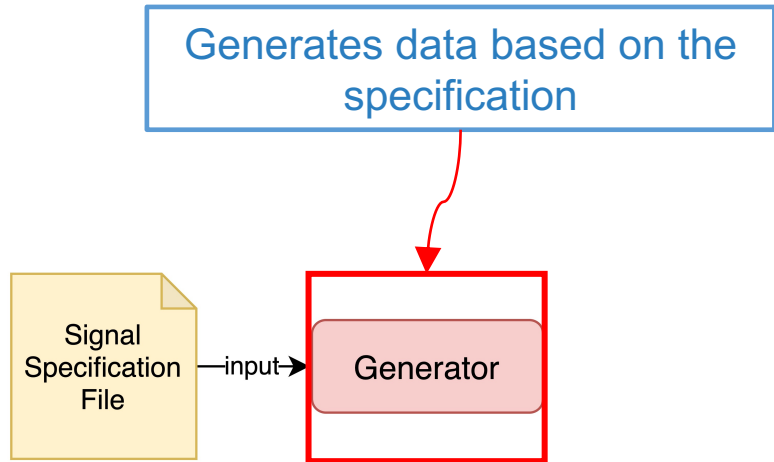
Benchmarking Methodology

- How many slow and fast changing signals?
- What is the duration of data to be generated?
- Single and vector valued events
Etc.



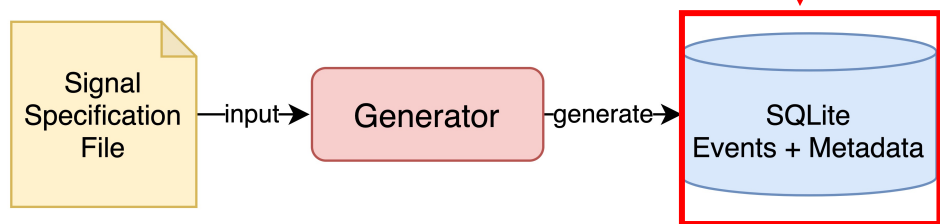
Signal
Specification
File

Benchmarking Methodology

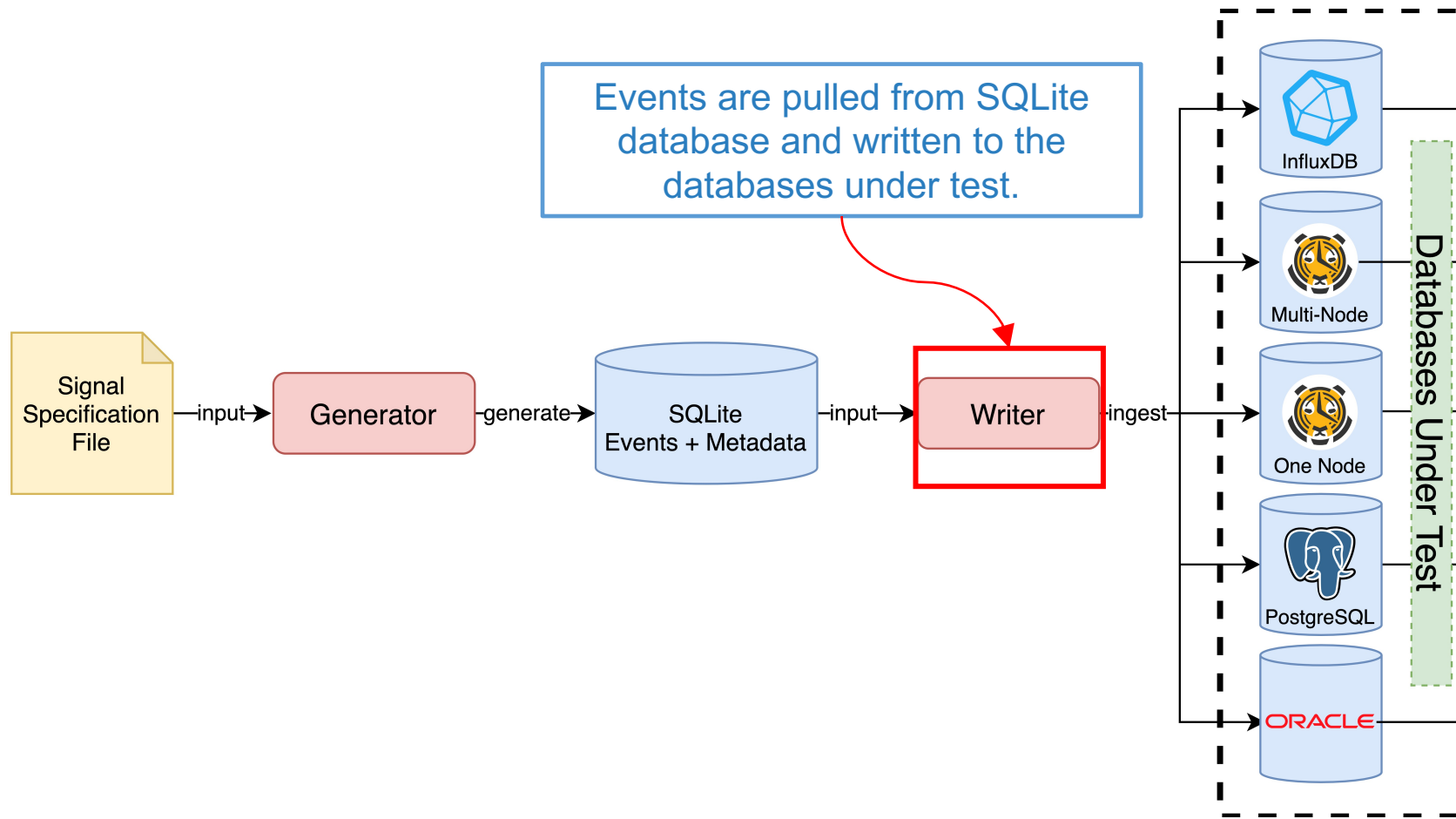


Benchmarking Methodology

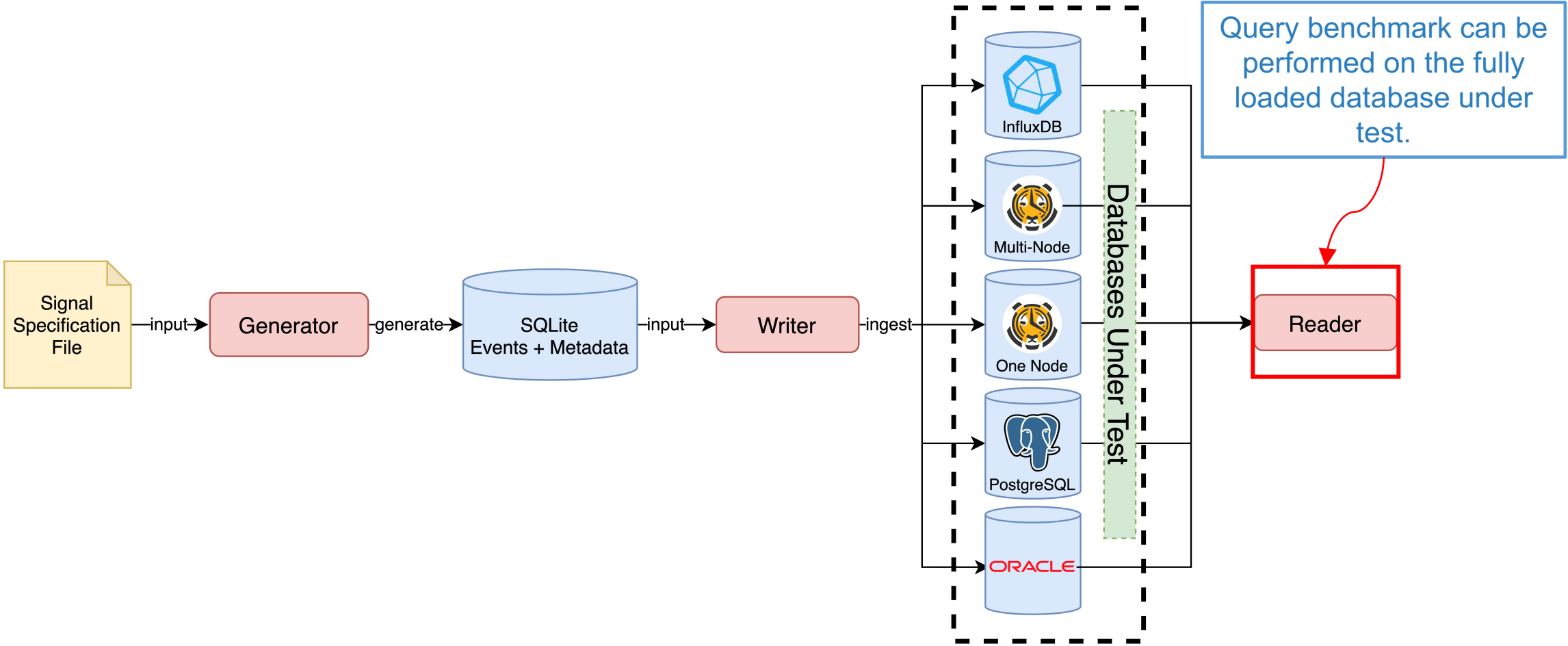
Generated data is stored here and allows us to do any sort of analysis through SQL. This way is better than storing in a flat file.



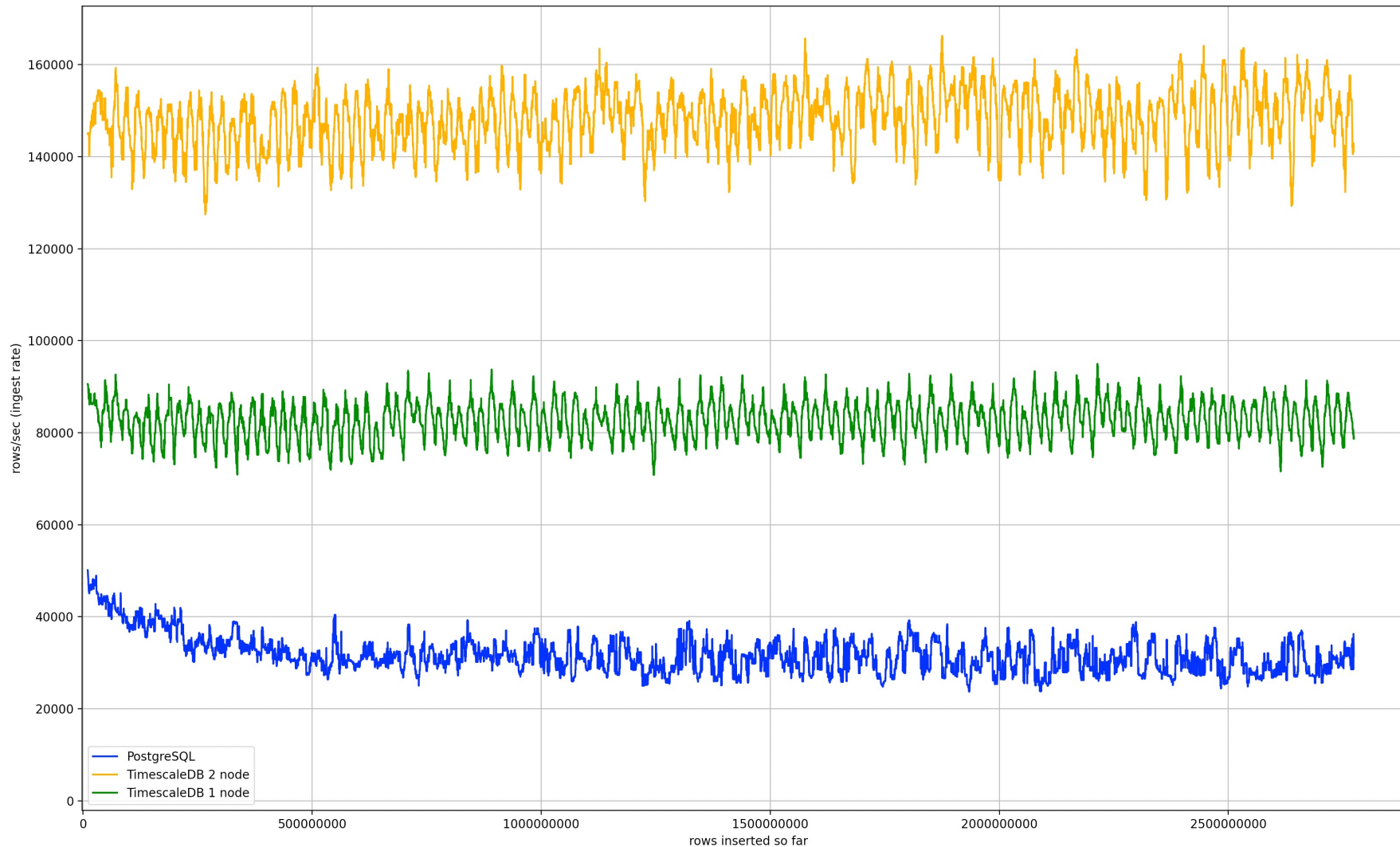
Benchmarking Methodology



Benchmarking Methodology



First Results: Data Ingestion Benchmark



Benchmarking Specs

16GB of Memory

1TB of Storage

250MB/Sec Storage Performance

8 CPU Cores

Disk Space Consumption

PostgreSQL	451GB
TimescaleDB 1 Node	427.8GB
TimescaleDB 2 Node	430.5GB

Summary and Future Work

Summary

- Benchmarking methodology
- Promising first results for ingestion benchmark of TimescaleDB
- Informal query tests show very good performance for some queries on TimescaleDB, on the other side some important queries have better performance on PostgreSQL

Future Work

- Perform read benchmarks
- Experiment with different schema variants
- Run the tests on more performant hardware; more nodes in the TimescaleDB installation
- Compare the results with other DBs supported by the NGA: Oracle and InfluxDB



QUESTIONS?

kmehant@gmail.com



<https://www.linkedin.com/in/mehant-kammakomati-1a0b41170/>



<https://twitter.com/mehant1>