



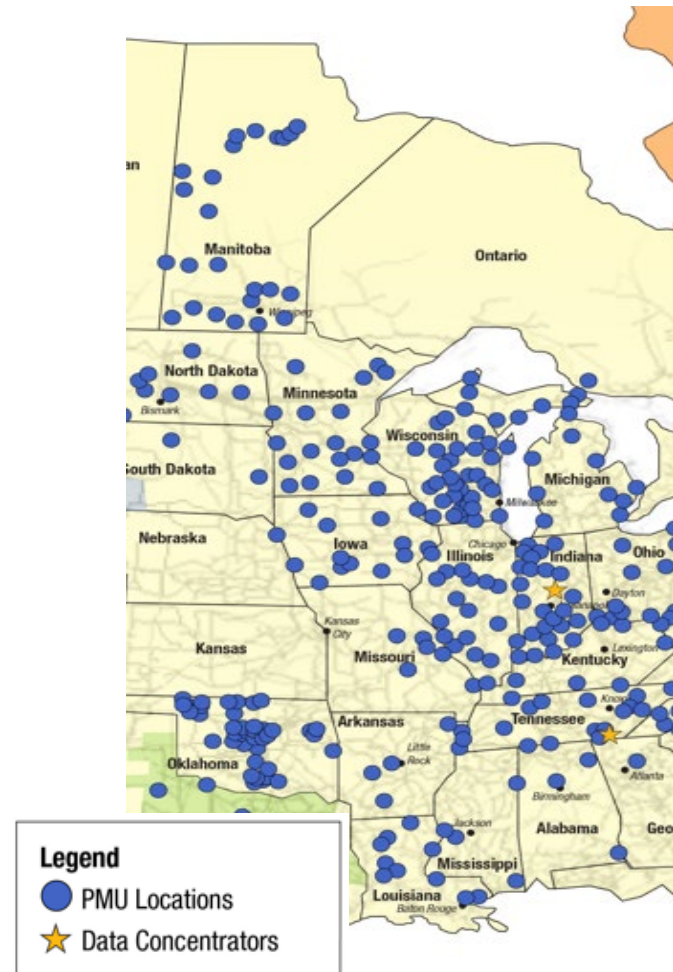
Shrinking MISO's Synchrophasor Data

NASPI Spring Conference

April 14, 2022

MISO Synchrophasors at a Glance

- MISO established its synchrophasor program in 2011 as part of a DOE grant
- Internal PMUs: 572
External PMUs: 70
- Not currently operationalized or in CIP



Current Data Footprint

- Current setup uses Oracle with daily partitions
- Nearly 200 TB of total data
- 250 TB of dedicated SAN
- Performance is mediocre
- This is expensive!

Measurement Type	Retention	Size
ABC	6 months	33 TB
PosSeq	1 year	60 TB
Down Sampled	2 years	5 TB
Total Per Site	-	98 TB
Grand Total	-	196 TB

Pivoting to a TSDB

Time-series databases (TSDB) have become mainstream since 2011

Historian vs TSDB has blurred

Can we do better?

Considered Storage Options

	Pro	Con
Upgrade Oracle	No rework	<ul style="list-style-type: none"> • Many...
OSIsoft PI	<ul style="list-style-type: none"> • Widely used for phasor data 	<ul style="list-style-type: none"> • Licensing • Better compression available
InfluxDB	<ul style="list-style-type: none"> • Excellent Compression • First class time-series 	<ul style="list-style-type: none"> • One-off support • Platform specific clients
TimescaleDB	<ul style="list-style-type: none"> • Postgres extension • Excellent Compression • First class time-series • SQL! 	<ul style="list-style-type: none"> • More manual setup
Parquet	<ul style="list-style-type: none"> • Standard big data format • Good compression 	<ul style="list-style-type: none"> • Manual integrations • Desirable features are not yet mainstream
Custom Format (Protobuf)	<ul style="list-style-type: none"> • Smallest data footprint 	<ul style="list-style-type: none"> • Many custom integrations
Others (Cassandra/OpenHistorian/Etc)	-	<ul style="list-style-type: none"> • One-off support • Reference Architecture

RDBMS vs Big Data

- Quickly narrowed down to Parquet and Timescale
- **Parquet: +Seq @ 1 PMU/day**
 - Raw (40 MB)
 - Gzip (25 MB)
 - Gzip + Delta Encoding (8 MB)
 - Limited implementations
- **TimescaleDB: +Seq @ 1 PMU/day**
 - Uncompressed: 267 MB
 - Compressed: 27 MB

Compressed size w/o
indexes only 0.5 MB

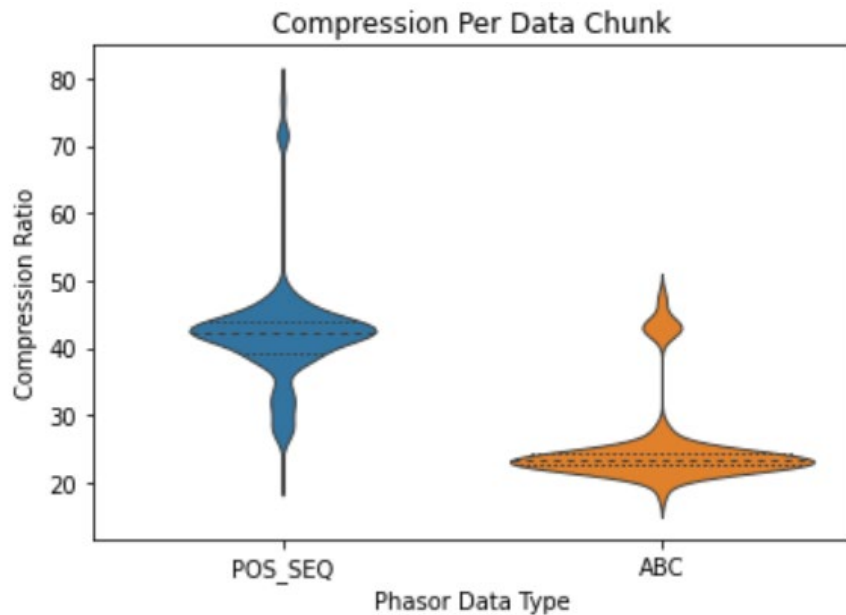
Choosing Column Types

	Float	Integer	Numeric
Angle - Current	13	44	11
Angle - Voltage	16	58	14
Magnitude - Current	18	60	17
Magnitude - Voltage	17	54	18
Frequency - Magnitude	26	54	20
Frequency - Delta	19	52	15
Quality	40	122	34

Compression Ratio	20	40	60	80	100	120
Compression Factor	0.05	0.025	0.016667	0.0125	0.01	0.008333

New Data Footprint

	ABC	PosSeq	Down	Total
Timescale Before	28.7 TB	55.7 TB	4.0 TB	88.5 TB
Timescale After	1.2 TB	1.3 TB	0.1 TB	2.6 TB
Compression Ratio	25	43	32	34
Compression Factor	4.1%	2.3%	3.1%	2.9%



Oracle's 98 TB
per dataset
to 2.6 TB!

TimescaleDB Experience

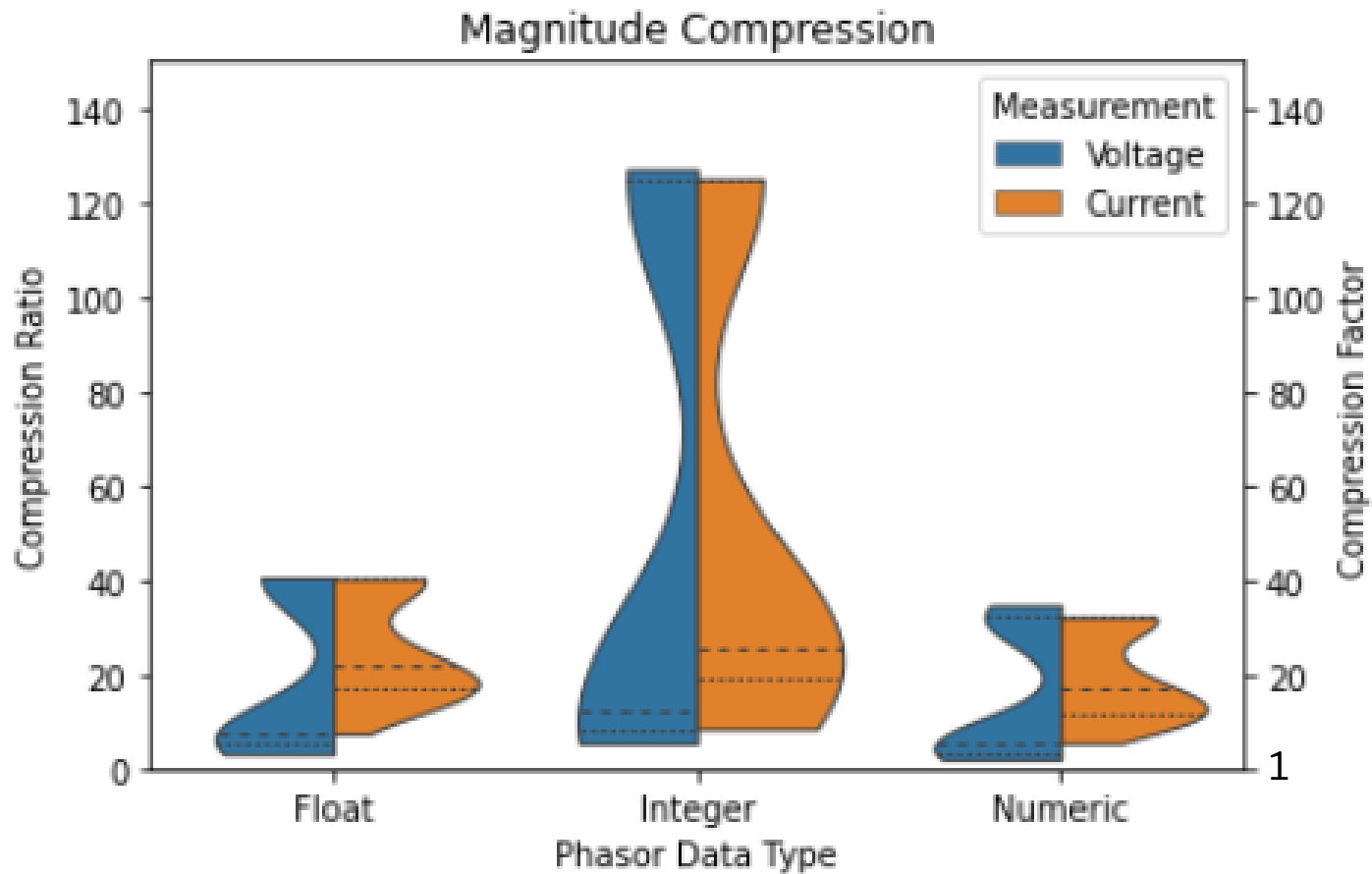
- Built on Postgres
 - Get a bonus mainstream RDBMS for metadata
 - Postgres replication and backup tools
 - Standard SQL and clients
 - Transition from Oracle RDBMS relatively easy
- Fine grain control over compression
- Easy SQL to convert to fixed-point integer storage
- Performance tested >1m rows/sec
- Massive compression savings with better performance
- Parquet looks very promising for cold-storage

Questions

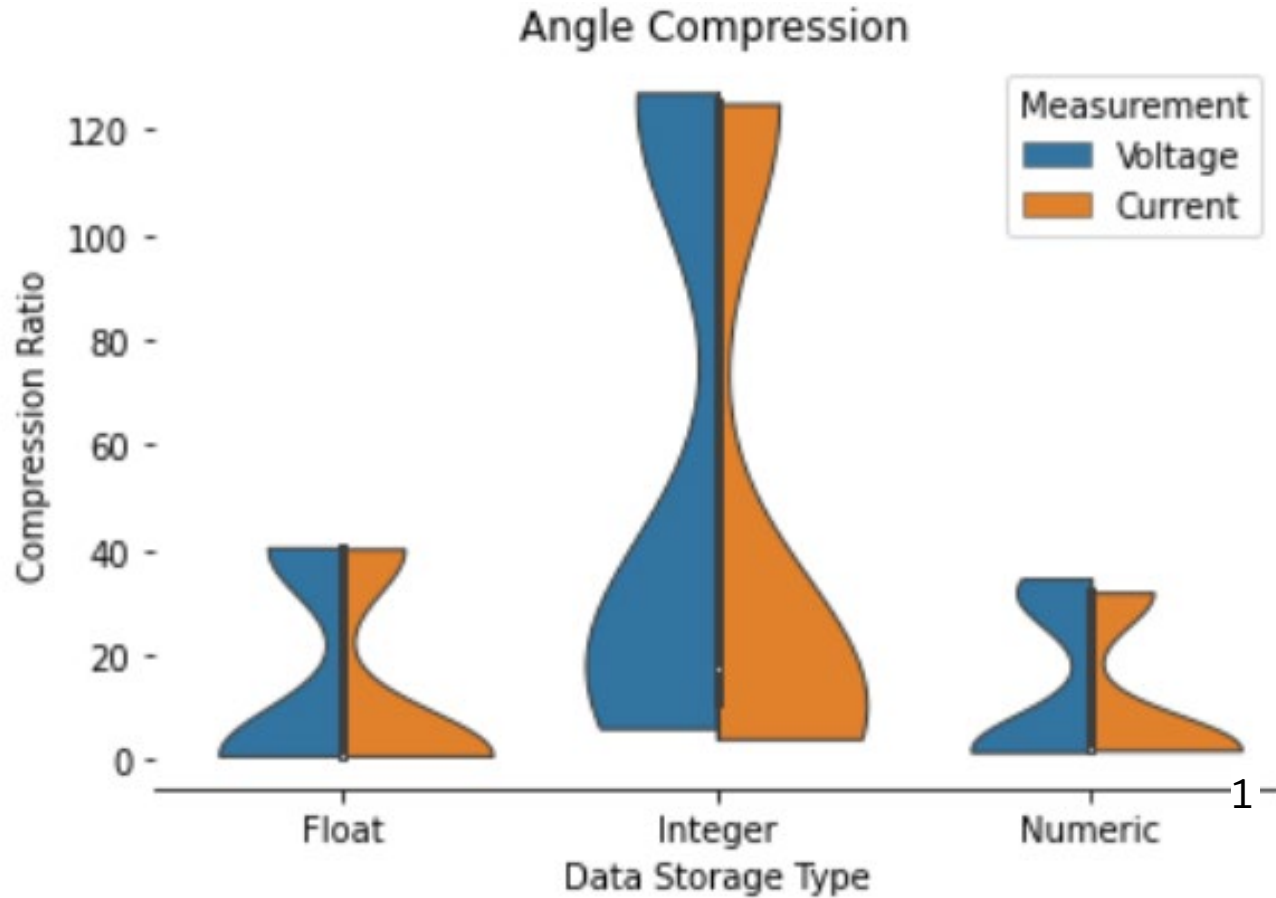
- bkiefer@misoenergy.org

Appendix

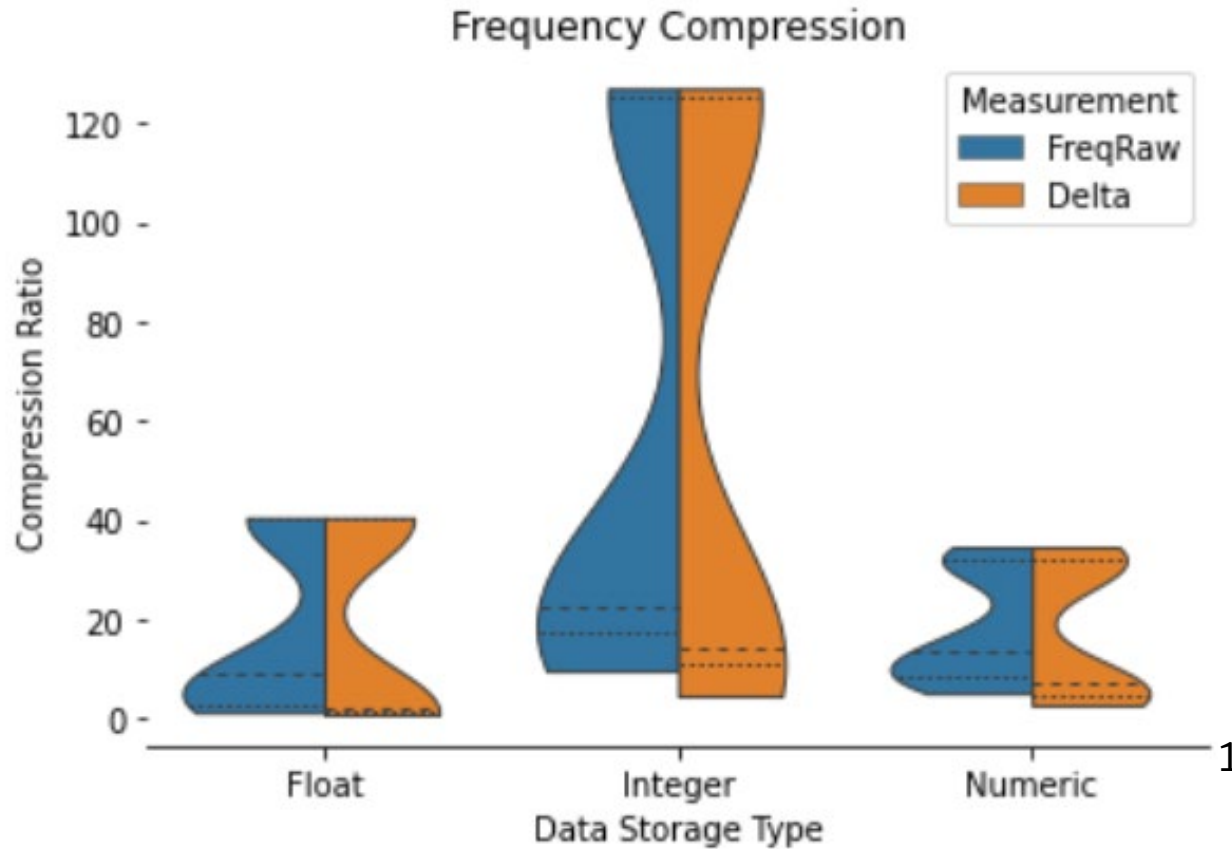
Compression Performance



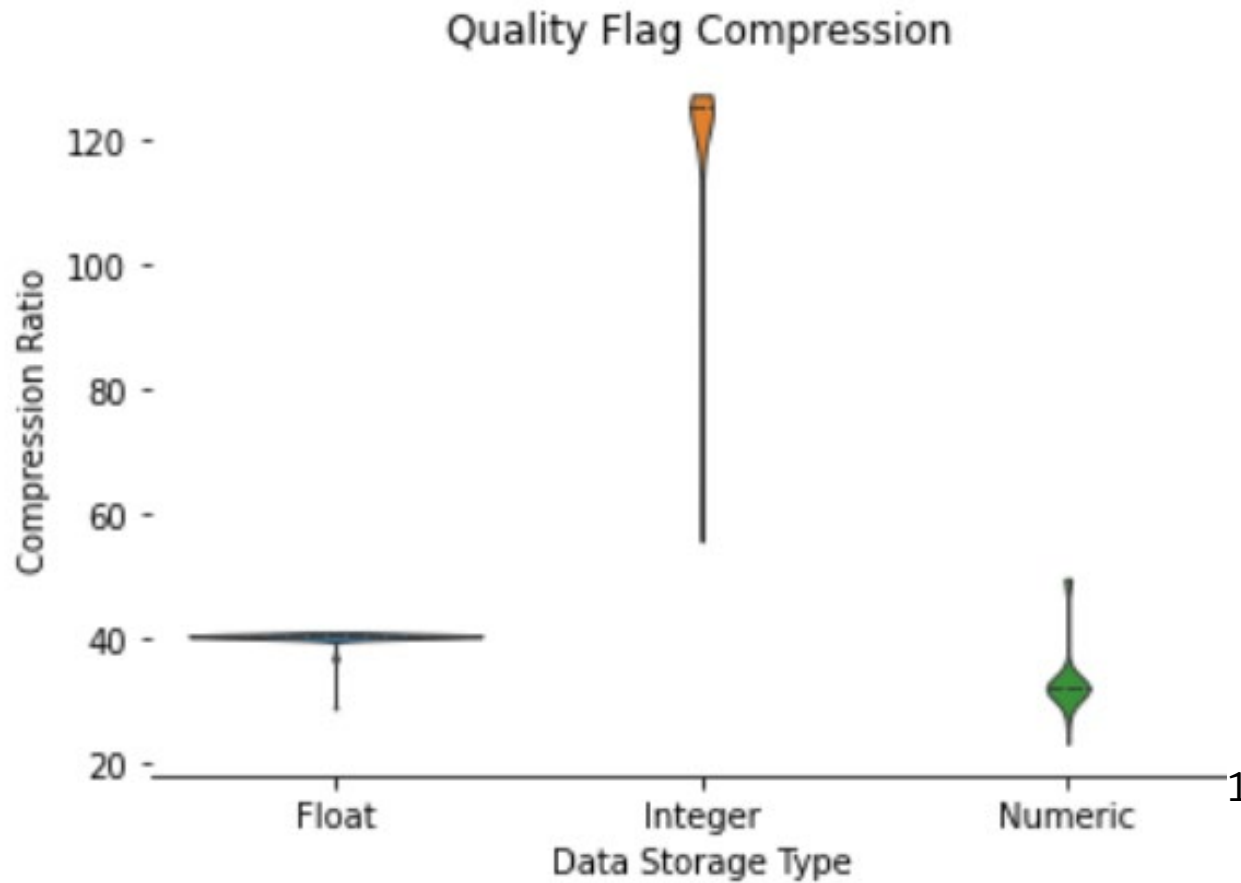
Compression Performance



Compression Performance



Compression Performance



TimescaleDB Compression

- <https://www.timescale.com/blog/time-series-compression-algorithms-explained/>