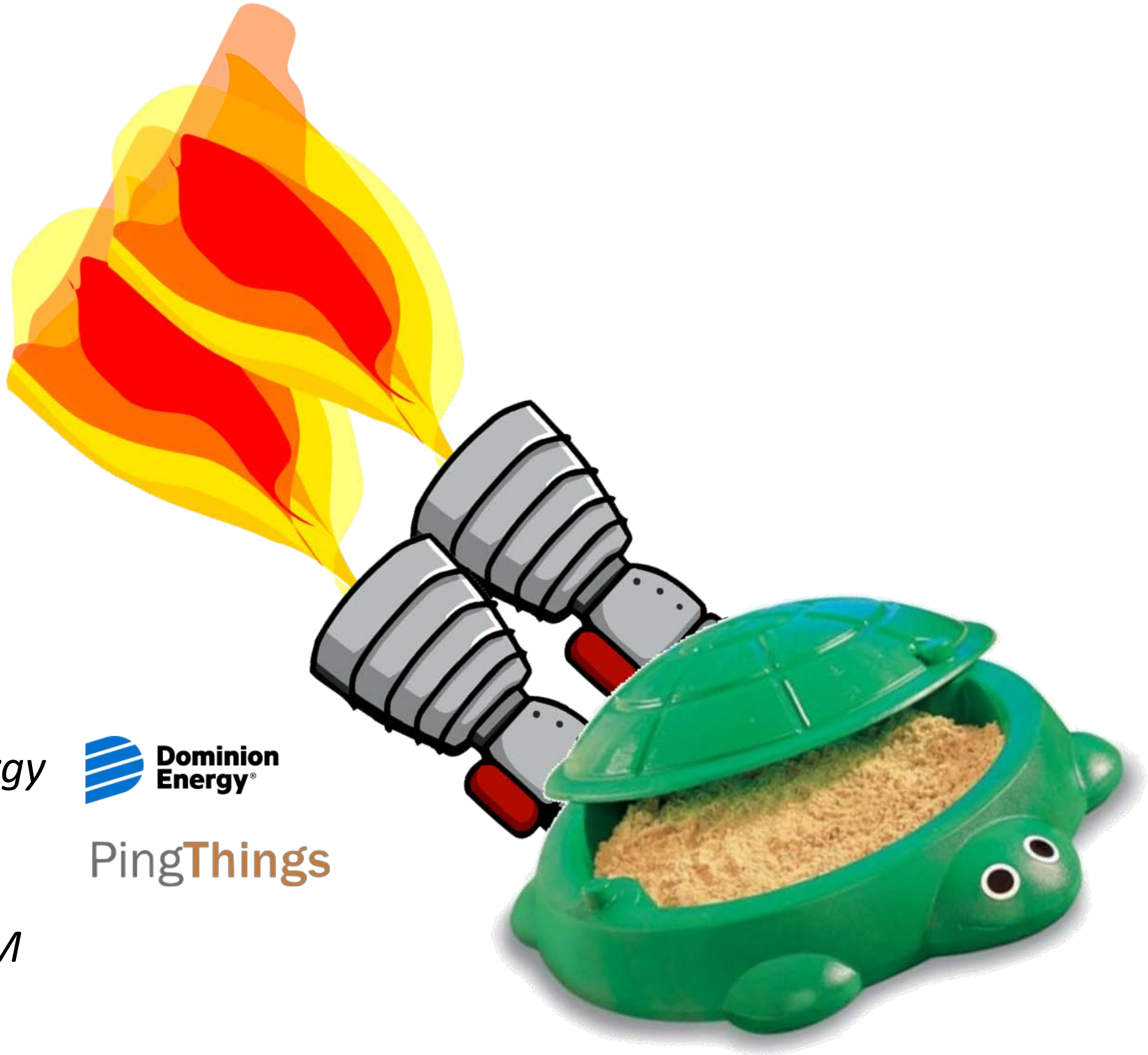
The Role of a High Performance Sandbox in Your Synchrophasor Analytics Pipeline



Kevin D. Jones, Ph.D. – Dominion Energy
Sean Murphy – PingThings

Dominion Energy

PingThings

NASPI Spring 2018 – Albuquerque, NM
April 24, 2018

Grid Scale Synchrophasor Data at Dominion

State of PMUs at Dominion

- **2009** – Kicked off synchrophasor initiative
- **2012** – began organic sensor deployment: *~350 PMUs & growing*
- **2013** – DOE Demonstration
  - Linear State Estimator released as OSS
- **2017**
  - DOE FOA970 - openECA Project
  - Began DFR-PMU configuration: *~total Transmission-system coverage*
  - Identified performance issues with new Enterprise Data Historian
  - Potential solutions measured in YEARS

Grid Scale Synchrophasor Data at Dominion

`(timestamp, value)`

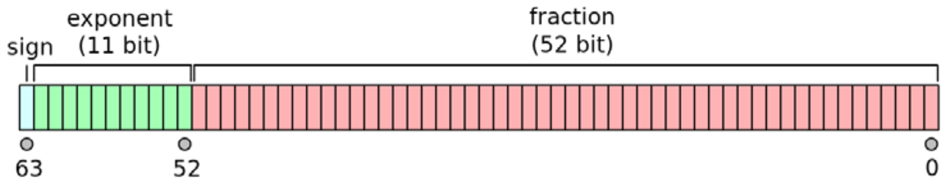
64-bit integer

$[-(2^{63}),\ 2^{63}-1]$

Or

[-9,223,372,036,854,775,808,
9,223,372,036,854,775,807]

64-bit float



$$(-1)^{\text{sign}}(1.b_{51}b_{50}\ldots b_0)_2 \times 2^{e-1023}$$

Grid Scale Synchrophasor Data at Dominion

Measurements Per Second

# of Streams	0.1 Hz	1 Hz	10 Hz	100 Hz	1 KHz	10 Khz	100 Khz	1 MHz	10 MHz	100 MHz	1 GHz
100000	10 k	100 k	1 M	10 M	100 M	1 B	10 B	100 B	1 T	10 T	100 T
10000	1 k	10 k	100 k	1 M	10 M	100 M	1 B	10 B	100 B	1 T	10 T
1000	100	1 k	10 k	100 k	1 M	10 M	100 M	1 B	10 B	100 B	1 T
100	10	100	1 k	10 k	100 k	1 M	10 M	100 M	1 B	10 B	100 B
10	1	10	100	1 k	10 k	100 k	1 M	10 M	100 M	1 B	10 B
1	0	1	10	100	1 k	10 k	100 k	1 M	10 M	100 M	1 B

Smart Meter SCADA PMU μPMU DFR

Sampling Rates

Ones	Thousands	Millions	Billions	Trillions

Grid Scale Synchrophasor Data at Dominion

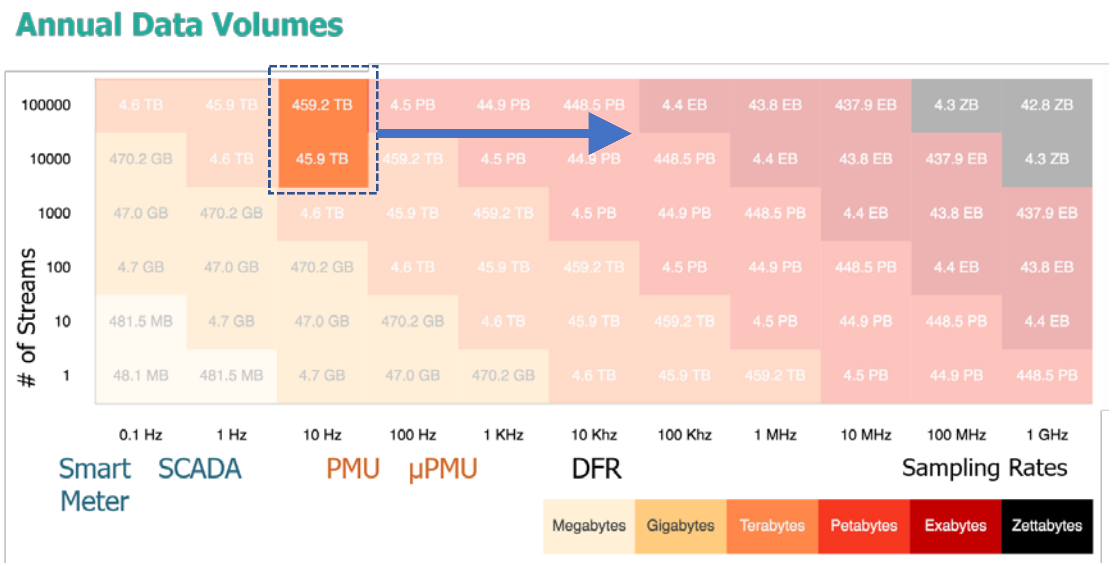
Annual Data Volumes

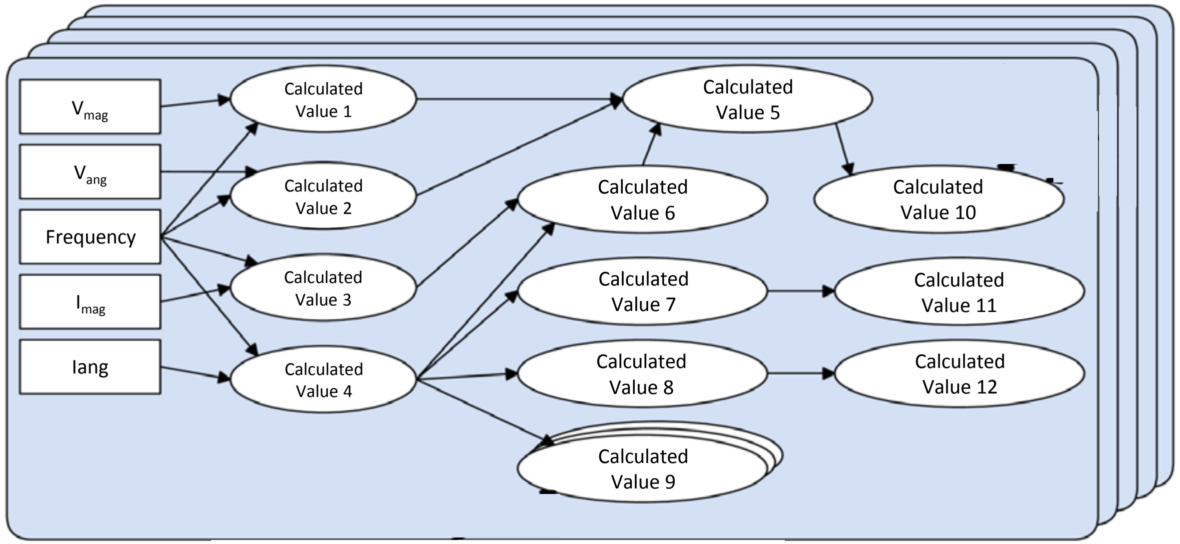
# of Streams	0.1 Hz	1 Hz	10 Hz	100 Hz	1 KHz	10 Khz	100 Khz	1 MHz	10 MHz	100 MHz	1 GHz
100000	4.6 TB	45.9 TB	459.2 TB	4.5 PB	44.9 PB	448.5 PB	4.4 EB	43.8 EB	437.9 EB	4.3 ZB	42.8 ZB
10000	470.2 GB	4.6 TB	45.9 TB	459.2 TB	4.5 PB	44.9 PB	448.5 PB	4.4 EB	43.8 EB	437.9 EB	4.3 ZB
1000	47.0 GB	470.2 GB	4.6 TB	45.9 TB	459.2 TB	4.5 PB	44.9 PB	448.5 PB	4.4 EB	43.8 EB	437.9 EB
100	4.7 GB	47.0 GB	470.2 GB	4.6 TB	45.9 TB	459.2 TB	4.5 PB	44.9 PB	448.5 PB	4.4 EB	43.8 EB
10	481.5 MB	4.7 GB	47.0 GB	470.2 GB	4.6 TB	45.9 TB	459.2 TB	4.5 PB	44.9 PB	448.5 PB	4.4 EB
1	48.1 MB	481.5 MB	4.7 GB	47.0 GB	470.2 GB	4.6 TB	45.9 TB	459.2 TB	4.5 PB	44.9 PB	448.5 PB

Smart Meter | SCADA | PMU | µPMU | DFR | Sampling Rates

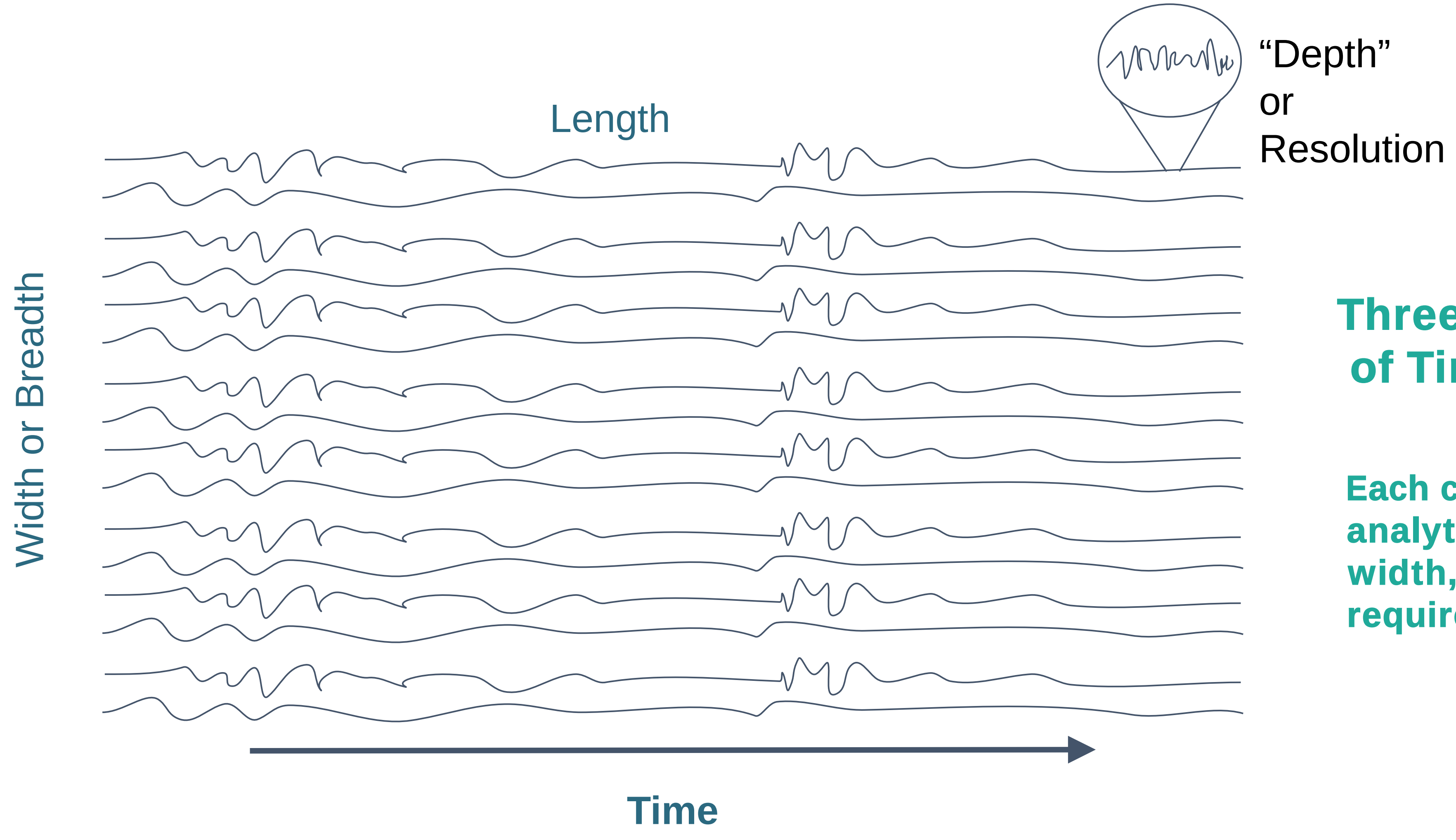
Megabytes | Gigabytes | Terabytes | Petabytes | Exabytes | Zettabytes

Grid Scale Synchrophasor Data at Dominion





Grid Scale Synchrophasor Data at Dominion



Three "Dimensions" of Time Series

Each class of time-series analytics has different width, length, and depth requirements

Grid Scale Synchrophasor Data at Dominion

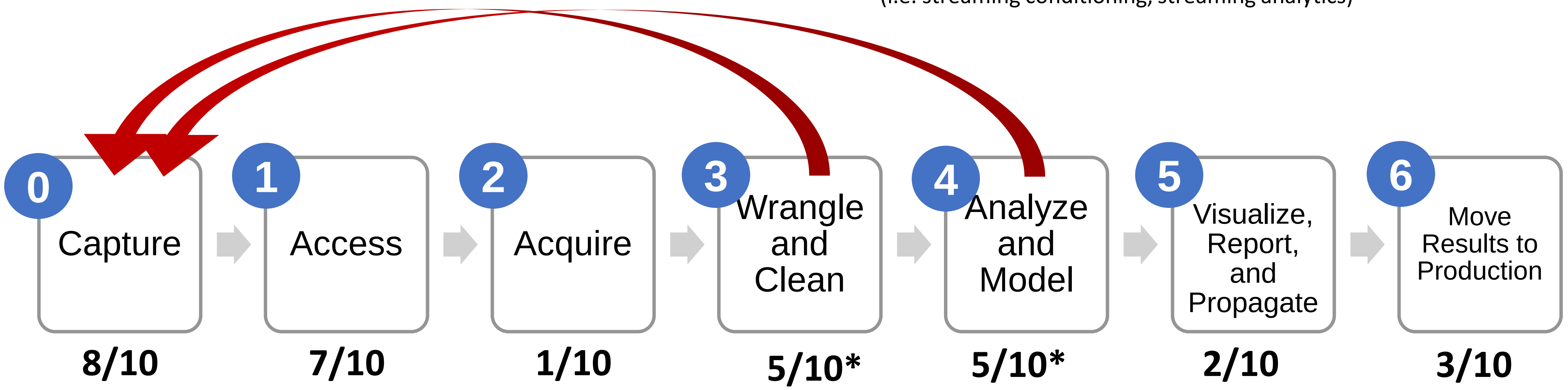
Some Existential Concerns

- Can we deliver on value promised? If so, can we deliver with a reasonable lead time?
- Are we pushing the boundaries of what is computationally feasible?
- What *is* computationally feasible? What, if any, are the true technological limitations at the bleeding-edge?
- How can we take control of our analytic goals and timetable?

To answer these questions and more, I began my search…

The [Synchrophasor] Analytics Pipeline†

*We get a 5 because of the success we've had some success pushing this back into the stream (i.e. streaming conditioning, streaming analytics)

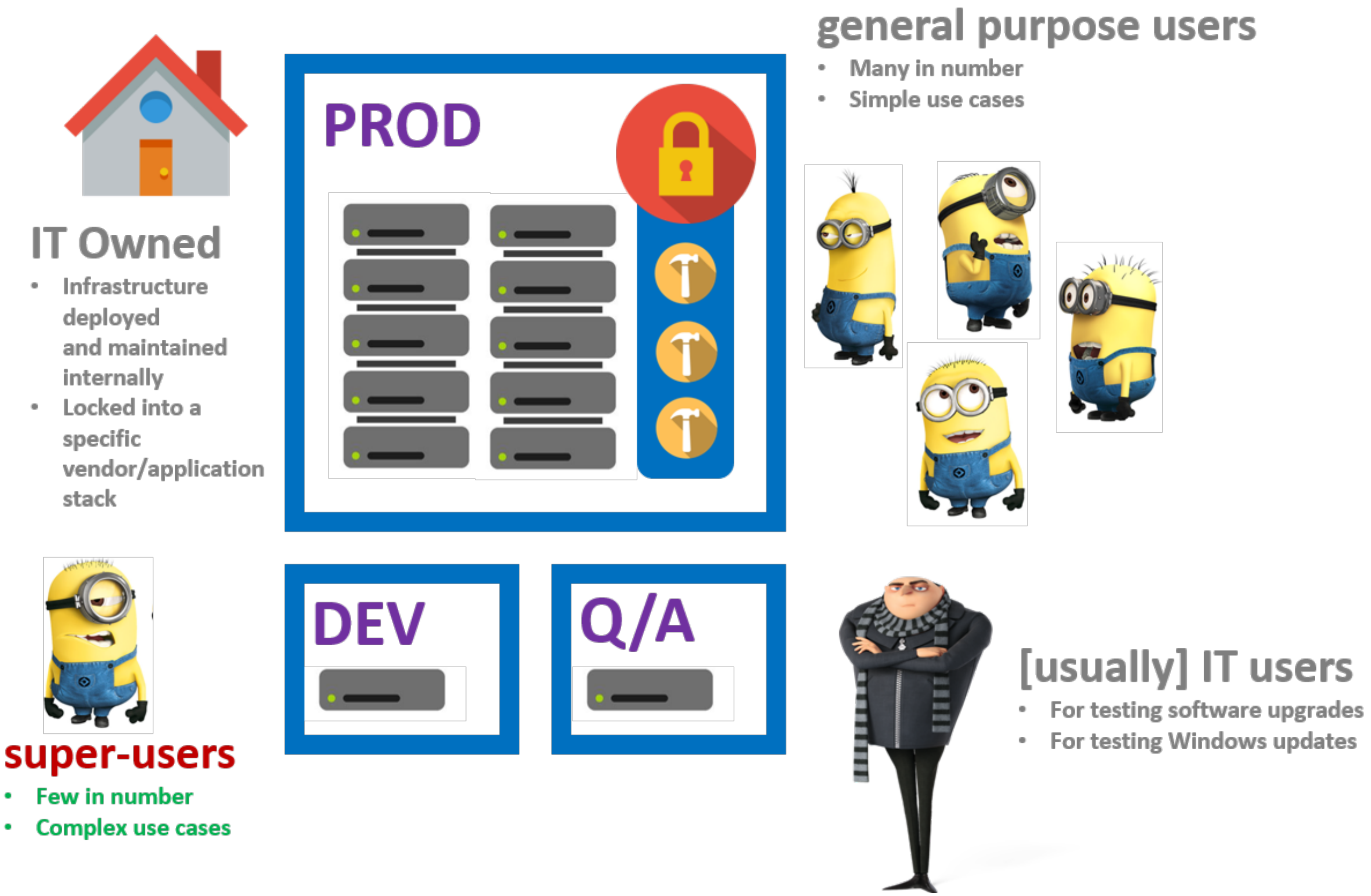


Each analytic must progress through all of these stages.

We don't "solve" one block for all time and only then move onto the next – we need to increase the speed and frequency that we iterate through the cycle

† Sean Murphy & Kevin D. Jones, *"Learning From Data"*, CIGRE Grid of the Future 2017

For [Synchrophasor] Data Analytics, We Build Systems Upside Down



Traditional Model

- IT owned and controlled.
- Can use only what the vendor makes available.
- The generalized "production" use cases limits growth and innovation.
- A Windows environment means bloated systems from IT oversight and limited privileges
- A "production" system requires YEARS to [painfully] modify and scale

Jevons Paradox

In 1865, a twenty-nine-year-old Englishman named William Stanley Jevons published a book, "The Coal Question," in which he argued that the bonanza couldn't last. Britain's affluence, he wrote, depended on its endowment of coal, which the country was rapidly depleting. He added that such an outcome could not be delayed through increased "economy" in the use of coal—what we refer to today as energy efficiency. He concluded, in italics, "*It is wholly a confusion of ideas to suppose that the economical use of fuel is equivalent to a diminished consumption. The very contrary is the truth.*"*

"occurs when technological progress increases the efficiency with which a resource is used (reducing the amount necessary for any one use), but the rate of consumption of that resource rises because of increasing demand"

https://www.newyorker.com/magazine/2010/12/20/the-efficiency-dilemma

A Vision for a High-Performance Sandbox

I needed to find a way to :

- Make exploration of data easier and quicker.
- Accelerate timelines of technology adoption and analytic value extraction
- Acquire full control over the analytic stack and computing environment
- Enable better engagement with collaborators.
- Develop internal expertise on contemporary analytic classes like machine learning
- Develop internal expertise on next-generation data systems.
- Avoid the pitfalls of “*one place for all of our data*”

A Vision for a High-Performance Sandbox

High-Performance

=

Most would think → *More CPU, More RAM, More HDD*

I think → *Whatever resources and architecture it takes to make a team high-performing*

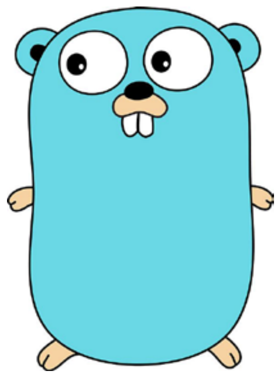
Sandbox

=

Most would think → *Non-critical, non-important, non-value creating, non-secure*

I think → *Flexibility, Imagination, Adaptation, Secure as it needs to be, high-value creating potential*

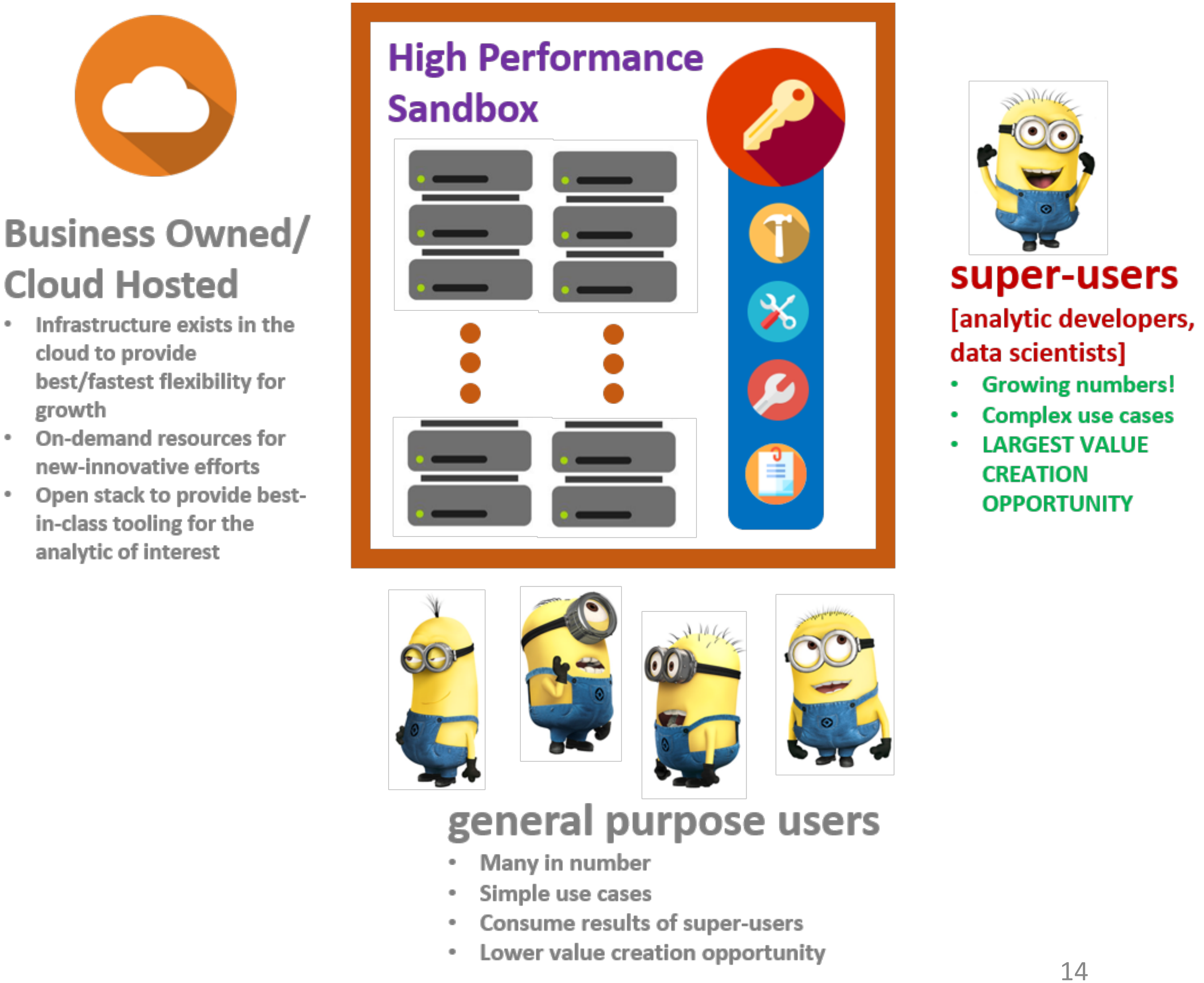
Example of this Philosophy: Google's **Golang**



The High Performance Sandbox Inverts the Traditional Model

High Performance Sandbox Model

- Business owned and controlled
- Emphasizes computing requirements of value creating activities for super-users
- “Super-users” takes on new meaning
- Important for synchrophasors because *the whole stack matters*!!!
- Cloud is not necessary for an HPS but very powerful and strategic.
- A true manifestation of the analytics pipeline is actually a high performance sandbox.

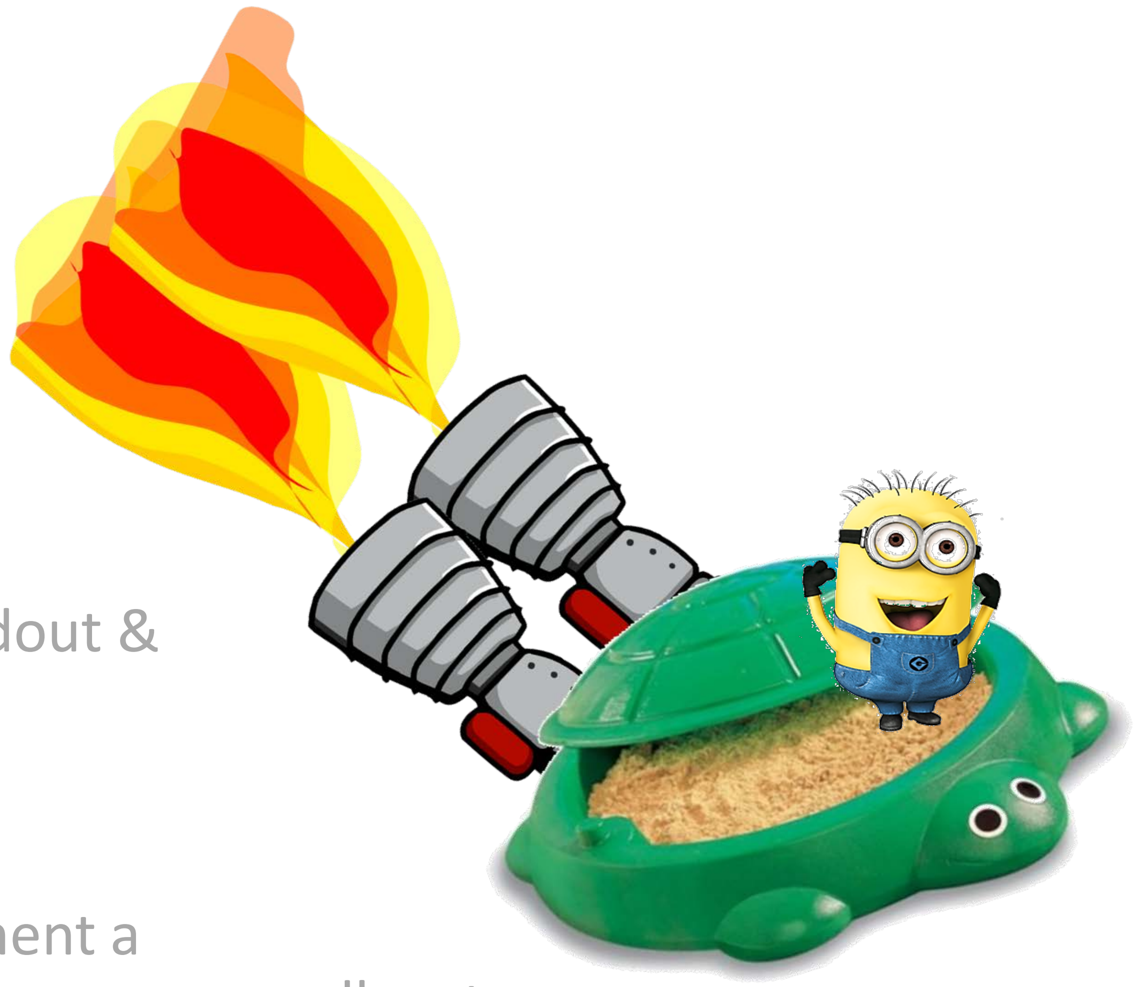


Until Next Time…

Lessons Learned

- We required a change in *philosophy* *AND* *technology*
- From the perspective of time-series data analytics:
  - Yes. Synchrophasors are special
  - We've only discovered the tip of the iceberg relative to what is possible
- Design/technology considerations are necessary at each level of the stack. Can't be bolted onto legacy platforms.
- Doing computing right for synchrophasor data analytics could help utilities do computing right for so many other applications.

Until Next Time…



Next Steps for Dominion

- Continue integration of DFR PMUs and relay-PMUs (ongoing)
- Working towards distribution PMU buildout & other types of sensor deployments
- Continue to utilize existing subsystems of our analytics pipeline
- Work with Sean & PingThings to implement a [preferably AWS cloud based] high-performance sandbox to complete our analytics pipeline.
- Work to integrate other high-resolution time series: DFR, distribution PMU