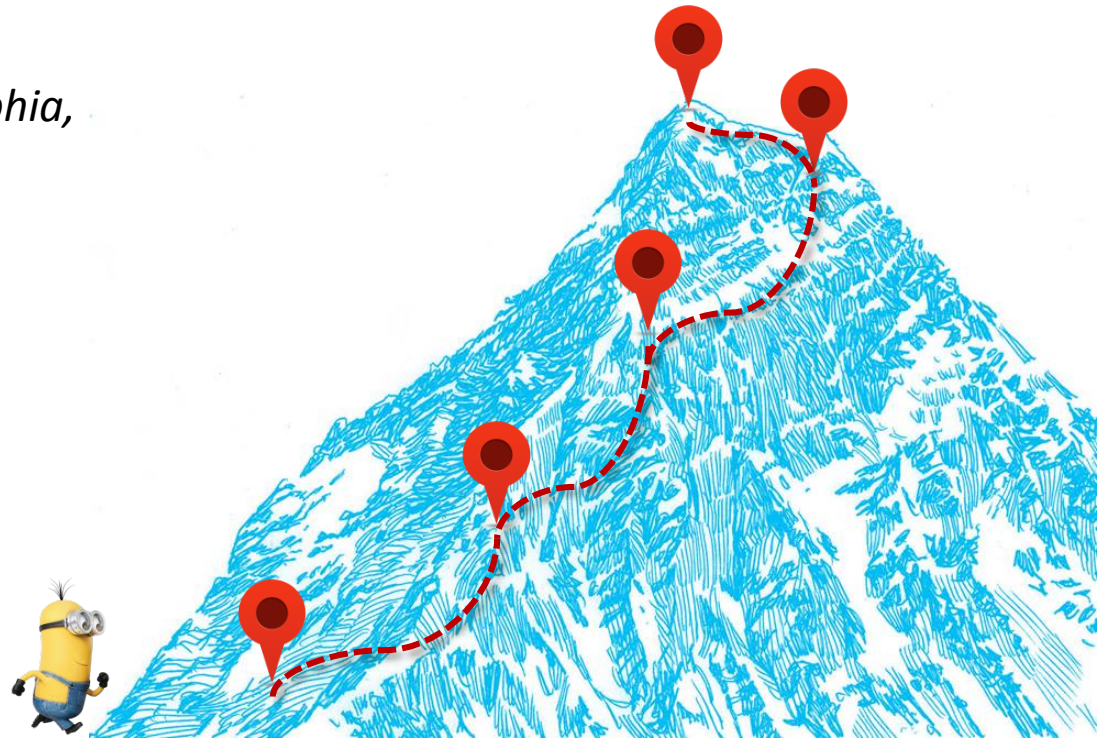
Getting Beyond Base Camp:
Scaling Your Synchrophasor Data Mountain

Kevin D. Jones, Ph.D.

October 24, 2018

2018 Fall NASPI, Philadelphia,



What Do I Actually Mean By “Scaling”?

- More PMUs??
- More Historical Data??
- More Compute Power??
- More Use Cases??
- More Success Stories??
- More ROI??

I suggest that all of these metrics of “scaling” are at best nonlinear and at worst superficial.

A Better Way to Scale?

- I still want:
  - More PMUs!!
  - More Historical Data!!
  - More Compute Power!!
  - More Use Cases!!
  - More Success Stories!!
  - More ROI!!

I propose that all of these metrics are derivative of a deeper, more molecular metric that, if considered, can maximize each dimension of interest.

Summit: ??

Camp 3: Analytic Experimentation/Evaluation

Camp 2: Infrastructure & Control

Camp 1: Growing Data Volumes

Base Camp: PMU Deployment

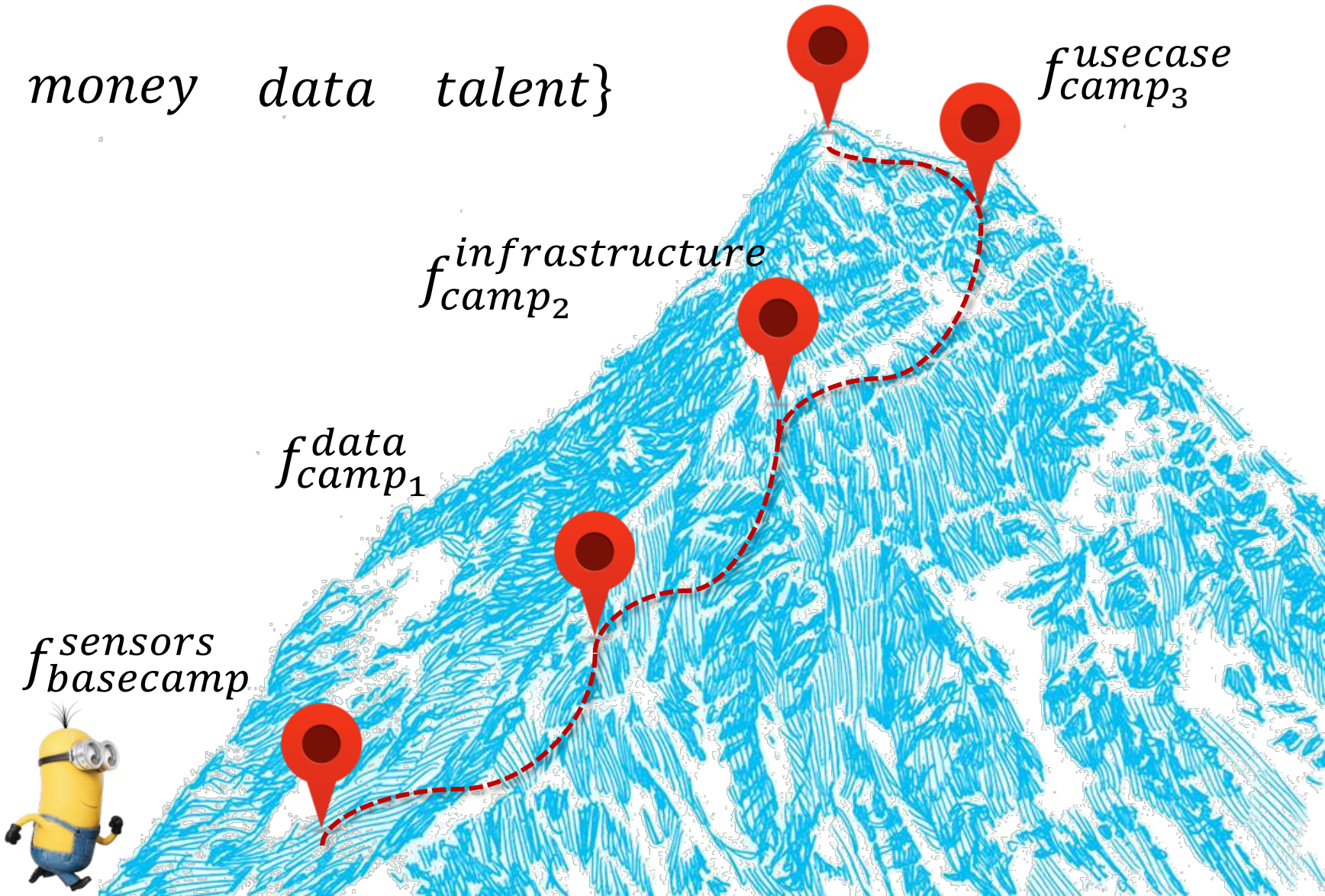
Why Did We Need to Climb??

`resources → value realized`

The function that maps between these two spaces changes as you ascend the mountain - as do the metrics that drive scaling.

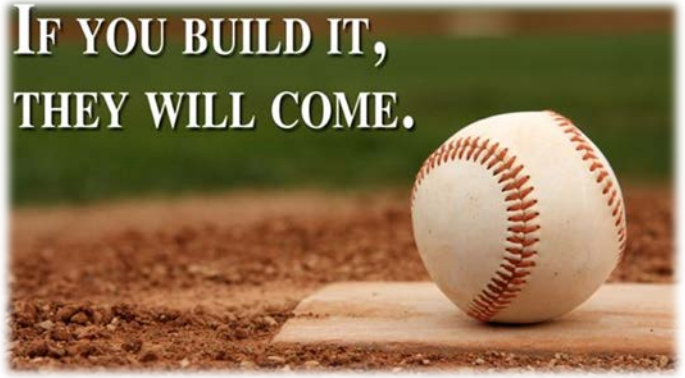
$$outcome = f_{system}^{metric}(resources)$$

$$resources = \{time \quad money \quad data \quad talent\}$$



Base Camp:
PMU Deployment

The Field of PMUs (Devices)



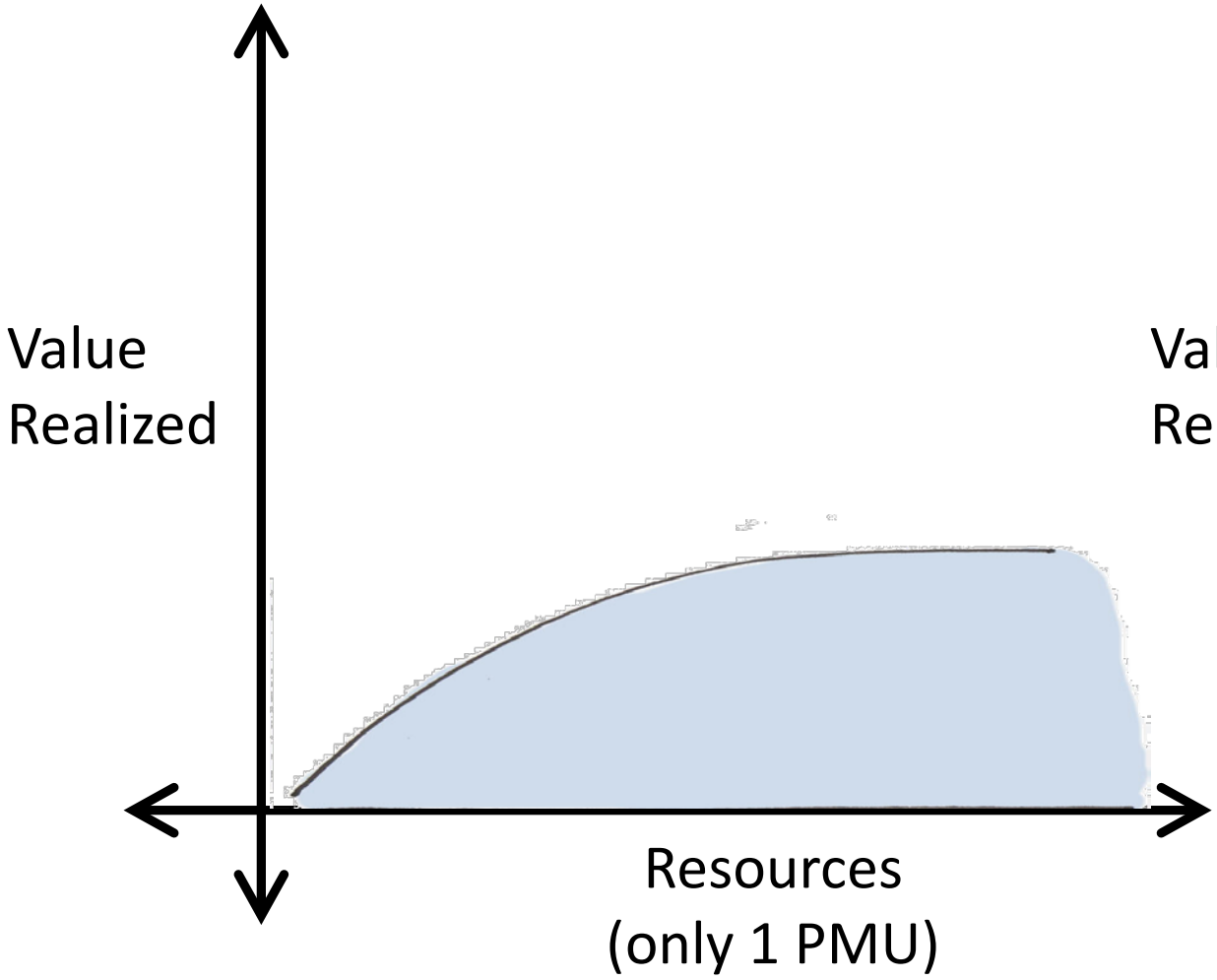
	DFR-PMUs (non-CIP)	Existing Relay-PMUs (CIP)	New Relay-PMUs (CIP)	Total Devices
Currently Streaming	161	40	0	201
Soon to be Streaming	119	488	72	679
Total Field Devices	280	528	72	880
Deadline	End of Q1 2019	EOY 2019	EOY 2019	EOY 2019

The Field of PMUs (Signals)

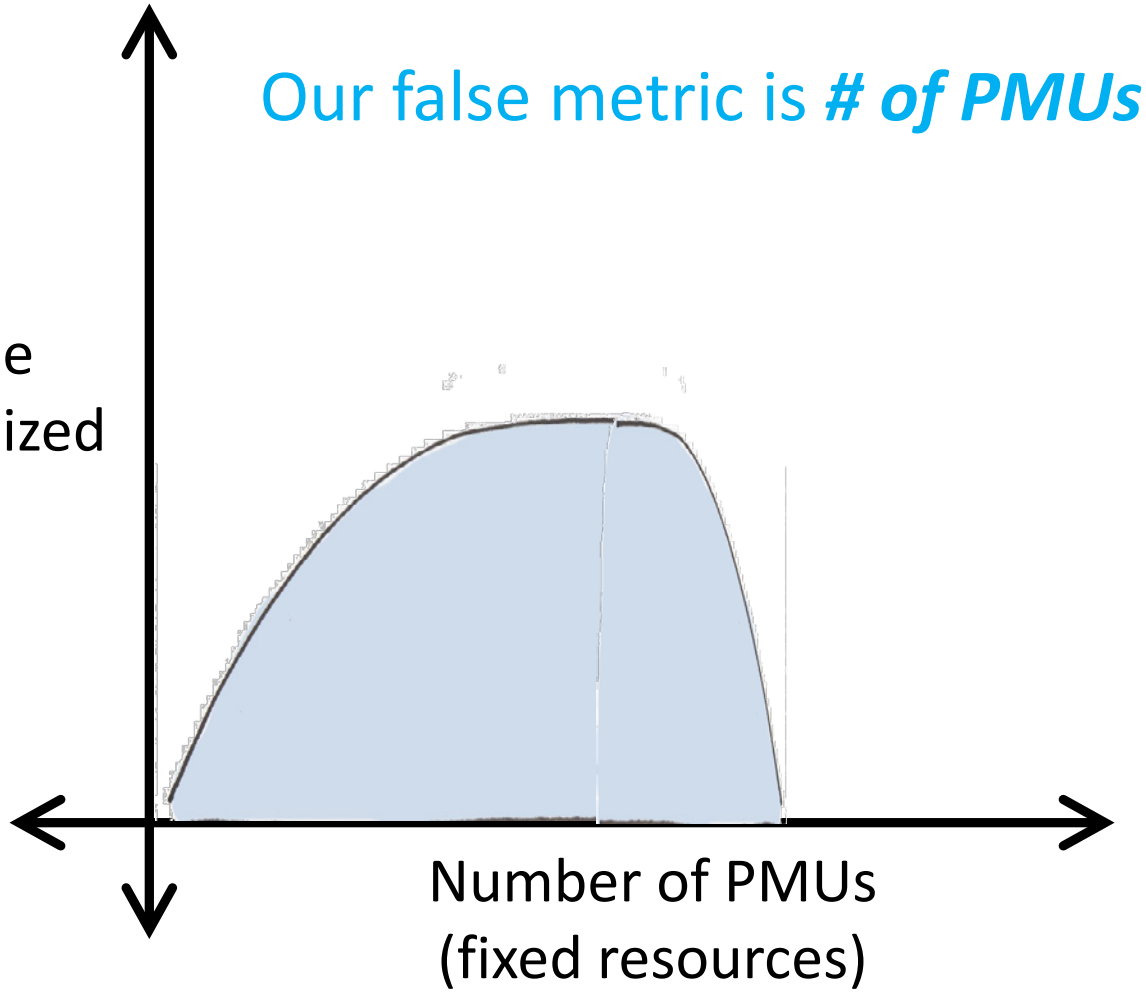
	Average Per DFR PMU	DFR-PMU EOY 2019 Total	Average Per Relay-PMU	Relay-PMU EOY 2019 Total	Total Measurements
Frequency	1	280	1	600	880
DFDT	1	280	1	600	880
Status	1	280	1	600	880
Voltage Magnitude	16	4,480	5	3,000	7,480
Voltage Angle	16	4,480	5	3,000	7,480
Current Magnitude	23	6,440	6	3,600	10,400
Current Angle	23	6,440	6	3,600	10,400
Digital	4	1,120	1	600	1,720
Analog	0	0	0	0	0
Totals	85	23,800	26	15,600	39,400

Scaling Only PMUs: A Proxy for a Bad Cost Curve

Imagine you have 1 PMU, 1 Relay Tech, 1 Engineer, and 1 Computer

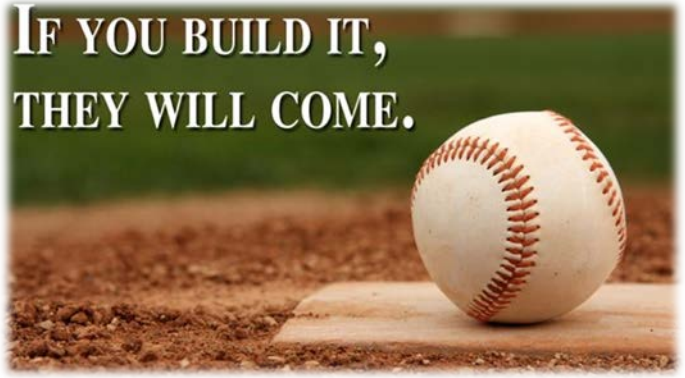


Our false metric is ***# of PMUs***

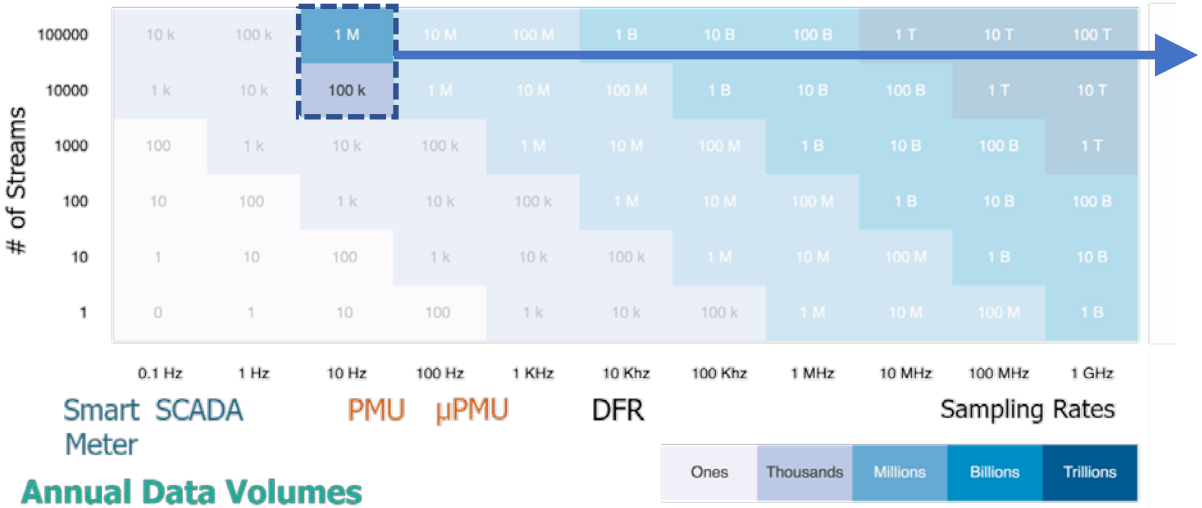


Camp 1:
Growing Data Volumes

The Field of PMUs (Data Rates & Volumes)



Measurements Per Second



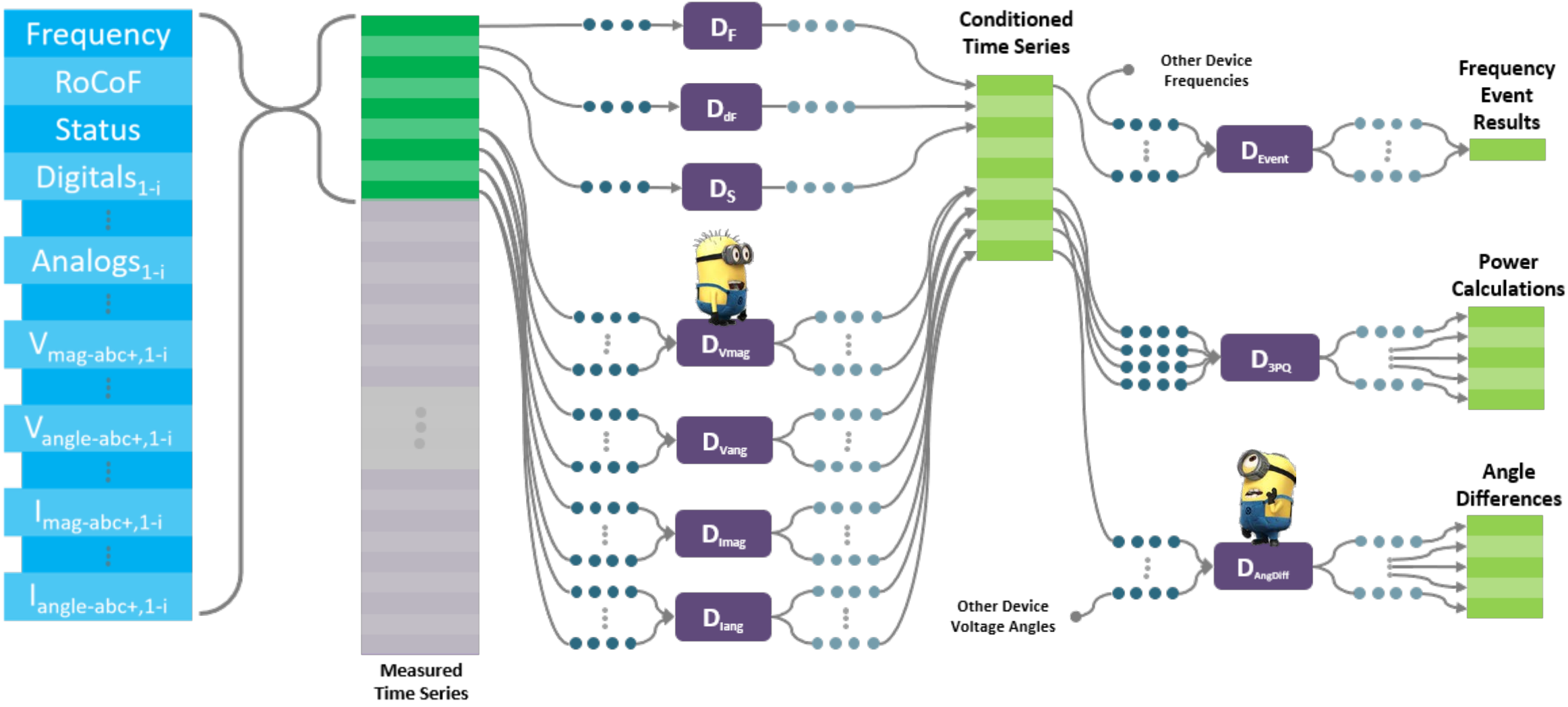
39,400 points * 30 Hz = **1.182M pps**

Annual Data Volumes

1.182M pps * 64 bits per point = 72.14Mbps (271 TB per year)

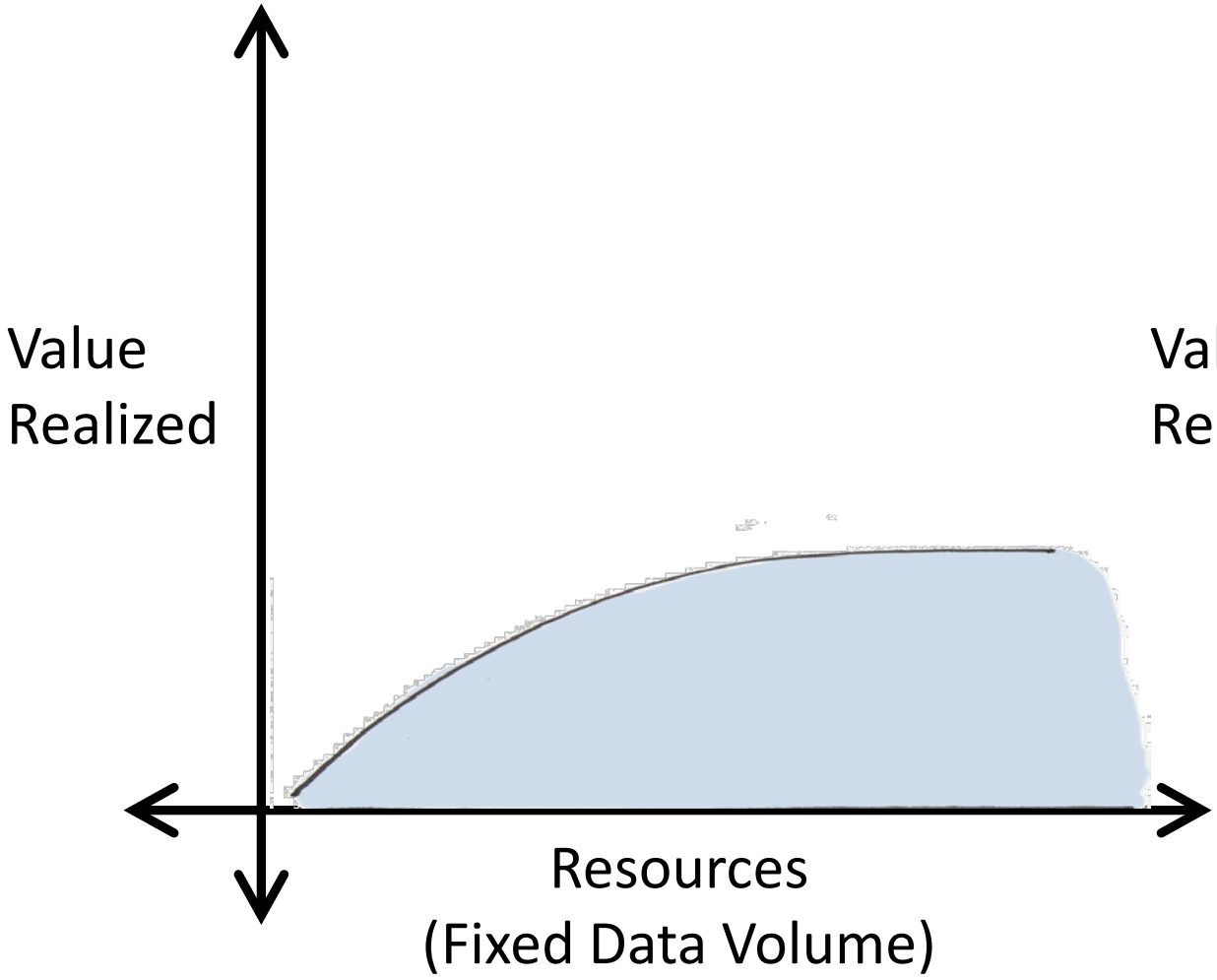
Exclusions: no redundancy, no compression, no ramp up, no calculated values

Calculated Values Multiply → Storage Requirements, Analytic Sample Space

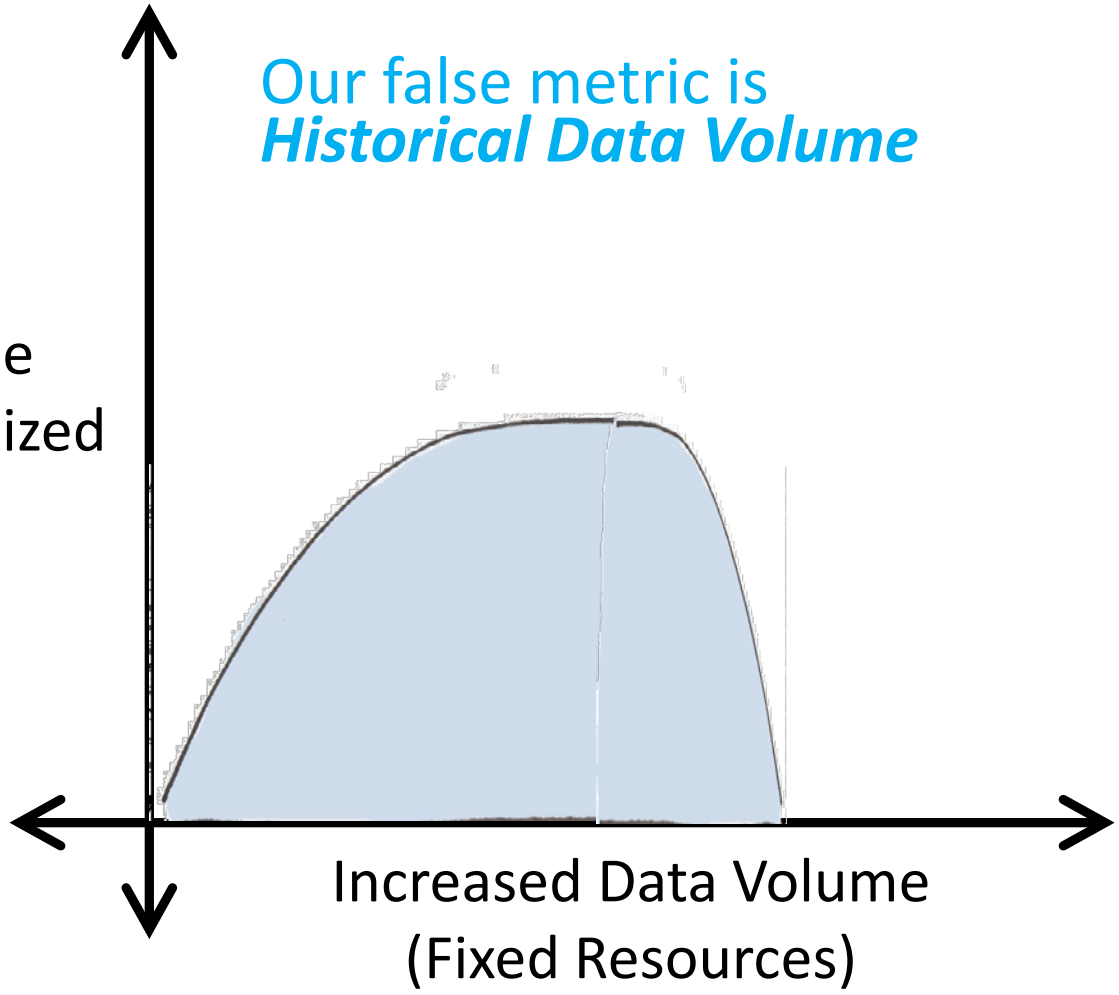


Scaling Diskspace with Single Machines

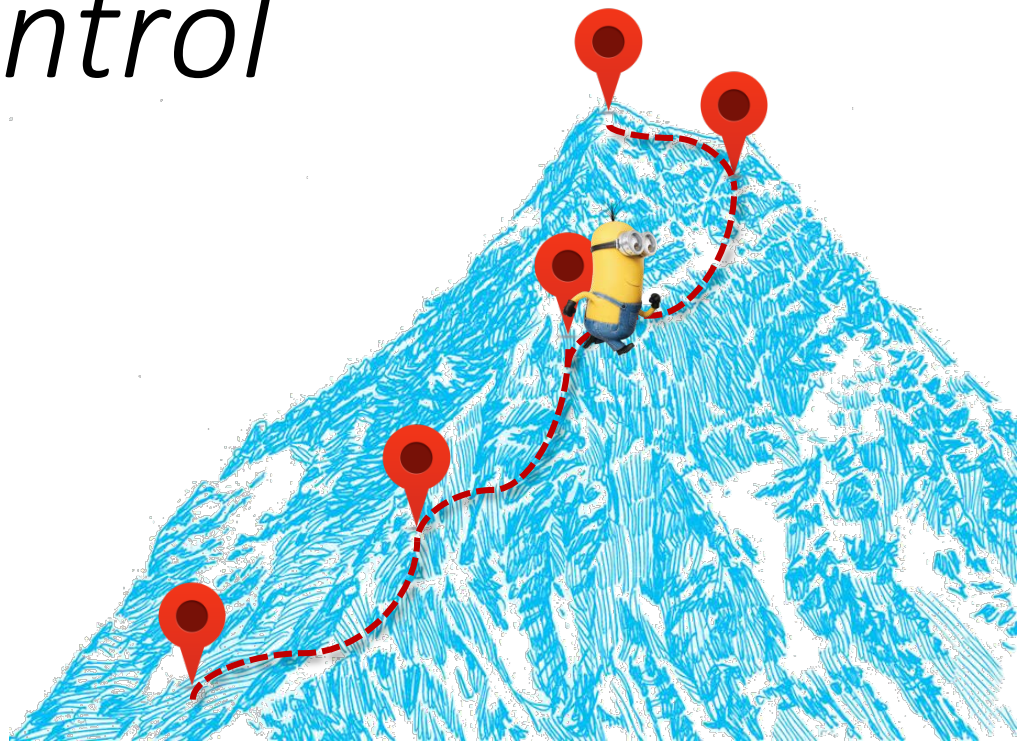
Scale out disk space until data becomes unsearchable and computation against the set becomes infeasible



Our false metric is
Historical Data Volume



Camp 2:
Infrastructure & Control



Cloud-First Approach to Infrastructure

- Dominion ET believes that contemporary analytic work is not sustainable with an on-premises solution.

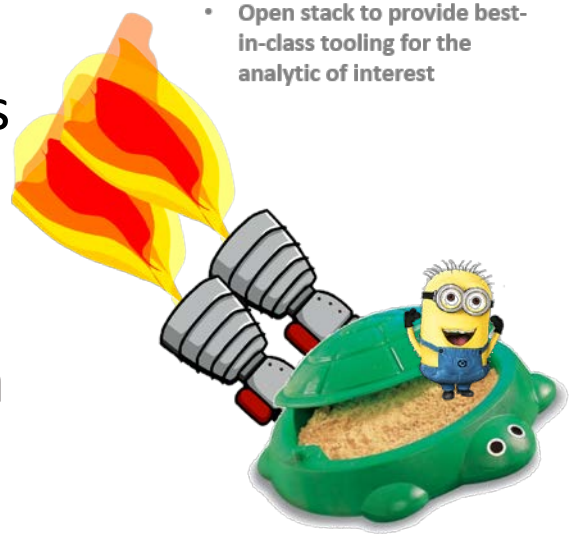
- Why Cloud-First Approach?
  - For Control
  - For Performance
  - For Scalability
  - For Flexibility
  - For Data Sharing

- In January 2019, we will begin streaming our synchrophasor data to PingThings' PredictiveGrid Platform hosted in AWS.

Minimize Control with Sandbox Environment

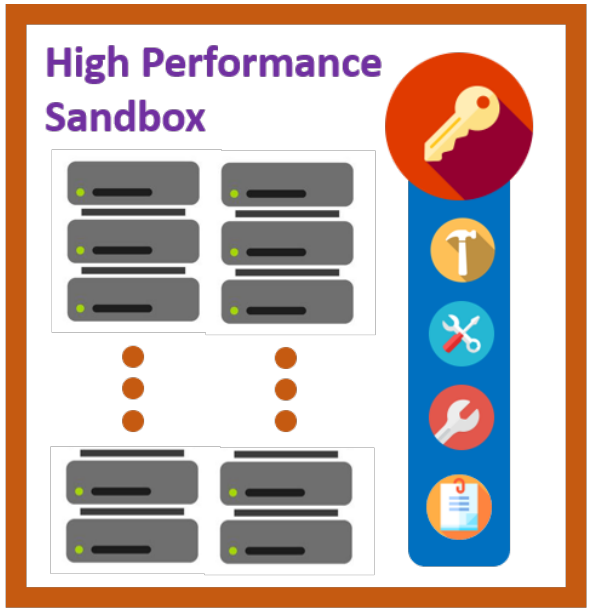
High Performance Sandbox Model

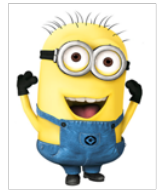
- Business owned and controlled
- Emphasizes computing requirements of value creating activities for super-users
- "Super-users" takes on new meaning
- Important for synchrophasors because *the whole stack matters*!!!
- A true manifestation of the analytics pipeline is actually a high performance sandbox.



Business Owned/ Cloud Hosted

- Infrastructure exists in the cloud to provide best/fastest flexibility for growth
- On-demand resources for new-innovative efforts
- Open stack to provide best-in-class tooling for the analytic of interest





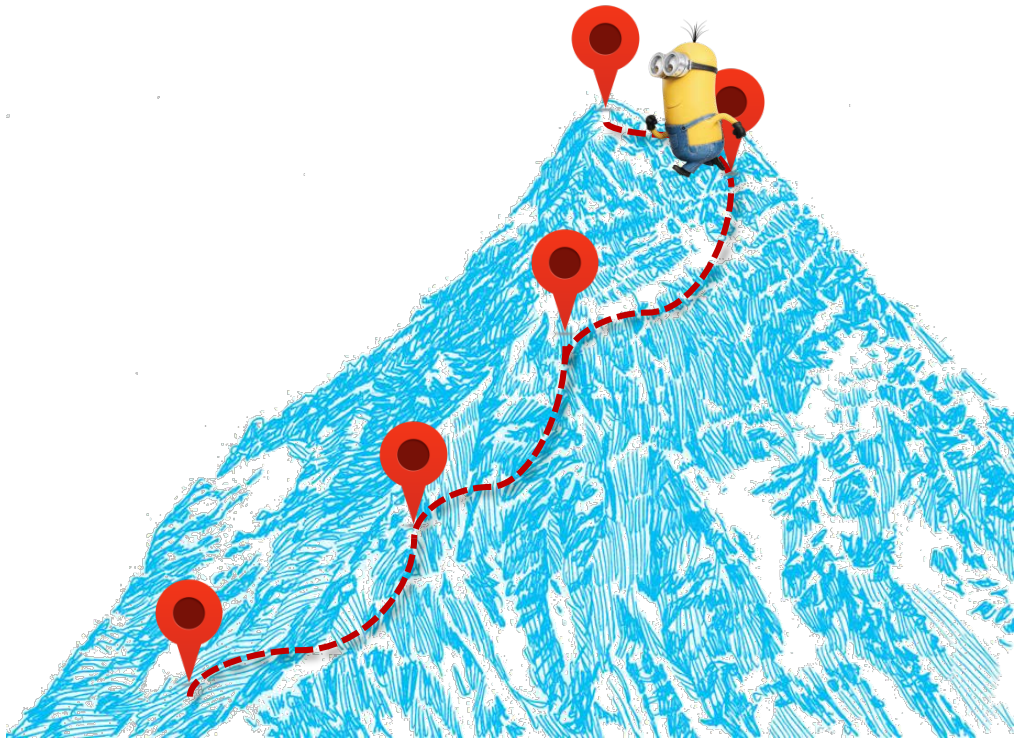
super-users
[analytic developers, data scientists]

- Growing numbers!
- Complex use cases
- LARGEST VALUE CREATION OPPORTUNITY

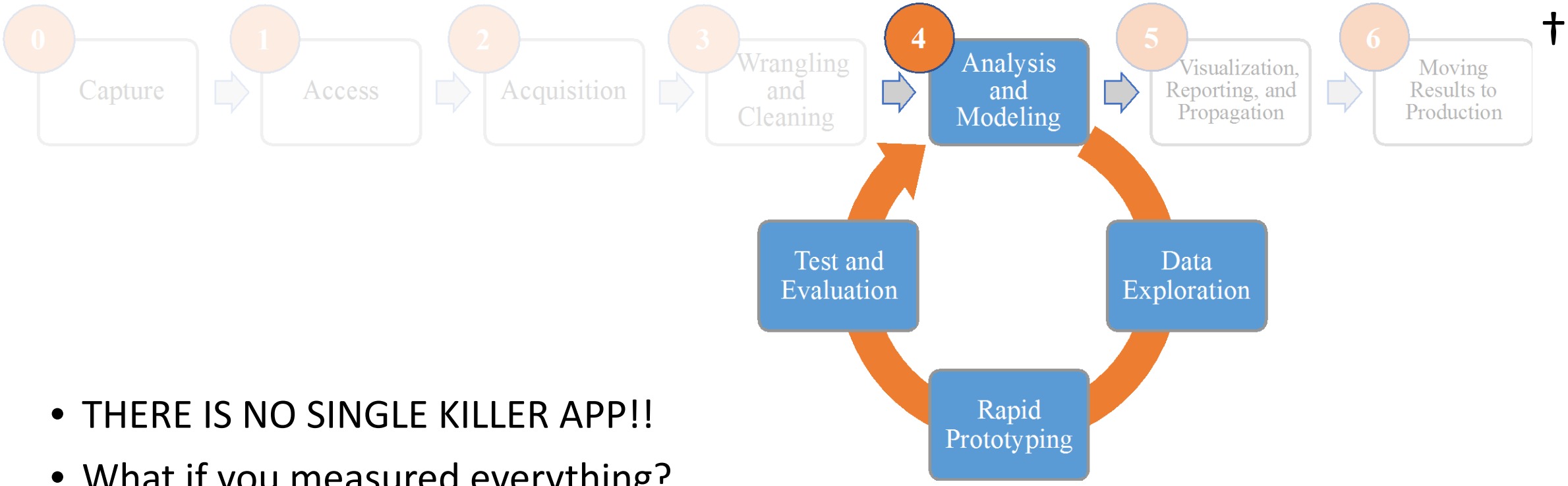
general purpose users

- Many in number
- Simple use cases
- Consume results of super-users
- Lower value creation opportunity

Camp 3 (Death Zone):
Analytic Experimentation & Evaluation



Clamoring for Use Cases



†

- THERE IS NO SINGLE KILLER APP!!
- What if you measured everything?
- What if you had virtually unlimited computational resources?
- So many use cases in the existing literature – so much **latent value**
- Exhaustively explore this space and lay foundation for ROI

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

Analytic Road-mapping: An Exhaustive Lit. Review

1. Identify existing use cases from the (Xplore) literature in an exhaustive fashion
2. Develop a strategic work plan that prioritizes the most promising analytic work relative to the population
3. Translate the desired analytic methodologies into a high level software architecture for ultimately building up a toolbox library

9,818 Total Records

Manual Filtering and Categorization

7,297 Actual Related Records

Manual Filtering

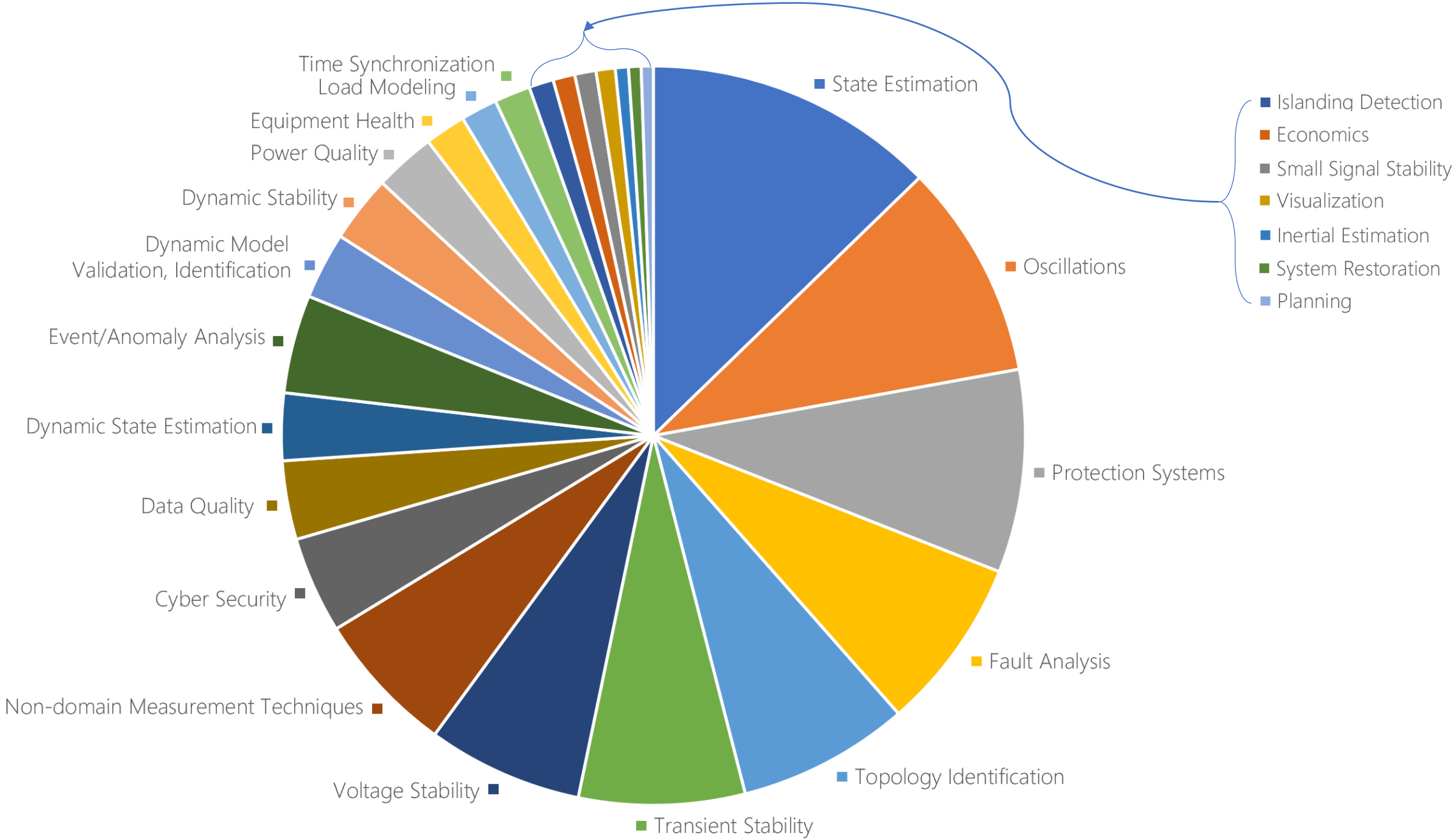
Category	Description	Count
Not Related	Not related to power systems or synchrophasors	2,521
Control	Power System Controls/Controllers Design, Active Islanding, Load Shedding	793
Overview	Overviews, Reviews, Surveys, Editorials	771
PMU Placement	Optimal sensor deployment	462
Communications	Protocols, PDCs, Networking, Security	346
Phasor Estimation	Estimation of phasors and frequencies from waveform data	284
Simulation	Simulators, Simulations, Co-simulation, Testbeds	166
PMU	Design/Build/Test of Measurement Devices	132
Real Time Operations	Control Room Only Applications	56
Standards	Standards Documents	33
	Total:	**5,564**

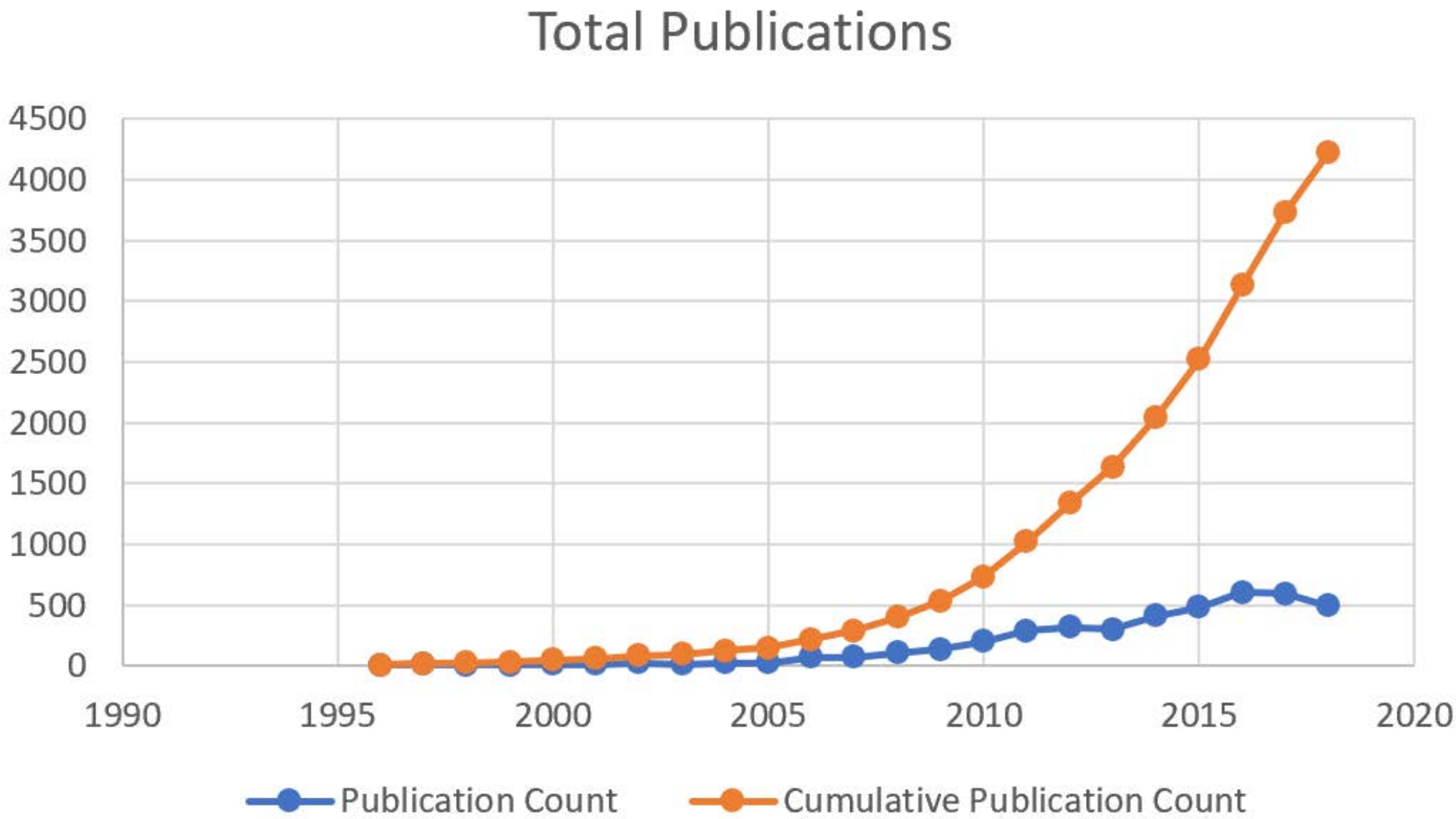
Valid Domain Categories (1 of 2)

Category	Description	Count
State Estimation	State Estimation, Bad Data Detection, False Data Injection	535
Oscillations	Detection, Localization, Identification, Mode Estimation, Mode Meter	397
Protection Systems	Relaying, RAS, Backup Protection, Algorithms	374
Fault Analysis	Detection, Localization, Classification, Identification	318
Topology	Network Impedances, Thevenin Impedances/Equivalents, Temperatures, Ratings, Calibration Metrics, Topology Estimation, Branch Status, Abstract Network Metrics, Matrix Inference	315
Transient Stability	Prediction, Metrics	304
Voltage Stability	Prediction, Metrics	285
Non-domain Measurement Techniques	A broad range of measurement-based techniques that do not clearly map to a specific, identified domain	263
Cyber Security	Detection, Prevention of Attaches, Data Manipulation	178
Data Quality	Metrics, Data Conditioning, False Data Injections	145

Valid Domain Categories (2 of 2)

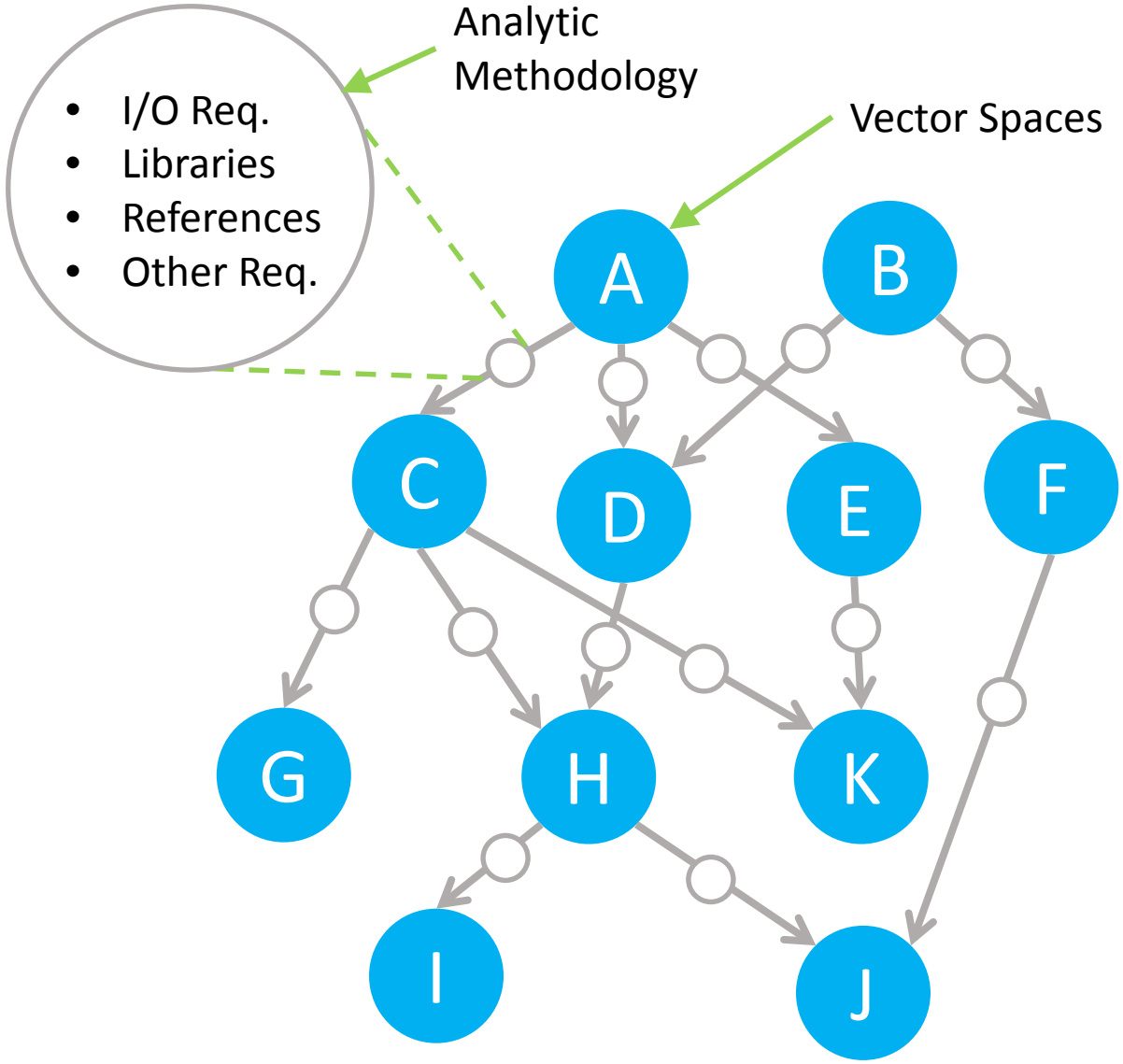
Category	Description	Count
Dynamic State Estimation	DSE, Distributed DSE	124
Event/Anomaly Analysis	Detection, Classification, Identification, General Analysis, Post-mortem Analysis	123
Dynamic Model Validation, Identification	Dynamic Models, Generators, Plants, FACTS Validation and Parameter Identification	123
General Stability	Prediction, Metrics	123
Power Quality		112
Equipment Health	Asset Performance and Lifespan Metrics	75
Load Modeling	Steady State and Dynamic Load Modeling	67
Islanding Detection		45
Economics	Power Systems Economics	40
Small Signal Stability	Prediction, Metrics	39
Visualization		35
Inertial Estimation	Estimating the System/Machine Inertia	24
System Restoration		23
Planning	Long Term Planning and Siting	22


State Estimation
Oscillations
Protection Systems
Fault Analysis
Topology Identification
Transient Stability
Voltage Stability
Non-domain Measurement Techniques
Cyber Security
Data Quality
Dynamic State Estimation
Event/Anomaly Analysis
Dynamic Model Validation, Identification
Dynamic Stability
Power Quality
Equipment Health
Load Modeling
Time Synchronization
Islanding Detection
Economics
Small Signal Stability
Visualization
Inertial Estimation
System Restoration
Planning


Total Publications
4500
4000
3500
3000
2500
2000
1500
1000
500
0
1990
1995
2000
2005
2010
2015
2020
Publication Count
Cumulative Publication Count

Next Steps for Analytic Road-mapping

- Depth Investigation of Domains
  - Identification of Subdomains
  - Extracting Requirements
    - I/O
    - Models
    - Metadata
    - Computation
    - Library Dependencies
  - Making value judgements
  - Determining feasibility and priority
- Developing Workplans
- Synthesizing High Level Software Architectures

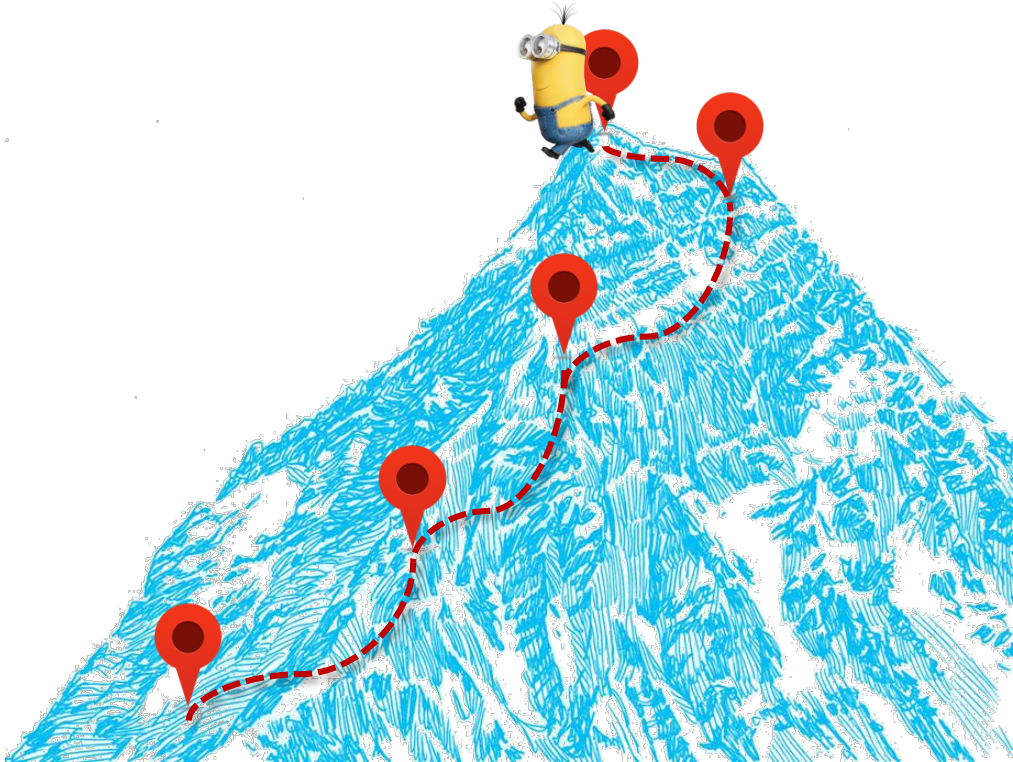


So, What is the False Metric?

- *ROI and Use Cases is the False Metric*
- This is true because:
  - We try to predict which use cases to invest in
  - The cost of use case development and potential ROI varies dramatically so it is not the best metric to minimize
- High ROI comes from cost effective experimentation
- Our New Cost Metric
  - Resources required to sample any chosen data point
  - Resources required to create a specific new data point

Summit:

Artificial Intelligence, Machine Learning, & Deep Learning



What is Waiting at the Summit?

- Sustainable exploration, and development of AI, ML, and Deep Learning analytics that has been enabled by:
  - Nearly full system PMU coverage
  - Lots of easily accessible data
  - Bureaucratically unburdened systems
  - Contemporary technologies
  - Fast feedback evaluation loops
  - Vastly increased analytic sample space
  - Low cost experimentation

Clamoring for Use Cases

Aggressive analytic goals for 2019

- Brute force – try everything!!

Jupyter Notebooks

- 50 by end of 2019

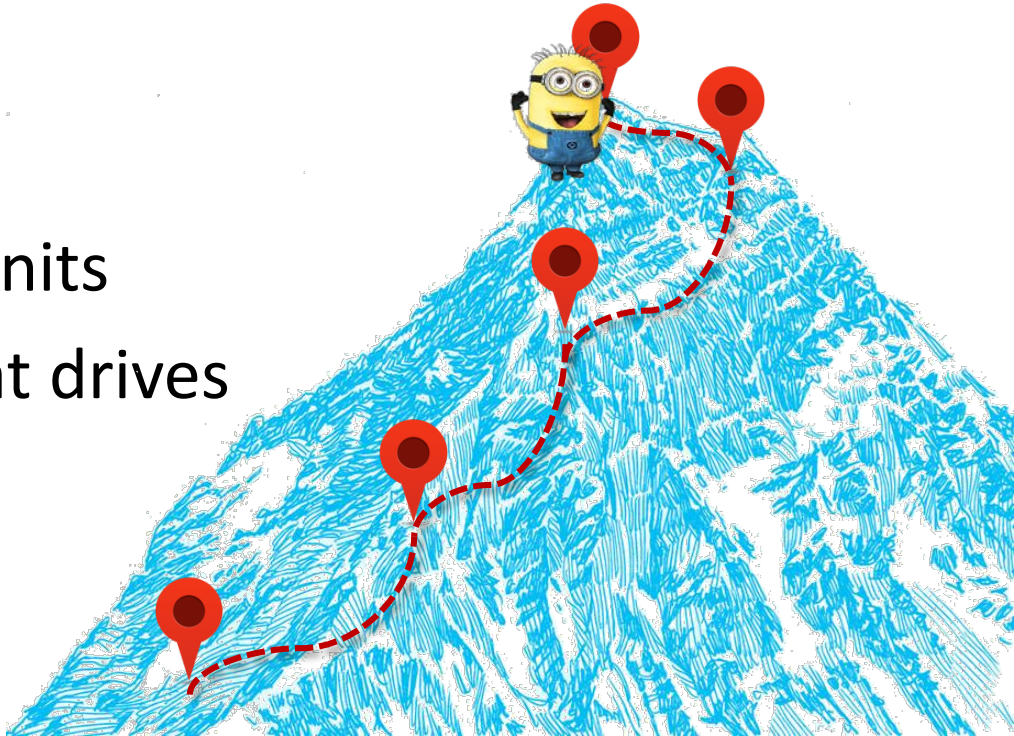
High potential ROIs
moved to “production”

I Want Help

- We are looking for talented graduate and undergraduate students to help us build out our analytic capabilities in 2019.
- If you're interested (and awesome!!),
  - Talk to me here at NASPI
  - Email me here: kevin.d.jones@dominionenergy.com
  - Or text me here: 304-767-4748
- In return:
  - Utility Experience
  - Access to Data
  - Publications!!

Conclusion

- The cost of analytic experimentation governs value extraction – but this is multi-dimensional.
- We can better improve all dimensions with a more molecular measure
  - The cost to access any random data point
  - The cost to create a new data point.
- It is helpful to use *time* as the cost metric units
- Remember, there is a different function that drives cost at each level of your ascent.



Thank You!

Kevin D. Jones, Ph.D.

kevin.d.jones@dominionenergy.com

304-767-4748