# PMU-Based Data Analytics Using Digital Twin and PhasorAnalytics Software (DE-OE0000915)


Philip Hart[1] (Presenter), Weizhong Yan[1], Tianyi Wang[1], Vijay Kumar[1], Pengyuan Wang[1], Gang Zheng[2], Maddipour Farrokhifard[2]

[1] GE Research
[2] GE Digital



Acknowledgment: "This material is based upon work supported by the Department of Energy under Award Number DE-OE0000915."

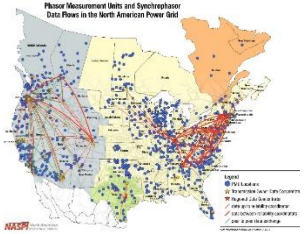




**The 2nd IEEE International Conference on Smart Grid Synchronized Measurements and Analytics (SGSMA)** *Virtual Event | May 24-27, 2021*

# GE Research – Big Data Analysis of Synchrophasor Datasets

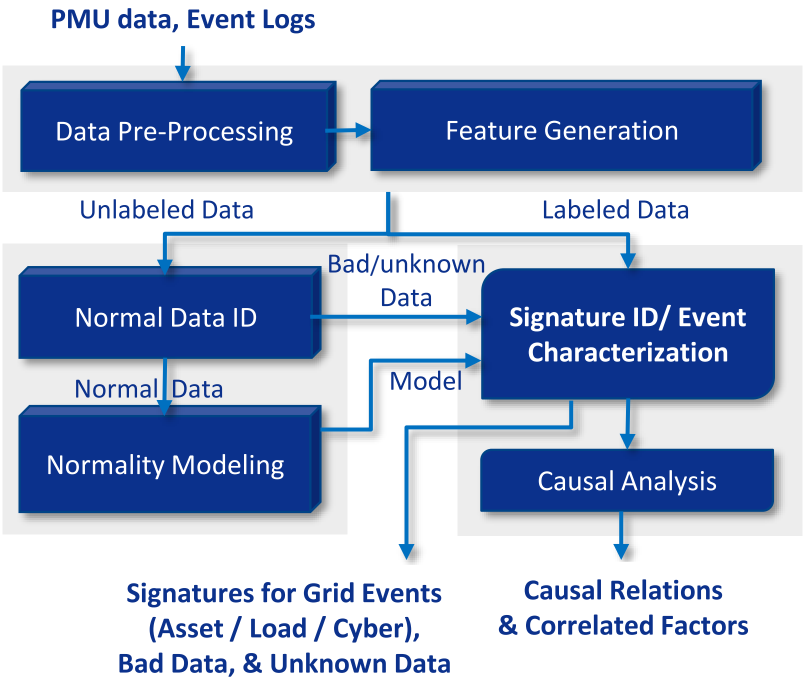
- US Dept. of Energy assembled a phasor measurement unit (PMU) dataset from 443 PMUs across Eastern, Western, and Texas interconnects, along with event logs w/ 1000s of recorded events (FOA-1861, DE-OE0000915).
- **Overall program objectives**:
  – Apply big data, AI & ML technology and capabilities to extract new insights, such as validated grid event signatures (generator trip, line fault, etc.)
  – Develop systems and tools for effective grid operation and management, with overall goal of improving system resiliency and reliability




| Interconnect | # PMUs | # records | Compressed data size (Terabytes) |
|---|---|---|---|
| IC_A | 212 | 160,809,031,796 | 2.9 |
| IC_B | 43 | 93,353,826,102 | 4.7 |
| IC_C | 188 | 241,437,700,843 | 11.0 |
| Total | 443 | 495,600,558,741 | 18.5 TB |

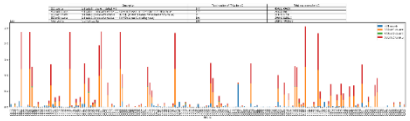
*[Image courtesy of NASPI, https://www.naspi.org/]*

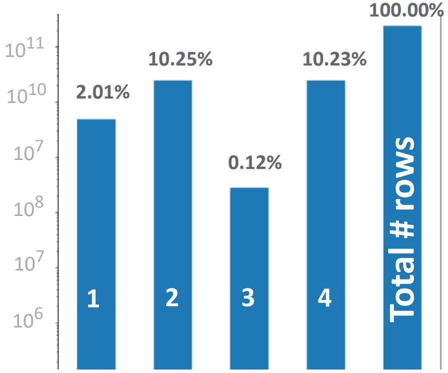
## High-Level Technical Strategy:



# GE Research – Data Pre-Processing and Feature Generation

## Data Quality (ICC):






1: # of rows with ‘status’ !=0
2: # of rows with >0 ‘unreasonable’ values
3: # of rows with >0 non-numerical values
4: # of rows with >0 missing values

## Feature Generation using Customized Big Data Platform

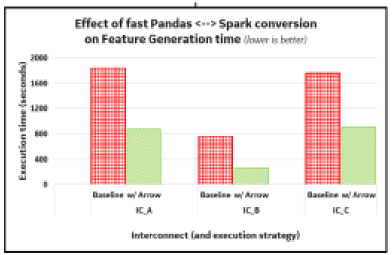
### *Raw Data:*

| Feature name | Raw signal channel | Description |
|---|---|---|
| f_diff_dn | f *(frequency)* | Maximum step down in 0.1 second |
| f_filter_p2p | f *(frequency)* | Peak-to-peak value after filtering out 1st principal component among all PMUs; used to characterize asynchronization with peers. |
| vm_diff_dn | vp_m *(voltage magnitude)* | Maximum step down in 0.1 second |
| vm_diff_up | vp_m *(voltage magnitude)* | Maximum step up in 0.1 second |
| vm_p2p | vp_m *(voltage magnitude)* | Peak-to-peak value |
| im_std | ip_m *(current magnitude)* | Standard deviation |
| im_diff_dn | ip_m *(current magnitude)* | Maximum step down in 0.1 second |
| im_RP | ip_m *(current magnitude)* | Exhibition of strong frequency components in the signal; used to characterize oscillations. |
| p_diff_up | p *(active power)* | Maximum step down in 0.1 second |

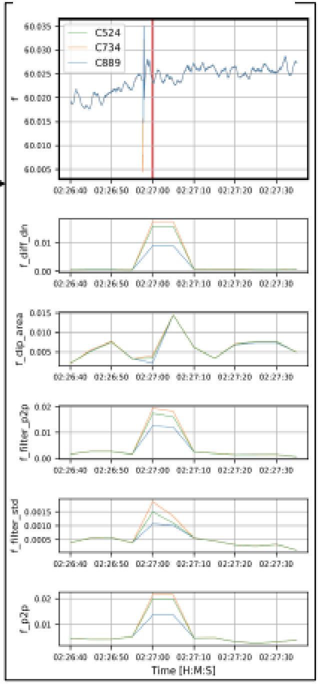
***Sample features***

- *Overall: 60+ feature functions to be calculated per every 5 seconds of raw data*
- *Across all PMUs in an interconnect; grouped into 7+ feature batches*

### *Performance Optimizations*




### *Feature Time-Series*




## Impact of Big Data Platform Performance Optimizations:

Resulting **productivity gains** offer power and grid systems researchers significant advantages

- e.g., 89 million feature values per PMU in IC_B (23.5 GB) in ~ 50 minutes
- flexible feature store (add, update, delete, query feature batches)

# GE Research – Normality Modeling and Signature Identification Example

**GENERATOR EVENT**

| Feature Name | Feature Score |
|---|---|
| **f_dip_area** | **41.39** |
| **f_p2p** | **38.76** |
| **vm_NSR** | **11.98** |
| **f_bump_area** | **10.53** |
| **vm_diff_dn** | **9.45** |
| **vm_step_ind_3** | **7.13** |
| **f_diff_dn** | **5.70** |
| **vm_diff_up** | **5.63** |
| vm_mean | 3.61 |
| vm_step_mag_2 | 3.13 |
| vm_max | 2.87 |
| vm_dip_area | 2.60 |
| vm_std | 2.16 |
| vm_step_mag_1 | 1.04 |
| vm_p2p | 0.95 |
| vm_step_mag_3 | 0.95 |
| vm_step_ind_2 | 0.32 |
| f_diff_up | 0.27 |
| vm_bump_area | 0.27 |
| vm_min | 0.20 |
| vm_step_ind_1 | 0.00 |

**TRANSFORMER EVENT**

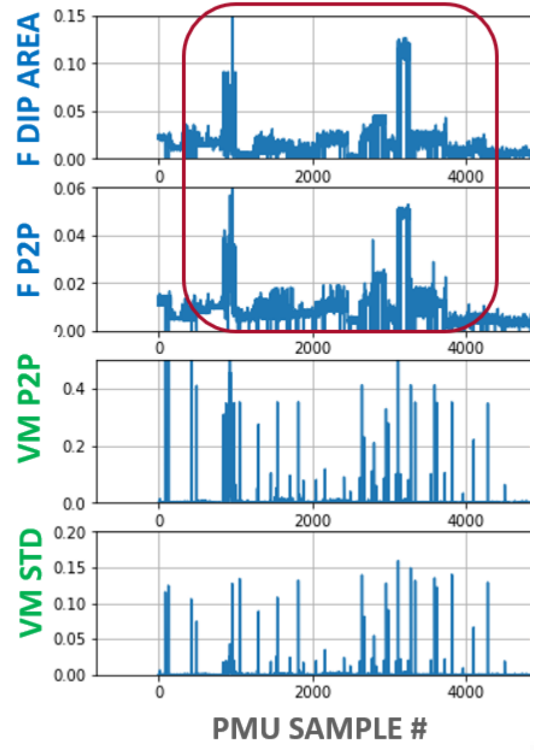
| Feature Name | Feature Score |
|---|---|
| **vm_p2p** | **772.79** |
| **vm_std** | **619.73** |
| **vm_min** | **260.72** |
| **vm_step_mag_3** | **201.98** |
| **f_diff_up** | **135.74** |
| **vm_diff_dn** | **76.62** |
| **vm_dip_area** | **71.16** |
| **vm_diff_up** | **69.56** |
| vm_step_mag_1 | 39.70 |
| vm_step_ind_3 | 21.61 |
| vm_step_mag_2 | 20.95 |
| f_diff_dn | 8.17 |
| f_dip_area | 6.45 |
| f_bump_area | 5.38 |
| f_p2p | 2.50 |
| vm_mean | 1.86 |
| vm_NSR | 1.59 |
| vm_step_ind_2 | 0.36 |
| vm_max | 0.13 |
| vm_bump_area | 0.05 |
| vm_step_ind_1 | 0.00 |

1. '**Overabundance of features**' is first generated
2. Then, a **trained normality model** is used to rigorously score features based on relevance to event

**Event Signatures Identified:**

- Generator event: ***f_dip_area, f_p2p, …***
- Transformer event: ***Vm_p2p, vm_std, …***

# GE Research –Event Signatures, Validation, and Conclusions




## Example Event Signature Results (Interconnect C):

| | generator | transformer | line fault | line equipment | line lightning | line not lightning | oscillation |
|---|---|---|---|---|---|---|---|
| f_diff_dn | 0.184 | 0.006 | 0.026 | 0.001 | 0.023 | 0.01 | 0.004 |
| f_filter_p2p | 0.115 | 0.007 | 0.026 | 0.002 | 0.014 | 0.009 | 0.024 |
| im_std | 0.291 | 0.072 | 0.045 | 0.056 | 0.05 | 0.047 | 0.152 |
| im_RP | 0.175 | 0.084 | 0.017 | 0.052 | 0.019 | 0.018 | 0.047 |
| im_diff_dn | 0.115 | 0.15 | 0.071 | 0.127 | 0.082 | 0.092 | 0.188 |
| im_diff_up | 0.036 | 0.089 | 0.039 | 0.123 | 0.093 | 0.126 | 0.008 |
| vm_diff_dn | 0.091 | 0.306 | 0.474 | 0.415 | 0.39 | 0.357 | 0.299 |
| vm_diff_up | 0.064 | 0.318 | 0.366 | 0.34 | 0.296 | 0.26 | 0.004 |
| vm_p2p | 0.085 | 0.188 | 0.174 | 0.184 | 0.162 | 0.127 | 0.21 |
| vm_step_mag_1* | 0.017 | 0.046 | 0.166 | 0.075 | 0.035 | 0.084 | 0.158 |
| vm_step_mag_2 | 0.044 | 0.128 | 0.147 | 0.056 | 0.048 | 0.106 | 0.142 |
| p_diff_dn | 0.081 | 0.125 | 0.043 | 0.065 | 0.073 | 0.174 | 0.412 |
| p_diff_up | 0.078 | 0.201 | 0.032 | 0.059 | 0.063 | 0.115 | 0.021 |
| q_diff_dn | 0.015 | 0 | 0.089 | 0.044 | 0.094 | 0.152 | 0.045 |

- Grid event signatures identified for >15 types of grid event across Eastern and Western Interconnects. Signatures give insight into ***event type, location, magnitude, & duration.***
- Signatures are 'transparent' & comprised of features suitable for real-time computation
- Developed binary event classifiers + decision fusion for a reduced event subset, applied it to dataset to discover and classify tens of thousands of unlabeled events (>90% accuracy)
  - Example—Eastern Interconnect two-year test dataset: 174 events 'generator' events & 14,668 'line' events discovered and characterized