Big Data Access, Analytics and Sense-Making

Zhenyu (Henry) Huang, Ph.D., P.E., F.IEEE
Laboratory Fellow/Technical Group Manager
Pacific Northwest National Laboratory
July 18, 2017
Deployment of a vast new phasor network is generating unprecedented real-time data

Today – SCADA data

Emerging – phasor data

<table>
<thead>
<tr>
<th>Variety</th>
<th>Voltage + Current</th>
<th>+ Phase Angle, ...</th>
<th>More Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Velocity</td>
<td>1 Sample / 4 Seconds</td>
<td>30-120 Samples / Second</td>
<td>~200x Faster</td>
</tr>
<tr>
<td>Volume</td>
<td>8 Terabytes / Year</td>
<td>1.5 Petabytes / Year</td>
<td>~200x More Data</td>
</tr>
<tr>
<td>Veracity</td>
<td>Unseen ms-Oscillations</td>
<td>Oscillations Seen at 10ms</td>
<td>Greater Accuracy</td>
</tr>
</tbody>
</table>
Smart devices and 2-way communication offer new opportunities, greater complexity

<table>
<thead>
<tr>
<th>Number of homes</th>
<th>100</th>
<th>1k+</th>
<th>10k+</th>
<th>100k+</th>
<th>500k+</th>
<th>1 Million</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compressed data size</td>
<td>2.5 GB</td>
<td>38.5 GB</td>
<td>366.3 GB</td>
<td>2.9 TB</td>
<td>13.6 TB</td>
<td>27.3 TB</td>
</tr>
</tbody>
</table>
More diverse data add to the complexity

► Weather/climate data
  (e.g. PNNL ARM* Data)
  ▪ 300 instruments, 2000 data streams 24/7
  ▪ 500 GB/day rising to multiple TBs/day
  ▪ Curating 20 years’ data

► Market/business data

► Cyber/communication data

► Simulated data
  ▪ Each contingency scenario generates 0.5M bytes data, adding up to TB scale

<table>
<thead>
<tr>
<th>Contingency Analysis</th>
<th>Number of scenarios</th>
<th>Serial computing on 1 processor</th>
<th>Parallel computing on 512 processors</th>
<th>Parallel computing on 10,000 processors</th>
</tr>
</thead>
<tbody>
<tr>
<td>WECC N-1 (full)</td>
<td>20,000</td>
<td>4 hours</td>
<td>~30 seconds 469x speed up</td>
<td></td>
</tr>
<tr>
<td>WECC N-2 (partial)</td>
<td>153,600</td>
<td>26 hours</td>
<td>~3 minutes 492x speed up</td>
<td>~12 seconds 7877x speed up</td>
</tr>
</tbody>
</table>

*ARM: Atmospheric Radiation Measurement
Data volume comparison: Grid vs. big science data
Making data accessible is a big challenge

Organizing and converting data to application specific formats

- SQL Tables
- NoSQL
- CSV
- PI
- PTI
- Text files
- Time series
- XML
- JSON

Redundancy: Underlined steps has to be performed by every application for each type of data

1. Connect to n data sources
2. Get Data * n
3. Combine * n
4. Transform
5. Analyze

Operational or Planning Applications
Organizing and converting data to application specific formats

- SQL Tables
- NoSQL
- CSV
- PI
- PTI
- Text files
- Time series
- XML
- JSON

1. Connect to n data sources
2. Get Data
3. Combine
4. Transform

GOSS

Operational or Planning Applications

1. Connect to GOSS
2. Request Data
3. Analyze
GOSS™: link data to applications

https://github.com/GridOPTICS/GOSS
Multi-layer data-driven reasoning
Computational challenges in keeping up with data cycles

- Current dynamic state estimation codes scale to ~1,000 cores
- Current computational performance meets the real-time requirement for regional systems
- **Challenge**: real-time performance (30 milliseconds) for interconnection-scale systems.
Mathematical challenges in handling non-Gaussian noise in power grid measurement

Noise extracted from PMU

Noise property analysis
Advanced modular visualization for easy exploration of large-scale data
Advanced visualization for improving hydro state awareness (Hydromap)

- Modernize displays for hydro planning and operations
- Develop new, novel visualization techniques and paradigms for analyzing dynamic data
- Develop modular framework for deploying and integrating new data visualizations

Interactive hydro map as a new visualization paradigm

Example modular visualization framework

Current data display in need of modernization
Multi-dimensional wind visualization (Glyphs)

Wind visualization showing multi-dimensional data in glyphs:
- Wind speed (length of tail)
- Wind direction (angle of tail)
- Generation (size of head)
- Uncertainty (color of tail)
- Forecast variability
- Wholesale price
- Capacity
- Last hour generation
- SCE error code
- Generation difference from forecast

Wind Visualization/Wind Forecast Visualization
Historical hydro view using radial visualization

- Compares hydropower generation across different projects along Columbia and Snake rivers.
- Alternative view shows generation across different sources such as hydropower, nuclear, renewables, and miscellaneous sources.
Data repository for data hosting (ARPA-E funded)

Data Repository

- Validated models & scenarios
- User-generated models & scenarios

Web Portal

- Download request
- Generation Processing
- Submission Processing

User

- Use existing datasets
- Generate new datasets
- 3rd-party datasets
- Submit datasets

Data Tools

- Dataset Generation/Anonymization Methods
- Dataset Metrics/Validation

Private Datasets

Anonymized Data

Case Published

New Case Request

Data Generation
Summary

• Grid data complexity is increasing with big volumes, diverse types, and various attributes.
• Such complexity poses significant challenges in data access, transformation, analytics, sense making.
• Math, computing and visualization technologies need to be developed to meet these challenges.
  – GOSS as a big data platform.
  – Multi-layer data reasoning and high performance computing.
  – Modular visualization as interface for information presentation.
Acknowledgement

• PNNL Researchers: *(Data and Computing)* Bora Akyol, Poorva Sharma, Yin Jian, Steve Elbert, Shuangshuang Jin, Bruce Palmer, George Chin; *(Power Engineering)* Ruisheng Diao, Yousu Chen, Mark Rice, Shaobu Wang, Karen Studarus

• Former PNNL Researchers: Terrence Critchlow, Ning Zhou, Ning Lu, Pengwei Du
Questions?

Further Information:
GridOPTICS: http://gridoptics.pnnl.gov/
GridOPTICS™ Software System (GOSS): https://github.com/GridOPTICS/GOSS
Interactive Visualization and Demo Center: http://vis.pnnl.gov/

Zhenyu (Henry) Huang, Ph.D., P.E., F.IEEE
Laboratory Fellow/Technical Group Manager
Pacific Northwest National Laboratory
zhenyu.huang@pnnl.gov