A CIM-Based Framework for Utility Big Data Analytics

Jun Zhu
Power Info LLC

John Baranowski
Andrew Ford
PJM Interconnect LLC

James Shen
Albert Electrical
System Operator

UCA International Users Group

CIM users group

Overview

- Opportunities & Challenges
- Patterns & Methodologies
- Practices & Applications

What is Big Data Analytics?

- Big data analytics is a new generation of technology that can be leveraged to extract business values from large volumes of and a wide variety of data.

- Based on analysis of big data, discoveries can be made to promote efficiency, optimize operation, and save cost, etc.

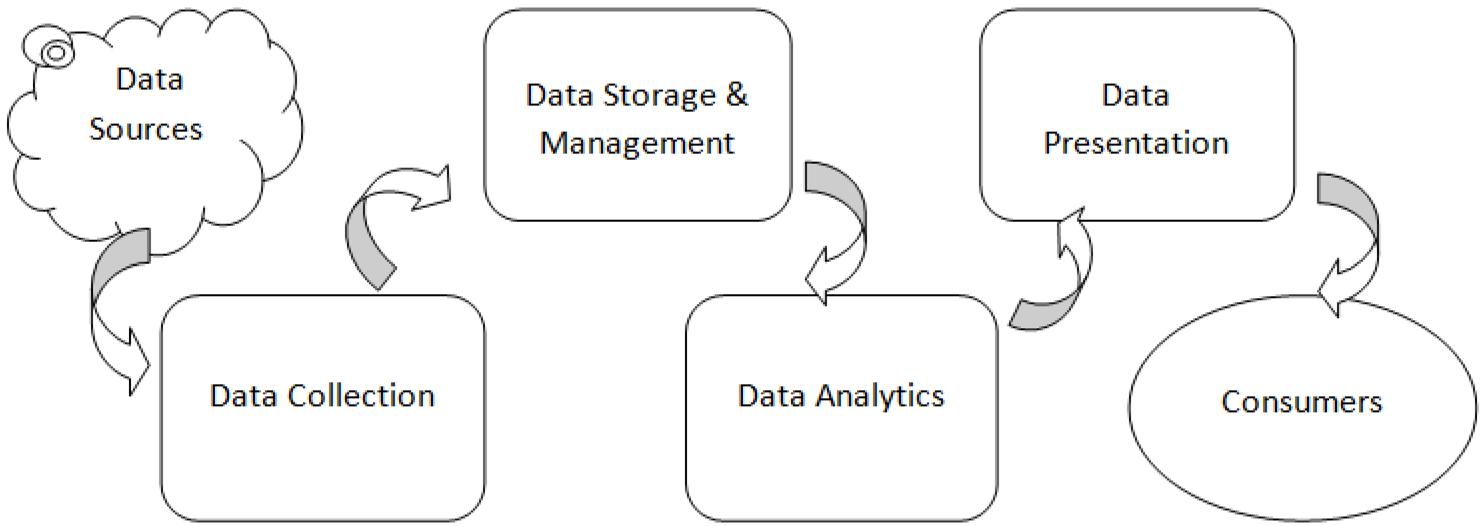
Big Data Analytics for Utilities

- As smart grid deployments created exponentially more data for utilities, the electric utility industry is now seeking new analytics tools and techniques to address their emerging big data issues.

- According to recently conducted surveys:
  - Companies in the utility industries have the highest expectations for generating returns on their big data investments than firms in any other industries. (http://www.tcs.com/big-data-study/Pages/default.aspx

  - It is predicted that global expenditure on utility data analytics will grow from $700 million in 2012 to $3.8 billion in 2020. (http://www.greentechmedia.com/research/report/the-soft-grid-2013)

Gap Analysis



- While tremendous progress has been made in big data collection and management, relying on big data to make decisions is still an industrial effort in its infancy.
- The traditional manual methods of data analysis, such as spreadsheets, ad-hoc queries, and database reporting, have proved insufficient to help analysts capture the patterns and insights from large and complex data sets.

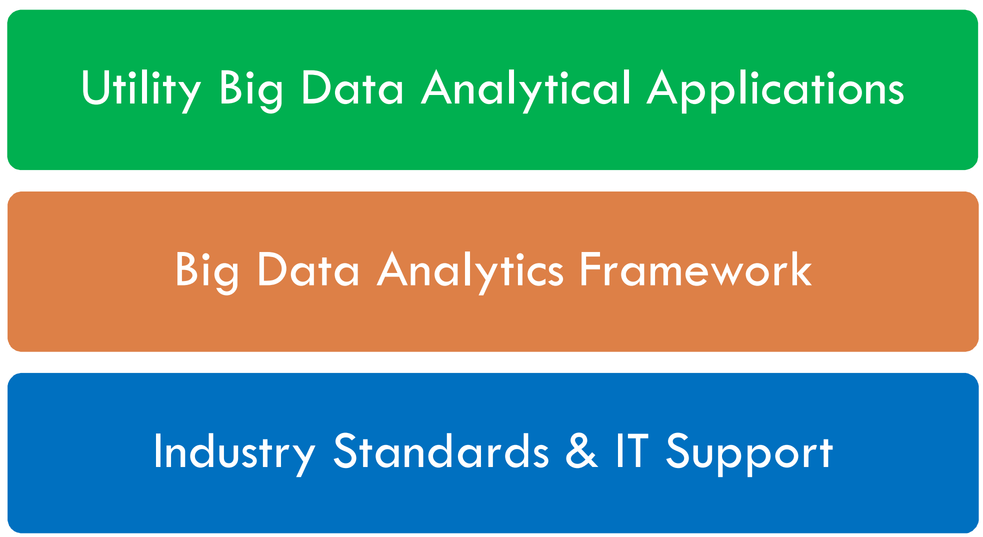
Biggest Challenges

- How to make disparate and incompatible datasets usable, interoperable, and valuable across the enterprise?

- The cost of building and maintaining a specific business-driven big data application is tremendously high due to the common obstacles. Each application must address many of those common issues, if not all.

Objective of the Project

- Developing a software framework to address key utility big data issues and foster development of utility big data analytical applications.



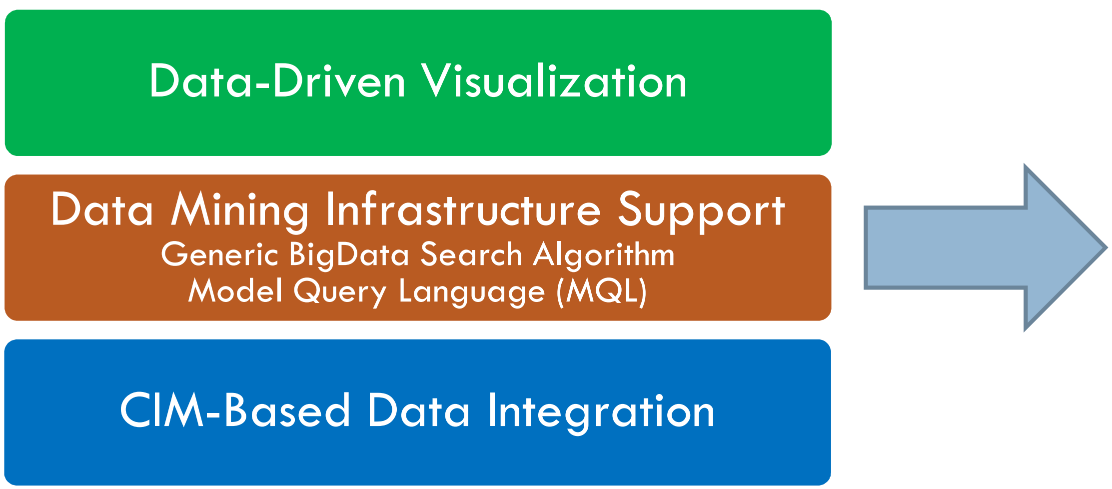
About the Project

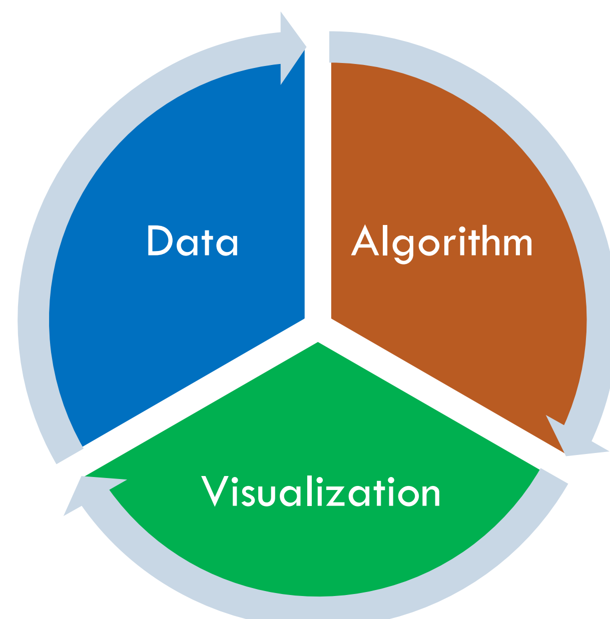
- Funded by US Department of Energy (DOE) under the awarded SBIR Grant (Award No. DE-SC0006347)

- The research results were leveraged by
  - BC Hydro
  - Albert Electrical System Operator (AESO)
  - PJM Interconnect LLC

Layered Architecture of CIM-based Big Data Analytics Framework

- Designed to address the common issues in three BIG areas of big data analytics: Data, Algorithm, and Visualization.
- Providing core data mining infrastructure support from which business-driven utility big data applications can be rapidly developed.





Common Information Model (CIM)

- Common Information Model (CIM) is a standard specification designed for promoting information exchange and fostering collaborations among various utility enterprise applications.

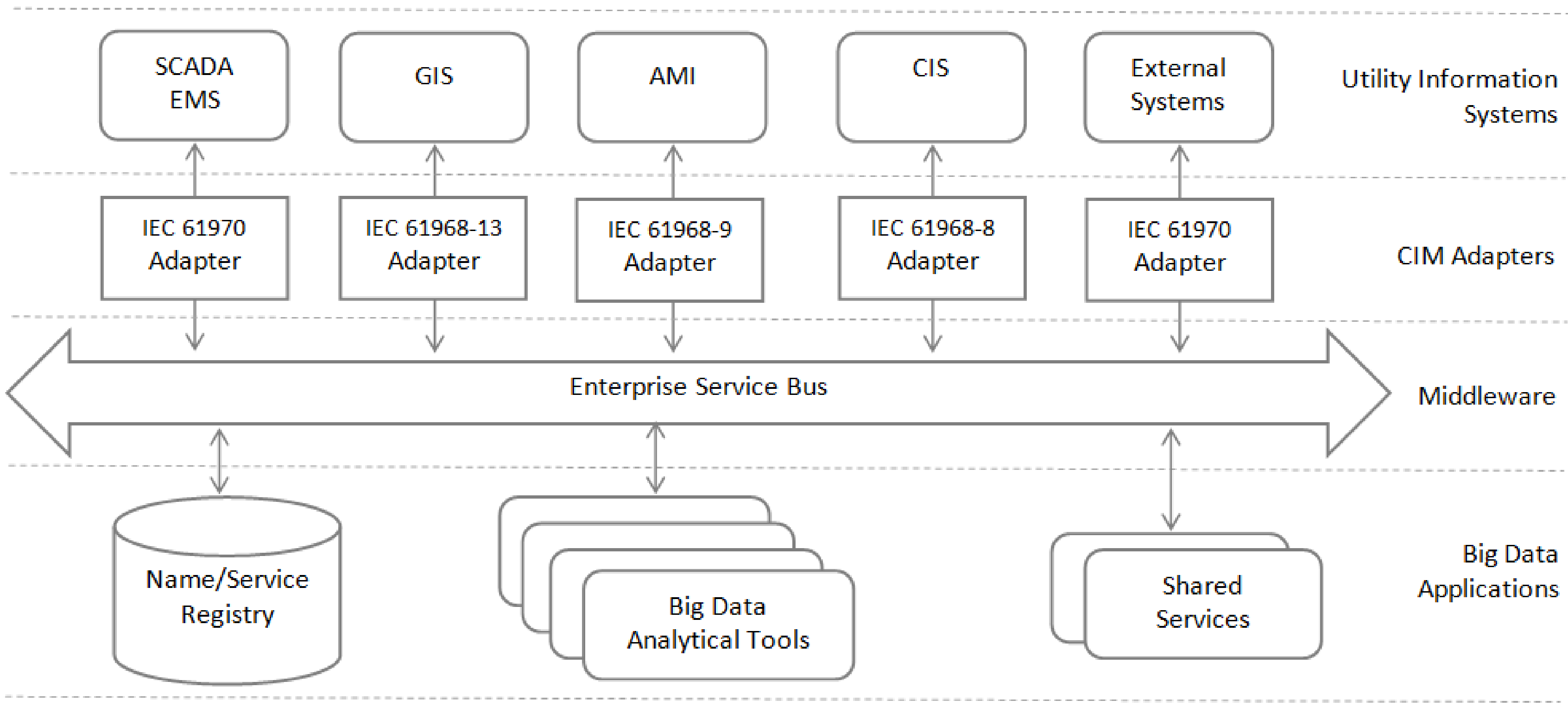
- Because CIM describes all aspects of power system, it provides a standard-based semantic foundation for Big Data Analytics in the utility industry.

CIM for Big Data Integration

CIM is designed/targeted to address various Big Data Integration Issues:

- Naming and Object Identification: Master Resource ID (MRID)
- Heterogeneous Modeling Practices: Operation vs. Planning, Node/Breaker (ConnectivityNode) vs. Bus/Branch (TopologicalNode)
- Data Grouping: Profile
- Data Correlation: Profile Dependency
- Data Composition: Frame
- Data Validation: Object Constraint Language (OCL)
- Data Ownership: Model Authority Set (MAS)
- Foreign Data: Extension
- Variations & Changes: Project

CIM-based Utility Big Data Integration



Data Mining for Business Intelligence (BI)

- Data Mining is the process of identifying new **patterns and insights** in data.

- Insight derived from data mining can provide tremendous economic value and critical decision supporting for cost-saving advantages and operation strategies.

- As the volume of data collected and stored in utility databases grows, there is a growing need to identify important patterns and trends and act upon the findings.

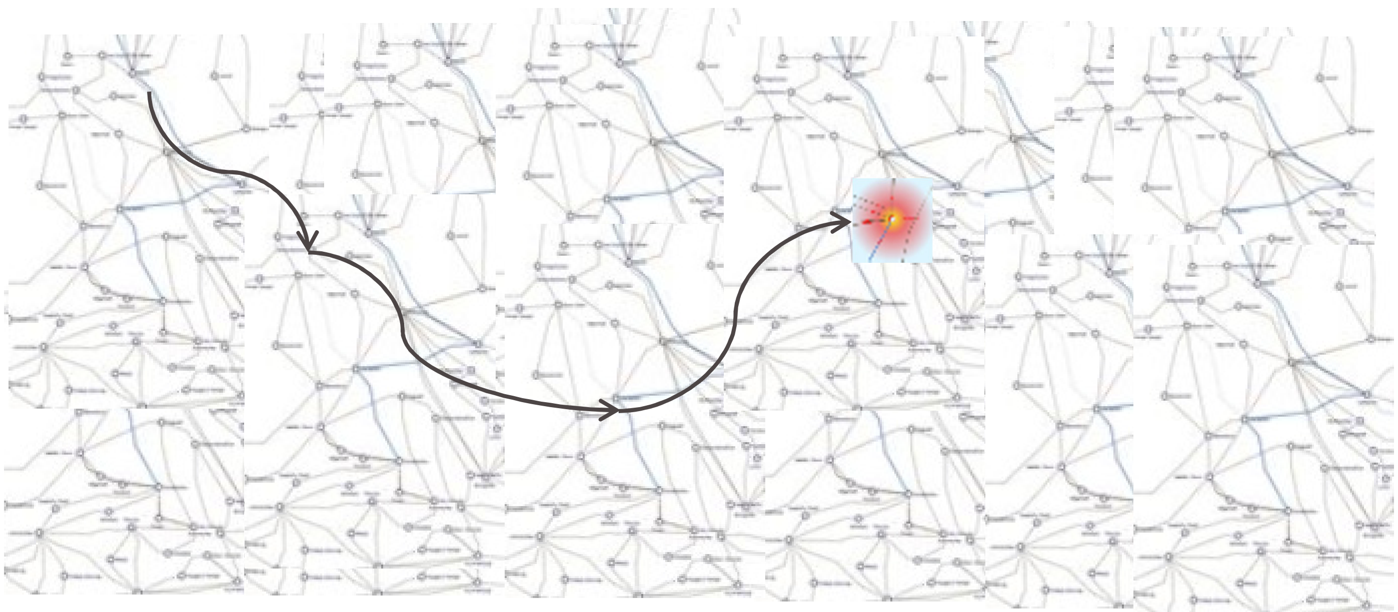
User-Centered Data Mining

- Focusing on providing end-users a capability to build their own analytical applications through **declarative programming** and **plug-in**.

- Providing core data mining infrastructure support from which business-driven analytical applications can be rapidly developed and seamlessly integrated.
  - ✓ Generic BigData Search Algorithm
  - ✓ Model Query Language (MQL)
  - ✓ Report Design

Generic Big Data Search Algorithm

- Designed to support CIM data structures
- Driven by CIM meta-data: CIM Profiles
- Leveraging Google BigTable data management techniques for information cataloging and indexing



Model Query Language (MQL)

Model Query Language is a declarative functional language, specifically designed to query, transform, and manipulate CIM-based data models.

- Most of the today's data query language, such as SQL, OCL, SPARQL, and XPATH, are either inappropriate for CIM models or too complicated for end users.

- MQL targets an object-oriented data model, such as CIM. It uses a simple syntax, similar to algebraic expression, to describe various kinds of data operations, including navigation, filtering, and transformation, etc.

```
ACLineSegment/[BaseVoltage->nominalVoltage]>=220 &&
[EquipmentContainer->Region->name]==“EAST”
```

- MQL was designed to work with Generic BigData Search Algorithm to performs the **user-declared** data operations.

Key Features of MQL

- **Declarative**: Users declare what need to be done rather than programming how to do it.
- **Recursive**: Making a complicated task easy to achieve.
- **Expressive**: Like an arithmetic expression, MQL is easy to construct.
- **Resourceful**: MQL supports
  - bi-directional dataset-to-dataset, object-to-object navigation
  - filtering, grouping, sorting, union, etc.
  - function and operator, built-in or user-defined
  - alias designed to facilitate the reuse of declarations
  - formatting & reporting

Visual Data Mining

- Visualization offers a powerful means of analysis that can help to uncover patterns and trends hidden in unknown data.

- Visual data mining focuses on the combination of visual and non-visual techniques as well as integrating the user in the exploration process.

Common Shortcomings of the Existing Power Grid Visualization Tools

- **Requiring the visual displays to be pre-designed**, thus hindering user's ability to discover. Most importantly, for a large grid, it is very labor-intensive and costly to build and maintain these pre-designed visual displays.

- **Focusing on displaying the data** rather than providing users with the interactive ability to facilitate information exploration.

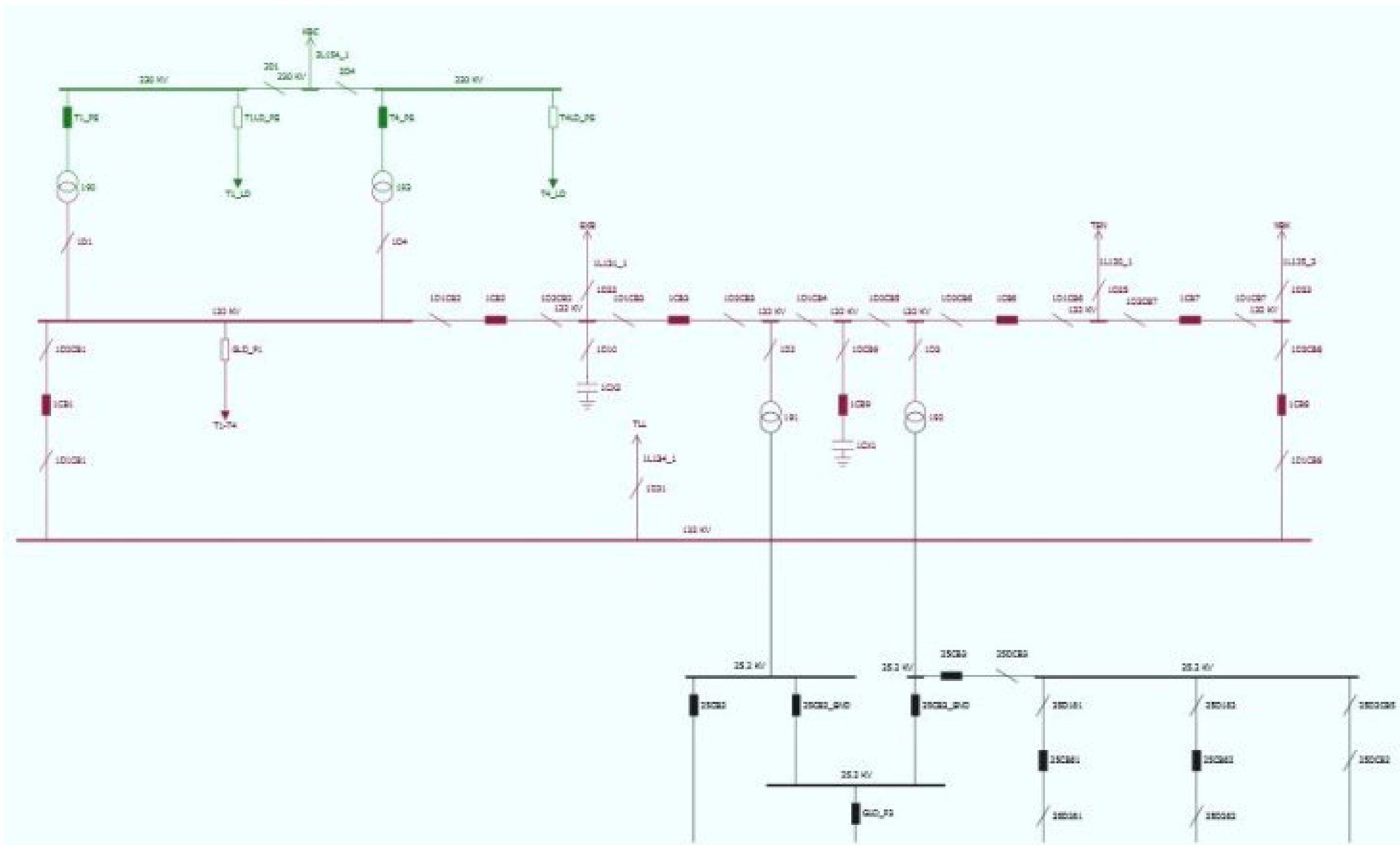
- Designed as a function of a monolithic system and built with proprietary technologies, thus **not interoperable with other utility information infrastructure.**

A Data-Driven Approach

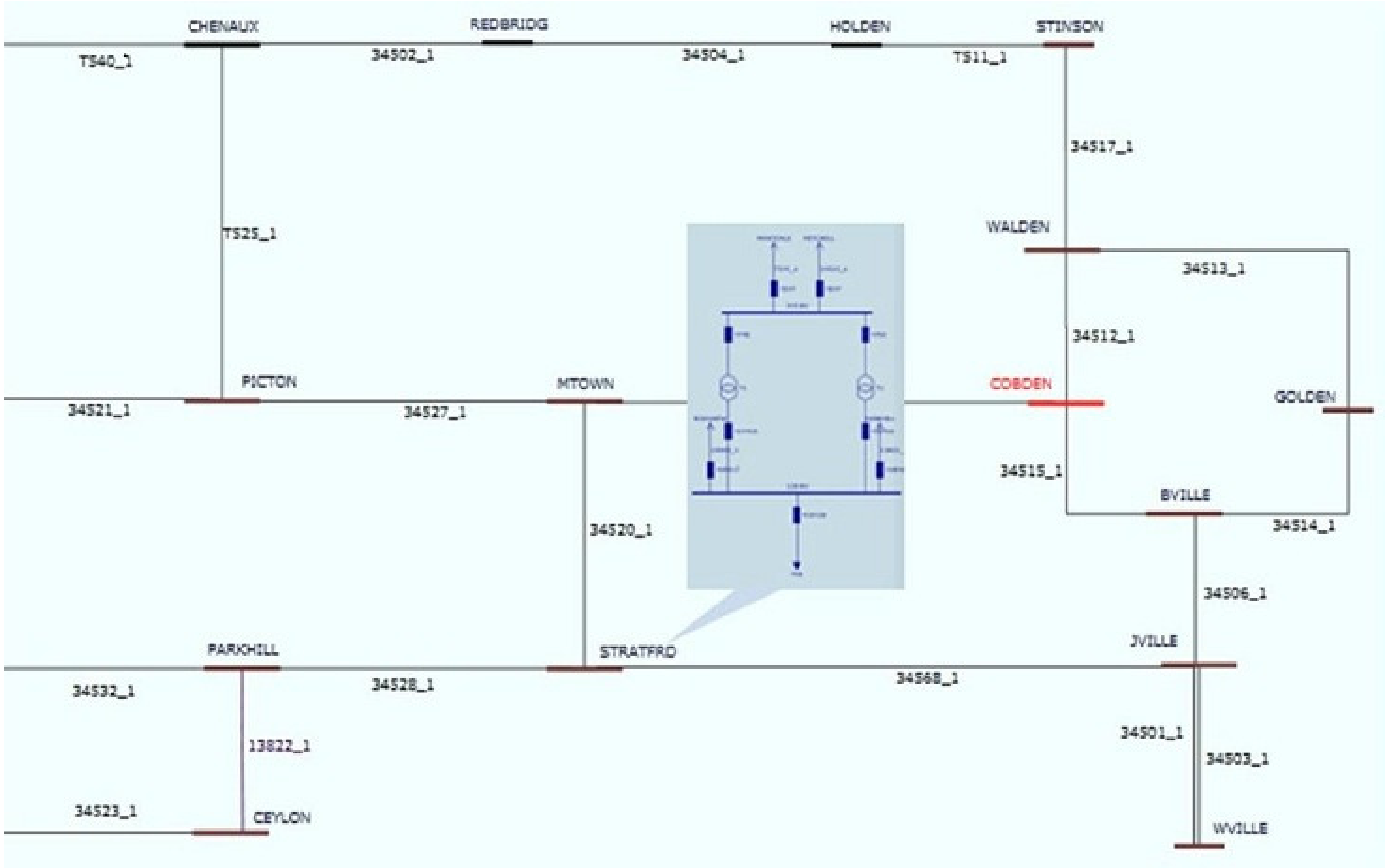
- A data-driven approach relies on developing sophisticated and powerful algorithms for manipulating the data under analysis and transforming it dynamically to create interactive visualizations. **It creates the visualization on-the-fly**, thus completely eliminating or significantly reducing the cost of building visual displays.

- The resulting visual presentations **emphasize what the data is** rather than how the data should be presented in a pre-designed manner, thus, **fostering comprehension and discovery**.

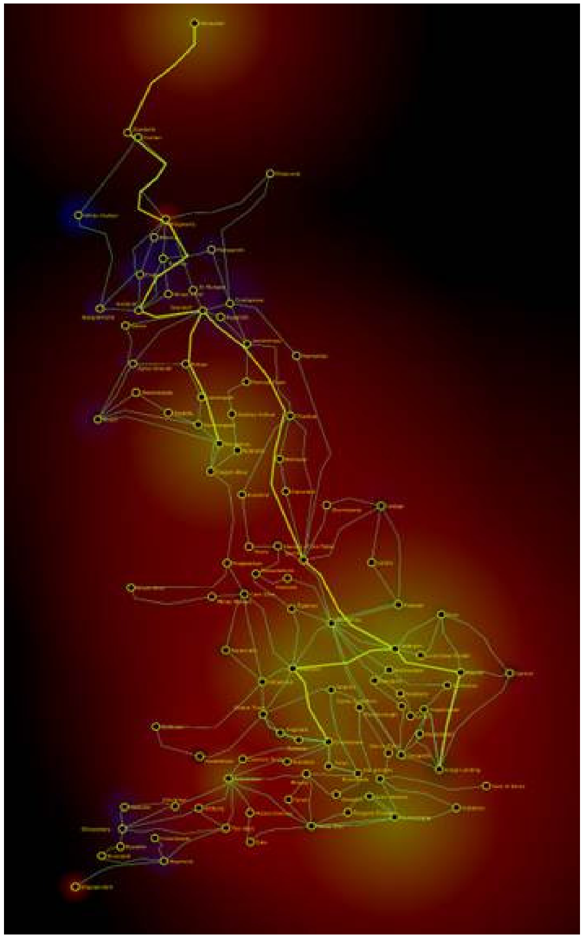
Example 1: Pattern Recognition for Auto Generation of Substation One-Line

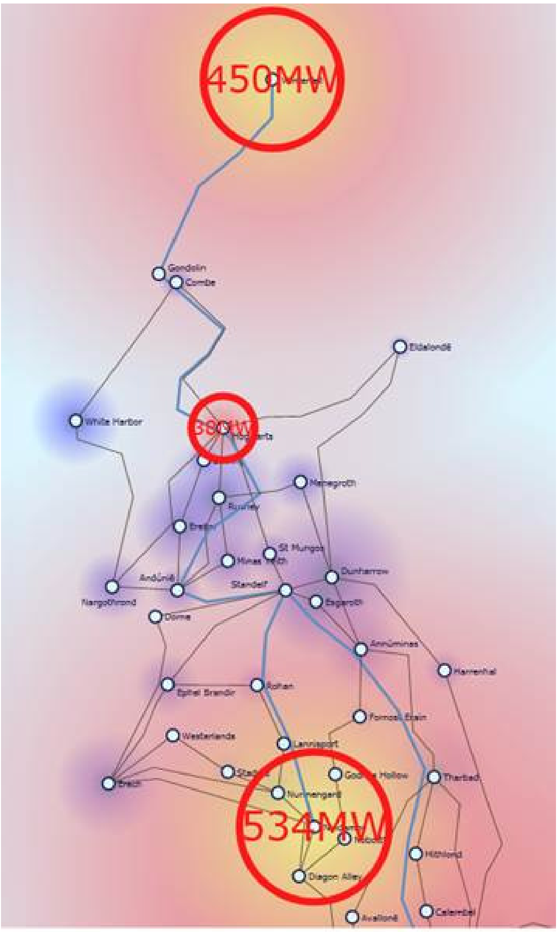


Example 2:Force Directed Layout of Line-Substation Connections

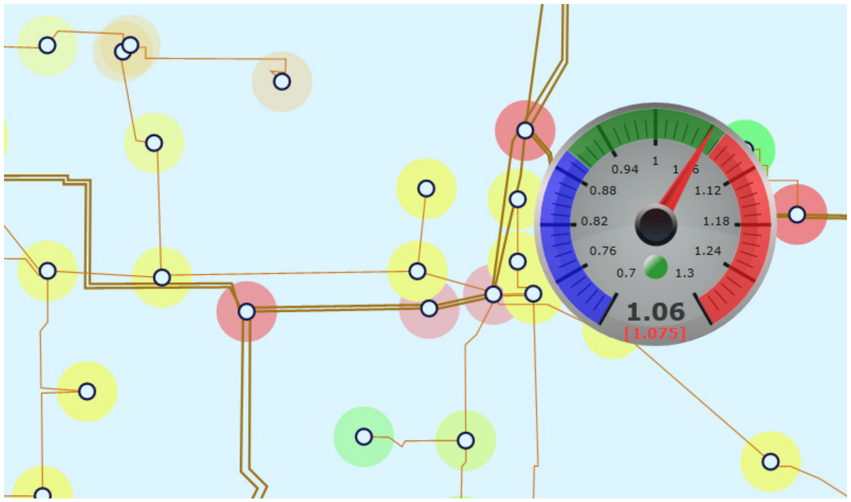


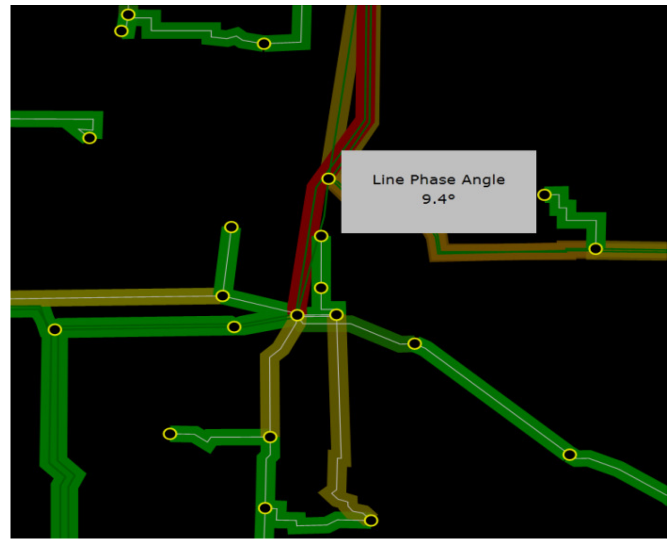
Example 3: Heat Map Representation of Energy Generation & Consumption

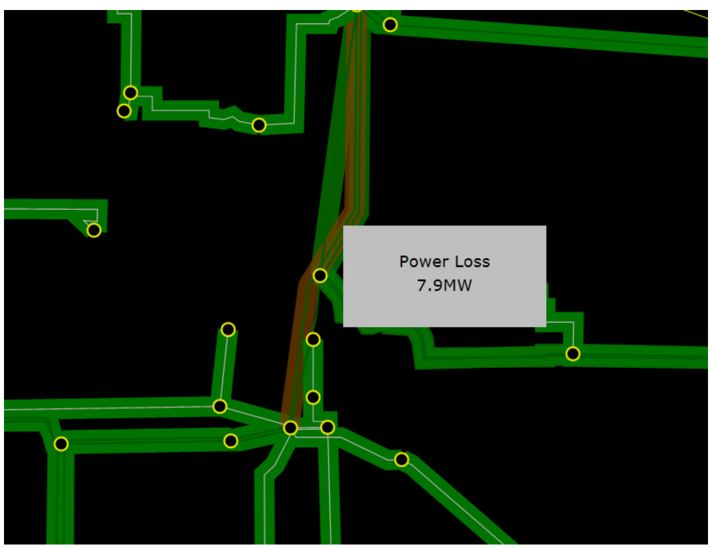


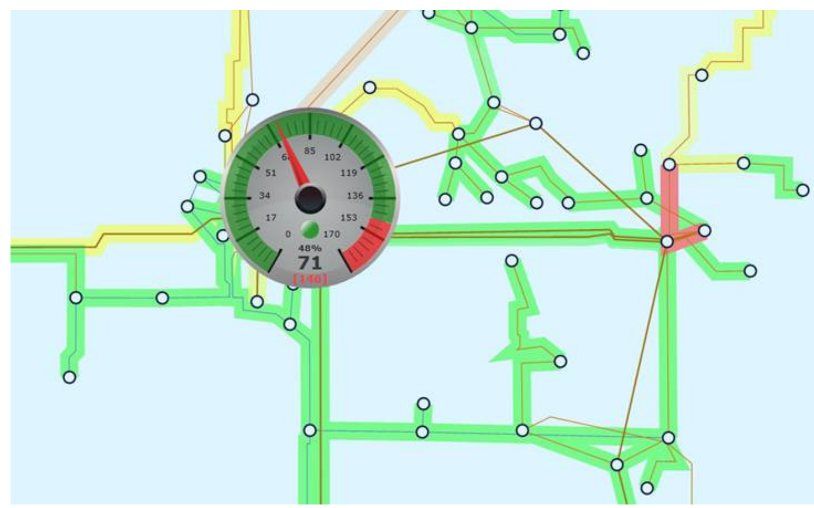


Example 4: Fuzzy Model Based Profiling









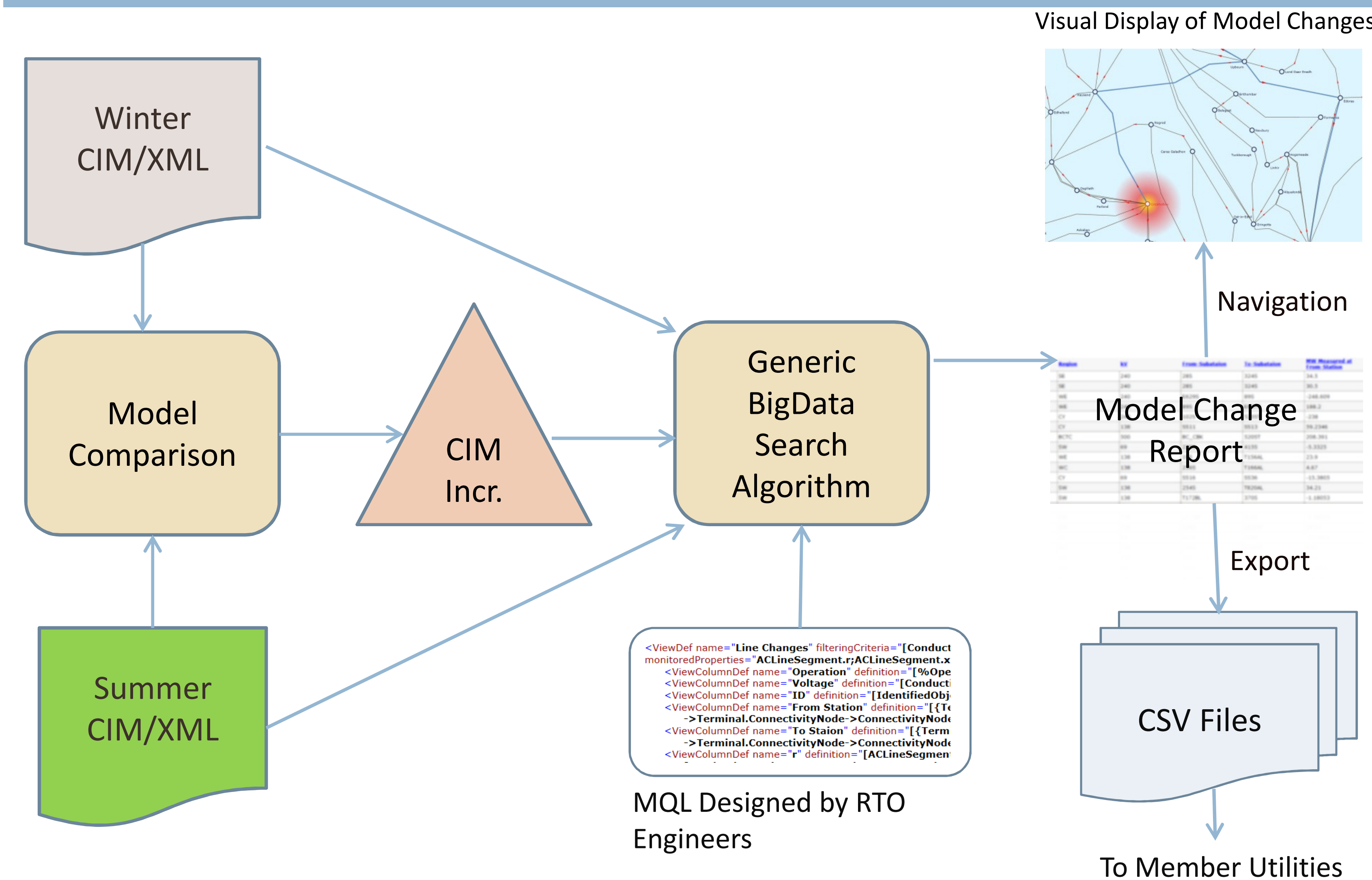
Application 1: Facilitating Model Sharing (Use Case)

Periodically, PJM need report model changes to their members & neighboring RTOs and have them confirm the accuracy of the model changes. This used to be primarily a manual process involving a lot of effort from both data providers and data recipients. The major challenges identified include:

- The model changes are represented in incremental CIM/XML format, which is not human friendly. CIM models grids at fine granularities, not suitable for reporting. Most importantly, many of the member utilities are lack of CIM expertise.

- The in-house developed spreadsheet-based model change reports are difficult and time-consuming to maintain. It is very difficult for member utilities to interpret these data-centric tabular reports.

Application 1: Facilitating Model Sharing (Solution)



Application 1: Facilitating Model Sharing (Proof-of-Concept)

- The model change reporting tool was derived by PJM modeling engineers with minimum engineering effort involved.

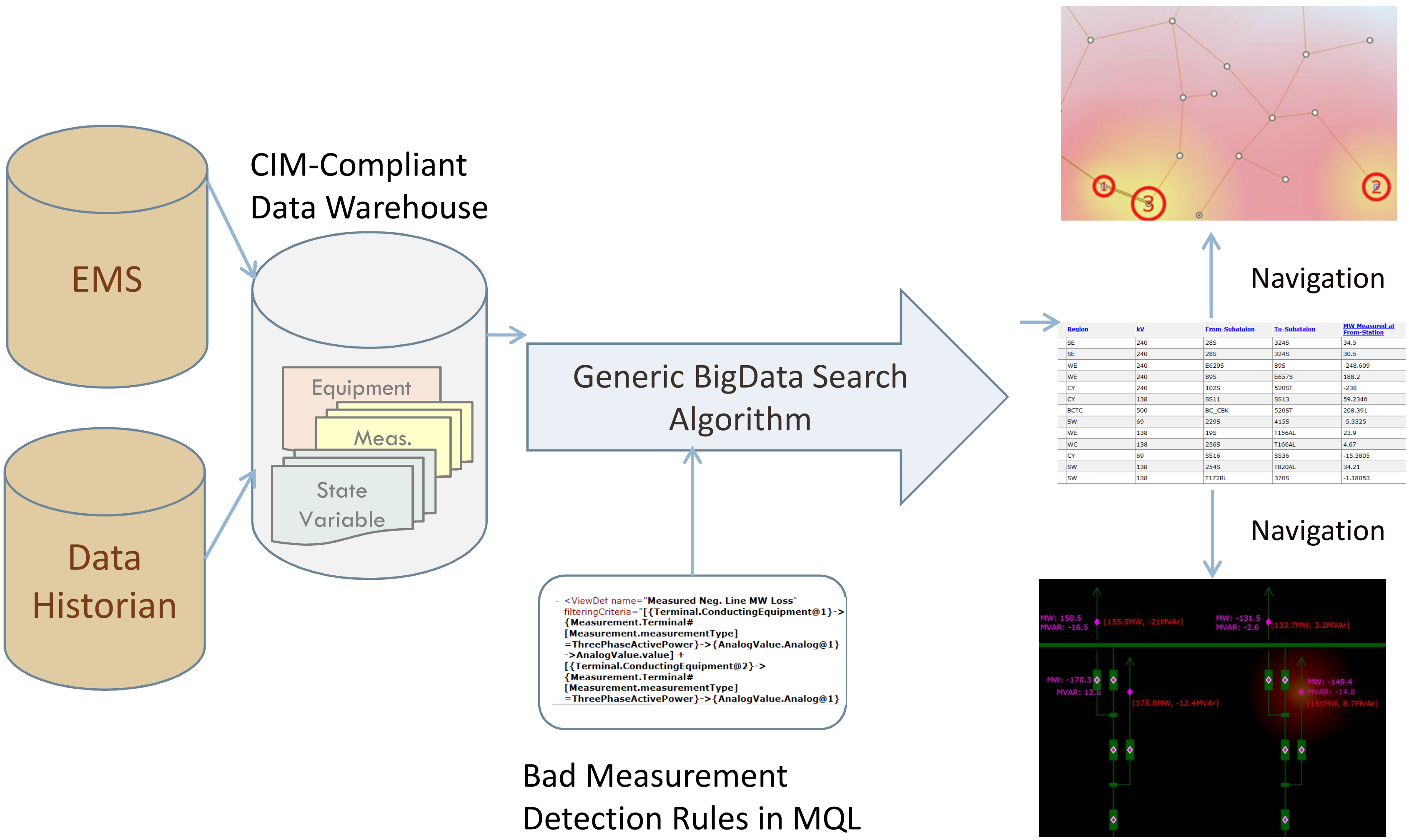
- In comparison with the legacy model change reporting tool, the new model change reporting tool has improved the business practice from the following perspectives:
  - Converting machine-friendly CIM Incremental to human-friendly tabular;
  - Transforming fine-granulated CIM objects to reality-based modeling entities familiar to model engineers;
  - Supporting advanced data presentation including classification, filtering, sorting, and grouping,
  - Visual display of model changes to facilitate interpretation and comprehension.

Application 2: Bad-Measurement Detection (Use Case)

Albert Electrical System Operator (AESO) recently extended their EMS network model to include a selected portion of the neighboring stations. During the project, engineers found that the quality of state estimation was downgraded due to the bad measurements from the external network.

- State estimator marks thousands of "suspect" measurements, most of which are either redundant or caused by noises.
- No indication of the severity and no classification.
- Many of the "suspect" measurements could be just temporary disturbances. The state estimation based solution is based on a single operational snapshot. It does not look at history.
- Tabular displays of "suspect" measurements are neither intuitive nor interactive.

Application 2: Bad-Measurement Detection (Solution)



Region	kV	From-Subataion	To-Subataion	MW Measured at From-Station
SE	240	28S	324S	34.5
SE	240	28S	324S	30.5
WE	240	E629S	89S	-248.609
WE	240	89S	E657S	188.2
CY	240	102S	520ST	-238
CY	138	SS11	SS13	59.2346
BCTC	500	BC_CBK	520ST	208.391
SW	69	229S	415S	-5.3325
WE	138	19S	T156AL	23.9
WC	138	256S	T166AL	4.67
CY	69	SS16	SS36	-15.3805
SW	138	254S	T820AL	34.21
SW	138	T172BL	370S	-1.18053

Application 2: Bad-Measurement Detection (Proof-of-Concept)

- Driven by a set of bad measurement detection rules specified in MQL, the query engine successfully detects a dozen of bad measurements combing the data from both EMS and data historian.
- Reports and visual displays are generated by visual display generator to indicate the location and impact for each of the detected bad measurements.
- After fixing the detected bad measurements either in model or in field, the quality of state estimation has been significantly improved.
- The effort involved in building the application is about two-man week from a consultant of Power Info and one-man week from a supporting engineer of AESO.

Conclusion

- CIM holds the key to Big Data Analytics in the utility industry.