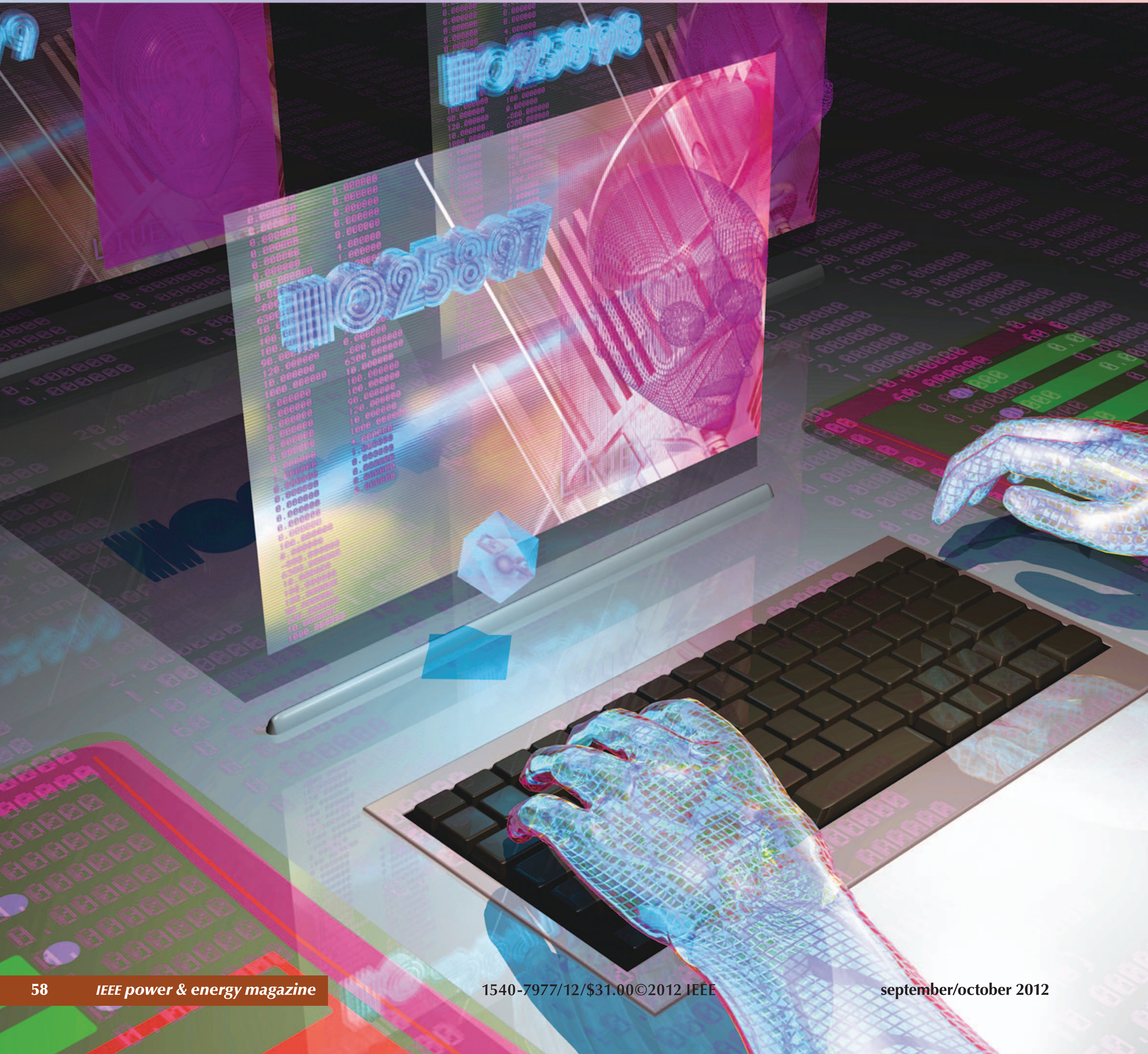
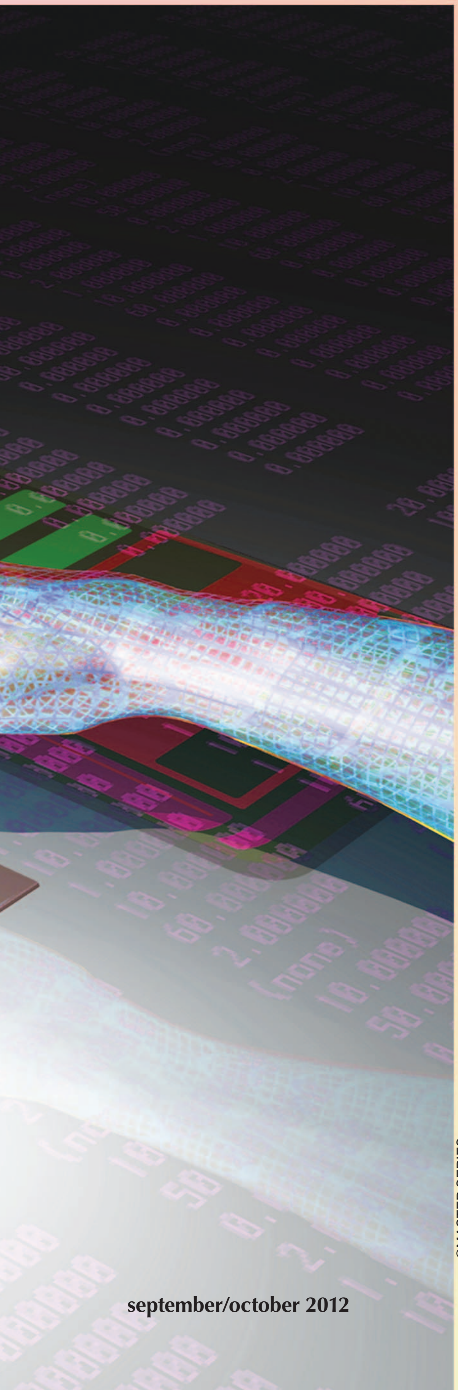
Measures of Value

By Tomo Popovic and Mladen Kezunovic



1540-7977/12/$31.00©2012 IEEE

Data Analytics for Automated Fault Analysis



THE POWER INDUSTRY IS EXPERIENCING AN ENORMOUS expansion of computer and communication devices in substations. As a result, a massive amount of measurement data is being continuously collected, communicated, and processed. This is partly due to the need for much better monitoring capability as power system loading and complexity of operation have increased. The installation of a large number of intelligent electronic devices (IEDs) to accomplish the monitoring task has created new challenges, such as cyber and physical security, time-synchronized data storage, configuration management, and efficient visualization. Automated data analytics solutions are the key to efficient use of IED recordings. This automated process includes conversion of measurements to data, processing data to obtain information, and extraction of cause-and-effect knowledge. This article provides real-life implementation examples of data analytics developed to handle measurements from digital fault recorders (DFRs) and digital protective relays (DPRs). The discussion addresses the implementation challenges and business benefits of such solutions.

Converting Field Measurements to Digital Data

When triggered, substation IEDs capture signals in a small time window that typically contains a few cycles of the prefault and up to three dozen cycles of the postfault data. These recordings consist of digital samples of multiple analog and status channels. A diagram of typical data sampling and processing in a modern IED is given in Figure 1. Prior to analog-to-digital (A/D) conversion, the input signals are sampled using the sample-and-hold (S/H) circuit at the times defined by the sampling clock. Synchronous sampling of all the input signals allows determination of phase angles among different analog input signals. This can be accomplished either by using one A/D converter serving all channels but having separate S/H circuits on each channel and a multiplexer that feeds another S/H circuit in front of A/D conversion (see Figure 1) or by using a separate S/H circuit and A/D convertor on each channel. Some older designs use a scanning method in which each channel is sampled and converted one at a time, causing a time skew among the corresponding samples on different channels. The quality of the data is affected by the conversion process and also by wiring, input transformer characteristics, clock accuracy, internal signal propagation, sampling rate, antialiasing filters, and so on. When implementing or using data analytics, it is very important to understand how the measurements are obtained and what the expected impact on the quality of the data is.

Data Analytics Starts with Data Integration

As the large-scale deployment of substation IEDs began to produce a "data explosion," it became obvious that data analytics solutions require several data integration functions, as shown in Figure 2.

Interface to IEDs

Software for interfacing with IEDs allows automated retrieval of newly recorded event data. The communication is typically implemented using vendor-specific software, which sometimes results in data's being stored in nonstandard and

Digital Object Identifier 10.1109/MPE.2012.2205319
Date of publication: 16 August 2012

> Correlated with the rest of the operational data, this knowledge can be used to enhance the decision making process when operating the system in real time.

proprietary file formats. File conversion into a nonproprietary format is then required when importing the retrieved data into the file repository to be accessed by variety of data analytics applications.

Data Warehouse

A flexible and standardized data repository called a data warehouse is used to support manual analysis needs as well as automated data analytics solutions. The data warehouse must be implemented using nonproprietary and standard formats. It should contain measurement data, configuration settings, and data analytics reports.

Data Analytics

The analytics functions can be implemented as stand-alone programs operated manually or in fully automated mode. The simplest form of data analytics reads the data from the repository, creates an output without corrupting the original data, and then sends the output results back to the data warehouse for storage. The data analytics sometimes utilize their own databases that may be decoupled from the substation data warehouse shown in Figure 2. This creates a challenge when a synchronization and integration of multiple data analytics is needed.

Visualization

While each data analytics function may have its own user interface, a universal approach for viewing results from all data analytics functions may also be desirable. Typical user interface options for fault analysis solutions include Web-based portals and event viewers; desktop-based event viewers and configuration editors; and various options for report dissemination, such as pagers, e-mails, text messages, printers, or faxes.

A Data Analytics Example: Fault and Disturbance Analysis

Fault and disturbance analysis entails taking measurements from IEDs triggered by the fault events and converting them

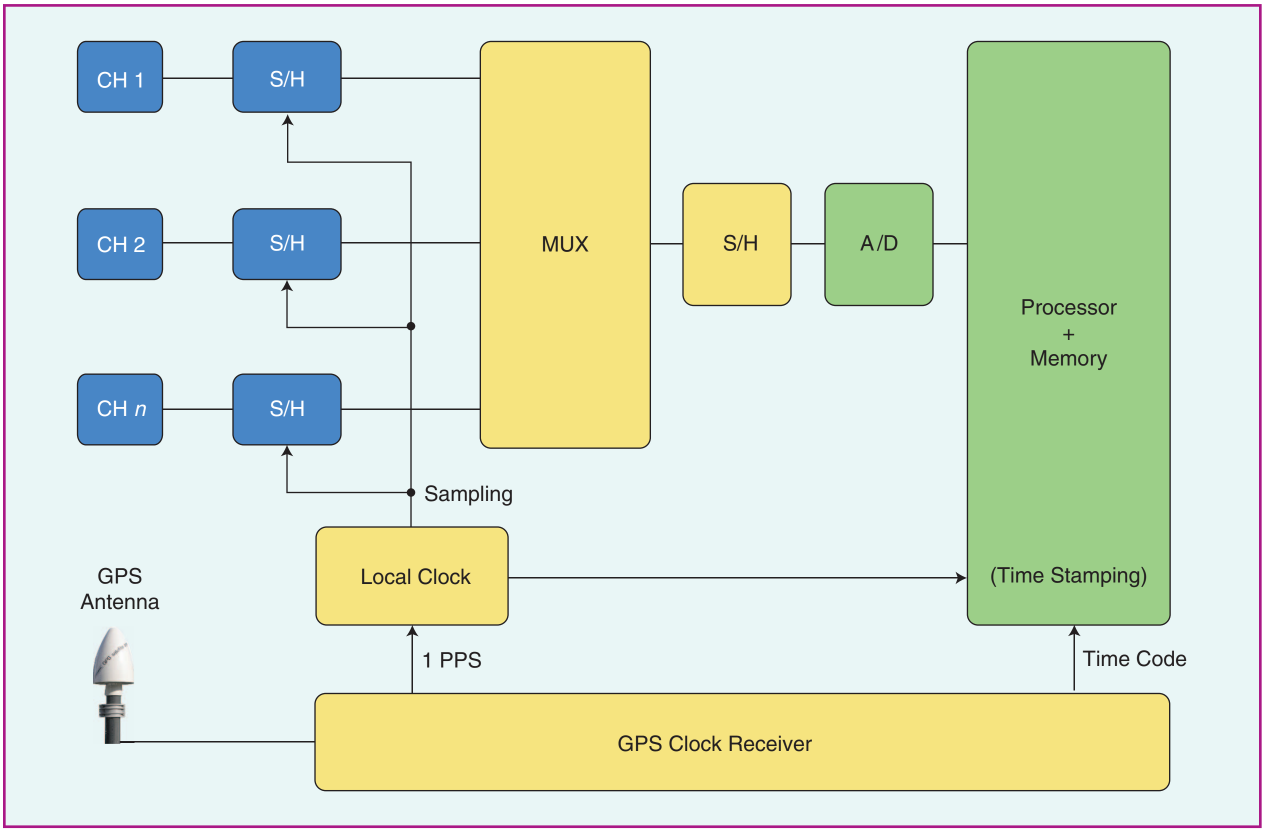


figure 1. Converting field measurements into digital data records.

> This article provides real-life implementation examples of data analytics developed to handle measurements from digital fault recorders and digital protective relays.

to data, processing data into information, and then using this information to extract knowledge about the fault event. The fault analysis can automatically provide various details, including identification of the affected circuit, whether the disturbance was a fault, fault type, fault location, duration, and evaluation of protection performance. All of this knowledge can be presented to the users and will help them take actions and make decisions more efficiently. This is especially important when there is a need for quick restoration of the system.

The implementation framework for an automated fault analysis is shown in Figure 3. The main components of the implementation example are the data analytics (fault analysis), the data warehouse, and visualization. Between the data analytics and data warehouse we have the following interfaces:

- ✔ IED data import/export
- ✔ configuration import/export
- ✔ analytics reports import/export.

The implementation of these interfaces should be independent of the types of IEDs used, and the same data warehouse should serve various data analytics functions. In this example, the fault analysis converts IED file formats, maps the configuration to imported IED data, performs digital signal processing, applies expert system logic for cause-effect analysis, and finally, prepares customized reports. The reports are then stored at the data warehouse and made available for later use. Visualization enables the user to directly communicate with and configure the particular analytics function. As shown in Figure 3, the visualization piece can interface with the data warehouse directly or, in some cases, be an extension of the data analytics function itself.

Sometimes the data analytics can be used to enable connections among different data management and processing systems. One such architecture that includes the fault analysis solution is given in Figure 4. There can be several substations, and each substation can have multiple IEDs: DFRs, DPRs, power quality meters (PQMs), and others. Substation data are collected and communicated via the substation PC and security gateways to the data integration system that belongs to the utility transmission group. The data are integrated, processed, and stored at the master station that also hosts the data warehouse. The data analytics can be used to connect to a wide-area measurement system, utility operations center, and other enterprise systems. If needed, it can even provide a connection to outside parties such as an independent system operator (ISO).

This example illustrates how the IED recordings that are traditionally considered nonoperational data can become operational. This is achieved by automatically downloading and processing the IED data and then sending the analysis reports to the plant information (PI) historian and SCADA. The reports provide information about affected circuits, fault types, calculated fault locations, assessments of fault-clearing performance, and conclusions as to whether faults were transient or permanent. Correlated with the rest of the operational data, this knowledge can be used to enhance the decision-making process when operating the system in real time.

The Configuration Challenge

Settings related to the power system include descriptions of monitored components such as transmission lines, buses, and transformers. There are also IED-specific settings, such as the details about the way particular IEDs are connected and configured. Finally, the analytics functions may have their own settings and configurations. All of these parameters are continuously experiencing both small and larger changes. These changes may be brought about by various upgrades in the system and equipment. In addition, there are changes to standards and recommendations issued by

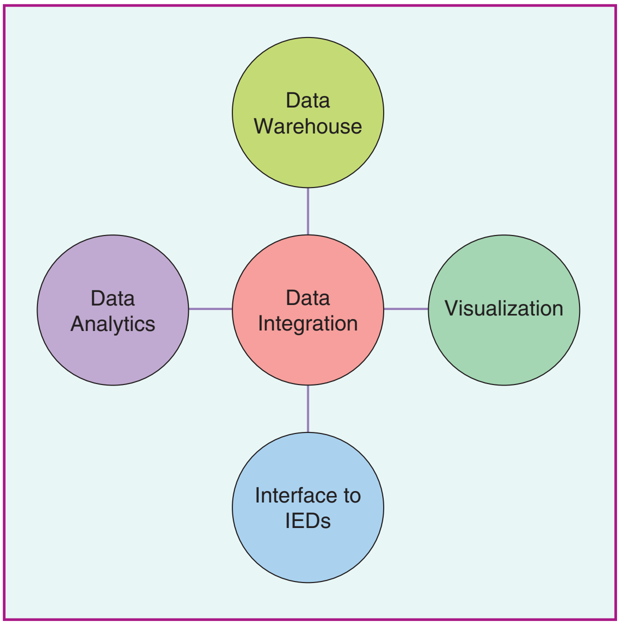


figure 2. Substation data integration as the foundation for data analytics.

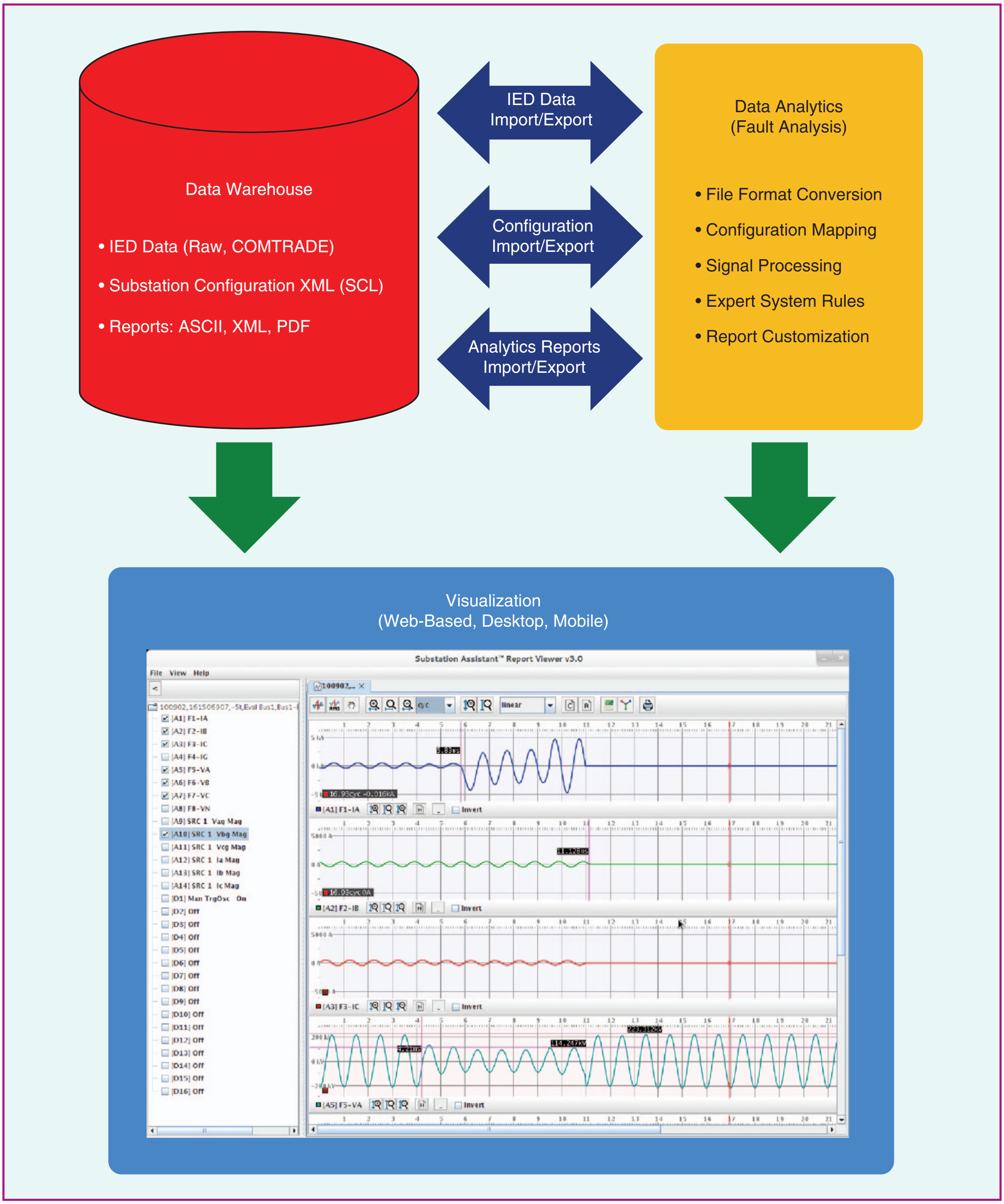


figure 3. Data analytics example: automated fault and disturbance analysis.

other entities such as IEEE, the North American Electric Reliability Corporation (NERC), and the Federal Energy Regulatory Commission (FERC), which are constantly evolving and affect various aspects and possible uses of substation data. Traditionally, such configuration changes have affected short-circuit study programs, simulation tools, the PI historian, and so on. Automated data analytics solutions are even more dependent on the correctness of these settings. All of the changes in the settings must be correctly captured in the configuration files using version control.

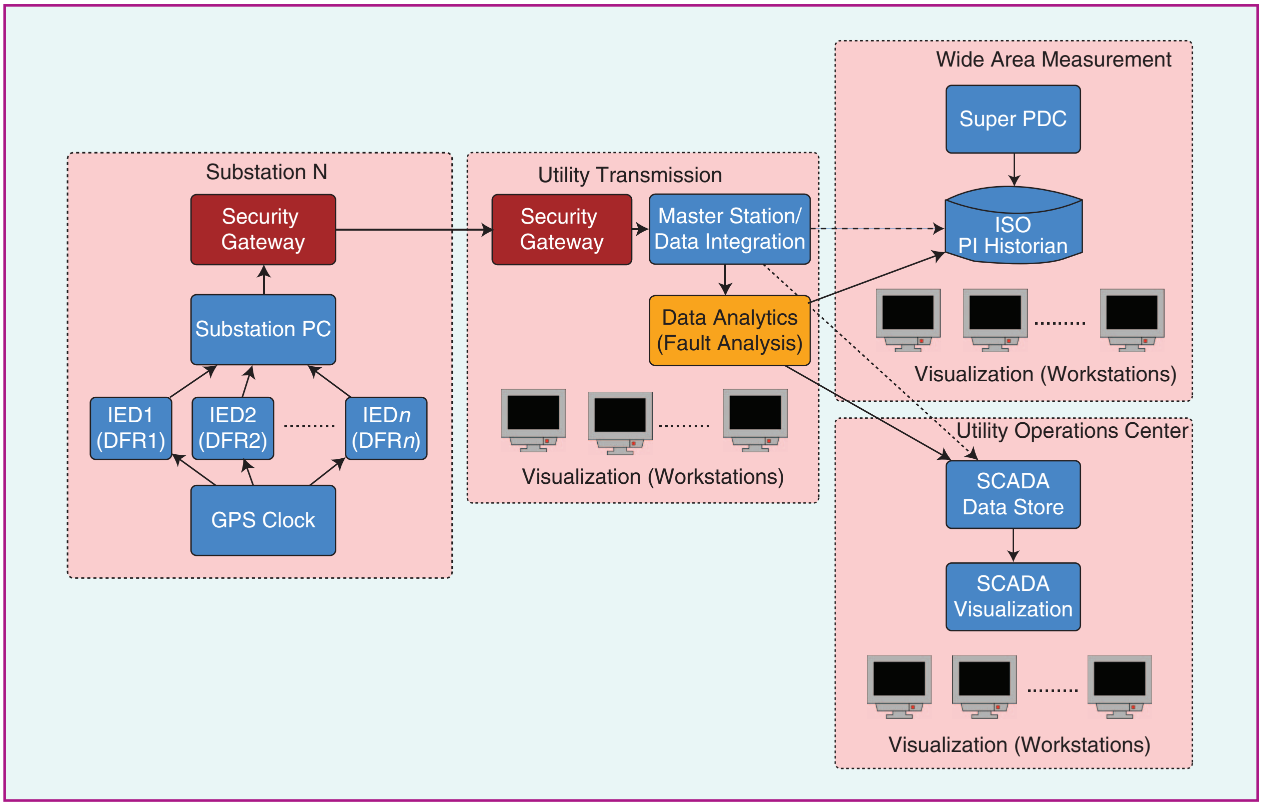


figure 4. Data analytics results can be sent to PI Historian and SCADA.

We will now provide an illustration of how substation monitoring with disturbance recording needs to be configured to automate the fault analysis application. A typical bus breaker arrangement for a transmission line is displayed in Figure 5. The example shows a breaker-and-a-half transmission line configuration. The measurement points of interest for fault analysis are marked with green labels. In order to enable automated operation of the fault and disturbance analysis for each circuit, we need to know the locations of the measurements of the voltage and current signals. We also need to map digital signals such as the breaker auxiliary statuses, relay trips, and protection scheme communication. The example includes monitoring of circuit breaker control status (element 52 in the IEEE standard naming convention), the associated protection of the transmission line (element 21), directional overcurrent protection (element 67), and the lockout relay (element 86). The protection scheme communications channels are presented with their respective transmit and received carrier and carrier frequency signals (TC and TCF). A detailed list of signals is provided in Table 1.

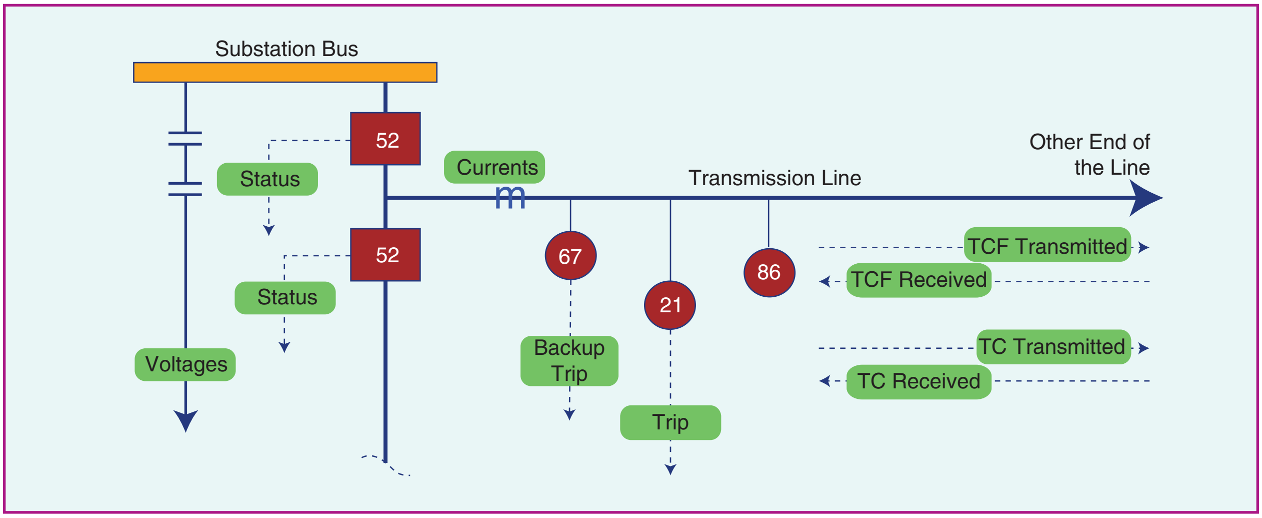


figure 5. Measurement points in a typical bus breaker arrangement.

> The data analytics converts a vast amount of raw substation data into useful information and subsequently into actionable knowledge, typically in the form of user reports.

table 1. Input signals for the fault data analytics applied to transmission lines.

Signal	Description	Type
I	Line currents: three phases or two phases and zero sequence	Analog
V	Bus voltage: three phases or two phases and neutral	Analog
PCB	Primary (bus) breaker contact status	Digital
SCB	Secondary (middle) breaker contact status	Digital
PRT	Primary relay trip status	Digital
BRT	Backup relay trip status	Digital
TCR	Blocking signal received status	Digital
TCT	Blocking signal transmitted status	Digital
TCFR	Breaker failure signal received status	Digital
TCFT	Breaker failure signal transmitted status	Digital

In the configuration settings, each transmission line is assigned the metadata describing how the signals from Table 1 are being monitored and mapped to corresponding analog or digital channels within the IED. The metadata may contain additional information, such as the line impedance and line length needed to automatically run fault location calculation. Even the GPS position of the line may be useful for displaying calculated fault locations on geographical satellite maps.

In general, the automated fault and disturbance analysis (see Figure 6) needs access to:

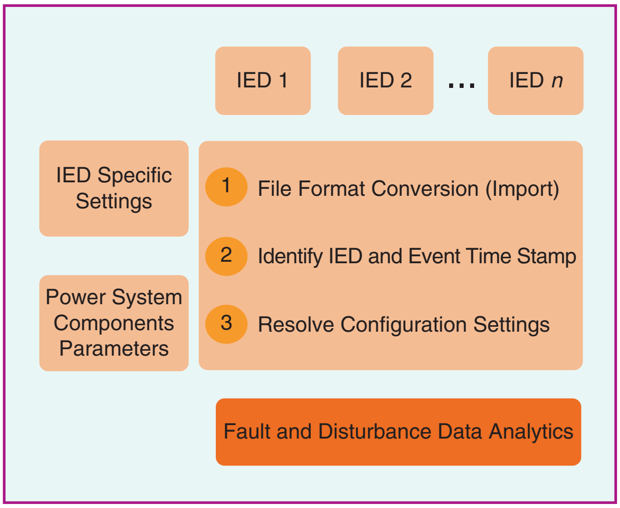


figure 6. Handling of the configuration settings is critical for data analytics solutions.

- ✔ fault and disturbance records as captured by substation IEDs
- ✔ IED-specific settings such as channel assignments and scaling
- ✔ power system transmission line parameters, such as line impedance, line length, mutual coupling, and GPS location
- ✔ the context in which the recorded disturbance and configuration will be analyzed, e.g., the circuit connection, the type of recording device, and the protection scheme used.

Recordings coming from the IEDs need to be matched with the corresponding IED settings as well as with the correct current power system component parameters. IED-specific settings sometimes come with the IED recordings, but it is not unusual to see those placed in a separate file or even kept on a remote computer. Easy access to the IED settings is critical in order to enable the fault analysis. Some of these issues are being addressed in current IEEE standards development work, including Common Format for Transient Data Exchange (COMTRADE) and Common Format for Event Data Exchange (COMFEDE).

Handling of the configuration settings can be implemented by interfacing to other systems such as short-circuit study program database, relay-setting coordination database, SCADA PI historian, or the International Electrotechnical Commission (IEC) 61850 Substation Configuration Language (SCL) files. The data analytics solutions can also have their own management and version control for the configuration settings.

Time stamping of the configuration settings is just as important as time stamping of the disturbance recordings. For each disturbance recording, we need to be able to locate the corresponding version of the configuration parameters to be used for the fault analysis. Figure 7 depicts an example of a simplified Unified Modeling Language (UML) sequence diagram for obtaining the configuration settings. In this case, each IED record is first converted into the nonproprietary COMTRADE file format. Then the preprocessing provides a unique IED identification ("id") and an event time stamp ("time"). The two parameters, id and time, are then used to retrieve the corresponding version of the configuration settings used by the automated fault analysis.

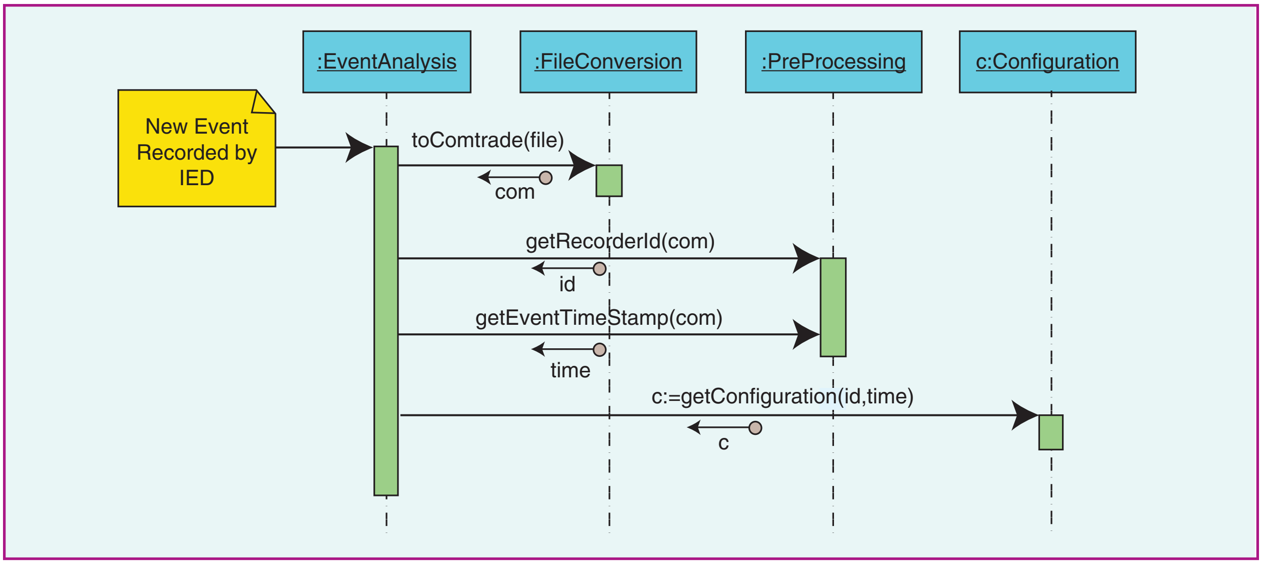


figure 7. Obtaining configuration settings corresponding to an event occurrence time stamp.

Inner Intelligence of the Fault Analysis

As discussed earlier, the data analytics consists of: 1) converting substation measurements into data, 2) translating data into information, and 3) extracting cause-effect knowledge from that information. For the fault analysis based on substation IED data, these main steps can be described as follows.

Measurements to Data

Measurements are being captured and recorded by substation IEDs. The recording is triggered by the occurrence of a fault or disturbance. The recording files are communicated and converted into nonproprietary formats defined by IEEE and IEC standards. The converted data are stored in the data warehouse. The event records are then matched with the configuration settings to perform extraction of the signal features.

Data to Information

The analysis identifies the affected circuit, such as a faulted transmission line or transformer, which further focuses the processing on the signals relevant to the selected circuit. The analysis determines the start and end of the disturbance and, based on that information, calculates prefault, fault, and postfault values for all the relevant signals. For a typical transmission line, these signals include current and voltage for each phase, as well as statuses for the digital signals, such as relay trip, breaker auxiliary contacts, communication, and so on.

Information to Knowledge

The information extracted in the previous step is used to acquire knowledge about the event. This is accomplished by applying the expert system rules for various steps of the cause-effect analysis: detecting the disturbance and determining the fault type, analyzing fault clearing, and evaluating the performance of protection relays, auto-reclosing logic, circuit breaker operation, and so on.

The inner intelligence of the expert system for automated fault analysis consists of the rules shown in Figures 8 and 9. A circle represents each rule subset, and each circle is related to a possible conclusion of the corresponding rule subset. Rule subsets are connected by directed lines, which means that one subset of rules produces a conclusion that is used by another subset of rules. In some cases, rules require both analog and digital quantities extracted from the fault recording files. The rules are designed to cover a wide range of possibilities, not to focus on special cases.

As shown in Figure 8, the event can be identified as "Not a Fault" disturbance, or through the Fault Type Detection, identified as a single line-to-ground, line-to-line, line-to-line-to-ground, or a three-phase fault. The relay operation can further be identified as a reclosing attempt. A breaker at the monitored substation can clear the fault, or the fault can be cleared by the protection at a remote substation. The disturbance can be a temporary fault, often called a "self-clearing" fault. Detection of a fault that was not cleared indicates a protection system failure. A reclosing attempt can result in either failure or success in clearing the fault. Even if the breaker auxiliary status is not monitored, its state (open or closed) can be determined based on the change and levels of the analog quantities representing phase currents.

The rules for protection system performance evaluation are shown in Figure 9. These rules are used to analyze the digital statuses from the breakers and communication gear used in protection schemes. The protection analysis evaluates the operation of primary and backup protection relays and the operation of the bus and middle breakers (shown earlier in Figure 5). Analyzing the states of the digital signals compared with the start and end times of the fault provides an in-depth evaluation of protection performance.

A report from automated fault analysis is given as an example in Figures 10 and 11. The report format was customized for the users in the protection group. This actual field recording triggered by a fault was analyzed, and the event was identified as a phase C—to—ground fault. The fault analysis determined the affected circuit, the disturbance's start and end times, the fault type, a fault location estimate, and protection performance. Figure 10 shows that the disturbance was about seven cycles long. The primary relay needed more than three cycles to trip,

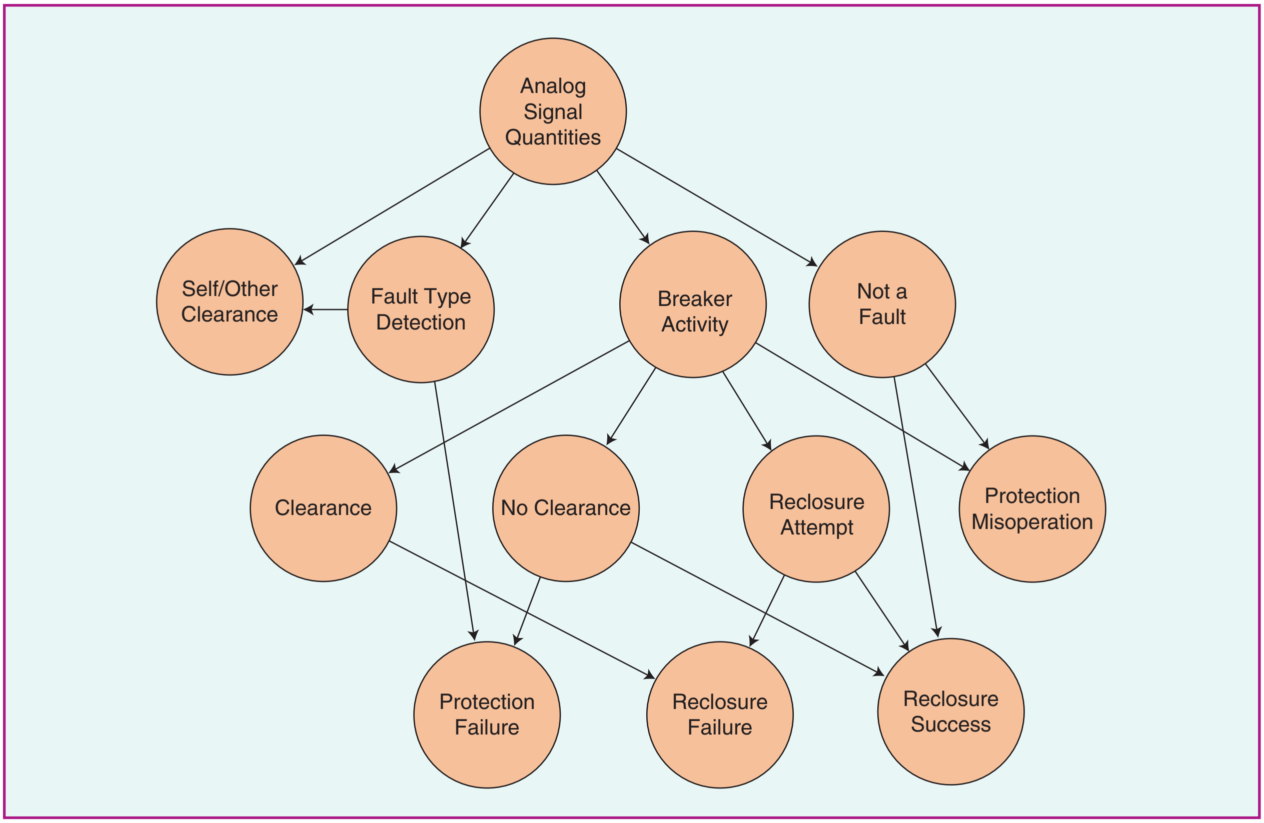


figure 8. Fault analysis rules based on analog signal inputs.

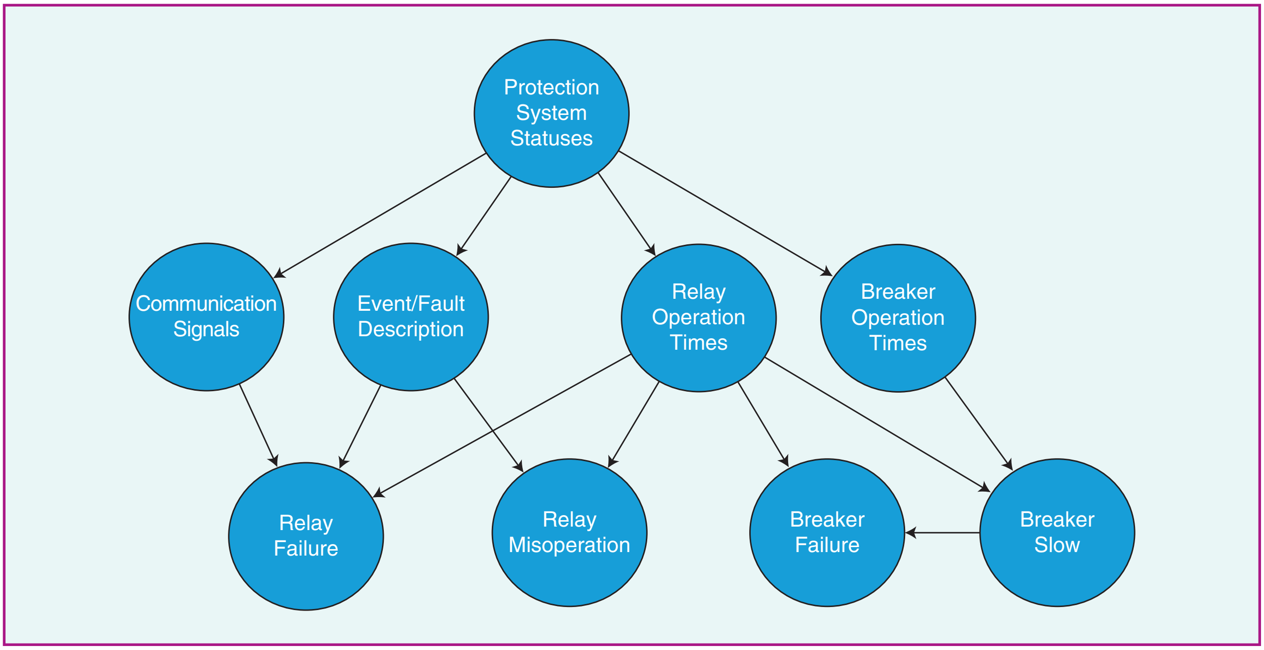


figure 9. Protection system performance evaluation rules.

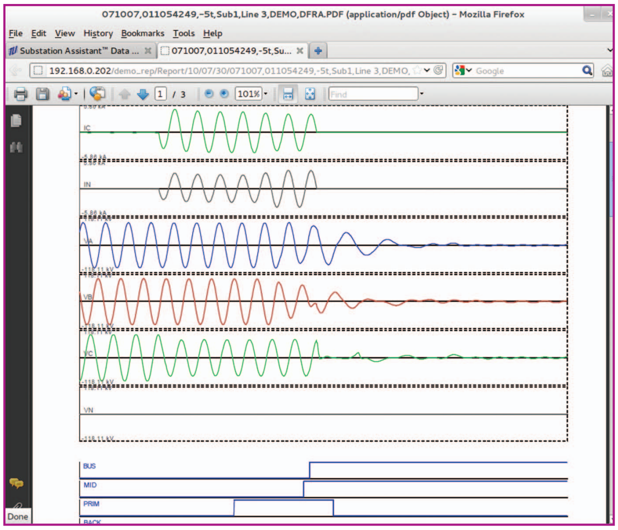


figure 10. Report example for viewing signals from the affected transmission line.

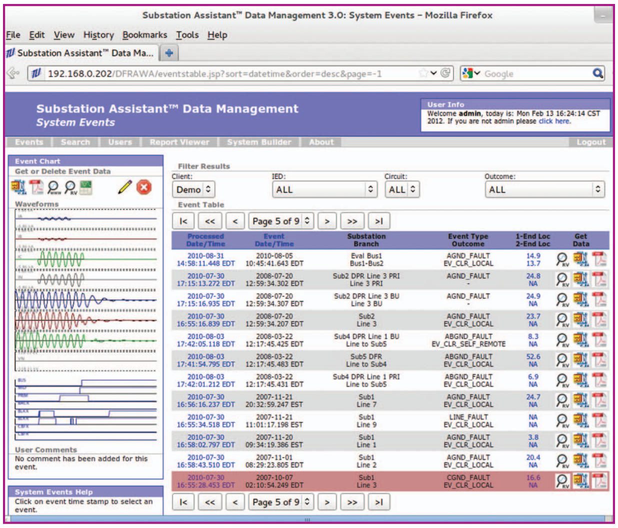


figure 12. Example of a Web-based user interface for fault and disturbance data analytics.

and it took the bus and middle breakers a little bit more than three cycles to open. The protection evaluation correctly indicated that the fault was cleared locally and also pointed out that the bus and middle breaker operations were slow (see Figure 11).

Data Analytics Visualization

The data analytics converts a vast amount of raw substation data into useful information and subsequently into actionable knowledge, typically in the form of user reports. Both the substation data and reports have to be made available to different user groups in a timely fashion. Different needs may result in the use of customized user interfaces and report formats. Figure 12 illustrates a Web-based user interface where a user can access the data analytics results in the form of an events table, using a standard Web browser. This Web solution displays the data and reports that are stored in the data warehouse. The reports are kept in an easily readable and nonproprietary file format (e.g., ASCII, XML, HTML, DOC, or PDF), as other automated data analytics may later use these reports as their input data. In the fault analysis example, this was accomplished using an IEEE file-naming convention (IEEE Standard C37.232-2007) and a standard SQL database engine.

The analysis reports can be sent automatically using SMS text, e-mail, mobile pagers, fax, and printers. Various notification options can be configured based on the event priority and user category. While the maintenance crew may appreciate a brief message identifying the substation, affected circuit, fault type, and location, the protection group may be interested in all the details about the fault and the related protection operations. When the automated reporting is combined with smartphone technologies, it can be a very powerful tool for "on the go" analysis (see Figure 13).

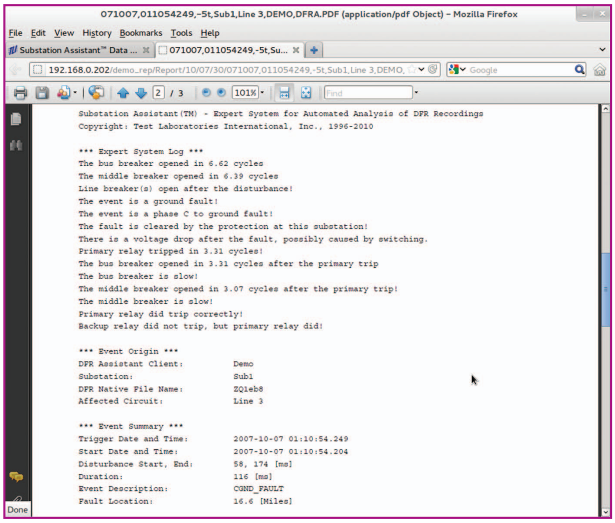


figure 11. Report example for viewing fault analysis results.

figure 13. Data analytics "on the go" using text/e-mail messaging and the mobile Web.

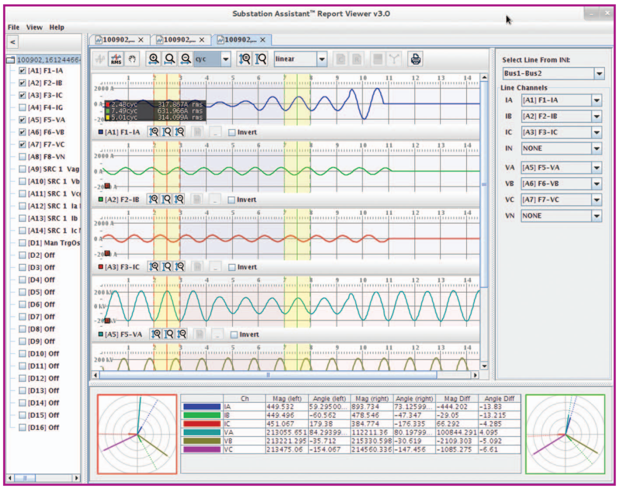


figure 14. Example of a desktop-based (rich client) user interface.

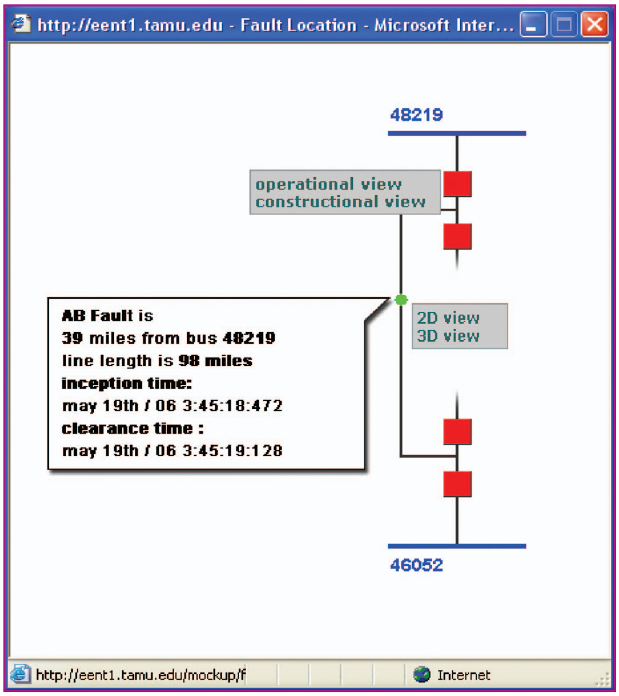


figure 16. Fault analysis integration with SCADA visualization.

The examples shown in Figures 12 and 13 illustrate "thin client" (Web, mobile) visualization. Another approach is to implement a "rich client" as the universal desktop-based user interface for fault analysis; one such tool, called Report Viewer, is shown in Figure 14. The viewer starts via the Web, using Java Web Start technology, and runs locally on the user's desktop or workstation. This tool may be used to manually inspect the signal waveforms as well as to access and read the fault analysis reports.

The rich client visualization enables a more native experience for the user and frees some of the server's resources. This is beneficial in situations when frequent user interaction and manual data manipulation are expected. Figure 15 depicts the use of manual fault location calculation. This tool lets the user interact with the results and modify the input parameters used in the fault location calculation algorithm. Parameters that can be changed are channel selection and prefault and fault measurements positions. The user can also change the line impedance and length, invert selected channels, and adjust scaling to further tune the data analytics results.

In addition to human users, it is possible that consumers of the data analytics results may be other systems, such as SCADA or PI historian systems. The data analytics function could export its results to the SCADA visualization, as shown in Figure 16. Additionally, the data analytics results can be interfaced with third-party visualization solutions. One such example is displaying calculated fault locations on a Google Earth satellite map, as illustrated in Figure 17.

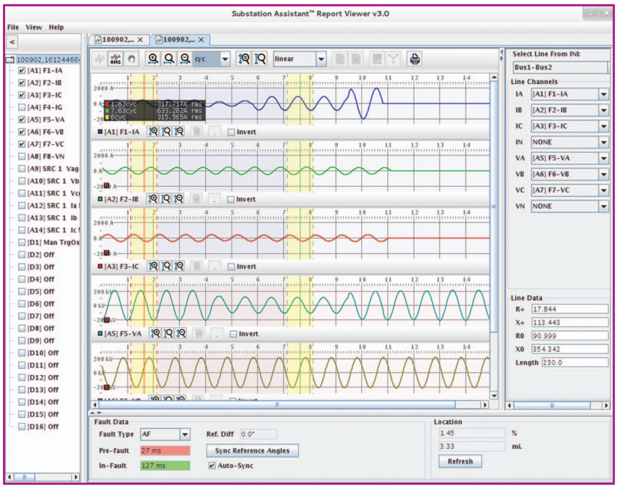


figure 15. A tool for manual fault location calculation.

figure 17. Integration with a third-party visualization: displaying the calculated fault location on a satellite map.

> In addition to human users, it is possible that consumers of the data analytics results may be other systems, such as SCADA or PI historian systems.

table 2. The benefits of automated data analytics.

Category	Improvements
Reliability	Reliability of assets, resilience to random events, reliability of operating decisions, robustness of system wiring and data
Productivity	Data integration, analysis, viewing, and archiving; event reporting
Capital investment	New data collection does not require new wiring, IEDs, communications, procedures, and so on; substation data analysis software is not as costly to install as hardware; hardware will not become stranded due to an inability to produce useful data.
Regulatory compliance	NERC, FERC, public utility commissions (PUCs), reliability coordinators, large customers

Benefits of Automated Data Analytics

There are several benefits of the automated data analytics coming from fault analysis based on substation IED data. These include:

- ✔ There is a major reduction of time spent on substation data handling, and analysis, either manual or automated, assures higher personnel productivity.
- ✔ Automated integration and archiving of the substation data using nonproprietary and standard data formats facilitate future data analytics implementations.
- ✔ Saving recordings from multiple IEDs corresponding to the same power system event provides the redundancy needed for improved data integrity checking.
- ✔ A standard data warehouse design keeps the solution open for implementation of different user interface tools, including integration with third-party visualization.
- ✔ The universal report-viewing and waveform inspection tools for accessing substation data regardless of IED type, model, and vintage make the data source transparent to all users.
- ✔ Providing the data analytics reports to multiple user groups in a format customized to fit their needs allows for a more focused and efficient decision-making process.
- ✔ Automated data analytics may be of great help in restoring the system by providing information and knowledge about the fault in a timely fashion.
- ✔ The inherent scalability of the proposed data analytics concept allows for the future addition of new IEDs as well as the implementation of new data analytics functions.
- ✔ The data analytics functions can be used to interconnect the systems within a utility enterprise, or even to connect with external entities, creating value for multiple users.

The data analytics value proposition is tied to many opportunities for return on investment such as reliability, productivity, capital investment, regulatory compliance, and standardization (see Table 2). Combining this fact with the trend toward large-scale deployment of substation IEDs makes automated data analytics solutions highly desirable.

For Further Reading

M. Kezunovic, B. Clowe, B. Ferdanesh, J. Waligorski, and T. Popovic, “Automated data retrieval, analysis and operational response using substation intelligent electronic devices,” in *Proc. CIGRE Session*, Paris, France, paper B5-206, pp. 1–8, Aug. 2012.

M. Kezunovic, “Translational knowledge: From collecting data to making decisions in a smart grid,” *IEEE Proc.*, vol. 99, no. 6, pp. 977–997, June 2011.

P. Myrda, M. Kezunovic, S. Sternfeld, D. R. Sevcik, and T. Popovic, “Converting field recorded data to information: New requirements and concepts for the 21st century automated monitoring solutions,” in *Proc. CIGRE Session*, Paris, France, paper B5-117, pp. 1–8, Aug. 2010.

T. Popovic and M. Kuhn, “Automated fault analysis: From requirements to implementation,” in *Proc. IEEE PES General Meeting*, Calgary, AB, Canada, July 2009, pp. 1–6.

Biographies

Tomo Popovic is with XpertPower Associates.

Mladen Kezunovic is with Texas A&M University and XpertPower Associates.

p&e