TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

JOONAS PUURTINEN

BIG DATA MINING AS PART OF SUBSTATION AUTOMATION AND NETWORK MANAGEMENT

Master of Science Thesis

Examiner: Professor Pekka Verho
Examiner and topic approved by the Faculty Council of the Faculty of Computing and Electrical Engineering on 5th May 2014

ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY
Master's Degree Programme in Electrical Engineering
PUURTINEN, JOONAS: Big Data Mining as Part of Substation Automation and Network management
Master of Science Thesis, 76 pages
May 2014
Major: Power Systems and Market
Examiner: Professor Pekka Verho
Keywords: Big Data, Data Mining, Substation Automation, Disturbance Recording, Maintenance

All fields of industry are constantly seeking ways to improve their efficiency. This is now especially true for power systems as they are facing one of the biggest challenges yet – how to cope with constantly increasing demands for electricity distribution with ageing power grid. Utilization of the big data mining in power systems presents one possible way to improve cost-efficiency and achieve higher level of reliability even with the ageing infrastructure. The target of this thesis is to research and develop ways to get additional information out of the currently mostly ignored disturbance recordings and history of process data.

The complexity of big data mining poses a great challenge for system developers. Power systems are among the best systems to get-started with big data mining solutions as they consist mainly of structured and semi-structured databases with vast amounts of information. The different naming conventions used in different systems along with great variety of different protocols hinders the easy comparison of information obtained from separate systems.

This thesis begins with the study of current naming conventions used in the power systems. Two standards, the COMTRADE and the IEC 61850, that define the organizing of data are looked into. This information is used to create a novel naming convention for future use within big data mining applications. The naming convention is chosen so that it supports the needs of current and future needs as well. The creation of a reliably structured central database is one of the key elements of practical data mining solution.

A system concept called Smart System Analyser developed for big data mining in power systems is presented next. It consists of relational SQL historian database and a novel calculation engine built around currently existing proven products. System components are described in detail and their operation explained.

The practical parts of this thesis is about the testing of this novel system first in simulated environment and then with actual power distribution company data. Even the early stages of the pilot testing show the potential for future development and benefit from power system data mining. An application is made for protection operation time calculation using the presented novel system. It is ran with data obtained from disturbance recordings and the results are visualized in a web interface

TIIVISTELMÄ

TAMPEREEN TEKNILLINEN YLIOPISTO
Sähkötekniikan koulutusohjelma
PUURTINEN, JOONAS: Big Data osana sähkönjakeluautomaatiota ja sähköverkon hallintaa
Diplomityö. 76 sivua
Toukokuu 2014
Pääaine: Sähköverkot ja -markkinat
Tarkastaja: Professori Pekka Verho
Avainsanat: Big Data, Sähkönjakeluautomaatio, Häiriötallenne, Kunnossapito, Tiedonlouhinta

Teollisuudelle ominaista on jatkuva pyrkimys prosessien tehostamiseen. Ilmiö koskee erityisesti sähkönjakelujärjestelmiä, sillä jatkuvasti kasvava sähkönkulutus yhdistettynä vanhenevaan sähköverkkoon aiheuttaa valtavan haasteen tulevaisuudessa. Big Data sovellusten hyödyntäminen tarjoaa erään mahdollisuuden tehokkuuden ja korkeamman käyttövarmuuden saavuttamiseksi nykyverkoilla. Tämän diplomityön tavoitteena on tutkia miten tiedonlouhintaa hyödyntämällä saataisiin lisäarvoa tällä hetkellä pääasiassa sivuutettavasta häiriötallenne- ja prosessihistoriadatasta.

Big Datan monimuotoisuudesta aiheutuu merkittävä haaste järjestelmäkehittäjille. Sähkönjakelujärjestelmät ovat yksi parhaista sovelluskohteista tiedonlouhinnalle, sillä suuria datamääriä säilövät tietojärjestelmät perustuvat pääasiassa jäsenneltyihin tietokantoihin. Erilaiset nimeämiskäytännöt ja eri aikakausin protokollat tekevät kuitenkin mahdottomaksi suoran yksinkertaisen vertailun tietokantatallenteiden välillä.

Tämä diplomityö alkaa perehtymisellä nykyään sähkönjakelujärjestelmissä vallitseviin nimeämiskäytäntöihin. Nimeämiskäytäntöjä tutkitaan sekä häiriötallenne COMTRADE-standardin että sähkönjakeluautomaatio-standardin IEC61850 osalta. Tätä kirjallisuustutkimuksesta saatua tietoa käytetään uuden nimeämiskäytännön määrittelemiseen. Määrittelyssä on tehty valintoja jotka tukevat nykyisiä ja tulevaisuuden vaatimuksia tiedonlouhintasovelluksille. Ennakoitavissa olevaan nimeämiskäytäntöön perustuvan keskitetyn tietokannan luominen on avainasemassa, kun lähdetään kehittämään käytännössä toimivia tiedonlouhinta ratkaisuja.

Seuraavaksi esitellään Smart System Analyzer konsepti sähkönjakelujärjestelmien tiedonlouhintaan. Konseptin mukainen järjestelmä perustuu SQL-historiatietokannan sekä uudenlaisen laskentaympräristön käyttöön. Järjestelmä hyödyntää jo olemassa olevia ratkaisuja mahdollisimman tehokkaasti. Sen osat sekä niiden toiminta esitellään työssä yksityiskohtaisesti.

Työn käytännön osuus koostuu järjestelmätestauksesta sekä simuloidussa ympäristössä että käytännön sähkönjakelujärjestelmässä. Sähkönjakelujärjestelmässä toteutettava pilottiprojekti osoittaa jo alkuvaiheessa konseptin luomat tulevaisuuden mahdollisuudet. Työn puitteissa kehitettiin ohjelma suojausaikojen laskentaan häiriötallennedataa käyttäen, jonka tulokset esitettiin web-käyttöliittymässä.

PREFACE

This is a master’s thesis written in Substation Automation Systems department in ABB Vaasa. The main goal was to look into ways of benefitting more from information already gathered from present power systems. The work was carried out with ABB development team situated in Tampere. I’d like to express my thanks for each and everyone involved with the project and especially to Antti Kostianen who also guided me through the process. Another special thanks goes to my supervisor Professor Pekka Verho from the Department of Electrical Engineering in Tampere University of Technology. I’d like to also thank my co-workers at Substation Automation Systems here at ABB Vaasa for lending me a hand whenever I needed help.

Last but certainly not least I’d like to thank my good old friend Marko Lamminsalo for his excellent advice regarding the thesis work.

Vaasa, May 14th, 2014.

Joonas Puurtinen

ABBREVIATIONS

API	Application Programming Interface. API specifies how software components should interact with each other.
CBM	Condition Based Maintenance. Maintenance strategy where components are serviced when need arises.
COMTRADE	Common format for Transient Data Exchange for power systems. File format for storing oscillography and status data related to transient power system disturbances.
DAS	Distribution Automation System. System consisting of all the remote-controlled devices at the substation level.
DMS	Distribution Management System. User interface systems providing operators grid status information.
EMS	Energy Management System. System of computer-aided tools used by operators of electric utility grids to monitor, control, and optimize the performance of the generation and/or transmission system.
GIS	Geographic Information System. Computer system designed to capture, store, manipulate, analyze, manage, and present all types of geographical data.
GPRS	General Packet Radio Service. Packet oriented mobile data service on the 2G and 3G cellular communication systems.
GPS	Global Positioning System. Space-based satellite navigation system that provides location and time information in all weather conditions, anywhere on or near the Earth where there is an unobstructed line of sight to four or more GPS satellites.
HSPA	High Speed Packet Access. Mobile telephony protocol that extends and improves the performance of existing 3rd generation mobile telecommunication networks.
IED	Intelligent Electrical Device. Microprocessor-based controllers of power system equipment
LTE	Long Term Evolution. Standard for wireless communication of high-speed data for mobile phones and data terminals.
OPC	Open Platform Communications. The standard specifies the communication of real-time plant data between control devices from different manufacturers.
RBM	Reliability Based Maintenance. Maintenance strategy where components are serviced according to their condition and criticality for grid operation.
RTDB	Real Time Database. The SQL database which is intended to store historical information.

SCADA	Supervisory Control and Data Acquisition. System to provide control of remote equipment in power systems.
SGEM	Smart Grids and Energy Markets. Finnish project aimed to speed up the development of international smart grid solutions.
SQL	Structured Query Language. Special-purpose programming language designed for managing data held in a relational database management system (RDBMS).
SSA	Smart System Analyzer. Name of the concept of performing automated big data related analysis based on substation measurement data.
TBM	Time Based Maintenance. Maintenance strategy where components are serviced periodically.
UHF	Ultra High Frequency. Range of electromagnetic waves between 300 MHz and 3 GHz.
VHF	Very High Frequency. Range of radio frequency electromagnetic waves from 30 MHz to 300 MHz.

1 INTRODUCTION

In this chapter the target objectives of this thesis with some background are described. After reading through the chapter the reader should have an overall picture of the problems in which this thesis aims to answer.

1.1 Background

Our current society has been built around the concept of constant economic growth. As result it has been steadily growing since 19th century, the early days of industrialization. This ideology has had the greatest impact on industry as its level often defines the pace of economic growth. To meet the demands for greater growth the industry has constantly streamlined and improved their processes. This is especially true for electricity distribution and generation because electricity has been our main method of transferring and consuming energy. Long gone are the days when simple analog relays were an improvement over the previous technologies. As we entered the information age the affordable personal computer became the main way of improving efficiency. Now that the use of PC has become widespread and matured as technology the solutions are becoming increasingly complex. To improve them poses even greater challenge as it requires considerable amount of research, development and risk taking. To maintain the growth, new ways of improving efficiency are constantly looked into of which the automated analyzing of huge data masses is currently deemed most promising – the dawn of big data mining is coming.

1.2 Objectives

The main objective of this thesis is to study and test different ways to use the data gathered from distribution network more effectively. The scope of this thesis is limited to examining ways to benefit more from process data and disturbance recordings. The similar topic was looked into in 2008 by Jaakko Yliaho in his Masters thesis Disturbance Recording Files Analyzing in Historian Database which gives a general view on how the information within disturbance recordings could be used (Yliaho, 2008). The added value of this information is sought from automated analyzing of data masses using a novel system developed at the ABB. The system consists of MicroSCADA Historian as a real time relational database and an external calculation environment. The goal is to develop, test and evaluate the system capable of gathering, storing, automatically analyzing the network data and visualizing the findings in an easy to read form in an actual electricity distribution network. The practical work of this thesis was designing, programming and testing an algorithm for automated protection time calculation based on the data obtained from fault recordings. In addition to this a novel naming convention was developed for use in future databases which utilize these automatic analyzing functions.

1.3 Structure

The chapter 2 addresses the state of today's electricity distribution environment. Current power system is reviewed and the expectations for it are listed. Its basic architecture and conventions of controlling it through substation automation systems is presented. This chapter is intended to give a baseline and motivation for which this thesis is all about. Some thoughts are given into why uninterrupted electricity distribution is so important in today's world and how the future looks from the smart grid perspective. When the need is established the chapter 3 moves closer into solution by giving a thorough introduction into one of the biggest phenomena affecting the automation systems – the big data mining. In this chapter big data is given a basic definition and some ways how to benefit from it. At the end of the Chapter 3 a scenario of how a future electricity distribution company might harness the big data to its benefit is presented. The chapter 4 deals with the biggest problem the big data analyzing systems have to deal with, which is the diversity of the data. In this chapter baselines for using already widely accepted standards as ways to structure and standardize gathered data are discussed. The chapter 4 describes the disturbance recording and standard defining its form as well as a novel naming convention developed specifically for automatic analyzing algorithms which bases heavily on the already widely adopted IEC61850 standard. In the chapter 5 the smart system analyzer concept is presented as a possible solution for big data mining in power system applications. The system components are introduced and their functions explained. At the end of this chapter a simple proof of concept testing is done in a simulated environment. The 6th and last major chapter of this thesis covers a practical pilot testing of the system in an actual electricity distribution system. Few analyzing cases are presented with the solutions for solving them. Finally a possible way to visualize the results is presented.

2 ELECTRICITY DISTRIBUTION SYSTEM

The Physical architecture of today's distribution networks is mainly radial, branching out in a tree like structure. Only in cities there exists multiple routes for power supply. In rural areas there are also some interconnections between parts of the network to help reliability issues but most of the consumers at these areas are supplied by only one route of network. In the case of a fault or an on-going maintenance on this connection, all customers supplied by this part of the network experience an interruption. The Figure 2-1 shows principle of current electric power architecture.

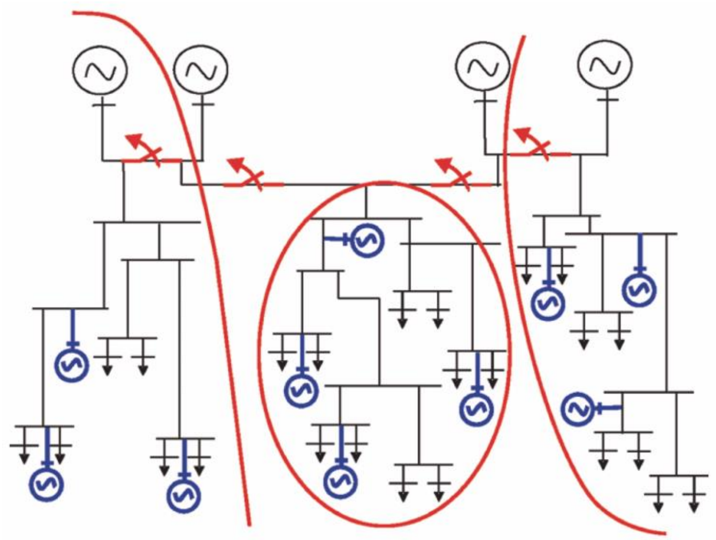


Figure 2-1 *Current electricity distribution network architecture. (Lukszo, et al., 2010)*

Power is produced in large centralized power plants often remote from load centers. Along an interconnection between source of electric power and a load, there are multiple crucial network components such as transformers, breakers and power lines. The actual status data out of these components is gathered by relays and metering devices.

2.1 Our dependancy on electricity

Our society is well past the point where electricity was just nice luxury to have. Of all the essential resources the electricity is the one on which a functioning of society relies the most and it is used almost everywhere and by everyone. Our lives are dictated by the constant availability of electricity. Water pumps are driven by electric motors, the majority of heating systems require at least some electrical power to operate and shops rely on electricity to preserve and even sell their goods. Many of the modern day jobs require the use of computer, which are run by electricity and the list goes on. The long disruptions of electric supply are disastrous for the current economies. All this places enormous requirements for the reliability of the electricity distribution systems. In the western countries the increasing reliability of these systems has caused many to take continuous electric supply for granted. The breakdowns of the electricity systems are usually minor and inflict only temporary discomfort for the users, but as recent events have shown, major blackouts can happen.

During autumn of 2003 there were 2 major breakdowns. First happened in the North America and as result approximately 50 million people were left without electricity. It took four days to restore the power. During this blackout it is estimated that USA's economy took a hit of between US$4 and US$10 billion. The second happened in Italy where a similar blackout left over 55 million people without power for only 3 hours but it still cost four lives (Lukszo, et al., 2010). In winter 2011 Cyclone Patrick (Tapani) hit the Scandinavia and caused widespread blackouts. Households which relied in electric heating had to survive without any heat source. Loss of power on large scale caused also GSM network to blackout after few hours when the reserve batteries died (Energiateollisuus Ry, 2012). The poor network resiliency is especially problem in the US where ageing network is trying to cope with extreme weather phenomena such as hurricanes and major storms. An inflation-adjusted estimated of US18$ and US33$ billion dollar costs to US economy have been caused by severe weathers. (US Department of Energy Facilities, 2013)

As these examples clearly show the reliability of the electricity distribution system is something to focus on. While completely avoiding major blackouts may require major refurbishments of the electricity distribution systems, still even minor improvements have direct economic impact and ultimately can be the matter of life and death.

2.2 Distribution automation

The regulation of the energy markets has caused electric power utilities to run their businesses as efficiently as possible. In particular, the owners of the power distribution networks are being required to improve areas of the network with substandard reliability. The owners are also being required to maximize the use and the life of their network assets by constant monitoring and maintenance. Power quality is also an important issue which is being monitored by authorities. Network control and automation systems have enabled network owners to adapt and succeed in the constantly evolving field of power distribution. (Automation In Power Distribution System: Present Status, 2012)

The distribution automation can be considered as an umbrella term for the combination of **distribution management system** and **distribution automation system**.

The DMS focuses on the control room, where it provides the operator the status of the network being controlled. It manages all of the functions needed to properly control and manage the network on regular basis. The DMS works through an organized network model database and it must have access to all supporting IT infrastructure.

The DAS consists of all of the remote-controlled primary devices at the substation and feeder levels, the local automation devices, and the communications infrastructure (Northcote-Green, et al., 2007). The following Figure 2-2 demonstrates a typical layout of a DAS.

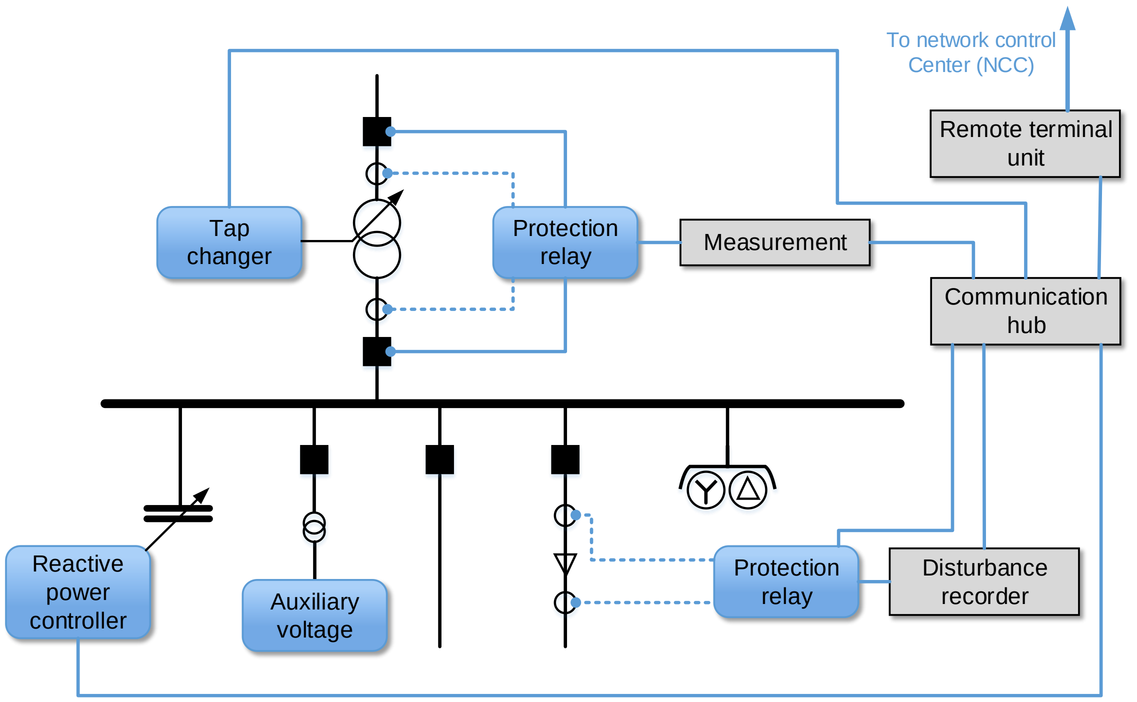


Figure 2-2 *Principle layout of a distribution automation system located in a substation.*

The primary devices directly connected to the processes such as circuit breakers are managed by relays while the transformers and compensators are managed by specialized controllers. The relays measure and conduct protective functions within the substation while being in connection to the control center through substation communication systems. Another major part of the distribution automation is on-the-fly adjustment of the substation equipment which is not directly involved into protection scheme. As an example the distribution system's voltage levels are maintained by moving the primary transformer's tap changer. An equally important task is to maintain a reasonable balance between real and reactive power flow. This power factor correction is done by controlling large capacitor banks.

In addition to protection and adjustment systems there are devices on the substations which measure power and energy flow and analyze the quality of the power system quantities. These measurements are crucial when trading within the electricity market. A distribution system owner is also responsible to provide electric energy with sufficient quality and through metering it can be verified. If analysis shows the quality being subpar then it can be corrected before it leads to bigger issues.

2.2.1 Relays in current distribution networks

The network automation units such as relays and specialized metering devises are responsible for gathering and sending data from individual network components which is then used to determine the network state. The safe and reliable operation of today's electricity distribution networks rely heavily on this data.

Numerical or intelligent electrical devices (IED) are able to communicate and send metering data through built-in communication interface. When a fault occurs the relay produces a time stamped alarm which is transmitted to the control center into the DMS. Additionally many modern relays perform secondary tasks such as power quality and network analyzing functions. While these relays form the backbone of the network relay protection there are still some older electromechanical relays in use. Older relays are very limited devices and as such they possess no extra features outside their specific function. The operating life of a substation automation device is usually tens of years and as a result there are still considerable amount of these older protection relays in use.

2.2.2 Communication

The communication link has played a critical role in the real time operation of the power system since the dawn of the substation automation systems as early as 1930’s. It provides a remote access to substation automation devices enabling centralized operation of electricity distribution network. While some systems are local and don't require communication, generally at least information about the state of the device is sent to the network control center First distribution automation systems were installed in the 1960’s. Early systems were able to provide status and control for a few points via telephone-switching based systems. As technology shifted into digital era the bandwidths of communication systems rose rapidly and ever greater variety of remote links became commercially available.

In distribution automation, the communication systems have been used for wide variety of applications for decades. The lifetime of systems is very long compared to normal IT infrastructure. In some cases, the systems installed in the 1970s are still in use today. The nature of DAS is that it undergoes a constant improvement cycle. To meet the constantly increasing requirements of the quality of the service, new systems are being developed and installed along the older systems. This often has the side-effect that many different media is being used to transmit signals ranging from copper circuits, radio, microwave, optical fibers and satellite communication. In addition the history weighs heavily on the communication scheme as often the communication protocols must extend, replace, support or include existing media and embed them into general communication architecture. (Northcote-Green, et al., 2007)

Different advantages and drawbacks are encountered among the different communication options available. The best communication option is usually determined by the application in question, as it depends on many different situation specific factors.

The variety of different communication options DA has to deal with is illustrated in the Figure 2-3.

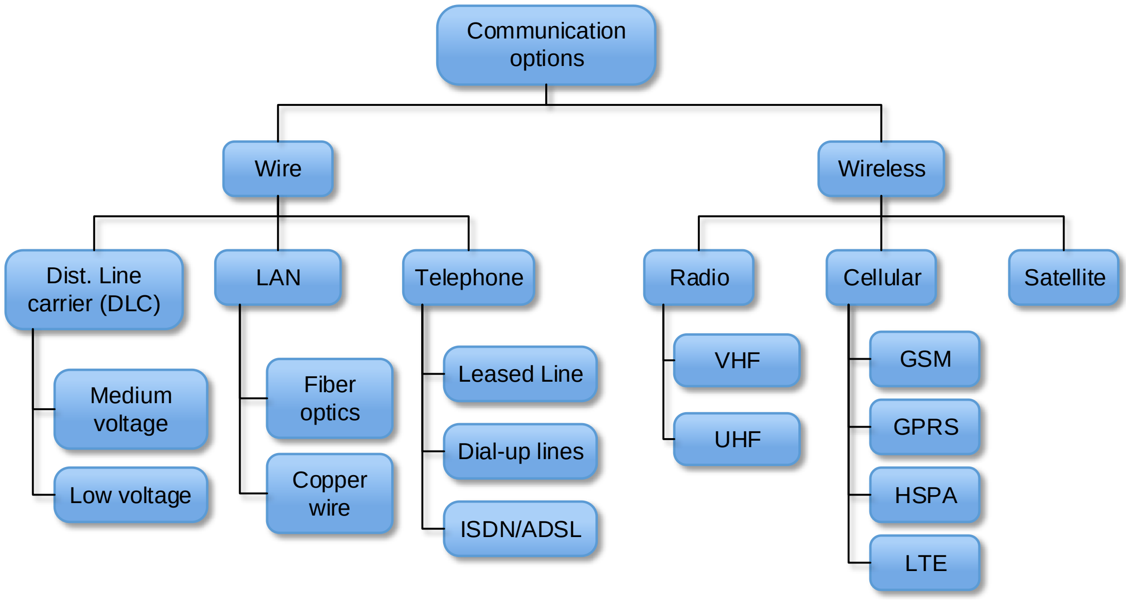


Figure 2-3 *Different communication options available for substation automation remote control. (Northcote-Green, et al., 2007)*

The communication can be done with either wired or wireless solutions. Systems relying on wires has been considered more robust, safe and reliable. Wireless solutions have usually been used in remote areas with only few devices where wiring would have been too expensive.

2.3 Maintenance strategies

Many of the current power grid components are coming to the end of their estimated life span. While the electric grid has been ageing the demand for electricity has constantly increased. All this has placed the maintaining of grid components and maintenance strategies in the spotlight in electric companies. The main focus of the maintenance is to avoid interruptions in power supply and to minimize the total costs (investments, interruptions, usage and the maintenance) of the network. The maintenance can be divided into two main categories: corrective and preventative maintenance. All of the major maintenance strategies used in today's industry are presented in Figure 2-4 (Lakervi, et al., 2008)

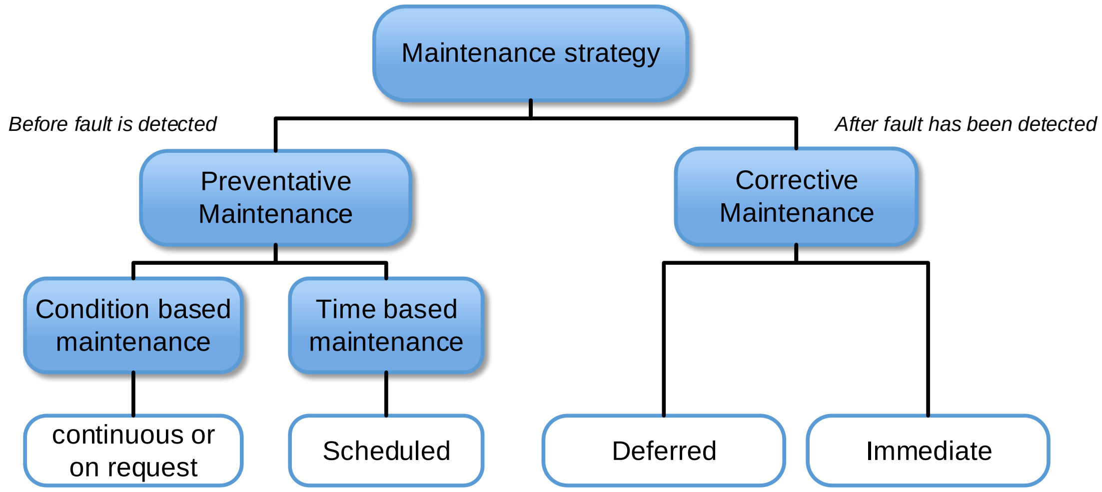


Figure 2-4 *Different maintenance strategies. (Lakervi, et al., 2008)*

Only a few decades ago the corrective maintenance as only strategy was feasible way to keep the grid in working condition and is still being used in cases where occurring faults cause only minor harm. Maintaining major grid components and whole systems this way is not practical anymore. Trend is to be able to predict and prevent upcoming component failures thus ensuring continuous operation. (Cadick, 1999)

2.3.1 Corrective maintenance

Corrective maintenance aims to repair the grid components after they have been damaged, meaning that the fault and an interruption has already occurred in the network. This of course cannot be used as main maintenance method today because it would lead to too frequent power outages. It is however used in some situations where preventative maintenance would be either too expensive or impossible due to the nature of the possible fault in the component. For example one cannot fully protect overhead power lines from trees falling on the and in these cases corrective maintenance is used. This is why corrective maintenance will always be part of the maintenance management because it is impossible to predict all upcoming failures. (Lakervi, et al., 2008)

2.3.2 Preventative maintenance

The importance of diagnostics and preventative condition management is getting higher in all areas of the electric power industry. The quality requirements for electricity and official oversight is steering the electric companies to minimize the amount of interruptions in power generation, transmission and distribution. The reliability of the power grid must be ensured but on the other hand the companies want to cut down on unnecessary maintenance work and focus it only where it is most needed.

The aim of the preventative maintenance is to detect defects in the grid components before they cause actual problems and fix them. This can be done by a simple time based maintenance (TBM) strategy where every component has a scheduled maintenance and it is done when it is due. This however leads to unnecessary servicing of some fully working components and is considered waste of resources. While time based maintenance achieves the goal of preventative maintenance, more advanced strategies have been developed, namely condition based maintenance (CBM). In CBM the actual condition of the component is diagnosed and the decision of servicing is done when need arises. The maintenance is performed after one or more indicators show that equipment is going to fail or that equipment performance is deteriorating. The Condition based maintenance is a good balance between efficient use of resources and maintaining components. The downside is that CBM method needs that condition information which in some cases might be impossible to get or getting it requires some expensive measuring devices to be installed. Therefore it can't be used in every situation.

The reliability based maintenance (RBM) strategy is used to determine the best way to maintain individual components on the grid. The idea of RBM is to assess the fault probability of individual component and to optimize the maintenance based on how critical the fault would be if one were to happen and how costly maintaining the component is. This can be seen in the Figure 2-5.

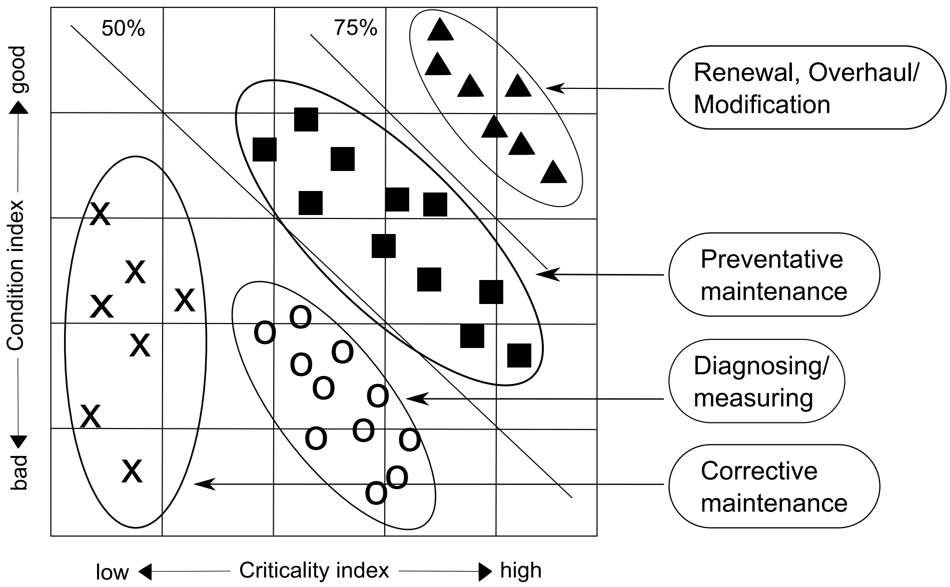


Figure 2-5 *Reliability based maintenance strategy. (Lakervi, et al., 2008)*

In RBM strategy every component is evaluated on how critical the component is for the grid and what is its actual condition. For example if components failure would lead to a minor problem in the grid the corrective maintenance becomes very feasible strategy. On the other hand the component might be crucial for the grid and in these cases its condition could be closely monitored via periodical inspections or continuous measurements. All of the more advanced maintenance strategies rely on the quality of the condition monitoring. (Lakervi, et al., 2008)

2.3.3 Condition monitoring

The Condition monitoring provides data on components state which is then used to determine the need for servicing. Ideally there would be comprehensive and accurate data of all the components in the system. (Aro, et al., 2003) The current level of technology would allow us to install devices which provide continuous condition information on every single component of the system, including the power lines. This would however cost so much that savings through maintenance optimization would never pay back the investment in measuring devices. Therefore in reality it is impossible to get the condition information from all of the grid components.

In practice the condition data is gathered by hand through periodical inspections of the network and for some of the most critical components by integrated measuring devices. However the advancements in the fields of computing power and information technology might present us a third option. The partial goal of this thesis is to look into the possibility of using already present process data, disturbance records and easily obtainable online data to make assessments on the grid components condition and state. This can be done only by storing historical data and developing data mining algorithms which provide us some additional information of the networks state. The following chapters will go into more detail of this manner of approach.

2.4 On the way to the Smart Grids

For more than 100 years, the basic structure of the electrical power grid has remained the same. Practical experiences have shown that the hierarchical, mostly manually controlled grid of the 20th Century is not suited for the needs of the modern world. Electric power distribution systems will be going through a profound changes driven by number of needs. There is the need for environmentally sustainable distributed energy resources and general energy conservation. Aging infrastructure sets demands for better grid reliability while at the same time there is the need for improved operational efficiencies. The changes required are significant for the electricity distribution systems, but can be achieved by adding automation to already existing systems and thus creating the “smarter electric grid”. The smart grid will be a modern electric power grid infrastructure for enhanced efficiency and reliability through automated control, high power converters and modern communications infrastructure sensing and metering technologies. The Figure 2-6 presents one possible structure of a smart grid system.

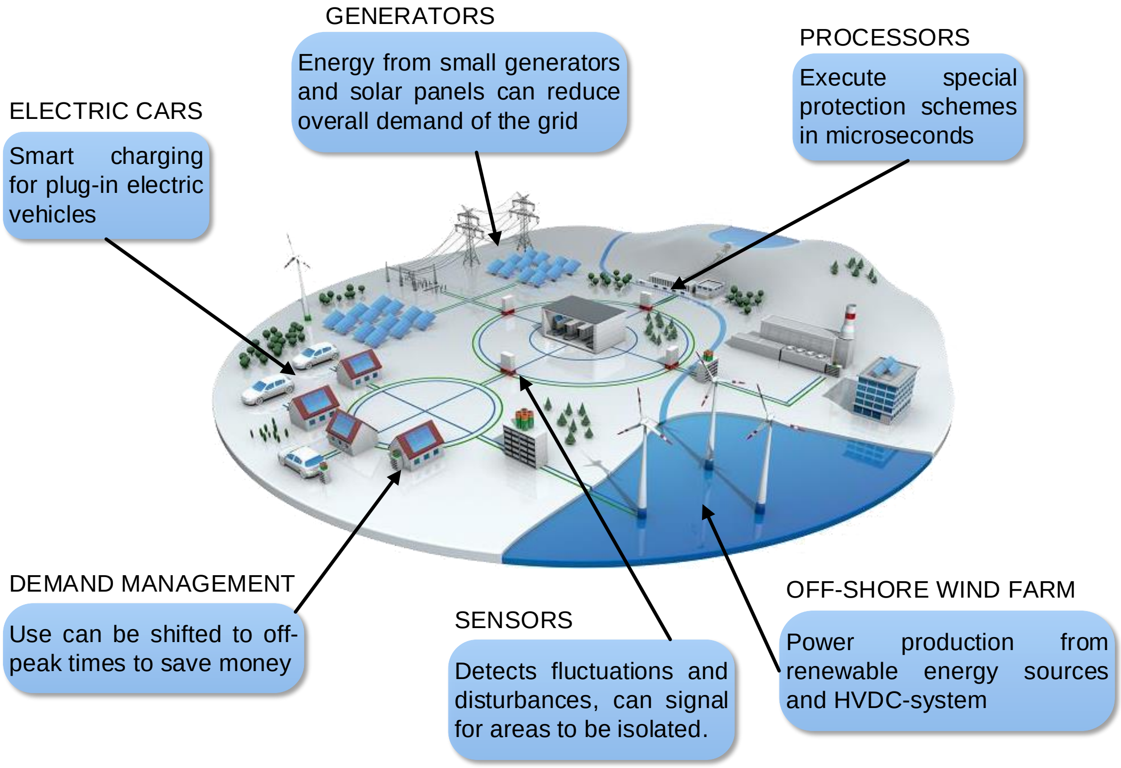


Figure 2-6 *Future vision of smart grids and possibilities it creates for network management.*

It's safe to say that these needs and changes present the power industry with the biggest challenge it has ever faced. However the changes will not happen overnight, but instead "naturally" as the ageing grid is being renewed.

There are three main types of on-going industry changes. The first is an organizational change. The electric supply has become competitive so that customers are now free to choose providers where it is cheapest. This has the diminishing effect for the role of the regional grid operator and opens up the market for competition and development.

The second driver of changes has been the question of the evident environmental issues and sustainable energy sources. Strive for cleaner renewable energy will lead eventually to decentralized generation.

The third driver being technological advancements like small scale distributed generation becoming cost-effective. The development of sensing and actuation technologies enable private customers respond to system conditions and prices of electricity making decentralized generation viable. Also improvements in distributed switching technologies for both transmission and distributions systems are driving the change. All this technology generates huge amounts of raw data which needs to be processed. This presents challenges for future information and communication technologies. For systems to work correctly all this data has to be managed and important bits to be found. The development of data mining becomes crucial.

3 BIG DATA

Big data is one of the most interesting and influential phenomena of today's world. Together with cloud services, such as storage and computing, it will play a major role in the upcoming revolution of the IT infrastructure and data handling. Mastering the big data processing could potentially lead to major cost savings with a minor additional investment into actual processes.

3.1 What is Big Data

There is no one comprehensive definition for Big Data as it varies depending who is describing it and in which context. However at general level almost every definition of big data concept boils down to huge and increasing data masses and the process of analyzing that data.

Already the Internet provides an easy access to varying databases and cost effective way to connect devices over long distances. These internet databases are created from the huge amount of data published by common internet users. In addition to this, technical development has led to an increased metering and sensor data to be gathered. Weather stations, surveillance cameras, smart phone sensors, routers and other systems like these are just example of how much and how diverse data is available. All of this data has naturally some kind of application already, but most of it is left unexploited and therefore most of the valuable information contained within it is lost.

All in all the big data concept tries to answer how to process increasing amounts of greatly varying data. It must cover how to transfer, store, combine if needed, versatilely analyze and most of all utilize all the data on hand. Instead of giving tools to make real time analysis and decision making the big data concept aims to help proactive planning, which is based on gathering, combining and innovative mathematical analysis of the history data that has been gathered over a period of time. (Salo, 2013), (Hurwitz, et al., 2013)

3.2 Three V's to define the Big Data

The big data as a concept is relatively new trend and it is still constantly evolving. People are unsure how to best describe it and its main aspects and opportunities. Most definitions of big data focus on the size of data in storage. However there are other equally important attributes that cannot be overlooked, such as data variety and velocity. The union of these three V's (volume, variety and velocity) is presented in the Figure 3-1.

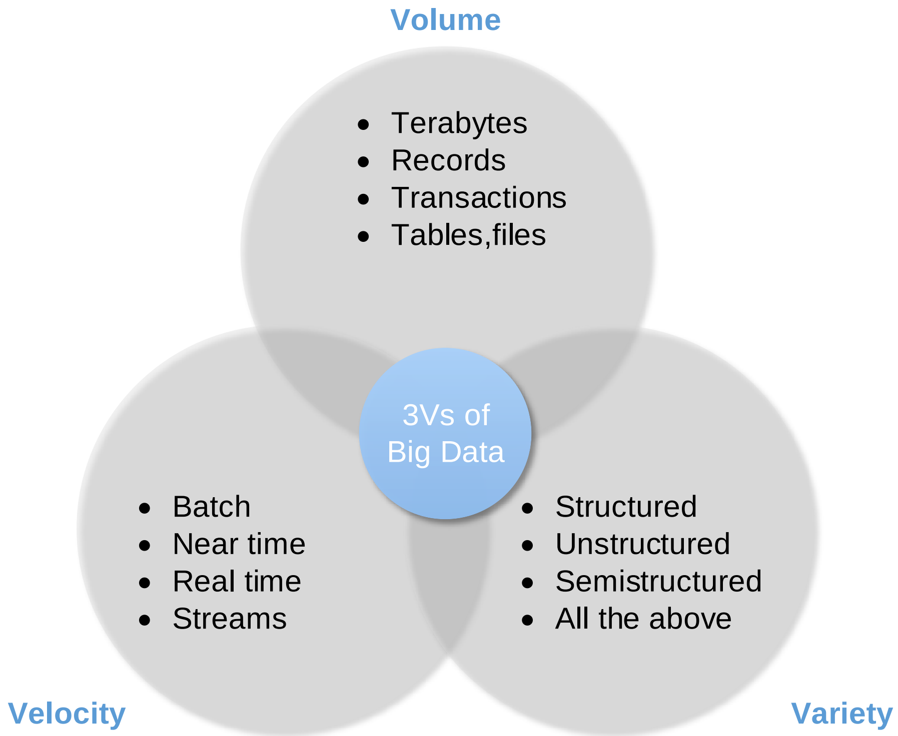


Figure 3-1 *Definition of Big Data concept through three Vs theory: Volume, Velocity and Variety.*

The volume of the data was a big problem in the early 2000s. The data masses started skyrocketing and the storage and CPU technologies were overwhelmed by the data flow. Now, a decade later, the IT infrastructure has become increasingly available and affordable, which has led to increase in devices able to generate and storage digital data. The scalability issues have been overcome even though the data volume has exponentially increased. Current estimations of the data generated daily revolve around 2.3 trillion gigabytes. The benefits gained from the ability to process large amounts of information is the main attraction of big data analytics. Having more data is considered to be better than having better models, as even simple mathematical approaches can produce excellent results given large amounts of data. Therefore it is obvious that the data volume is the main attribute of big data.

Variety refers to steep increase in data types algorithms need to handle. Conventionally we have been used to store and process data from structured sources like spreadsheets, databases and lists. Now the data is coming from great variety of sources such as e-mails, photos, videos, text-files and audio recordings. The variety of this unstructured data poses a serious challenge for actual big data applications. Only on rare occasions the data presents itself in a perfectly ordered form and ready for processing. Big data mining begins with the extraction of ordered meaning from unstructured data for humans or applications processing it further. When moving source data to processing application some information is lost as parts of source data is being discarded. Potential information loss is another major side-effect that comes from data variety.

The velocity aspect of big data deals with the pace at which the data flows in from sources like production processes, robotic manufacturing machines, measuring devices and human interactions with computers. It describes the frequency of data generation or the frequency of data delivery. The real time nature of the data means that it has to be captured and stored right away and failing to do that leads to its loss. This is not a problem on its own, but with big data the volume and variety make it challenging. In addition the analytics that go with streaming data have to process and take action in real time. (Normandeau, 2013), (Russom, 2011)

3.3 Different ways Big Data can create value

In the last few decades we have seen a significant increase in productivity and it is mainly thanks to widespread adoption of IT infrastructures as means of managing processes and businesses. The use of big data applications will be the next significant way to further increase productivity. It might even become the key way for companies to outperform their rivals. We will look into few ways how a big data can create value. It should be noted though that not all sectors of industry can benefit from all of them and some sectors are naturally poised for greater gains and that there are of course other ways than just these mentioned, but these are some of the most likely to be used in the power industry sector.

Creating transparency: Making the data a company holds and gathers more easily accessible between all of the parties who might benefit from it as this can create significant value for the company. For example in electricity distribution, integrating data from grid control center, weather reports and sub-contractors conducting maintenance could significantly cut down interruption times as field crews could work more autonomously.

Innovating new business models, products, and services: Big data enables companies to create new products and services. In addition it helps companies to understand their customer segment better and through that knowledge to improve their ways to improve efficiency and effectiveness, enabling organizations both to do more with less and to produce higher-quality products.

Replacing/supporting human decision making with automated systems: Complex analytics can significantly improve decision making by finding valuable insights that would otherwise remain hidden. This would help to minimize risks as human's limit to handle big quantities of continuous information flows is limited at best. This kind of analytics would have applications from common retailers to process industry. In some cases such as electric grid control the decisions might not be necessarily automated but instead supported by big data technologies and techniques.

Discovering unforeseen needs and exposing variability: Organizations can collect more accurate and detailed performance data from their processes. Which in turn can be used to assess the natural variability in performance in those processes. Understanding the roots of the phenomena better can be profited from by improving overall performance of these processes. (Manyika;ym., 2011)

3.4 Techniques for analyzing Big Data

The techniques for analyzing big data can be adopted, developed and up-scaled from wide variety of techniques in use today to manipulate, analyze and visualize the current databases. These techniques combine the knowledge from several fields such as statistics, computer science and applied mathematics. This nature of big data analytics means that interdisciplinary expertise is required to derive value from big data masses. The research is being carried out continuously to develop new techniques to analyze new combinations of data. The techniques available are too numerous to go through of them all in the scope of this thesis, but the most suitable ones for electric power industry are looked into.

Data mining itself is an integral part of many different big data analyzing methods. The aim is to extract patterns from large datasets by combining methods from statistics and machine learning with database management.

Association rule learning is a database management technique for discovering meaningful relationships among the variables in large databases. This approach of filtering the data is based on making relations between events which seemingly don't related to each other. An easily understandable example of this kind of approach would be a supermarket customer who after buying product A is likely to buy product B too. The similar associations could be adopted into electricity distribution and used to tie together events in the network.

Close relative to Association rule learning is the machine learning. It is a process in which algorithms are created so, that they evolve based on the behaviors of empirical data. Main focus of machine learning is to enable computer systems to recognize complex patterns and make seemingly intelligent decisions based on data on hand.

Data fusion and data integration presents a set of techniques that combine and analyze data from multiple sources to provide additional information. The goal is to provide insights that are more accurate or even undetectable if the datasets were analyzed one at the time. One example of an application is the metering data collected from smart meters combined with the real-time process data from relays to provide better perspective on the performance of a complex distributed system.

Predictive modeling draws from a set of techniques in which mathematical model is created to best predict the probability of an outcome. One way to achieve this is to choose a suitable model and analyze how the value of the dependent variable changes when one or more independent variables is modified. Predictive modeling could be used for example to determine which manufacturing parameters have greatest impact on customer satisfaction.

One of the most crucial tasks of the big data analysis is the visualization of the results produced by analyzing techniques. There is a distinct limit at how much of data human beings can perceive effectively at one instance. Presenting the information so that people can consume it effectively becomes a key factor in successful big data analyzing. (Manyika;ym., 2011), (Rajaraman;ym., 2013)

3.5 Big Data in next-generation utility systems

Like on every other field of industry, for electricity distribution the big data presents a very powerful way to improve effectiveness. While electrical utilities possess a great deal of structured data collected from their network measuring systems they also have to deal with unstructured data sources such as maps, photos and utility's history data. Turning this data into more useable form for big data mining can be quite a challenge. The nature of big data in power systems varies depending where the analyzing is done. In the current systems the amount of raw signal data decreases when moving into higher levels of the system but at the same time the amount of data sources becomes greater. This is illustrated in the Figure 3-2.

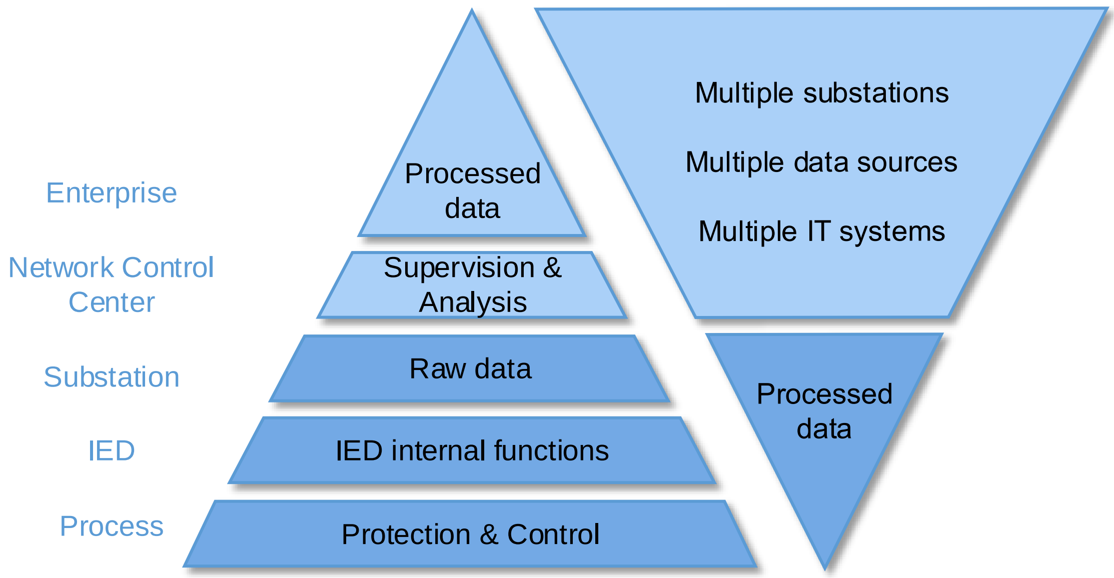


Figure 3-2 *Amount of data sources available at different levels of power system.*

It is clear that at every level of operation the big data problem comes into question. The future trends of electrical vehicle usage and integrating grid systems increasingly to electricity markets mean that utilities have to process an increasing amounts of complex events. The dawn of distributed energy resources further complicates the data management challenge. However it is clear that a broad variety of data available creates opportunities to improve operations and decision-making on many different systems and business processes.

The Figure 3-3 lists some sources for big data mining in an electric utility company. It also illustrates how the findings of big data is continuously used to improve the company's operation. Big data from these systems can be used to improve planning for outages, predicting equipment failures, responding to weather events and optimizing the flow of energy across the network.

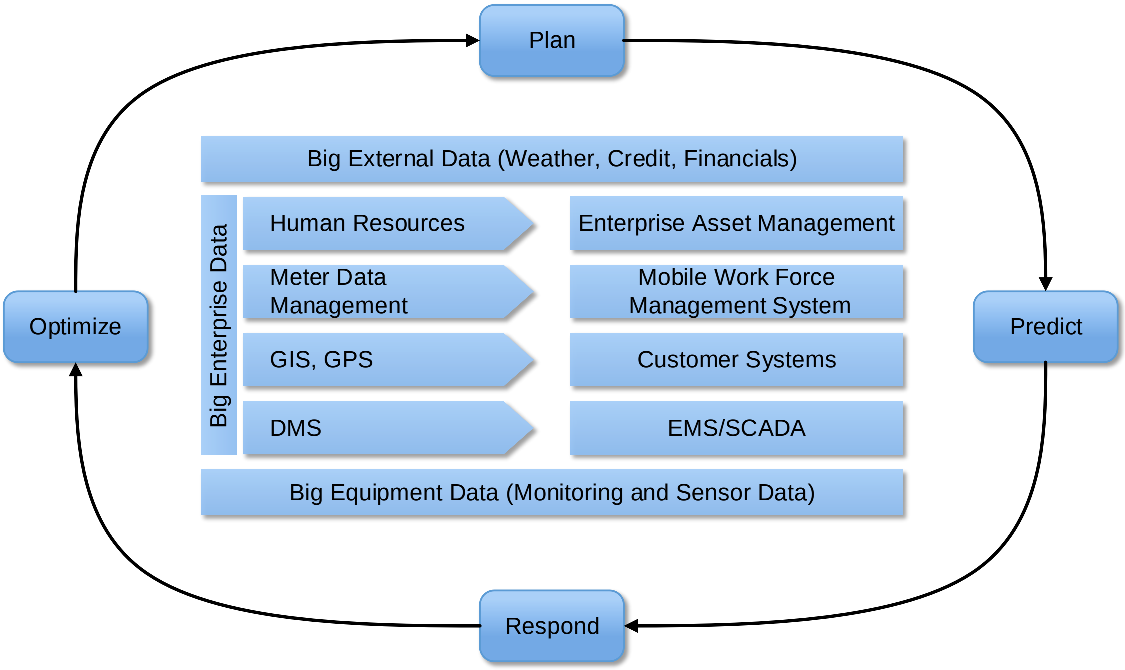


Figure 3-3 *Process how big data can benefit a next-generation utility, and some of the big data sources available listed.*

The autumn of 2013 came with two more severe storms than what we are normally used to here at northern Europe. Both of them lead to interruptions of which the longest ones lasted more than three days. Big data offers electric utilities ways to react more efficiently to infrequent events. The analysis of freely accessible weather data can help shape utility's response to fast-changing weather conditions. Whereas the history data from previous years combined with asset health and network reliability data can be used to better prepare for potentially disruptive events. A next generation solution would be the use of flyover data from RC-drones to map the fallen trees, downed lines and flood areas and more to optimize restoration. Furthermore, decisions that normally required multiple skilled workers could be automated. The resulting flood of unstructured data from drone usage means that big data techniques must be used.

The effective handling of big data can be used to produce more accurate forecasts about hourly and daily customer-level loads. The analysis of smart meter data combined with customer profiling can be used for customer load optimization. Shifting energy consumption from peak-hours to lower priced segments of the day would benefit both the customers and the distribution company. (Bane, et al., 2013), (Srikanth, 2013)

The given examples are just few possibilities and many more like them can be found with little effort. This thesis revolves mainly around predictive maintenance and how to use the maintenance resources more effectively. Predictive solutions such as asset health estimation use real time-analytics to detect potential or developing situations. Real-time measurements and operating history can be used to flag assets that are trending toward failure so that action can be taken before actual failures. In the practical part of this thesis we will look ways how the big data analysis could be used to avoid doing unnecessary testing of breakers. In Finland the electric distribution companies are required by the Finnish Safety and Chemicals Agency (Tukes) to conduct periodical proofing tests to demonstrate that their grid protection is in working order. (Tukes, 2011) One possible way to use history data is to create automated reports on every installed circuit breaker which have had to operate during the three year period.

4 DATA UNIFORMITY FOR AUTOMATED ANALYZING

The industry has developed standardization system to help address the incompatibility issues caused by vendor specific implementations. Before the universal standards were adopted, the customers of substation automation system vendors were almost completely dependent on their chosen system provider. It was nearly impossible to combine two different systems from two different vendors as their communication and operating principles differed too much from each other. At present the international standards regarding substation automation are freely accessible so that every system developer can benefit from them and make sure that their systems work in unison with other supplier's systems and that their systems meet the current requirements.

From the perspective of big data mining to be able to rely on standardized databases can help a lot. The analyzing algorithms could be developed to assess and detect the right signals and data structures for higher level of automation during commissioning.

It also makes the using of same data mining applications in different systems more efficient by minimizing the additional database and signal engineering required.

In this thesis two types of big data sources are used: The real-time process data from SCADA system and disturbance recordings from IEDs. The process data is covered by IEC 61850 standard while the disturbance recordings are covered by IEEE Power & Energy standard C37.111.

4.1 Disturbance recordings

4.1.1 General

The concept of disturbance/fault recording is not a new invention. Recording devices have existed for many years stating back to the first ink chart recorders. These old recorders were analog systems and highly specialized in monitoring one single task. Reading and analyzing of the recordings was done manually and thus it was too costly to incorporate disturbance recordings into electricity distribution grid protection scheme. Modern digital equipment however has the capability to monitor a large number of analog and binary signals which can be collected to a centralized location via remote communication links. The analog signals such as voltage and currents of the transmission lines are used as primary source of data to determine fault type and duration. The digital signals such as circuit breaker position, relay output contacts and lockouts are added to recordings to give analysts a better understanding of the event. Possibility to add and sync digital inputs with the main events as well as increased capabilities of disturbance recording devices and IEDs with disturbance recording capability allow more thorough analysis of the power system disturbances. Time stamping and time synchronizing of the records is a necessary task of today's disturbance recorders and it is made possible by using GPS clock signal as synchronizing value.

As automatic analyzing of disturbance recordings is one of the two main goals of this thesis, some of the main attributes of the disturbance recordings are discussed in this chapter with emphasis on the COMTRADE standard itself. This is done to give the reader a better understanding of the information available when using disturbance recordings as the main source of measurement data.

4.1.2 Types of disturbances of interest

The types of interesting disturbances from the grid protection and analysis viewpoint are generally divided into four main categories by the event duration. Transients are very short in duration and typically are cleared by the operation of protection equipment such as circuit breaker. These events last not more than fractions of a second, but they provide a lot of measuring data. The collected data can be used to analyze if the protection operated correctly or calculate the fault location. High-speed recording is essential to capture the individual samples of the voltage and current waveforms with enough resolution to display power system faults and transients. The recordings of this type usually consist of only few seconds as it basically covers the whole event and it is a way to keep file sizes as small as possible.

Almost in the same event group as the transient is the short term disturbances. These generally include all other time-delayed fault clearing and reclosing events where system operation is not affected. Short term events last longer than the transients but are usually in the order of few tens of cycles.

The long term and steady state disturbances consist of events that affect the system stability such as power fluctuation, frequency variation and voltage quality problems as well as events that do not directly affect the system stability like harmonics produced by the loads and the interaction of power system components. These events can be analyzed to find the source of the problem. Low-speed recording is used to capture short term and long term events. To detect long term and steady state disturbances a continuous recording is required which differs these types of analysis from actual fault recording analyzing performed using COMTRADE files. (Strang, et al., 2006)

4.1.3 Triggering methods

Disturbance recordings are only captured when fault recorder senses an ongoing fault on the line. Recording events may be triggered by changes in measured analog values or by the change of external digital input. The triggering from digital inputs is usually straightforward as the fault recorder is simply started when binary signal changes state.

Analog triggers operate either directly from measured values or by calculated analog channels which can be tripped by any combination of triggers such as change in signal magnitude, harmonic content of the signal, rate-of-change of the signal or protection function. The signal magnitude trigger covers most of the disturbance recorder operations recorded from electricity distribution network as it observes if the measured signal either exceeds or falls below the set-point. A typical application of magnitude trigger is over current event of the current channel. The rate-of-change trigger works like the magnitude by observing if the signal rate-of-change differs from the threshold values. This kind of triggering is useful for example when analyzing the long term variations in power system frequency. The harmonic trigger activates when the harmonic content of the channel is out of threshold values for a specified time delay. (Strang, et al., 2006)

4.1.4 COMTRADE format

The rapid development of digital devices for fault and transient data recording and testing generated the need for standard format for the exchange of data. The COMTRADE standard defines a common format for files containing transient waveforms such as disturbance recordings and simulated event data collected from power systems or power system models. Its main goal is to provide an easily interpretable form for use in data exchange needed for systems operating with the interchange of various types of fault, test and simulation data. The standard does not define any means to compress or encode the data as it just covers the files stored in physical media such as hard drives. (Ryan, et al., 2005)

4.1.4.1 COMTRADE files

Each COMTRADE record consists of up to four files associated with it. The complete set is made up from header file (.HDR), configuration file (.CFG), the measurement data file (.DAT) and information file (.INF) of which the configuration and data files are only required while header and information files are optional.

The header file is meant to be read by a human user and it has no predefined form. The creator of the file can include any information and in any order he desires. The header file is meant to be used as an introduction file which would give the analyst more background information about the event recorded. The format gives some examples as what the header file might contain: The description of the power system prior to disturbance, length of the faulted line or parameters of the system behind the nodes where the data was recorded, etc. The header file is not intended to be manipulated by the application program.

The configuration file is defined to be an ASCII text file which has to be in format defined by the COMTRADE standard. The file is readable by using any word processing program. This file is needed for a human or computer program to successfully read and interpret the values recorded into the data file so it must be included in every set of recording data. All the required data of the configuration file is listed in Table 4-1.

Table 4-1 *Setting parameters a COMTRADE configuration file is required to include.*

Station name, identification of the recording device, and COMTRADE Standard revision year.
Number and type of channels
Channel names, units, and conversion factors
Line frequency
Sample rate(s) and number of samples at each rate
Date and time of first data point
Date and time of trigger point
Data file type
Time Stamp Multiplication Factor

The following is an example of a partial configuration file taken from an actual REF615 protection relay currently in use in a substation.

```
REF615,192.168.50.32,1999
73,9A,64D
1,IL1,A,,A,0.3125,0,0,-32767,32767,80,1,P
2,IL2,B,,A,0.3125,0,0,32767,32767,80,1,P
3,IL3,C,,A,0.3125,0,0,-32767,32767,80,1,P
…
…
64,Unused BI,,,0
50
1
1600,8
30/09/2013,15:26:39.807106
30/09/2013,15:26:40.807106
BINARY
```

The configuration file tells that the recording is named after the relay REF615 which identifier is its IP address and that the COMTRADE standard revision is dated to 1999 - the older version. The second row summarizes that total of 73 measurement channels were used out of which 9 were analog and 64 digital. The configuration parameters of individual channels are listed next ending in the 64th binary channel. These parameters are described in the Figure 4-1. The last part of the configuration file lists the line frequency, sampling rate, time stamps for the first data value in the data file and for the time of the trigger point respectively. The final row describes the data file type which must be either ASCII or BINARY.

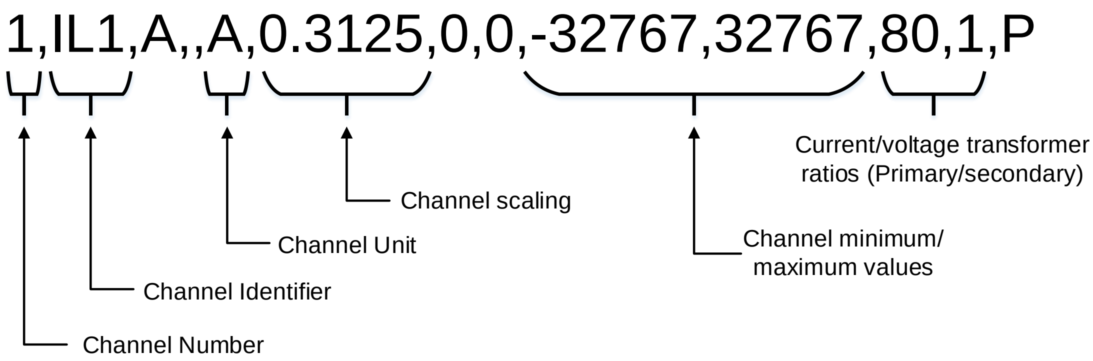


Figure 4-1 *One of the configuration (.CFG) file rows in a fault recording following according to COMTRADE standard.*

The Data file (.DAT) contains the recorded event data in either ASCII or BINARY form and conforms to the format defined in the configuration file. The data file is divided into rows and columns. One row contains one sample of every recorded channel and the number of the rows varies with the length of the recording. Each row is made up of the sample number, time stamp, and data values for each analog and digital channel. An example of data file row is presented in Figure 4-2.

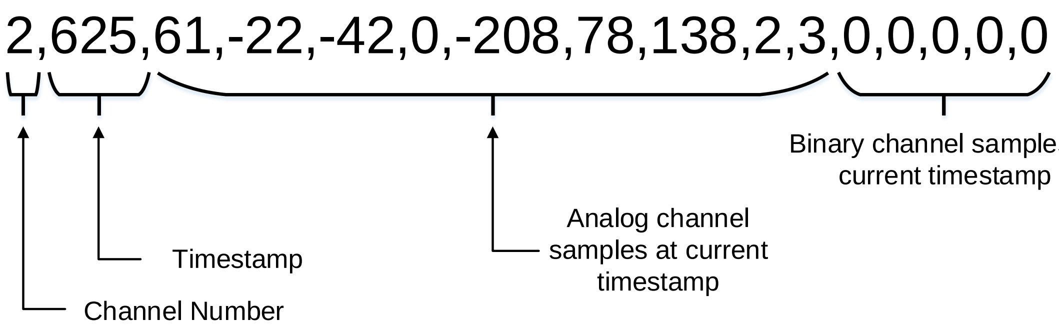


Figure 4-2 *One of the data (.DAT) file rows in a fault recording following according to COMTRADE standard.*

While the disturbance recordings can be read by using a text-editor, provided the files are in ASCII format, it is very hard to make a lot of sense of them simply by reading the instantaneous values row by row. The manual analysis of the recordings is made a lot easier using a program that is designed to interpret and visualize the recording signals. The contents of a real substation recording are plotted in Figure 4-3 using COMTRADE viewer WaveWin.

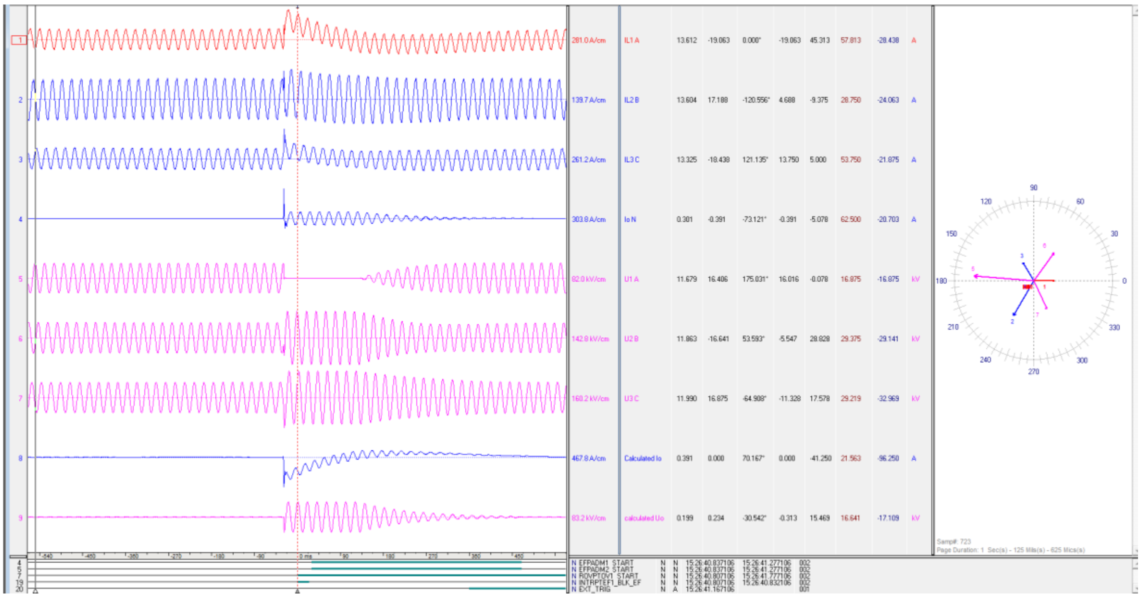


Figure 4-3 *Example of how fault recording looks in WaveWin COMTRADE viewer.*

The viewer plots all the recorded signals and some additional ones such as residual voltage and current which are purely calculated values.

The information file is one of the two optional files defined in the standard. As such it is meant to store additional information of the event recorded in the .DAT files, which could enable enhanced manipulation and analysis of the data. Data in the .INF file is in a computer-readable format consisting of public and private entries following formatting defined by the standard. Public entries contain information in a form that is meant to be able to be used by equipment or software made by more than one manufacturer. In contrast the private section is defined to contain manufacturer-specific information in a format of vendor's own choosing. In most cases this private information can only be used with a specific vendor's software or hardware. The information file can however contain multiple private sections allowing private implementations from multiple vendors in a single file. (Ryan, et al., 2005)

4.2 Developing naming convention for historian database variables

The naming of data points becomes a focus point in a wide area system which is intended to combine different data sources into a single data storage. If collected data is to be comparable between each other, then the data point names must be unique as well as recognizable. The proposed naming is used in the practical parts of this thesis and concerns every data source added to the Historian database. Consistent naming of IEDs and their data channels is essential to aid post fault analysis, providing an easy method for identifying what channel information is relevant for particular operation. The naming proposition draws heavily from already well-established standards IEC 61850 and IEC81346.

4.2.1 The signal naming concept of IEC 61850

Among the existing automation standards the IEC 61850 is unique because instead of just specifying how the bytes should be transmitted on the wire it provides a comprehensive model describing how power system devices should organize data. As a result the data is organized consistently across all types and brands of devices. (ABB, 2014)

The IEC61850 states that in the case of hierarchically structured objects of the substation or process structure, both name and description attributes for each object contain only that part which identifies the object within this level of the hierarchy. This means that full path name of a mapped process object is a concatenation of all name parts of higher hierarchy levels up to this level. The uniqueness of full names is ensured by using syntax conventions as specified in another IEC standard the IEC81346-1. In addition to the mandatory use of IEC 81346-1 standard for name syntax the 61850 standard recommends the use of whole 81346 series for derivation of functional and IED product names as technical keys. The only allowed separator between SCL names is the dot (.).

The IEC 61850 device model begins with a physical device which can be described as the device that connects to the network and is typically defined by its network address. The physical device contains the logical devices (LD) whose amount is not fixed. The logical device divides into logical nodes (LN) as each LD can consist of multiple LNs. An LN is a container for data and associated services that are logically linked to a specific process function. The name of the logical node tells what kind of signal is in question and into what kind of context the node refers to. For example there are logical nodes for metering and measuring whose names begin with letter “M”. The standard name of the logical node for measurement unit for 3-phase current is IMMXU and the postfix numbering is used to delineate between multiple measurements at the same level. Each logical node contains one or more elements of Data related to the practical function of the device the logical node is representing and is defined in the IEC61850-7-3 part of the standard. This means the data variables available vary between different logical nodes.

According to IEC 61850-7-2, signal identifications are built from the following parts:

1) User defined part identifying the logical device LD in the process (LDName).

2) (Function-related) Part to distinguish several LNs of the same class within the same IED/LD (LN-Prefix)
3) The standardized LN class name and the LN instance number, which distinguishes several LNs of the same class and prefix within the same IED/LD
4) Signal identification inside a LN consisting of data and attribute name as defined in IEC 61850-7-3 and IEC 61850-7-4

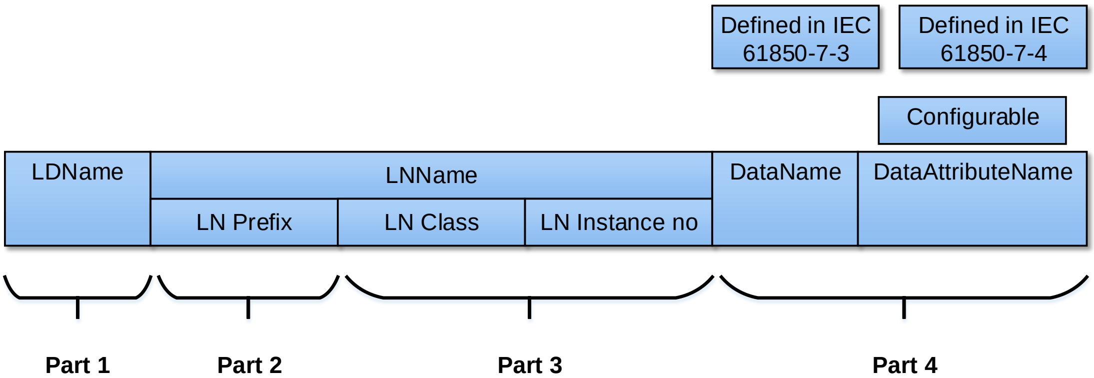


Figure 4-4 *Elements of the signal identification as defined in IEC 61850-7-2. (IEC61850, 2009)*

The LN name is formed in combination of parts 2 and 3 in the Figure 4-4 which makes the different LN instances effectively distinguishable within the same logical device of an IED. The LN class forms the body of the signal name giving it the main description as what kind of signal is the one in question and what kind of properties does it have. LN prefix is used to classify different parts of the LN classes into functional groups such as supervisory control (C), Protection (P) and sensors (S). The instance number is added in a case where two or more of the same instances of same LN name would be formed otherwise.

The DataTypeTemplates definition section of the SCL and the standardized names are defined in IEC61850-7-3 while the IEC61850-7-4 defines possible values for name paerts 3 and 4 in Figure 4-4. The name parts 1 and 2 of the Figure 4-4 have several options on how to form them and of which the standard directly mentions two of the most important ones.

In **the product related naming** the LN prefix, class and instance number attributes are fixed by the vendor of the device as can be seen in the Figure 4-5. In product related naming the project engineer is left free to choose the best name for IED and therefore IED names vary depending who has made the configuration. This makes interpreting the signal properties harder as it presents randomness to the naming scheme. IEC61850 however offers another option for naming called **function-related naming** which is more suitable to the needs of the system being built.

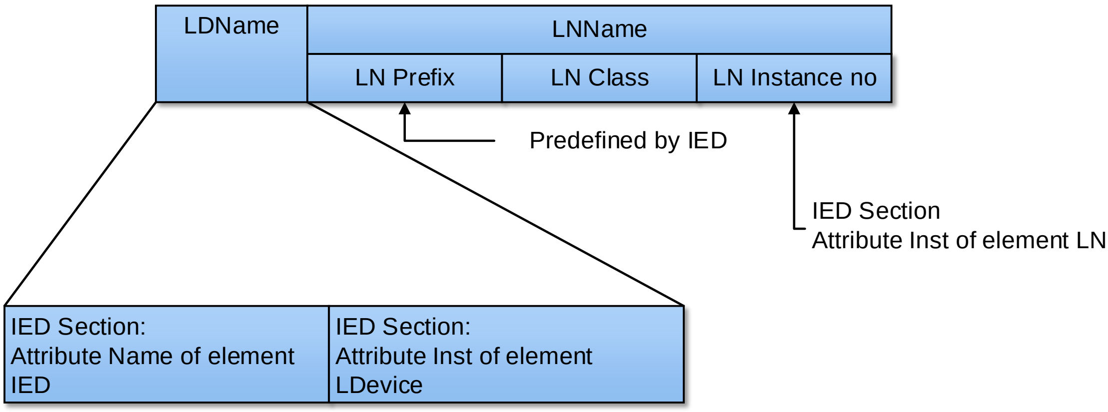


Figure 4-5 *Elements of the signal name using product naming. (IEC61850, 2009)*

In function-related naming the idea is to name IED and logical device according to its function as illustrated in the ***Figure 4-6***. The IED name is created by assembling a complete name from sub parts such as voltage level of the installed IED, bay identification and IED's function within the bay. This concept automatically guarantees uniqueness of the IED's name within the substation. There is no specification for standardized naming of voltage levels, bays or power system devices in the IEC61850 but the standard directly refers to IEC 81346 for naming of different levels and objects within the substation. The major drawback of a naming convention such as this is its bulkiness when applied to a complex substation. For example let's assume we have a substation with 110kv voltage level (E1) which has a bay (Q1) and IED named SB1. In the IED there are multiple logical devices (LDx) which hold the signals for circuit breaker operation. The complete name tag would be the concatenation of these and would produce tag of E1Q1SB1LD1CSXWI.pos.stVal. (IEC61850, 2009)

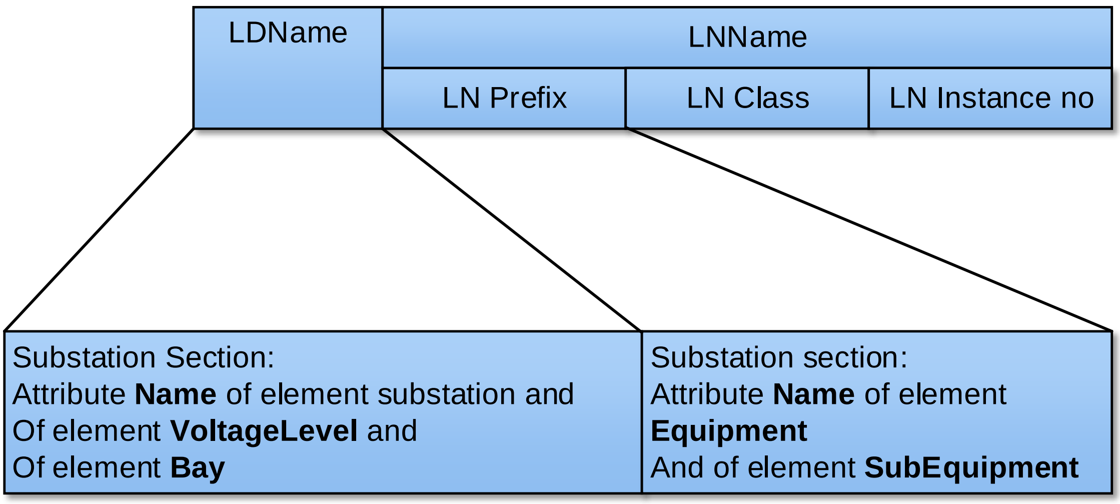


Figure 4-6 *Possible elements of the signal name using functional naming. (IEC61850, 2009)*

The data structure definition according to the IEC 61850 of a circuit breaker is illustrated in the Figure 4-7 with an example of a fully constructed name of the signal.

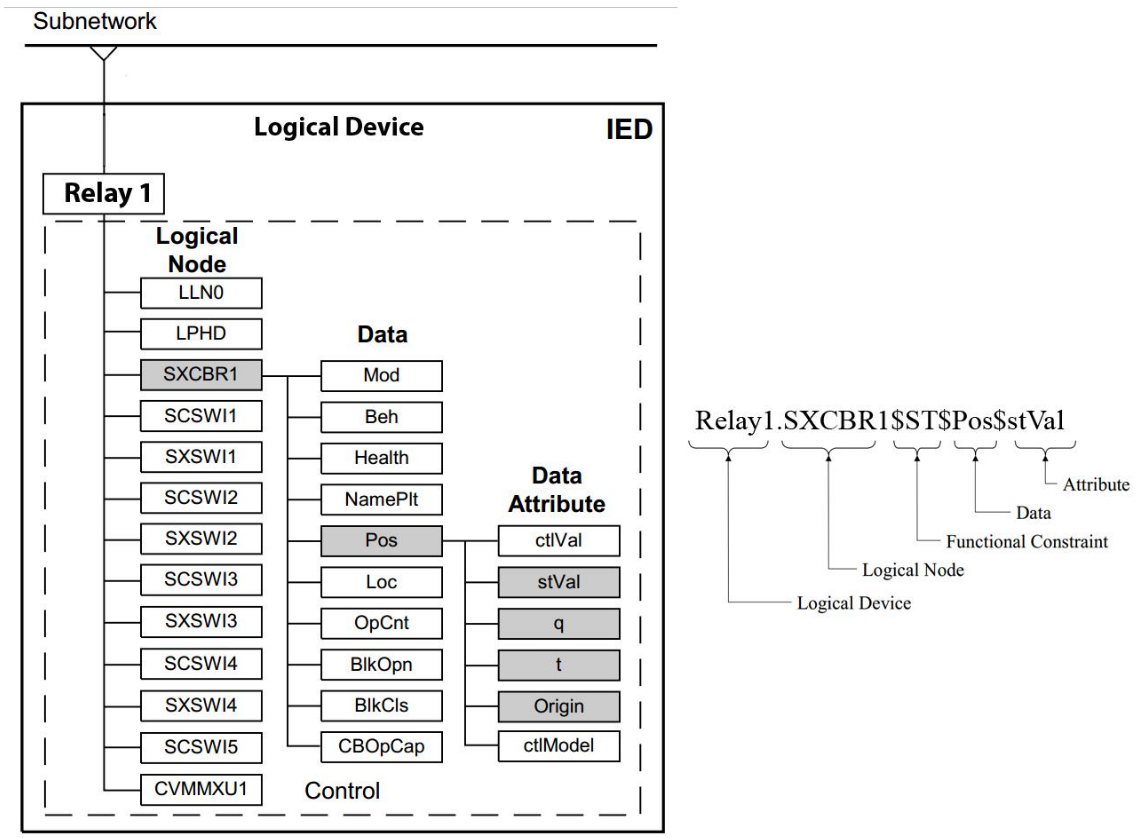


Figure 4-7 *Signal naming logic of IEC61850-7-2 standard.*

The naming scenario in Figure 4-7 represents a typical substation relay configuration scheme where one IED consists of multiple LNs. The circuit breaker is modeled here as an SXCBR1 logical node. The XCBR is defined in the standard as a logical node for a switch with circuit breaking capacity. The prefix “S” comes from LN prefix table and stands for sensors. The logical node contains variety of data such as Health for telling the external equipment health, NamePlt for external equipment name plate information, OpCnt for operations counter and BlkOpn/Cls for beaker open/close command blocking. The path highlighted shows how signal is mapped for circuit breaker position status. (ABB, 2013b) It is clear that the use of naming convention of the IEC61850 standard as a base for database variable naming has various benefits. The multi-layer hierarchical nature of the signal naming is enough to cover majority of naming scenarios. In addition as the signal’s qualities can directly be read from its name, the mapping of correct variables into analyzing functions can be automated which naturally decreases work load when commissioning the system. One advantage that cannot be overlooked is also the already wide spread usage of IEC61850 standard as process database organization convention. This means that algorithms crated following its conventions are easily portable into new systems and therefore the system under development has already a larger market sector.

4.2.2 Need for novel naming convention

While configuring an IED, the user has to come up with and enter a number of electronic designations that uniquely identify the IED and each channel within it. Current standards try to steer users in right direction by demanding mandatory information fields to be filled for every data point. However in many cases they are described as unformatted fields and it is therefore up to the user to compose and specify these fields without any given standardized practice. Such fields may include IED names, install locations, phase identifiers, voltage classes, etc. Users and utilities are often forced to develop their own naming conventions which best suit their own needs without the thought of greater picture. This in addition to the historical weight of the constant and fast paced evolution of substation automation equipment explains why we have so many different types of naming conventions in circulation today.

The above mentioned information fields are essential for successful fault and disturbance analysis as well as any attempt to make more sophisticated automatic analyzing of grid events. For example without accurate and informative names it is impossible for an automated application to associate voltage and current phases together in order to calculate fault location or a missing phase.

Currently the use of disturbance/fault recordings is mainly done by hand by the transmission grid owner. The recordings are downloaded on site from the IEDs and then manually analyzed. After some major blackouts there have been a growing interest of automatic analyzing of the events. This is especially beneficial if the event includes the tripping of many grid components. The more widely the standardized naming convention spreads the more comprehensive analyzing can be done on events that have impact on wide area. Current GPS time synchronization enables users to combine triggered data from multiple IEDs into one single record and give a wider perspective of an event. As electricity distribution systems are becoming more complex due to implementation of distributed power generation the need for a wide area view of events is becoming greater.

Nowadays we have literally thousands of analog and digital signals available to us from a single IED and means to gather and store them all. Constantly growing numbers of signals to be mapped has raised the costs of signal engineering while at the same time the hardware has gotten cheaper. Because of this the laborious task of signal configuration and documentation has become the major cost component of the substation automation projects. One of the key tasks in the development of an automated analyzing system is to try to minimize the signal engineering needed while applying analyzing functions to the customers network components. Cutting down this manual task is only achieved through relying on higher level of automation in signal mapping phase, which in turn can only work if the data is in some prefixed form such as the signal naming convention introduced next.

4.2.3 Proposed unique naming convention

Theoretically an IED can compose of one channel or, as more commonly, can contain multiple channels. The naming convention described next attempts to identify the fields required to uniquely name an installed IED and all of its possible analog and digital channels. The idea of nametag proposed is to start describing the data point location from larger scale and ending in specific signal name. The naming is based on standards IEC61850 and IEC81346. The format is composed of a sequence of fields using the comma "." character as the delimiter between consecutive fields:

Country.Vendor.StationID.VoltageLevel.Bay.Device/Relay.SignalName

Country information is pretty self-explanatory and names the country where the signal is originating.

Vendor information names the company who owns the signal source such as substation and its equipment.

StationID is the identification of the substation where the signal is originating from. For example a fictive substation called Herwood could have StationID "HER", which would instantly tell the user from which geographical area the signal has been measured.

VoltageLevel *and* ***BayName*** are the identifications inside the substation. The different feeder bays have unique names within the station to distinguish them from each other. In addition to this the different voltage levels are separated as some of the calculation functions can only be performed on certain voltage levels. These tags can be separate or merged into one. An effective way to achieve this is to use single letter expression of voltage level combined with the running numbering for consecutive bays. Some commonly used voltage level expressions are listed in the ***Table 4-2***.

Table 4-2 *Some of the most commonly used voltage levels in distribution networks and their corresponding letter expressions. (IEC81346, 2009)*

Letter expression	Voltage level
E	110 kV
J	20 kV
K	10 kV
M	1 kV

RelayType/DeviceName is used to describe the grid asset being monitored. This describes the physical device where the measurement signal was received. If the physical device in question is the relay which for example measures currents and voltages on a feeder line then it should be listed. On the other hand if the relay operates a circuit breaker or earthing switch then the breaker/switch should be named.

SignalName is used to describe and uniquely identify the signal the relay has measured. IEC61850 includes already a widely accepted way of naming substation automation related signals and it is intended to be used here without changes. SignalName is built from IEC61850 standardized names of the Logical Node, functional constraint, Data and Attribute fields like discussed in the chapter 4.2.1.

The practical use of the signal name depends on the level in which the database holding the signal information is located. It should be established that not all of the name tag parts are required on every level of operation. The analyzing can be done at the substation level at lowest where only the information from voltage level on is relevant. In the future there might be some interests in going even greater scale by taking a step further and combining the data from different electricity distribution companies as a vast database to address large scale faults and phenomena. This is where the “larger scale” parts of the name tag become essential.

5 SMART SYSTEM ANALYSER

This chapter gives a detailed description of the system components used for smart system analyzer concept. All of the parts are reviewed individually and after that they are brought together to form the automatic analyzing suite. At the end of this chapter a proof of concept testing is done by linking a SCADA system with the SSA. Using simulated signals from SCADA some simple calculations are done with the visualization of the results to test the concept in a controlled environment.

5.1 System overview

The traditional model of substation automation is based around IEDs which handle the protection and control of the substation assets on the bay level. This means that IEDs primarily operate according to the data gathered from its own bay and perform tasks in that bay only. A system that has access to all of the bays of the substation simultaneously has a distinct advantage over this presently dominating model. When an event occurs on some of the feeders it can benefit greatly from the measurement data gathered from adjacent bays and gain a greater view of the incident. This is what the smart analyzing concept aims for. The reason why systems like this haven't been used more widely before is the lack of a system wide bus-based communication protocols such as IEC 61850. In a system with bus-based communication all the data is available to all devices connected to it. It should be noted that even without communication bus there is normally a data concentrator device that has access to every signal in the substation, but mapping signals from that node into station computer is very labor intensive. The other major obstacle in utilizing station wide functions has been the computing power available. When moving into bigger scale the memory and capacity of the processors have been the limiting factor.

This kind of analyzing is mainly intended for non-critical situations, where the results are not needed in a hurry. Some examples of what kind of functions could be performed at the station level are fault prediction through component lifespan and ongoing event analysis, power quality monitoring and statistics collection. If a substation were equipped with a station computer performing these tasks some of the overlapping functions from relays could be switched off and their CPU capacity released for critical functions which concern safe operation of the network. Computer running the substation wide analysis is not in any way critical to the safe operation of the network so it can be upgraded regularly without risk of interruptions to the electricity supply. In contrast if the functions were performed in traditional way in the IEDs the upgrade process would always require proofing tests which take up several hours per IED and naturally cause significant costs. New updated functions could be tested and adopted easily which would lead to better analyzing results and constantly developing system.

The system divides into three logical entities each having its own area of responsibility. The database act as a centralized databank for all data regardless of its origin and as the core of the system where raw data is being fetched and the results of the analysis are stored. It must be able to handle data from different sources simultaneously, some of which is real-time and some periodic in nature. The database used in this concept is a historian database, which enables storing the data long times reliably. The second part is analyzing itself and it is done in a novel external calculation engine. The calculation engine never stores data more than what is needed to complete the task at hand and afterwards results are saved into the historian database. The third part of the system makes the results readable for human users. It is widget based web user interface which can be linked into the database variables and makes reading the results easier through graphs, reworked tables and various graphical presentations.

5.2 System components

5.2.1 Real time database (RTDB)

The RTDB Real Time Database is a relational database that is designed and optimized for industrial process information management and history recording. It is of a modular design with stress on high performance and reliability. Main function of RTDB is its database capabilities, but along with it comes modules which enable implementation of full process maintaining solutions like tools to implement the business logic to handle and refine database data. However in Smart System Analyzer concept only the database functions and those supporting it are used, while the actual calculation is done externally. The RTDBs responsibilities are storing and managing the process data and engineering configurations as well as providing means for data acquisition and exchange with other applications such as the calculation engine and the web UI for visualization. (ABB, 2009)

5.2.1.1 RTDB architecture

The RTDB is built on data abstraction interface called VtrinLib based on Microsoft .NET technology. It is used in various functions to support the modular structure of the RTDB which can be seen in the Figure 5-1. The abstraction converts the data structures from various data sources into unified object model, thus simplifying their use internally. According to the manual, the Data Abstraction Interface is extendable semantic free way of publishing data object for visualization, reporting, upper level public interfaces, development tools, applications and external systems.

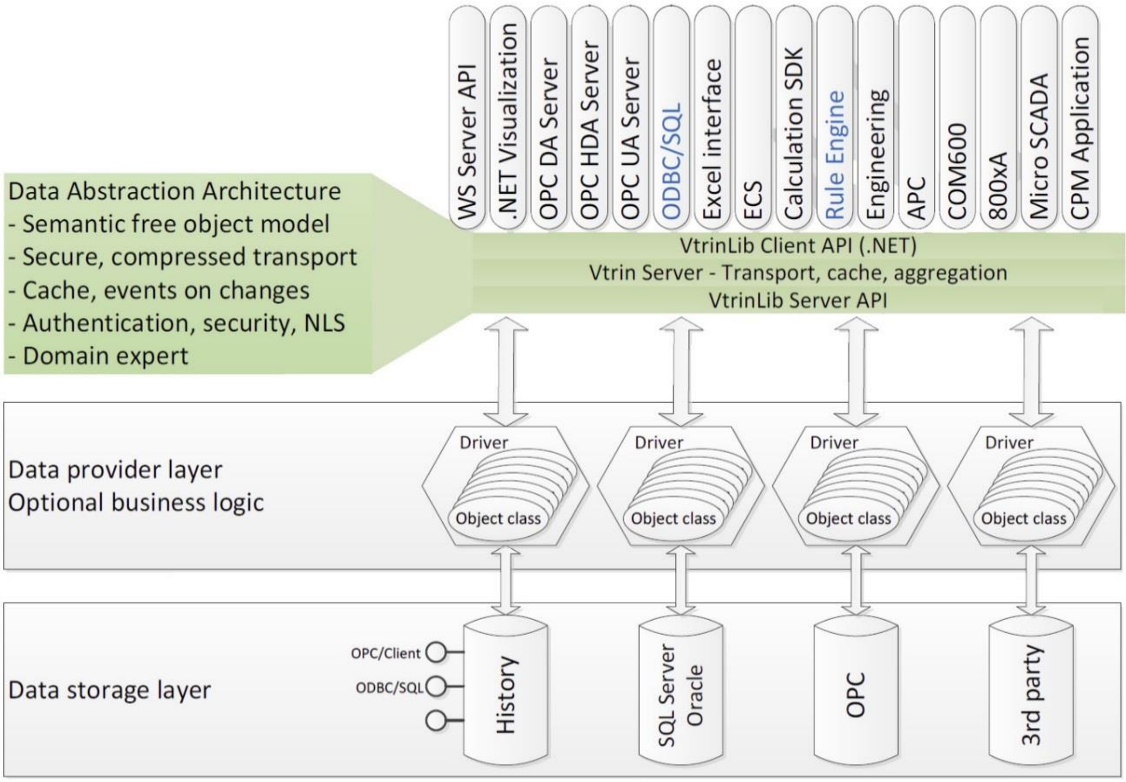


Figure 5-1 *RTDB architecture with its different layers illustrated.*

The data abstraction hides the actual internal data storage, i.e. History, OPC, or other data source, and the client-server model provides remote access to them. It makes accessing multiple data sources from one application possible and even integration of external databases into the main RTDB.

The user authentication is also managed by abstraction layer. This allows for different user classes with different privileges into the database information. In big data applications this is crucial as vendor specific systems are things of the past and 3rd party developers should be supported by giving them access to the same raw data. However the processed information might be something with limited access.

The Data Abstraction Interface enables also the implementation of system level connectivity and configuration exchange. For example with MicroSCADA the linking of process data into RTDB is simplified as variables are created automatically with intended RTDB side variable naming achieved by SCIL script definitions.

5.2.1.2 History recording

The RTDB stores process data as samples in database variables. These variables can be of the data type double (floating point), 64-bit integer, binary or string (text). The variable only accepts new values from matching data type. The measurements from primary process are recorder by pairing a sample value of the measurement with corresponding timestamp and this pair is then saved into the variable. The resolution of the timestamp is supported up to the accuracy 100ns which is considerably higher than the resolution available from measurement devices in the substation.

The history table of a variable can consist either of one or several RTDB core tables. The end user however always sees the tables as one. This allows general optimization of the database as older values with lower importance can be compressed to save storage space. On the other hand the current values can be stored in a table that is optimized for fast recording. The length of the latest history table, the length of the compressed tables and other compression related settings are defined independently for each history table.

5.2.1.3 Database interfaces

To connect the database into systems providing data the RTDB has couple of supported options:

- OPC DA/HDA/AE and UA clients for data acquisition from control systems into the database
- OPC DA/HDA servers to provide access to real-time, history and aggregated history data.
- OPC UA server to provide standard platform independent interface to the data available in the Data Abstraction Interface.
- Data Abstraction Interface VtrinLib.
- JavaScript API for web browsers to enable platform independent user interfaces.
- Web socket server API to the Data Abstraction Interface.
- Direct interfaces to the data storages such as direct SQL queries/commands.

The access for 3rd party systems is mainly provided with the use of OPC interfaces. The OPC stands for Open Platform Communications and is a standard from 1996 with name revised in 2011. It specifies the communication of real-time process data between devices to better the interoperability of different manufacturers' solutions. The OPC design is such that after the OPC server is configured for a particular hardware the clients with same configuration can access all the data the server is sending. The OPC Data Access (DA) is a subpart of the OPC standards and is focused on dealing with the real-time data while the Historical Data Access (HDA) deals with the handling of history data. Events and alarms are handled with the OPC EA standard. The JavaScript and Web socket server interfaces are related to the visualization and are more closely examined in the later parts of this chapter.

5.2.2 Calculation engine

The calculation engine developed is a solution for running non real-time applications on a PC computer working with data from station level or higher. As another substation automation building block running internal logic it is intended to supplement existing systems. The calculation environment focuses on functions which don't need to guarantee short response time i.e. are not time critical. This is what differs it from existing IED applications. Additionally this novel environment is intended to utilize historical data instead of live data.

Architecturally the calculation engine is extensive environment meaning it is built so that it allows adding new functionalities through addition of pluggable components without the need of rebuilding the whole environment. 3rd party application development is made possible but a security mechanism for prohibiting this can also be used if solution requires. The environment reuses as many existing components as possible and currently it draws from existing relay configuration tool PCM600 and RTDB for historical data storage. This allows for faster development and makes end user experience resemble as much as possible other environments for definition and execution of substation applications. The calculation environment is configured using a model where logic and data flow is graphically composed from function blocks which can be bundled into function block libraries.

5.2.2.1 Architecture overview

The main component of the calculation environment is the server which is a Windows service running continuously on a PC computer. It is not an actual server per-se but a kind of "application server" hosting user-definable applications processing data from RTDB running on the same computer. While calculation engine is primarily fed with data from RTDB it can also access OPC servers directly for special purposes. An overall view of the architecture is presented in the Figure 5-2.

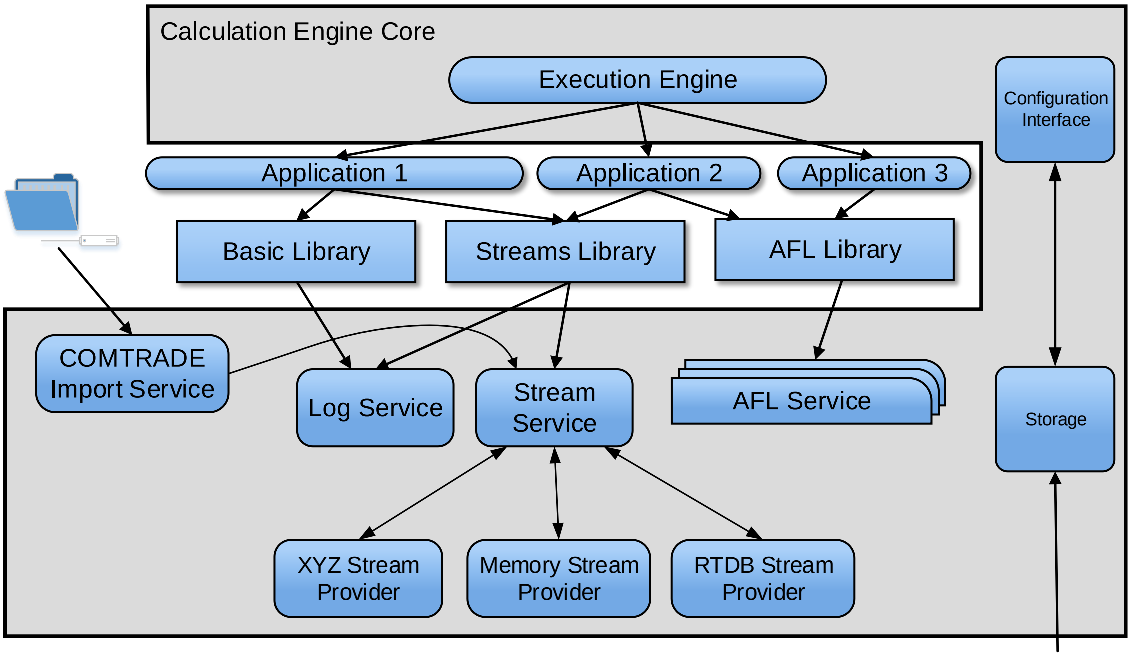


Figure 5-2 Calculation environment architecture with its major components named.

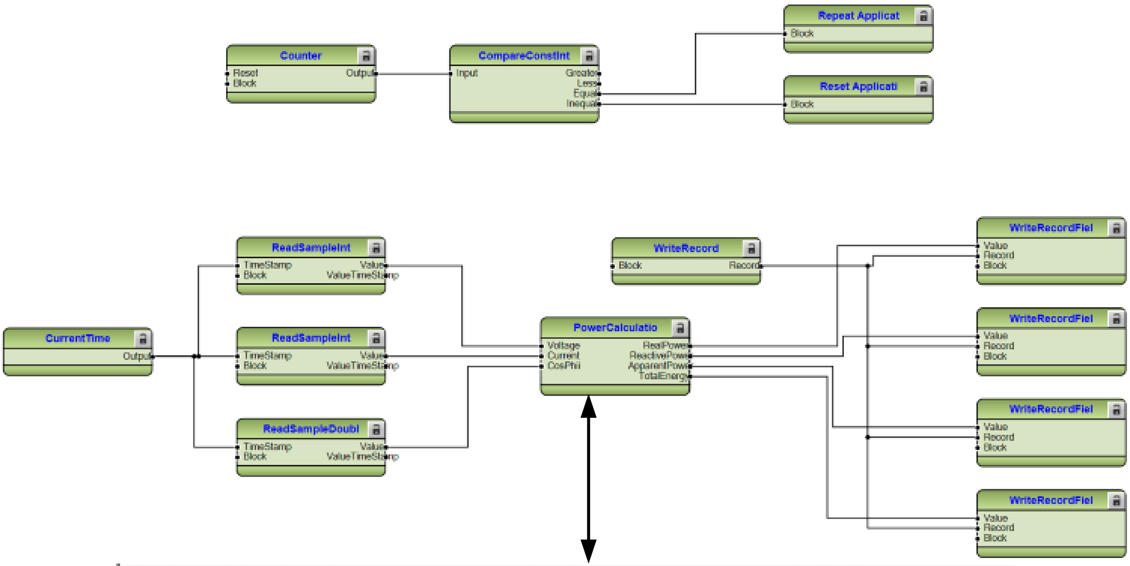
The execution engine of the calculation environment is responsible for running the applications intended for data analyzing. Its task is to keep track of application status, triggering and running etc. The applications are created and customized using PCM 600 relay configuration tool.

In addition to applications the environment hosts also various services which provide some specific functionalities such as debug logging and COMTRADE import service, which imports disturbance recordings into the RTDB whenever new file is detected in the file system. The data input and output from the applications is handled by a service called Streams Service. Along with it comes interfaces accessing any external data from application point of view. Whatever has been collected to RTDB via OPC is visible to the applications. The AFL services provide ways to run functionalities developed for ABB IEDs in the calculation environment.

The plugins are the final major part of the environment's architecture. They are designed to add flexibility and easing of deployment of the system. A plugin is a self-contained component that adds features to the environment. Function block libraries, services and stream providers are pluggable as an example.

5.2.2.2 Applications and function blocks

The main purpose of the calculation engine is of course its capability to provide ways to automatically analyze the gathered data from substation into meaningful information. This is achieved through running applications under calculation engine that process information gathered from primary devices i.e. IEDs. The calculation engine allows for multiple applications to be run simultaneously and independently from each other. The applications are made using function block logic and thus require very little knowledge of actual programming. The calculation engine has a library of generic ready-to-use function blocks each performing a simple task and by combining those more complex analyzing is made possible. The major advantage of function block logic is that it satisfies both the needs of large flexibility requirements concerning the applications and at the same time relative ease of use. Figure 5-3 shows an example how applications are formed and how blocks' operation can be tuned by changing its parameters.



Group / Parameter Name	IED Value	PC Value
PowerCalculation: 1		
Configuration parameters		
calculationMode		1
StationNameSetting		Herwood
BayNameSetting		J27
currentTransformerPrimaryRatio		4000
CurrentTransformerSecondaryRat		5
HighLimit		1500
LowLimit		200

Figure 5-3 *Example application made using calculation engine function blocks and how blocks' operation can be affected through setting parameters.*

After an application diagram like in ***Figure 5-3*** is built the execution engine calculates a proper running order for all of the blocks. The order is decided so that no block can be initiated before all of its inputs are calculated. This effectively prevents using feedback loops within same diagram. Application is triggered by blocks with triggering capabilities which range from periodic triggering to condition based such as RTDB variable value change. When triggered the application is run once through and the results are recorded as specified. If solution requires the use of looping logic such as iterative calculation, there are blocks that can execute the application again and again until the desired conditions are met. Users can add complexity by cascading applications so that after one application finishes it triggers another which takes the analyzing further. There is no limit in the amount of cascaded applications. If the default function block libraries are not enough to achieve some desired logic then users have the option of programming additional function blocks using C#. By being pluggable objects these new libraries extend the functionalities of the calculation engine while making no changes into the core of the calculation engine. This makes 3rd party solutions independent from the development of calculation engine itself.

5.2.2.3 Data handling

The major challenge that substation wide calculation faces is its mixed nature of how the data flows into the database. Only some of the devices in the substation operate in real-time while others may have unpredictable delays in their operations. Generally the primary devices and their designated IEDs operate in real-time, but the records of their signal outputs can still arrive at the history database in random order and at random intervals. This can be the result of problems in the communication line or TCP/IP protocol's characteristic of how information is transferred using multiple IP packets. These examples are situations where continuous stream of data gets mixed up on the way to the RTDB, but some of the measurements are not even intended to be recorded continuously. One example of this kind of data is the disturbance recording which is saved only during faults and covers only a few seconds. Same data transfer related problems affect the fetching of disturbance recording into the database. To combat this problem the internal clock of the system differs from the real-time with an assumed safe margin offset which ensures that no data will appear in the input stream after the stream has been processed.

We have already established that data from substation comes in various forms and each data source has its own characteristics which require different approach when executing calculations. This is why the data output and input is being handled as streams of data instead of direct value transfer. Streams hide the varying nature of substation data from end user and provides uniform interface for the streams. Because of this abstraction the person designing the applications has to only concern himself with the direction of data transfer – the output or input stream.

5.2.3 Web user interface for visualization

Probably the most critical part of big data analysis is the visualization of the results. This is the phase where raw and even calculated data becomes information for the end user. In this smart system analyzer concept the visualization is intended to be done using a newly developed web user-interface platform. It is designed to be independent from existing ABB solutions and as such it is possible to connect it to various different environments. The platform client is based on HTML5 and Websockets. The support of websockets makes greater level of interaction between browser client and server possible when handling live content such as streams of process data. The data access from client side browser to the server side data is provided by JavaScript client application programming interface (API). This API makes connecting to databases possible without the need of actually copying data. Currently it is possible to connect to SQL databases and those that support Object Linking and Embedding (OPC). The Smart System Analyzer concept uses MicroSCADA Historian as database which is accessed through data abstraction layer illustrated in the Figure 5-4.

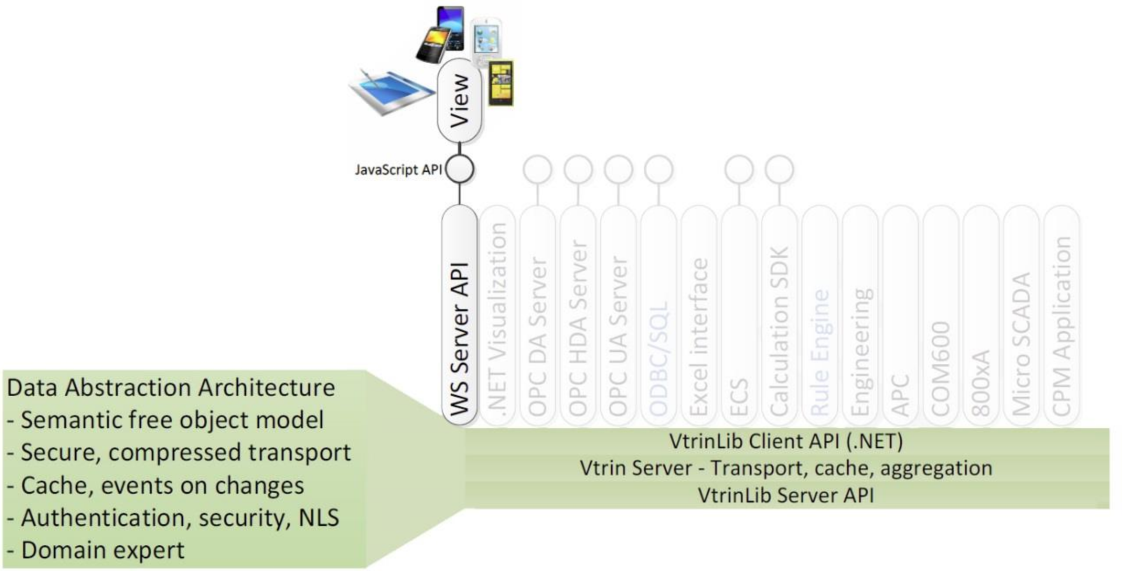


Figure 5-4 *The Web UI's (View) principle of connecting into the RTDB through data abstraction layer.*

The visual look of different UIs is made by using Dasboard Editor presented in Figure 5-5. The editor provides basic building blocks for UIs through widget library. As default the library includes various trend charts, lists, gauges and pie charts. Along with the premade objects comes base widgets with examples and tutorials on how to implement one's own solutions. As the default library grows the need for programming decreases when creating new UIs.

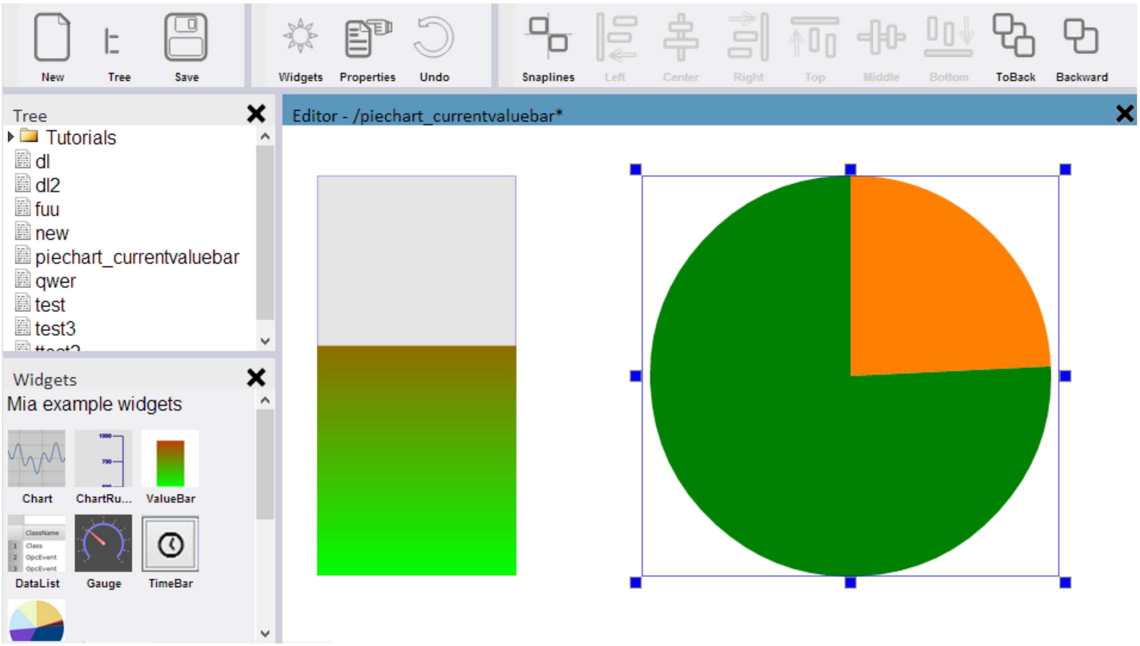


Figure 5-5 *Editor for generating and configuring WebUI's for RTDB.*

5.3 Proof of concept testing in virtual environment

In this chapter the system is tested as a whole functioning unit with all pieces put together. The case is assumed to be an ideal one with process database completely done by the specifications of IEC61850. This doesn't reflect the typical real life scenarios but it makes the database handling a lot simpler. It should be pointed out that this test is purely intended to test the system instead of the big data handling principles. The data is generated through simulating the process points in SCADA system which are then relayed into the Historian database. The raw data is then read and processed in the calculation environment after which the results are saved back to the database. The system functions just like it would in a real scenario. A simple circuit breaker condition monitoring was chosen as the case for study.

5.3.1 Circuit breaker

Circuit breakers are devices, which are used to open and close electrical circuits. In electricity distribution the circuit breakers form the backbone of network protection being the only component, which has the ability to disconnect live lines without taking damage, which would prevent further operations. Circuit breakers differ from conventional switches in their capability to break currents multiple times the nominal operation current. The most important task of a circuit breaker is to interrupt fault currents caused by short circuits or earth-faults and thus protect rest of the infrastructure and its users. In these cases the opening command is usually issued by an IED which has detected an anomaly e.g. currents that exceed the safe levels. The disconnecting and reconnecting after fault has been cleared should be carried out so that the system stability is maintained at all times. The operational requirements of the circuit breakers have steadily grown at the same pace as the electricity distribution systems have evolved to meet the ever increasing power requirements. The requirements on live tank circuit breakers may be as high as 80kA current interrupting capability at 800kV rated voltage. The earliest circuit breakers were based around oil and compressed air as the insulating medium to extinguish the electric arc. Nowadays SF_6 gas operated circuit breakers have replaced the traditional methods almost completely at high-voltage applications. This is because in general circuit breakers containing SF_6 gas are less bulky at the same ratings and therefore more economical choices. (Haarla, et al., 2011) The Figure 5-6 shows a cross-section of typical SF_6 circuit breaker and its components. The red line shows the current path under normal operation i.e. when the circuit breaker is in closed position.

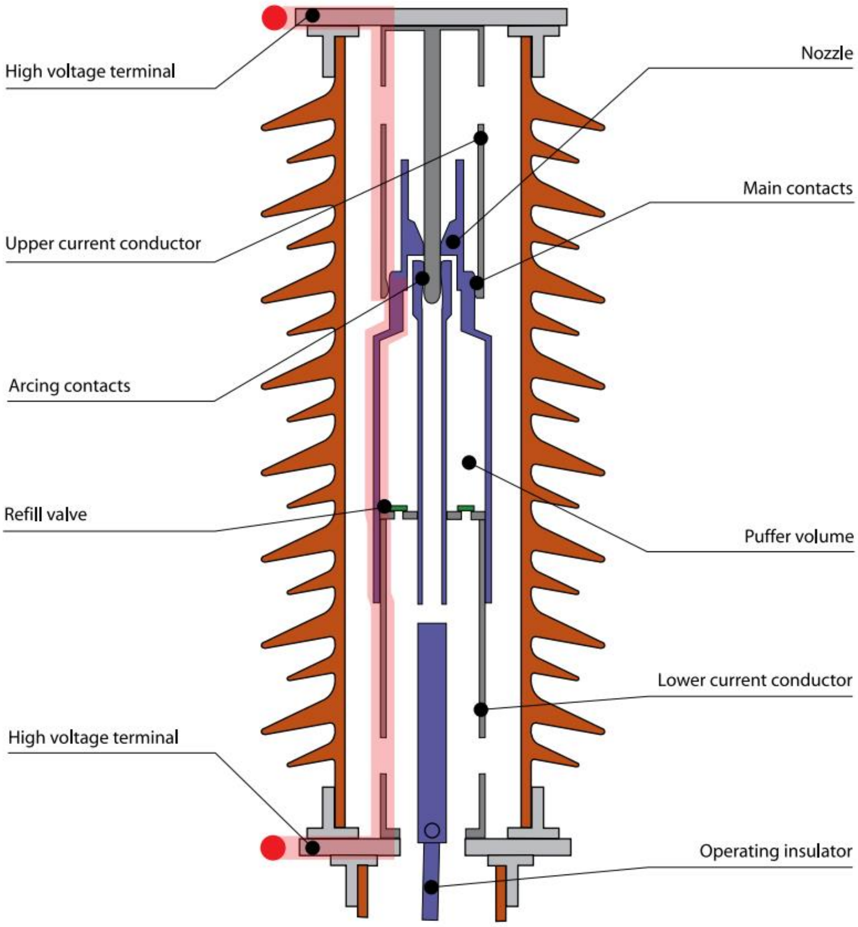


Figure 5-6 *Main components of a typical high-voltage SF_6 circuit breaker. (ABB, 2013a)*

All kinds of switching devices have two different states of operation – the normal and the operational state. Most of the time the switch has to conduct currents with as low resistance as possible to minimize the losses. However the switch has to be able to turn from conductor into insulator or vice versa at moment's notice. This change of state only takes few tens of milliseconds. The current interrupting sequence is illustrated in Figure 5-7.

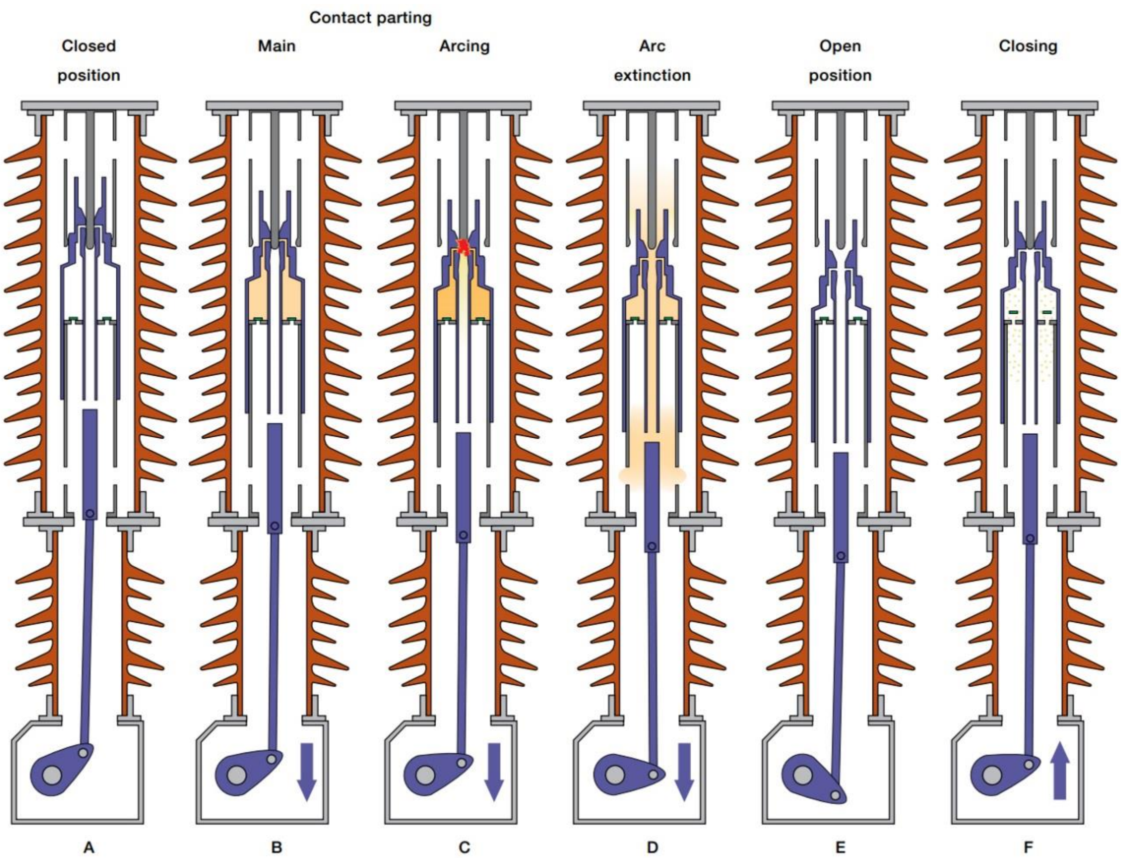


Figure 5-7 *SF_6 circuit breaker current interruption operation cycle. (ABB, 2013a)*

The current interruption sequence begins with the separation of main contacts which carry the current through the breaker under normal operation. Separation causes arcs to form which in turn start to heat up the gas inside used as insulator medium. The pressure increases in the puffer volume because of this. When the contacts move far enough from each other a channel for high-pressure gas is opened and it is blown out from the puffer volume through the nozzle. The blast cools the arc and eventually extinguishes it. It is worth noting that current flow doesn't stop right after the contact surfaces separate because the electric circuit is maintained through the arc. At the very moment of separation the contact surfaces partially melt and vaporize which in turn creates favorable conditions for arc to form. Because of this chain of events every current interruption sequence takes its toll on the circuit breaker. Eventually mechanical wear from repeated operations leads to the failing of circuit breaker. (ABB, 2013a)

5.3.2 Circuit breaker maintenance

The ABB made circuit breakers can have a service life exceeding 30 years and 2000 or 10 000 mechanical operations depending, which class of circuit breaker is in question. This kind of lifespan is achieved if the circuit breaker serviced and maintained appropriately. For each circuit breaker type the operating and maintenance instructions define time intervals and activities for three types of maintenance: ocular inspection, preventive maintenance and overhaul. The ocular inspections are done normally at 1-2 year intervals to keep in track with circuit breakers overall condition. This method naturally reveals only major problems as internal condition can't be evaluated. Preventative maintenance is either performed after 15-years of service or earlier if the breaker reaches maximum permissible electrical wear. Finally after 30 years the circuit breaker is taken out of service for a complete overhaul where contact surfaces and other parts subjected to mechanical wear are replaced. This major overhaul is expected to add another 10 to 15 years of service for circuit breaker depending on the environment stresses it is exposed to.

A circuit breaker's life cycle costs (LCC) are the sum of the initial cost of acquiring the breaker, all the maintenance during lifetime and repair costs after failures in present values. Typical LCC calculations for ABB circuit breaker show that the cost of maintenance and repair donate only about 1/6th of the whole. As an example the Figure 5-8 shows how LCC forms for circuit breaker type HPL 420B2. The time span was assumed to be 30 years with interest rate of 5%. The overhaul costs after 30 years are included in the maintenance portion. (ABB, 2013a)

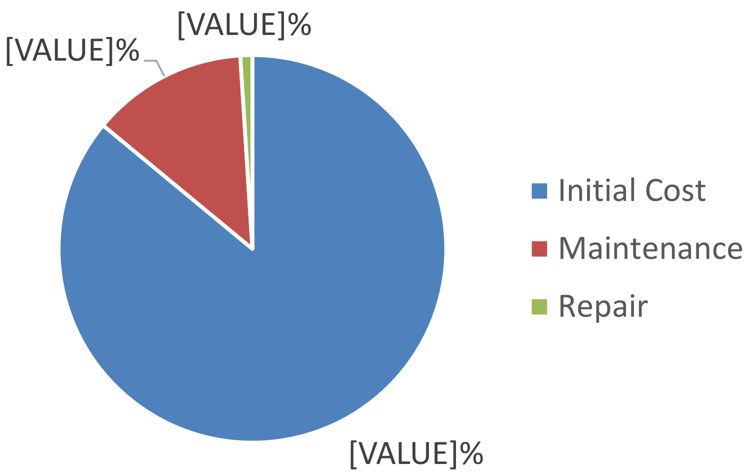


Figure 5-8 *Circuit breaker's life cycle costs divided into initial purchase, maintenance and repair costs. (ABB, 2013a)*

There is an ongoing trend of shifting from purely time based maintenance into condition based maintenance or a combination of these two. To be able to incorporate condition based maintenance one must first get reliable information about the real condition of maintainable device. Solutions for on-line condition monitoring of circuit breakers have existed for some time now, but they were mostly used in special cases. Generally on-line monitoring was applied to either very important circuit breakers or whose manual inspection would have been too time consuming and expensive. Nowadays however the IEDs have built-in functions for circuit breaker condition monitoring. This analysis could be shifted to substation level computer with access to all measurement data as well as possibility to link additional information for active analysis of circuit breaker's condition. System could supervise power supply voltage and current, motor circuit, contact travel (giving information about speed, overtravel and damping), operating times and speeds, damping time, ambient temperature and count of the number of operations to name a few. Some of the mentioned functions can be performed with normal setup of the substation and some require additional sensors to be installed.

5.3.3 Application for circuit breaker condition monitoring

To be able to analyze anything some data points are needed. In this case it is assumed that a substation called Herwood with 4 outgoing 110kV bays requires solution for circuit breaker condition monitoring. It is also assumed that the fault current measurement can be acquired into the SCADA system. In a real scenario this is rare as current information is usually gathered only for the steady-state operation of the network. However if this system were installed on a substation computer it would have access to fault current information through disturbance recordings. These simplification are made, because simulating SCADA signals is much easier than creating plausible disturbance recordings.

First task is to determine how to evaluate circuit breaker condition. The characteristic electrical weariness curve seen in Figure 5-9 for type LTB 170 D1/B circuit breaker can be found from its manual. It represents the maximum permissible electrical wear of the device.

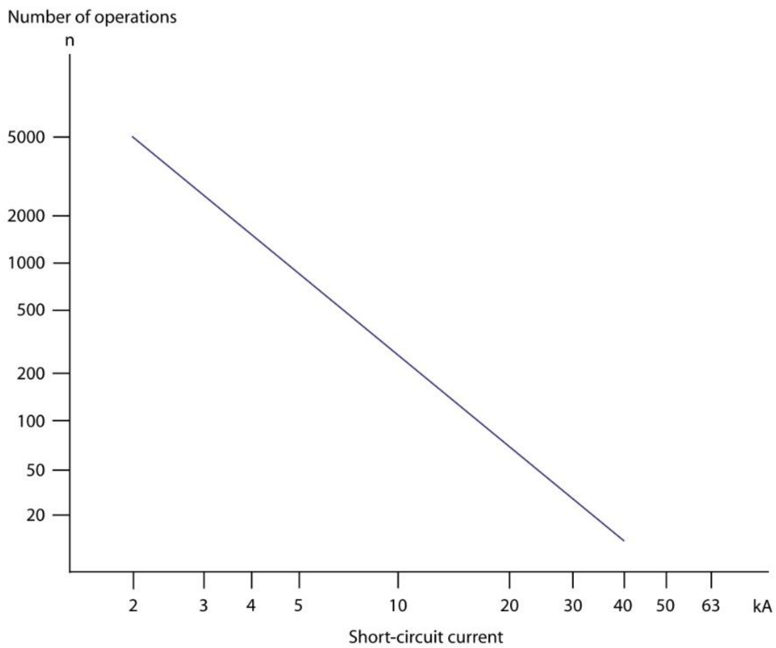


Figure 5-9 *Maximum permissible electrical wear of a 170D1/B circuit breaker (ABB, 2013a)*

Graph shows that the electrical wear depends on not just the number of operations, but also from interrupted current. Logically interrupting higher short-circuit currents result in higher amounts of electrical wear. Simplest way to evaluate circuit breaker condition would be to count its number of operations, but the graph clearly shows that this method could produce misleading results. A more sophisticated method is to take into account the interrupted currents by using following equation:

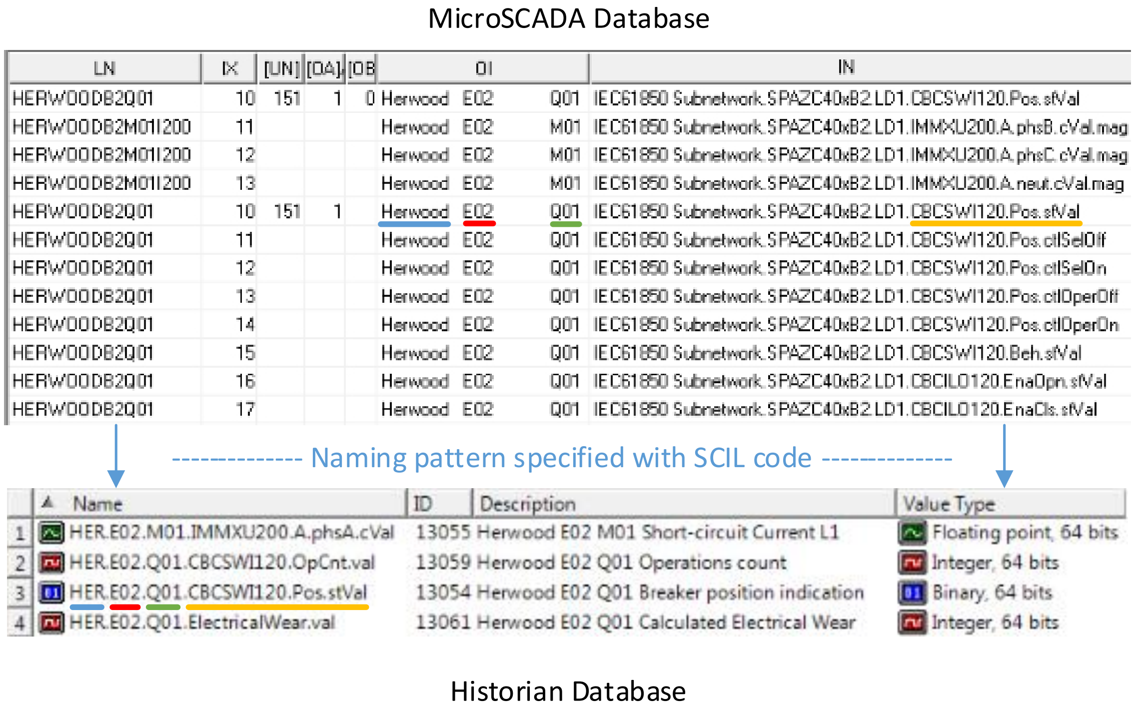
$$\sum n * I^k = T \tag{1}$$

Where:

n	Number of short-circuit operations
I	Short-circuit current, kA(rms)
k	An exponent in the order of 1.8-2.
T	A total permissible number, specific for each circuit breaker type

Now that the needed measurements are defined the configuring of SCADA database can begin. Some database engineering is required to create the process points and to link them into the historian database. MicroSCADA acts as SCADA system in this simulation while MicroSCADA Historian act as RTDB.

Nowadays these two products come together and their configuration is streamlined. In MicroSCADA user can create logging objects which are used to link process points into the historian database. Whenever a process point is linked with logging object a variable is automatically created into the RTDB and is named according to user specified conventions. These conventions can be configured using simple SCIL scripts to suit the needs of different databases. In addition single SCADA system can have multiple logging objects and therefore multiple different ways to compose a variable name tag to help to merge database sections with different naming conventions. The name tag can be created in any way by combining the different attribute fields of a single SCADA process point. For example in our case the name tag composes of the Object Identifier (OI) and the Item Name (IN) attributes like shown in Figure 5-10. OI attribute holds the hierarchical information of the process point's location divided into bay, voltage level and device fields. IN attribute is used for signal mapping of the IEC61850 signals and therefore contains the signal name. To get a name tag like described in the chapter 4.2.3 only these two attributes are needed, assuming the system is built around the IEC61850 standard. For described calculation purposes the circuit breaker state information and a supposed short-circuit current information is only needed.



LN	IX	[UN]	[DA]	[OB	OI			IN
HERWOODB2Q01	10	151	1	0	Herwood	E02	Q01	IEC61850 Subnetwork.SPAZC40xB2.LD1.CBCSWI120.Pos.stVal
HERWOODB2M01I200	11				Herwood	E02	M01	IEC61850 Subnetwork.SPAZC40xB2.LD1.IMMXU200.A.phsB.cVal.mag
HERWOODB2M01I200	12				Herwood	E02	M01	IEC61850 Subnetwork.SPAZC40xB2.LD1.IMMXU200.A.phsC.cVal.mag
HERWOODB2M01I200	13				Herwood	E02	M01	IEC61850 Subnetwork.SPAZC40xB2.LD1.IMMXU200.A.neut.cVal.mag
HERWOODB2Q01	10	151	1		Herwood	E02	Q01	IEC61850 Subnetwork.SPAZC40xB2.LD1.CBCSWI120.Pos.stVal
HERWOODB2Q01	11				Herwood	E02	Q01	IEC61850 Subnetwork.SPAZC40xB2.LD1.CBCSWI120.Pos.ctlSelOff
HERWOODB2Q01	12				Herwood	E02	Q01	IEC61850 Subnetwork.SPAZC40xB2.LD1.CBCSWI120.Pos.ctlSelOn
HERWOODB2Q01	13				Herwood	E02	Q01	IEC61850 Subnetwork.SPAZC40xB2.LD1.CBCSWI120.Pos.ctlOperOff
HERWOODB2Q01	14				Herwood	E02	Q01	IEC61850 Subnetwork.SPAZC40xB2.LD1.CBCSWI120.Pos.ctlOperOn
HERWOODB2Q01	15				Herwood	E02	Q01	IEC61850 Subnetwork.SPAZC40xB2.LD1.CBCSWI120.Beh.stVal
HERWOODB2Q01	16				Herwood	E02	Q01	IEC61850 Subnetwork.SPAZC40xB2.LD1.CBCILO120.EnaOpn.stVal
HERWOODB2Q01	17				Herwood	E02	Q01	IEC61850 Subnetwork.SPAZC40xB2.LD1.CBCILO120.EnaCls.stVal

	Name	ID	Description	Value Type
1	HER.E02.M01.IMMXU200.A.phsA.cVal	13055	Herwood E02 M01 Short-circuit Current L1	Floating point, 64 bits
2	HER.E02.Q01.CBCSWI120.OpCnt.val	13059	Herwood E02 Q01 Operations count	Integer, 64 bits
3	HER.E02.Q01.CBCSWI120.Pos.stVal	13054	Herwood E02 Q01 Breaker position indication	Binary, 64 bits
4	HER.E02.Q01.ElectricalWear.val	13061	Herwood E02 Q01 Calculated Electrical Wear	Integer, 64 bits

Figure 5-10 *Principle how SCADA database variables are linked into the RTDB and how their name tags are formed there from SCADA attributes using SCIL definitions.*

The Figure 5-11 represents flow of data within the application. Total of four RTDB data points are needed for this type of condition monitoring. Two of which are actual measured data streams from SCADA system and two are used to store results in between the calculations as calculation engine is not intended for data storage but temporarily during analysis.

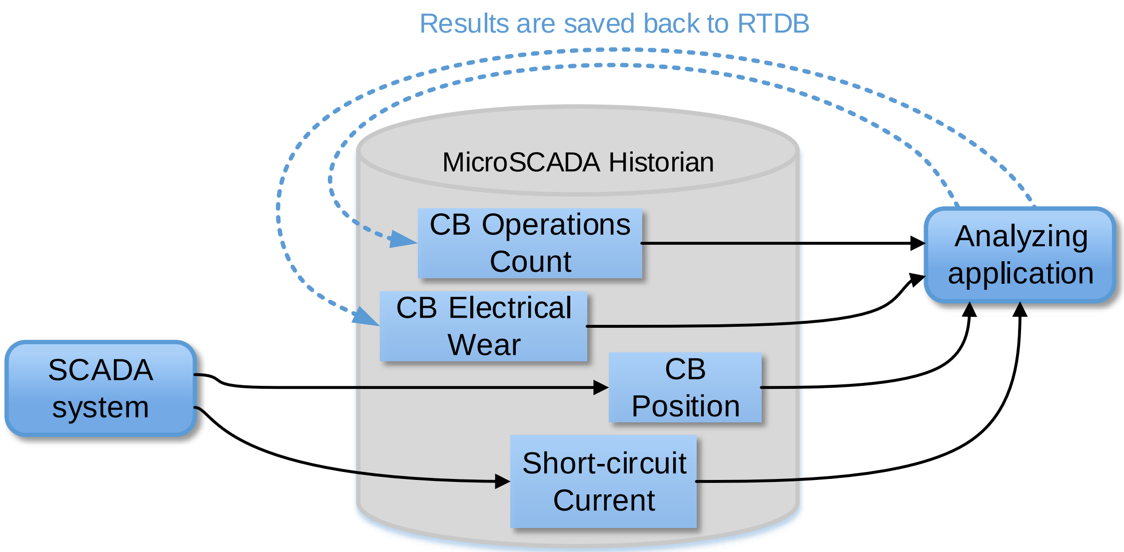


Figure 5-11 *Dataflow diagram of the application for circuit breaker monitoring.*

Now that the database is in working order and named properly the application itself needs to be designed with PCM 600. There are four main parts that need to be addressed: triggering, inputs, calculation and outputs. Triggering is done with “waitForSignal” block which triggers the calculation after every change in the circuit breaker position. This means that calculation is triggered not only after every circuit breaker opening but at closings too. The signals indicating circuit breakers current position and its cumulative amount of openings are linked with a custom block called “CBoperationsCounter”. The block detects if the breaker opened or closed at current trigger. With every opening the operations count value is increased by one and stored to the database with current time. After this the “ElectricalWearAnalysis” block reads the signal value of short-circuit current and calculates the electrical wear caused using eq. 1. The result is then added to the current electrical wear read from the RTDB and new value is saved into the database. The function block design of the application is shown in the Figure 5-12.

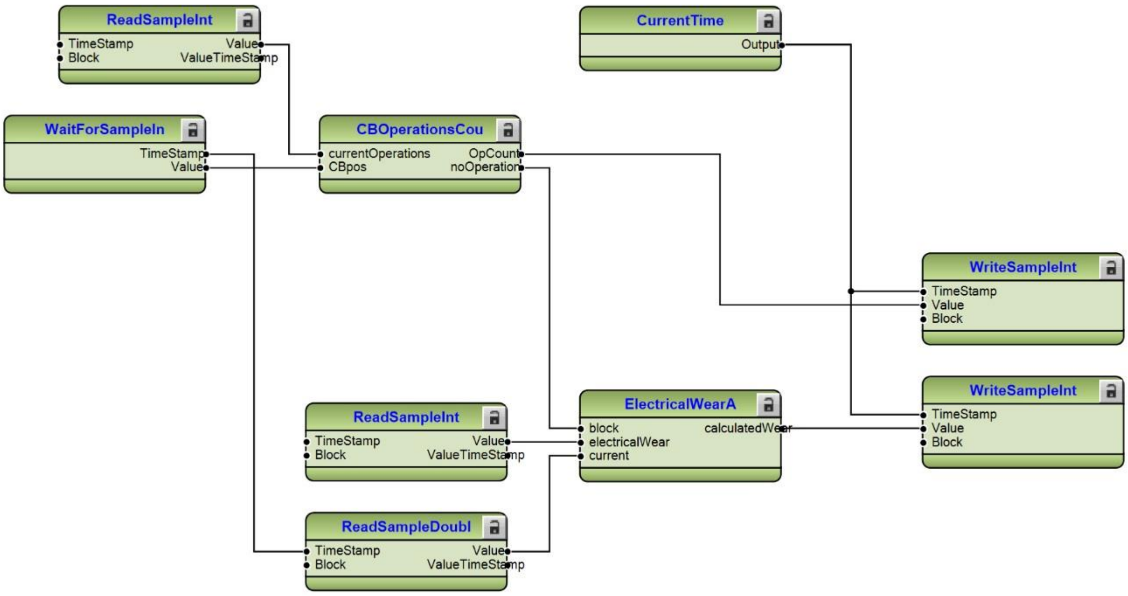


Figure 5-12 *Circuit breaker condition monitoring application design with function blocks.*

Application is then made active and some 800 circuit breaker openings and closings are simulated from SCADA with random short-circuit currents ranging from 1kA to 10kA. The Figure 5-13 shows the cumulative electrical wear and the count of openings plotted in a same graph. The Historian supports adding warning and alarming limits to database variables through newly added “Rule Engine”. With Rule Engine it is possible to create conventions, rules, according which information is displayed. If warnings and alarms were set correctly the user would get an advance notification when a circuit breaker is nearing its overhaul.

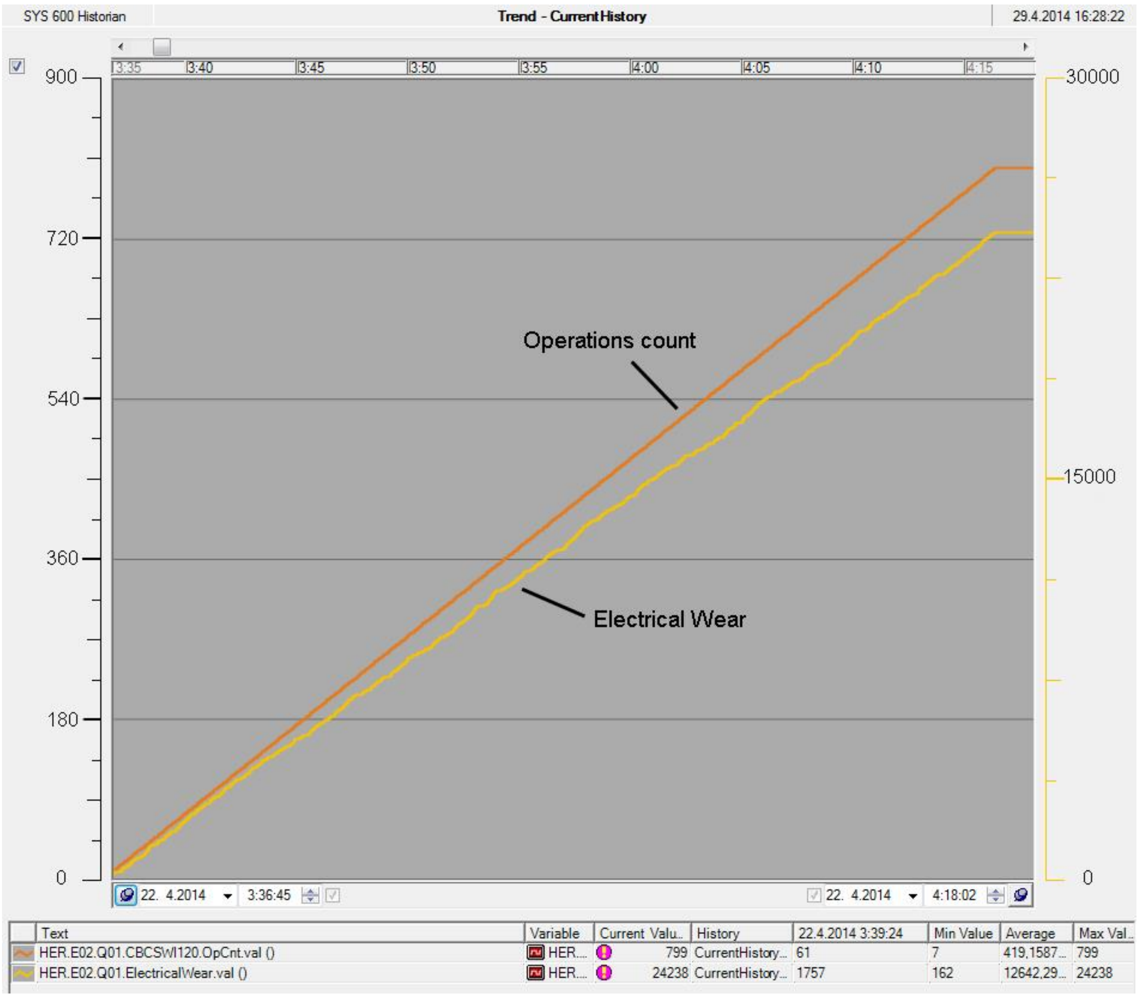


Figure 5-13 *Simulated results of circuit breaker operations count (orange) and corresponding cumulative electrical wear (yellow).*

This simple case demonstration proofs that the system functions as intended when brought together. In addition it shows clearly the potential value of data mining applications in power systems. These systems are usually operated by a handful of people with limited time to go through historical statistics. There is no doubt that a human operator could reveal the same results as automated analyzing does, but it takes dedication. Because of this, even simple applications can create results which lead to cost savings. Similar studies using historian database has been made by Mikko Söyrinki in his Master thesis Transformer Condition Monitoring in SQL Database. (Söyrinki, 2008)

6 PILOTING THE SMART SYSTEM ANALYZER

6.1 Distribution network of Elenia Ltd and the pilot project

The Elenia Ltd. is a Finnish electricity distribution company operating from Tampere. Elenia supplies electricity for over 410 000 customers covering over 100 municipalities with a network spanning from Tavastia to Ostrobothnia. This network covers over 50 000 km^2 and is illustrated in the Figure 6-1. (Elenia Ltd., 2014) From the point of big data the Elenia has approximately 140 primary substations with an average of 8 feeders per substation. In additions to this some 5000 remotely controllable disconnectors and 170 main power transformers. To approximate the total amount of processable data we can assume that there are around 5 to 10 data points per disconnector and few tens of data points per feeder. This leads to a total amount of 100 000 data points if the analyzing functions would be applied to every substation in the network.

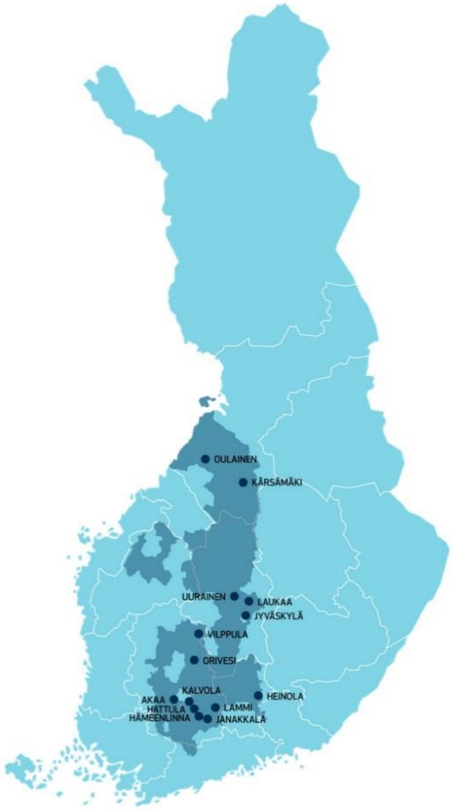


Figure 6-1 *Elenia Ltd. area of operations illustrated with darker blue color on the map of Finland.*

The Elenia has agreed to participate in a joint project testing the analyzing concept described in the chapter 5. The pilot is run under the Finnish Smart Grids and Energy Markets (SGEM) project. The main goal of the SGEM is to speed up the development of smart grid related solutions which can be demonstrated in practical environments. The pilot is planned to run in three distinct phases with increasingly demanding goals after each completed phase. The first phase is to test and demo the system with offline disturbance recording data provided by Elenia. The main goal of this phase is to iron out the bugs in the system and develop a few working functions for analyzing disturbance recordings automatically. After all parties are satisfied that the basic functionalities are working, the system is then installed into the Elenia's servers and run on real-time data from one substation. The final step will be increasing the coverage of the system into multiple substations. The incremental nature of the pilot was chosen to lessen the challenges of big data related problems and to get the project underway as fast as possible.

There are three different Master's thesis workers involved with the Elenia pilot. This one concentrates on the single feeder solutions with the goal of developing working algorithms for automated analysis of the events. Ville Jokela's work is to develop scripts and tools on how to apply working functions into greater amounts of feeders/bays. With 100 000 data points in Elenia's network it is clear that the mapping can't be done by hand. Third Master's thesis worker Tuan Vu has the responsibility of developing the ways for visualization of the results through Web User-Interface. The system principle intended for Elenia's pilot case is shown in the

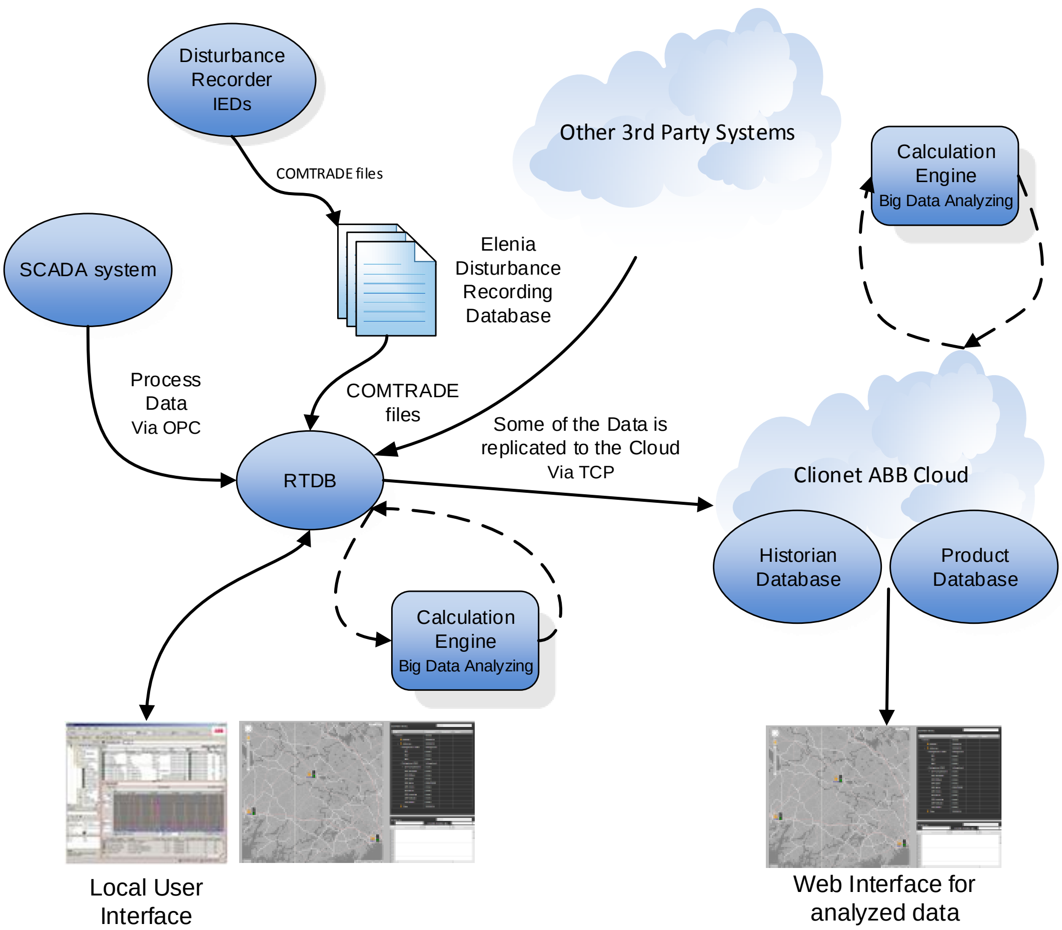


Figure 6-2 *System overview diagram used in the Elenia pilot case when analyzing is performed at the control center level.*

The goal is to gather data primarily from SCADA system and disturbance recordings with additional configuration data, which is essential for some analyzing functions, read directly from file system. Elenia has already a system in place which fetches disturbance recordings from relays whenever a new recording is made. This recording is then saved in an orderly manner in which the directory path represents the equipment model of the Elenia's network hierarchy. When importing the disturbance recordings, the directory paths can be used as means to sort out where the recordings are originating at.

Possibility of replicating the data from Elenia pilot to the ABB file server is also looked into as it would give an enormous source of testing data for prototyping purposes. The ABB Clionet service could be an ideal solution for this as it is already an implemented service. The present Clionet is data-sharing and backup service primarily intended for secure keeping of technical documentation of protection relays throughout their lifespan. Access can be granted through web-interface so that maintenance staff has always access to up-to-date information. Naturally some of the big data mining results could be linked to the existing Clionet services and used to support the maintenance procedures.

6.2 Fault Location Analysis

The fault location calculation is not a new invention in the power systems and grid management. Many relays today perform fault location analysis based on the short-circuit currents and line segment impedances which are obtained from the relay setting parameters. The fault-current, fault-impedance and the calculated fault distance is then sent to the DMS system for The main objective of fault location calculation is to achieve fast, selective and reliable operation for faults on a protected line section. Besides this the information is of great importance when operators are trying to determine where to send the repair crews. Accurate results decrease greatly the downtime of the protected lines.

6.2.1 Calculating the Fault Location

The calculation of fault location in relays provides impedance-based fault location as result. The fault location can be calculated with relatively high accuracy from short-circuits and even earth faults. However earth faults have some limitations when calculation is applied to isolated networks. The calculation is based on the fundamental frequency component of the fault current and phase voltage phasors. The calculation is done in two steps. First the type of the fault is detected because it rules which kind of mathematical solution is required to get the correct result. Identification of the faulty phase is based on combined impedance and current criteria. There exists three different fault types which require different method for calculation: single-phase earth faults, phase-to-phase short circuit faults and three phase short circuit faults.

As soon as a fault condition is recognized the fault distance calculation starts with the correct one of the seven impedance measuring elements modeled as fault loops. The one used for three-phase short-circuit is presented in *Figure 6-3*.

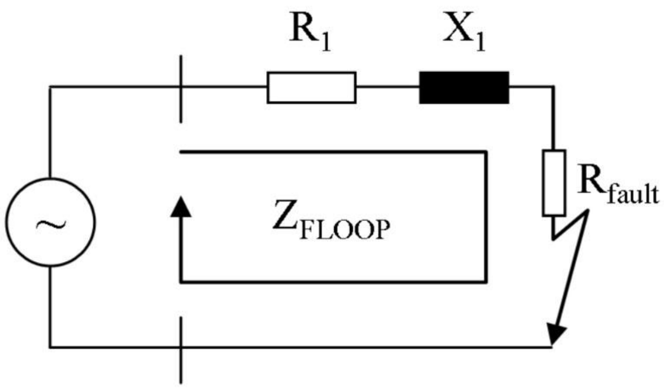


Figure 6-3 *the substitute circuit for three phase short-circuit calculation.*

The fault loops are parameterized with the positive and zero sequence impedance settings as an initial data. The short-circuit calculations require only the positive sequence impedances. The impedance values can be obtained from the datasheets, but they are only good for certain tower configurations. To correct the values the values are adjusted according to the actual install configuration of the power lines.

The inherent result of the fault distance calculation is the ohmic fault loop impedance value:

$$Z_{FLOOP} = R_{FLOOP} + j * X_{FLOOP} + R_{fault} \quad (2)$$

Depending on the fault type the composition of terms R_{FLOOP}, X_{FLOOP} and R_{fault} is different.

The accuracy of the fault location calculation is affected by steady-state asymmetry of the system. In reality, power systems are never totally symmetrical, which produces errors if assumed so. To mitigate the effect of steady-state asymmetry the calculation uses delta quantities instead of relying on the actual values. The delta quantities describe the change in the measured quantities caused by the fault:

$$\Delta x = x_{fault} - x_{preFault} \quad (3)$$

To be able to calculate the change in quantities the pre fault values must be sampled before the fault condition occurs. Another major source of error for correct distance calculation is the current caused by the load which is made even worse with higher fault resistance values. The delta quantities can be used to compensate for the effect of the loading. There are other factors that have serious impact on the accuracy of the results. The most problematic being the fault resistance. Basically the smaller the fault resistance the more accurate the calculation gets. This is why earth-faults are the most problematic as in some cases the fault resistance can be quite high. Also the saturation of the current or voltage transformers at some harmonics deteriorates the fault distance estimate. (ABB, 2012)

6.2.2 Benefits of FLOC analysis done at higher than bay level

As already mentioned, the fault location calculation is currently performed mostly in relays and results sent back to the control center of the network. The information is then linked with the distribution management system which can pinpoint the possible fault locations from the map. This makes sending repair crews more effective as there exists a good estimate where the repair crew should start the search.

Up until now the calculation of fault location of earth-faults has been impossible in networks that are separated from ground potential. The majority of Finnish 20kV networks are built like this. However recent developments have made possible the fault location calculation in these kind of networks and more advanced algorithms are being developed to allow for calculation in compensated networks. To be able to benefit from these new developments requires updating the software in the relays, which in turn leads to mandatory proofing tests to that relay and its primary devices. It goes without saying that taking out feeders for overhaul just for software updates is not feasible and surely not financially arguable. However if the calculation were done centrally at substation computer the upgrading wouldn't affect the normal operation of the network. The owner would benefit from all the developments with as fast pace as possible. As the algorithms evolve to handle more complex situations the processing power requirements increase too. If the relays are expected to last for several decades in the network the processors in them will be severely outdated at some point compared to the day's standards. Updating the one centralized substation computer is an easier option compared to updating all of the substation IEDs

Another major advantage of FLOC calculation performed at higher levels than at the IEDs is the wider perspective it allows for. When fault occurs the fault location accuracy can be improved by combining data from various sources. For instance if network contains measurement units at secondary substations the calculation can benefit from their measurements. Additionally integration with DMS could provide always up-to-date network topology settings which are needed for accurate results. Another example of combining data sources is the use of lightning locator services, which are already openly available online. As one of the most common cause for faults are the lightning strikes to and near the transmission lines, it would be very beneficial to be able to rule these faults out as recognized to be caused by thunder storms. This is because it would give the operators instant knowledge that the occurred faults were indeed of passing nature and not part of some major issues building up. In addition recognizing the faults helps when making analysis based on historical data.

The mentioned methods are all based on calculations with real-time data, but one of the major differences between station level computing is its ability to gather and store history of the events. Fault location results can be of value if used to predict upcoming faults. Performing the location calculation not just for permanent faults but also for passing faults could reveal incidents that would otherwise be overlooked. One example of this is a tree barely hugging a transmission line. It would produce frequent disconnects fixed by autoreclosures until it eventually falls down on the line severing the cables completely. Such a case is demonstrated in the Figure 6-4. If the fault location would be calculated from each of these disconnects preceding the actual fault a pattern emerges, which can be used to give an early warning for users. If the repair crew is send before the tree falls completely down the repair costs would be considerably lower. Vesa Hälvä established in his master's thesis commissioned by Elenia that a need for a system analyzing repetitive reclosings exists. (Hälvä, 2013)

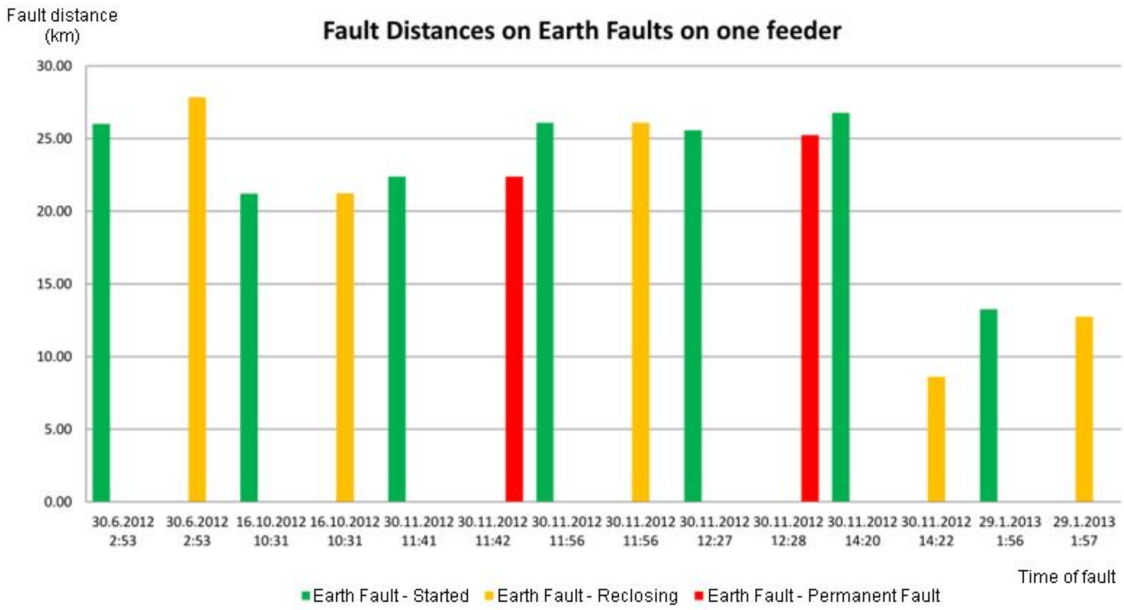


Figure 6-4 *Earth-fault distance calculation results of one feeder/bay. The fault distance is on the y-axis and the event occurrence time on x-axis.*

The green bars represent the moment when an earth-fault is detected in the system. It is always followed by another bar either yellow or red. Yellow bars indicate that the fault passed after reclosing of the circuit breaker and the red represents faults that were not corrected by such action. It can be seen that before each permanent fault there are passing faults which originate at approximately the same distance as the permanent fault was eventually calculated. Slight variation of the distance is to be expected as earth-fault calculation accuracy is affected by other events as discussed in the chapter 6.2.1.

6.3 Network protection operation time analysis

6.3.1 Protection functions within an IED

Modern IEDs contain a great variety of different protection functions designed to detect and react to various different unwanted situations. The amount of protection functions IED has can vary from only a few to several tens. In this chapter only the most commonly used feeder protection functions are presented while there exist many more. The majority of feeder protection functions deal with over currents caused by short-circuits or earth faults as these are the biggest threats for the safe operation of the electrical grid. There are also functions that monitor over- and under-voltage situations**.**

Overcurrent protection deals with currents that exceed the levels of safe operation. These are usually caused by a short circuit on the power transmission line. Most common causes for short-circuit faults are caused by lightning strikes on the lines, mechanical failures of the system or ice and snow building up on the cables. Modern IEDs offer multiple options for overcurrent protection which can be divided into two categories according to their way of operation: directional and non-directional. Non-directional functions simply monitor the current magnitude while the directional functions also calculate the direction of the current flow.

Another major source of overcurrents is earth-fault. Earth-fault is caused when phase conductor gets in contact with earth potential either directly or through other part of the system. This usually happens as a result of a failure in the insulation or as an after effect of some other type of fault. IEDs have functions for detecting these types of faults just like with the short-circuit situations. The nature of the earth fault makes it harder to detect than the short circuit induced overcurrents. The earthing resistance is in some cases so high that it is impossible to make a distinction between a fully working and a faulty system just by current measurement. Because of this the earth-fault detection relies on residual current measurement which should be zero or close to zero in a balanced load situation. The earth fault disturbs this balance and causes residual current to rise. In many cases it is difficult to achieve selective earth-fault protection based on the magnitude of residual current only. To obtain a selective earth-fault protection scheme, it is also necessary to take the direction of the residual current into account.

Detecting earth-faults in isolated neutral, resistance earthed or reactance earthed systems is impossible by using the conventional method of measuring only the residual current. In isolated neutral systems the earth-fault current can only form electrical short-circuit through the line capacitances of transmission lines. This leads to undistinguishable low fault currents which are hard to detect. To detect faults in these kinds of systems the used method is to measure the rise of system neutral voltage, the residual voltage, which tends to rise during earth-faults.

A more specialized case of earth-fault is the intermittent earth-fault encountered especially in compensated network with underground cables. As the current trend is to use underground cabling when building new transmission lines there is a growing need for this type of protection. Intermittent earth fault is caused by cable insulation deterioration which causes it to lose its insulation qualities. Eventually cable insulation fails to withstand the voltage difference between phase-to-earth and the fault is initiated. The intermittent earth-fault differs from ordinary earth-fault the nature of the fault current. During intermittent earth-faults the fault extinguishes itself when the current drops zero as illustrated in *Figure 6-5*. This causes very short transients in residual current and residual voltage waveforms which can be picked up by IED logic. There are several factors which affect the magnitude and frequency of these transients such as fault resistance and the time at which the fault occurred in respect the voltage's waveform. (ABB, 2012), (Haarla, et al., 2011)

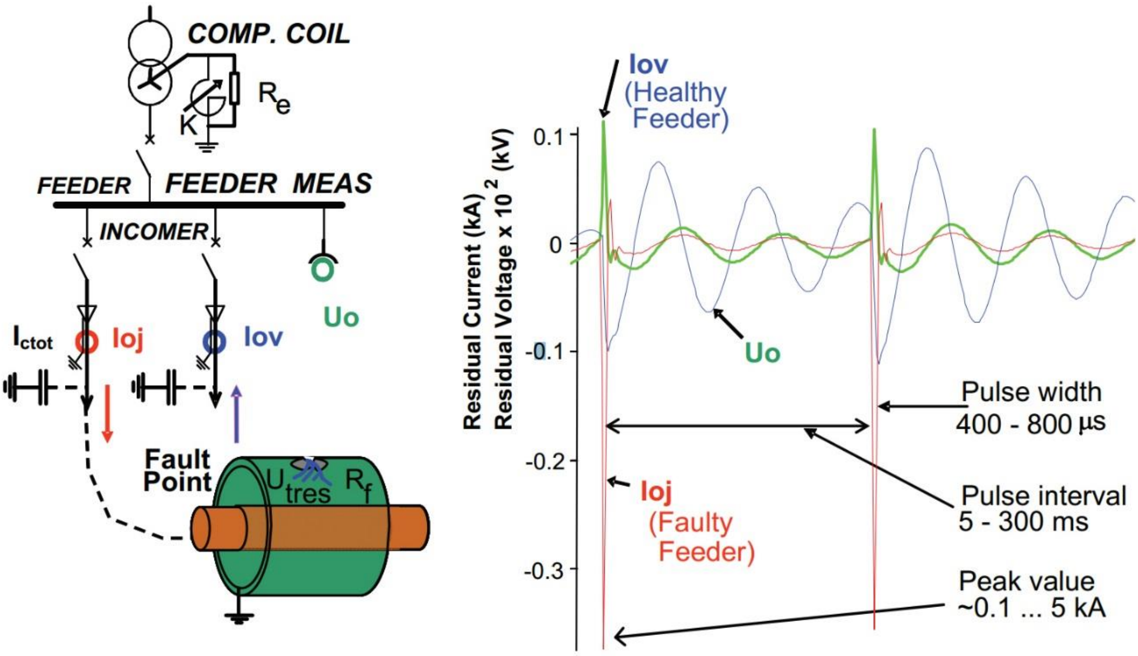


Figure 6-5 *Typical intermittent earth-fault characteristics. (ABB, 2012)*

6.3.2 Calculating the operation time

First thing is to define what the operation time of protection is from IED's perspective. The sequence of events happening internally within an IED during fault is shown in Figure 6-6.

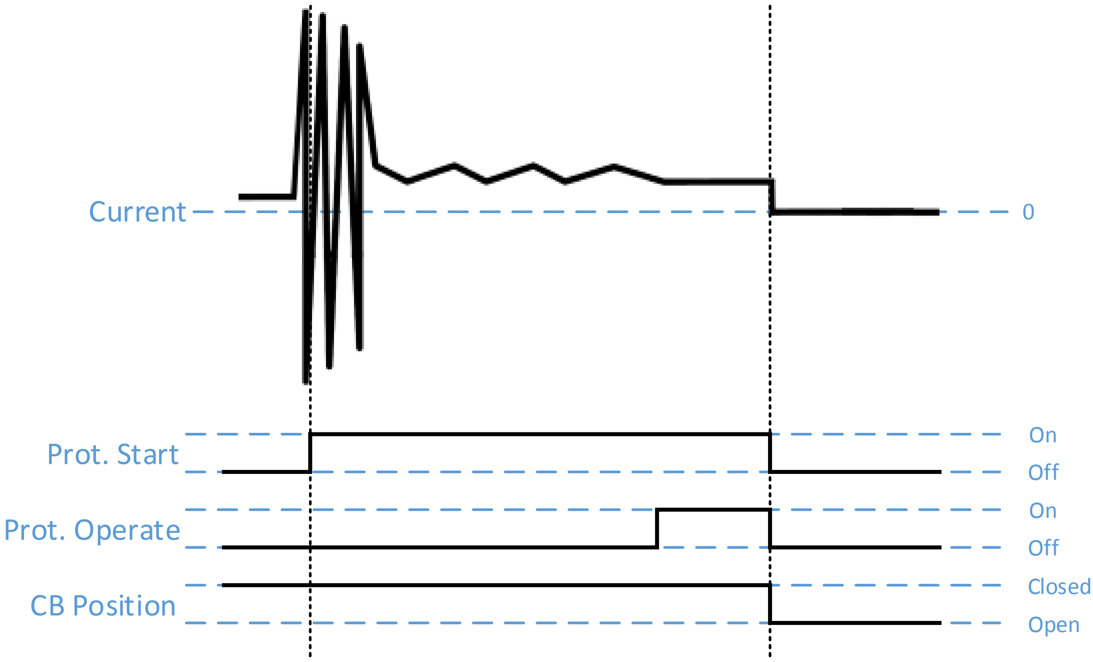


Figure 6-6 *IED's internal signals in an overcurrent protection event.*

The relay records the current signal which exceeds the allowed limit. This leads to internal start of protection functions which are designed to counter the fault in question. As the start signal rises the internal clock also starts to run. IED stays in dormant mode until enough time has passed from the point of start and switches into operate mode. The correct amount of time is determined from the configured settings which are set so that the protection as whole works selectively and reliably. If the fault is cleared during the start phase the IED returns back to normal state where all the protection function are dormant. However if a protection function gets to the operation mode the IED sends signal to the circuit breaker to open and the line segment managed by the IED is disconnected from the rest of the network. If protection was designed and configured right the fault is cleared from the system. The protection operation time is the time between the start signal going active and circuit breaker being registered as opened.

Now that the goal is clear the focus can be shifted into real world application, which is seldom that straightforward. After getting the disturbance recording into the calculation environment the first thing is to read through the binary channel signals and check if the circuit breaker open signal has gone active at any point. If it hasn’t and no protection start signals are detected either then there is no need for further calculations as current recording contains no fault event. In other case the analyzing can begin. Analyzing starts from the circuit breaker opening time calculation. This is achieved by finding the last moment from the recording where circuit breaker opened, and the fault perceived, and the moment before that when the circuit breaker was last registered as closed. The circuit breaker state is marked with double binary, which is stored into the fault recording as two separate binary channels. The circuit breaker operate time is calculated as time between points A and B shown in the Figure 6-7.

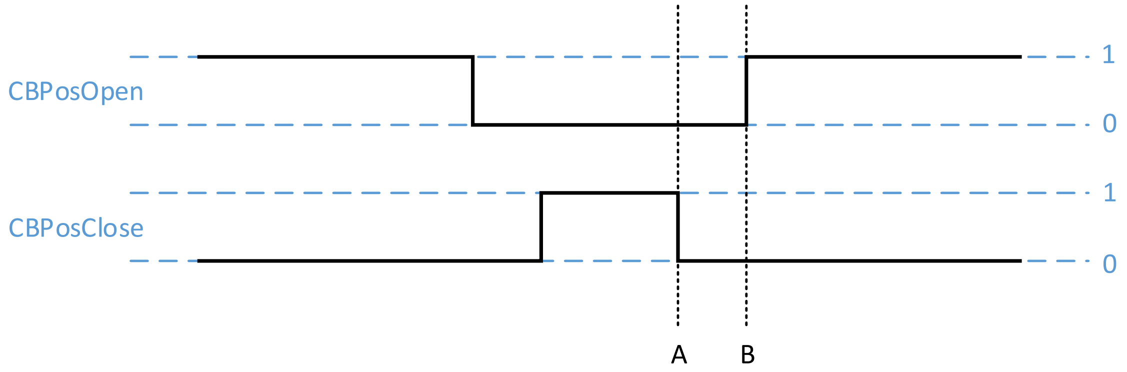


Figure 6-7 *Signals indicating circuit breaker state and its operate time marked between points A and B.*

There are some faulty situations to consider when calculating the circuit breaker opening time. The algorithm can detect if the open and close signals overlap which is physically impossible as this would mean that the breaker is open and closed at the same time. This can be the result of either circuit breaker micro-switch malfunction or internal IED fault. In any case it is recorded to the result log.

In a real case multiple protection functions are activated instead of just one like presented in the example Figure 6-8. The next task is to pick right one of them because only one got to the operate stage causing the circuit breaker to open. The simplest method to achieve this is simply by finding the operate signal first which there should be only one. After the correct protection function is found and its start signal time of origin detected the application can calculate the total time needed for protection to operate

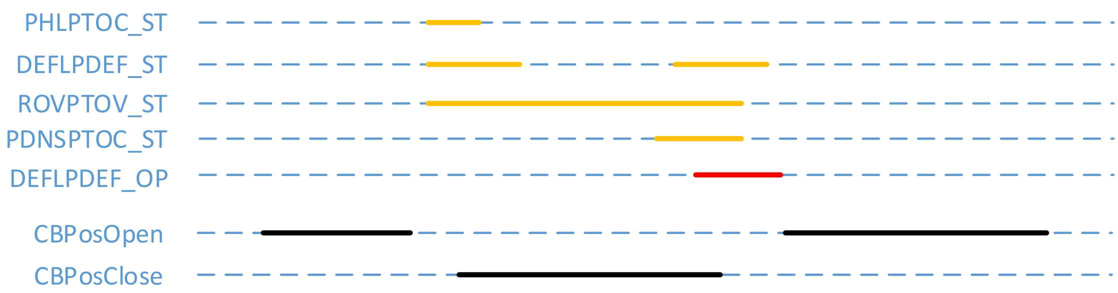


Figure 6-8 *Events recorded by an actual IED during earth-fault with multiple functions active of which directional earth fault protection operated. Yellow lines indicate protection START-signals and red line indicates OPERATE-signal.*

The next thing is to compare the calculated operate time against the setting values specified for the protection function in question. If the time exceeds set values the analyzing result is tagged as an alarm or warning depending on the overshoot. Finally a record is made into the RTDB table, which can be then linked onto the user interface for visualization purposes.

6.3.3 Benefits of the protection operation time analysis

The main motivation behind this kind of analysis is the ability to supervise if the protection is working as intended. Currently the information about the protection operation time is obtained only after mandatory proofing tests or by manual analysis of disturbance recordings. To be able to detect defects in protection operation as soon as they emerge is extremely beneficial as failing protection can have very serious consequences. The worst case would be a loss of life, but majority of cases damages would be limited to unnecessary disconnections of correctly working lines.

Reading the signals necessary to determine protection operation time yields also information of circuit breaker's operation time. This can be used to evaluate the cause of failing to meet the protection time limits set for the IED. Which in turn gives operators an advance warning of circuit breaker's condition deterioration. Using simple open/close signals measured by micro switch located at the main axle of circuit breaker only gives information about the breaker as a whole unit. If circuit breaker operate time would be calculated from the current signals instead of the state signal the calculation would get the performance results of every phase contacts individually. Adding this to already existing analysis wouldn't need much further development as all the information necessary is already read into the calculation environment.

With history database comes always the option of detecting long term changes in operation. Following the operation times of a circuit breaker over an extended period of time could reveal problems causing the circuit breaker operation time grow steadily. It is not enough to simply monitor the operation times of circuit breaker as there are other things affecting the performance such as the ambient temperature which in part justifies the use of a system like presented in this thesis. All this analysis about the circuit breaker operation can add value to the primary circuit breaker condition monitoring described in the chapter 5.3.3.

The protection operation time analysis detects which of the relays internal protection functions lead to the disconnection. This is valuable information when combined with results of other analysis and information obtained later. The fault location calculation for example can detect the type of fault and by comparing the results of these both analysis it is possible to evaluate if the correct protection function operated. The IED configuration can be revised if discrepancies are found and the protection of the network would be improved.

6.4 User interface for visualization of results

The importance of proper visualization of the results has been mentioned a several times already and it can't be stressed enough. Even the most sophisticated big data mining applications are close to being useless if the important results can't be highlighted. Tables and lists are seldom the best option when dealing with masses of information. Information flow generated by current systems is so high that users become easily overwhelmed. An event list of process automation system is a great example of this, as it is flooded with entries when something goes wrong in the system. It is a great tool in making decision in real-time, but for a historical presentation of system's state it becomes unreadable as events tend to pile in.

The user-interface done for this pilot project consists of a main view showing the state of the whole distribution network and substation level views for showing more detailed information. An example of how the main view could look like is shown in Figure 6-9.

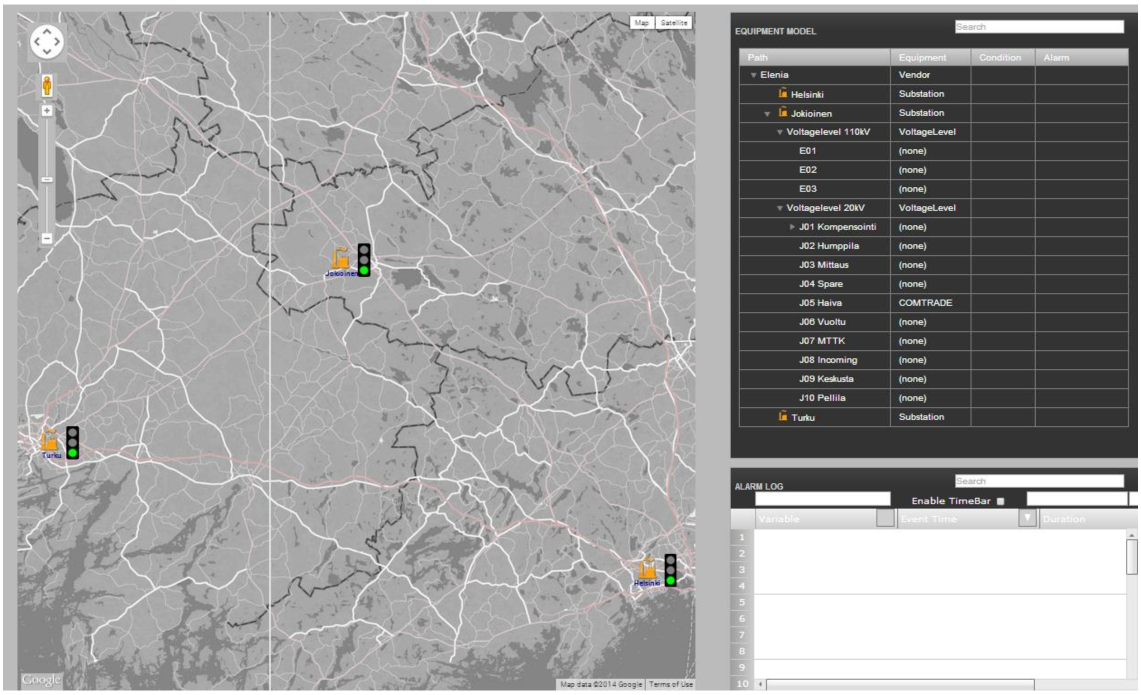


Figure 6-9 *Proposed main view of the web user-interface for Elenia pilot.*

The main view divides into three main sections. The map being the most visible one, the equipment model tree on the right side of the view and an Alarm log on the right bottom side of the view. The map is intended to give a quick overview of the system state at all times. It is made freely zoomable to suit the needs of different sizes of systems. Another distinctive feature of map view is the use of status symbols tied to every substation. These are represented by traffic lights. The green light indicates that everything is performing as it should, the yellow sign is reserved for warnings and the red is for alarms. It is intended that warnings are given from situations where an event is detected which is not considered as an obvious fault e.g. condition monitoring analysis has found a pattern emerging with circuit breaker operation. The red is on the other hand for clear faults which need immediate action. An example of an alarm could be the protection operation failure. Another way of browsing through alarms is the use of Alarm Log, which lists all active alarms detected in the system. The equipment model is used to navigate within the system. It lists all the substations divided into voltage levels and finally corresponding bays. The user can jump to any substation by clicking the substation name from the equipment model list. This would produce a detailed view of the substation like shown in Figure 6-10.

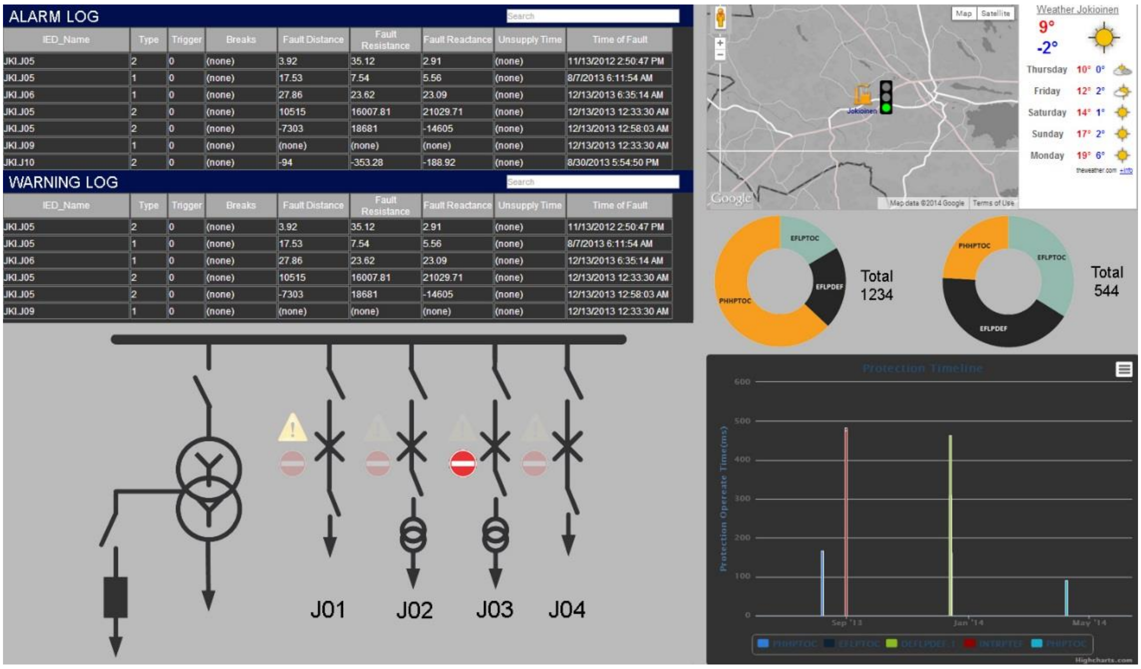


Figure 6-10 *Detailed view of the substation events and analysis reports.*

When user opens the substation view he is firstly presented with two logs for alarms and warnings. These both hold the active alarms currently in that substation with filtering and search functions to find specific item if log has multiple entries or when going through log histories. Below the logs is a clear presentation of substation diagram used for structuring information. Every bay can have its own presentations e.g. alarm and warning symbols lit up if there are any concerning that bay. By clicking the bay symbol (or warning/alarm log entry concerning the bay) the user gets into more detailed view of the bay. On the right upper corner a more detailed map view of the substation and its surroundings is presented. With future systems it could show the network topology linked from DMS for example. To the right of the map is real-time weather information of the area where the substation is located. The weather data is streamed directly from the weather forecast company Ilmatieteenlaitos. Below the map some general information can be displayed. In the Figure 6-10 a model of stacked bar graph is shown presenting the results of protection operation time calculation. Another way of showing the results could be the use of a pie chart which would show the total number of operations within substation with every segment showing the ratio of each operated protection type. The page can be made to look whatever suits the needs of the end-user the best. In any case the goal is to be able to give some kind of general picture what is happening at the substation with detailed information hidden until user specifically wants to see it. A prototype of how the detailed bay information view could look like is shown in Figure 6-11.

Figure 6-11 Detailed view of the substation bay and analysis reports concerning it.

The bay information view gives user a more detailed view of individual bays of the substation. The page contains the results whatever analyzing functions are run on the substation in question. Some examples could be future adaptation of power quality calculation, which would be best presented as graphs to see how they develop. The results of fault location history is also presented here with some probability estimates for predicted reoccurring fault. From every fault location calculation the original fault recording is stored in the RTDB and can be displayed for manual verification. The protection log would list all the protection events with their operation time margins from the set time limit.

7 CONCLUSIONS

This thesis began by examining the current situation and practices of today's power systems. The grid itself is ageing while at the same time requirements for its reliability are getting higher. It was established that electric grid resiliency is something that has serious impacts on our everyday life and even economy. During these times of recession every new way of improving efficiency and productivity is always welcome. A new trend of big data was presented as an option to achieve this. It is clear that this analyzing of huge data masses will eventually spread to all fields of industry and might even be main way of companies to outperform their rivals.

The field of substation automation systems has always been about the efficient handling of big databases. Therefore developing big data related analyzing for power systems is only a logical step to take if the aim is to increase productivity.

The main goal of this thesis was to give a closer look into a novel system proposed for automatic analyzing of substation data. This system would generate another automation level at the substation between the current IEDs and the network control center. The system has centralized historian database and a novel calculation environment for conducting data mining. This system is built around already proven products such as MicroSCADA Historian and PCM 600 relay configuration tool. All this makes development and configuration of the system easier and more familiar for project engineers. Even though the system is still under development, it shows great promise. Basic functionalities are in place and new are being implemented. Elenia's pilot testing revealed that the proposed system can be used successfully to analyze actual network data. However as the product development and the testing is still underway, it is clear that to get a fully working product, there is still work to be done.

The major benefits this kind of system is its increased view of the substation events as it is not limited to one bay only, instead it has access to every data source within the substation. This gives it a distinct edge over current network analysis done at the IED level. Additionally it would ease off the work load of current IEDs by taking some of the non-time-critical functions to handle. While being non-integral part of the protection scheme, the system can be taken off-line at any moment for upgrading and tinkering. Because of this the system would be always up-to-date and running the latest analyzing functions.

The proposed system also opens up new possibilities for conducting power system analysis as its database is designed to record and manage history data. This makes statistical analysis possible with chance to reveal slowly developing events. When analyzing functions are upgraded it is also possible to run them with old already analyzed data to get even more accurate results and to make results comparable with newer entries.

An intriguing possibility to speed up the development of big data analytics would be a model where analyzing functions would be given freely for customer and as return service the developer of those functions would get access to the raw data. This would provide an enormous resource for further research and development.

Finally some ways to visualize the results of the analysis were looked into. It was stressed that finding the right ways for visualization is one of the most critical parts of successful data mining solutions.

8 REFERENCES

ABB. 2014. ABB review - The Corporate Technical Journal. Special Report IEC 61850. [WWW] January 1, 2014. [Accessed on: April 15, 2014.]. Avalaible at: http://new.abb.com/fi

ABB. 2009. COM 600 Manuals. ABB Homepage. [WWW] February 13, 2009. [Accessed on: April 15, 2014.]. Available at: http://new.abb.com/fi

ABB. 2012. Feeder Protection and Control REF630 IEC. ABB - Product Guides. [WWW] September 18, 2012. [Accessed on: April 20, 2014.] . Avalaible at: http://new.abb.com/fi

ABB. 2013a. Product Guides: BLK Spring - Operated Mechanisms. ABB - Product Guides. [WWW] February 6, 2013. [Accessed on: April 20, 2014.]. Avalaible at: http://new.abb.com/fi

ABB. 2013b. Substation Automation Protection & Control. ABB Homepage. [WWW] April 9, 2013. [Accessed on: April 10, 2014.]. Avalaible at: http://new.abb.com/fi

Aro, Matti, et al. 2003. Suurjännitetekniikka. Jyväskylä : University Press Finland Otatieto, 2003. p. 520.

Automation In Power Distribution System: Present Status. Gaur, Neha and Gaur, Priyanka. 2012. II, Jersey : Journal of Engineering Research and Studies, 2012, Vol. III.

Bane, Andy and Bower, Paul K. 2013. How to Master Big Data for the Next-generation Utility. Electric Light & Power. [WWW] May 1, 2013. [Accessed on: November 13, 2013.] http://www.elp.com/articles/powergrid_international/print/volume-18/issue-5/features/how-to-master-big-data-for-the-next-generation-utility.html.

Cadick, John. 1999. Condition Based Maintenance - How to Get Started. Garland : Cadick Corporation, 1999.

Elenia Ltd. 2014. Elenia: Electricity distibuition and Heating Company - About us. Elenia Homepage. [WWW] Elenia Ltd., January 1, 2014. [Accessed on: April 20, 2014.] http://www.elenia.fi/yritys/about.

Energiateollisuus Ry. 2012. Loppuvuoden sähkökatkoista kärsi 570 000 asiakasta. Energiateollisuus.fi. [WWW] Energiateollisuus Ry, January 19, 2012. [Accessed on: May 13, 2014.] http://energia.fi/ajankohtaista/lehdistotiedotteet/loppuvuoden-sahkokatkoista-karsi-570-000-asiakasta.

Haarla, Liisa and Elovaara, Jarmo. 2011. Sähköverkot II - Verkon suunnittelu, järjestelmät ja laitteet. Tallinna : Helsinki University Press, 2011.

Hurwitz, Judith, et al. 2013. Big Data for Dummies. Hoboken : John Wiley & Sons, Inc., 2013.

IEC61850. 2009. Communication Networks and Systems for Power Utility Automation. Geneva : International Electrotechnical Commission, 2009. Vol. VI.

IEC81346. 2009. Industrial Systems, Installations and Equipment and Industrial Products - Structuring Principles and Reference Designations. IEC 81346. Geneve : International Electrotechnical Committee, 2009. Vol. II.

Lakervi, Erkki and Partanen, Jarmo. 2008. Sähkönjakelutekniikka. Helsinki : Helsinki University, 2008. p. 285.

Lukszo, Sofia and Deconinck, Geert. 2010. Securing Electricity Supply in the Cyber Age. Virginia : Delft University of Technology, 2010. p. 193.

Manyika, James, et al. 2011. Big Data: The Next Frontier for Innovation, Competition, and productivity. New York City : McKinsey Global Institute, 2011.

Normandeau, Kevin. 2013. Beyond Volume, Variety and Velocity is the Issue of Big Data Veracity. Inside Big Data. [WWW] September 12, 2013. [Accessed on: November 10, 2013.] http://inside-bigdata.com/2013/09/12/beyond-volume-variety-velocity-issue-big-data-veracity/.

Northcote-Green, James and Wilson, Robert. 2007. Control and Automation of Electrical Pwer Distribtution Systems. Miami : Taylor & Francis Group, 2007. p. 465.

Rajaraman, Anand, Leskovec, Jure and Ullman, Jefferey D. 2013. Mining of Massive Datasets. Cambridge : Cambridge University Press, 2013. p. 511.

Russom, Philip. 2011. Big Data Analytics. Chicago : TDWI Research, 2011.

Ryan, Bob and Shank, Charles. 2005. IEEE Standard Common Format for Trainsient Data Exchange (COMTRADE) for Power Systems. New York : The Institute of Electrical and Electrics Engineers Inc., 2005.

Salo, Immo. 2013. Big Data - Tiedon Vallankumous. Jyväskylä : Docendo, 2013. p. 147.

Srikanth, R P. 2013. Big Data analytics can help in reducing power shortage in India. Information Week. [WWW] October 21, 2013. [Accessed on: November 2013, 13.] http://www.informationweek.in/informationweek/interviews/285791/analytics-help-reducing-power-shortage-india.

Strang, William and Pond, Jeff. 2006. Considerations for Use Of Disturbance Recorders. s.l. : System Protection Subcommittee of the Power System, 2006.

US Department of Energy Facilities. 2013. Economic Benefits of increasing Electric Grid Resilience to Weather Outages. Energy.gov. [WWW] August 1, 2013. [Accessed on: May 1, 2014.] http://energy.gov/downloads/economic-benefits-increasing-electric-grid-resilience-weather-outages.