systems

MDPI

Article

An Integral Model to Provide Reactive and Proactive Services in an Academic CSIRT Based on Business Intelligence

Walter Fuertes [1,*], Francisco Reyes [1,*], Paúl Valladares [1], Freddy Tapia [1], Theofilos Toulkeridis [1,*] and Ernesto Pérez [2]

[1] Department of Computer Sciences, Universidad de las Fuerzas Armadas ESPE, Av. General Rumiñahui, S/N, Sangolqui 171-5-231-B, Ecuador; bpvalladares@espe.edu.ec (P.V.); fmtapia@espe.edu.ec (F.T.)

[2] Corporación Ecuatoriana para el Desarrollo de la Investigación y la Academia, La Condamine 12-109, Ecuador; ernesto.perez@cedia.org.ec

* Correspondence: wmfuertes@espe.edu.ec (W.F.); fxreyes@espe.edu.ec (F.R.); ttoulkeridis@espe.edu.ec (T.T.); Tel.: +593-2-3989-400 (ext. 1906) (W.F.)

Received: 1 October 2017; Accepted: 14 November 2017; Published: 23 November 2017

Abstract: Cyber-attacks have increased in severity and complexity. That requires, that the CERT/CSIRT research and develops new security tools. Therefore, our study focuses on the design of an integral model based on Business Intelligence (BI), which provides reactive and proactive services in a CSIRT, in order to alert and reduce any suspicious or malicious activity on information systems and data networks. To achieve this purpose, a solution has been assembled, that generates information stores, being compiled from a continuous network transmission of several internal and external sources of an organization. However, it contemplates a data warehouse, which is focused like a correlator of logs, being formed by the information of feeds with diverse formats. Furthermore, it analyzed attack detection and port scanning, obtained from sensors such as Snort and Passive Vulnerability Scanner, which are stored in a database, where the logs have been generated by the systems. With such inputs, we designed and implemented BI systems using the phases of the Ralph Kimball methodology, ETL and OLAP processes. In addition, a software application has been implemented using the SCRUM methodology, which allowed to link the obtained logs to the BI system for visualization in dynamic dashboards, with the purpose of generating early alerts and constructing complex queries using the user interface through objects structures. The results demonstrate, that this solution has generated early warnings based on the level of criticality and level of sensitivity of malware and vulnerabilities as well as monitoring efficiency, increasing the level of security of member institutions.

Keywords: CSIRT; data warehouse; cyber-attacks; ETL; OLAPS; Kimball; SCRUM; vulnerability analysis; Incident Managers

1. Introduction

According to International Telecommunication Union (ITU) [1], Information Security need to be managed from an integrative perspective. The lack of coherent and rigorous management has led to government institutes, companies and universities to be targeted by cyber-attacks [2]. In addition, the speed with which an institution has been able to recognize, analyze and respond to an incident will decrease recovery costs and reduce the potentially generated damage [3].

For the treatment of incidents, as reported by NIST [4], exists the Computer Security Incident Response Team (CSIRT) or the Computer Emergency Readiness Team (CERT), which provide the services of: analysis, coordination, support and response to computer security incidents [4]. However,

when dealing with an academic CSIRT, the conglomerate of information collected may be a cause of partial or non-compliance of these services.

Based on this context and particularly because Cyber-attacks have increased in severity and complexity [5], our study has been focused on the design of an integral model based on Business Intelligence (BI), which provides reactive and proactive services in an Academic CSIRT (A-CSIRT), in order to alert and reduce any suspicious or malicious activity on information systems. To accomplish such purpose, a system has been assembled, that generates information stores, being compiled from a continuous network transmission of several internal and external sources of an organization. This research has been divided into two modules. The first Module, has been intended mainly for A-CISRT Incident Managers, attempts to improve the visualization of events such as malicious code, which will enable member institutions to take actions and decisions. Module two, has been aimed to analyze vulnerabilities and to recognize real-time attacks, through the development of a BI system through SCRUM Agile Methodology [6], specifically for A-CSIRT members, allowing real-time visualization of the problems that may occur in the network to establish control measures.

These solutions consider a data warehouse (DW), which is focused like a log-correlator, being formed by the information of feeds with diverse formats. Additionally, it is able to analyze an attack detection and port scanning, obtained from sensors such as Snort [7] and Passive Vulnerability Scanner (PVS) [8]. These are stored in a database, where the logs have been generated by the systems. With such inputs, we designed and implemented BI systems using the phases of the Ralph Kimball methodology [9], Extract, Transform, Load (ETL) and On-Line Analytical Processing (OLAP) processes. Furthermore, a software application has been implemented using the SCRUM, which allowed to link the obtained logs to the BI system for visualization in dynamic dashboards, with the purpose of generating early alerts and constructing complex queries using the user interface through objects structures.

Hereby, the main contributions of this project have been to propose new extracting information algorithms based on DW, BI, ETL and OLAP from diverse data sources with different formats (i.e., Internet feeds, Snort and PVS), in order to homogenize them. Furthermore, we adapted the methodology of Ralph Kimball, which has been created exclusively for the analysis of performance to obtain business decisions, now addressed into the Incident Tracking System as part of preventive and proactive services of an A-CSIRT.

The remainder of the article has been divided into four further sections. In Section 2, we present studies related to our research, while Section 3 presents the design and setup of the research, which links the specification of requirements, the design used in the solution and its implementation. Later in Section 4, we demonstrate the evaluation of the results. Finally, Section 5 lists the conclusions and describes potential future research.

2. Related Work

The Information Technologies (IT) have optimized the ways of accessing and managing information. A clear example of this has been BI, which has become a suitable channel for optimizing decision support systems (DSSs). The new trends and risks are based on scenarios and in real time, where the business processes need to be executed under certain norms and good practices, which guarantee the reliability and availability of information [10,11]. As for the growth and sophistication of cyber-attacks, new scenarios have been presented in the face of certain security incidents, where control and reduction of response times are key aspects [12,13]. Therefore, it needs to be emphasized, that the technologies associated with BI have been growing and gaining improved positioning in certain techniques such as Data Mining and DW. That has been based on the fact that the current BI systems are linked to the structure of the data, which allows data extraction and analysis.

The construction of a DW/BI solution appears to be complex, which has led Ralph Kimball to propose a methodology [9], which simplifies this complexity. This starts with an understanding of the business requirements and is followed by how the organization has been able to be valued through

obtained and/or generated information. In order to achieve this purpose, the initial activity may be initiated with the determination of the information priorities, defining which business processes produce data, the scope of the project, the level of granularity to be obtained, the data dimensions and the facts to be analyzed, together with the analytical processes that need to be executed with such data (i.e., design and implementation of a database model and ETL).

However, by arranging all the aforementioned aspects towards a globalized business environment, a management and control software will be required, in order to articulate and implement the present investigation, which will be part of the universal principle of prevention [14]. Therefore, we propose to implement mechanisms and/or early warning services, which should prevent malicious activities. Such goals have been emphasized within a variety of case studies, such as: the elimination of vulnerabilities, which have been focused on the predictive aspects, low level inspection operations, (i.e., providing different attributes to the software) [15]; Trend analysis, behavior patterns and data mining techniques [16,17]; Issues with Cybersecurity, based on adequate knowledge management, governance and organizational psychology [18]; Adequate use of Information Security Managers, through systems protection, information security architecture and event management (SIEM) [19]; Intrusion Detection System (IDS), supported through Snort as a tool to generate a large number of alerts [20].

During an extensive literature review, only a very few studies have been encountered related to the design and development of a dimensional BI/DW data model to deploy early alerts in a CSIRT. The few studies, which have been considered to be related to this context, addressed better incident management techniques in CSIRTs. For a more detailed explanation, we have divided our argumentation into the following aspects.

There are novel techniques that have been used in CSIRT to improve the treatment of computer security incidents, such as the study proposed by Hellwig [21], which establishes new challenges for CERT-communication while Yang [15] proposes a novel vulnerability prediction approach based on the CERT-C secure coding standard. Bollinger [22] explained the development of a threat intelligence and incident detection strategy. Kruidhof [23] discussed the evolution of CSIRT due to trends in technology and society, while Bhatt [24] discussed how to structure SOCs around SIEM Systems. Osorno [25] presented an integrated incident management model, which combines elements of an incident life cycle, while Qian [26] presented a formal model of incident response capability-building during the operation transitioning. Finally, Belsis [27] presented a framework based on advanced database techniques for the correlation of incidents in order to keep track of a complete attack. Considering how information security may be aligned with business objectives, recent studies such as that by Elmellas [28] have used threat intelligence as a technical solution that includes a proactive security, while Grobler [29] provides guidance for the implementation of thread detection and incident recovery for SOCs and CSIRTs. Sharkov [30] presented a cyber-resilience management model for cyber risks and Mejía [31] described a proposal related to those content and security controls of the CSIRT website that need to be considered. In reference to studies that analyze the importance of Business Intelligence in risk management for institutions and for early warnings to prevent future disasters, the study proposed by Rajasekharaiah [13] presents a decision algorithm. Such algorithm has been employed for data Mining and to ensure data authenticity, with a Commutative RSA, Wu [32] providing a review of the state-of-the-art research in business intelligence used in risk management.

Finally, in relation to the benefits of Big Data Analytics and Business Intelligence to review security and privacy challenges in large data environments, Gahi [33] presented some available protection techniques and tracks that enable security and privacy in a malicious big data context, while Zuech [34] presented a background on Intrusion Detection and some big data implications and challenge. Mahmood [35] presented a research focusing on the application of Big Data Analytics techniques to cybersecurity. Jaramillo [36] presented a preventive, detective and corrective accountability and data risk management, based on the usage of control policies and a model-driven engineering, while finally Ahmad [37] proposed a double loop model for incident learning, in order to address potential systemic corrective action in such areas like risk assessment and policy development processes.

3. Materials and Methods

Our study has been leveraged in a conceptual framework, as illustrated in Figure 1. The research has been based on the methodology of Ralph Kimball [9,38] (1), which itself has been centered on the definition of requirements for the planning of the project (2). Furthermore, the design of the ETL programming (3) and the cohesion with the database has been established (4). Additionally, the OLAP cubes, the reports and Dashboards have been developed (5). Finally, we have designed and implemented a Web BI System, which interacts with refined data, focused on automating the reactive and proactive service of the Academic CSIRT (A-CSIRT). A brief description of the phases performed is explained below.

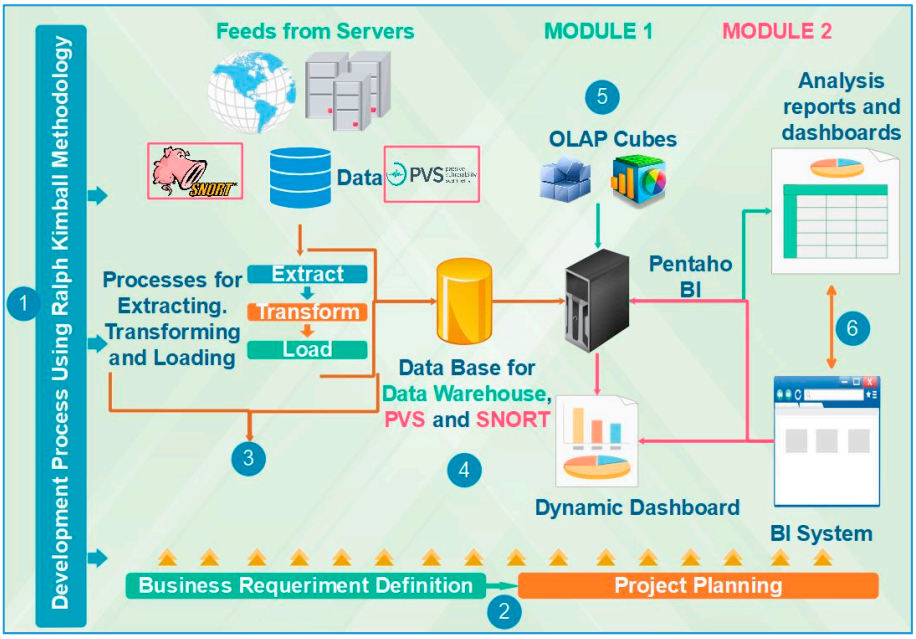


Figure 1. Illustration of the conceptual and methodological framework of this study, which reveals the foundations used to design and implement the BI system using the phases of the Ralph Kimball methodology, ETL and OLAP processes. In addition, a Web software application has been implemented using the SCRUM methodology.

3.1. Project Planning

At this stage, the objectives, the scope and the purpose of the project, have been determined. The main objective of this project has been to improve reactive and proactive services for an academic CSIRT, establishing an internal and external solution. The scope focuses on providing two modules, of which the first has been to provide all the incident and alert information reported to an Academic-CSIRT in a single dimensional database for incident trends analysis, while the second Module, aimed towards the generation of a BI system, which assembles two complementary solutions (i.e., Snort and Passive Vulnerability Scanner) in order to handle alerts and warnings, vulnerability analysis, leading to help prepare, protect and secure systems of future attacks, or computer security incidents [39].

In order to accomplish this aim of the project, the following techniques have been used: (1) ETL processes, which allowed to extract information from different data sources (feeds), in order to generate a desired structure through the execution of commands, enabling the extraction from different data sources. (2) Dimensional database, in which the information has been debugged and integrated from one or more sources, in order to be able to process and analyze through different perspectives; (3) OLAP, which allowed the interactive analysis of large volumes of information; (4) Dashboards, which have been visual summaries of business information, providing an enlarged and improved understanding

of the global conditions of the business through the use of metrics and key performance indicators; and (5) Reports, which generated a better visualization of data, in order to keep present administrators informed about the security status of their business.

3.2. Definition of Requirements

The Academic CSIRT aims to: (1) Support members to implement preventive and proactive measures, with the intention to reduce the risks of computer incidents; (2) Support the community, in order to respond to appearing incidents; (3) Carry out monitoring and issuing of alerts; (4) Support for computer incidents. With the purpose to accomplish this, the Academic CSIRT has: (i) *reactive services*, which have been designed to generate alerts and warnings, as well as provide adequate management of incidents and vulnerabilities; (ii) *proactive service*, with the objective of providing technological surveillance, security assessments, development of security tools, intrusion detection service, security-related dissemination, configuration and maintenance of security tools, applications and infrastructure; (iii) Security *Quality Management Services*, which establish security consultancies, security awareness, in addition to security education and training [2]. However, due to the amount of malicious event data, by which the Academic CSIRT handles non-automatically and together with the need to provide to their members with a tool, which allows them to effectively analyze the traffic of their respective network, this project presents a solution by using BI, in order to generate incident management services based on security presentation and data analysis.

3.3. Design and Implementation of the Solution

The present study has been based on the life cycle developed by Ralph Kimball [9] (see Figure 2), in which three different phases have been identified. The first phase refers to a selection of the applied hardware and software components. The second phase establishes the dimensional model and the elaboration of ETL processes. The third phase involved the specifications of the BI application taking into account the adaptation of the SCRUM methodology. Additionally, in order to accomplish an accurate development, Kimball established four basic principles: (1) Focus on the business, which identifies the business requirements and associated value to develop links with the business; (2) To build an adequate infrastructure, for which a high performance information base need to be designed, at which the business requirements will be reflected; (3) Deliveries in significant increments, which establishes an increase in terms of 6 to 12 months, resulting to a considerable evolution of the project; (4) Offer the complete solution, where all elements, which provide value to the users of the business are delivered. Based on the aforementioned considerations, the present project has been divided into two modules, of which their design and development will be detailed explained below.

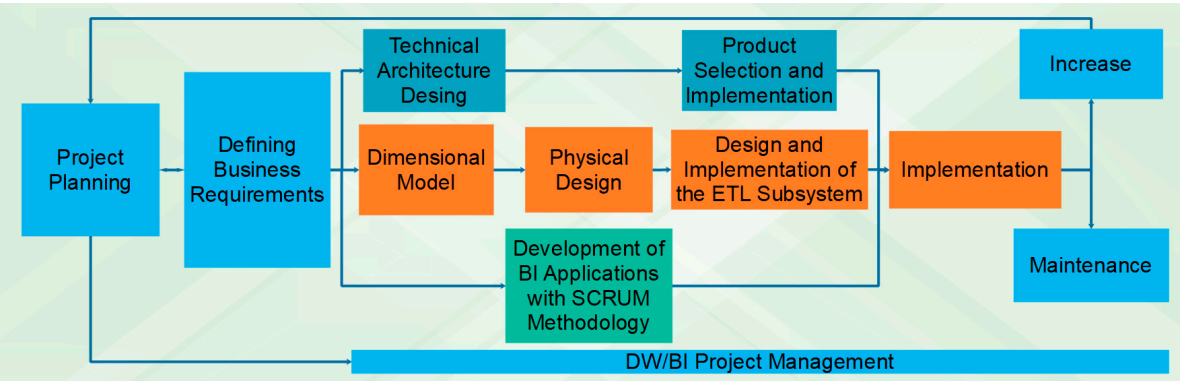


Figure 2. The Ralph Kimball Methodology Life Cycle assembled with the development of BI applications with Agile SCRUM. This illustrated combination details the different phases for the accurate process with which we have been able to streamline and establish a system being focused strictly on the needs of the CSIRT.

3.3.1. Data Sources

The first Module obtained its data from feeds, which received information from different sources such as: Zone-h, TurkBot, Clean MX Foundation, CERT.br, Team Cymru, Nessus, Netcraft, Shadowserver, etc. The A-CSIRT stores information of malware, electronic fraud, phishing, brute-force attacks and botnet activities, DDoS, Spam, among others. These malicious events appear on the public networks of institutions through the Internet. The main issue has been, that these information sources have a varied format, which depend on the source. This generates delay, incompatibility and difficulty in developing dashboards and reports, which usually present in a clear and orderly form the values that have been processed. Nonetheless, this module has been able to homogenize the origins of the sources through the use of extraction, transformation and load (ETL) algorithms.

The second Module collects information from two complementary systems: (1) Passive Vulnerability Scanner (PVS), which is a product of the Tenable Network Security company that stands out for analyzing all the occurring traffic in a certain network (i.e., by ports, services, hosts, applications, operating systems and vulnerabilities); and (2) Snort, which is an open source Intrusion Detection and Prevention System (IDS/IPS), which has been capable to generate real-time traffic analysis. PVS generates two information files, of which the primary is a plain text file focused on real-time events, while the second is an enriched Xml (.nessus) file that exposes the vulnerabilities of a particular network environment. Furthermore, Snort has the three following configuration modes: (i) sniffer, which illustrates the traffic in console mode; (ii) packet logger, which is applied to specify the directory, where the log files will be stored; (iii) Network Intrusion Detection System (NIDS), which is intended to specify the configuration directory [6]. For this study, Snort has been configured as a NIDS. In addition, Snort has the external system *Barnyard2*, which is responsible for transporting the generated logs to a relational database with a defined schema. The challenge has been to collect the information focused on obtaining data of interest, as well as to find a way to filter the information avoiding an overload of the database with irrelevant information.

3.3.2. Dimensional Data Modeling

According to Kimball [9,38], a dimensional data model is a database design technique, which is intended to support end-user queries. For the first Module, the generation of reports and dashboards, has been a complex task, as it deals with data with a high variety of different formats. Therefore, in order to correctly handle the attached information to the needs of the A-CSIRT, it has been necessary to obtain the data in a standardized method, for which we designed a format with the intention to store the recorded events. This format consists of each of the dimensions. Dimensions and fact tables are illustrated in Figure 3.

The *dimensions* represent the different points of view through which the information has been analyzed. Additionally, they are conformed by different attributes that have been analyzed and structured in a hierarchical way. On the other hand, the *facts* represent business indicators, which establish the direction of the analysis of the dimensions. The facts are composed of the indicators associated with a certain business process and the keys of the involved dimensions. For the second Module, by means of the implementation of ETL algorithms, the generation of flat databases (see Figure 4a) has been achieved for the PVS system, by storing the processed information in specific fields. These bases are intended to generate a real-time event analysis and vulnerabilities of a given network. As previously mentioned, Snort has a system that allows the transmission of collected information to a database. Therefore, the *Barnyard2 system* provides a model, which has been edited in order to obtain a relational process (see Figure 4b).

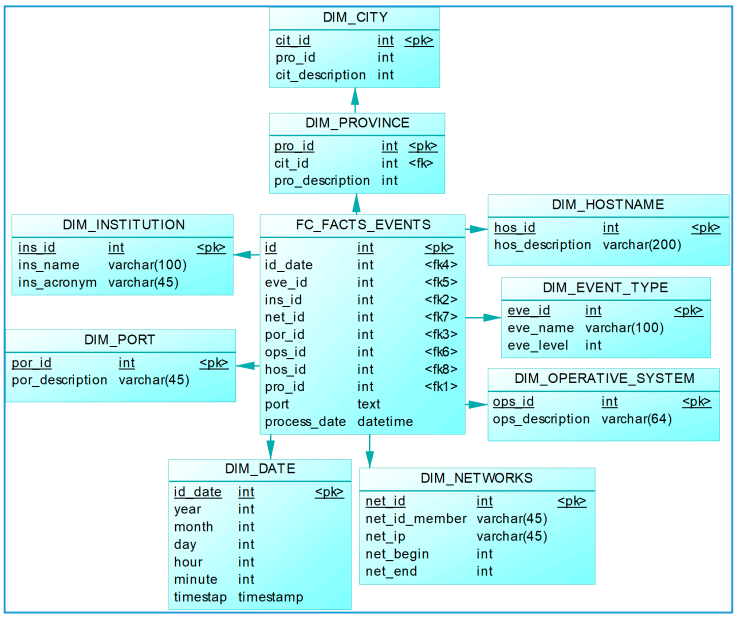
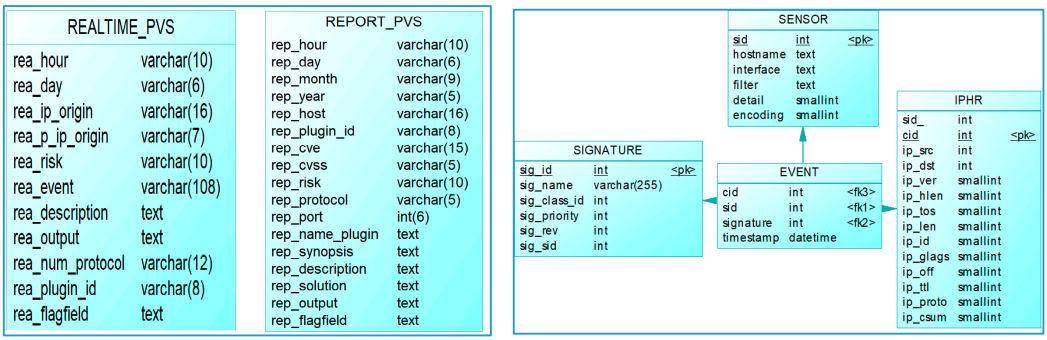


Figure 3. The Dimensional Data Base. The dimensional model for BI establishes a star topology, which allows a correct relationship between the facts and its associated dimensions.



(**a**) (**b**)

Figure 4. (**a**) Although the PVS is a proprietary system, the deployment of information is limited. Therefore, the flat databases for the PVS system provide timely and relevant information for vulnerabilities and events in real time; (**b**) The database for Snort is composed of the most significant tables of the original non-relational model of the Barnyard2 system.

3.3.3. ETL Processes for Capturing and Data Filtering

For Kimball, ETL processes have been a key part of generating a BI project, since they establish the flow of data and the structure that will be later processed. However, it is described as a process that demands greater use of efforts and resources, being costly, time-consuming and complicated to generate [38]. This process has been responsible for the extraction of the information from the source systems, establishing a consistency and quality of the data, homogenizing the information from different sources enabling a parameterized use (i.e., applying transformations flow). Finally, it generates output data that are able to be used by different information analysis tools. When the Incident Managers apply a dimensional database, then the ETL systems will be responsible for the load generation, which is known as *data stage*. This process accounts for 70% of the resources allocated to the maintenance and development of the project, generating a significant value for the data. This leads to the assumption, that poorly defined or poorly validated processes will be able to adversely affect a well-designed BI system.

ETL jobs have been designed in order to automate the loading of information into the database, which have one or more transformations. The difference is that this engagement does not handle

records, rather than a sequence of tasks, besides the fact that the transformations have a parallel execution, whereas the duty may consider a repetition loop. For the first Module, a data warehouse has been generated, which has as a main component a dimensional database with star topology. Kimball presented the following four phases for the development of a dimensional database: (1) selection of the business process, (2) definition of data granularity, (3) selection of analysis dimensions, (4) identification of facts or business indicators. In addition, it establishes within the design of the database that the information need to be treated to carry all the facts to the maximum level of granularity [40]. This work has several transformations (Figure 5). In principle, we have the dimensions (Job Dimensions) that are able to be visualized in Figure 6, which are intended to fill the different dimensions. Afterwards, we proceeded to extract the information from the different data sources to generate a defined format. This step has been fundamental, as it generates a correct data insertion in the fact table. With the consideration that each event has a different output, a case city has been developed by means of SQL algorithms, in charge of validating the province and the city in a format that purifies the data for a better presentation to the user.

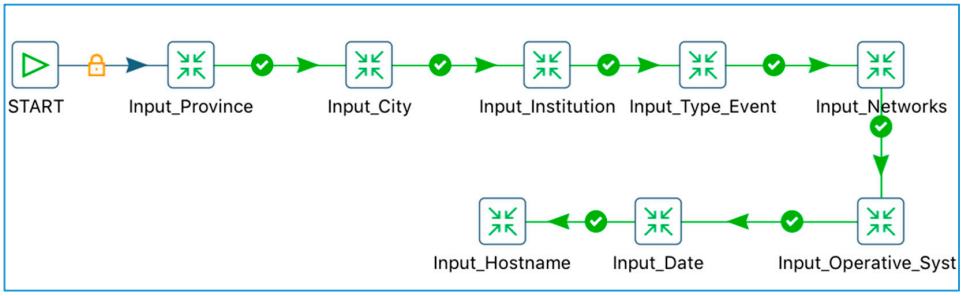


Figure 5. Job for data warehouse Load. Pentaho's data integration system allows a robust generation of processes that lead to consecutive jobs and transformations.

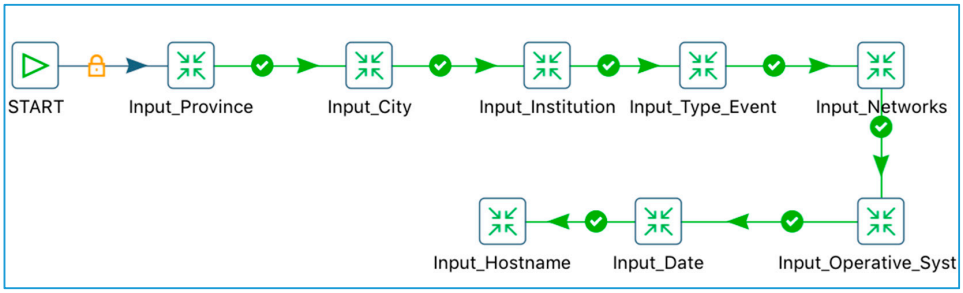


Figure 6. The Dimensional Load Work, where the workflow is set for one run every 24 h.

The CSIRT has a case classification, which is a document issued by the Forum of Incident Response and Security Teams (FIRST) [41]. This describes the rules that need to be followed enabling the personnel in charge to classify the category of cases, the level of sensitivity and the level of criticality for each case of the CSIRT. The classification document is considered for cases to be entered into the dimensional database. Based on information from FIRST and CSIRT Incident Managers, events are able to be categorized in three ways: (1) *Level one*, are incidents that affect critical systems or information with a negative impact on customers or their earnings; (2) *Level two*, are incidents that without impacting customers or profits, do affect non-critical systems; (3) *Level three*, are possible incidents for non-critical systems. This study provides valuable information, which alerts institutions about their security level. Additionally, it provides to the CSIRT potential actions of improvement.

For the second Module, ETL processes have been developed focused on each system. In this way, three ETL solutions has been developed for PVS, which have been focused on vulnerability analysis, real-time activities and risk filtering. Figure 7 demonstrates the transformation that has been

generated for the real-time events, highlighting the generation of flows that support to optimize the processed packets, while avoiding a data redundancy or delay in the process. Afterwards, it is possible to generate a job that executes the previously mentioned transformation, highlighting the notification of possible errors presented in the execution of the transformation via e-mail. This occurs during a process that allows to filter the risks due to established parameters.

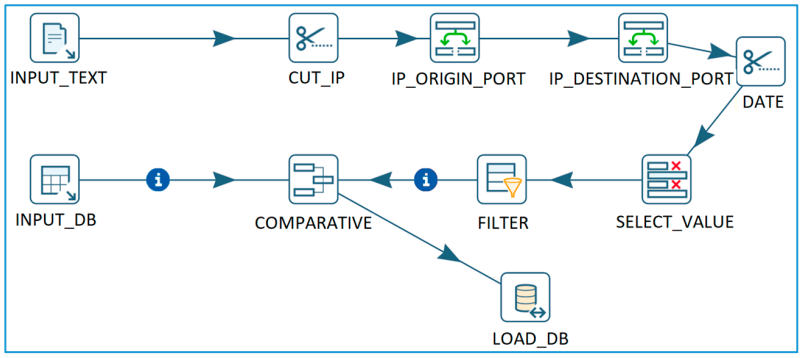


Figure 7. Transformation from a flat file. The filter of the transformation supports the elimination of data that are able to generate drawbacks in the data collection.

For vulnerability reports, the transformation generated to process vulnerabilities collected from enriched XML (.nessus) has been highlighted in Figure 8. Within this programming, a control of the events has been established, specifying a classification to be stored in the database. In Figure 9, the generated result has been captured, allowing the interaction with the transformation of vulnerabilities, having additionally a control of the flow through e-mail warning.

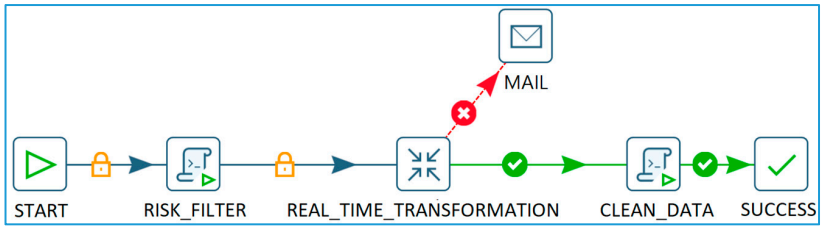


Figure 8. Schematic illustration of the algorithm of the generation of a filter with possible control of transformation errors. Pentaho's data integration tool allows the execution of shell scripts. For this case, the script has been programmed with regular expressions.

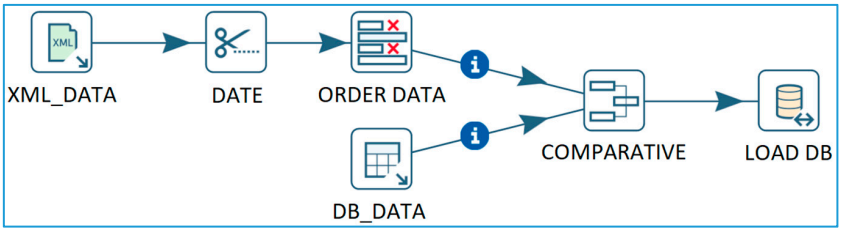


Figure 9. Transformation from an enriched xml. Pentaho's data integration tool allows to extract the information of an attribute or a specific node from an XML file. In addition, the comparative establishes a flag avoiding data redundancy.

The Snort system has a tool that automates the logs to a database. However, in order to generate a complement, an ETL process has been developed allowing the filtering of information, using parameters established in a file, achieving user relevant data generation (see Figures 10 and 11).

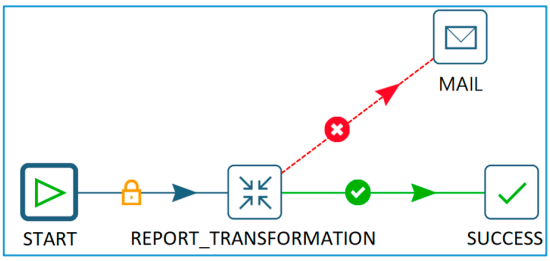


Figure 10. The data loading algorithm with defined time loop execution. Reports are set every 6 h with a repeated loop.

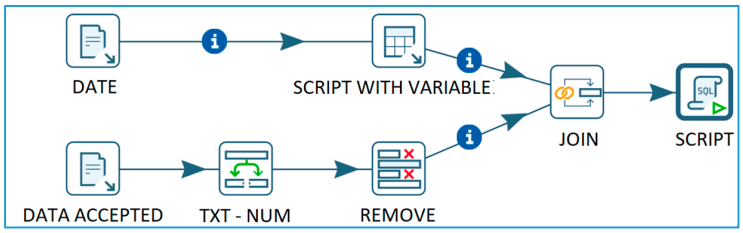


Figure 11. Transformation for filter generation from a .txt file. Pentaho's data integration tool allows to operate with the metadata of a document. This establishes the knowledge about the modification date of a certain document, hereby generating changes within the database.

3.3.4. Development of OLAP Cubes and Reports

On-Line Analytical Processing (OLAP) cubes interact only in the first Module of this project. Therefore, OLAP cubes have been structures of information that have been used to obtain data of interest for a business. Based on these cubes, queries are able to be generated in an organized and multidimensional way for further analysis.

OLAP cubes receive the information given by several data sources, from which it is possible to extract the relevant information with the help of algorithms specific to this research (Figure 12). Hereby, the main components are XML files with dimension schemes, which are able to use any text editing tool or a specialized tool in the generation of these files. One such tool has been the Pentaho Schema Workbench, which provides a fairly user-friendly graphical interface for building Mondrian schemes. In addition, it allows a publication on the BI server. However, for the generation of reports, the Pentaho Report Designer tool has been used, which is a set of open source tools that supports the construction of reports in various formats, accepting multiple data sources such as JDBC, Olap4J, Pentaho Analysis, Pentaho Data Integration, XML, etc., including system-defined metadata.

3.3.5. Custom BI Dashboards Development

According to Lynne [42], Dashboards are control panels, where a dense array of information need to be displayed in a reduced appearance. A BI dashboard is a data visualization that displays the current status of metrics and key performance indicators for an organization. For the present study, the Pentaho BI tool has been used, based on the guidelines of the Community Dashboard Framework (CDF) [43].

For the first Module, three dashboards have been developed, of which (1) the first focused on providing information regarding the behavior of an event classified by year and month of a given institution; (2) the second displayed the types of events that have been presented per year in an institution. Finally, (3) the third dashboard has been able to see the percentage of institutions according to the level of security (i.e., institutions with more incidents), by selecting the year and the month. Institutions have not been able to visualize them.

On the other hand, the second Module has dashboards focused on each component of the system. Thus, for PVS, two dashboards have been developed, of which the first focused on real-time

events and the second for vulnerability reports. Meanwhile, a dashboard has been created for Snort, which allows the visualization of the last detected events, complementing the real-time events stored in PVS. These dashboards stand out by the options of printing and exporting tables to Excel.

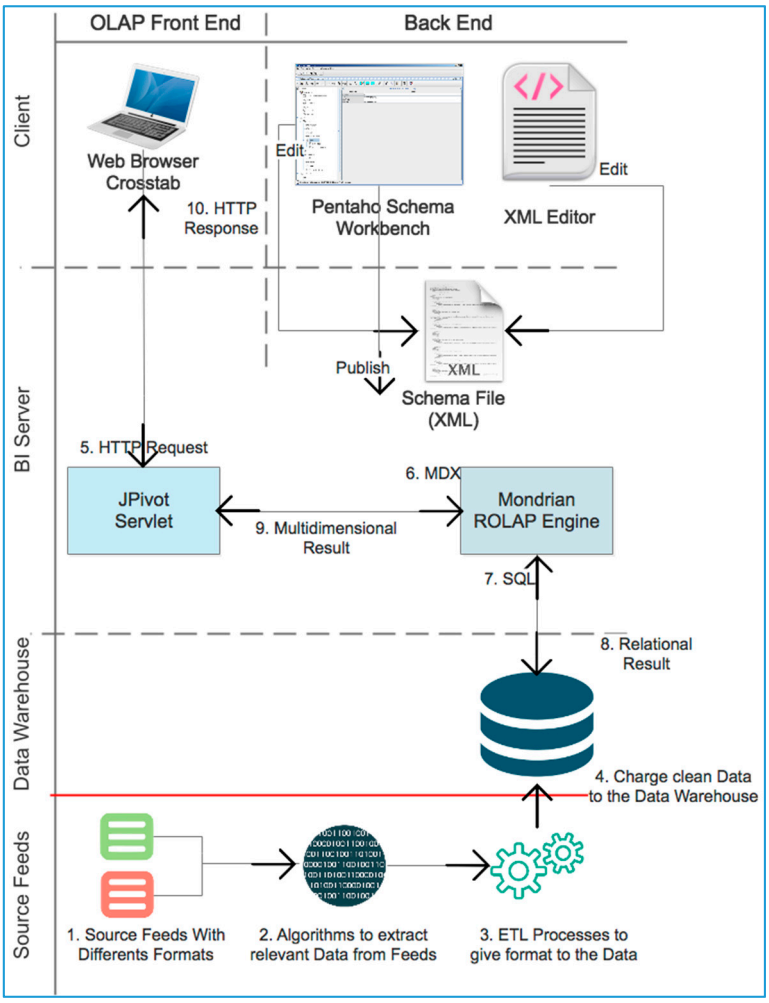


Figure 12. OLAP cubes, as adapted from Pentaho Analysis Services Architecture. Source: Adaptation from [14]. OLAP cubes are based on the correct schematization of the database. In addition, they generate data by MDX and SQL.

3.3.6. Development of the BI System

For the development of BI system, we have proceeded to use the SCRUM methodology, for the following reasons: (1) It is independent of technologies, which means that it is able to be coupled to different programming models; (2) it is focused on people, due to the fact, that the user and the developers have defined roles; (3) it provides rapid results; (4) the customer is active; (5) the time management is considered; (6) it is iterative; (7) and it responds to changes [44]. The development of the application takes place in the second Module of this study.

The BI systems use the Pentaho Plugin "Integrator," which generates a secure connection between the application and the Pentaho BI server. This application provides two roles defined as User and Administrator, which allows to view all dashboards within the application, with the difference that the Administrator is able to enter new users to the application. In addition, an access to Pentaho is generated with the same benefits, so that the role User is able to perform only tasks that have been assigned by the Administrator.

3.3.7. Preliminary Discussion

In spite of the aforementioned, this has not been a modest implementation of ideas by Kimball, rather than an adequate adaptation combined with SCRUM, which allowed to resolve certain limitations faced by the BI projects of computer networks security that the traditional software development route has not been able to face. These restrictions related to the validity of the data, specified the different formats, unrelated feeds, assorted the style of development, adverted of several cyber-attacks and additionally allowed to a change in the types of unit tests, which have been needed by the domain.

However, Kimball established a series of steps for the development of BI solutions, with a bottom-up approach, where the exhaustive analysis and the successive development of specific parts over the entire solution do predominate at the same time [40]. Nonetheless, the development of a BI solution oriented to Cybersecurity may not be something static and definitive, since not only the business needs change, or the frequency and hostility of attacks on the networks but it certainly does also require advanced technology. Due to this context, the development of the BI solution needs the means that allow its updating and modification without significantly affecting the business and the end users. Furthermore, a BI solution requires a project control system, which lacks the rigidity of classical project management, the difficulties and ambiguities of which appear from the scope of the project.

Even so, BI projects have a tendency to be complex, consuming time and computational resources. This may lead to an undesired delay in the delivery of the solution. The application of agile development has been one of the alternative methodologies, which have been studied in order to support teams to accelerate the delivery of commercial value. SCRUM has been one of the agile methodologies, which focuses on project management and allows iterative and incremental development. According to Mulder [45], "The goal of SCRUM is to deliver as much quality software as possible within a series of short time boxes called Sprints, which lasts about a month. SCRUM is characterized by short, intensive, daily meetings of every person on a software project." The central idea of SCRUM has been that the requirements have been more likely to change during the development process, which creates the development process to be complex and unpredictable.

Based on the described considerations, we have assembled to solve the detected limitations in this study, by using the SCRUM agile methodology with Kimball. Furthermore, in [46,47], the possibility of developing an agile BI solution with SCRUM had already been explored. Specifically, in the current study, SCRUM allowed the facility to interact directly with the client, which streamlined the development process that has been focused on the needs of the CSIRT. Additionally, this has been checked at all times by the client, allowing to reduce the time required to changes in the case when errors may have been detected. This has been fulfilled coherently with the following principles of the agile manifesto [48]: "(1) Our highest priority has been to satisfy the customer through early and continuous delivery of valuable software; (2) Welcome changing requirements, even when late in development. Agile processes harness change for the customer's competitive advantage; (3) Deliver working software frequently, from a couple of weeks to a couple of months, with a preference to the shorter timescale; (4) Business people and developers need to work together throughout the project; (6) The most efficient and effective method of conveying information to and within a development team has been face-to-face conversation; (7) Working software has been the primary measure of progress; (9) Continuous attention to technical excellence and good design enhancements agility; (11) The best architectures, requirements and designs emerge from self-organizing teams."

SCRUM has been used to minimize the risks during the implementation of the project but in a collaborative manner with the technical staff of the CSIRT. In addition, SCRUM has been applied in order to generate only the necessary documents for the configuration and maintenance of the application, thus achieving to give decision support.

3.3.8. Proposal to Permit Scalability in a Scenario of Significant Data Growth

A new proposal has been presented, which would be applied to implement a model oriented to mass storage of information (Figure 13). This scheme would be focused on the case of a potential

exponential growth of data in the future. In the current proposed model, a Big Data analysis tool known as Apache Storm [49] has been highlighted, as Storm generates a processing of unlimited data flows. This stands out mainly for being a distributed and real-time open source computing system. Furthermore, it has been applied to an ETL process of the information and stored in a relational database, so that the BI flow of the system is able to be generated. Storm allows reliable processing of large volumes of data in analytics, distributed RPC, ETL processes, among others. Storm lacks a process with an origin and an end, as the system has been based on the construction of Big Data topologies for its transformation and analysis within a continuous process of constant input of information. Based on the described reasons, Storm has become a firm system of Big Data analysis, being a complex event processing system (CEP) [50]. These types of solutions have been those that allow companies to respond to the arrival of data promptly and continuously, collecting it in real time with sensors for millions of comments generated by social networks, bank transfers and other origins.

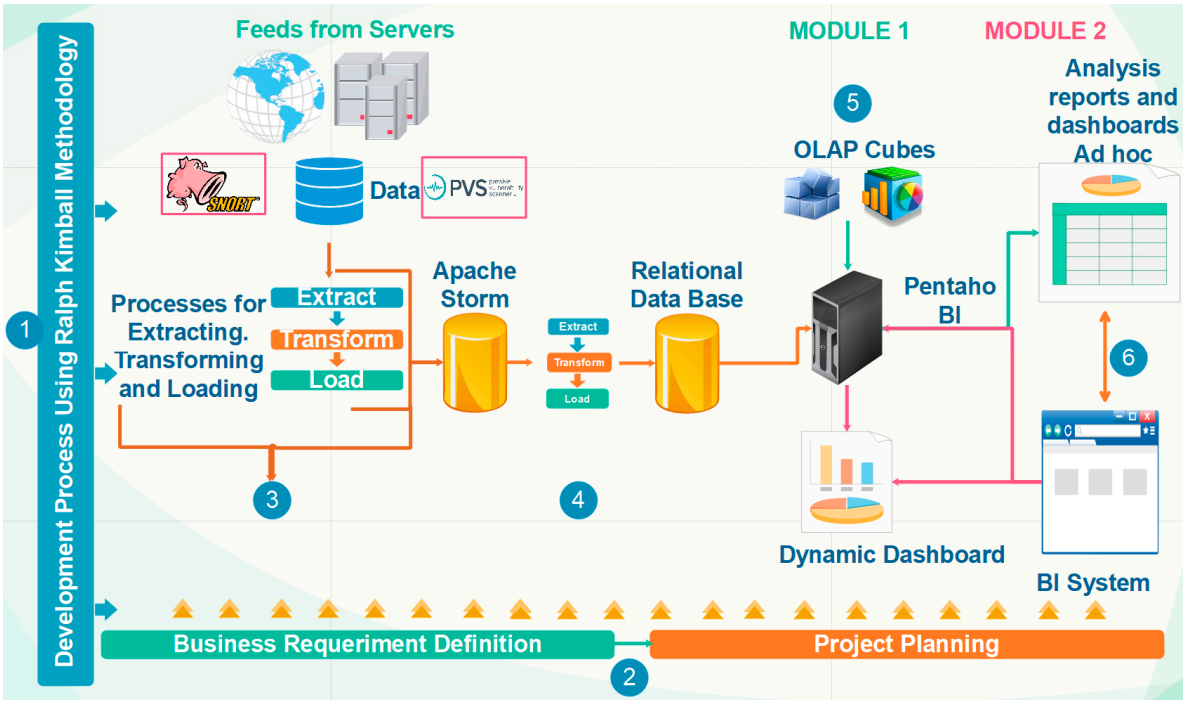


Figure 13. A schematic illustration of the proposal to generate scalability using Big Data techniques. The proposed system includes the use of Apache Storm to centralize the obtained information, streamlining the process in case of the appearance of scenarios which include a large amount of data. Therefore, lots of data may be stored, processed and relevant data obtained from it, which will subsequently be stored in a corresponding database.

For developers, Storm reflects a special interest for several reasons, as it has been able to be used in several programming languages, offering compatibility with components and applications written in several languages such as Java, C #, Python, Scala, Perl or PHP. Additionally, it has been also scalable, fault tolerant, easy to install and operative.

4. Results

4.1. Proof of Concept

For the proof of concept, consistency tests have been performed, with which it has been possible to determine that the information entered in each database has been valid corresponding to its sources avoiding data redundancy or inconsistency. Therefore, a comparison has been performed between the

database and the sources of information. In addition, a verification has been established, where the created ETL processes ensure its correct behavior.

The proof of concept has been conceived in two differenced cases, being for the A-CSIRT Incident Managers and for the A-CSIRT members. For the first case (see Figure 14), a CentOS version 7 server with the Pentaho BI and MYSQL Database tool has been installed, which involved the dimensional model. The Platform Manager has been responsible for creating, editing or executing the ETL processes, through the generation by the Schema Workbench of new OLAP cubes or new reports by the Pentaho Report Designer.

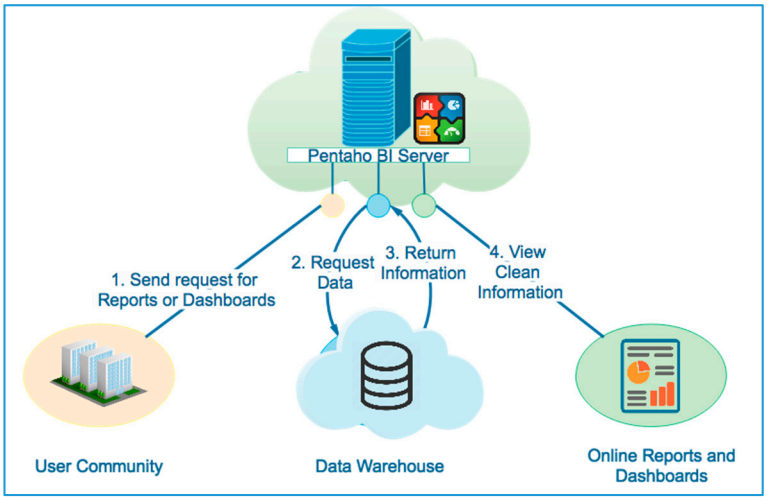


Figure 14. The Module One with the Proof of Concept. Module one has been focused on the general deployment of information, being daily updated.

For the second case (see Figure 15) a CentOS version 7 server has been installed, with all the components involved in this phase (Snort, PVS, MYSQL, Pentaho BI and BI System). Through the BI system, a request has been generated to the BI platform of Pentaho, which establishes two paths. The first path allows the Administrator to be able to edit the dashboards and the second has been focused on the deployment of the dashboard in the BI system for its particular impression or export to Excel.

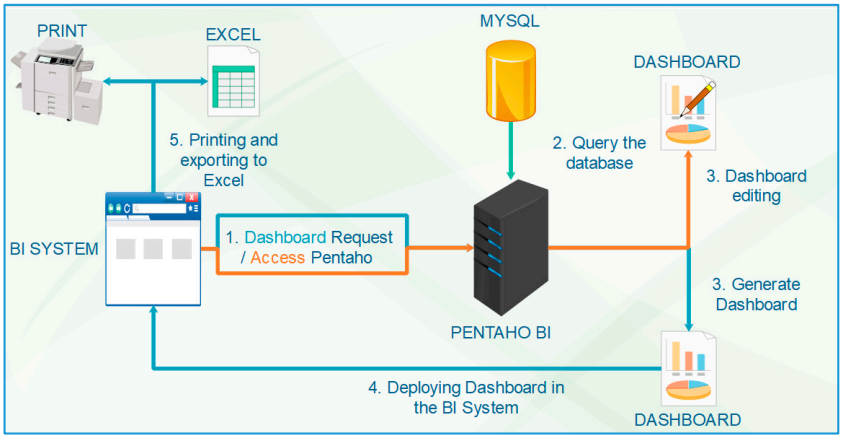
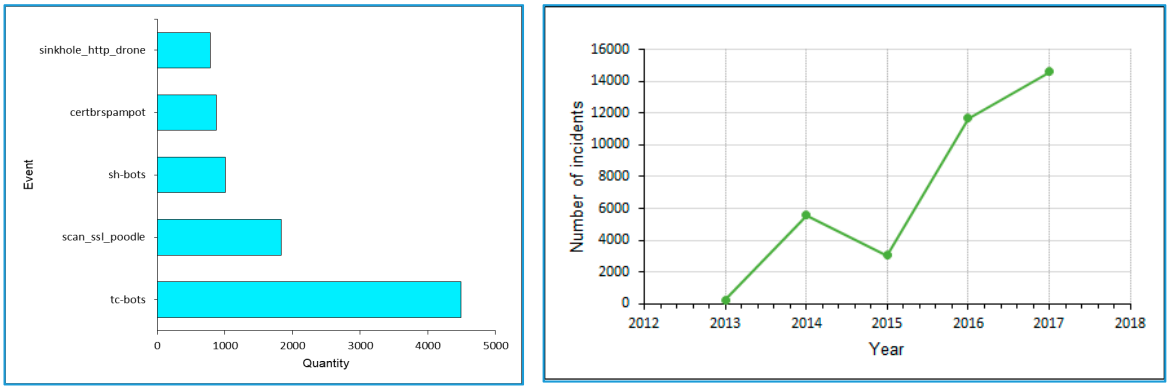


Figure 15. The proof of concept of module two. This module conceives the data management in real time. Therefore, the ETL processes capture a greater use of the resources for its continuous execution.

4.2. Evaluation Results

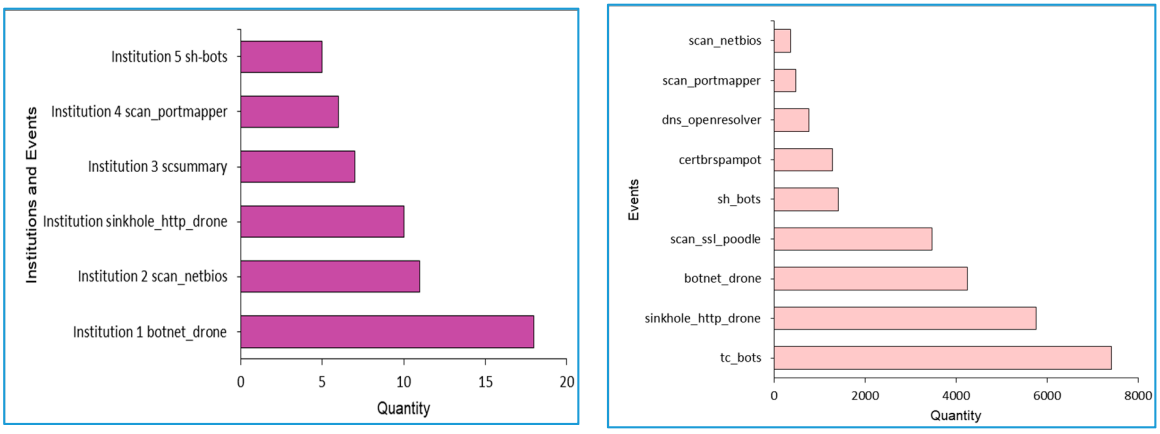
For Module One, Figure 16a presents the most representative events of the year, which supports A-CSIRT to generate corrective warnings that need to be followed by members in order to guarantee a better security. The total number of incidents per year has been illustrated in Figure 16b, demonstrating the obtained evolution in the elapsed time of several years.



(**a**) (**b**)

Figure 16. (**a**) Most occurred events in 2016; (**b**) Total number of incidents per year.

Figure 17a illustrates the events, which needed most time to be solved in 2016. This allows the A-CSIRT to generate a better advice to the institutions. Figure 17b reflects the most generated type of event. Bots are software applications, which run automated scripts over the Internet.

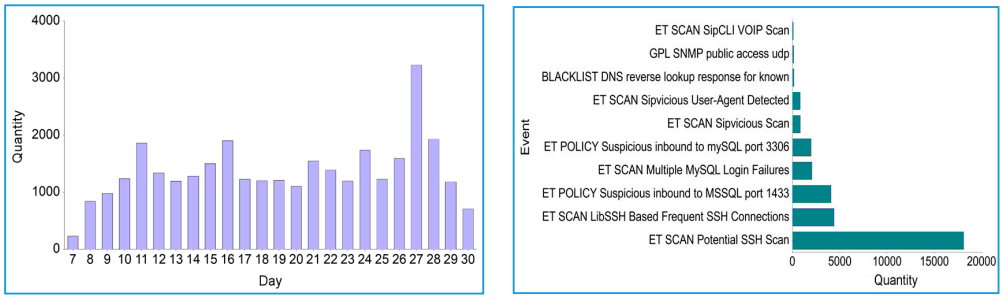


(**a**) (**b**)

Figure 17. (**a**) Most delayed events to be solved; (**b**) Number of events by type.

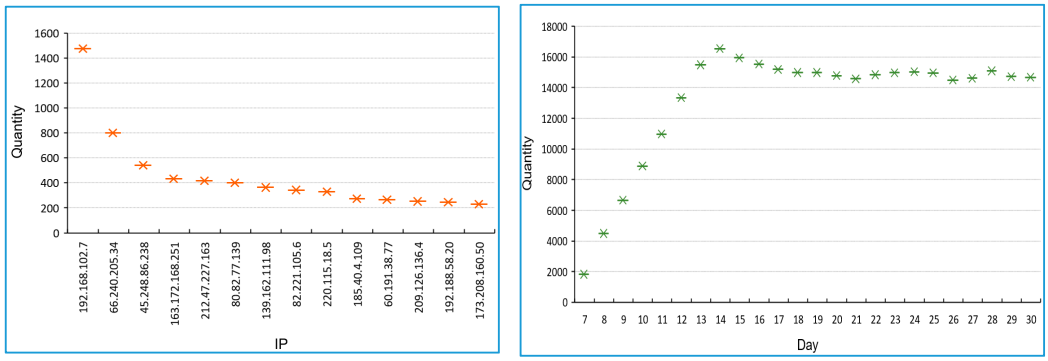
For Module Two, the application has been executed in the time period within 7 to 30 June 2017. Clearly, the variation of events is lacking, except for 27 June, where a significant elevation occurred. However, based on its origin (i.e., potential SSH scan), similar values to the initial ones have been reached within the following days (see Figure 18a,b).

Consequently, an exponential value has been encountered in relation to the detected vulnerabilities within the IP addresses. This resulted due to open ports and services, which are unable to be closed during operating regulations. It has been also possible to visualize an increase of vulnerabilities registered on day 14th, followed by a variable flow of information in the succeeding days. This event has been originated by the increase of the host's analysis (see Figure 19a,b).



(**a**) (**b**)

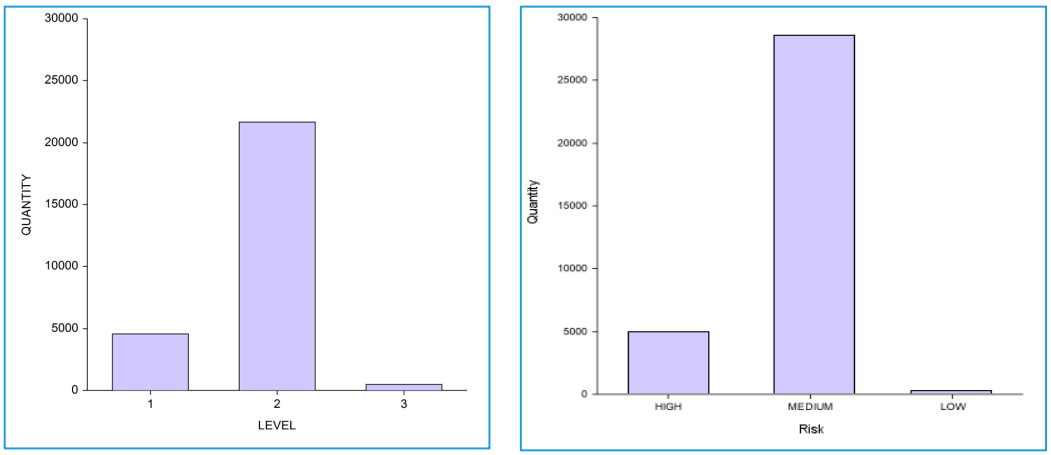
Figure 18. (**a**) Evaluation of events per day; (**b**) Evaluation type of event and frequency.



(**a**) (**b**)

Figure 19. (**a**) Evaluation of vulnerability per host; (**b**) Vulnerability assessment per day.

Finally, a conclusive analysis has been derived from Module 1, representing in Figure 20a the degree of criticality and sensitivity. There, it appears that there is a high amount of level 2, which does not generate a chronic alarm. Nonetheless, this will need a more detailed evaluation in order to be resolved. However, a high number of alerts cataloged as Medium-ranged risk, have been derived from Module Two as illustrated in Figure 20b. This means, that the administrator faces in the network a permanent increase of vulnerabilities. Therefore, both modules represent a considerable improvement for the A-CSIRT, as they provide an overview of the current security level. In addition, a permanent network monitoring identifies immediate problems, malware and vulnerabilities.



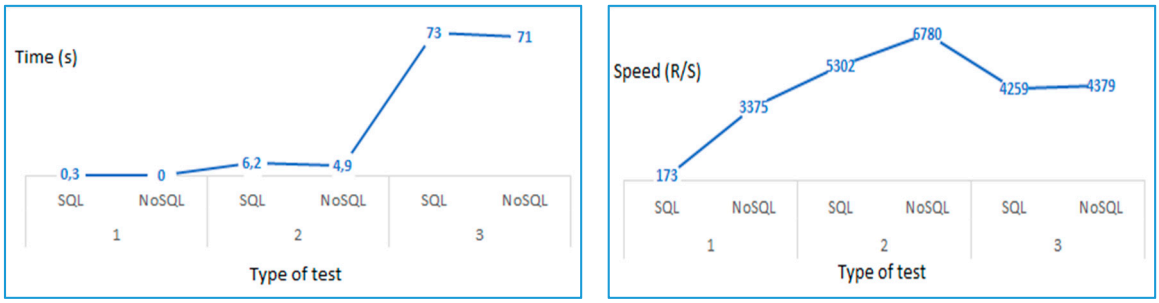
(**a**) (**b**)

Figure 20. (**a**) Number of events classified by level of criticality and sensitivity; (**b**) Bar illustration with the Total Evaluation of Risk.

4.3. Comparison of Results Using MySQL against a NoSQL Environment

As part of the proposal using Big Data techniques (see Section 3.3.8), the MySQL database environment has been analyzed against a NoSQL environment. Hereby, the responsibility of the Academic CSIRT has been to provide the services of analysis, coordination, support and response to computer security incidents.

For this new proof of concept, three groups of data have been used. Test 1 has been carried out with 54 data obtained by real time of PVS. Test 2 has used 33,021 data obtained for the reports generates, while test 3 has been composed of 314,002 data acquired from the values of Snort. We opted for a relational database, as it has been adequate in the information processing. The separation gap between the NoSQL databases and the SQL databases has been very short, surpassing the NoSQL for a few seconds as illustrated in Figure 21a. Such statement has been based on the evolution in the amount of data, which have been in the tests.



(**a**) (**b**)

Figure 21. (**a**) Execution time in seconds of a SQL and a NoSQL database. This establishes the time that has been spent writing the files in a SQL and a NoSQL database; (**b**) Speed measured in registers per second of a SQL and a NoSQL database. This sets the speed at which files have been processed for a SQL and a NoSQL database.

Furthermore, the NoSQL database has been faster in the processing of information (Figure 21b). However, in the current study, we opted for a relational database, due to the fact, that when working with critical information, the concepts of ACID (atomicity, consistency, isolation and durability) have been fundamental to safeguard the data [51]. This recommendation also establishes that most of the NoSQL databases have not been located in mission critical applications, which in turn require high security, high availability, disaster recovery, modes of replication, implementation modes, besides others. These characteristics have been demanding to implement and require several years to be robust enough in NoSQL databases. In addition, in the NoSQL databases the integrity of the data has been compromised by the little support there is for transactionality. Such activity accelerated the process but lacks security when executing queries. Thus, NoSQL database may not be optimal to work with systems that use real-time data analysis as BI needs to increase the cybersecurity level.

5. Conclusions

This research has been focused on the design and implementation of an integral BI system that generate added value to the reactive and proactive services of an Academic CSIRT. Hereby, the current study has been divided into two modules. The first Module aimed mainly to A-CISRT Incident Managers, in order to improve the visualization of events such as malicious code, allowing member institutions to take actions and decisions. The second Module, aimed to analyze vulnerabilities and to recognize real-time attacks, through the development of a BI system through SCRUM, specifically for A-CSIRT members. This enabled real-time visualization of the issues that may occur in the network, in order to establish control measures. To fulfill such purpose, the phases of Ralph Kimball's methodology have been exploited to generate a BI product. The SCRUM methodology has also been

used to obtain a rapid development oriented to the A-CSIRT. Additionally, previously unpublished ETL algorithms have been designed to standardize the raw data from different sources. The obtained results demonstrate that both Modules have been useful for an A-CSIRT, since they maintain a structure of the data. Simultaneously, they present a supporting visual aid to attend malicious activities generated in the network. As an innovative projection, a new proposal focused on the scalability of the A-CSIRT has been presented in a scenario of massive data growth. As a Big Data technique, the Apache Storm has been considered, with its advantages and possible implications. Finally, three new evaluations in form of tests have been added, including a comparison between the MySQL and the NoSQL databases. Based on the nature of a BI solution oriented to Cybersecurity and taking into account the evaluation made by the three tests, it has been demonstrated that for this current being it would not be advisable to apply Big Data techniques in this project.

As future work, we have planned to complement the quality security management service, generating a module that will support Hardening People. Additionally, we have proposed to join CSIRT systems that will provide security audits or assessments.

Acknowledgments: This study has been partially funded by the Universidad de las Fuerzas Armadas ESPE in Sangolquí, Ecuador within the research project entitled "Training Computational Platforms, Experimentation, Management and Mitigation of Cybersecurity Attacks", Code ESPE-2015-PIC-019, in its Phase 3. Co-funding was given by the Consortium for Advanced Internet Development (CEDIA).

Author Contributions: Walter Fuertes is the mentor of the project. He proposed the application of the scientific method connected to the Methodology of Ralph Kimball. He also obtained financial and economic resources. Francisco Reyes and Paúl Valladares designed and implemented Modules One and Two respectively. In addition, they conceived, designed and performed the experiments; Freddy Tapia analyzed the related Work; Ernesto Pérez contributed with information of different sources and the technological infrastructure for proof of concept. Additionally, he supervised the development of the BI system; Theofilos Toulkeridis edited the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wamala, F. *ITU National Cybersecurity Strategy Guide*; International Telecommunications Union: Geneva, Switzerland, 2011.
2. Ruefle, R.; Dorofee, A.; Mundie, D.; Householder, A.D.; Murray, M.; Perl, S.J. Computer security incident response team development and evolution. *IEEE Secur. Priv.* **2014**, *12*, 16–26. [CrossRef]
3. Cichonski, P.; Millar, T.; Grance, T.; Scarfone, K. Computer security incident handling guide. *NIST Spec. Publ.* **2012**, *800*, 61.
4. Johnson, C.; Badger, L.; Waltermire, D.; Snyder, J.; Skorupka, C. Guide to cyber threat information sharing. *NIST Spec. Publ.* **2016**, *800*, 150.
5. Watkins, B. The Impact of Cyber-Attacks on the Private Sector. Available online: http://www.amo.cz/wp-content/uploads/2015/11/amocz-BP-2014-3.pdf (accessed on 22 September 2017).
6. Sandberg, A.B.; Crnkovic, I. Meeting industry: Academia research collaboration challenges with agile methodologies. In Proceedings of the 39th International Conference on Software Engineering, Buenos Aires, Argentina, 20–28 May 2017.
7. Freet, D.; Agrawal, R. A virtual machine platform and methodology for network data analysis with IDS and security visualization. In Proceedings of the Southeast Con, Charlotte, NC, USA, 30 March–2 April 2017.
8. Wedgbury, A.; Jones, K. Automated asset discovery in industrial control systems: Exploring the problem. In Proceedings of the 3rd International Symposium for ICS & SCADA Cyber Security Research, British Computer Society, Ingolstadt, Germany, 17–18 September 2015.
9. Kimball, R.; Ross, M.; Becker, B.; Mundy, J.; Thornthwaite, W. *The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence Remastered Collection*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
10. Power, D.J.; Sharda, R. Decision Support Systems. In *Springer Handbook of Automation*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 1539–1548.
11. Lim, E.; Chen, H.; Chen, G. Business intelligence and analytics: Research directions. *ACM Trans. Manag. Inf. Syst.* **2013**, *3*, 17. [CrossRef]

12. Pipyros, K.; Thraskias, C.; Mitrou, L.; Gritzalis, D.; Apostolopoulos, T. A new strategy for improving cyber-attacks evaluation in the context of Tallinn Manual. *Comput. Secur.* **2017**. [CrossRef]
13. Rajasekharaiah, K.M.; Dule, C.S.; Srimani, P.K. CRSA cryptosystem based secure data mining model for business intelligence applications. In Proceedings of the International Conference on Electrical, Electronics and Optimization Techniques (ICEEOT), Chennai, India, 3–5 March 2016.
14. Valladares, P.; Fuertes, W.; Tapia, F.; Toulkeridis, T.; Pérez, E. Dimensional data model for early alerts of malicious activities in a CSIRT. In Proceedings of the 2017 International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS), Seattle, WA, USA, 9–12 July 2017.
15. Yang, J.; Ryu, D.; Baik, J. Improving vulnerability prediction accuracy with secure coding standard violation measures. In Proceedings of the 2016 International Conference on Big Data and Smart Computing (BigComp), Hong Kong, China, 18–20 January 2016.
16. Macas, M.; Lagla, L.; Fuertes, W.; Guerrero, G.; Toulkeridis, T. Data Mining model in the discovery of trends and patterns of intruder attacks on the data network as a public-sector innovation. In Proceedings of the 2017 Fourth International Conference on eDemocracy & eGovernment (ICEDEG), Quito, Ecuador, 19–21 April 2017.
17. Manuel, J.; Cordeiro, R.; Silva, C. Between Data Mining and Predictive Analytics Techniques to Cybersecurity Protection on eLearning Environments. Available online: https://link.springer.com/chapter/10.1007/978-3-319-67621-0_17 (accessed on 5 September 2017).
18. Tisdale, S.M. Cybersecurity: Challenges from a Systems, Complexity, Knowledge Management and Business Intelligence Perspective. *Issues Inf. Syst.* **2015**, *16*, 191–198.
19. Gabriel, R.; Hoppe, T.; Pastwa, A.; Sowa, S. Analyzing malware log data to support security information and event management: Some research results. In Proceedings of the First International Conference on Advances in Databases, Knowledge and Data Applications, DBKDA'09, Gosier, Guadeloupe, 1–6 March 2009.
20. Harang, R.; Guarino, P. Clustering of Snort alerts to identify patterns and reduce analyst workload. In Proceedings of the Military Communications Conference, 2012-MILCOM 2012, Orlando, FL, USA, 29 October–1 November 2012.
21. Hellwig, O.; Quirchmayr, G.; Huber, E.; Goluch, G.; Vock, F.; Pospisil, B. Major Challenges in Structuring and Institutionalizing CERT-Communication. In Proceedings of the 2016 11th International Conference on Availability, Reliability and Security (ARES), Salzburg, Austria, 31 August–2 September 2016; pp. 661–667.
22. Bollinger, J.; Enright, B.; Valites, M. *Crafting the InfoSec Playbook: Security Monitoring and Incident Response Master Plan*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2015; ISBN 9781491913598.
23. Kruidhof, O. *Evolution of National and Corporate CERTs-Trust, the Key Factor*; IOS Press: Washington, DC, USA, 2014; pp. 81–96.
24. Bhatt, S.; Manadhata, P.K.; Zomlot, L. The operational role of security information and event management systems. *IEEE Secur. Priv.* **2014**, *12*, 35–41. [CrossRef]
25. Osorno, M.; Millar, T.; Rager, D. *Coordinated Cybersecurity Incident Handling: Roles, Processes and Coordination Networks for Crosscutting Incidents*; Laurel Md Applied Physics Lab.: Laurel, MD, USA, 2011.
26. Qian, Y.; Fang, Y.; Jaatun, M.G.; Johnsen, S.O.; Gonzalez, J.J. Managing emerging information security risks during transitions to Integrated Operations. In Proceedings of the 2010 43rd Hawaii International Conference on System Sciences (HICSS), Honolulu, HI, USA, 5–8 January 2010; pp. 1–11.
27. Belsis, M.A.; Simitsis, A.; Gritzalis, S. Workflow Based Security Incident Management. In *Panhellenic Conference on Informatics*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 684–694.
28. Elmellas, J. Knowledge is power: The evolution of threat intelligence. *Comput. Fraud Secur.* **2016**, *7*, 5–9. [CrossRef]
29. Grobler, M.; Jacobs, P.; van Niekerk, B. Cyber Security Centres for Threat Detection and Mitigation. In *Threat Mitigation and Detection of Cyber Warfare and Terrorism Activities*; AISPE Book Series; IGI Global: Hershey, PA, USA, 2016; p. 21.
30. Sharkov, G. From Cybersecurity to Collaborative Resiliency. In Proceedings of the 2016 ACM Workshop on Automated Decision Making for Active Cyber Defense, Vienna, Austria, 24–28 October 2016; pp. 3–9.
31. Mejía, J.; Muñoz, M.; Ramírez, H.; Peña, A. Proposal of Content and Security Controls for a CSIRT Website. In *New Advances in Information Systems and Technologies*; Springer International Publishing: Basel, Switzerland, 2016; pp. 421–430.

32. Wu, D.; Chen, S.H.; Olson, D.L. Business intelligence in risk management: Some recent progresses. *Inf. Sci.* **2014**, *256*, 1–7. [CrossRef]
33. Gahi, Y.; Guennoun, M.; Mouftah, H.T. Big Data Analytics: Security and privacy challenges. In Proceedings of the 2016 IEEE Symposium on Computers and Communication (ISCC), Messina, Italy, 27–30 June 2016; pp. 952–957.
34. Zuech, R.; Khoshgoftaar, T.M.; Wald, R. Intrusion detection and big heterogeneous data: A survey. *J. Big Data* **2015**, *2*, 3. [CrossRef]
35. Mahmood, T.; Afzal, U. Security analytics: Big data analytics for cybersecurity: A review of trends, techniques and tools. In Proceedings of the Information Assurance (NCIA), Rawalpindi, Pakistan, 11–12 December 2013; pp. 129–134.
36. Jaramillo, E.; Munier, M.; Aniorté, P. Information Security in Business Intelligence based on Cloud: A Survey of Key Issues and the Premises of a Proposal. In Proceedings of the 10th International Workshop on Security in Information Systems, Angers, France, 5 July 2013.
37. Ahmad, A.; Hadgkiss, J.; Ruighaver, A.B. Incident response teams—Challenges in supporting the organizational security function. *Comput. Secur.* **2012**, *31*, 643–652. [CrossRef]
38. Kimball, R.; Caserta, J. The Data Warehouse? ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data. John Wiley & Sons: Indianapolis, IN, USA, 2011.
39. Mohd, N.; Yunos, Z.; Ariffin, A.; Nor, A. CSIRT Management Workflow: Practical Guide for Critical Infrastructure Organizations. In Proceedings of the 10th European Conference on Information Systems Management, ECISM 2016, Evora, Portugal, 8–9 September 2016.
40. Kimball, R.; Ross, M. *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*; John Wiley & Sons: Hoboken, NJ, USA, 2013.
41. Schieber, D.; Reid, G. *CSIRT Case Classification (Example for Enterprise CSIRT)*; Technical Report; FIRST: Manchester, NH, USA, 2004.
42. Roberts, L.D.; Howell, J.A.; Seaman, K. Give me a customizable dashboard: Personalized learning analytics dashboards in higher education. *Technol. Knowl. Learn.* **2017**, *22*, 317–333. [CrossRef]
43. Bouman, R.; van Dongen, J. *Pentaho Solutions: Business Intelligence and Data Warehousing with Pentaho and MySQL*; Wiley Publishing: Hoboken, NJ, USA, 2009.
44. Sugumaran, V.; Sangaiah, A.K.; Thangavelu, A. (Eds.) *Computational Intelligence Applications in Business Intelligence and Big Data Analytics*; CRC Press: Boca Raton, FL, USA, 2017.
45. Mulder, S. An Action Research Study on the Use of SCRUM to Provide Agility in Data Warehouse Development. Ph.D. Thesis, University of Pretoria, Pretoria, South Africa, 2011.
46. Goede, R. Agile Data Warehousing: The Suitability of SCRUM as Development Methodology. In Proceedings of the 5th IADIS Multi Conference on Computer Science and Information Systems, Rome, Italy, 24–26 July 2011; pp. 51–58.
47. Scholtz, I.I. Inmon versus Kimball: The Agile Development of a Data Warehouse. Ph.D. Thesis, North-West University, Potchefstroom, 2016.
48. Highsmith, J.; Alistair, C. Agile software development: The business of innovation. *IEEE Comput.* **2001**, *34*, 120–127. [CrossRef]
49. Iqbal, M.H.; Soomro, T.R. Big data analysis: Apache storm perspective. *Int. J. Comput. Trends Technol.* **2015**, 9–14. [CrossRef]
50. Dayarathna, M.; Malshan, M.; Perera, S.; Jayasinghe, M. Scalable Complex Event Processing on a Notebook. In Proceedings of the 11th ACM International Conference on Distributed and Event-Based Systems, Barcelona, Spain, 19–23 June 2017; pp. 327–330.
51. Gessert, F.; Wingerath, W.; Friedrich, S.; Ritter, N. NoSQL database systems: a survey and decision guidance. *Comput. Sci. Res. Dev.* **2017**, *32*, 353–365. [CrossRef]