2013

Using Splunk to Manage Integration System's Log Files

A Graduation Internship Thesis

Hogeschool ICT

accenture

High performance. Delivered.

Fransiskus Prayuda
Fontys Hogeschool ICT
6/10/2013

GRADUATION REPORT

FONTYS UNIVERSITY OF APPLIED SCIENCES

HBO-ICT: English Stream

Data student:	
Family name , initials:	**Prayuda, F.**
Student number:	**2157006**
project period: (from – till)	**4 February – 30 June 2013**
Data company:	
Name company/institution:	**Accenture Technology Solutions**
Department:	**Advance Technology & Systems**
Address:	**Versterkerstraat 6, Almere, The Netherlands**
Company tutor:	
Family name, initials:	**Hagemans, G.**
Position:	**BPM Consultant**
University tutor:	
Family name , initials:	**Aarts, Henk N.H.J.M.**
Final report:	
Title:	**Using Splunk to Manage Integration System's Log Files**
Date:	**10 June 2013**

Approved and signed by the company tutor:

Date:

Signature:

Foreword

This report is written for Fontys Hogeschool as a thesis for my graduation. The aim of this report is to give detailed information of the assignments that I performed during my internship, and how I managed to accomplish all the assignments in this project. Besides that, this report is also written to help the graduation juries assess the whole process that I have done in my final project.

The main reason I decided to do my graduation project at Accenture was because I am interested with the system that handles business operations data at large enterprises. As one of the biggest IT consultant company, I believe that working at Accenture will be an opportunity for me to get more knowledge about these kind of system.

The opportunity to be able to work in a company like Accenture is priceless; hence I want to thank them for giving me the chance. I would like to give a big appreciation to my company tutor, Guy Hagemans, who had given me support and guidance during the project. I also want to thank Peter van Gulik whom had given me detailed insight and description of the Oracle BRM system. Without their assistance, I would never be able to complete all the assignments in this project.

I would also like to thank my school tutor, Henk Aarts for his feedbacks during the creation of this thesis. His feedback ensures a high level of quality from all documents that will be submitted to Fontys and also gave me valuable knowledge on how to write a good report.

I also want to thank my family in Indonesia and all of my friends, that always give support to each other to get through hard times. Finally, I want to give my appreciation to people who have willingly given up their time to read and give feedback to this thesis.

Fransiskus Prayuda
June, 2013

Table of Contents

Summary

This thesis aimed to give a detailed explanation of the process from a graduation project that was carried by Fransiskus Prayuda, a 4th year ICT student of Fontys Hogeschool in Eindhoven, The Netherlands. The graduation project took place at Accenture Technology Solutions, an IT consultancy company located in Almere, The Netherlands. The graduation project took exactly 20 work weeks, which began on February and ended in June 2013.

As an IT consultancy company, Accenture offers a lot of products and services. Enterprise Application Integration (EAI) or system integration, along with the services needed to implement the system integration into their client's current systems are one of them. The system integration product already contains various default tools, but usually the client has special wishes for a type of tools that are not provided by those default tools. Because of this, Accenture and their clients have to spend a lot of time and money to develop custom tools to fulfill the special wishes.

On the other hand, time to market is also an important factor for improving their client satisfaction, which made Accenture realize that there can be a significant time and cost saving if they are able to build a library of 3rd party tools that can be reused for different client situations with similar request. Because of this idea, Accenture is interested to see a proof of concept from one of the 3rd party tools to see the possibility of using reusable tools in their projects.

The tool that was chosen for this proof of concept is called Splunk, Splunk is a computer software that is able to manage and indexes machine generated big data. The big data that is used in this proof of concept came from the log files of a Dutch telecommunication company's billing system. These log files have very large size and extremely unstructured which make them rarely used even though they holds valuable information about the system.

The proof of concept built in this project is able to help manage log files coming from the billing system and analyze the events recorded in the log files. In general analyzing big data, even with traditional data processing software, is proven to be extremely difficult and time consuming because of their size and complexity. But by using Splunk, one will be able to manage big data files and it also allow them to analyze the information it holds.

The proof of concept also proved that utilizing 3rd party tools to fulfill Accenture's client special wishes is possible. Using 3rd party tools will also help Accenture to reduce the time and cost required to prepare the whole client solution. Because with 3rd party made-ready tools, Accenture will only need to configure the tool and it will directly able to work with the desired system.

Glossary

3rd party software

A software that developed by a different developer than the original vendor of the platform (in this project the platform will be an integration system)

Accenture Knowledge eXchange (KX)

A knowledge library only accessible to Accenture employees which holds information from previous projects, which can help other employees in future projects

Accenture Technology Solution (ATS)

A group of Accenture which focus on providing technology solutions and services to their clients

Begin anchor

Regex patterns used by Splunk to mark the beginning of the characters that have to be matched

Big data

Collection of very large data that make it impossible to process them using traditional database management tools

Boolean expressions

Expression in programming language that will produce Boolean value (either true or false)

Client solution

A term used by Accenture consultant for the complete client's system and the integration system that will be implemented

Components

Software that build up the whole integration system (in this project the DM and CM are the example of components inside Oracle BRM)

CSV(Comma Separated Value)

A file that stored data in plain characters and each of the value are separated using commas

Dashboard

A tool that give graphical presentation of business process or other particular object, this is usually presented in a single page and specifically made to be easy to read

EAI (Enterprise Application Integration)

Software and systems that are used to integrate a set of enterprise applications. In this paper, EAI is also written simply as integration systems

End anchor

Regex patterns used by Splunk to mark the end characters that have to be matched

ESB (Enterprise Service Bus)

A method of distributed computing that is used for interaction and communication between two or more software

Fields

A set of values (characters and / or numbers) that extracted by Splunk via the user given regex

Flatlist

A format of storing data that are divided by their category and are put in hierarchical structure

In-house software / development

When a company needs a piece of software, in-house development means that that software is developed by programmers within the same company

Information Technology (IT) Systems

Term that used in this paper to describe the whole system that a company depends their business process on (e.g. finance / accounting, customer relationship, manufacturing system, etc.). This kind of system is usually called Enterprise Resource Planning (ERP)

Log files

A file where all the events happening are recorded by the systems. This allow the people who maintain the system to know what activity happens inside it and it can also be used to diagnose problem

On cloud

Term used to describe system that runs on another computer system which is connected via network. The term is used because usually the user only need to run the client software in their own system to use the whole software without knowing where actually it is stored, therefore it is called running software on cloud / cloud service

On premise

The opposite of cloud service, to be able to use the program the user have to installed the whole software package in their system

Open source

A term used for a program (with their source code) that is available to use or modifications by the general public

Operation code (opcode)

A method used by Oracle BRM to communicate between each component (intern or extern)

Oracle Communications Billing and Revenue Management System (BRM)

A revenue management system that specially built by Oracle for communication and media service providers

Oracle Connection Manager (CM)

A component of Oracle BRM that provide an interface between clients and the rest of the system. All client applications connect to the BRM system through a CM

Oracle Data Manager (DM)

A component of Oracle BRM which is used for accessing data in a storage system or legacy system by translating requests from CMs into a SQL query that the database can understand

Parser engine

A term used in computer science for a type of analytic formal grammar which hold a set of rules for recognizing strings (characters and numbers) in the language

Proof-of-concept

A demonstration which in principal is used to verify some ideas / concepts that it has the potential of being used

Regular expression (Regex)

A set of characters (which have literal and / or symbolic meaning) that can automatically identify the similarity of textual material from a given pattern

SAP AG

Multinational software based on Germany who makes enterprise software (ERP) to manage business operations

Scalability

The ability of a system to adapt the growing amount of work or the ability to be enlarged to accommodate the growing amount of work

SOA & BPM cell

A cell (or group) inside Accenture technology solutions which focused on service oriented architecture (SOA) and business process modeling (BPM)

Splunk

A software for searching, monitoring and analyzing machine generated data via a web-style interface

Splunk Knowledge Database

The place where Splunk store the information on what it will do with your data (e.g. field extraction, data transformation, data classification, etc.)

Structured Query Language (SQL)

A programming language specially created for managing data stored in relational database manager system (RDBMS)

Sumo Logic

A cloud based system used for managing and analyzing big data to deliver real time insight

Tibco Software Inc.

A company that provides infrastructure software for companies to use on premise or on cloud environments

Total Commander

A freeware orthodox / command based file manager that has tabbed interface to improve the speed of file operations by user (e.g. comparing file, moving file, renaming file, etc.)

Type of tools

Other functionalities that can be add into integration system, without the need of hardcoding it directly into the system

1. Introduction

All companies nowadays have become more and more dependent to IT systems for their daily business operations. Because of this dependency, most of the large companies have a very big application landscape. This landscape consists of more than one system and applications running together at the same time which bring difficulties if the company wants to make a change in the landscape. This difficulties will cause a problem because IT systems need to follow the flexibility of business conditions to be able to support them. To overcome this problem, there is an emerging need to integrate all of the applications and systems inside the company's application landscape.

The system that provide integration with other applications or systems is called 'enterprise application integration'. These kind of systems usually placed into separate dedicated hardware because of the huge amount of data that need to be extracted, transformed and loaded between two or more different systems and / or applications. Enterprise integration system is one of the products that are offered by Accenture Technology Solutions. Besides selling the product, Accenture will also help with the implementation of this integration system into the client's existing system and architecture.

Accenture is able to deliver enterprise integration system from multiple vendors e.g. Tibco, SAP AG, Oracle and many more. This system also provides default tools that can be used along with the integration process. But most of the time, their clients have special wishes for a set of tool that cannot be acquired from those default tools. Because of this, Accenture has to spend a significant amount of time to custom made several types of tools that are commonly asked by their clients.

As Accenture realized the importance of time to market, they come out with an idea of a time saving method by using reusable tools. By building a set of commonly used tool, the time to market will be greatly reduced which will also lead to cost saving of deploying this integration systems. These are the reason why Accenture interested in a proof of concept to see if it is possible to use reusable tools for multiple client situations with similar request. My graduation project was to choose one of the tools that are interesting and to build the proof of concept using that tool.

This report is organized as follows: Chapter two will be about Accenture, including the company's history and how it is organized at this moment. Chapter three will give a detailed explanation about what this project about, chapter four will be about all the information that I researched during the beginning of the project to get familiar with the problems relating with this project. The fifth chapter will be about the approach / method that were used to complete the whole project. Chapter six to nine will explain about what happened during each phase and all the problems that were met and how I solve them, and in chapter ten I will give conclusion and recommendation of this project.

Finally after chapter ten, I will be giving an evaluation of my experiences from this graduation project. The references that are used inside the report will be put inside square brackets with number ('[]') with all the sources can be found in the References chapter, and to close this report there are also some attachment that will give more information to support this report.

2. Company Profile

2.1. Company History

Accenture started around 1950s along with the installation of the first commercial computer system in United States. The company reputation was built mainly as technology consultant and system integrator. In late 1980s, Accenture began to offer new way of business integration solution that aligned organization's technologies, process and people with their own strategy.

In a few decades Accenture expanded its service including management consulting, beside the already existing technology consulting which bring benefit its client. The company pioneered systems integration and business integration; led the deployment of enterprise resource planning, customer relationship management and electronic services; and has established itself as a leader in today's global marketplace. [1]

2.2. Accenture today

Accenture is a global management consulting, technology services and outsourcing company, with approximately 259,000 employees and clients in more than 120 countries. Combining their experience and extensive research on the world's most successful companies, Accenture work together with their clients to help them become high-performance businesses

Accenture's "high performance business" strategy focuses on the expertise in consulting, technology and outsourcing to help clients perform at the highest levels, which make them able to create sustainable value for their customers and shareholders. Accenture also help identify new business and technology trends which can be used to develop solutions to help their clients around the world to:

- Enter new markets.
- Increase revenues in existing markets.
- Improve operational performance.
- Deliver their products and services more effectively and efficiently.

By enhancing their consulting and outsourcing expertise with alliances and other capabilities, Accenture help move clients forward in every part of their businesses, from strategic planning to day-to-day operations. [2]

The organogram below will give detailed overview of the department's organization inside Accenture:

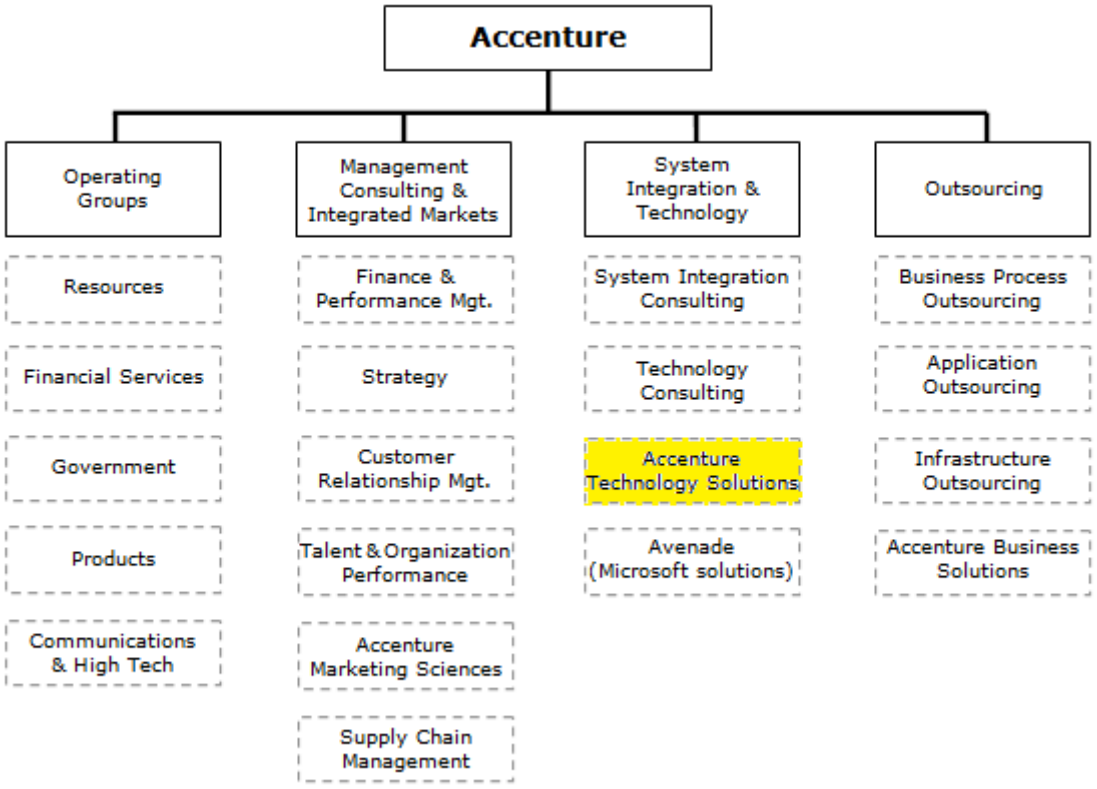


Fig 1. Accenture Organogram

I worked at Accenture Technology Solutions for this graduation project, more information about this department will be explained in the section below.

2.3. Accenture Technology Solution (ATS)

As a 100% subsidiary of Accenture, Accenture Technology Solutions focus on delivering IT services and maintaining IT systems. Accenture Technology Solutions is part of the worldwide Accenture Solutions Center Network. This network is based on three important components:

- Multidisciplinary teams of resources in order to fulfill client-specific requirements;
- A global network of delivery centers at regional and international levels;
- Uniform methodologies and tools.

The key focus of ATS is on providing technology solutions and services to Accenture clients. ATS most important capabilities are SAP, Oracle (including Siebel, PeopleSoft and JD-Edwards), Business Intelligence, Java, Application Testing and Project Services. From their client locations and from their offices in Almere and Den Bosch, they deliver application development and application maintenance for multiple organizations varying in size. [3]

The organogram below show how the departments are organized within Accenture Technology Solutions:

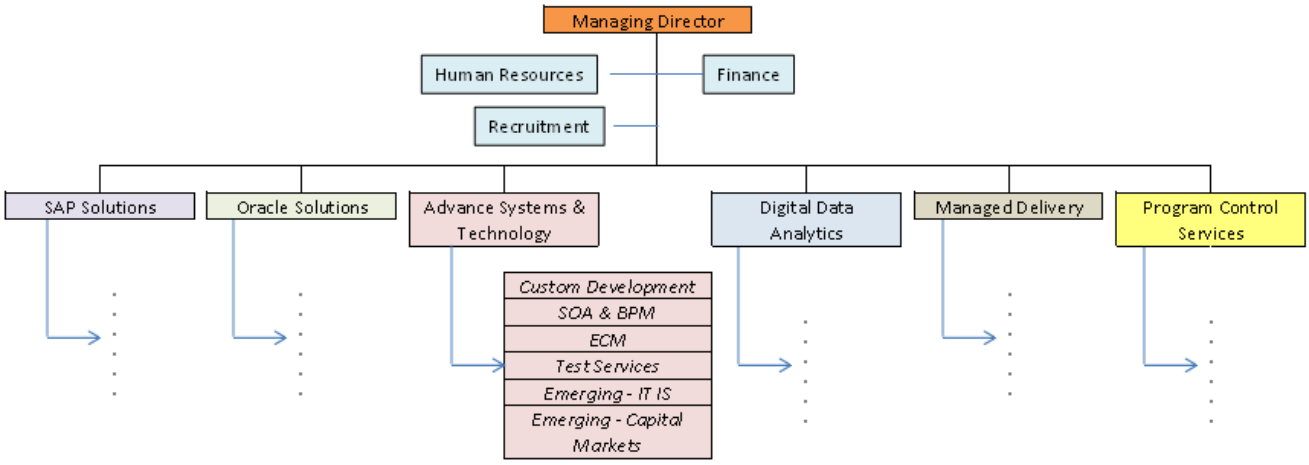


Fig 2. Accenture Technology Solutions (ATS) Organogram

From the organogram in fig. 2, we can see that ATS is divided into several clusters and each cluster is further divided into cells. Clusters and cells are the term Accenture used for departments and groups respectively. This graduation project is located inside the Advance Systems & Technology cluster and inside the SOA & BPM cell to be more specific. SOA & BPM cell is specialized in delivering business process management that is connected with SOA solution by optimizing and managing business processes through design, automation and, ultimately, business activity monitoring.

3. The Project

This chapter explains about the problems that exist in company's current situation and the desired situation in detail. The difference between the two situations is the initiative behind my graduation project. Besides that, I will also explain the approach that I planned to be able to complete this project.

3.1. Current situation

As explained in previous chapter, enterprise application integration is one of the products that are offered by ATS. There is a lot of software vendors who provide complete integration system that provide support for the whole Enterprise Integration stack (from basic messaging up to orchestration of the complete business processes). Because of the diversity of their client, Accenture are not biased to a specific enterprise integration system.

In reality, what happen most of the time is the clients have more requirements of tools that are not supported by the default tools from the whole integration stack itself. This is why Accenture has to make the custom tools by themselves that go on top of the default integration system to fulfill the customer wishes. These tools sometime have to be remade again for different projects or customers which can take a lot of time before the whole solution is ready for implementation.

3.2. Desired Situation

Time to market is also important to improve client's satisfaction, and Accenture know that there can be a lot of time saved if they build up a library of tools that can be re-used and implemented in different client situations. Beside time saving, this kind of library will also reduce the cost required for implementing an system integration. The differences between the current and desired situation are the reasons why this project was initiated.

3.3. The Assignment

My assignment is to look for the possibility of a reusable set of tools, and choose one of the tools that can be used as proof of concept of reusing this tool into several client solutions. Reusing tools will help Accenture to save time they need to complete a whole integration system, which ultimately will also reduce the cost of the system. The detailed steps that have to be completed for this project are listed as follow:

1. Investigate the availability of tools that can be used along with enterprise integration system that already exist at mature stage in the market today.
2. Select one of the tools which seem interesting and useful for the clients as a proof of concept.
3. Build the proof of concept by implementing the tool into a real life case.
4. Assess the added value the tool bring and the feasibility for the tool to be used across solutions.

3.4. Approach

A plan of approach is needed to be able to control and finish this project with the desired quality. But for this project I did not directly go into making a project plan, because the assignment is still a bit vague for me at the beginning. This is why I decided to spent about the first two weeks for some research to get basic understanding about the terms and the way integration system works. Even though I have not made a definite plan for the whole project, I already draw up what steps I have to do to be able to complete this project. The steps that I planned were:

1. Look for type of tools that can be used for the proof of concept
2. Make a comparison from the tool to help choosing the appropriate tool
3. Look for data that can be used in the proof of concept
4. Build the proof of concept by implementing the data inside the chosen tool
5. See what are the advantages / benefits the tool bring to the data

I did not put the steps directly into phases because it was still very early in this project, but I was able to create the plan after I finished the research mentioned before. This research will be explained in more detail in the next chapter, and the methodology I used in this project will be explained in chapter five. The detailed planning that I made can also be seen in my project plan which can be found in Appendix B.

4. Research

This chapter will explain the research that I performed to get more background information about what the assignment is about. Because of the complexity of the system environment I have to work with, the information I gathered in this research helped a lot during the creation of the project plan and during the implementation of the proof of concept. This research was done by using various methods (e.g. by internet research, via small interview) and sources (e.g. the internet, and when applicable I also looked into the Accenture KX or asked to one of the colleagues in the company). The research questions that emerged during the beginning of the project are formulated as follows:

1. How EAI system works? What kind of data that they use for exchanging information?
2. What kind of tools that are mostly requested by the clients of Accenture?

4.1. EAI works using an ESB

The first research was done to get more insight on how integration system works nowadays. This research was also aimed to get knowledge of the method used by the integration system for exchanging data. This knowledge helped me a lot on the whole project, especially when I was building the proof of concept as this research give basic understanding of the integration system. Other important knowledge that I acquired was the understanding of 'components' term inside integration system scope. This research was done by using internet research method, which means all data (texts and pictures) found in here were already available on the internet.

Modern EAI works using Enterprise Service Bus (ESB) as their backbone to carry all the messages from one place to another. As bus-based EAI, a number of other necessary functionalities were identified (e.g. security, transaction processing, error handling); because rather than hardcoding all these features, the bus architecture allowed these functions to be enclosed in separate components. These components then could also be grouped in various configurations files to handle any integration scenario in the most efficient way possible and could be hosted anywhere within the infrastructure. [4]

Fig 3. ESB as the backbone of the integration system

The type of data that is mostly used (in the scope of my graduation project) is CSV and flat list file. CSV (or comma-seperated-values) file is a file that store text and number in a sequence of characters, without data on how to interpret it. The CSV file consist of records that seperated by line break, and each records usually consist an identical sequence of fields that are sepereated with comma [5]. It also works the same with flatlists, but instead of line break and commas the data inside a flatlist are stored in some kind of hierarchial lists. Both of this files are used in EAI because they are lightweight which make it easy and faster to transporting them across the ESB. Even though it will make the corresponding system need to work harder (i.e. more processor power and memory is required) to interpret the data, but CSV / flatlist file will not required a lot of the ESB limited capacity. [6]

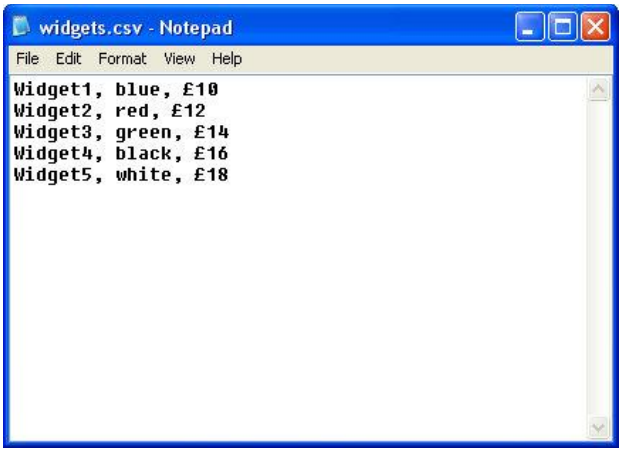


Fig 4. Example of a CSV File

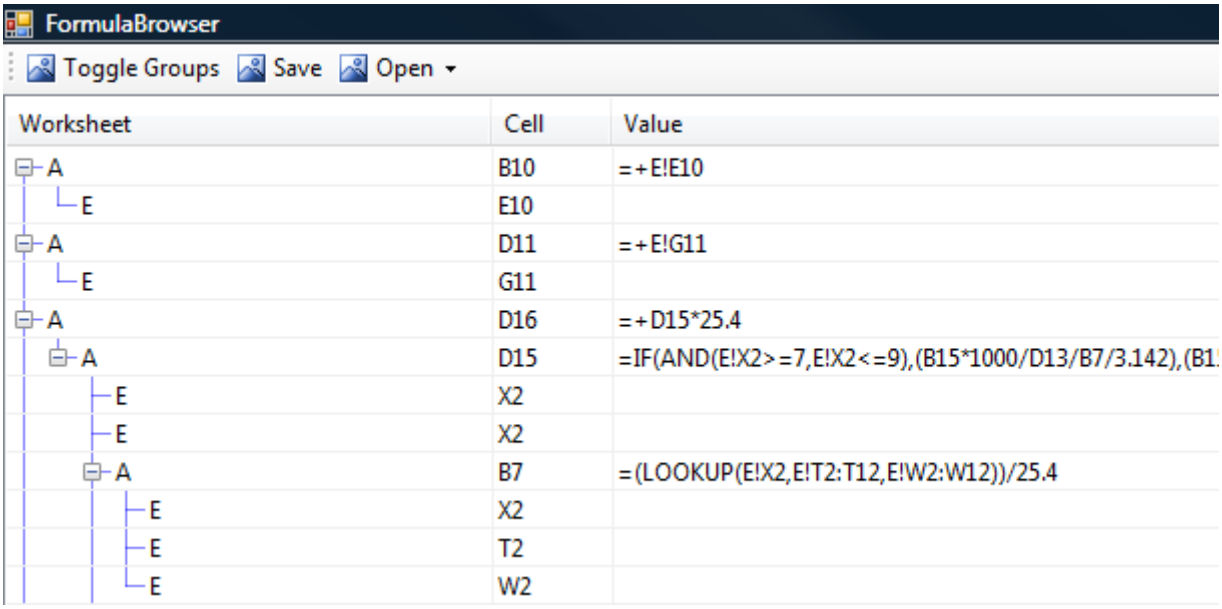


Fig 5. Example of a flatlists with its hierarchical structure

4.2. Most requested type of tools

The second research was done to have a basic idea of what kind of tools that are usually asked by Accenture's clients. Information that was gathered from this research is very important because it will specify the scope for the whole project. The result of this research also helped me during looking for other type of tools in the initiation phase, because the type of tools that I found here can be used as the starting point from where I can expand my search. The methods used for this information gathering were by researching into Accenture Knowledge eXchange (KX) and via interview of Accenture colleagues who have the knowledge about this matter. The data that came out from this research is a list that gives overview what kinds of tools are commonly asked by the clients. An excerpt from the list will be shown below:

- Error-Handling
  - Tools that will record all error happening during the integration process
- Logging
  - Tools that will handle and manage all the log from components inside the integration system, including log from the error-handling component
- Implementation Patterns
  - Tools that provides guidance about the implementation patterns of products that together will represent a comprehensive EAI platform
- Development Tools
  - Tools that can be used to develop new components

5. Methodology

This chapter gives short explanation of all the phases that were used in this project. For easier progress control, the project will be divided into 4 phases in which each of them have their own activities and deliverables. For this project I choose to use standard waterfall method because the activities in the next phase sometimes require the result from the previous phases, which means the phases have to be done sequentially. To ensure the quality of all the deliverables, I had a discussion every week with my company tutor, who also acts as the formal client, about the progress of deliverables on each phase.

5.1. Initiation

The first phase was aimed to get information of what kind of tools are available in the market that can support the whole integration process. This phase took about 4 weeks for me to be able to complete the deliverable and I took one extra week to make final changes according to my company tutor's feedback. More detailed information about this chapter will be explained in chapter six. The approach used in this phase is by researching on the internet for available type of tools that are already at mature stage in the market. Steps that were taken on this research are: look for any information about type of tools that can be used inside an integration platform. From those tools, choose which type of tools was interesting to be realized in this project and finally, look for products that can fulfill the tasks of the chosen type of tools. The milestone of this phase was a report that gave detailed description of a list of type of tools that available in the market and products that can fulfill their tasks, this report will also act as the deliverable for this phase and can be found in appendix C.

5.2. Analysis

This phase began with a set of products that already chosen from the previous phase. More detailed information about this chapter will be explained in chapter seven. This phase took about 5 weeks for me to be able to complete the final delivery and I took one extra week to make final changes based on my company tutor's feedback. The first step for this phase was to experiment with those tools and look for differences in key features between them. The risk in this phase arose from the accessibility of the program, because most of the program have paid license which means I will not be able to use them for a short time experimentation. This risk was successfully overcome because I found that both programs have trial license with complete features for a short period of time, which was more than enough for my experiment.

From the differences I found in their key features, I tried to set a list of criteria that can be used to shown all the key feature of each program. Then a product comparison report was made using those set of criteria and also it has to be written following the Accenture's template because of the possibility that it will be use inside Accenture in the future. The last step in this phase was to choose a product from the product comparison report that will be used as the final product of this project. The milestone for this phase was the comparison report that shows all differences and similarities in the interesting products, this report act as the deliverable for this phase.

5.3. Build

More detailed information about this chapter will be explained in chapter eight. This phase took about 4 weeks for me to be able to build the proof of concept and in the end and I took one extra week to make final changes following my company tutor's feedback. The first step in this phase was to look for a real life case that can be used for the proof of concept; the possibility of a real life case was become a risk at the beginning of this phase. But this risk was able to be resolved by communicating

with the company tutor and some Accenture colleagues, which in the end my tutor and I found a case that can make use of Splunk.

After acquiring the case, I try to analyze what problems were exist inside the case and how the tool can solve some of them. The last step in this phase was to build the proof of concept by implementing the data from the case inside the product. The milestone for this phase is the fully working product with the data from the case which will show the features and advantages the product can bring, this demo will also be the final delivery for this phase.

5.4. Assessment

This final phase can only be done after the case was able to be implemented into Splunk. This phase took me around 3 weeks to complete the deliverable along with the completion of this thesis. The aim of this phase is to make a summary of all the features of Splunk and what are the added values that Splunk can bring. Besides that, there will be also information about the feasibility of Splunk to be used in other projects. The milestone of this phase is a product report which contains all the information of the summary. This report will also act as the final deliverable for this phase and can be found on appendix D.

6. Initiation Phase

As explained in the previous chapter, initiation will be the first phase in this project. This phase was aimed to look more into the type of tools that can be used with the integration system nowadays. It was required to have the list of commonly requested type of tools by Accenture's clients before I began with this phase. This phase was further divided into 2 big tasks that need to be done to gather the information which, will be delivered to Accenture using a report at the end of this phase. This chapter will be organized as follow, first I will explain the steps that were done and the data that are acquired from those steps. And after that I will explain all the difficulties that were met during this phase and what I did to solve or minimize the effect to the end product of this phase. On the next phase we will see how I choose the criteria to compare the selected product and which product will be chosen to be built into proof of concept.

6.1. Tools in the market

After collecting a list of tools that are mostly requested by Accenture's client, we (my company tutor and I) decided that it will be better to look for more type of tools that are already on mature stage in the market today. The purpose was to expand the list so that we can have a wider overview of all those tools (e.g. are they stand alone tool or connected work with some other types of tools, are they specific or general purpose tool). This information gathering was done by using internet research and from this research I succeeded looking for more types of tools that can be used in the integration system nowadays. The results are listed as follows (description and the usage of each tools can be seen in Appendix C):

Load Balance
System Management
Audit
Recovery
Parse
State Management
Resource Management
Event Monitor
Error Raising
Error Handling
Archive
Message Queue
Transformer
Router
Filter
Validate
Map
Configuration Management/Versions
Notify
Track
Monitor
Report

Fig 6. List of types of tools that are available in market at this moment

The next step was then to make a list of sample products that can serve the purpose of the type of tools. This sample products were vary from their size, stand alone or packet program, open source or paid license, on cloud or on premise, and many more. This was done to oversee the availability of the products to overcome the type of tools in the market.

The final step was to look deeper to each type of tools for the value they will bring to the whole integration package. In the end, my company tutor and I decided which types of tools are interesting to be implemented into proof of concept. In this project, we select to look more into monitor, dashboard and reporting features; the reason behind this decision was because the graphical feature of these type of tools will make things easier to see what are the added values that they able to bring, even if it's only from proof of concept.

For the type of monitor, we decided to look into monitor for log files generated by the integration system. The reason we choose this specific product is that logging has become one of the main components in data integration. All data that goes in and out, successful or not are all recorded in the log files and stored in a server. But usually this kind of log are very big and have no pattern at all which makes it difficult to use them as an analysis choice of a problem that happened inside the system.

6.2. Third-party vs. in-house software

Before going to the next step, my tutor and I had a short discussion whether I am going to look for 3rd party or develop the tool by myself for this project. After several considerations, we decided to look for 3rd party tool. The reasons we choose this decision are explained as follow:

- It is difficult to keep the program up to date. Technology are evolving every day, this means if I am going to develop the tool by myself it should be able to work with the technology available nowadays. But after the project is done, Accenture will need a person that will control and update the program to follow the latest technology to be able to use the tool. By using 3rd party software; the provider of the tool will be the one who is responsible of updating their product, and because they are a commercial company they will absolutely keep their product up-to-date.
- Complex environment. Because integration system is a very complex system, it is difficult to find / set up a development environment along with the data that are suitable for developing the tool that I have to build.
- Time constraint. Because of the tool that I have to be built have to be generic enough to be able to be implemented to different client situation, this means I have to know understand the basic of integration system architecture. Seeing from the complexity of this kind of system, it will be very time consuming and will be rather impossible to be done in the time frame of this project.

After deciding for third-party tool / software, I started to look on the internet for company whose product can fulfill the task of the chosen type of tools.

6.3. Products that can fulfill the task

The research on this part was aimed to find products that can act as a log management system. This product should also support huge files because of the number of transactions that go through everyday

via the integration system. From this research, I came out with a list of products ranging from small to big and free / open source to paid license. After discussing the list of choice with the company tutor, we decided to choose the paid license because it has the smallest set up and maintenance costs. Beside that the cost that have to be paid for the license are not that big for any large company, especially if this tool able to perform better than the others. With better performing tools, Accenture and their clients can further save their time and money even though the license costs are higher. Finally, from this research we had two products that stand out among the others; the two products are SumoLogic and Splunk.

6.4. Challenges during this phase

There were some difficulties that were met during this phase; the first one was occurred when I had to look for the types of tools. The problem arose because of no international standards in the naming of a type of tool. This means that every company who made the same type of tool can have different names which also bring the next problem of difficulty on finding the products. This problem also happens because of the same reason: each company can name their product differently even though their purpose is the same with other tools. Because there was no actual solution for this, I tried to minimize the effect of this problem to the end product by gathering information from trusted sites. There is a lot of information on the internet about this and on various sites (e.g. forums, articles), but I only took the information I got from reliable website (e.g. Microsoft, Oracle, IBM sites and some other EAI companies).

The next challenge emerged when I was looking for sample products for each of the type of tools. The problem was that a product can support the task of more than one type of tools (e.g. monitor usually will also go with notify and report components). Instead of showing the sample products to each type of tools, I grouped the related tools and list the products that go with them to solve this problem.

The third problem was met while making preliminary comparison of the sample products. After looking for sample products, I was able to acquire a long list of them. It would be nice to experiment with all of them to get a deeper look with each program, but this was not possible because of the complexity of the program (in setting up, loading data, etc.) and comparing each of them will require a lot of time even more than the time range of this project. To solve this problem, I just looked for more information about the features of each programs from their website and use it to make preliminary comparison. But most of the time the features seen in a product's website are very limited which makes the data gathered in this way can be a little bit shallow.

The last problem was met during selecting a product that is interesting to be built. This problem arose because not all of the type of tools is 100% interesting to the Accenture's clients. But this problem was easily solved by discussing with my company mentor and several colleagues of Accenture about which products will be valuable to Accenture's clients.

7. Analysis Phase

The second phase of this project was aimed to do more extensive research to the programs that were chosen in the previous phase. This was done by making another comparison, but this time with the detailed capabilities of each product. The detailed feature was achieved by experimenting with both of the products. As in the previous phase, this phase was also organized in similar structure: first I will explain the step I took and it will end with the problems I met during this phase. In the next phase we will see what case I used for the proof of concept and how to implement them into the chosen product.

7.1. Criteria to compare products

After having 2 products chosen from the previous phase, the next step was to make a detailed comparison of all the features that can be found inside both products. The first step taken in this research is by experimenting with both of the programs by using the manual book to see what the features each of them has. From the experiment, I started to set a list of criteria that can be used to compare those products. Some criteria that were used are listed as follows:

- Price / Cost
- What kind of log files supported
- Interactive user interface
- Customizable dashboard
- Speed / Performance

From those criteria, I made a table that gives detailed overview of all the features of Splunk and SumoLogic. The feature of both products will be explained briefly in the following subchapters.

7.2. The Interesting products

Splunk and SumoLogic are tools from companies with the same name that specialize in big data management. Every messages and routines that happened inside the integration system is recorded in the log files. Because of the high number of transactions that go through the integration systems every day (± 2 million messages), the log files from the integration system also considered as a big data. More detailed explanation about the features and capabilities of Sumo Logic and Splunk can be found in chapter 7.2.1. and 7.2.2. respectively.

7.2.1. Sumo Logic

Sumo Logic is a cloud based service log management system. Their service include timely and actionable information, taken from any type or source of log data, that not only helps solve operational, security challenges but can also be used to provides critical business insights as well. Their cloud-based approach overcomes the problems of premise-based solutions, including limits on scalability, inefficient and uncontrolled costs. [7]

Feature of Sumo Logic [8]:

- Sumo Logic modeled its approach of Google, which uses advanced machine learning algorithms to whittle down mountains of log file data into common groupings
- Sumo Logic Use search engine-like syntax which can quickly find records with relevant keywords
- LogReduce™ which can reduce hundreds thousands of results into a single page of meaningful patterns.

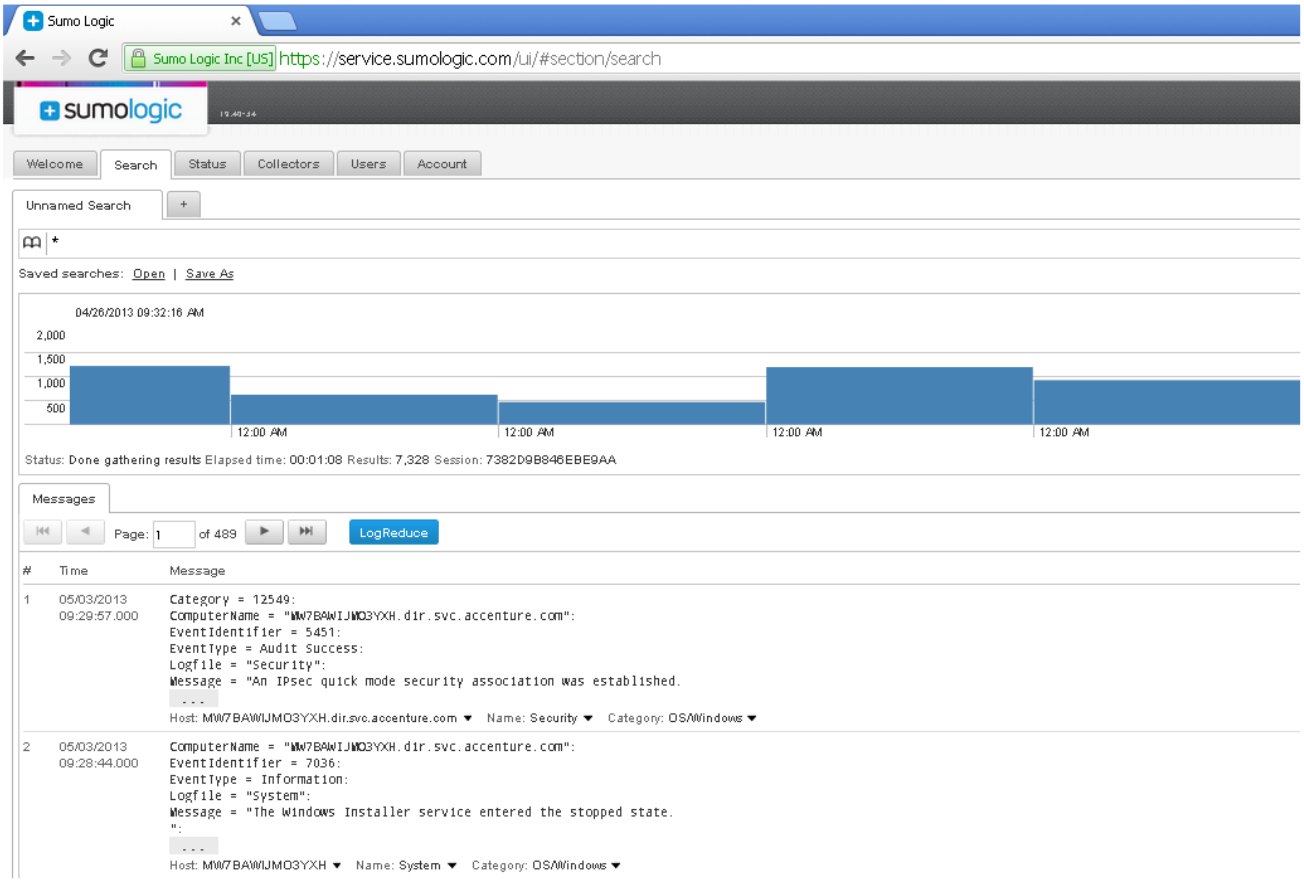


Fig 7. Search menu of Sumo Logic

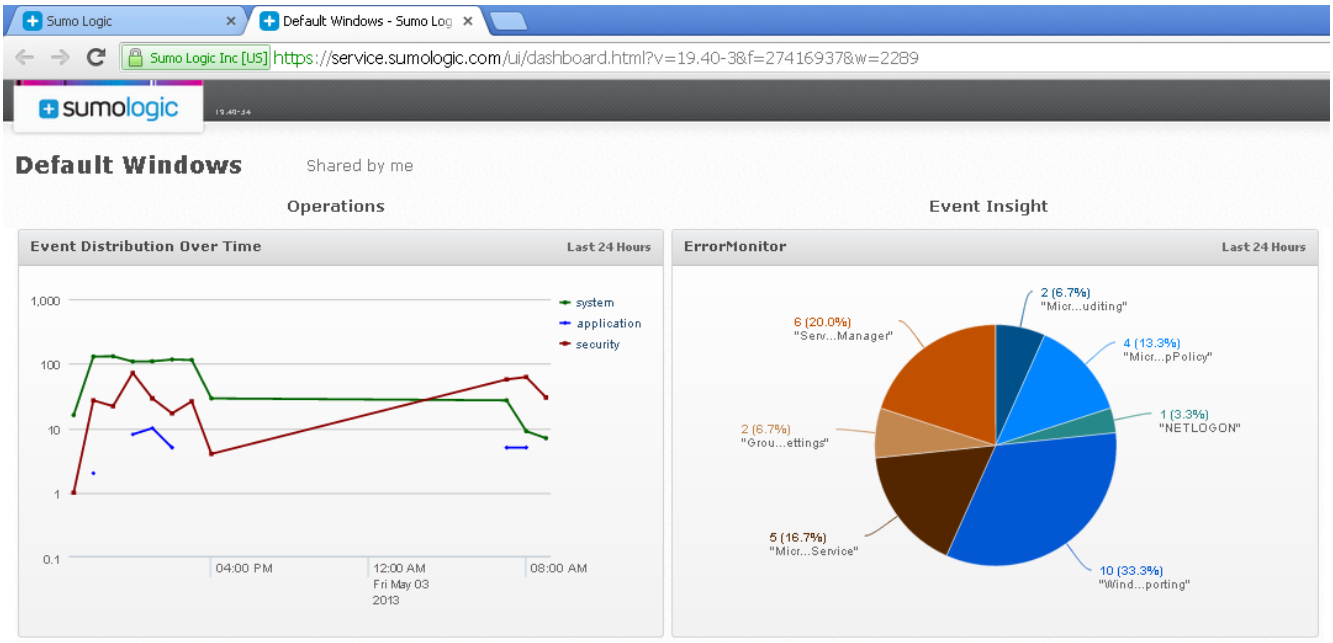


Fig 8. Dashboard features of Sumo Logic

7.2.2. Splunk

Splunk is a software for searching, monitoring, and analyzing IT data from any application, server or network device that makes up your IT infrastructure by using web browser as their user interface. It can capture, index and correlates real time data into searchable repositories which can also be used to generate graphs, reports, alerts for statistical analysis. Splunk can index either structured or unstructured machine-generated big data. [9]

Features of Splunk [10]:

- Searches feature in Splunk were made based upon Unix Piping and SQL which support searching, filtering, modification, manipulation, insertion and deletion.
- Search Acceleration which makes the future search using the same query run faster
- Information-rich views and dashboards that can fit the wide-ranging needs of the user

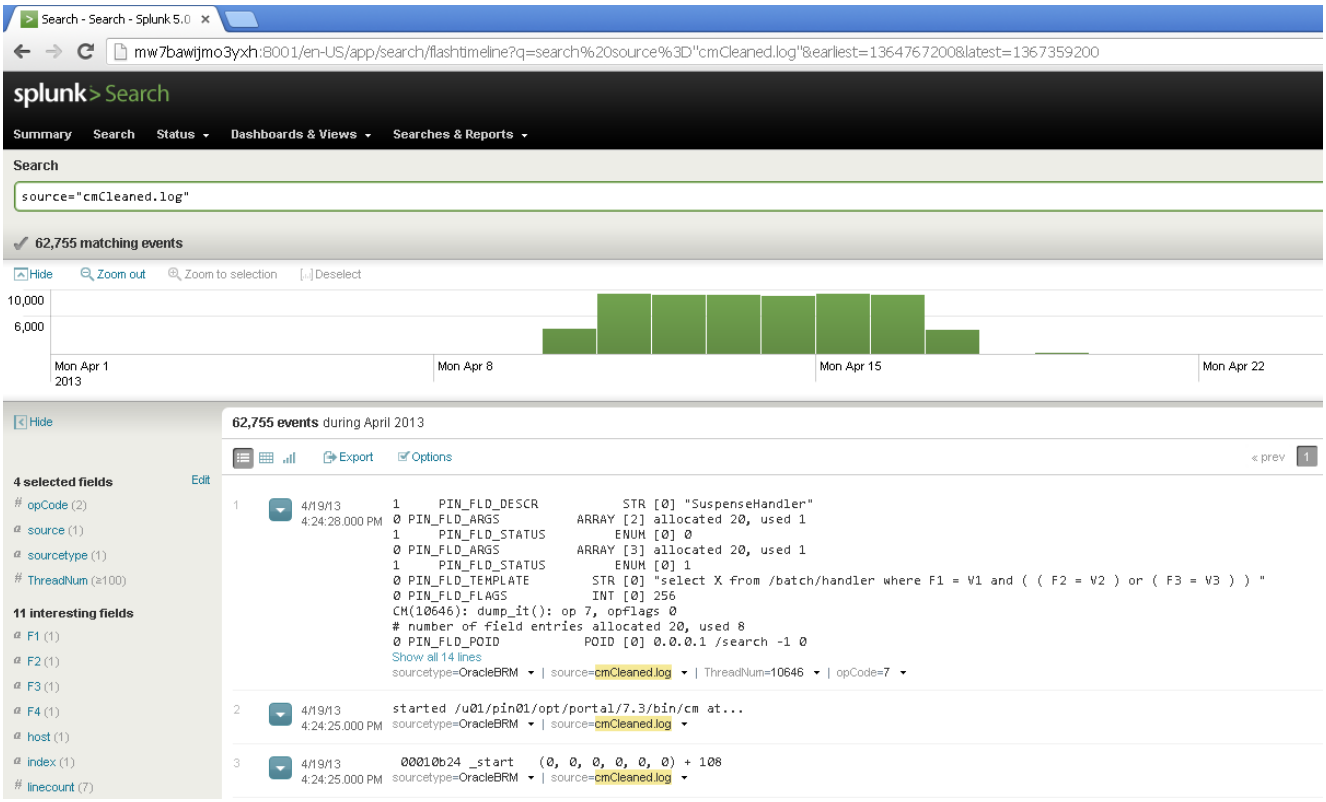


Fig 9. Splunk can capture and index a machine generated log file

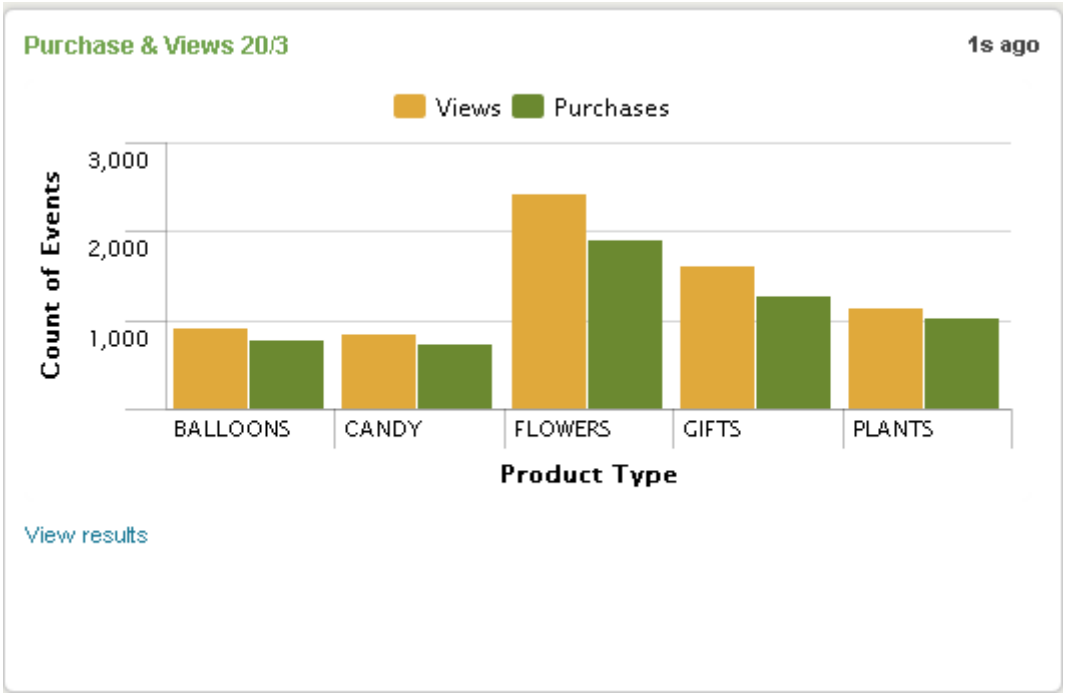


Fig 10. Splunk can also make graph from the data in log files to support analysis of business process

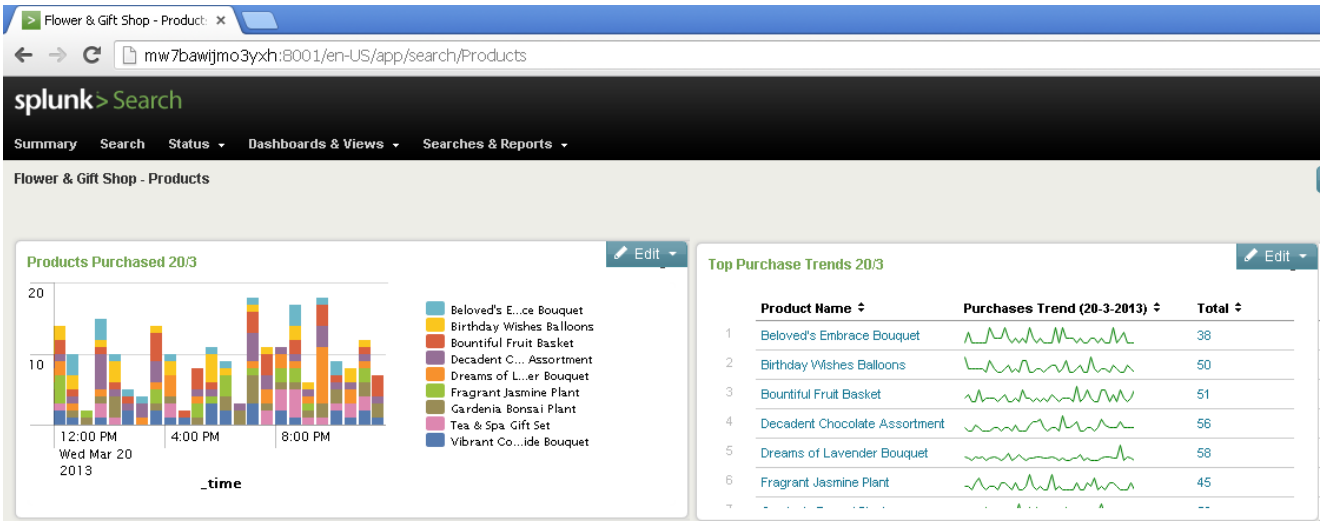


Fig 11. Splunk rich and interactive dashboard feature

7.3. The chosen product

After experimenting and made a product comparison report from both of the program, my tutor and I had discussion over which product will be the most interesting for Accenture (e.g. new and emerging type of tools) and also usable for their clients. As the result from our discussion, we decided to use Splunk for the proof of concept that will be built in this project. Even though the price of Splunk is higher than Sumo Logic, the features that Splunk has is far more superior from the other and outweigh the price differences. That's why for this proof of concept, we choose to go with Splunk because with its rich features it will show completely what kind of values that this kind of product can bring.

7.4. Challenges during this phase

There are two difficulties that I encountered; the first one is when I had to experiment with all the programs that came out from the preliminary lists. At the beginning there were 8 programs that can fulfill the task from the type of tools of monitor, track alert and notify; but most of them cannot receive offline log files. This means that they need an already running system to monitor, and then they can start indexing all the events and logs that happened with the system. The solution of this problem is to make all the tools that do not support offline log files are not applicable to this project. This workaround was chosen because it will be difficult to find or set up a fully working integration system in the development environment that can be used for this project.

Second problem that I met in this phase is the limitation of the products. This happened because both products are paid license product, to experiment with them I have to make use of the trial versions that they provided. For Splunk it means that all the features can be used for free within 2 month free trial and for Sumo Logic the data that can be transferred to their server are very limited (500 Megabytes (MB) per day). There is no actual solution / workaround to solve this problem, the only way is to make use of the given spaces, be smart with time management (especially for Splunk) and filter / clean up from unnecessary data for files that need uploaded to the server (especially for Sumo Logic).

8. Build phase

This phase was aimed to build a proof of concept that will become the end product of this project. The first step that was done for this phase was to look for a case from a real life integration system environment that can make use of the chosen product, which is Splunk. After I acquired a case, the next step was to understand the problem that exist inside the case, and finally implement the data from the case inside Splunk. More detailed steps will be explained below alongside with all the problems that were met during this build phase. The next phase is the final phase where I asses the value Splunk bring to the case and whether it is possible or not to use Splunk inside different client situations.

8.1. Oracle BRM

My tutor found a case that will be perfect to be implemented into Splunk and make use of its features. He found a log data from the billing system use by Dutch Landline and mobile Telecommunications Company to send their invoice to the customers. The problem in this case is the log data that are stored in the server were never been used by people that maintain the system. The main reason was because the log files do not have good structure which makes it extremely difficult to understand / analyze them. It will take a big amount of time and manpower for a company to be able to use them, which make the company think that it is an impractical thing to do.

Before I go any further in explaining the logs, I will explain first about Oracle Communications Billing and Revenue Management (BRM). Oracle BRM provides a fully convergent charging and billing system to manage the entire revenue management lifecycle. From a single modular platform BRM supports charging and rating for any service, any network, any payment method, any geography and supports all customer and partner type. [11]

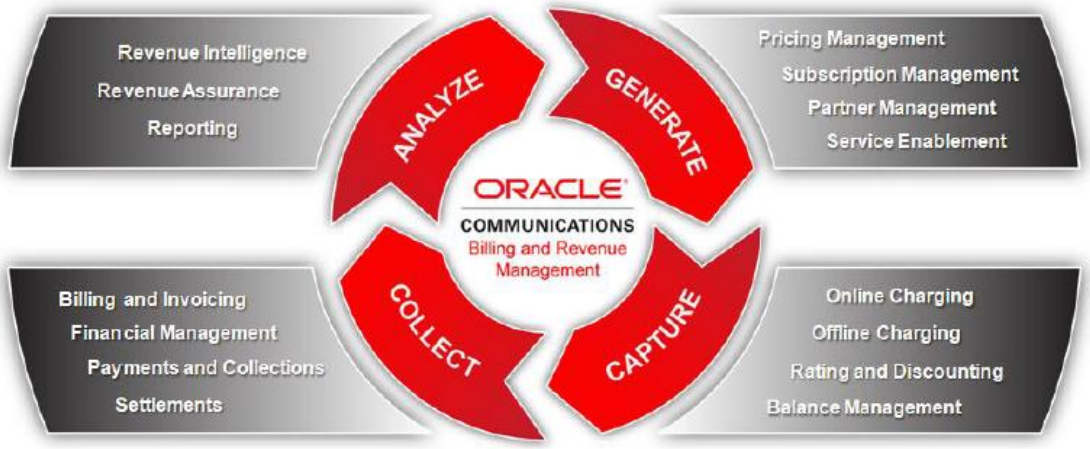


Fig 12. A brief overview of the usage of Oracle BRM

There were 3 log files from this system that I can implement in Splunk, they are : cm.log, cm.pinlog and dm_oracle.pinlog. The contents of each log files will be explained in detail in the next subchapter.

8.1.1. Cm.log

This is the biggest log that I received, the size itself reach around 3, 3 Gigabytes (GB). This log contains all the operation that was run by the BRM's CM (Communication Manager) system. BRM communicates (intern and extern) with help of opcodes (operation codes). The data that are required for the operations are sent through the parameter of the opcodes in the format of flatlist which are also recorded here. This log file is also quite special because there are no warning / error event recorded in there, this means that this logs record all the function that were run by the Oracle system but did not record the result of the call. Below I will explain an event from the file to give an overview of how the how whole contents of the log looks like.

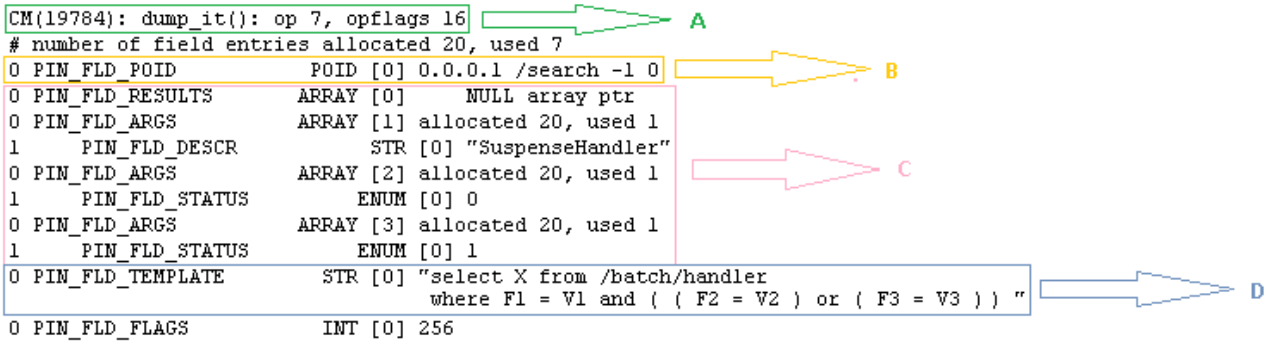


Fig 13. Snippet of an event from cm.log file

Figure 13.A shows that this log was from CM thread number 19784 and this event was recorded by dump_it method. The opcode number is 7 which is used to look / search for object inside the system. Figure 13.B is the Portal Object ID (POID) field which basically tells the system where the object's location using the IP addresses. Figure 13.C are the parameters that were required to run the opcode which are recorded in flatlist format and on this case it was looking for a SuspenseHandler with value 0 or 1. In the end the opcode will form a SQL like statement (Figure 13.D) to be run by the system. The SQL statement can be seen by replacing the value of Fx and Vx so it will look like this:

```
select X from /batch/handler where PIN_FLD_DESCR = "SuspenseHandler"
and ( ( PIN_FLD_STATUS = 0 ) or (PIN_FLD_STATUS = 1 ) )
```

Fig 14. SQL statement like that stored inside flatlist

8.1.2. Cm.pinlog

Unlike the previous log which contain all the operations that have to be run by the Oracle BRM Communication Manager, this log contains the result from those operations. This log consist a lot of important data because it recorded which operation run successfully without errors, and which operation returns an error. It also gives detail information on which class and lines that cause the error. A snippet of the log will be given and explained below.

```
E Wed Apr 17 00:10:38 2013  sz0304  cm:22281  cm_child.c(115):4669 1:sz0304:pin_cycle_forward:22277:3:102:1366150237:553
        op PCM_OP_SUBSCRIPTION_CYCLE_FORWARD returned an error
        <location=PIN_ERRLOC_RTP:19 class=PIN_ERRCLASS_SYSTEM_DETERMINATE:1 errno=PIN_ERR_BAD_VALUE:46>
        <field num=0:0,0 recid=0 reserved=0 reserved2=0 time(sec:usec)=0:0>
        <facility=0 msg_id=18013 version=0>
```

Fig 15. Snippet of an event from cm.pinlog

From the log event above we can see the time, which thread number encountered the error (in this case (cm:22281) and on which class and lines it happened (in this case cm_child.c(115):4669). The second line gives a custom string that written by developer to help them debug the error. The next lines show the location and more technical information about the error (e.g. PIN_ERRLOC_RTP:19 means the error was happened inside the real- time pipeline).

8.1.3. Dm_oracle.pinlog

This log contains all events that came out from Oracle Data Managers (DM) which means most of the errors happened here have connection with the underlying database of the system. Some of the errors in this log also consist of Oracle error code which connected to the SQL run by the Oracle database. An example of the event from this log file will be shown and explained below.

```
E Sat Mar 23 11:41:04 2013  sz0304  dm:13915  dm_subr.c(144):2722 1:sz0304:<no_name>:14346:1:1:1364035264:38
        ORACLE error: do_sql_insert: PINStmtExecute: code 1536, op 0
        =ORA-01536: space quota exceeded for tablespace 'PIN01'
    "insert into config_t ( poid_DB, poid_ID0, poid_TYPE, poid_REV, hostname, name, program_name, account_obj_DB, account_obj_ID0, account_obj_TYPE, account_obj_REV,
created_t, mod_t, read_access, write_access, descr, value, version ) values ( :poid_DB, :poid_ID0, :poid_TYPE, :poid_REV, :hostname, :name, :program_name, :account_obj_DB,
:account_obj_ID0, :account_obj_TYPE, :account_obj_REV, :created_t, :mod_t, :read_access, :write_access, :descr, :value, :version )"
```

Fig 16. Snippet of an event from dm_oracle.pinlog

From the snippet above we can see that the thread number that report this log was in this case dm:13915. In the second and third line we can see the Oracle error and the details of the error which in this case is an error in Insert SQL statement caused by space that want to be inserted were exceed the allowed value. And in the last 3 lines, the log shows the SQL statement that encountered the error.

8.2. Implementing into Splunk

After understanding the syntax and meaning of all the log files, the next step was to upload them into Splunk. After uploading them, I started to analyze and identify data that are be valuable to be shown and finally implement them to Splunk so that it can show the desired data. The subchapters below will give more explanation of the steps taken to accomplish this part.

8.2.1. Looking for interesting data

The first thing to do before developing the proof of concept in Splunk was to look for data that can be interesting to be shown to the customer. Searching for the usable data was a little bit challenging because most of the events in the log files are very technical. Because of this situation, I did a consultation with another Accenture consultant, Peter van Gulik, who is familiar with the technical and functional working of the Oracle BRM. In the end, I came up with a list of interesting data that can act as requirement for the proof of concept of Splunk; but because most of the data is technical, there are not a lot of graphical pattern which can be shown.

8.2.2. Make Splunk recognize patterns with Regex

Recognizing patterns is one of the features in Splunk, this feature work by using Regex (Regular Expressions). First we must insert the regex into the Splunk knowledge database to make it recognize a value. After inserting the regex, set the target file and stored them inside the database, Splunk will directly look for this patterns and store it on separate fields (works like a variable) of the value that match the patterns. This fields then will directly calculated the occurrence in percentage and numbers which furthers help analytic process.

Regular expression (common abbreviation for Regular expression including regex and regexp) is a term used in computer science for a specific pattern that provides concise and flexible means to specify and recognize strings of text; e.g. particular character of characters, words, or patterns of characters. Regex provides grammar for a formal language which will be interpreted by a regular expression processor (a parser engine that examines text and identifies substrings that are member of the specified formal language). The concept of regex was mainly utilized inside Unix system (editor ed and the filter grep), but currently it is also used by many text editors and programming language to search and manipulate text patterns. [12]

Inserting regex into Splunk knowledge was easy, but making Regex to recognize pattern was a little bit tricky. Here is one of the regex examples that I stored in Splunk knowledge database:

```
:\d+\s(?<correlationID>[\d]+:[^\t]+:[^\s]+)\s
```

Fig 17. Example of regex used to extract a value

The regex above was used to recognize correlation ID that can be found inside the error or warning event in cm.pinlog and dm_oracle.pinlog. To understand the regex above, it is better to look at the one of the example of the event where the ID resides.

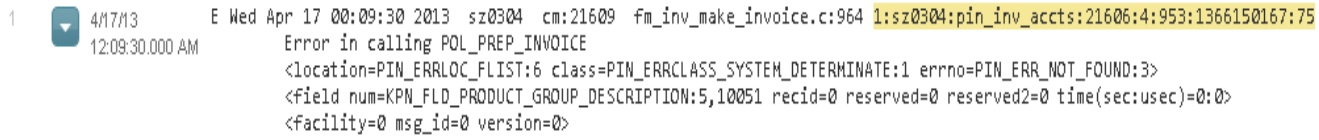


Fig 18. Snippet of an event with the value that want to be extracted

Figure 18 show one of the events along with its correlation ID (the value is highlighted in yellow). The first step to make regex to match the value is to find the beginning and end anchor, anchor is a value where the regex start / finish matching the patterns. For this regex, I use a ':', number, and space as the beginning (in this case was ':964 ') and a space as the end anchor. That's why from the regex above we can see :\d+\s in the beginning and \s as the end. The \d and \s means decimal and white space respectively, meanwhile the plus sign means that there is one or more value available (in this case one or more decimal value).

The next part of the regex above are marks by open ['('] and close [')'] parentheses, this symbol used to separate the begin and end anchor with the values that want to be extracted. The field name where the values will be extracted into was located inside the '?<>' symbols and in this case it was named corellationID. The rest of the regex is used to give patterns to the values that Splunk will extract, the

'[]' symbol means a character set. Splunk will look for a values where begin with one or more decimal sets '[\d]+', a ':', then all value as long as it's not a :, and finally all value except the space which act as the end anchor.

After the regex was done and stored inside the Splunk knowledge database, Splunk will directly recognize the value if the patterns occur in the events. As seen in the figure 19, the correlationID are stored and shown below each event.

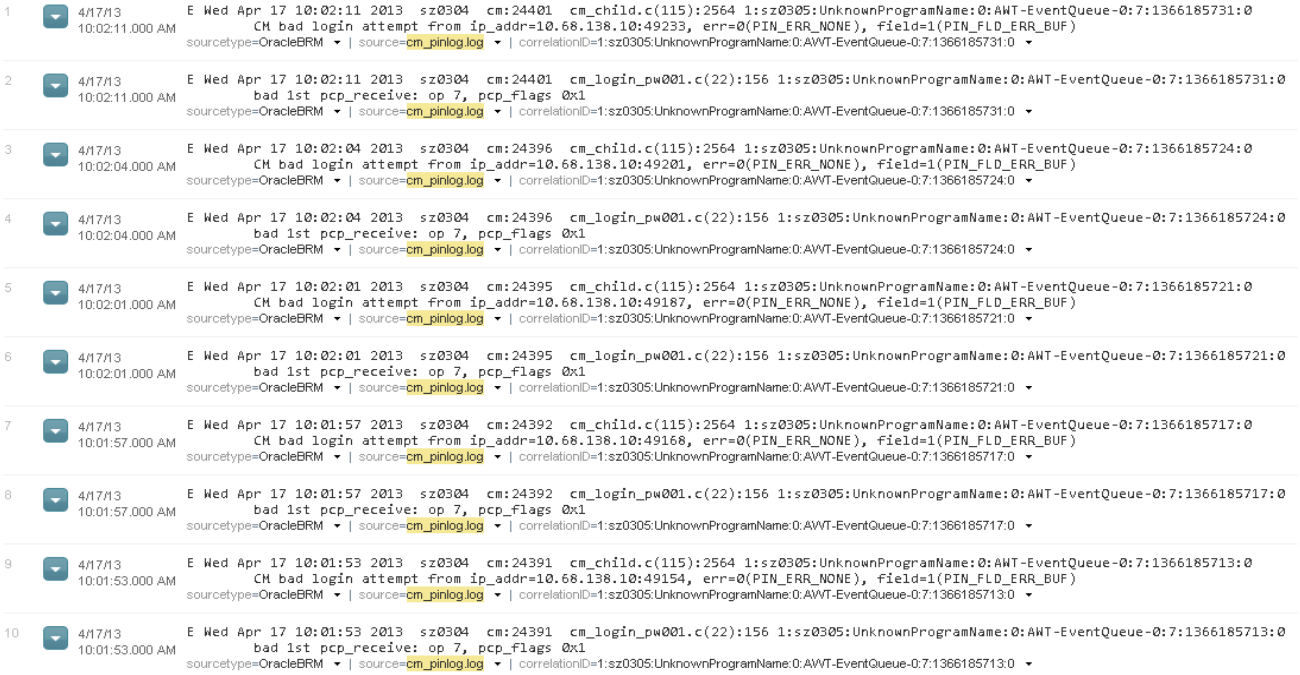
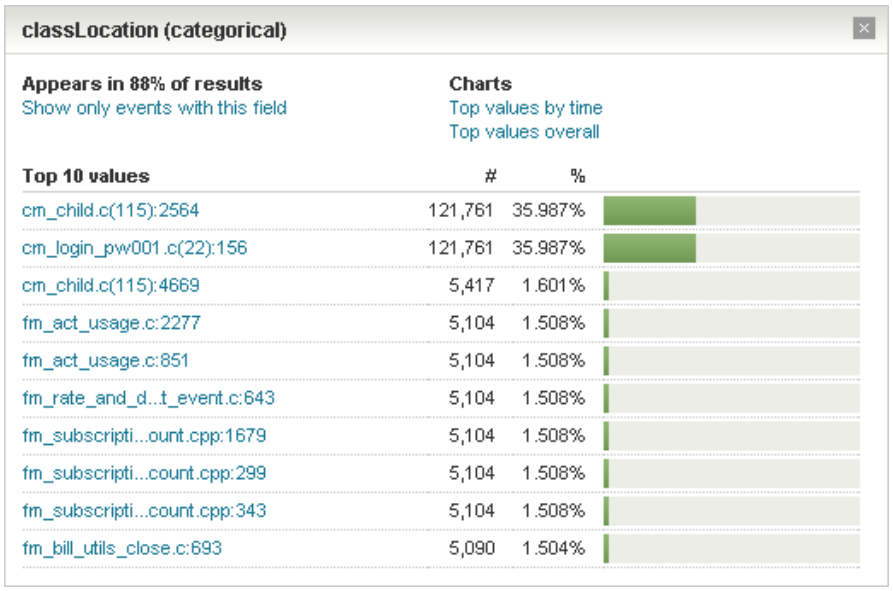


Fig 19. Screenshot of Splunk recognized the value of correlationID in each event automatically

8.2.3. Using recognized value to help analyze data

After Splunk was able to capture the value that match the given pattern, it will directly summarize them. The data that are summarized by Splunk can then use to help analyze events recorded in the log files. For better explanation, I will give an example that I took from cm.pinlog below. All error events in cm.pinlog also record from where the error comes from, it gives detail information of the class and which lines that causes the error event. The figure 20 shows that after Splunk are able to recognize the data, we will directly see the number of occurrence of the corresponding data.



classLocation (categorical)

Appears in 88% of results
Show only events with this field

Charts
Top values by time
Top values overall

Top 10 values	#	%
cm_child.c(115):2564	121,761	35.987%
cm_login_pw001.c(22):156	121,761	35.987%
cm_child.c(115):4669	5,417	1.601%
fm_act_usage.c:2277	5,104	1.508%
fm_act_usage.c:851	5,104	1.508%
fm_rate_and_d...t_event.c:643	5,104	1.508%
fm_subscripti...ount.cpp:1679	5,104	1.508%
fm_subscripti...count.cpp:299	5,104	1.508%
fm_subscripti...count.cpp:343	5,104	1.508%
fm_bill_utils_close.c:693	5,090	1.504%

Fig 20. Splunk will automatically make statistics from the recognized values

As we see from figure 20, we can directly get the idea which class and on what line that cause a lot of error. With this information, we can prioritize which one that have to be dealt first and on this case are the first two values which causing around 70% of all the errors recorded. Furthermore, with simple search command (more about search command will be explained in the next chapter); Splunk can transform this data into graph which will give easier way to interpret the data and graphs in figure 21 can also be put into dashboard that can be used for monitoring.

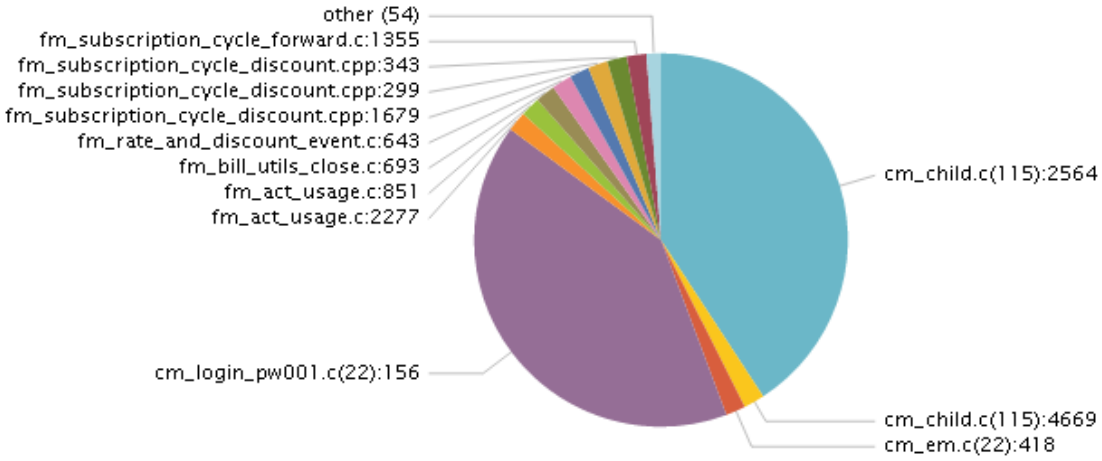


Fig 21. With simple search query, Splunk can summarize the recognized values

8.2.4. Using Splunk search feature to correlate data

Splunk has a search language that was created based upon Unix piping and SQL. Because of Unix piping, all Splunk commands are chained together using a pipe character '|'. The pipe character tells Splunk to use output of one command (left of the pipe), as the input of the next command (right side of the pipe). This enable user to refine the data step by step until they get the data they want. The search term that can be use including keywords, phrases, Boolean expressions, key/value pairs, etc.

Finally the search feature can be used into 2 different main purposes: to investigate the root cause of an issue from the data that we indexed and to summarize the data into a report (in tabular or visualization format).

This search feature was used to correlate the events between cm.pinlog and dm_oracle.pinlog. Before using Splunk, correlation between log files are very time consuming and nearly impossible. But with Splunk this task become possible, figure 22 will give example of the correlation of log event that exist in cm.pinlog and dm_oracle.pinlog by using their correlation ID.

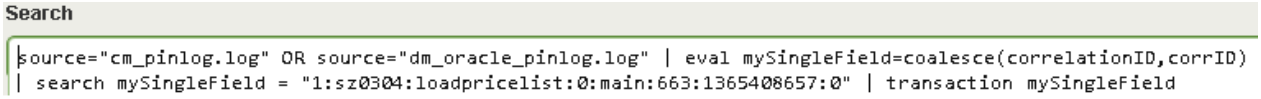


Fig 22. Screenshot of the search query to correlate between two logs

Fig 23. The result from correlating events from 2 different sources by the help of correlation ID

From figure 23 we can see that correlating log events is more efficient with Splunk. Errors that are recorded in two different places means that they were caused or happened in multiple components of the integration system. With the complexity of the integration system, it will be very time consuming to investigate this kind of error. That's why this combining logs capability will be very helpful to analyze and investigate error that occurred in multiple components of integration system.

8.3. Challenges met during this phase

There are some problems that I met during this phase; most of the problems met in this phase happened during implementation of the case sample into Splunk. This phase will explain the problem in detail, what the causes were and how I solved them. The first problem was concerning the log file, as explained before the log file from the Oracle BRM system is not very well structured. This became a problem because it was difficult to understand the log file, and it was also hard to find interesting data in the log file. To solve this problem, I asked one of Accenture's consultants, Peter van Gulik, who understands how Oracle BRM systems work. With his help and the Oracle BRM manual book, I was able to understand the meaning of the log and found which data were interesting to be shown with Splunk.

After I was able to understand the log, the next problem came during inserting the log into Splunk. Splunk user interface are web-based which means they are accessible using web browser, this cause problem while uploading the cm.log file (it size was around 3, 3 GB). The problem arose when I want to send the file to be indexed in Splunk, the system just say 'Saving' without any progress or time left. After waiting few hours, I knew that this will not work maybe because the web browsers have problem processing file size that big.

The only solution that I had was to take some part of the log so I will just use smaller part of the log for the demo. To be able to do this I have to open the log and this bring another problem, because most of the text editor (Notepad, Notepad++, Wordpad) are not able to open file this big. I looked for another program that support big files (gVim, Emacs, Large Text Viewer) but it was also unsuccesful. This happened because they need to load the log contents to RAM, which proved to be an impossible task as the log itself had 3,3 GB in size and the RAM in my system are only 4 GB.

In the end I used program called Total Commander (an Orthodox File Manager, an old way of managing files and data) that are able to give preview of the file. From here I found that most of the log files consist of whitespace, and in the end what I need to do was just to clear the whitespace. Unfortunately, Total Commander does not support editing feature so I have to copy the text manually part-by part-and save them into a new text file. After removing the whitespace, the log file size was reduced into around 500 MB and I can successfully upload it into Splunk.

After all the log files were indexed in Splunk, the next step was to input the regex so Splunk can find the patterns in the log files. Making the pattern, of which value Splunk has to extract was not a big problem, but finding the correct begin and end anchor with the regex was trickier. To explain it more clearly I will took one example from the event below. For this event, I wanted to take the "fm_bill_mb_apply_charges error" string. This is the string that the Accenture's developer made by themselves to know more about the error, and as we see the pattern of all the value that we want is character.

```
E Wed Apr 17 00:10:38 2013  sz0304  cm:22281  fm_bill_make_bill.c:7400 1:sz0304:pin_cycle_forward:22277:3:102:1366150237:553
        fm_bill_mb_apply_charges error
        <location=PIN_ERRLOC_RTP:19 class=PIN_ERRCLASS_SYSTEM_DETERMINATE:1 errno=PIN_ERR_BAD_VALUE:46>
        <field num=0:0,0 recid=0 reserved=0 reserved2=0 time(sec:usec)=0:0>
        <facility=0 msg_id=18013 version=0>
```

Fig 24. Snippet of an event from with the value that want to be extracted

Taking the end anchor is easy as we only need to take the '<' before 'location'. The begin anchor was a little bit tricky, cause if we took a ':' then a digit and then the space before the first character (in this

case ":553 "), then Splunk will took the beginning value as ':22281 ' because it has the same pattern and will extract all the string from "fm_bill_make_bill.c:.." until the end anchor. To solve this problem I use more value as the beginning anchor, so not just the ":553 " but the whole ID (in this case "1:sz0304:…:553 ") and finally Splunk can recognize the correct value. But this is just one problem of the pattern recognition from regex, other problem was also emerge because not all value have the same pattern in log (e.g. the ID can have 8 or 7 character sets). The only solution to solve this incorrect patterns recognition was to follow the rule of thumb, which are:

- Be sure to give the regex for beginning and end anchor as specific as possible
- Meanwhile, try to be as flexible as possible (e.g. use wildcard character that can be used to substitute any characters in a string) for the regex that are used to match the pattern of text we want to capture

If Splunk recognized the wrong values, it will also affect the accuracy of our analysis from the log files.

The last problem that I met during this phase was emerged during correlating between two logs (cm.pinlog and dm_oracle.pinlog). The search query used to look for this value was not complex, but it caused efficiency problem because of the big number of events that has to be checked. Mathematically, there will be around 285 billion events have to be checked which made Splunk took about 15 minutes just to find the logs that exist on both of the logs. Because of this, I have an idea to find another way of checking both of the logs to increase the efficiency of correlating logs.

At first I used a method where every event in first log file has to be checked against every event in the second log file. But as I explained above, this method proof to be inefficient because every change we made in the query used for searching for this correlating logs (e.g. time range) made the system redo all the checking again. The new method I used is more efficient as I divide the query into two smaller parts. The first step of the new method is to list all the ID that exist on the first and second log files, and then do a distinct count of the sources (because it recorded in two different places the value have to be 2) and finally filter all the event that only have value 1.

The first step will give us all the ID that existed in both logs, the second step is to use the ID to look for the specific events. The new method need only around two to three minutes to completely give the value we looking for, even though it will absolutely bring more work to the user (2 vs. 1 step query). But looking from the efficiency point of view and considering of the small log events that can be correlated (Only around 10 events that can be correlated from 384.000 events on first log and 85.000 on second log), it was not worth the time it took by the first method to find them. Because of this, I think the second technique is more valuable, because it will cut drastically the time needed to do the analysis.

9. Assessment phase

This last phase was aimed to assess / evaluate the value that Splunk bring to the problem inside the case. I used two different approaches to complete this phase: The first approach was by analyzing what features Splunk has and what kind of solution that its features can bring to the problem. The second approach was by interviewing person who has knowledge of how the Oracle BRM system works about his view of Splunk. I will explain them in detail in this chapter, more information about what advantages that Splunk brings and what possibilities Spunk has can be found in the appendix D.

I had a good understanding what Splunk capable of after completing the proof of concept. From these capabilities, I try to analyze what kind of value that Splunk bring to the problem existing in the Oracle BRM log files. One of the added values from Splunk is the ability to manage unstructured data coming from the log files (more values can be found in the appendix D). The second approach was done by an interview I did with a Accenture's consultant, Peter van Gulik, who familiar with the technical and functional operations of Oracle BRM. This interview was began by me giving a presentation / demo of the capabilities of Splunk, and after that I asked his opinion what kind of useful features that Splunk can help with the Oracle BRM log files based from his experience.

All the information that I gathered from this phase are summarized into a product report. This report will give Accenture the overview of all the benefits that Splunk able to bring to the Oracle BRM case in this project. Besides that, I will also assess my experience with Splunk to see what are the values that it can bring on general. This was done because Splunk is a big data management tool which means it can manage all big data files and log files is just one example of them. This part can also be use give insight of what Splunk can be used for beside for the log files. Finally this product report also acts as the end deliverable of this phase which is also the final deliverable of this project.

10. Conclusions and Recommendations

Integration systems are used by the companies to integrate all their systems and applications. Integration is a crucial step because companies, especially the large ones are very dependent with IT to support their daily business operations. Business conditions are usually unpredictable which means there can be a lot of changes in short period of time, so IT systems should also be flexible to changes to be able to support them. Without integration, modifying disparate IT system will be nearly impossible, because of all the effort and time needed to accomplish this.

Accenture offers various integration systems packet from multiple vendors and as an IT consultant company; they will also help their client to implement it. Most of the time their client have a special wishes for tools that cannot be acquired easily from the default tools provided in the packet. This cause Accenture to custom made these tools for each client, even though some of them are commonly asked by the clients, which takes lot of time and cost a lot of money before the whole integration systems are ready. Accenture know that there can be a lot of time and cost saved if they can reuse tools into different client situation, and they want to assess the possibility of these reusable tools via a proof of concept which become my assignment.

During my research I discovered that there are a lot of 3rd party tools available in the market today. These tools are vary in size, license (open source and paid), and type (e.g. message queue, monitor, error handler, etc) are able to work along with most of the integration system available these days. After I had a discussion with my company tutor, we decided to build a tool that can help to monitor, notify, track, and report log files from the integration system. For this purpose we chose a tool called Splunk to be built into the proof of concept.

Splunk is a big data management software, that help to manage and index all machine generated big data (e.g. log files). We decided not to use dummy data to make the proof of concept look more real, instead my tutor was able to acquire several log files from a billing system use by a Dutch telecommunication company. The billing system they use is Oracle BRM who have very large (one log have size over 3 GB) and unstructured log files which makes them unusable.

After implementing the log files into Splunk, I was able to uncover a lot of valuable data, e.g. an overview of all the errors and what cause them. At the end of the project I also did an assessment of all the benefits Splunk can bring, and I can conclude that Splunk can bring values to all company that need helps in managing and analyzing their machine generated big data. Finally, after the approximately 5 months doing this project, I am also able to conclude that using 3rd party tools along with the integration system to be able to fulfill Accenture's client special wishes is feasible.

To close this chapter I have two recommendations for Accenture that I can give based on my experience from this project. The first one is about the usage of Splunk in general, Splunk is a very good tool with very intuitive UI which make it easy to use and learn. This is a perfect tool for companies that need help with big data, because it will not only bring benefits to technical but also to the business side of a company depending on what kind of data uploaded into Splunk,. The second recommendation is to make use of 3rd party tools for type of tools that are not provided in the default integration package to help Accenture save time needed to deploy a client solution. From my experience in this project, I was able to make Splunk work with Oracle BRM logs within a month and I believe this will also apply to other 3rd party tools. The time that saved in here will also ultimately reduce the whole cost of implementing an integration system.

Evaluations

Working at Accenture was a priceless experience for me, because I am able to work with integration system for my graduation project which was also an amazing opportunities for me to have more knowledge on that field. At the beginning of the project, I was still a bit vague of what I have to deliver in the end. This was caused by the complexity of integration system and my lack of experience with them, that's why I took some time to do basic research about the integration systems. The theory of ERP that I got from school helped speed up the process of my research, I can understand the basic of integration systems within 2 weeks after the start of my project .After I get the basic understanding, the project became clear to me and I can began drawing up the project plan.

After my project plan was complete, I began the first two phases with researching. It was a little bit difficult at the beginning because the field that I researched was not well known on the internet; this means that I was not able to find information easily about the subject. I need to rephrase my search query multiple times before I can get the information that I looking for. From this research I discovered something interesting, there is still no naming standard for components in the system integration field (e.g. load balance component is also called network balancer in some company) which can lead to confusion.

The most memorable moment was when I received the log at the first time; I was surprised seeing that it has 3 GB in size. It took me around few hours to be able to open that log files, and few days to be able to load it into Splunk. It was also an interesting moment to saw what Splunk can do with those log files via their graphical user interface, because usually log files are text based which means people have to fully understand the syntax used to be able to analyze them. Other memorable moment was the positive reaction from other consultants when I gave them a short demo of the proof of concept I made.

Finally the creation of this report was a bit challenging for me. I am not good at explaining things, and there are a lot of terms and abbreviations in this report which make things more complicated. But with this report, I also learned a lot about how to make a good report which will be very useful for me in the future. Beside that I am fortunate to get help from my school and company tutor, and also from friends who I've asked their time to read and give feedback to my report; which in the end made me able to delivered this detailed report of my project. Overall, I was happy with all the experience that I got from working on this graduation project at Accenture. Beside experience, I also got knowledge and skill that will be very useful for my career in the future.

References

[1]http://www.accenture.com/us-en/company/overview/history/Pages/index.aspx

This is the official Accenture USA company website, the part in the report about history of Accenture were rephrase from this link

[2]http://www.accenture.com/us-en/company/overview/description/Pages/index.aspx

This is the official Accenture USA company website, the part in the report about the situation of Accenture nowadays were rephrase from this link

[3]http://www.accenture.com/nl-en/pages/service-ats-technology-solutions-summary.aspx

This is the official Accenture Netherlands company website, the description of Accenture Technology Solutions in the report were rephrase from this link

[4]http://www.mulesoft.com/resources/esb/enterprise-application-integration-eai-and-esb

Mulesoft is a company that offers ESB software packet as their main product. The information about the usage of ESB inside the integration system today was referenced from their site.

[5]http://en.wikipedia.org/wiki/Comma-separated_values

A link to Wikipedia site that explained about CSV files, the description about CSV in this report was summarized from this page

[6]http://publib.boulder.ibm.com/infocenter/spsstafs/v4r0m1/index.jsp?topic=%2Fcom.ibm.spss.tafs.help%2Ftm_import_cf_flat.htm

This is the URL from a help file from one of the IBM products, there is a short information in this site about the flatlist file format that was referenced in the report

[7] http://en.wikipedia.org/wiki/Sumo_logic

An URL to Wikipedia site where the short description about Sumo Logic was summarized

[8] http://www.sumologic.com/product/capabilities/

The page above is the official Sumo Logic website, features of Sumo Logic that are written in this report were referenced from the URL above. Some of the detailed explanations were also taken from Sumo Logic manual book

[9] http://en.wikipedia.org/wiki/Splunk

A link to Wikipedia page that contains the description of Splunk which was summarized to be used in this report

[10] http://www.splunk.com/view/splunk/SP-CAAAG57

URL to the official page of Splunk where its features are referenced to be used in this report. Some of the detailed explanations were also taken from Splunk manualbook

[11] http://www.oracle.com/us/industries/communications/comm-brm-management-wp-1637483.pdf

Link to the Oracle website that holds the detailed overview of their BRM products, this overview were summarized and used to introduce Oracle BRM in this report

[12] http://www.regular-expressions.info/tutorial.html

URL to the regex official page, the detailed basic information of regex in this report was referenced from this site

List of figures

Attachment List

Appendix A: Project Survey

Appendix B: Project Plan

Appendix C: Detailed information about all the components found during Initiation phase

Appendix D: Information from product report about advantage that Splunk can bring

Appendix A : Project Survey

Fontys
University of Applied Sciences

Appendix A : Graduation Project Survey
HBO-ICT: English Stream

Data student:
Name student : Initials: F. Name: Prayuda
First name:Fransiskus Studentnumber.: 2157006

Telephone: 0647221292 E-mail.: p.prayuda@student.fontys.nl

Data company:
Name company/organisation: Accenture Nederland B.V.

Visiting adress : Versterkerstraat 6, Almere

Company mentor : Initials: G.M. Name: Guy Hagemans

Telephone: 0652367895 E-mail.: guy.hagemans@accenture.com

Department/ position: Advanced Technology and Systems / Consultant

Start date Graduation project : 1 February 2013

Duo Graduation project : ~~Yes~~ / No

If duo name of buddy: -

Accepted by student: 14-11-2012 **signature:**

Accepted by company: date: 13/11/12 **signature:**

accenture
Gustav Mahlerplein 90
Postbus 7799, 1070 AT Amsterdam
Tel: +31 20 493 83 83 · Fax: +31 20 493 80 80
accenture.com

Hand in date Graduation Project Survey:

Approved by graduation project coordinator: yes/no date: signature:

Remarks : _____

PLEASE SEND THIS FORM BY EMAIL TO THE INTERNSHIP COORDINATOR IMMEDIATELY AFTER THE INTERNSHIP INTERVIEW HAS TAKEN PLACE.

Description of the graduation project:

1. **Describe the problem analysis:**
   Enterprise Integration Platforms are extremely important in today's business; applications need to communicate with each other within the Enterprise Domain. There are many software vendors who provide out-of-the-box toolbox that can be used to build-up any given integration solution.
   But most of the time, Accenture still have to build and rebuild common components that go on top of the default toolbox provided by the software vendor; for example: Error-Handling, Logging, Implementation Patterns, Development Tools etc. Nowadays time to market becomes more and more important and Accenture decided that there can be a lot of re-use and time-saving if they build up a fixed library of re-usable components and standards that can be implemented in different client situations.

2. **Describe the graduation assignment.**
   The objective of the assignment will be to set up generic components and standards for a specific platforms that can be easily plug-and-played into any given execution architecture. Specific task that have to be done in this project will be listed below:
   1. Investigate availability of components - Enterprise Integrations platforms existing in the market today.
   2. Define a generic execution architecture common for any Integration Platform.
   3. Select one of the platforms which seem interesting as a proof of concept.
   4. Design and implement a set of components which can be used as a proof of concept.
   5. Asses the feasibility for the build-up of a common set of assets to be used across solutions.

   This project will also be a great opportunity for me to know more about enterprise domain because I am very interested about how an architecture, that used mostly by companies nowadays, works.

3. **What is the research component of this assignment?**
   There are three research areas that have to be done in this project, the first one is to investigate the availability of components within Enterprise Integrations platforms that exists in the market today. The second one is done after all the components ready to be deployed, this research is concerning which platforms (for example: IBM, TIBCO, SOFTWARE AG, ORACLE etc.) is good and interesting to be used as proof of concept. And the last research is to see whether the result of the proof of concept can also be implemented into other solutions.

4. **What are the methods and tools?**
   Accenture offers the following methods and tools: The Accenture Delivery Methods (ADM), Enterprise Search functionality and vendor documentation.

5. **How and by whom will you be guided by the company?**
   Who: Guy Hagemans
   How: Both face to face and via teleconferencing.

6. **What fields of Study play an important factor in realizing the graduation assignment?**
   There are several fields of study involved in this project. The first one is information analysis, I have to analyze the enterprise integration platform systems that are available nowadays, then figure out what components will be made and finally decide on which platform it will be deployed.
   The next one is design; like other software development project the components that will be made have to work as efficient is possible, this can be achieved through a good design. The last one is realization where the result from design phase is implemented and finally all the components have to be tested so it will run perfectly.

OTHER DETAILS:

Preference university tutor:
1. Ad Maas
2.

Appendix B : Project Plan

Project Plan

Enterprise Integration Framework

Author: Fransiskus Prayuda

Version 1.1 (27 May 2013)

Table of Contents

1. Introduction

Naturally, enterprise architecture consist from a large number of applications and systems that are relied by the company to conduct day-to-day business operations. A company can have systems that are acquired from either third party or developed in-house to support their employee information, costumer relationship, business logic and etc. In theory, breaking these complex business processes into smaller tasks are desirable to make the systems that support them easier to implement and more flexible to business and technologies advancement. This separation make enterprise integration platforms extremely important in business today as it helps applications to communicate with other systems in and outside of the enterprise domain.

Accenture are able to deliver integration packet (software and systems) from various vendors. Even though out of the box integration platforms provide a toolbox to build-up any given integration solution, most of the time they have to custom made some tools that go on top of the default toolbox by the client demand. Making these tools took a lot of time and on other hand the time to market is also important, which in the end initiate this project to start building a proof-of-concept from one of the commonly asked and valuable tools to see if Accenture can start reusing in multiple projects in the future. This document will serves as the base plan during the whole duration of the project.

2.Project Statement

2.1 Formal client

The formal client for this project is Accenture Nederland B.V. which is represented by Guy Hagemans.

2.2 Project leader

Fransiskus Prayuda will act as the project leader and is responsible for all communication between the project participants and the external parties.

2.3 Current situation

Today's medium to large business use a very vast application landscape to support their daily operations. This landscape can consist of more than one systems running at the same time. Recently there is an increasing need to have all those system integrated to make them more flexible and responsive to business changes. Usually integration system need more than one dedicated system / platform to fulfill the integration process because of the size and complexity of the data stored inside each system. The separation between the main and integration system is also necessary to keep the performance of the main system unaffected by the integration process.

Integration systems, or usually called Enterprise Application Integration (EAI), are one of the products that are offered by Accenture Technology Solutions. In this case Accenture can deliver a lot of integration platforms (e.g. IBM, TIBCO, SOFTWARE AG, ORACLE, etc.) because of the diversity of their clients. These integration platforms also include some default tools that can be used to support the integration process.

2.4 Project justification

Even though the integration platforms packet consist a lot of tools that can support the whole process, most of the time there are specific wishes from the clients for some types of tools that cannot be obtained from those default tool in the packet. Accenture has to spend a significant amount of time to custom made (sometimes even from scratch) those tools that are commonly used for several clients. Beside that the time to market is equally important, and Accenture know that time and money can be saved with the re-use of commonly used tools that can be implemented into different client situations.

2.5 Project product

Accenture Technology Solutions want to see a proof-of-concept from one of these tools. This working demo will also be used to see the possibility of reusing other tools into different client situations in the future.

2.6 Project deliverables and non-deliverables

The project deliverables for Accenture are listed below:

Deliverables	**Description**
Analysis Report	This will contain the result from the analysis phase which consists of the available tools / component that are mostly requested by the client and available in the market nowadays. This report will also contain overview of products that can be used to handle the tasks from those tools.
Product Comparison Report	This deliverable will give more explanation of the features from several products that are interesting to be implemented

Product Report	This report will give detailed information of the effect from implementing one of the tool into a real life case. This report will also explains if the tool that will be implemented is working well or not, and what are the advantages that the tool brings to Accenture and its clients.
Setup guide / manual to install the tool (Optional)	This deliverable will help the implementation / set up of the tool on another project in the future

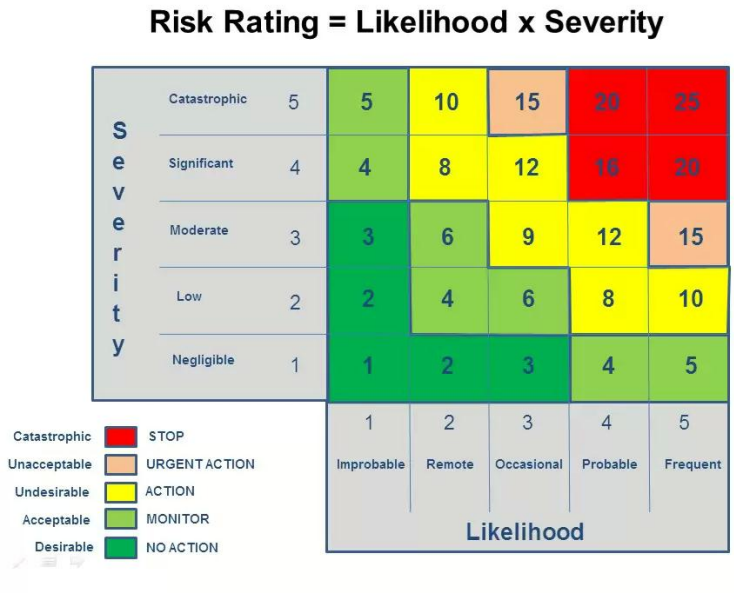
2.7 Project constraints

The project have to be completed in 100 working days (about 20 weeks) and the documents / report must follow the Accenture's format so it can be published on Accenture internal website so the information can be used around the company. Other constraint is some basic knowledge about the integration system itself (e.g. how the system work and what kind of data types being transfer from one system to another) because in the end the tool have to be able to work together with them.

2.8 Project risk

Risk	Severity	Likelihood	Mitigation
Insufficient work hours	Significant	Probable	Make a set of goals that want to be achieved every week and compare the project progress with the main project plan
Lack of knowledge about the commonly requested tools	Moderate	Improbable	Try to look for information about the components on the internet, look for book for reference or follow online training and ask among Accenture colleagues
Lack of communication with school and company mentor	Catastrophic	Probable	Make communication plan and try to contact them with email or telephone frequently, also make personal visits to Almere and Eindhoven
Changes in project requirement	Significant	Remote	The end product that must be delivered is already set from the beginning of the project
Availability of integration tools that can be used as proof of concept	Significant	Remote	Look in the analysis phase if there will be sufficient tools in the market to support the rest of the project
Availability of sample case that can be used for the proof of concept	Catastrophic	Probable	Communicate with the formal client few weeks before to make sure a case sample is available. If it is not available, dummy data also possible to be used for the proof of concept

Legends:



3. Project Phasing

The projects will be divided into phases for easier control, which a product / sub-product will be delivered at the end of every phase. I will use the standard waterfall method in this project because most of the phases need the information resulted from the previous phase. At the beginning of each phase there is also a possibility to meet with the formal client to talk about information of the products that will be delivered and some problems that occurred in previous phase that still need to be fixed. The project itself will be divided into 4 big phases, which are:

1. Initiation
2. Analysis
3. Build
4. Assessment

Each of the phases will be explained in more detail below.

3.1 Initiation

The aim of this phase is to gather more detailed information about the project which is: the way of integration system works and the tools that are mostly desired by the clients of Accenture. The steps that will be taken in this phase are listed as follow:

1. Find out the commonly requested type of tools by clients via Accenture employee
2. Do a market research for related type of tools that already at mature stage. This research will be done by using the sources listed below :
    - Accenture KX (knowledge bank)
    - Google
    - Accenture global network
    - If required, an interviews with one of the Accenture employee
3. Make a detail overview for every type of tools (e.g. what it is used for, what are the advantage it bring)
4. Decide with the formal client which kind of tool are interesting to be built in this project

The milestone for this phase is a report in which consist the list of type of tools that are gathered during this phase with a short description for each of them. This analysis report will also act as the end deliverable of this phase.

3.2. Analysis

The purpose of this second phase is to choose a specific product that in the end will be built as proof of concept. The steps that will be taken are listed below:

1. Look for products that can take care the tasks of the chosen type of tools
2. Experiment with them to know what are the features and highlights of each product
3. Make a list of criteria that can be used to compare all the products
4. Make comparison from all the products by using the criteria created in the previous step
5. Decide with the formal client which program will be used to make a proof of concept / working demo

The milestone for this phase is a comparison report which shows the key features of the related products, which also serve as the end products of this phase. The report will also give an overview to help the formal client to choose which program is interesting to be built in this project.

3.3 Build

This phase will be specifically used to build the proof of concept / working demo, the first step that have to be taken is to look for a sample case. This can be done by communicating with the formal client to look for a possible case that can be used for making the proof of concept. The next step after receiving the sample case is to start building the demo by following these steps:

1. Analyze the data from the sample case to look for possibilities of which features from the program that can be shown in the demo
2. If required, do an interview with other consultant who are experienced with the system to look for interesting data that could be used for the proof of concept
3. Implement them one by one into the program

The final deliverable for this phase is a working demo / proof of concept using the data that is provided by the client from the sample case. Because this phase will be quite long, the development will be divided into several milestones. Each milestone will deliver a prototype from the final product which can be used to get feedback from the formal client; this prototype will also allow the client to control the development progress. Finally, the feedback from the client can be used to further improve the quality of the next prototype so in the end we can have a good quality proof of concept.

3.4. Assessment

This is last phase and to do this phase all the program's functionalities have to be working completely. The aim of this phase is to make a product report that covers overview of the result from the whole project. The steps that have to be taken for gathering required information will be explain as follows:

1. Assess the added value that Splunk bring to the project
2. If required, consult with another consultant who understands the system from the sample case by giving small demo from the proof of concept.
3. From there, try to also assess the value that Splunk can bring to another project in general

All of this information will also be put into the product report which will act as the milestone and end deliverable of this phase. Finally, if there is still time left in the project a manual will be written to help another employee to use and set up the tool in the future.

4. Management plan

With the limited resources that are available for this project, this section will cover the plan to manage all the resources as efficient as possible and made all the deliveries that delivered have high quality. This section divided into 6 which are Money, Skills, Quality, Information, Time, and Organization.

4.1 Money

There is no money involved in the project, so this part of project plan is not applicable for this project.

4.2 Skills

Skills that are needed to complete all the deliverables on each phase of the project will be explained in the table below :

Phase	Skills needed
Analysis	Person who have knowledge about the existing tools the company have nowadays, which can be represented by formal client
Design	Knowledge about the system and requirements that are needed by the client
Build and Deploy / Implementation	Programming skills to make sure all the functionality working as desired

4.3 Quality

To ensure the quality of the deliverable, there are several techniques used in this project:

- Application quality will be ensure by using functional testing / black box testing to make sure the system are able to do what it required to do. Because of the system already built by the company that provides them, it will be a waste of time to do white box testing (unit and / or system testing) because of the complexity and knowledge of the source code needed from the tool itself.
- Project plan quality will be checked using S.M.A.R.T. which means:
    - All the phase will have their own period of time
    - After each phase, a check will be done to make sure the progress is on time
    - All the steps that will be taken will be discussed first with the client to make sure that the time can accommodate with the work load
- All document will be created using Accenture format so it can be used within the company

4.4 Information

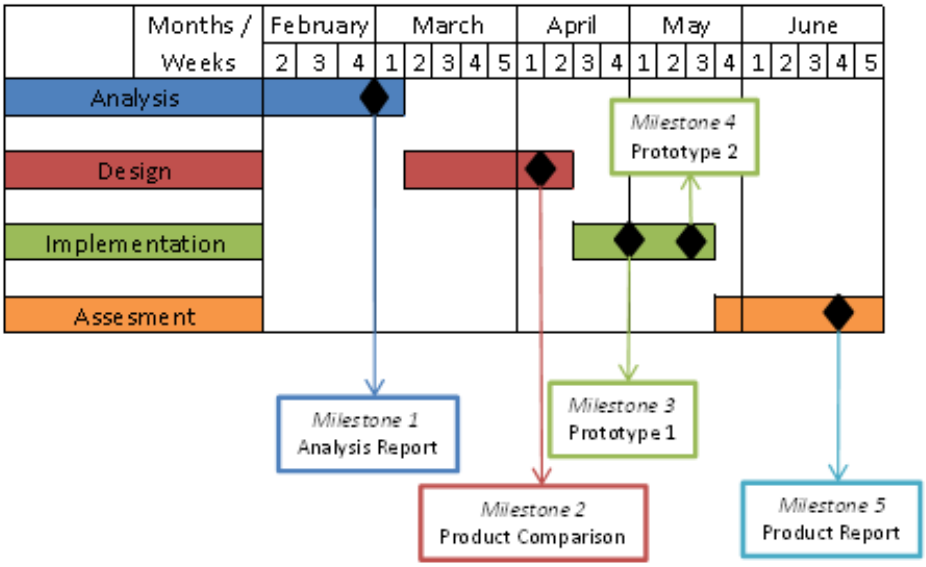
This part will explain the paperwork that resulted from this project and the role of each parties.

	Project Leader	Formal Client	School Tutor
Project Plan	Dr S Ar	Di A R	Di A R
Analysis Report	Dr S Ar	Di A R	-
Product Comparison	Dr S Ar	Di A R	-
Product Report	Dr S Ar	Di A R	-
End products	Dr S Ar	Di A R	-
Final Process Report	Dr S Ar	Di A R	Di A R

Legends: **Dr** Draw up **S** Send
Di Discuss **R** Receive
A Approve **Ar** Archive

4.5 Time

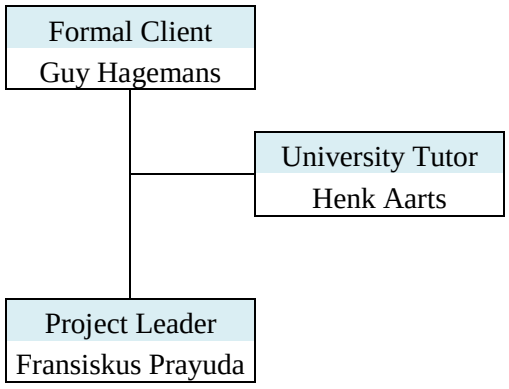
The project will be started on 5th February 2013 and finish on 30th June 2013, the Gantt chart below will explain in more detail how the time divided between each phases and when is each milestone will happened which mark the deadline of each deliverables.



The milestones are deliberately put before the end of each phase so if a problem arise, there is still time to take corrective action to fix the problem.

4.6 Organization

The organizational structure of this project is shown in the following organization diagram:



Position	Responsibilities
Project Leader	Responsible for the whole project progress
	Regular contact with the formal client
	Make appointment with the school tutor
Formal Client	Approve the documents done by project leader
	Accept the deliverables of the project
School Tutor	Gives suggestion / feedback over the documents and reports for Fontys

5. Communication Plan

Introduction

This Communication Plan mentions all parties with an interest in this project and the way in which they will be involved plus the means of communication that will be used.

Interested Parties in this project

Who	On Behalf of	Interest	Means of communication
Guy Hagemans	Accenture Nederland	Formal Client	Phone : 0652367895 Email : guy.hagemans@accenture.com
Henk Aarts	Fontys Hogeschool	School tutor	Phone : 0885071026 Email : n.aarts@fontys.nl
Fransiskus Prayuda	-	Project Leader	Phone : 0647221292 Email : fransiskus.prayuda@accenture.com p.prayuda@student.fontys.nl

Means of Communication

From	To	Information	Medium	Frequency of data
Project Leader	Formal Client	Process of the project	Report	At the end of each phase
Project Leader	School Tutor	Progress of the project / Log book	Report	Every 2 weeks
Formal Client	School Tutor	Overview of the assignment	Meeting	2 times during the whole project
Project Leader	School Tutor	Final Process Report	Email	At the end of every month

Appendix C : Detailed information about the components

Types	Usage	Benefits
Load Balance	Distributes work across multiple physical systems	Improve performance by distributing work into several servers
System Management	Provides health monitoring and deployment facilities	Monitor the status of the whole integration system
Audit	Records EAI system and service events, both normal and abnormal	Same benefits with notify, track, monitor, and report but in this case it involves all processes in enterprise system wide not only the integration ones.
Recovery	For a process that fails, this service recovers the state of the process to a previously known status that preserves data integrity	Any process that was in the middle of a process and fails will need to be recovered to a previously known status to maintains the integrity of the data
Parse	Takes a stream of input data from the network and creates structured data from it	The other part of ETL process, this components take care conversion of data(flat files, table, etc.) into a suitably format for transportation
Event Monitor	Monitors all system events except integration events	Overview of all process in the EAI that are not connected to integration process
Error Raising	Detects abnormal events and raises them as error events	Raising error that are detected as abnormal event to the handler
Error Handling	Processes an event that the Error Raising service has raised as an error	Handling error that raised by the error raising component
State Management	Manages the integrity of the status of processes	Overview the status of operations on applications that are integrated by the EAI system.
Resource Management	Enlists resources for the transaction manager	To maintain the status of the internal processes of the EAI system
Archive	Periodically archives data from the metadata services	Archive message during transport so if the system fails, it can be used as backup and recovery
Message Queue	Manages and orders the persistence of messages	A mechanism to ensure that data can be reliably stored until the EAI tool is ready to process the data further. This service is a basic element of asynchronous messaging inside EAI. It also allows the EAI tool to manage resources efficiently and makes the EAI tool more resilient and scalable than synchronous integration mechanisms.
Transformer	Uses the rules specified in the map of each data element to transform the contents of each element of input data to the corresponding element of output data	The main process of EAI which is basically a part of ETL process, transforms sources data, filter unnecessary data and finally load it to the destination systems.

Router	Allows the EAI service to represent a route to facilitate integration	
Filter	Provides a mechanism for users to filter out information from certain data	
Validate	Can be used to validate many elements of the data, such as syntax, format, and range	Make sure the data that is being transported is valid to make sure data integrity between systems
Map	After structured data has been validated, the Map service tries to map it to the output data	Useful to map the data that will be transported, so EAI knows what kind of transformation process the data needs
Configuration Management/Versions	Tracks configuration changes to an EAI component or modifications of EAI metadata, and maintains old versions	Manage modifications that happen to EAI, it's also useful to maintain old versions of the EAI
Notify	Electronically informs a business user of a particular event in the EAI tool service through a standard communication channel	Monitor all events that are happening inside the EAI, this information can be used to the report or notify components to the business user. In the end the business user can see any anomalies happened in the process and track back which processes cause it.
Track	Provides a way to analyze the log created by the Audit service to provide a trace of a complex series of related data	
Monitor	Monitors all processes in the EAI tool and tracks the service levels of the processes dynamically	
Report	Provides reports to both business users and system administrators	

Appendix D : Assessment over the benefit that Splunk can bring

Value that Splunk brings to the billing system

We already saw that log files from the Oracle BRM are very big and complex, which make them unusable even though it contains a lot of valuable information about the system. The first value that Splunk able to bring to the log files is the help manage them. After uploading the log files, Splunk will directly indexed all the log events and separate each log based on their time it occurred. In Splunk search menu, we can directly see a timeline with a graph bar of how many events occurred per day (time range can also be set). This feature will directly bring the next value which is helping to analyzing the log files; the timeline with number of occurrence will help analyzing the cause if there is an event spike on a specific time range.

Splunk search feature also capable of making summary with the help of (distinctive count, group by, etc) to further help with the analysis process. In the proof of concept I made a summary of all errors happened and what causes them, this summary will help developer to prioritize in which error / bug should be fix first. Besides that, Splunk also help to discover the underlying cause of several errors that were previously unknown (e.g. data constraint violations, errors in custom developed and extended modules, etc.). Another feature that will bring a big help on analyzing the log is the ability of Splunk to correlate events between two logs.

Integration system in nature is a very complex system, there is a possibility that an error happened in one component are caused by another component and vice versa. Without Splunk this correlating task will be very time consuming because not all log events can be correlated. But with a correlating query, Splunk is able to return log events that are already correlated from multiple sources and this will bring a big help in analyzing complex errors that happened within the system.

Beside of all the features explained above, Splunk also have real-time monitoring and alerting capabilities which are able to further improve the quality and stability of the whole integration system. Finally Splunk can bring changes to the methodology developer investigate error from reactive to pro-active approach. Usually developers investigate using top-down method by looking at functional error to find the underlying technical errors that cause them. By using Splunk, developers are able to discover technical errors and fix them before they are causing functional error. In short, the developers can maintain the system more efficiently and stay ahead of all the errors. In the next chapter we will see on how Splunk can also benefit to other type of files it indexes.

Valut that Splunk can bring in general

Splunk is a big data management which means it can index not only log files but all kinds of machine generated big data. In the previous chapter we already saw what kind of benefit Splunk can bring to Oracle BRM's log files; in this chapter we will see what value Splunk can bring in general (big data) view. As a big data management, the first and main usability of Splunk is to manage big data and help user to analyze / gather information from those data.

Beside the search feature that already mentioned above, Splunk also feature a very interactive user interface. The user interface is made very simple so that the users who have no previous experience with Splunk can still use the program intuitively. This simple yet powerful user interface offer a big help in simple data mining operations, because the user will still be able to find the data he is looking for even without using the search language that Splunk has. Even though the search language is still needed for complex data mining operations (e.g. making summary or overview of a category).

In the previous section, we already saw about correlating data between two logs files. But Splunk capability of correlating files is not only for log files, Splunk also able to correlate value from extern data storage file (e.g. database file, csv file, etc). With this capability, if there are business data in the log file we can correlate them with more detailed business information which makes it possible to do business analysis and monitoring only from the log files of the system.

Finally, Splunk has monitor ability which able to help us keep track of all these business and technical information. This monitor will also be able to be put into dashboard along with other monitor from various sources that we made to give quick overview of all things happening inside the system or with the business itself. Moreover, the monitor feature are also able to give alerts which means we can set Splunk to let it monitor for certain category and let Splunk send an alert to us if a certain value in that category is reached. From all the added values explained above, we can see that Splunk will bring benefits to any kind of business by helping them manage big data that are generated by their systems.