Institute of Parallel and Distributed Systems

University of Stuttgart

Universitätstraße 38

D-70569 Stuttgart

Masterarbeit Nr: 3585

Analysis and Design of Document Centric Workflows for automating tasks in a multi-tenant Cloud archive solution

Pradeep Parameshwaran

Course of Study : INFOTECH

Examiner : Prof.Dr.-Ing.habil.Bernhard Mitschang

Supervisor **:** Dipl.Inf. Tim Waizenegger

Commenced : October 28, 2013

Completed : April 29, 2014

CR Classification: C.2.4, D.2.13, H.3, H.4.1, K.6

ABSTRACT

Information Lifecycle Governance (ILG) is a cross functional business initiative intended to align the cost of information with its value to the enterprise, increase transparency and control and reduce the risk of legal and regulatory obligations for data. It is this dynamic workload system that enables the users to analyze, formalize and optimize for a cloud environment such for being able to provide a fully managed "Archive as a Service" in private and public clouds. In this context of the Master Thesis a research on the possibilities on how to improve and optimize the information lifecycle governance workloads especially in the context of cloud environments. It looks for a formal definition of the individual ILG workflows using Process management concepts with a Process Engine can be used. The main goal is to allow the definition of generic ILG tasks in a declarative way and to guarantee transactional integrity and check-point restarting capabilities. An end user subscribes to SaaS archive service in the cloud has to move data off-premise and delete data management processes to the service provider without comprising data security and privacy. The first scenario is to evaluate on various workload management solution with document centric workflows. The second scenario to investigate describes the use case where a recurring batch load system periodically imports valuable business data in to the SmartCloud Archive. The thesis also proposes the architecture for the required uses to create the batch load and disposal sweep tasks in an enterprise perspective by eliminating administrative client for SmartCloud Content Management System. The architecture proposed moves the data off the premise into a cloud environment and thereafter managed in an automated way. The management of the data had been made to flexible, easy, reliable and efficient.

Keywords: data archive, SaaS, workload management, batch load, document centric workflows

TABLE OF CONTENTS

List of Figures

List of Tables

1. INTRODUCTION

Information Lifecycle Governance (ILG) is a cross functional business initiative intended to align the cost of information with its value to the enterprise, increase transparency and control and reduce the risk of legal and regulatory obligations for data. Its target is to reduce litigation risks and consumption of storage and infrastructure and helping to optimize its unit cost to the utility of the information stored. ILG solutions are built around the concept of an unstructured data archive and its main purpose is to locate, identify and classify unstructured data coming from Email, Fileshare, and SharePoint or from legacy Content Repositories. All unstructured archive tasks are driven by an elaborated workload management system that has to perform activities like: load, classify, index, transform, archive, retain, secure, report, notify, export and delete. It is this dynamic workload system that is used to analyze, formalize and optimize for a cloud environment such for being able to provide a fully managed "Archive as a Service" in private and public clouds. In this context of the Master Thesis a research on the possibilities on how to improve and optimize the information lifecycle governance workloads especially in the context of cloud environments. An approach for a formal definition of the individual ILG workflows using Process management concepts with a Process Engine can be used. The main goal should be to allow the definition of generic ILG tasks in a declarative way and to guarantee transactional integrity and check-point restarting capabilities. An end user subscribes to SaaS archive service in the cloud has to move data off-premise and delete data management processes to the service provider without comprising data security and privacy. The first scenario is to evaluate on various workload management solution with document centric workflows. The second scenario to investigate describes the use case where a recurring batch load system periodically imports valuable business data in to the SmartCloud Archive. The thesis also proposes the architecture for the required uses to create the batch load and disposal sweep tasks in an enterprise perspective by eliminating administrative client for SmartCloud Content Management System. The architecture proposed moves the data off the premise into a cloud environment and thereafter managed in an automated way. The management of the data had been made to flexible, easy, reliable and efficient.

1.1 Information Life Cycle Governance (ILG)

ILG is just more than an acronym in the enterprise context. It is the best way of managing the enterprise related information over time with the necessary control to advance good business practices [46]. It has certain policies, processes, technologies and practices which are used to manage the stored information throughout its creation and its end with deletion or disposal. This seems to be simple task but there is lot of complexity involved in it. Many organizations struggle with the complexity of managing lot of electronic information more laws that dictate how information is managed and greater consequences by failing it to get it right. ILG is responsible for promoting the business productivity and efficiency of an organization. It allows the organization to be legally compliant in case of information management.

In the broader enterprise sense information is the data in any electronic or paper form which is created or received by an organization during its day to day business operations. There will be some information which has greater value and some may not be useful after few years. Thus the value of the information is determined based on its content rather than the location where it is stored or the system through which it travels. Our business world is replete with more information in more and more place screaming out for better management. The value of the information is determined either during the generation of the information and reevaluated in the near future to check the validity of the concerned information. As soon as the information is created or received by an organization, the value must be immediately identified so it can be properly managed in accordance with all the governing rules utilizing the right technology.

The important thing that is clear from these aspects is that an organization cannot keep everything for a long period and expect to continue to be an efficient business; because there are also an increasing volume of laws and restrictions which emphasize the organization to properly manage the information. As a counter effect each organization need to actively manage their information if they are going to continue to be an essential asset for running an organization by abiding the laws. As the information technologies moves forward, the problem of maintaining enterprise information is still getting worse [37].

"If we look strictly at the growth of digital information, estimates indicate the total stock of information is growing over 50% a year"[46]

Enterprise data, measured in terms of data volume, is estimated to be growing at 45% a year [46].This drastic growth of information is mainly due to the factors driven by government regulations, legal issues faced by the organizations every year, legal e discovery mechanisms, IT services, automation and machine to machine communication etc.

"By an estimate, Walmart, whose data warehouse already exceeds 500 terabytes or half a petabyte , has increased its data storage capacity by an average of 32% each year since 2009"[46]

It is axiomatic to state that future IT capabilities will be different and driven by a better and more effective use of information in the firm. Information Life Cycle Governance (ILG) is a policy based IT systems strategy conceived by the data storage industry to address the problem of enterprise data mobility and storage based on the value of the data. It's an ambitious concept, one requiring important changes in the management, data policies and hardware and software technologies to be effective [7].

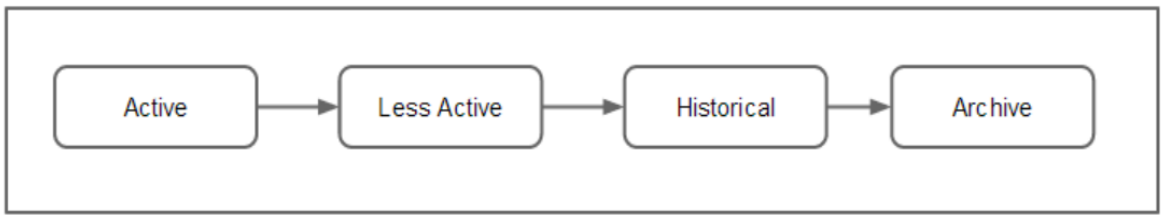

Figure [1] Lifecycle defined by Data Access (Frequency)

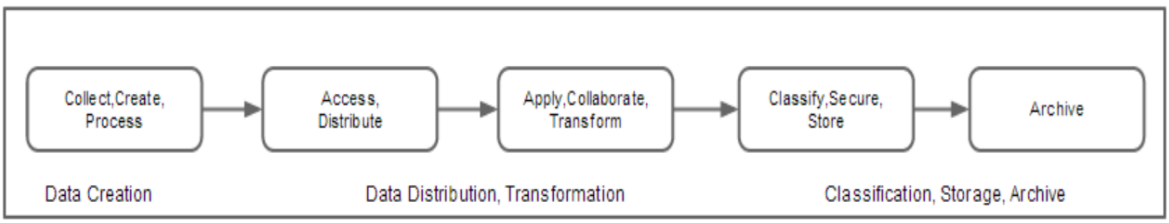

Figure [2].Lifecycle defined by Data Archives

The above figure describes the way in which the information is generated and retained. As discussed earlier the necessity to retain the information increases when its value increases. IBM has developed a solution for the ILG solution stack which is into production already. IBM provides the ILG solution stack in the form of IBM SmartCloud Content Management (SCCM), wherein the users can have an automated governance of information in an enterprise. Before describing about IBM SCCM system its necessary to know about the cloud computing aspects used in an Enterprise.

1.2. Cloud Computing for Enterprises

Cloud computing is emerging as a persuasive and consumer driven utility and attracts interests from various technology experts and public in large scale. Cloud computing for Enterprises addresses the need for a single point of reference for state-of-the-art cloud computing design and implementation techniques. It considers cloud computing from the perspective of an Enterprise Architecture.

The Enterprise architecture provides a high level corporate view of an enterprise. The concept of Enterprise Architecture and the associated architecture design principles and practice of service orientation has gained popularity of late. The present day Enterprise architecture is mainly focused on business process which takes us to the concept of Business Process Centric Enterprise Architecture. The primary objective of such a concept is to adopt a business oriented approach to integrally enhance and manage business processes by converting the actual business objectives into the realistic operations including the IT support. This kind of approach will result in the whole organizational processes to be organized by designing business processes in a more abstract manner without implementation considerations and more focused on design of the relevant information systems and the organization in parallel.

1.3. Dynamic Workload Management in Cloud

A workload in cloud computing can be defined as a set of batch jobs and programs executed in real-time and adapted to the modern data center. Workload management is the evolution of job scheduling with workload management capabilities for the dynamic data center. These workload management capabilities can be designed using the business processes. The aspects of scheduling workloads here include automatically resolving complex dependencies on various platforms and application tiers and triggering

workloads based on both IT and business events. This Master Thesis deals with analyzing the problems in the Dynamic Workload Management in an Enterprise Content Management system and proposes a framework to have an efficient workload management system. It also considers some use cases from IBM SCCM system and discusses about the workload management solution implementation on top of the IBM SCCM System.

1.4. IBM SmartCloud Content Management System

The IBM SmartCloud Content Management System (SCCM) provides a platform for managing the document compliance of an organization to avoid legal and corporate governance issues the companies might face in the future. The SCCM solutions are built on top of the IBM Enterprise Content Management Information Lifecycle Governance (ILG) solution stack. The SCCM provides the administrators a good control of the contents and its value to deal with. It also makes sure that the electronic information with increased value are searched, captured and retained for the organization [14]. The information managed are in compliance with record policies and the source of the information. IBM SCCM offers document and email content collection, preservation and lifecycle management such to automate the legal discovery process.

1.4.1 Services provided by IBM SCCM

IBM SCCM provides services for ILG solution stack for making bulk load operations into the cloud, deleting bulk data from the cloud and to automate the legal document discovery.

Batch Load:

The SCCM system provides the bulk load option to the users to upload the documents as batch files to the user. The client uploads the document batch files to the staging area of the private cloud by a batch load service running inside the content management cloud [14]. The client can monitor the batch load status and he can also access the batch load reports concerned to a batch load task. The batch file is a compressed content which has the files to be loaded to the cloud and an definition file which specifies the document classes, record category and meta data values for the files to be loaded.

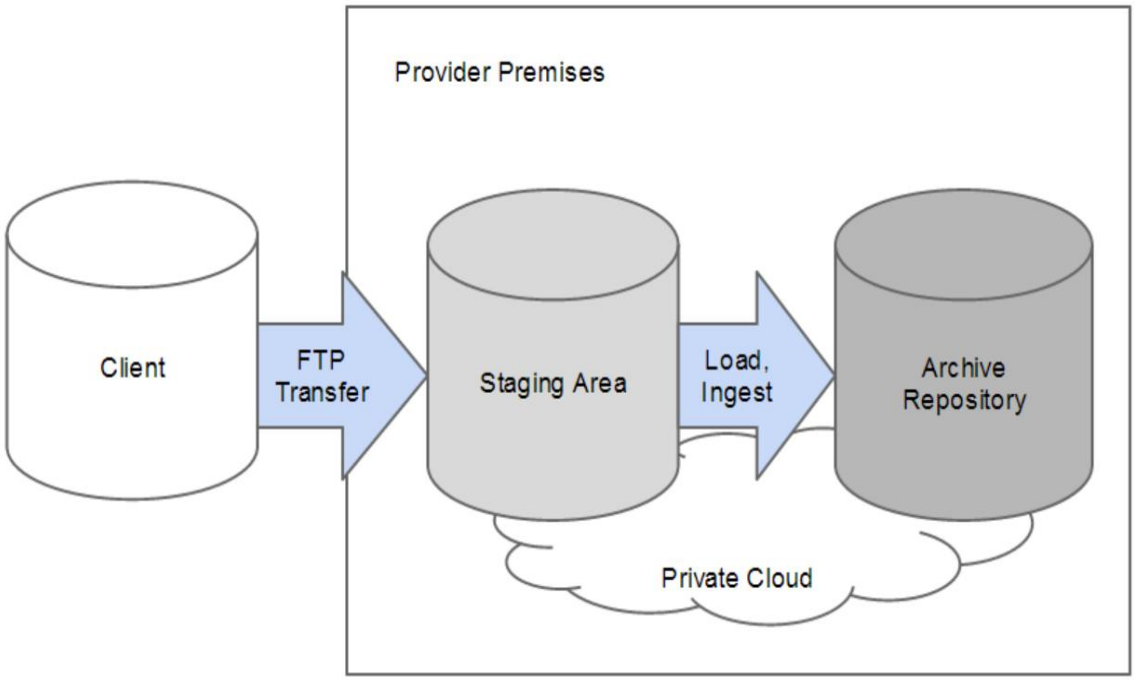

Figure [3]. Batch Load operation

The client can transfer the batch file contents to the landing area of the server with some FTP protocol. The batch loader then reads the definition file located in the compressed batch file contents and ingests the files into the Archive repository by keeping track of the Meta data of the ingested contents.

Archive Management:

In order to archive the contents in a repository the SCCM systems uses three main techniques namely Indexing, Classification and Retention [14]. The documents which are in the archive are indexed using some predefined document classes. These document classes follow certain hierarchy to form document class hierarchy. The predefined document class makes the job easier by avoiding us to create new document classes for each batch load. The document classes define the metadata associated with archived documents. The document classes contain numerous properties which can be used for indexing.

In case of classification technique a file plan is used for classifying documents for records management. It contains a hierarchy of records that serve as a container for filed records. The records in the archive have some retention policy which determines the

retention period of the record. Based on the retention period the disposition of the record is decided.

In case of Retention technique the records are governed using the retention policies of the record. A retention policy has certain set of properties like Aggregation, Retention event, Retention start date and Retention period.

Reporting

The SCCM system provides a set of predefined reports which can be used to verify the action status of the tasks which are performed on SCCM systems [14]. These reports range from Audit reports, System Task reports, Archive usage reports etc.

This thesis begins with the Introduction to the ILG solution stack and IBM deliverable called IBM SCCM system. Chapter 2 gives a brief introduction to the Enterprise Content Management operations and FileNet which is an Enterprise solution by IBM. Chapter 3 discusses the related works in dynamic workload management and gives an idea about the possible approaches to be followed. By defining the problem statement in the Chapter 4, its becomes necessary to discuss about Case Manager Integration points in Chapter 5 ,which helps us to understood the possibilities of solving the given problem. Chapter 6 explains possible use cases and solution framework followed to solve the workload management problem. Finally the Thesis concludes with the possible recommendations for future work.

2. CURRENT TRENDS IN ENTERPRISE CONTENT MANAGEMENT

2.1. Introduction

The terms often used for the dynamic contents are often confusing. In particular, the terms knowledge management, content management and delivery management are often misunderstood. There are discrete definitions for the above three but they conceptually overlap each other at times [24].

“Knowledge management” is the way and means that organizations create store and access reusable knowledge to accomplish enterprise goals. It has three possible sub sections: organizational processes and rules, knowledgeable and engaged individuals, appropriate technology to support knowledge sharing. Loss of knowledge can be prevented with the right collaboration and data storage tools.

“Content management” is the creation, publishing and management of company information and documents on the web. This can be done by distributing contents to the owners in each business unit. Therefore, “distributed content management” is decentralized content creation, publishing and management by multiple authors and groups throughout the organization,

“Delivery management” is often used in the web arena. It deals with the contents which are already prepared and making it available for access.

2.2. Enterprise Content Management Operations

Create

You will want your content owners to only edit their content, not the format or navigation. Start with a standard header and footer and plan out the overall navigation for the entire site. The site must look and work consistently across the organization. Next create page templates for content owners to use. The content owners can inherit policies to their documents so that its access can be restricted to certain set of users.

Approve

Most of documents generated need an approval process. You will need to provide a workflow to satisfy the business needs. Most business units have a business analyst who performs the content authoring task and will then want to have a manager to approve the

content before it goes live on the web. Most often the site owner or business analyst will need to assist the business unit in improving their workflow instead of mapping the existing cumbersome process.

Deliver

The deliver stage of content management is basically the process for how the content, which has been filled into the appropriate template, is published to the web. The more sophisticated software packages have automated publishing to production as soon as the content has been approved through the workflow.

Manage

The contents that are created in the content management system should be managed properly. All data must be easy to find for all users. The important factor which makes the searching of data for the users more simple is to have data tags (Meta data) for the contents.

2.2.1. Objectives and impacts of ECM Systems

Varying types of organizational objectives for ECM and resulting impacts were identified in the case studies:

- Improved internal and external collaboration, involved knowledge creation and sharing among the enterprises.
- Value-added or new customer services and products involving digital contents.
- Reliability and quality of information content which minimizes the errors in the services.
- Efficiency, effectiveness and flexibility of knowledge work and business processes, including reusable contents and templates.
- Meaningful knowledge work involving easier and less tedious human routines.
- Organizational memory recording the practice, history and transactions of the enterprise.
- Direct cost savings in information processing operations and facilities.
- Satisfying external regulations and standards directly or indirectly governing the enterprise's information management.

- Platforms and capabilities to develop and maintain targeted content management applications quickly for emerging purposes [8].

2.3. Components of Enterprise Content Management

The following figure depicts the components of an Enterprise Content Management system.

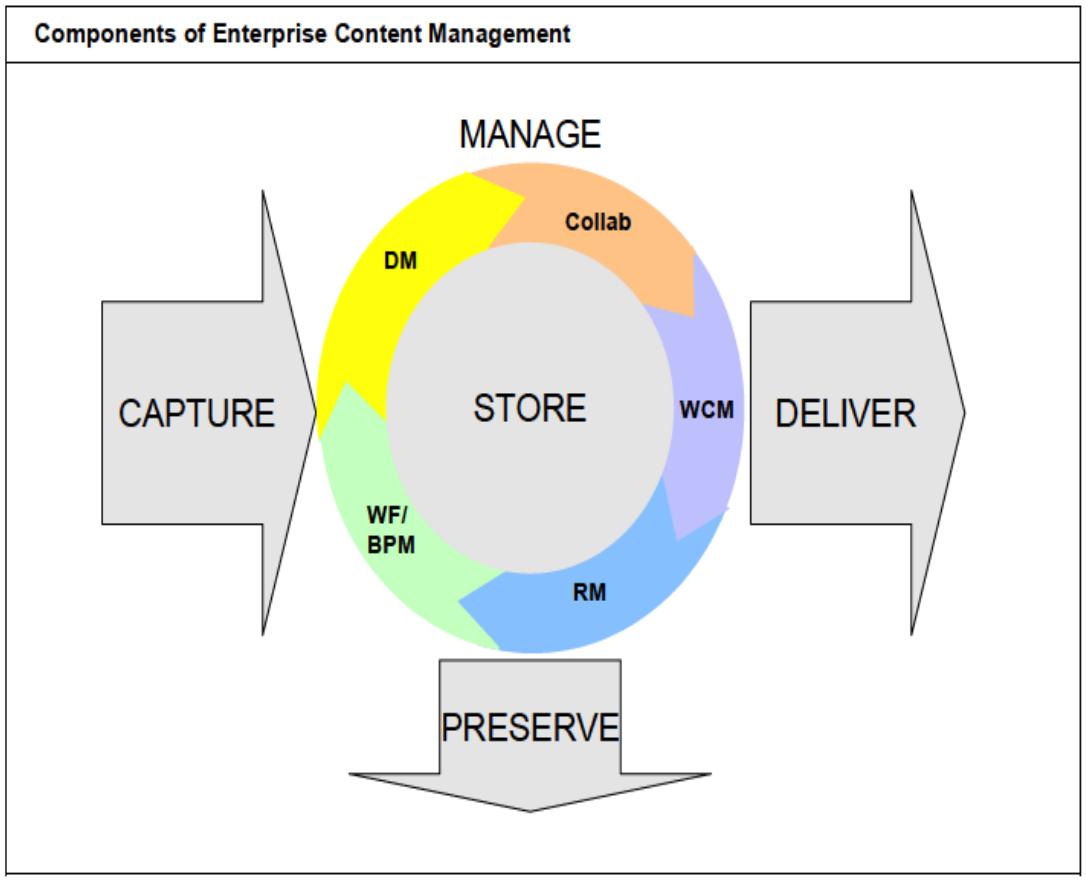

Figure [4].Components of Enterprise Content Management [8].

The individual categories and their components will be examined in the following:

2.3.1 Capture Component

The Capture category of ECM system contains functionalities and components for generating, capturing, preparing and processing analog and electronic information [8]. There are various levels involved to classify simple and complex information using various classification methods (especially automatic classification method). The capture components are otherwise called as Input components.

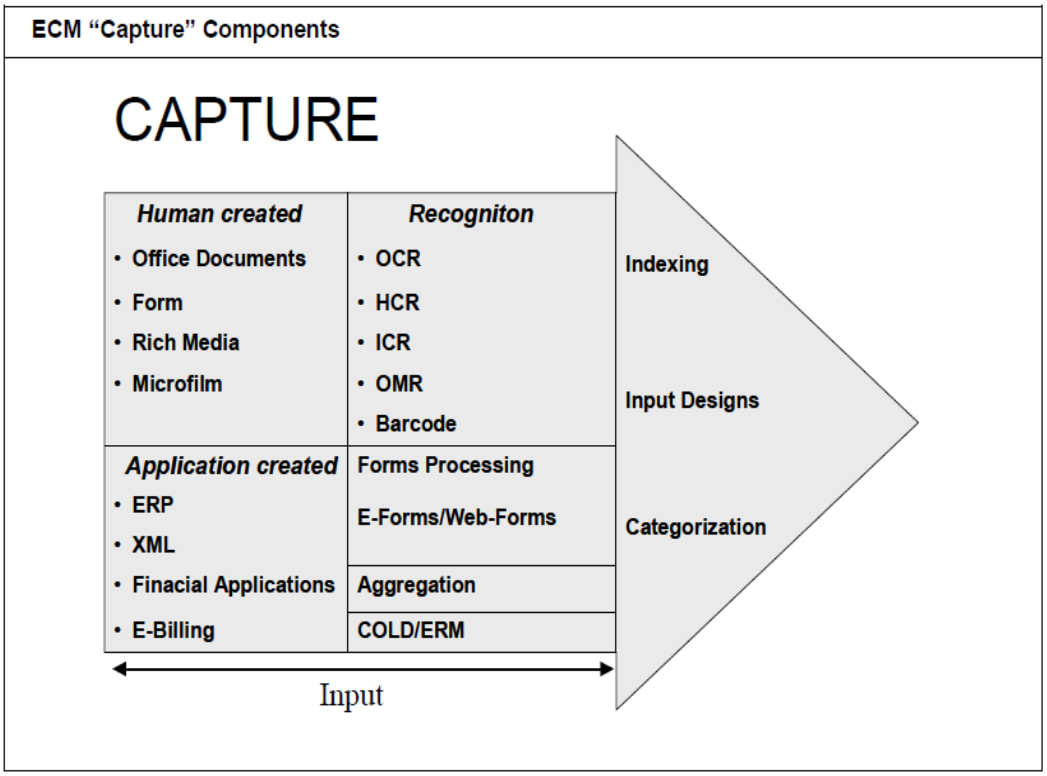

Figure [5].ECM "Capture" Components [8].

The manually captured information can be mostly involved with the paper documents, electronic documents in offices, emails, digital video and audio's. Thus this type of information should be manually classified or should be converted into a format which can be classified automatically. The automatic or semi-automatic capture involves usage of XML documents, some application like ERP systems to generate and capture information.

Nowadays all information is available via Forms and/or E-forms/ Web-forms. Thus capturing form information is an important aspect in the present enterprise systems. In forms capture there are two different technologies, although the information content and character of the documents may be identical.

- Forms Processing

  In this type of processing the capture of information can be done from industrially or individually printed forms via scanning. There are different image

recognition technologies which can be used to often capture the information from the scanned forms.

- E-Forms/Web-Forms
  In this type of processing the capture of electronic form data can be done by using an automatic processing unit which can often classify the data accordingly [8].

Components for indexing of captured information

The capturing of data from various sources should often be indexed. Thus ECM systems incorporate further components for indexing and gathering the information.

Some indexing techniques are discussed below:

- *Manual Indexing:*
  In case of manual indexing often there will some indexing attributes assigned to the information in the database. This is done by the database administrator.
- *Profile based Indexing:*
  Since manual indexing is often recognized as complex it's always easier to have some preset profiles. This makes the indexing much easier and less cumbersome. These profiles can describe document classes that limit the number of possible index values, or some criteria for indexing.
- *Automatic Indexing:*
  Based on the information in the web forms automatic classifier algorithms can classify the information based on document class, file category etc. These type of systems can engineered in such a way that the classification can be done based on some self learning algorithms [8],[24].

2.3.2 Manage Component

The Manage component are for the management, processing and use of information [8].It incorporates

- Databases for administration and retrieval, and
- Access authorization systems

The following figure depicts how the manage component of an ECM system is structured.

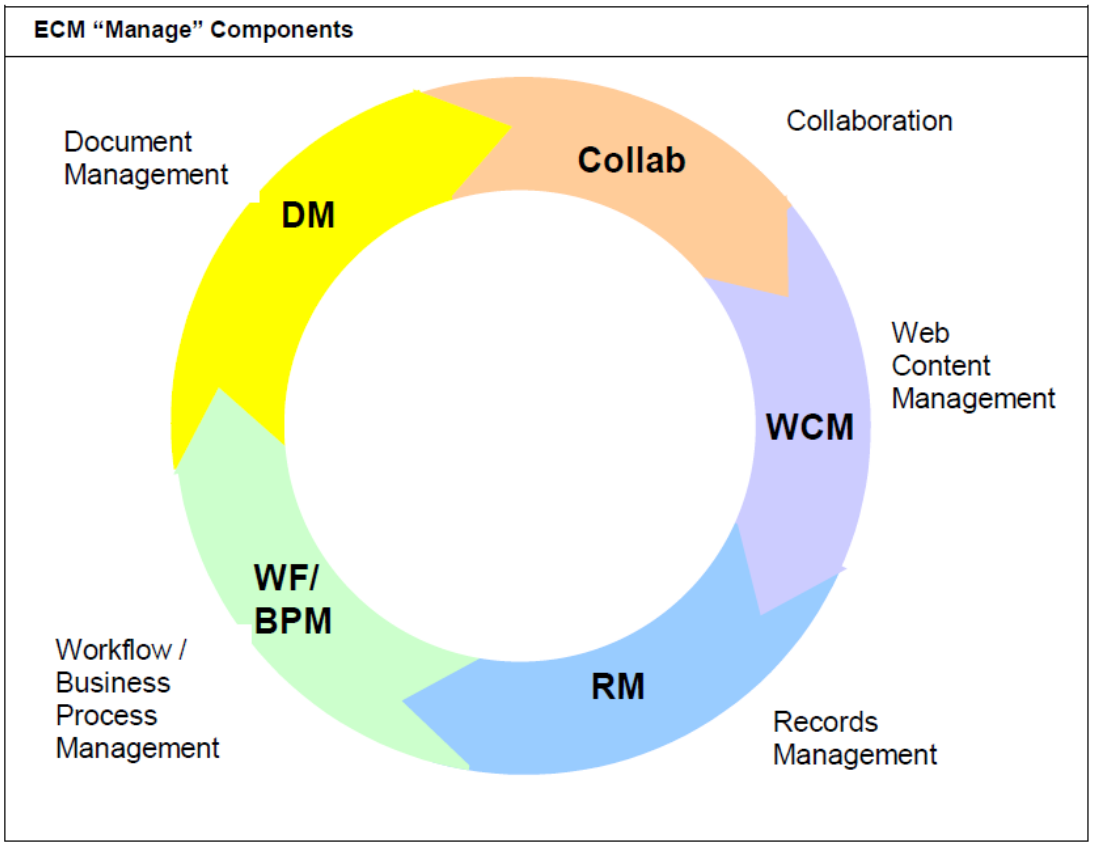

Figure [6].ECM Manage Components [8].

DM-Document Management

The Document Management provides the functionalities which helps in managing the documents in an application space. The main functionalities includes,

- Check in / Check out
- Version management
- Search and navigation
- Visualizing

Collaboration

Collaboration means coupling the components and its functionalities together. Collaboration helps in using jointly usable information databases, controlling the

information processing paradigm, administering information management components etc.

WCM- Web Content Management

Enterprise Content Management is tightly integrated with Web Content Management. Since the end user content operations are mostly based on web client based systems such as browsers it's always necessary to have a tight integration with the WCM components. The WCM includes functionalities such as creating or editing new or existing information in a controlled generation and publishing process, administration of web based information and its presentation, separate the authoring rights of the public and non-public information and visualization for Internet presentation (browser, HTML, XML etc).

Records Management (Archive Management)

Record management refers to the administration of records, important information of the organization that are supposed to be archived. Record management is considered as an independent storage media in an ECM system which manages the information about the records such as access policies, retention management and archiving policies etc. The components in the Record Management system use an standardized meta-data format to indentify the information. This avoids the ambiguity in the information that is retrieved [14].

Business Process Management / Workflow Management

The workflow components can be either Production workflow or Ad-Hoc workflow. The production workflow has certain set of predefined workflow steps which will be executed based on some events. But in case of Ad-hoc workflows the workflow sequence is determined on the fly by the user or the system based on some events.

The BPM components aims at the complete integration of all affected application within an enterprise [23].The main BPM functions are

- Complete workflow functionality
- Process and data monitoring at the server level
- Enterprise Application Integration to link different applications

- Business Intelligence with rules structures, integration of information warehouses and utilities that assists user in their work [8].

2.3.3 Store Component

The Store components help the enterprise to hold the information for usage temporarily. The store components can have three variants by which information storage and retrieval can be handled. The storage component can be a repository in a file system or databases.IBM ECM systems uses repository based storage and retrieval mechanisms. Other than repository there can be library services such as version control and audit trails.

2.3.4 Preserve Component:

Compared to the store components which is used for preserving the data temporarily where as preserve component stores the data permanently. This can be referred as "Electronic archiving".

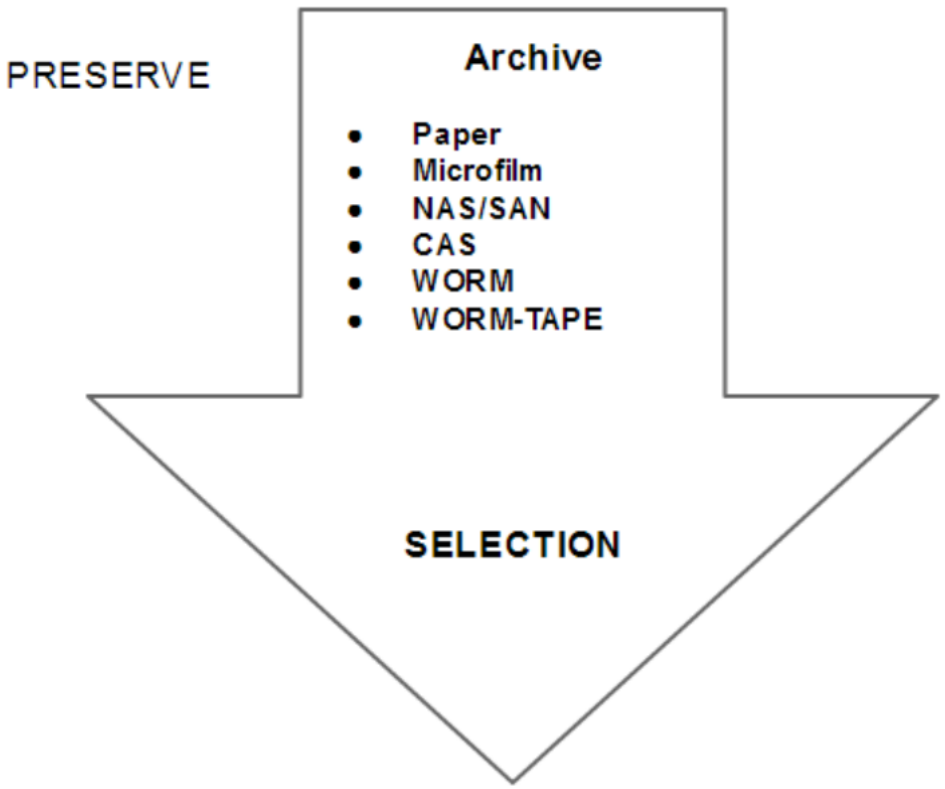

Figure [7].Preserve component

2.3.5 Deliver Component

The deliver component can be considered as an representation component where in its used to present the information from the “Manage”, “Store” and “Preserve” components. The deliver component has the functions which help to ready the data for the store and preserver components. It has two primary design aspects

- Layout/ Design : which has the tools for layout and formatting the output.
- Publishing: which has the applications responsible for presenting the information

2.4. FileNet Content Manager

IBM FileNet Content Manager was mainly developed to provide extensive document handling capabilities for an Enterprise. It provides full content lifecycle and extensive document management capabilities for digital content. This chapter discusses about the FileNet Platform and IBM FileNet Content Manager [22].

IBM FileNet Platform is a unified enterprise foundation for the integrated IBM FileNet products. It includes enterprise content management, business process framework and extensive compliance capabilities to address a wide range of content related business requirements.

IBM FileNet Platform is the key element in creating the adaptable enterprise content management environment necessary to support a dynamic organization that must respond quickly to change [22].

Businesses today face an information explosion due to plethora to data and are struggling to manage these unmanageable data. Mainly data which we consider here are documents which are unmanageable and unsearchable documents. This volume of information leads to several obstacles like productivity issues, storage costs escalation and legal issues related to disposal of data etc. Thus the enterprises requires a solution that is reliable, actively available and consistent which leads to the enterprise content management system.

FileNet is an enterprise level platform with modular solutions that manage content, optimize business processes, and enable compliance and smart archiving. The FileNet Platform is agile and highly scalable. It enables organizational leaders to capitalize on

this information by aligning its use to business directives. By implementing Enterprise content management strategies these business benefits can be achieved in a cost effective way.

The FileNet System is capable of analyzing and providing a deeper insight into the knowledge stored. It also helps to index and classify so that the data can be found intelligently. The FileNet helps to keep important information automatically and dispose of it at the end of the life cycle.

It also provides integration to provide investigation tools and reports to support auditing and business intelligence. The content life cycle management capabilities are expanded with the capabilities such as Workflows, content analysis and classification, and business intelligence that raise solutions beyond the file shares. Now FileNet platform has become the integral part of the corporate operation and decision making process.

2.5. IBM FileNet Platform Overview

The FileNet Platform is a unified enterprise content management and business process management platform that supports all of these requirements. It resolves content, process, compliance and discovery issues in one single solution. The solutions tight integration increases the efficiency of operations by reducing the usage of different components for content management in an enterprise.

The platform architecture provides interoperability across standard organization applications, file stores and data repositories, which reduces integration costs, increases accuracy and accelerates values. The integrated approach effectively addresses the complex demands of managing content across an enterprise [22].

2.5.1 FileNet Enterprise Reference Architecture

The FileNet Platform is implemented based on the generic Enterprise Reference Architecture which contains the blueprint for designing content and business process management solutions. The reference architecture has various blocks ranging from management services to development services.

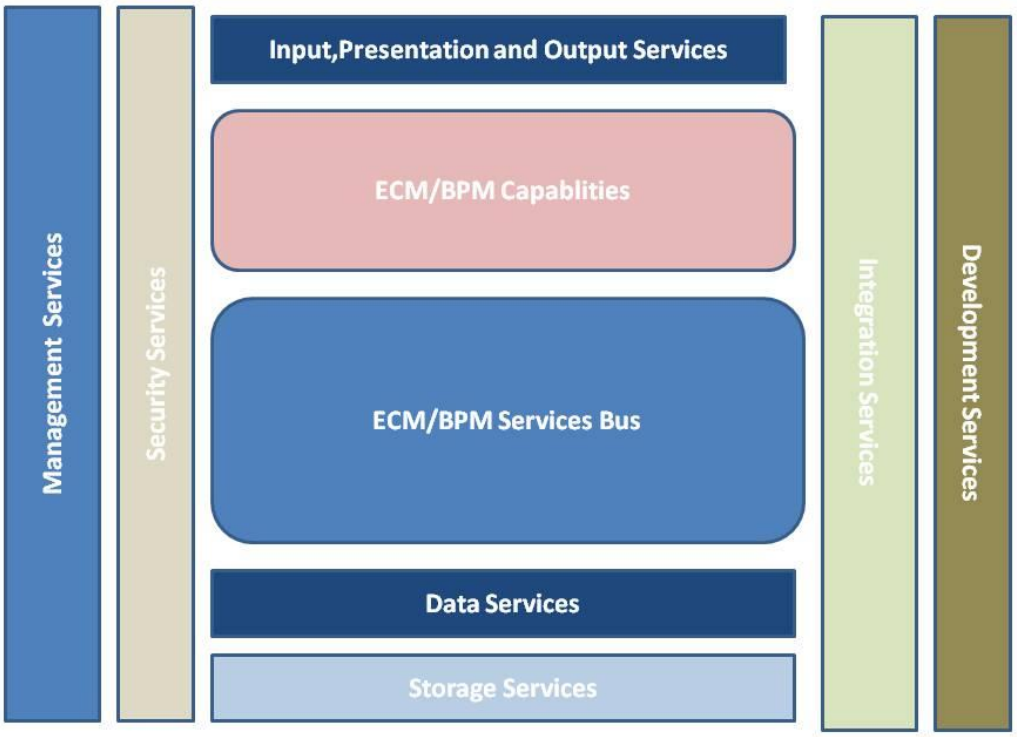

Figure [8]. Enterprise Reference Architecture [22].

The high level architecture has seven different blocks grouped into two areas:

- Central block which defines capabilities and services:
    - Input, presentation and output services: User interfaces, application building blocks, capture and acquisition of content and output
    - ECM/BPM capabilities and services: Core of FileNet Platform, providing content management and business process functionality
    - Information infrastructure: Data and storage services
- Vertical Blocks:
    - Management services: Process and system management
    - Security services: Access control and management
    - Development services: Expedites application creation
    - Integration services: Integration with data systems, repositories and other applications

Central capabilities and services

The layers input, presentation and output services, ECM/BPM capabilities and services, and the information infrastructure are the center of the framework. Starting from the bottom up, they support each other in forming the innermost structure that supports the solution architecture.

Information infrastructure

The foundation of the architecture contains data services and storage services. Data Services include integrity, encryption and other functionality related to data integrity. The data service layer abstracts call to data stores and applications. Storage services provide an abstraction between FileNet and a variety of storage, data and content cache services.

ECM/BPM capabilities and services

Building on the information infrastructure, Content and Process Management Services provide content and process management modules. These services are provided by core engines [23].

Input, presentation and output services

Input and Presentation Services collect, capture and archive assets from a variety sources into the ECM system. This layer also provides display and interface support for users and applications.

Management services

Management Services include system and process management and also give insight into the health of the system through monitoring tools. High availability, disaster recovery and capacity planning functions are located here as well as cloud support.

Security services

Security Services include modules for identity management, perimeter and security models.

Integration services

The Integration Services layer includes APIs, SOA and orchestration functions, Tools, such as Master Data Management, Business Intelligence and collaboration offerings, increase insight and productivity in business operations.

Development services

The Development Services layer has located process, application, form, object, and design and development tools. Test and debugging features and application deployment structures accompany this rich tool set.

2.5.2 IBM FileNet Platform core components

There are three core components in IBM FileNet which are basically formed by ECM/BPM Capabilities and Services.

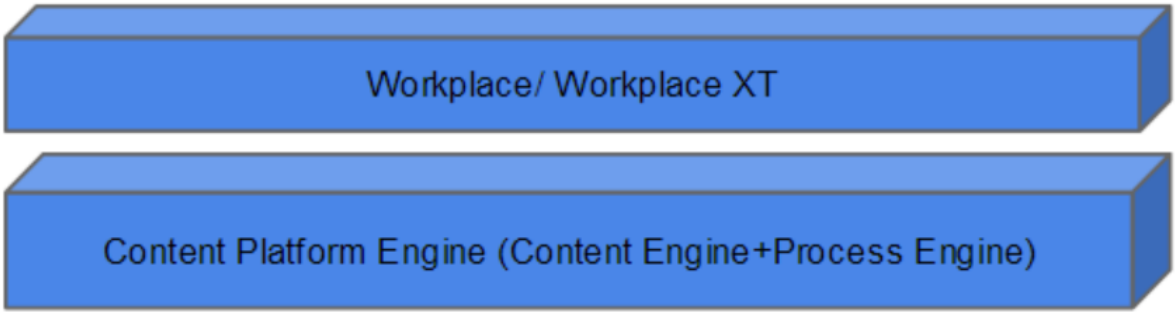

Figure [9]. Core FileNet Platform Core Components

- Content Engine: Provides the core content management capability, including versioning, security, and life cycle management.
- Process Engine: Provides the core workflow capability (business process management) including modeling, electronic forms and monitoring tools.
- Workplace / Workplace XT: Provide user interface to access content and process engines [22].

2.5.3 FileNet Application Interfaces

The platform has several application interfaces which goes straight head with the platform components. The application interfaces provides the framework or integration

points to integrate the FileNet components with external applications. The application interfaces helps IBM FileNet to integrate with the Microsoft office so that contents can make use of Microsoft office features for creating or editing a document.

IBM FileNet eforms

The FileNet eforms provides an alternate option instead of the traditional paper forms. Here the form data is an electronic data which can be handled with ease. With the increased presentation and visualization environment the eforms provides an good alternative for paper forms. Eforms doesn't require any complex coding or scripting.

IBM FileNet Business Process Framework

This is main application interface which is discussed in thesis. The Business Process Framework of the FileNet can be extended to provide wide variety of customized business processes over the FileNet platform. The business process framework provides case based functionality, security and roles management [23].

2.6. FileNet Core Components

The core components of FileNet are tightly integrated to form a platform which is being bundled together as FileNet Platform. The core components are responsible for managing the enterprise content and enterprise process management. The core component also bundles a web client with which the enterprise contents are being accessed.

As discussed earlier the core components are Content Engine (CE) , Process Engine (PE) and Workplace XT.

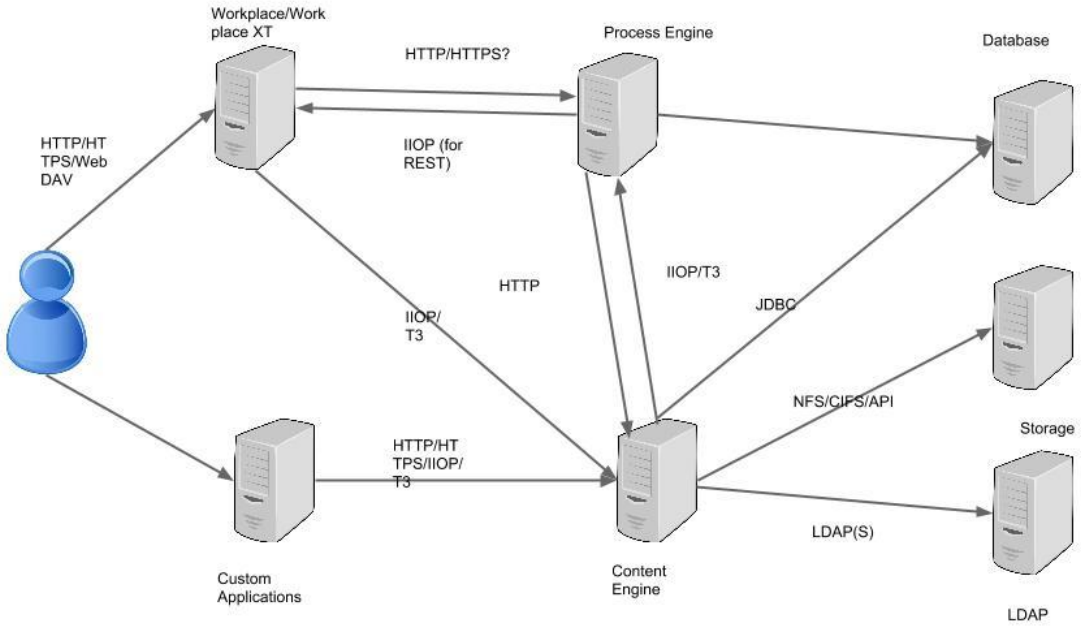

Figure [10].IBM FileNet core components

The above figure depicts the high level architecture view of the key components of FileNet and their interactions.

2.6.1 Content Engine

The content engine is responsible for storage and retrieval of all content within an IBM FileNet system. The content engine makes itself possible for content handling by providing services for creating, retrieving, updating, deleting and securing content. It also provides events for handling actions over the content, document life cycle events, subscription events and integration with storage mediums.

In case of FileNet the content engine runs on top of the Websphere Application Server or Weblogic as an Enterprise application. The content engine is developed to run on J2EE application server. The data model of the contents in the content engine has documents with certain document types, classes and Meta data information. In addition to storage of contents and handling of contents, content engine also provides a rich security model for the contents. This robust security model makes uses of LDAP to authenticate users and to use, view, update and delete a given object in the repository.

The content engine authorizes the users with the help of the Access Control Lists (ACL's) which has some predefined authorization permissions for the object access.

2.6.2 Process Engine

The Process Engine in the IBM FileNet enables the workflow management capabilities. The process engine enables the users to have his own work based inbox, work queues, process tracking, and process configuration console and orchestration mechanisms [20].

The Process Engine is again an Enterprise application running on top of WAS (Websphere Application Server) or Web logic application servers. The process engine uses object stores in the repository to store the workflow related information.

To help facilitate logical separation between lines of business and applications, the Process Engine offers the ability to segregate workflow data into subdivisions of the repository called isolated regions.

Architecture:

The Process Engine has the single process, multi threaded architecture which makes the workflows to enable orchestration. The single process architecture of the Process Engine also enables to have multi-tenancy within a Process Engine Server. This enables the FileNet to have multiple instances of process engine accessing different object stores in a single enterprise system.

The architecture of the Process engine is explained in the upcoming pages. The components of the Process Engine are updated in order to support the latest architecture which involves Content Platform Engine which is a combination of both Content Engine and Process Engine.

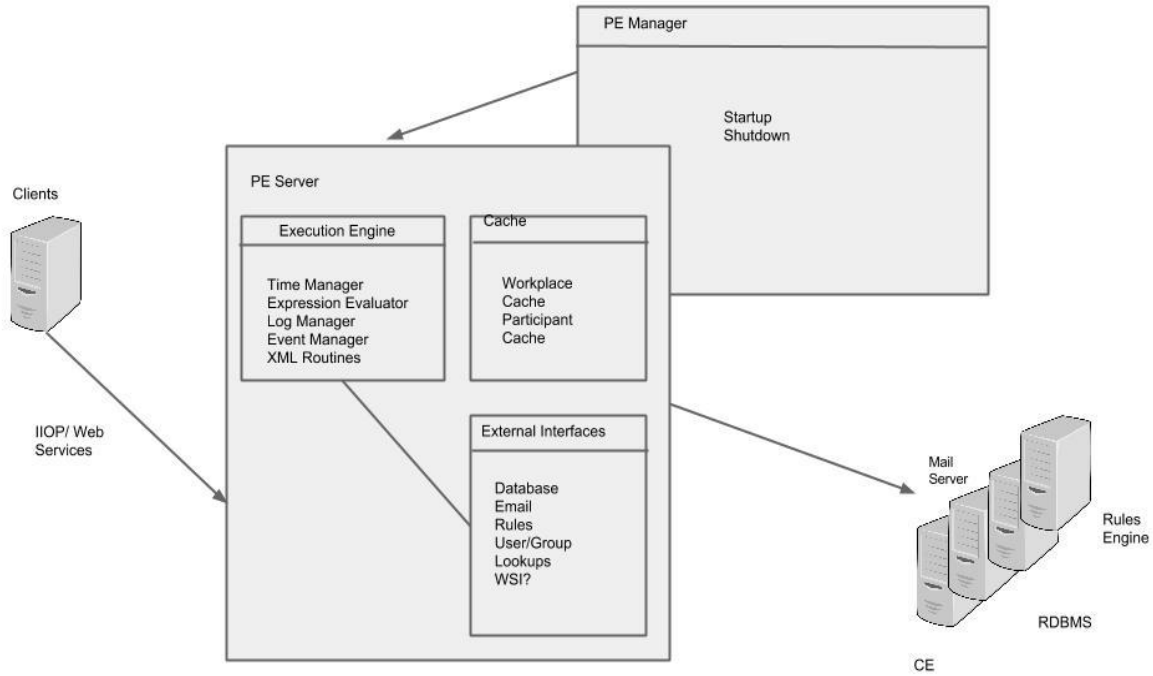

Figure [11]. The Process Engine Architecture

The Process Engine server handles two different types of tasks: PEServer and PEManager. A FileNet with Process Engine server instance is nothing but an instance of PEServer itself. Thus it is responsible for handling database persistence, workflow tasks, orchestration etc. Whereas the PEManager is responsible for starting and shutting down the Process Engine Server instances.

A Process Engine can usually be leveraged with the content engine object store by installing the Process Engine adds over the object stores. Thus the system will install the PE components in the object store which enables us to handle workflows in the object stores. These object stores can be later used to manage the business processes.

3. RELATED WORK

3.1. Evolution of Enterprise Solutions offering with Cloud

The cloud computing is a paradigm that is transferring the IT industry by transforming the way in which the Enterprise services are developed, deployed, sold, delivered and consumed. Now most of the enterprises focus on cloud computing because it minimizes the resources usage. Instead of managing a complex IT infrastructure enterprises can focus on core competence in their organization and obtain those IT functions as cloud services [28]. In the perspective of the Cloud service provider the scalability, down time and reliability matters a lot. Thus now there exists a paradigm shift towards providing the enterprise services with the help of cloud based infrastructure. As quoted in [39] this leads to the evolution of the Common Cloud Management Platform (CCMP). There are many research outcomes in the process of adapting the enterprise solutions to the cloud. But still there are some interesting issues which are blindfolded. This thesis describes one such issue and provides an adaptive solution by considering IBM Smart Cloud Content Management as an example. The related works discussed in this chapter are mainly focused on promoting scalability, reuse and modularity.

With the growing demand of the large scale computing resources , nowadays cloud computing became the buzz talk around the industries by providing different kinds of services ranging from Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) [39]. The key for these services to be exposed to the world are via datacenters, which are built across various parts of the globe by different providers. A cloud provider has plethora of datacenters distributed across various geographical parts of the world.. Each data centers can have thousands of servers and thousands of Virtual machines in it to increase the scalability and performance to the clients.

The elastic behavior of the cloud systems makes it a good player to provide the Enterprise application as a cloud based service. Thus the cloud based service also ensures the increase of resource availability and reduced costs for an Enterprise. Cloud computing offers to business a heavy outsourcing model for computational resources, where service availability, security and quality guarantees all become essential features.

3.2. Data Intensive workloads in a Cloud

The concept of "infinite resource availability" explained by the cloud computing paradigm has let us to explore on large scale data intensive computing. The main focus will be on the enterprise contents in which data intensive computing is more relevant. Considering the fact that the data intensive computing in an enterprise requires infinite resources there must be some management function that regulates application workload on these resources. Thus it's always the major area of focus for workload management in a data intensive cloud is on scheduling and provisioning of resources [12].

The [29] discusses about a framework for providing the data intensive workload management paradigm in cloud. The data intensive computing presents a new challenge for systems management in cloud. The workload management is an important component among the systems management in cloud. The workload management is mainly responsible for scheduling requests and provisioning resources in cloud. Since the load of a data service offered in cloud can vary for time to time it's impossible for an administrator to manually configure or maintain the workload so it's highly essential to automate these workloads on a data service. This eases the management work of a system administrator.

3.2.1 Data Intensive Computing Systems

Most of the data intensive computing system independently performs the task scheduling and data replication tasks. These tasks want the data to be as close as possible. The better option to do this is to have replicas of the data as close as possible to the tasks since moving large data consumes time and affects the performance as well. Thus data intensive systems have their primary focus on the data in the cloud. The data centers built by the cloud service providers mainly provides scalability by providing hundreds to thousands of nodes. Thus it's essential to have good fault tolerance option. It’s recommended not to use the low quality or low cost hardware's to build the data centers.

The data intensive systems use workflows to route the tasks and to manage the scheduling processes on the fly. But these workflows not dynamic they decide the tasks to be executed on the fly but resource allocation are predetermined by the system administrator. Thus this paper [29] suggests that predictive-execution of workload tasks

may reduce the performance and it also might cause bottle necks. The workload execution must have minimal execution time but maximum scalability factor.

3.2.2 Provisioning Systems

In case of cloud based offerings the provisioning is done for the allocation of resources. Most of the provisioning aspects focus mainly on scaling. In most of the cases the scaling is dealt by the user defined rules where in the user specifies the steps to the performed on certain situations e.g. bottle necks or overload. The mechanisms for current provisioning techniques to handle varying workload demand may not scale for large scale data processing. Arm brusht et al pointed out that there is a need to create a storage system that could harness the advantage of elastic resources provided by a cloud while meeting existing storage systems expectation in terms of data consistency, data persistence and performance.

[29] Proposes the following architecture for the workload management framework in a cloud.

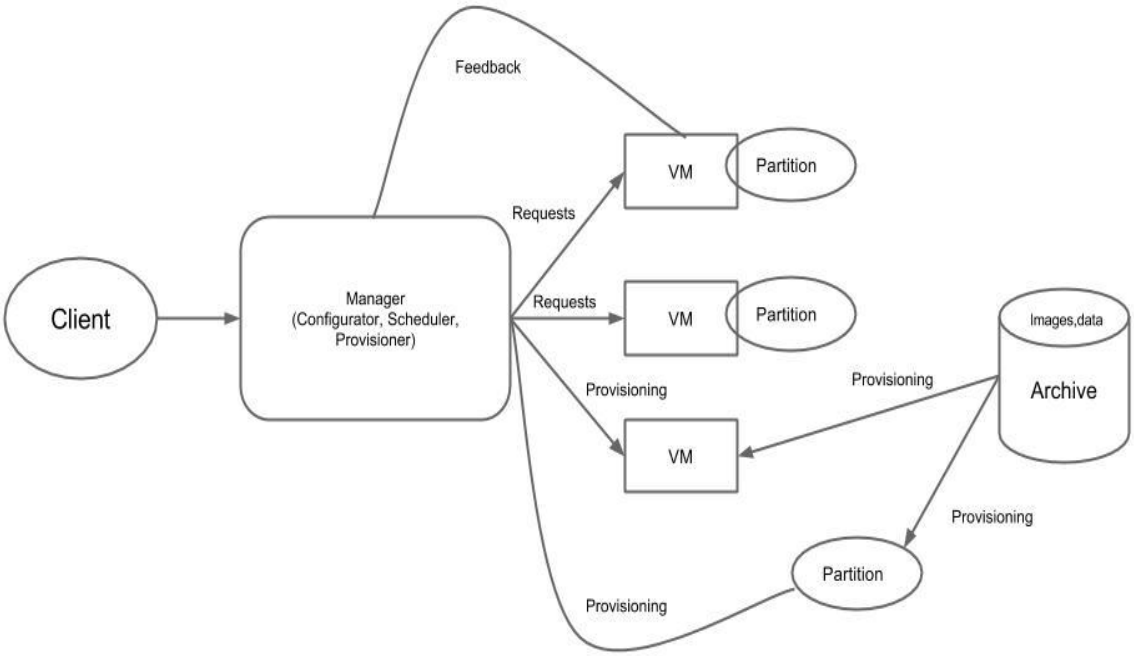

Figure [12]. Data-Intensive workload management in Cloud.

The architecture discussed above has four main components: a client, a manager, storage and execution resources. The execution resources require some settings which can be retrieved from the archive. The client has some application workload that needs to be executed. The manager manages the workload that should be executed.

The client submits the workload to be executed to the manager. The manager has three main components in it: Configuration, Scheduler, Provisioner. The Configurator chooses the suitable configuration settings to execute a workload. This configuration is then sent to the proviosioner which prepares the execution environment (in our case its VM's) based on the configuration settings. Then the scheduler maps the workload to the specific execution platforms which executes the workload. The manager receives periodic feedback from the resource platforms during execution phase.

The provisioning process facilitates the manager in meeting the objective of the workload execution. At the start it provisions storage and execution resources and as the time goes on, it acquires more resources or releases held resources as needed.

3.3. Black box based Workload Management Solution

Due to the dynamic nature of the cloud , changes in the usage pattern , additional resource requests and the dynamic nature of cloud application itself its necessary to identify the possible bottle necks in the enterprise solution offered as cloud solutions and to rectify them. Traditionally most of them use the model based workload management which faces a big challenge due to the high dynamic nature of the cloud applications. For instance [37] introduces way to provision resources for N-tier web applications in the cloud with the assumption that non-CPU resources are adequately provisioned. [37] Proposes a self optimizing workload management solution which adapts to the dynamic nature of the cloud applications. There are various performance management solutions to determine the utilization of the various resources and some performance metrics. The main drawbacks of all this threshold -based technique is that monitoring agents have to be deployed in the computing system, along with a need for a scalable monitoring infrastructure. Thus its essential to enhance the model based workload management solutions to handle dynamic nature of enterprise contents in the cloud.

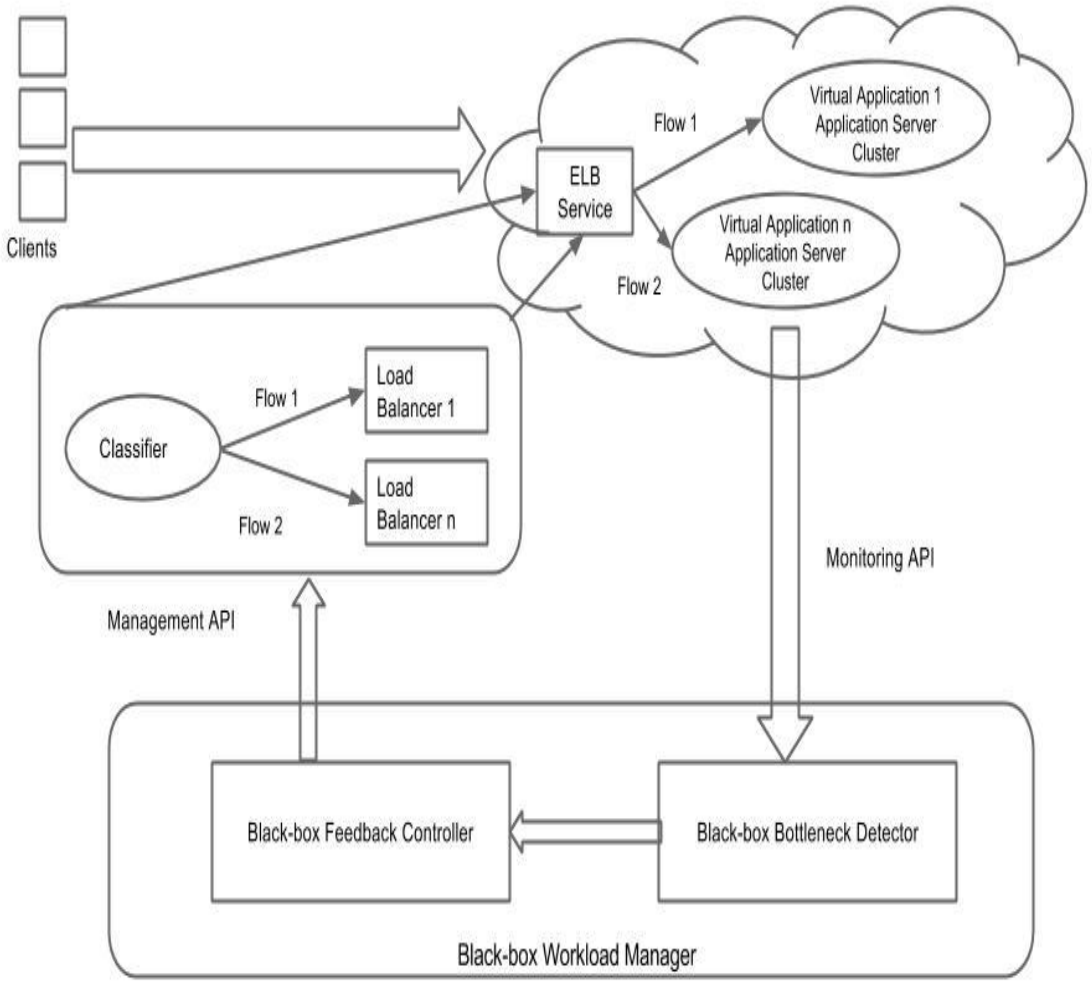

Figure [13]. Black box Workload Management Solution

Each of the Admission Control components has a black box workload manager assigned to it. The workload manager consists of a Black-box feedback controller and a detector. The Admission Control component utilizes the functionalities of a Black-box Workload Manager by using the Management API's. This type of architecture is easy to scale since the controllers works independently of each other. The Black-box feedback controller regulates the load admission rate to the application and the Black-box bottleneck detector captures source saturation at runtime. The workload is optimized or monitored with the help of Monitoring API's from the ELB (Elastic Load Balancing) [37] which is a monitoring service. The monitoring services keeps track of the performance of the workload based applications by keeping getting feedback values such as Throughput from the Application clusters. To address the challenge of bottleneck dynamics for cloud

application management, [37] proposes a target-less bottleneck detector, which treats the system as a black-box and detects the bottleneck without any human defined target thresholds, such as response time target or resource utilization target.

3.4. Automating business process lifecycle management in hybrid clouds

The Business Process Management (BPM) system helps to execute business processes in an organization, which in turn helps the organization to carry out the task with increased efficiency and continuously. But modeling, deploying and executing these kind of tasks requires lot of effort, thus its always essential to have automation of these business processes. In contrary the growing popularity of cloud computing poses challenges on how business processes can leverage resource outsourcing and elasticity. [27] Considers these challenges and provides a framework for automating business processes in hybrid clouds.

Certain research efforts have made it realistic to have isolated solutions for the different life cycle phases of business processes in a cloud. These inter dependence between lifecycle phases requires all the mechanisms to be integrated in an organic manner in order to create consistent and optimized BPM solutions. The prototype described in this paper was the underlying reason for us to use IBM Case Manager for use cases which will be discussed in the future chapters. The platform discussed in this paper has been prototyped and installed as Rackspace instances. It defined some solution templates as the business process model and a solution as a solution template instantiation. The user can define his own templates or solution packages and use it. The search options can be based on some keywords, category, provider, creation date etc. The user can create his own templates with business process in it. The tool used for generating these templates will have a web based editor and uses the dragging and dropping new services, deleting or modifying existing services in the middle canvas [27].

After creating the templates the user can create a data flow and assign roles for the steps in the business process. This can then be considered as solution and it can be deployed and tested in a cloud environment. A user can view and operate on the tasks assigned to him/her defined in the solutions by going to his personalized task page. The user has also the ability to delegate or pass his work to another user based on the user availability. The other user has the option to take tasks from other users as well. This makes the users to cooperate with other users and share the tasks within them.

3.4.1 Business Process design tool

The framework also poses a platform that assists the business users to design the business processes. Users can submit high level descriptions of business processes in the form of diagrams which represent the orchestration of services. This platform just provides the high level diagrammatic description of the business processes with some abstract services and control flows [27]. The business user can define roles to the steps in the workflow. But it lacks the description of data flows and data objects that are passed between the services. The platform then helps to specify data flow between the services, but also detects the missing edges and services which are not defined or available, providing an opportunity for users to validate the workflow after design phase.

Preconditions

The platform consists of service registration mechanisms and data type hierarchy allowing the classification of services into service types and their instances. The general reason for having service registration is to enable interoperability between the services and setup a common agreement on service semantics and usage. The registration process also enables business process reusability by allowing the manipulation of abstract service types instead of service instances in the business process design.

Data flow binding suggestions

A business process is often designed as an end-to-end transformation of data objects. Thus, identifying the transformations that each service performs on a data object as it moves through the business process will give an indication of the parameters that should be bound together.

Users will likely apply the same combinations of services in different business processes so that keeping track of the ways the services are bound in the past will be useful to guide future compositions. Thus some abstract services can be later defined while designing workflows for the document centric tasks.

3.4.2 Solution Deployment

The solution which was developed in the solution platform which was proposed has the workflows developed using the process design tool. This solution should be deployed in

the cloud to make it highly available. There are certain mechanisms to optimize the path in which deployment takes place [27].

3.5. A Content Management System for Cloud Computing Environment

The amount of digital contents to be stored and retrieved is increasing day by day at an accelerating speed. These contents are mostly in heterogeneous systems and handling these type of contents manually is a big problem. Thus it becomes necessary to have a content management system with good performance in a Cloud computing environment. [2] introduces a content management system that provides a uniform way to manage the content stored in all of the devices with the same user friendly and efficient principles. The system which was proposed was designed and implemented with the characteristics of the distributed and heterogeneous environment in mind.

The VisualRest system is designed to provide a uniform way for users to manage the contents stored in different devices with same user friendly and performance related principles. This section deals with the precise overview of VisualRest and the components in it [2].

3.5.1 REST Architecture

REST is architecture for web services for building an network-based software. In case of REST based architecture the resources are exposed by the servers and are consumed by the clients using HTTP methods. Typical HTTP methods which are used are GET, PUT, POST etc. An important feature of REST based services is the clear division of application state between the client and the server. The notable feature in VisualRest based implementation is the usage of push based REST model , which enables increased scalability [2].

Eventually, three main properties of REST such as the layered use of client-server style, the uniform interface and the possibility to negotiate a suitable representation for a resource. The first important property in to have caches and gateways as intermediaries. The REST based architecture allows us to enable intermediaries. These intermediaries can be caches and gateways as discussed above. The second property is to enable scalability, REST based architectural style permits a wide range of clients to utilize the system. The third and the final property is the negotiable representations allow humans and programmed clients to simultaneously use the system.

3.5.2 XMPP

The Extensible Messaging and Presence Protocol (XMPP) is a protocol designed for near real time and request response services streaming XML messages [2]. The services implemented using XMPP are mainly used for instant messaging, system control and lightweight middleware. The XMPP is implemented based on the client - server architecture wherein the client connects to the provider and this communication will persist until the request service is completely offered. Thus this kind of services is highly useful when we want to send notifications of a changed state or monitor. This is one of the underlying concept of the In basket workflow notification which will be discussed in the upcoming chapters.

3.5.3 VisualREST

This section gives a short overview about the VisualREST architecture discussed in [2]. It's an architecture with converges with the content management system that facilitates users to access the contents distributed across heterogeneous platforms. The users are facilitated to register the devices which they use with the VisualREST servers to access their contents via a single entry point. This entry point can be a VisualREST server. The VisualREST server provides a web interface to access the contents and to view the contents. This kind of architecture is simply based on the client-server model. The communications between the different components are done essentially through the HTTP protocol. The VisualREST server keeps track of the contents in the repository for the newer version once the newer version of the content is available its notified to the users via the XMPP protocol. New versions are automatically created when content is updated thus the contents in the VisualRest is never lost.

The VisualREST server acts as a middleman between different clients to store the meta data of contents from all the clients in a single repository. Thus the content management is done based on the predefined roles for the user. Clients of the VisualREST can have three different roles namely 1) The user who uses the service with a browser 2) the container program which provides contents to the users and 3) A loosely coupled program which uses REST based interface.

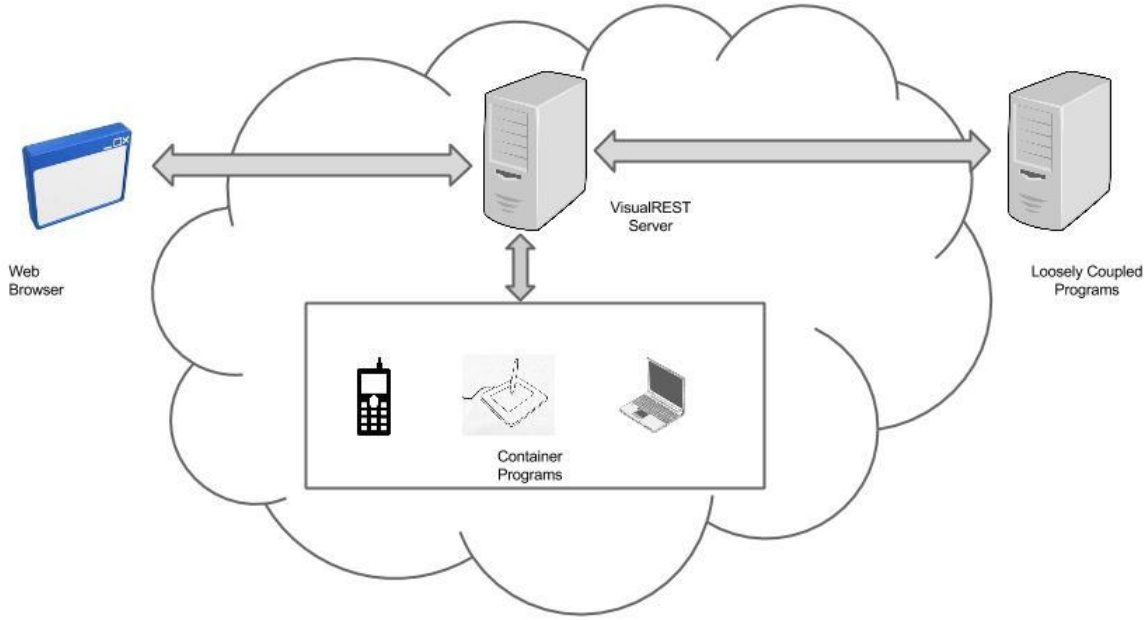

Figure [14]. VisualREST Architecture

The above figure describes the architecture of VisualREST implementation. The content programs generate contents and it can be any devices connected to any heterogeneous system. The client can access and modify contents using a web browser. The loosely coupled programs are used to manage the resources by using CRUD operations and visualizing the representation of the content from VisualREST.

All the above works discussed in this chapter forms the basis for the thesis to solve the problems which many developers face in managing the workloads in a Enterprise Content Management System.

4. CHALLENGES IN WORKLOAD MANAGEMENT IN ECM

Before discussing about the possible challenges in workload management in an Enterprise Content Management system, it's essential to know about the differences between the traditional Workflow and BPM.

4.1. Workflow and BPM: What's the Difference?

When considering the case management applications, the differences between workflow and business process management (BPM) becomes highly important. Even some of the advanced workflow capabilities are sufficient for an ECM system operation.

Workflow is standard functionality included as a process management feature in the many Enterprise Content Management systems. In contrary a BPM suite (BPMS) of tools includes the workflow and should be used when additional functionality is required, as it is within most case management applications [18].

4.2. Process Management in ECM

An unstructured content often referred as the electronic documents has a specific document life cycle within an enterprise. In the enterprise terms a life cycle is commonly the period where the content in the document is highly useful. These kind of electronic documents often flow between the various workers of a business, and many workers often need to reference the documents to accomplish their work. A workflow can then be defined as the steps involved in routing a piece of content through its life cycle.

Most of the Enterprise Content Management systems provide a Process Engine and a set of associated components and applications that enable easy implementation of robust business process management solutions designed for use by business users. This process engine also provides a set of API's to enable custom programming for automating processes and a flexible environment for process improvements. The process management feature enables the user to model and design the document centric workflows.

4.3. BPM feature in ECM

The BPM is a foundation for orchestrating business processes as well as a business discipline for continuously optimizing the way a business operates. To satisfy comprehensively the BPM requirement, a suite of integrated business process tools is required. A BPMS enables the understanding, automation and optimization of business processes. The routing of documents is just one element of BPM.

The major objective of workflow is to replace the physical movement of documents with the electronic movement of document images in a predefined sequence. The objective of BPM is overall business process management and optimization using process-centric technology.BPM starts with quantitive analysis of existing business processes-referred to as Business Process Analysis (BPA). To perform BPA, your first need to establish a baseline of characteristics for the existing processes. The baseline is established by discovering the details of the current state of the business processes.

With the current state discovered, the next step is to develop an improved future state process design. By understanding the properties of the current state process along with capturing any issues or opportunities for improvement, the business analyst gains an understanding of what changes are required to optimize the process design. When BPA tools are part of the BPMS, the proposed process design can be imported directly into the process builder or modeling interface, where the business logic is configured.

BPM provides more than electronic document movement: It is a technology based approach for understanding, automating, and optimizing business processes. Complex business processes are a mixture of human and automated activities [25].

When a user performs manual operations by interacting with an BPM application, the interface must be flexible, easy to use and intuitive in a ways it should be empowered for an user to be used to make decisions or judgments and also it should be in a way a user can interact with the application either directly or using another application such as an email client.

The main issue to deal with regardless of where it's an workflow or BPM, is to reduce the overall complexity of designing the workflows and also to propose a framework with integrated platform which is flexible enough to orchestrate any type of process, open

enough to integrate with other applications easily, and robust enough to start small and scale to handle large volumes of information inside a repository.

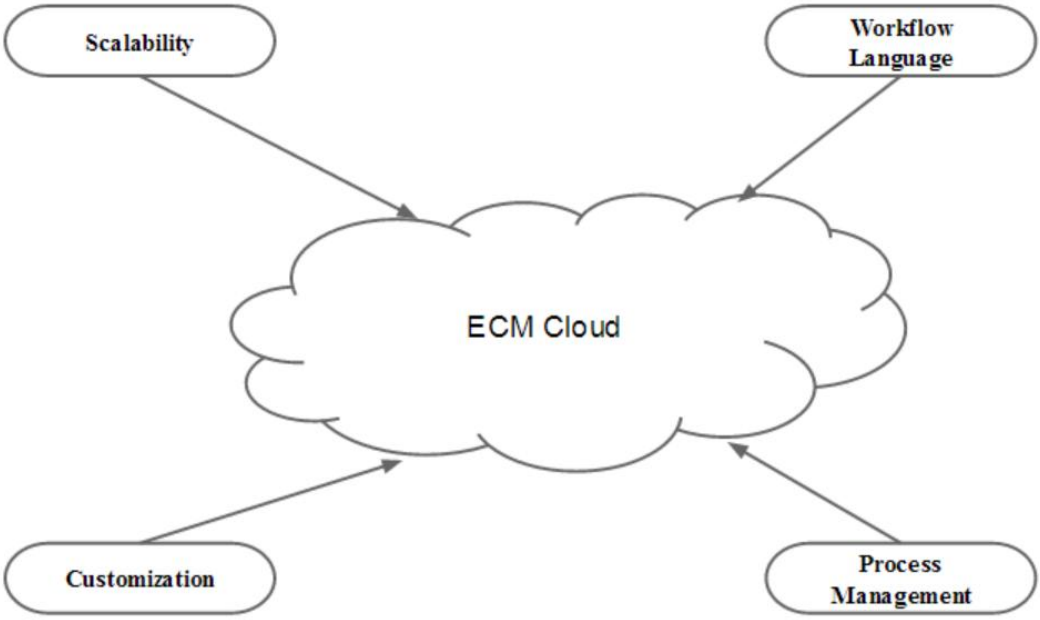

Figure [15]. Challenges in ECM Cloud

4.4. Scalability in the Cloud

This section discusses about the scalability notion in Cloud and also possible challenges in scaling Enterprise Content Management applications in Cloud. Scalability is a desirable property of a system, which indicates its ability to either handle growing amounts of work in a graceful manner or its ability to improve throughput when additional resources are added. A system in which the performance increases or improves after adding hardware, proportionally to the capacity added, is said to be a scalable system. Thus the document based Content management systems has the problem in scaling the applications to the end users. The possible challenge is to provide the end users the content which is supposed to be given to the particular user. It's always a performance hurdle to provide the contents which are not useful for that particular end user.

Similar in principle to the management of an application system under varying amount of work load to ensure that the performance of the system is not affected by surges in demand for the application, organizations should be able to increase or decrease computing power without adverse delay [19].

The content demands varies dynamically by certain time periods like days, days of the week, months of the year etc and these variations affect the performance of the organization while not meeting these demands. Thus there is a necessity to have some workflows which can be used to provision the resources. A provisioning of resources is the concept of allocating the required resources. The ability in terms of capacity to manage these variations results in three different states namely: under provisioning of resources, on-provisioning and over-provisioning of resources.

An under provision occurs when there is a deficit for resources. In this case the resources are not sufficient to cater for surges in demand. On- provision occurs when the resources are just enough to satisfy the increased demand, while over provision is a state where the resources level is available or in use are more than necessary to meet the increased demand. The over-provisioning mechanism is outside the scope of this thesis. From these perspectives it's clear that when the demand increases, resources should be increased to meet the demand and as the demand decreases, use of resources should be decreased to have just the right of amount of resources to meet the demand [35].

4.5.The complexity of Workflow Language

In order to utilize the workflow technology to achieve some tasks its necessary to understand the workflow language which conform to the grammar of some workflow description language to workflow engine. At present there is plethora of workflow languages like BPEL (Business Process Execution Language), MoML (Modeling Markup Language) to name a few. These kind of languages are generally complex so some users find it difficult to program in this language. In general the developers find it difficult to design document centric workflows using BPEL just because of the complexity of the workflow language involved [35].

4.6. Problem in Customization

Most often the user specific enterprise applications should be highly customized. The customization can be done based on the user roles. The increased level of customization

can help the enterprise to safeguard the information based on the roles. The customization of an application generally refers to the provides the services which the user is entitled to have. This kind of user based application is still under the perspective of research. In this thesis a framework is proposed to have customized applications based on roles.

4.7. Need of a Paradigm shift

Generally the cloud services and resources are provided based on the pay as much as you strategy. The cloud can be an Infrastructure as a service where in the infrastructure can be delivered via virtual machines. These virtual machines can have some application servers which runs the services which can be provided as the cloud services. Thus its highly necessary to leverage the cloud applications as a service to the users with ease. Thus there is a necessity to have a paradigm shift wherein the user can be provided with customized application and he pays for the application which he uses and the service related to that application [35].

5. IBM CASE MANAGER INTEGRATION POINTS

5.1. Overview

The case management provides a flexible way of performing business applications for the knowledge workers. The knowledge workers can develop the tasks in much flexible way and those tasks can be carried out with ease [20]. The traditional ECM solutions combined with business processes are not sufficient to meet requirements of certain applications. Most often if it's a document centric workflows then it's way too complex to design business processes. In industries such as health care, insurance and banking sectors case management is well understood and used. However the traditional case management approach can be extended to support some document centric tasks which can be considered as case, when discussing about the case management perspective.

The underlying methodology of the case manager is around the concept of processing a business task which can be called as a case. A case can be a collection of business related activities such as assigning roles, distributing tasks between the users and also performing some document centric operations. Each case has a case folder which can be solution folder where in the document related to the solution and the document centric workflows are stored. This case folder also keeps track of the work done and deadlines to be met [20].

From the above discussion it's clear that the Case management solution is a management technology to solve a business problem which is closely related to the records management. Thus it's clear that case management solutions can be a good integration for performing some IBM SCCM tasks.

The case solutions are knowledge intensive and highly flexible. The workflows in the solutions can be document centric and can coordinate with other BPM processes to perform tasks. These workflows are capable to exposing the document centric services which other BPM tasks can use. Most of the document centric solutions require higher about of flexibility which makes these kinds of solutions hard to implement using the traditional BPM processes. Since the BPM processes should have easily predictable and repeatable tasks its complex to implement some document centric tasks which are spontaneous and flexible. Since the user the decides of when the workflows should execute these kind of solutions are less predictive by the system which makes it flexible

for users to use it. The participants who interact with the document centric workflows are organized by roles. In order to implement a robust and effective document centric solution its necessary to have a software platform that consists of range of functionalities. These functionalities include content management, process management, business rules, collaboration and analytics [20]. The user interface (IBM Content Navigator) in our case should be sol flexible and should allow greater collaboration with the FileNet platform [21]. For instance the users should be able to execute certain tasks from his personal workbasket itself.

5.2. Case Management Solution versus BPM Solutions

The case management solutions are often document centric solutions. It's very important to realize the intent of the solution to be developed. The scenarios in which the solutions is to be used is also equally important. Most of the Case management solutions have the primary focus on the data around the case and what is needed to complete it, but in case of a BPM solution is the primary focus is on a business process and to determine the ways to complete it. In case of a document centric case management solution the processes are short and it will be used for some specific functionality like fetching some information, invoking a service, checking in and checking out of a document from the file system and processing business process rules etc.

Case Management Solutions	BPM Solutions
The workflow path depends more on the user's decision instead of the workflow system itself	The workflow path depends more on the system control or business rule based control flow
The main goal is to design a flexible process to drive a business goal	The main goal is to design business process which is repeatable step by step to solve a business problem
No need to have modeled processes to solve the business processes. The user can design his own business processes	Most of the BPM solutions have the business problems solved using some modeled processes.
The main focus of the workflow will be in the data and document meta data	The main focus of the workflow will be exclusively on processes and how to achieve those processes successfully
The workflow can have indeterministic life	The workflow should have deterministic

cycle which is the more flexible way of designing a solution for a business problem	life cycle and each process should be short lived
The case management solutions can be content based , because the users decides the next process to be executed based on the contents	It's not necessary to have content based decisions for a business process in BPM

Table [1]. Case Management Solutions versus BPM Solutions

From the above table it's clear that the document centric workflow solutions can be implemented in the similar way as an Case management solutions. Thus the business goal of the case management solutions are always achieved but the path to accomplish it is non deterministic.

5.3. Modeling

Process modeling and Case modeling are different. A model for a solution is the definition of that solution. The design tool in the case management is used to model a solution. In case of BPM process modeling is well understood and formalized by standards like the Business Process Modeling Notation (BPMN) [23]. In case of document centric workflow modeling it's important to understand the document characteristics and the workflow processes to be used. For instance for automatic classification of documents its necessary to understand the properties of the document.

5.4. IBM Case Manager Environments

The case manager components are built on top of the IBM FileNet platform. This enables the case manager to use the content engine and platform engine with ease. In order to have fully fledged development and production ready solutions its necessary to know about the different environments in a case manager system [21].

5.4.1 Development Environment

Out of three phases in case management the first is the development environment in which the case solutions with workflows are engineered. The development environment consists of design object store, target object store , an isolated region and a connection point which enables the case manager to connect to the object store. The connection

point can be considered as a connection definition for communicating with the target object stores.

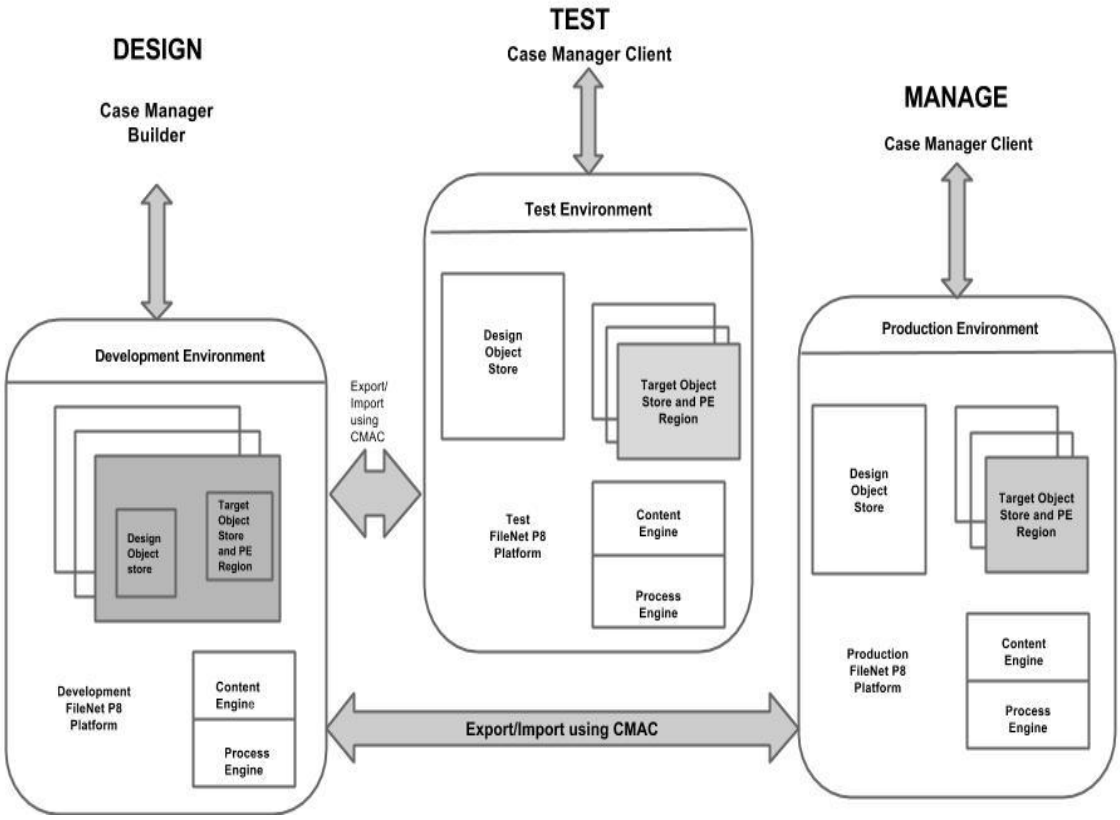

Figure [16]. Case Manager Phases and components

In the design environment it's possible to create business solutions with specific business rules. It's also possible to define workflows for the cases which can be done using process designer component in WorkplaceXT. The notable solution artifacts of a case management solution are properties, rules, roles, case types , pages etc.

The development environment allows for distinct project areas, each with their own target object store and Process Engine region. There can be multiple project areas and each project area can have many solutions assigned to many users.

5.4.2 Production Environment

The production environment uses content engine and process engine components. It's not necessary to have case management builder and WorkplaceXT in it. The production environment uses process engine components to perform workflow steps. The case management solutions are deployed, run and managed in a production environment. The

solution and its assets are exported from the development environment and imported into the production environment using the Case Manager Administration Client tool. The solutions which are imported into the production environment are deployed into the target object store and process engine isolated region for proper execution.

5.4.3 Test Environment

The development environment can be considered as a test environment by itself. Sometimes it's necessary to have a separate test environment to test the test cases which must not be influenced by development environment. Thus a test environment is used to perform the user - acceptance tests, performance tests and pre-production steps. The test environment uses both design and target object stores, process engine extensions to execute the test cases. As similar to the production environment the test environment.

5.5. IBM Case Manager Components

The Case Manager has been bundled with several other components to enable the simplified way to create case manager solutions on top of FileNet platform. The table below gives the overview of building blocks of a case manager.

Abbreviation	Component
CMAC	IBM Case Manager Administration Client
FDM	FileNet Deployment Manager
FEM	FileNet Enterprise Manager
CE	Content Engine
PE	Process Engine
BSpace	Business Space

Table [2]. Building blocks of Case Manager

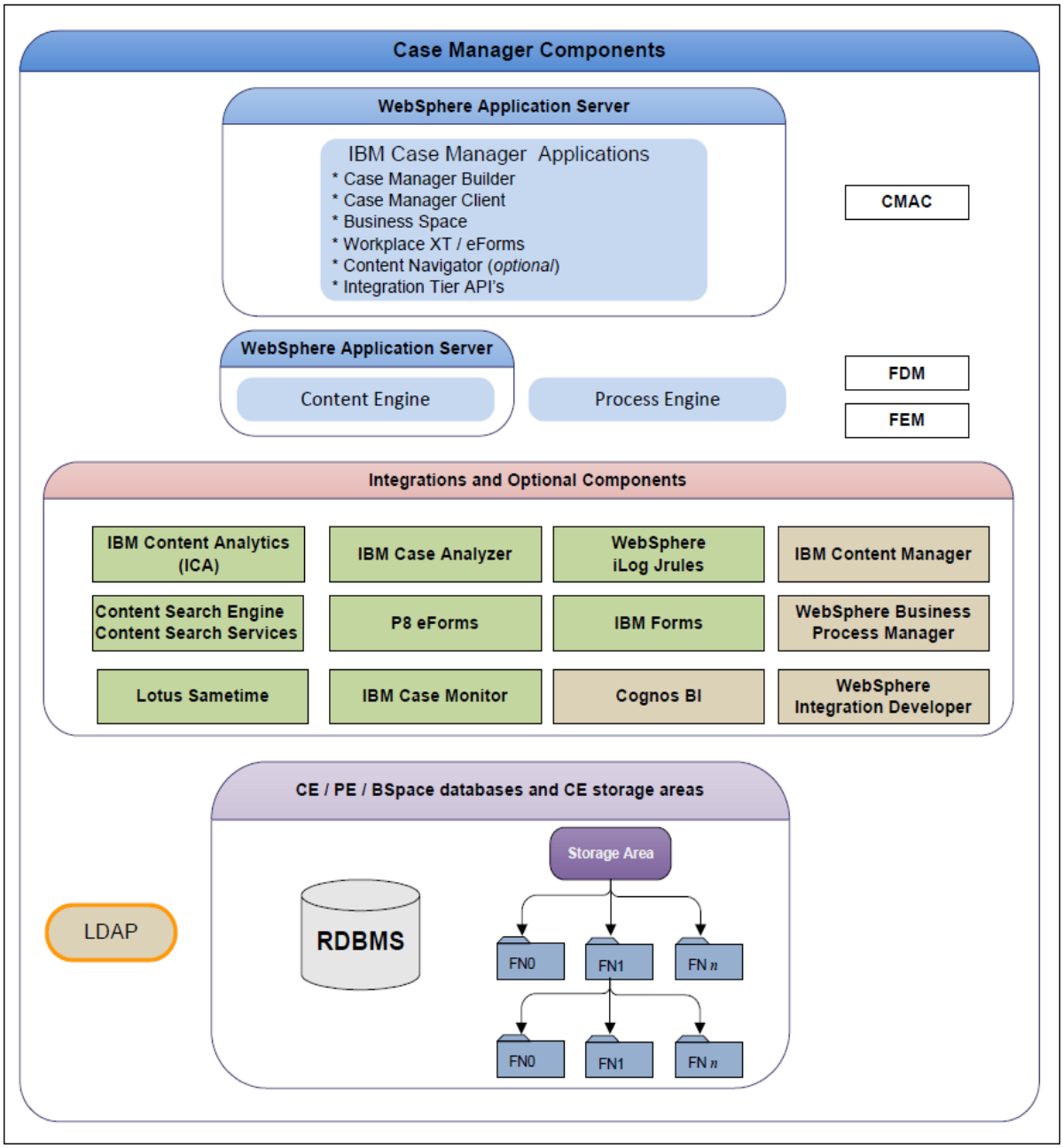

Figure [17]. Case Manager Components [21].

The case manager components are generally divided into two different categories i.e Core components and IBM Case manager extension components. The core components are used to developing the case solutions and the extensions provides additional support to the case solutions.

5.5.1 Core Components

The above figure describes some of the core components of a Case Manager. These core components are responsible for providing core functionality to the case solutions. The

most important of those are the content engine and process engine which will provide an extension point for the SCCM solution integration in the upcoming use cases. Some of the core components are listed below.

- Case Manager Builder
- Case Manager Client
- Case Manager API (Provides REST API and FileNet Integration API)
- CMAC
- Content Engine
- Process Engine(Responsible for designing document centric workflows)
- WorkplaceXT (Provides process designer component to design document centric workflows)

Case Manager Builder

This development environment component is responsible for designing the case solutions for the business level users. The users can create their workflows and business rules by themselves with ease. The solutions can then be deployed into the target object store and tested with ease with the help of case manager desktop on top on ICN (IBM Content Navigator).

Case Manager Client

The case manager client is used to test the case solutions. The SCCM system must have content navigator to be replaced as case manager client which makes it easy for the SCCM users to use the case manager solutions on the fly. The case manager client uses Case Manager Representational State Transfer (REST), Process Engine REST and Content Management Interoperability Services (CMIS). The REST API helps the case manager components to communicate with the underlying FileNet Platform [21],[23].

Case Manager API

The Case Manager API is considered as an main integration entry point for SCCM system with Case manager components. The aim of the thesis not to provide an infrastructure for SCCM system with Case manager installed in it. The sole purpose is to efficiently utilize the core components of case manager in the SCCM system. Thus case manager API library is a good starting point for it.

The Case Manager provides two types of API's one for the REST calls to the FileNet platform and other is the standard java API library of the Case manager. The Case REST API has REST based APIs and CMIS for Process Engine. The REST API in the case manager follows the REST based architectural style and uses the Java APIs from the process engine and content engine.

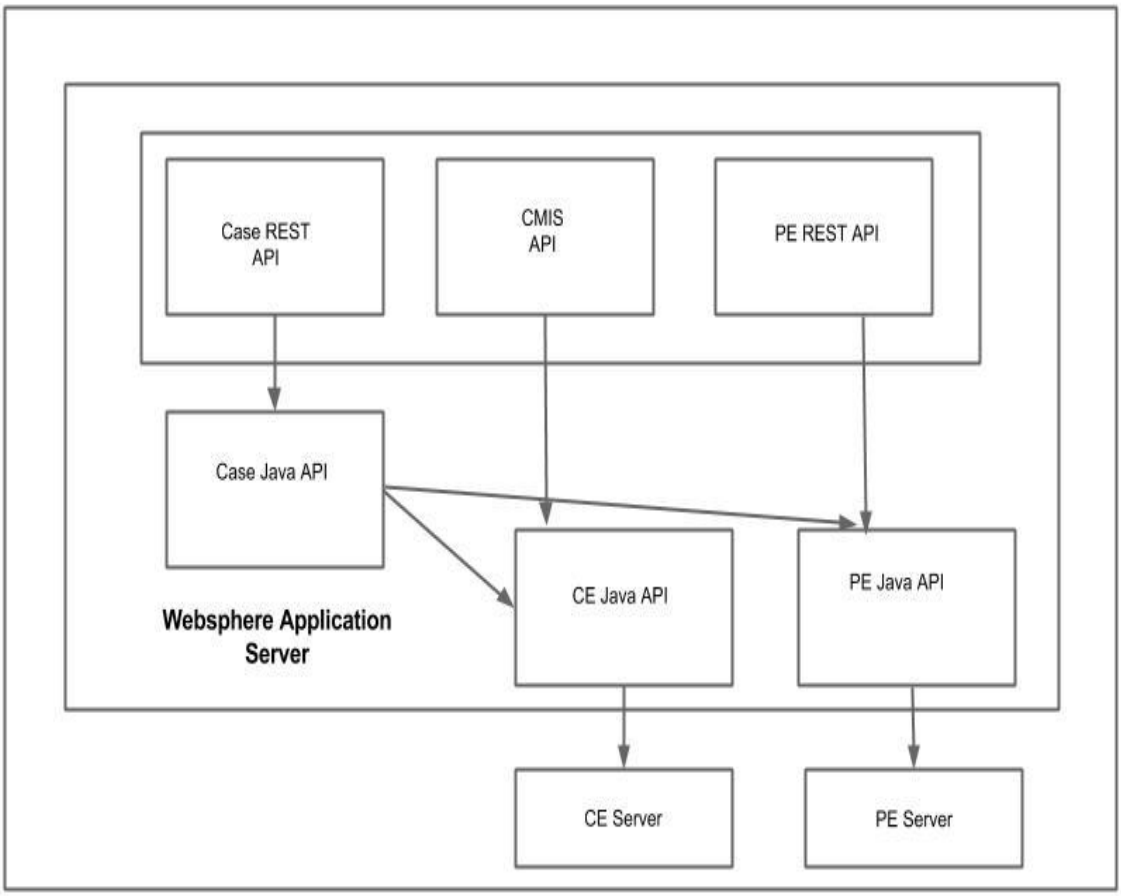

Figure [18]. Case Manager API architecture

Content Engine

The content engine supports the case manager operations by providing object stores to store the case manager solutions. The content engine has the design and target object store with which the case solutions are designed, deployed and tested. The design object store has an solution page which has the workflow definitions, solution definitions and case pages which can be later deployed into the target object store for usage.

Process Engine

The process engine is the heart of the case manager execution environment. It supports case manager mainly in two aspects.

Firstly, in a case solution it's necessary to have some business rules and role based process in it. Thus the process engine helps the users to design workflows with rules and to assign processes based on certain roles. It also supports authoring of workflows in case builder environment. The process engine provides a good interface for integrating the existing BPM workflows with the case manager workflows.

Secondly, the Process Engine is responsible to invoke the case items to be executed in a case when the case solution is executed via the SCCM desktop client. This involves provisioning of resources and updating the in baskets of the users with the next consecutive steps or work items [22].

5.5.2 Case Manager Extension Components

Some extension points in the case manager are IBM Business Process Manager (IBM BPM) and IBM Content Manager (CM8) [21].

IBM Business Process Manager (IBM BPM)

As discussed in the previous sections the case manager provides facilities to integrate with the external BPM processes. This can be done when the case manager is bundled with IBM BPM component. This enables the developers to use BPM processes to case solutions in the case manager builder environment.

The IBM BPM server has the BPM processes running in an external environment. This BPM processes are mapped with the case solution steps using case builder. During the deployment phase of the solution creation, it checks for the process applications , snapshot and process mapping with the BPM server settings. The content engine also provides the event handler with which the BPM processes in the IBM BPM server are invoked.

IBM Content Manager

In our scenario it requires the document centric workflows to have access to the contents in the repositories. The integration of case manager components with content manager

helps us to access and use the content in the repositories. This integration is not specific to a specific design and object store repositories it enables the workflows to use the contents in any repositories in which the case solutions are deployed seamlessly. The documents created by the case solutions are also handled by the content manager to check in it into the respective repositories.

5.6. Environment Configurations

As discussed earlier the case manager can have a development environment, test environment or production environment. The configuration of each environment differs based on the components available for those environments.

5.6.1 Development environment profiles

The case manager development environment normally has single instances of each component, thus it does not require any load balancing mechanisms or clustering. If the development environment is a cluster based environment then its recommended to have IBM Content navigator and IBM Case Manger in the same cluster. Normally a development environment profile consists of, a Websphere Application Server profile in which the Case Manager is to be deployed, a FileNet Content Engine Websphere Application, a FileNet Process Engine , an RDBMS system and an LDAP system.

5.6.2 Production environment profiles

The production environment is mainly for deploying the solutions and running it so it's not necessary to have the development environment components in it. Based on the requirement the production environment can have one or more target object store. There can be one or more solutions in a single object store as well. The figures below will describe various approaches of deploying a solution.

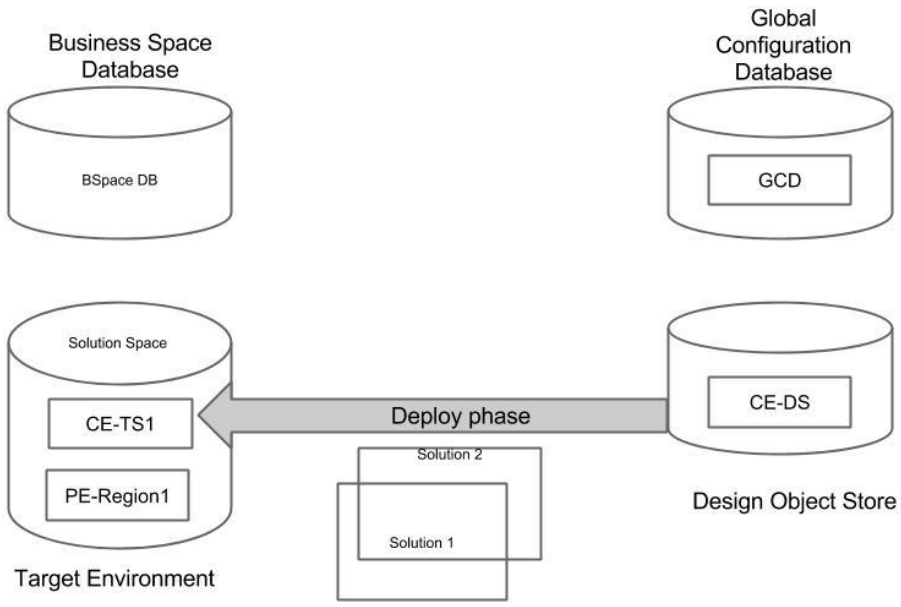

Figure [19]. One target object store hosting two case solutions

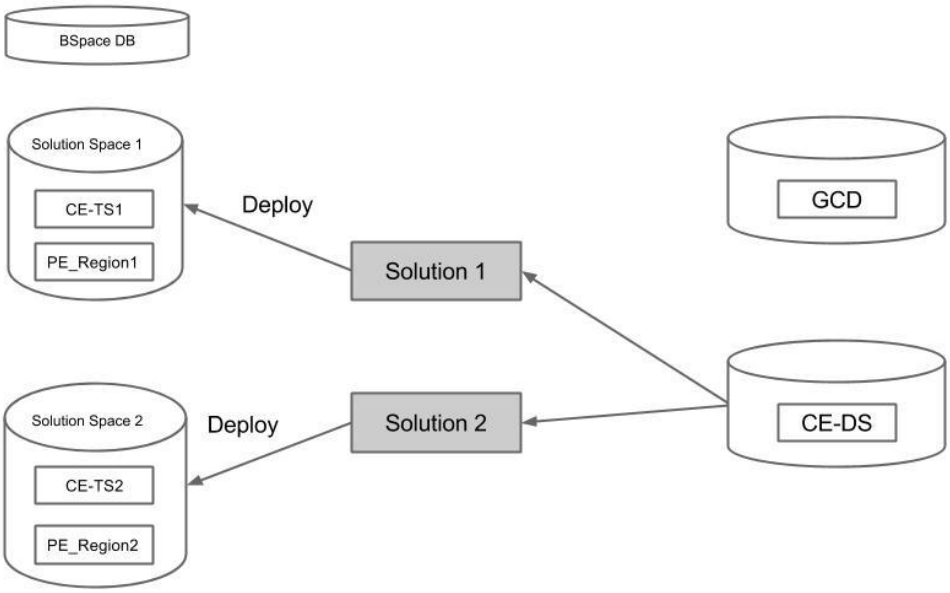

Figure [20]. Multiple Object Stores with One solution each

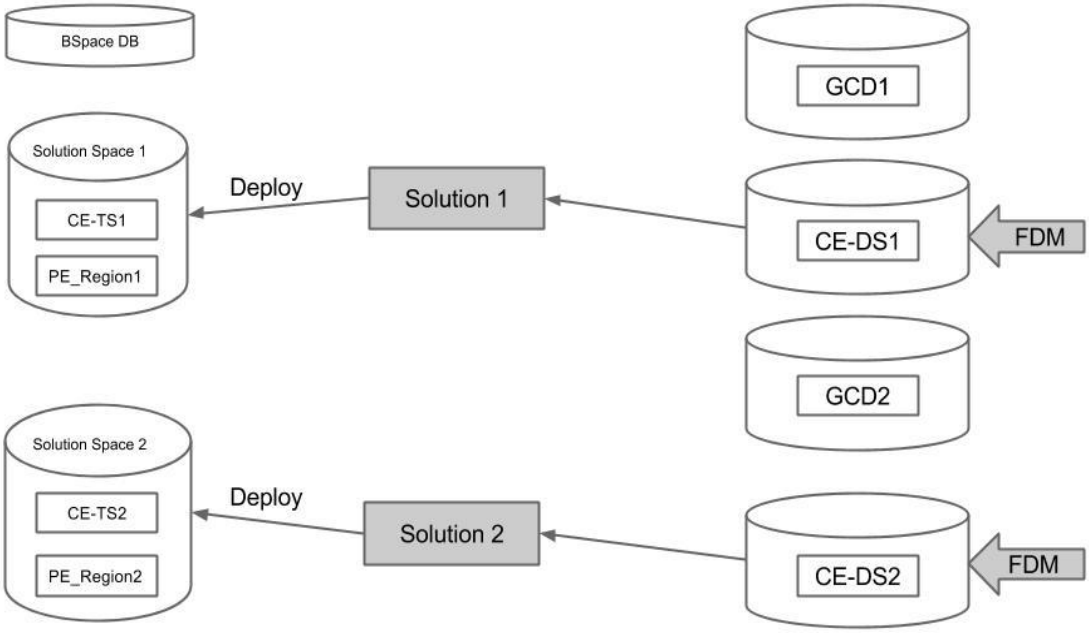

Figure [21]. Dedicated FileNet domain for each solution

Thus production environment provides higher availability to the solutions to be deployed into the SCCM cloud, which enables flexibility to the business users as well.

6. USE CASES AND IMPLEMENTATION

6.1. Introduction

The challenges which were discussed in Chapter 4 and the integration points discussed in Chapter 5 gives an idea how to approach these challenges with a use case. For the use cases the IBM SmartCloud Content Management System which was introduced in the initial chapters is considered. The IBM SCCM system provides cloud services for the ILG solution stack. It has some important services like Batch Load, Disposal Sweep and LDAP Synchronization services. The traditional IBM SCCM faces scalability issues when many users try to access the IBM SCCM administration portal. Thus it becomes necessary to integrate the services provided by SCCM system into the IBM Content Navigator (ICN) thus the end user can access it from his ICN portal itself. As discussed in the chapter 4 is also necessary to provide high level of customization in the solutions. Thus the plan is to import the solution package into the SCCM system and use it as an add on.

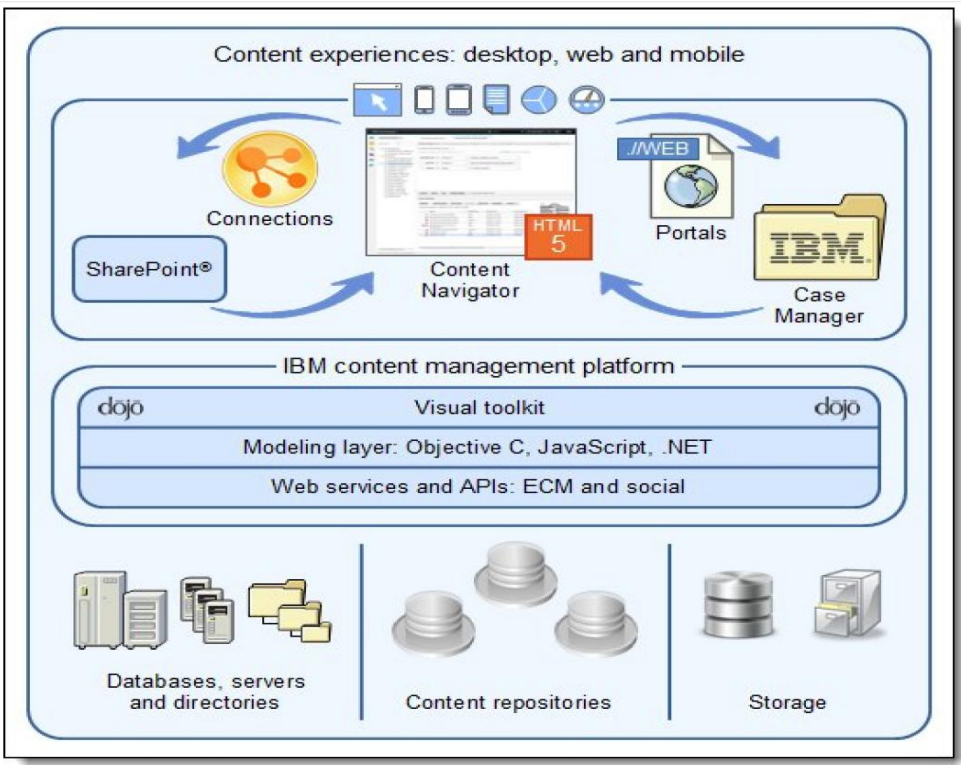

Figure [22]. ICN Integration with Case Manager

As the figure implies ICN provides a good integration point for IBM Case Manager Solutions. Thus the idea is to make the SCCM cloud services as solution packages in

IBM Case Manager and import it into the SCCM system as a solution package. Thus the solution package can be an add-on for the ICN and it eliminates the possibility of having separate administrative user interface for the SCCM system.

6.2. IBM SCCM Services

Chapter 1 gave a short overview of the services offered by the IBM SCCM System. The services ranges from Batch load tasks and Disposal Sweep tasks.

6.2.1 Batch Load Use-Cases with Workflows

The figure below explains the use cases to be handled by the user who has access to the system. The existing SCCM system doesn't have the possibility of approving the report and the system itself is a static code running to perform the batch load task [14]. In our case the Process Designer is used to create the workflows which can run on the cloud. The workflow is designed in such as way that it supports scalable application access. There are two kinds of users 1) End user and 2) Administrator who can be considered as the approver as well.

The scenario is that once the batch load task is complete the approver and the end user is notified about it. The approver will get a notification about the approval process to be done in this personal in-basket.

Roles in-basket

The roles in-basket is a bucket that contains the work items that the users or groups that belong to the role can access. The role in-basket is embedded on top of the IBM Content Navigator and each role has a separate in-basket. The process of creating in-baskets for each role is taken care by the IBM Case Builder.

For each role and associated in-basket there will some solutions which can be defined. In our scenario these solutions can be the IBM SCCM services. The Case Manager Builder defines a Process Engine role, an in-basket and a work queue for each role.

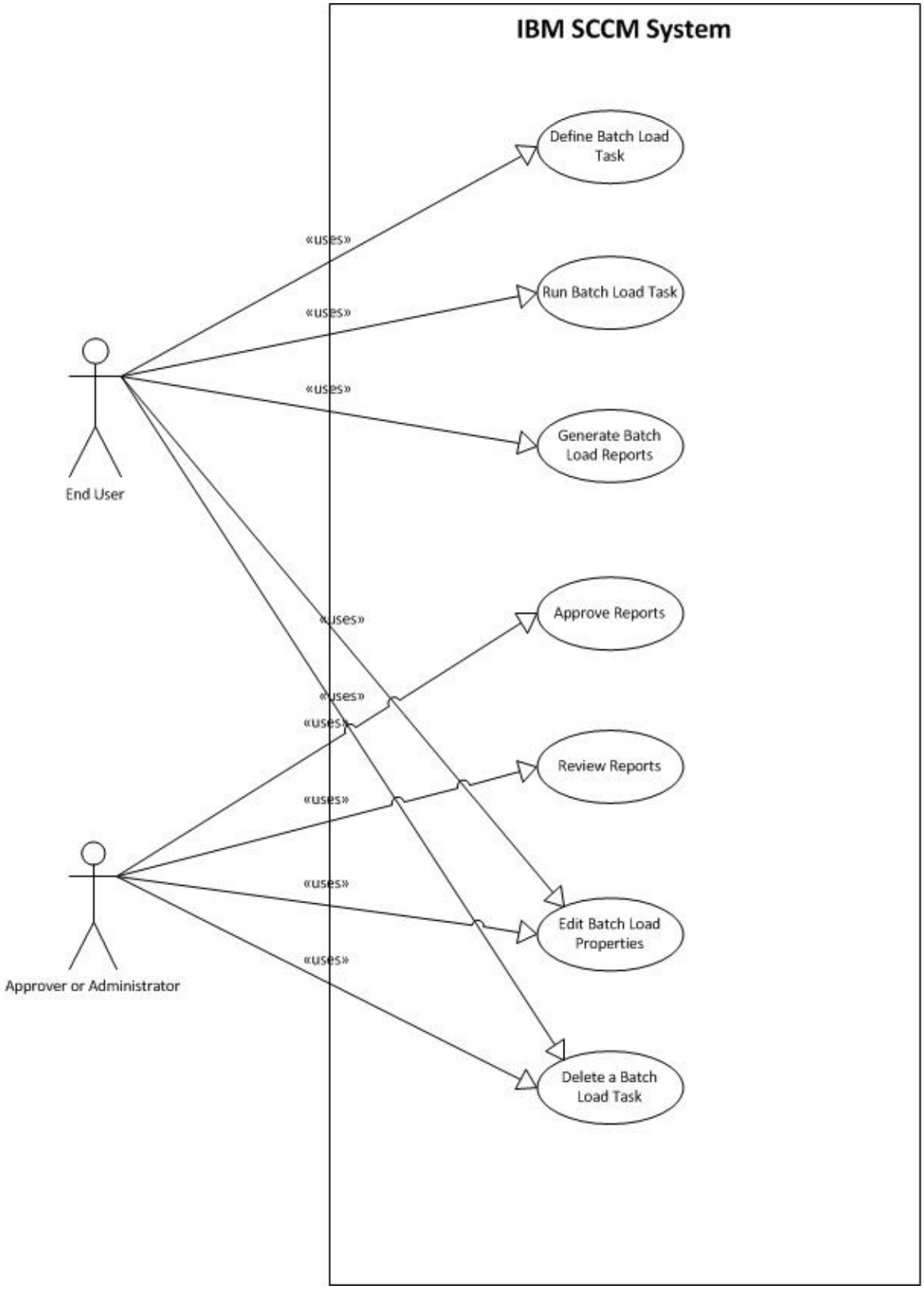

Figure [23]. Batch Load Use-Case Diagram

The role can be mapped with any number of users. In our scenario there exists two roles as defined in the use case diagram and each role can have many users mapped into it. It also possible to instruct the Process designer to create a role without creating a work queue. Additionally, a process designer can be used to delete a work queue which was created for a role that is defined in Case Manager Builder.

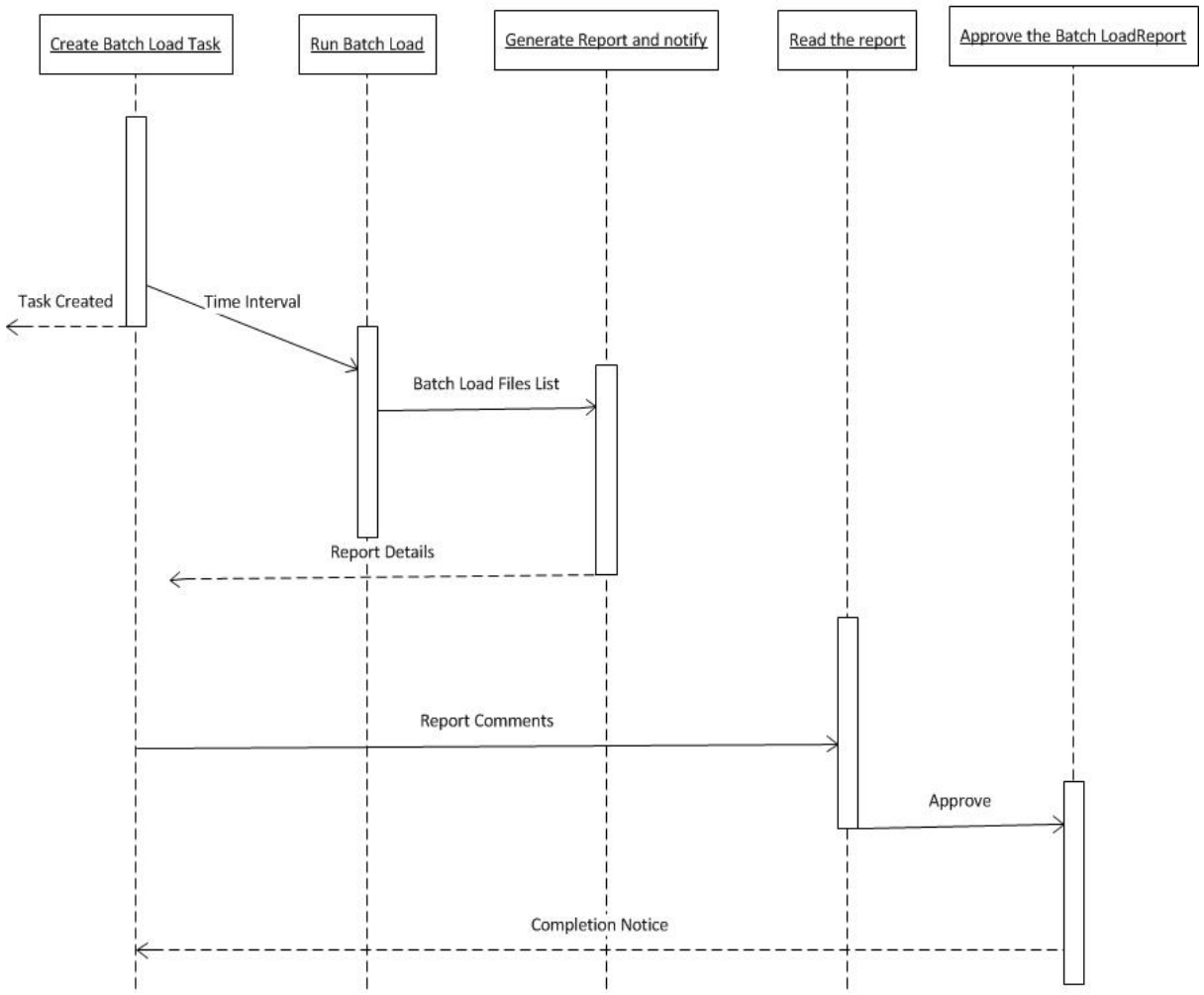

Figure [24]. Batch Load Sequence Diagram

The diagram above describes the sequence of activities to be performed by the SCCM batch load task.

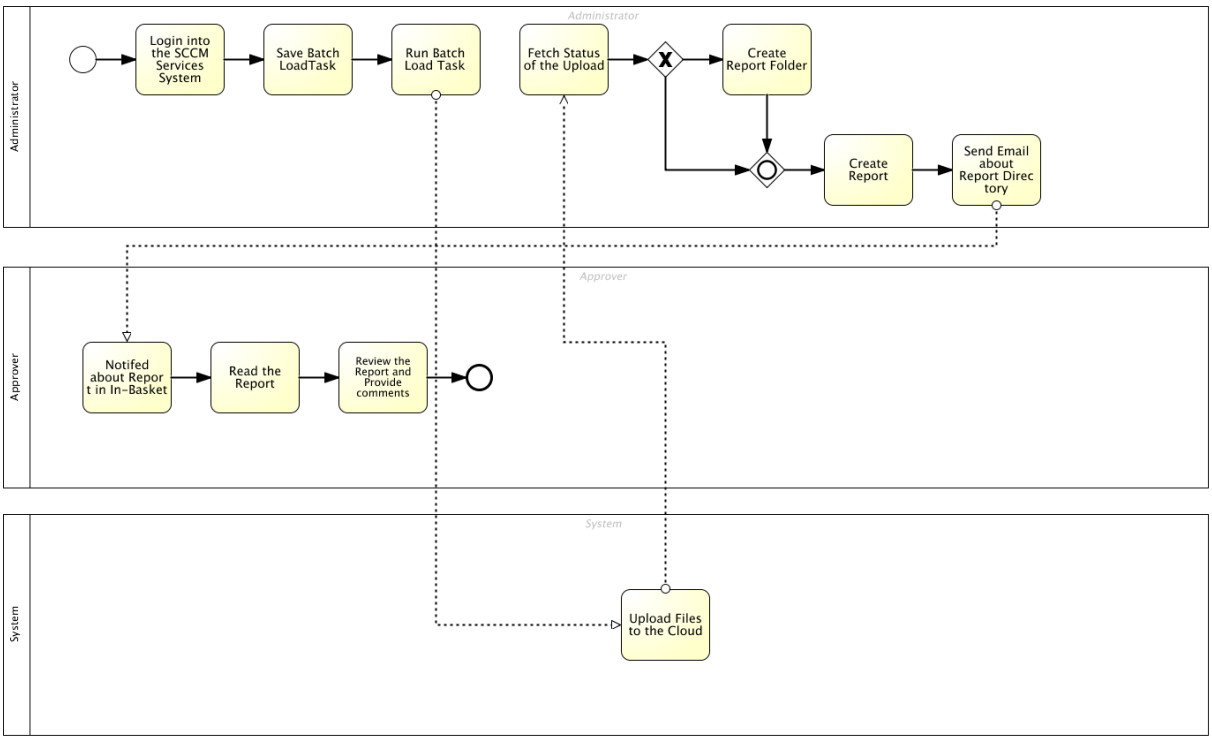

Figure [25]. Batch Load Workflow

6.2.2 Disposal Sweep Use-Cases with Workflows

A disposal sweep task can be otherwise called as bulk deletion. A disposal sweep can be scheduled to delete documents of particular type. Sometimes it becomes necessary for an approver to verify if the correct documents are being deleted [14]. Thus approval process is very important in case of a disposal sweep. Once the approver approves a disposal report and the deletion takes place in the system.

The sequence diagram below describes the disposal sweep operations and the sequence in which the system processes the request.

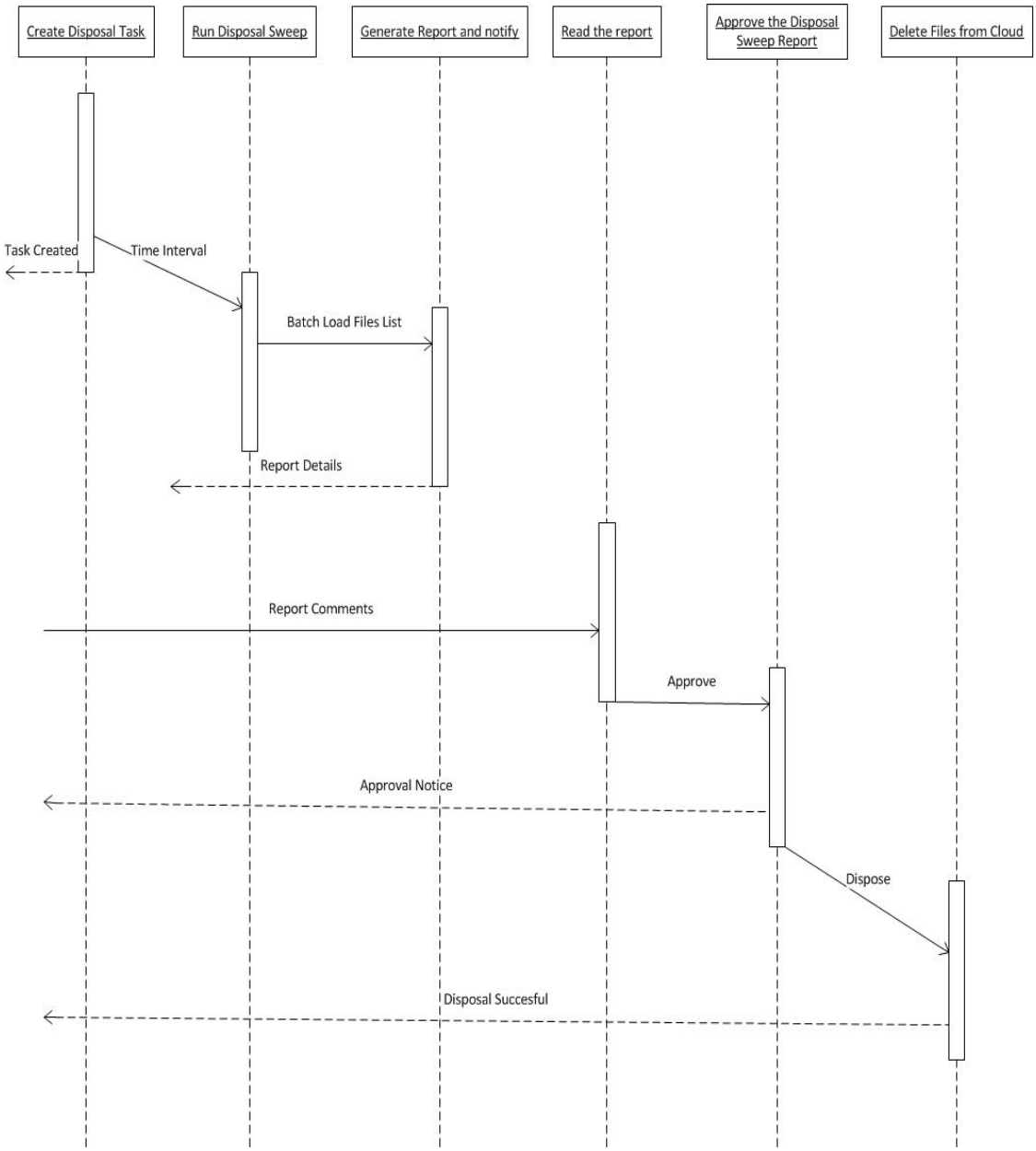

Figure [26]. Disposal Sweep Sequence Diagram

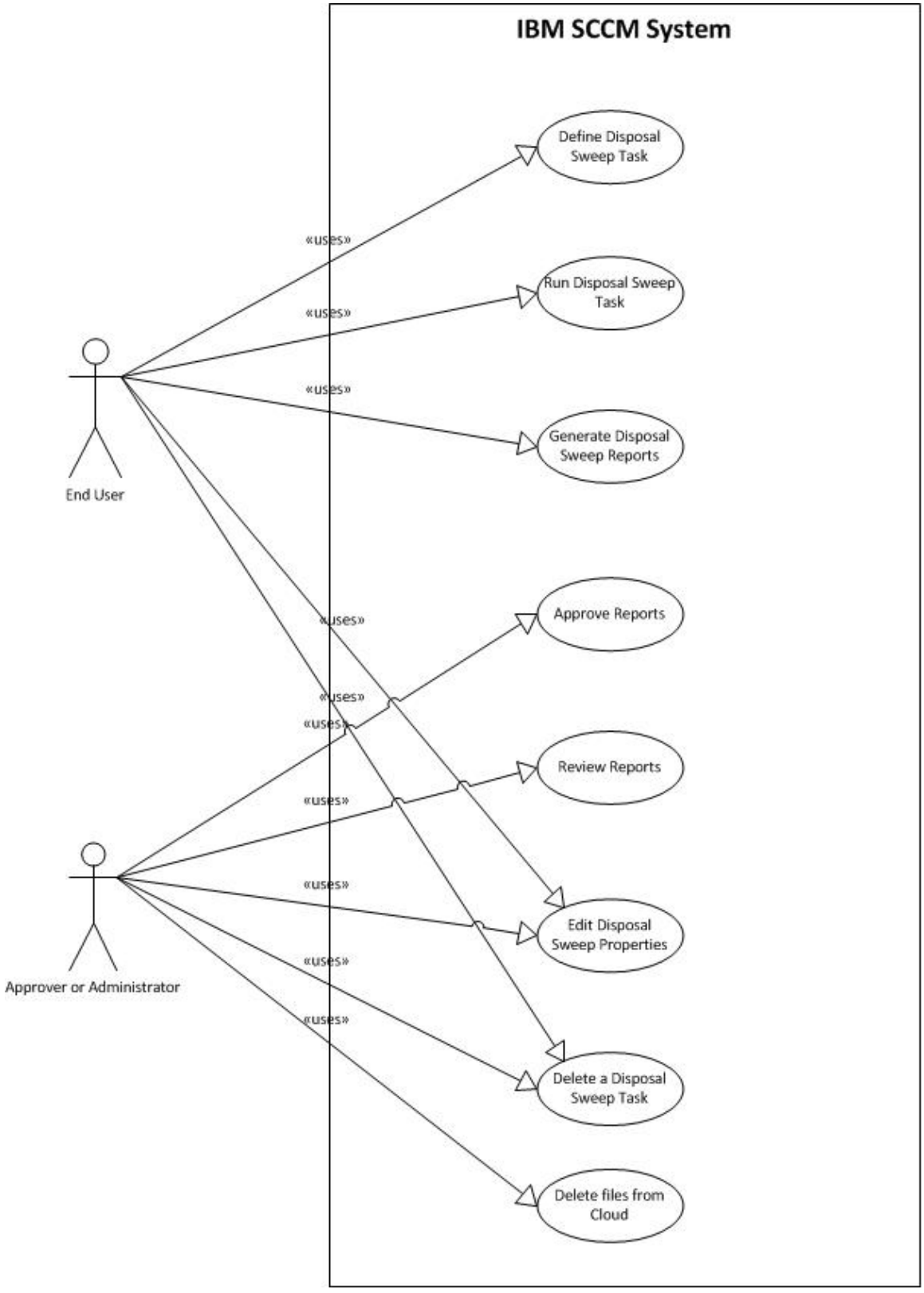

Figure [27]. Disposal Sweep Use-Case Diagram

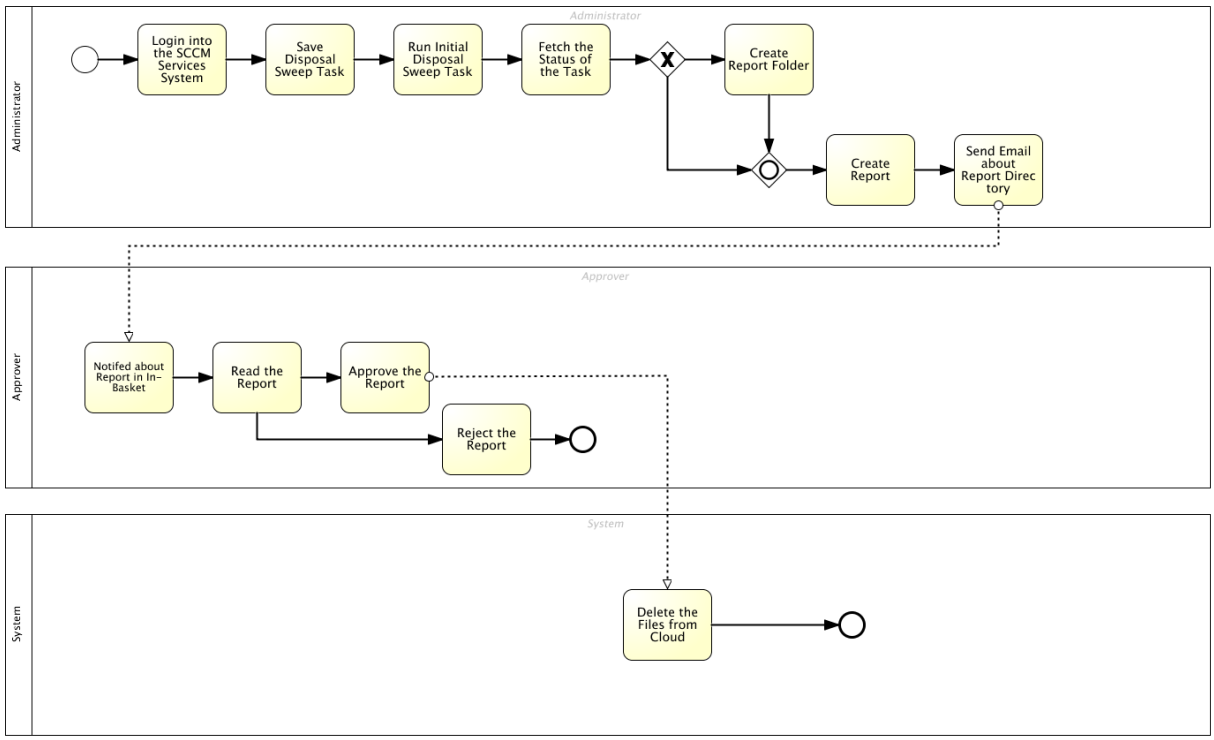

Figure [28]. Disposal Sweep Workflow

In case of a Disposal Sweep operation with an approval process, it's important to have an reviewer who can approve the reports and the sequence of operations to be performed is quite different compared to the batch load operations. The end user creates a disposal sweep with a predefined folder hierarchy to be deleted and the document types involved. The proposed system collects the information about the files which are eligible to be deleted based on the retention period. The file plan is assigned with a retention period by a background task which is not known to the end users. The workflow evaluates to file plan to find the documents which has the specific retention period.

The workflow then creates a report of the documents which matches the criteria of the retention period. Once this is done it notifies the approver about the report. This can be notified via the personal in-basket of the approver. The approver can step into the process and approve the report which by checking the documents to be deleted. This approval by the approval can be considered as an authorization for deletion of the documents from the cloud. The workflow then revisits the disposal sweep task and runs it again. This time after the approval the workflow deletes the files from the cloud. Thus workflow has the provision to check the approval status, if the approval is not done or if

the report is not approved by the reviewer then the deletion will not take place and the user in-basket will be notified with the next possible step.

6.3. Proposed Architecture for the SCCM System with IBM Content Navigator

This thesis also proposes an feasible architecture which uses the Case Manager components, IBM SCCM System and IBM Content Navigator which gives an scalable approach do deal with current IBM SCCM System challenges. The main factor to consider is the scalability issue wherein the user has to login into the SCCM Administrative client to do the administration tasks such as Batch load and Disposal sweep, so the proposal is to integrate the SCCM tasks into IBM Content Navigator which provides increased flexibility and ease of access for such tasks. The figure below gives the proposed architecture which diagrammatically explains how the integration of the IBM Content Navigator and the SCCM System will work. It has three important components namely SCCM System, Case Manager Component and the FileNet Platform Services Component.

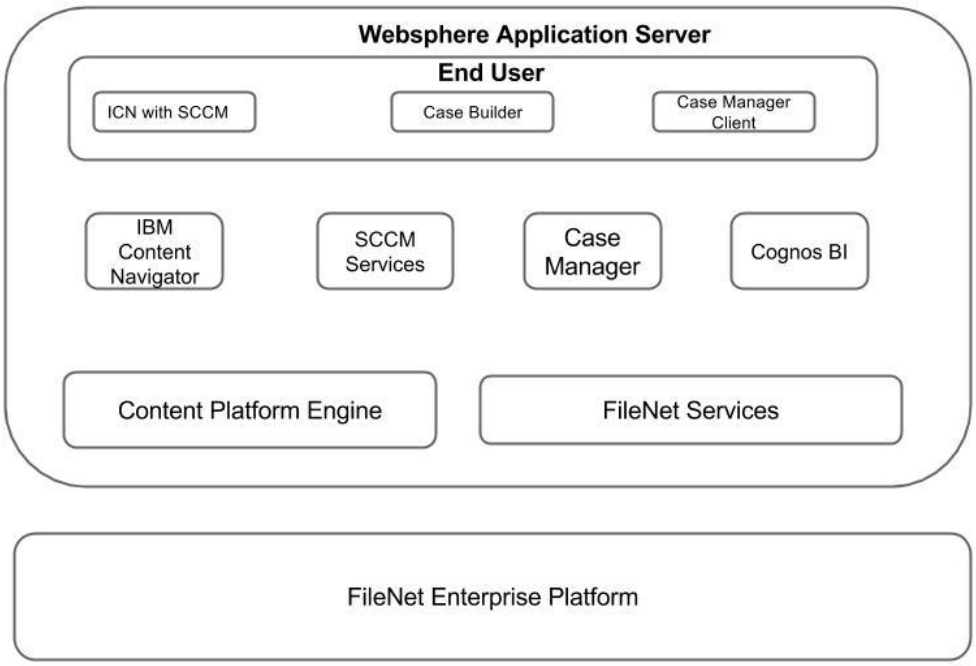

Figure [29]. High Level Architecture

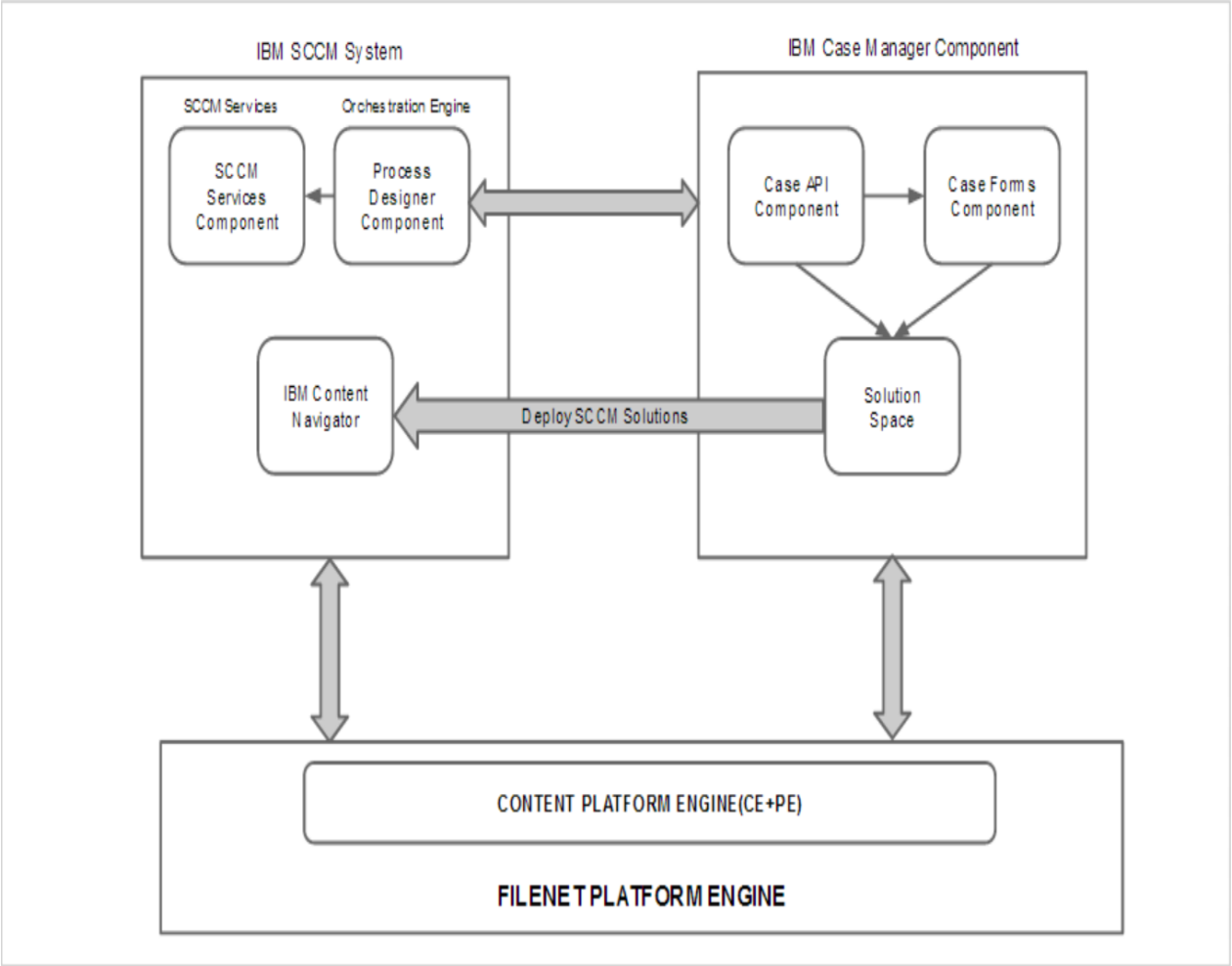

Figure [30]. SCCM with ICN Architecture

The FileNet platform provides the content engine services and process engine as Content Platform Engine (CPE) services. The Case Manager components use the CPE to develop case solutions. CPE also provides Process designer which can be considered as an orchestration engine where in the document centric workflows are defined. The properties of the workflow can be obtained from the case API component in the Case Manager. The workflow uses REST calls to consume the SCCM services. The Case solution is then converted into a solution package. This solution package can then be deployed into the production system as an add-on. This solution package can then be accessed via IBM Content Navigator component.

6.4. Structure of an SCCM Solution with Workflow components

An SCCM solution package can have the services offered by the IBM SCCM system for the ILG solution stack. A solution package has the pages to be used in the IBM Content Navigator as an alternative to the SCCM administration client, List of document classes to be used in the SCCM tasks, List of properties to be used for the tasks and other solution related information. The solution package follows an hierarchy where in the each solution has many case type which can be SCCM tasks like Batch load, Disposal Sweep etc and many roles. The solution space contains the pages to be shown to the end user based on his roles. The case type can tasks as described in the use cases and sequence diagrams. Roles can have In-Basket which displays the steps to be performed. This solution package can then be exported from the development environment. The solution can be made available in a market place wherein the customers can download the solution and import it in their production environment. The customer can then map the roles to the end users and use the solution. The figure below describes the hierarchy of a Solution package and the contents in it.

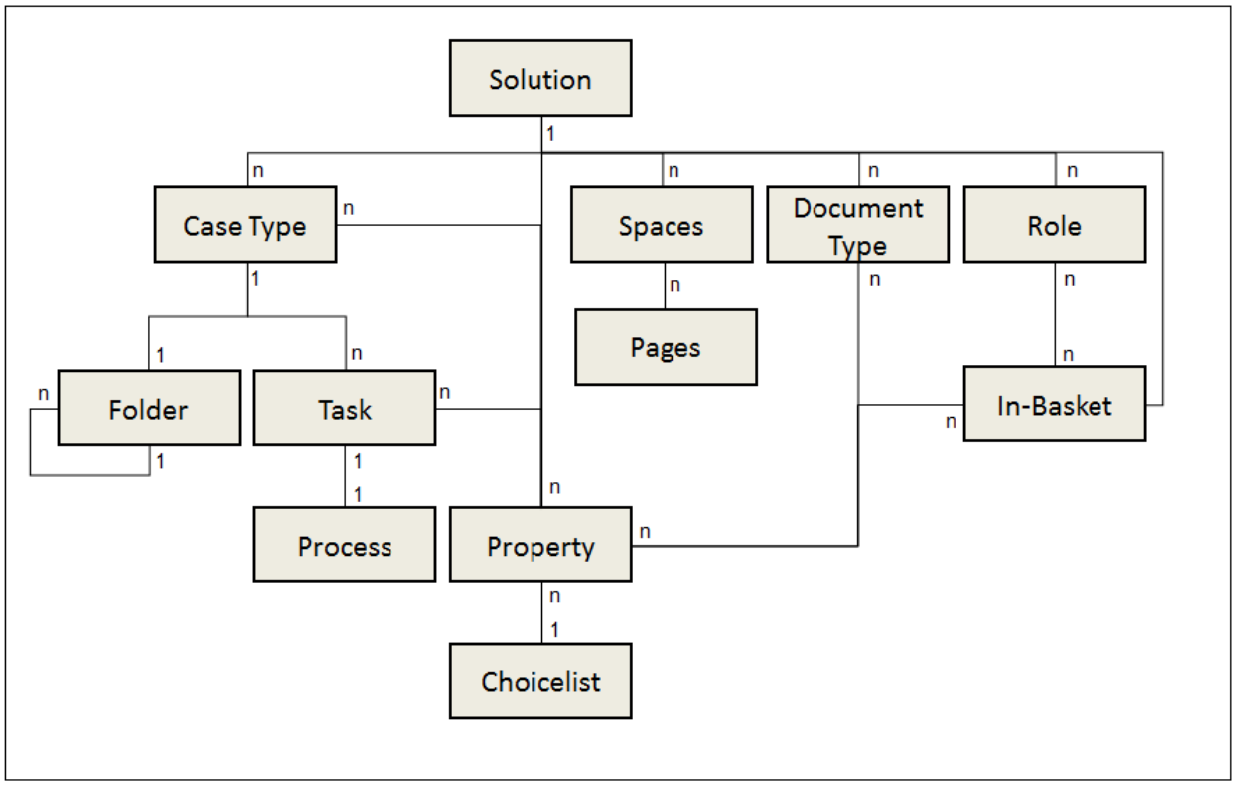

Figure [31]. A SCCM Solution Package Structure

6.5 Orchestration Engine in SCCM

The proposed SCCM system contains an orchestration engine which is responsible for the design and execution of the workflows. An orchestration engine can invoke many kinds of action in parallel across any number of nodes. Several useful actions are available by default, and you can easily add and use many new actions. In the SCCM scenario these actions can be document centric actions such as check in, check out, delete, download etc. The number of nodes supported by the orchestration engine enables the SCCM system to be scalable. Thus the scalability problem of the SCCM can be solved with ease. The process orchestration lets to integrate and coordinate web service operations across and beyond the enterprise.

Modularization and Reusability

One of the benefits of having a process orchestration engine is that atomic business components is already wrapped with a service which can be re used over again and again in the high level process design. The Orchestration engine makes it easy to accomplishing ILG based workflows by standardizing the ways of services which can be incorporated with each other. A central process which can be possibly a webservice which publishes various functionality of the ILG solution stack and coordinates the execution of different operations on the Web services involved in the operation.

6.6 Workflow In-Basket

The workflow In-Basket contains the next set of steps to be performed in the document management lifecycle process. In the SCCM scenario it has the approval process notification window and a step processor to process the approval process. The step processor can be configured to use the Content Navigator for the approval process. The approval process shows up the attachments such as report and other documents which are required by the approver to approve a process. Thus In-Basket plays a greater role in tracking the process engine operations. The In-Basket can be configured to be a public In basket, a group of user In basket or an individual In basket.

7. FUTURE WORK AND CONCLUSION

The business process management platform helps the business analysts to design, implement, deploy and execute the business processes quickly and efficiently. As discussed earlier the Data-Intensive computing presents new challenges for systems management in the cloud. One such challenge is that most of the data intensive applications may be built upon conventional frameworks, and requires variety of resources for efficient performance. The second challenge is that the parallel nature of large-scale data intensive applications requires that the provisioning and resource allocation should not performance bottlenecks. The next challenge is the need to have a framework to support effective scaling of resources when large amounts of data is retrieved or involved.

This thesis analyses the problems in Dynamic document centric workflow concepts and proposes the tools to be used for providing an efficient performance outcome. This thesis also took into consideration of the related works in the Enterprise Content Management as a cloud service and discussed about various methods used for automating the workload management, designing a workload management solution etc. Further the case manager integration points were also discussed in depth. In order to describe the solution the use cases from the IBM SCCM System is considered and proposed a solution for the challenges faced by the system.

The architecture also facilitates the optimal instantiation of the business processes by selecting concrete service implementations to maximize users utility, considering multiple implementations of abstract services and multiple possible deployments in the hybrid cloud environments.

7.1. Conclusion

The architecture proposed for the SCCM system use case has been used to develop document workflows with a small number of services. As the next phase, the platform will be piloted by a group of selected users in which roles concept can be clearly distinguished.

In conclusion, the Chapter 1 gave a brief overview of the introductory concepts in ILG solution stack. This part explains the solution scenarios which are required in the enterprise world. One such solution, IBM SCCM was also briefed. The services offered

by the IBM SCCM system were also explained. Chapter 2 gave an background about the Enterprise Content Management system and the Content Management operations like Create, Store or Preserver and Retrieve in an Enterprise perspective. In order to map the back group with an real world application, a brief description about IBM FileNet Content Manager was also discussed. With consideration to the Workload management view a survey on the related works was discussed in Chapter 3. This chapter also explains the underlying concepts in Data intensive workloads in cloud and some important specifications in provisioning of systems in the cloud. One such solution which was discussed was "Black box based Workload management solution", where there was a proposal to have an self optimizing workload management solution for cloud applications. The most important work on Automating business process lifecycle management in hybrid cloud solution was also discussed. This provided an underlying concept for the use cases discussed. This solution proposed the usage of the Business Process design tool and data flow bindings. In order to give a clear idea about the REST based content access, this thesis discussed about the VisualREST content management architecture which gave plethora of idea about XMPP protocol as well. The Chapter 4 discussed about the possible problems on providing Enterprise Content Management solutions in Cloud and some factors affecting. The integration points on the Case Manager tool was discussed in depth in Chapter 5 which gave an idea about the process designer and case solution concepts in the Case Manager architecture. With the Case Manger component in mind, the use cases for the SCCM system were defined. The sequence of operations to be performed and roles were also explained in Chapter 6. This thesis also proposed a feasible architecture for the SCCM system integration with ICN and clearly slated the way in which these solutions overcome some of the ECM in cloud problems.

Bibliography

1. 2010 19th IEEE International Workshops on Enabling Technologies: Infrastructures for Collaborative Enterprises, Pankaj Goyal: Enterprise Usability of Cloud Computing Environments Issues and Challanges, Larissa, Greece.
2. 2011 37th EUROMICRO Conference on Software Engineering and Advanced Applications (SEAA). Niko Mäkitalo,Heikki Peltola,Joonas Salo and Tuomas Turto: VisualREST: A Content Managment System for Cloud Computing Environment,Oulu, Finland.
3. 2011 IEEE International Conference on Web Services (ICWS), Hau Liu, Yasmine Charif, Gueyoung Jung, Andres Quiroz, Frank Goetz, Naveen Sharma: Towards simplifying and automating business process lifecycle management in hybrid clouds, Washington, DC, USA.
4. 2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), Rizwan Mian, Patrick Martin: Executing data-intensive workloads in Cloud, Ottawa, Canada.
5. 2012 IEEE 7th International Conference on Networking, Architecture, and Storage (NAS), Jianzong Wang, Rui Hua, Yifeng Zhu, Changsheng Xie, Peng Wang, Weijiao Gang: C-IRR: An Adaptive Engine for Cloud Storage Provisioning Determined by Economic Models with Workload Burstiness Consideration, Xiamen, China.
6. CGOC_ILG_LeaderReferenceGuide, by CGOC
7. Xiao Ding, Ruobing Luo, Ming Hui: A Multi-Tenant Oriented Cutomizable Compensation Mechanism For Workflows-2010 IC-BNMT,Beijing,China.
8. Dr.Ulrich Kampffmeyer: Trends in Records, Document and Enterprise Content Management, Germany.
9. Keep It Moving: Proactive workload management for reducing SLA violations in large scale SaaS clouds.
10. Tim Dörnemann,Ernst Juhnke,Bernd Freisleben: On-Demand Resource Provisioning for BPEL Workflows Using Amazon's Elastic Compute Cloud, Marburg, Germany
11. Trieu C.Chieu, Lianhzhao Zeng,Ajay Mohindra: An Extensible Enterprise Content Management System with Service Component Architecture,IBM T.J Watson Research Center,NY ,USA
12. Barker, Adam; Weissman, Jon B.; van Hemert, Jano: Orchestrating Data-Centric Workflows, w: 2008 8th IEEE International Symposium on Cluster Computing and the Grid (CCGrid). Lyon, France, s. 210–217.
13. Bing, Han; Dan-Mei, Xia: Research and Design of Document Flow Model Based on JBPM Workflow Engine, w: 2009 International Forum on Computer Science-Technology and Applications. Chongqing, China, s. 336–339.
14. IBM SCCM Cloud, Architecture Document, IBM, USA.
15. Bruneo, Dario; Distefano, Salvatore; Longo, Francesco; Puliafito, Antonio; Scarpa, Marco (2013): Workload-Based Software Rejuvenation in Cloud Systems. w: *IEEE Trans. Comput.* 62 (6), s. 1072–1085. DOI: 10.1109/TC.2013.30.

16. Chieu, Trieu C.; Zeng, Liangzhao: Service-Oriented Approach for Implementing an Extensible Content Management System, w: 2008 IEEE Congress on Services Part II (SERVICES-2). Beijing, China, s. 96–103.
17. Gao, Ming; Jiang, Jichen: Towards Content-Oriented Integration for Process-Centric Knowledge Management Systems, w: 2009 Asia-Pacific Conference on Information Processing, APCIP. Shenzhen, China, s. 529–532.
18. Goyal, Pankaj: Enterprise Usability of Cloud Computing Environments: Issues and Challenges, w: 2010 19th IEEE International Workshops on Enabling Technologies: Infrastructures for Collaborative Enterprises. Larissa, Greece, s. 54–59.
19. Guo, Yuanxiong; Fang, Yuguang; Khargonekar, Pramod P.: Optimal Workload and Energy Storage Management for Cloud Data Centers, w: MILCOM 2013 - 2013 IEEE Military Communications Conference. San Diego, CA, USA, s. 1850–1855.
20. IBM: Advanced Case Management with IBM Case Manager.
21. IBM: Customizing and Extending IBM Content Navigator.
22. IBM: IBM FileNet Content Manager Implementation Best Practices and Recommendations.
23. IBM: Introducing IBM FileNet Business Process Manager.
24. Kochut, A.; Deng, Y.; Head, M. R.; Munson, J.; Sailer, A.; Shaikh, H. i wsp. (2011): Evolution of the IBM Cloud: Enabling an enterprise cloud services ecosystem. w: *IBM J. Res. & Dev.* 55 (6), s. 7:1. DOI: 10.1147/JRD.2011.2170920.
25. Li, Xiufang; Liu, Qingxiu; Hou, Ruilian: The Application of Workflow Technology on Dynamic PDM Document Management, w: 2012 International Conference on Computer Science and Electronics Engineering (ICCSEE). Hangzhou, Zhejiang, China, s. 559–562.
26. Lin, Jiazao; Zhao, Zhili; Li, Shoubo; Zhang, Huajian; Li, Lifen; Li, Caihong; Li, Lian: Application-Objected Workflow Management System Based on Abstract Service, w: 2009 Eighth International Conference on Grid and Cooperative Computing (GCC). Lanzhou, Gansu, China, s. 144–152.
27. Liu, Hua; Charif, Yasmine; Jung, Gueyoung; Quiroz, Andres; Goetz, Frank; Sharma, Naveen: Towards Simplifying and Automating Business Process Lifecycle Management in Hybrid Clouds, w: 2012 IEEE 19th International Conference on Web Services (ICWS). Honolulu, HI, USA, s. 592–599.
28. Makitalo, Niko; Peltola, Heikki; Salo, Joonas; Turto, Tuomas: VisualREST: A Content Management System for Cloud Computing Environment, w: 2011 37th EUROMICRO Conference on Software Engineering and Advanced Applications (SEAA). Oulu, Finland, s. 183–187.
29. Mian, Rizwan; Martin, Patrick: Executing Data-Intensive Workloads in a Cloud, w: 2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid). Ottawa, Canada, s. 758–763.
30. Mogill, Jace A.; Haglin, David J.: Toward Parallel Document Clustering, w: Distributed Processing, Workshops and Phd Forum (IPDPSW). Anchorage, AK, USA, s. 1700–1709.
31. Palm, Chad: The IBM Software Story Customer Presentation – for seller input DRAFT / Discussion Document April 19, 2010.

32. Pathirage, Milinda; Perera, Srinath; Kumara, Indika; Weerawarana, Sanjiva: A Multi-tenant Architecture for Business Process Executions, w: 2011 IEEE International Conference on Web Services (ICWS). Washington, DC, USA, s. 121–128.

33. Rahaman, Mohammad Ashiqur; Roudier, Yves; Schaad, Andreas: Document-Based Dynamic Workflows: Towards Flexible and Stateful Services, w: 2009 World Conference on Services - II (SERVICES-2). Bangalore, Karnataka, India, s. 87–94.

34. Rimal, Bhaskar Prasad; El-Refaey, Mohamed A.: A Framework of Scientific Workflow Management Systems for Multi-tenant Cloud Orchestration Environment, w: 2010 19th IEEE International Workshops on Enabling Technologies: Infrastructures for Collaborative Enterprises. Larissa, Greece, s. 88–93.

35. Tran, Hai V.: Data Management Challenges in Cloud Computing, w: 2013 13th International Conference on Computational Science and Its Applications (ICCSA). Ho Chi Minh City, Vietnam, s. 19–27.

36. Wang, Jianzong; Hua, Rui; Zhu, Yifeng; Xie, Changsheng; Wang, Peng; Gong, Weijiao: C-IRR: An Adaptive Engine for Cloud Storage Provisioning Determined by Economic Models with Workload Burstiness Consideration, w: 2012 IEEE 7th International Conference on Networking, Architecture, and Storage (NAS). Xiamen, China, s. 79–83.

37. Wu, Haishan; Tantawi, Asser N.; Yu, Tao: A Self-Optimizing Workload Management Solution for Cloud Applications, w: 2013 IEEE International Conference on Web Services (ICWS). Santa Clara, CA, USA, s. 483–490.

38. Yang, J.; Yu, T.; Jian, L. R.; Qiu, J.; Li, Y. (2011): An extreme automation framework for scaling cloud applications. w: *IBM J. Res. & Dev.* 55 (6), s. 8:1. DOI: 10.1147/JRD.2011.2170929.

39. Zaigham Mahmood, Richard Hill: Cloud Computing for Enterprise Architectures (Computer Communications and Networks).

40. Zhang, Hui; Jiang, Guofei; Yoshihira, Kenji; Chen, Haifeng (2014): Proactive Workload Management in Hybrid Cloud Computing. w: *IEEE Trans. Netw. Serv. Manage.*, s. 1–11. DOI: 10.1109/TNSM.2013.122313.130448.

41. Zhang, Yongliang; Ren, Huaiqing: Transaction in Document Approval Process Workflow, w: 2013 Fifth International Conference on Computational and Information Sciences (ICCIS). Shiyang, China, s. 458–461.

42. A report for the society for Information Management Advanced Practices Council,Storage Networking Industry, Data Management Forum.

Declaration

I hereby declare that the work presented in this thesis is entirely my own.
I did not use any other sources and references than the listed ones. I have marked all direct or indirect statements from other sources contained therein as quotations.
Neither this work nor significant part of it was part of another examination procedure. I have not published this work in whole or in part before.
The electronic copy is consistent with all submitted copies.

Signature

(Pradeep Parameshwaran)
(Stuttgart, 09-04-2014)