Literature Study

Business Process Management in the cloud: Business Process as a Service (BPaaS)

Author: Evert Duipmans
Supervisor: Dr. Luís Ferreira Pires

April 1, 2012
## Contents

1 Introduction 3  
1.1 Motivation 3  
1.2 Objectives 5  
1.3 Approach 5  
1.4 Structure of the report 5  

2 Business Process Management 6  
2.1 BPM lifecycle 6  
2.2 Design 7  
2.2.1 Business Process Management and Notation (BPMN) 8  
2.3 Implementation 10  
2.3.1 Orchestration vs. Choreography 11  
2.4 Enactment 11  
2.5 Business Process Management System 12  
2.6 Conclusion 13  

3 Cloud Computing 14  
3.1 General benefits and drawbacks 14  
3.2 Service models 16  
3.2.1 Infrastructure as a Service (IaaS) 16  
3.2.2 Platform as a Service (PaaS) 18  
3.2.3 Software as a Service (SaaS) 19  
3.3 Cloud types 21  
3.3.1 Public Cloud 21  
3.3.2 Private Cloud 21  
3.3.3 Hybrid Cloud 21  
3.3.4 Community Cloud 21  
3.4 Cloud providers 22  
3.4.1 Amazon 22  
3.4.2 Microsoft Windows Azure 23  
3.4.3 Google App Engine 26  
3.5 Conclusion 28
4 BPM and cloud computing
  4.1 General benefits and drawbacks .......................................................... 29
  4.2 Service Models .......................................................... 29
    4.2.1 IaaS .............................................................. 30
    4.2.2 PaaS .............................................................. 30
    4.2.3 SaaS .............................................................. 31
  4.3 Combining on-premise and cloud .......................................................... 32
    4.3.1 Architecture .............................................................. 33
    4.3.2 Optimal distribution .............................................................. 34
  4.4 Social BPM .............................................................. 35
  4.5 Environments .............................................................. 36
    4.5.1 Product overview .............................................................. 36
    4.5.2 Oracle Fusion .............................................................. 36
    4.5.3 Appian .............................................................. 37
    4.5.4 Fujitsu InterstageBPM.com .............................................................. 38
    4.5.5 Barium Live .............................................................. 38
  4.6 Conclusion .............................................................. 38

5 Research directions .............................................................. 40
  5.1 BPEL engine deployment in the cloud .............................................................. 40
  5.2 Social BPM .............................................................. 40
  5.3 Process decomposition .............................................................. 41

6 Conclusion .............................................................. 44
Chapter 1

Introduction

This work investigates the possibility of integrating Business Process Management (BPM) with cloud computing. Both subjects are described in more detail and already existing research about the subject and already existing tools are investigated. This chapter gives the motivation behind the subject, the objectives and approach of the research, and explains the structure of the report.

1.1 Motivation

BPM has gained a lot in popularity the last two decades. By identifying and managing business processes, companies get insight in what they are doing, which parts of the process can be automated or can be optimized. This may lead to lower costs, better customer satisfaction or optimized processes for creating new products at lower cost [30].

Business Process Management Systems (BPMS) are able to capture business processes and keep track of running instances of these processes. A business process can be described using workflows. Workflows consist of activities which are either human-based, system-based or a combination of the two in case of human-system interaction. A BPMS contains a workflow execution engine which coordinates the execution of a business process step by step. Each process instance is monitored by the BPMS and business users are able to look into these process instances. The process instances give insight into the progress of the processes and show if processes are completed successfully, or have failed. In case of failure, the process instance shows in which part of the process the failure occurred. By monitoring, evaluating and changing business processes, companies are able to optimize their processes.

To be able to control and monitor a business process, a BPMS needs to
communicate with the participants in the process. Introducing BPM into a company might lead to integration issues, especially since legacy software may not have standardized interfaces. Not all software that participates in the business process offers clear interfaces for controlling or monitoring the software.

Nowadays a lot of software vendors do offer their products as a service. The software is hosted on a server and customers are able to login to the software using the Internet, mostly by using a browser. One of the big advantages here is that companies can change their product relatively easily without having to distribute the updates to all their customers. Instead, they only have to adjust the software that is running on their own servers. Many of these services are built using the Service-Oriented Architectures (SOA) principles. SOA[27] allows designing and developing software services in a uniform way, so that reuse is enabled. This gives developers support for integrating and composing services.

Offering software as a service gives providers also new challenges. In the previous situation, where software was shipped to the user, the hardware on which the software would run was provided and managed by the user. In the new situation the software runs on the machines of the provider. This means that the provider is responsible for the hardware and that maintenance has to be performed by the provider. The provider has a choice to invest in servers and personnel to perform these tasks, or the choice can be made to outsource the hosting of the software to a third-party provider. Cloud computing [9] is an example of a model where computing resources are offered to the user as a service.

The NIST describes cloud computing as a model for shared configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction [25]. Cloud computing provides users with 3 service models:

**Software as a Service (SaaS)** SaaS is a model in which software is offered as a service to the user. The software is hosted on a server and users access the software by using a web browser.

**Platform as a Service (PaaS)** PaaS is the offering of a computing platform as a service. Users are able to deploy their applications on such a platform. The platform offers auxiliary functionality such as a web server, databases, load balancing and more.

**Infrastructure as a Service (IaaS)** With IaaS, the cloud offers platform virtualization to the customer. The user is offered a virtual machine with some storage. Instead of buying servers and other network equipment, users just rent these resources.
1.2 Objectives

The goal of this work is to investigate the possibility of combining BPM and cloud computing. By moving BPMS software to the cloud, companies do not have to buy and maintain expensive servers to manage and coordinate their own business process. Instead, the business process can be uploaded to a service in the cloud that performs and monitors the process.

This report gives an overview of problems and solutions of business process management in the cloud, by looking at existing solutions and scientific papers that discuss the topic. The ultimate goal of this work is to identify research issues for a master assignment.

1.3 Approach

The following steps are taken in order to pursue the defined goals:

1. Study the literature about BPM and identify the phases and tools within BPM that can benefit from cloud computing.

2. Study the literature about cloud computing, in order to identify the benefits and drawbacks of cloud computing in general and also the specific benefits and drawbacks of the 3 service models.

3. Study literature about the combination of both cloud computing and BPM.

4. Discuss existing products where BPM is already combined with cloud computing.

5. Devise possible opportunities for further research.

1.4 Structure of the report

Chapter 2 explains business process management by considering the BPM lifecycle and investigating several tools and languages that are useful in the management process. Chapter 3 elaborates on cloud computing. The fundamentals of cloud computing are described and an overview of cloud platforms is given. Chapter 4 investigates the combination of BPM and cloud computing by discussing scientific literature about the subject and by giving an overview of existing software. Chapter 5 identifies research directions for a master assignment. Chapter 6 gives our conclusions.
Chapter 2

Business Process Management

The goal of BPM is to identify the internal business processes of an organization, capture these processes in process models, manage and optimize these processes by monitoring and reviewing them.

Business process management is based on the observation that each product that a company provides to the market is the outcome of a number of activities performed \[30\]. These activities can be performed by humans, systems or a combination of both. By identifying and structuring these activities in workflows, companies get insight into their business processes. By monitoring and reviewing their processes, companies are able to identify the problems within these processes and can come up with improvements.

This chapter discusses business process management by considering the BPM lifecycle. Each of the phases of the lifecycle is investigated. The chapter concludes by giving an overview of the BPM phases that are relevant for this work.

2.1 BPM lifecycle

The BPM lifecycle is an iterative process in which all of the BPM aspects are covered. The BPM lifecycle is shown in Figure 2.1 and consists of the following phases:

- **Design**
  The design phase consists of identifying existing processes and capturing the business processes in process models.
• **Implementation**
  In the implementation phase, the designed process is implemented in an executable process language, which can be deployed in a BPMS.

• **Enactment**
  The enactment phase is the runtime phase of the lifecycle. The business process is deployed and monitored by a BPMS.

• **Evaluation**
  In the evaluation phase the monitored information that is collected by the BPMS is used to review the business process. The conclusions drawn in the evaluation phase are input for the next iteration of the lifecycle.

In the remainder of this chapter, the first 3 phases of the lifecycle are explained in more detail. More information about the 4th phase is omitted in this report, since the activities involved in this phase are not relevant for this work.

### 2.2 Design

In the design phase the business processes within a company are identified. The goal of the design phase is to capture the processes in business process models. These models are often defined using a graphical notation. In this way, stakeholders are able to understand the process and refine the models
easily. The activities within a process are identified by surveying the already existing business process, by considering the structure of the organization and by identifying the technical resources within the company. BPMN[21] is the most popular graphical language for capturing business process models in the design phase.

When the business processes are captured within models, these models can be simulated and validated. By validating and simulating the process, the stakeholders get insight into the correctness and suitability of the business process models.

2.2.1 Business Process Management and Notation (BPMN)

BPMN is a notation for defining business process models, developed by the Business Process Management Initiative [31]. The first version of the standard was released in 2004. Recently, in March 2011 BPMN 2.0 was released. BPMN has been designed to be understandable for all the business users that have to deal with process diagrams. This means that business analysts need to be able to create the first sketches of the models, but the diagrams also have to be understandable for technical developers who have to implement the business processes represented in the diagrams and employees who have to monitor the processes.

Language constructs

BPMN diagrams are modelled as flowcharts, consisting of activities that are connected to each other to determine their relations. The most important constructs available in the language are:

- **Event**  
  Events effect the flow of a process. Three types of events are defined in BPMN, namely Start, Intermediate and End events.

- **Activity**  
  Activities can be described as work performed by a company. There are two types of activities: tasks and sub-processes.

- **Gateway**  
  Gateways are controls that are available for changing sequence flow. Condition/decision paths are examples of situations that can be modelled by gateways.

- **Sequence Flow**  
  A sequence flow is used for ordering flow elements, and is used to show the sequence of activities that are performed in the flowchart.
• **Message Flow**
Message flows are flow elements that indicate the flow of messages from one participant to another participant. Message flows are used to model the communication between two separate flow charts.

• **Association**
Associations are available for associating data, text and other artefacts with flow objects.

• **Pool**
Pools can be used to represent the participants of a process. Activities that are related can be grouped into a pool.

• **Lane**
Lanes are partitions within a pool. They can be used for organizing and categorizing activities.

**Example**

The following example is adapted from OMG’s BPMN examples document [20]. The BPMN diagram of the example is shown in Figure 2.2. The example shows a Business-To-Business-Collaboration between a pizza vendor and a customer. Both participants are modelled using a pool. Inside these pools, the internal business process of the participants is shown. The pizza vendor has 3 employees: a clerk, the pizza chef and the delivery boy. Each of these employees is represented within the pizza vendor pool by a lane.

The business process of the customer starts when a customer is hungry and wants to eat a pizza. The customer has to perform two activities, namely a pizza needs to be selected and ordered by sending a message to the pizza vendor. The next step for the customer is to wait for the pizza. This has been modelled by an event based gateway (diamond shaped), which indicates that two possible events can occur. The first possibility is that a customer waits for 60 minutes and still has no pizza. In this case, the customer contacts the pizza vendor and asks when the pizza will be delivered. After the call, the customer starts waiting again. In the second case, the pizza is delivered within 60 minutes. In this situation, the customer pays for the pizza and eats the pizza. After these activities the customers hunger is satisfied and the business process terminates.

The business process of the vendor starts when an “order received” event is received by the clerk. The gateway after the event introduces parallelism to the process. A “Bake the Pizza” order is given to the pizza chef and the clerk self is waiting for an incoming “Where is my pizza?” event. As soon as the pizza chef finishes his pizza baking activity, he orders the pizza boy
to deliver the pizza. The pizza boy delivers the pizza to the customer and waits for the payment. When the pizza money is received, the pizza boy gives the customer a receipt and the process terminates.

![BPMN diagram of a pizza vendor and a customer](image)

Figure 2.2: BPMN diagram of a pizza vendor and a customer [20]

### 2.3 Implementation

After the business process models are validated and simulated, they have to be implemented. The implementation of these models can be done in two ways. One can choose to create work lists, with well defined tasks, which can then be assigned to workers within the company. This is often the case when there is no automation within the business process execution. The disadvantage of working with work lists is that the process execution is hard to monitor. There is no central system in which process instances are monitored, this has to be done by each employee within the company that is involved in the process.

In a lot of situations information systems participate in a business process, in which case a business process management system (BPMS) can be used. A BPMS is able to use business process models and create instances of these models for each process initiation. The advantage of using a BPMS
is that the system gives insight into the whole process. The system is able to monitor each instance of a business process and gives an overview of the activities that are performed, the time the process takes and its completion or failure.

Business Process Management Systems need executable business models. The models defined in the design phase are often too abstract to be directly executed. Therefore, they need to be implemented in an executable business process language, such as BPEL[1]. In addition, collaborations between business processes can be implemented by using a choreography language, such as CDL[24]. Below, the difference between orchestrations and choreographies is explained in more detail.

### 2.3.1 Orchestration vs. Choreography

An Orchestration describes how services can interact with each other at the message level, including the business logic and execution order of the interactions from the perspective and under control of single endpoint [27]. BPEL[1] is an XML-based language that can be used to describe orchestrations. The language gives its users the freedom to describe business processes in two ways: executable or abstract. Abstract processes serve a descriptive role, whereas executable processes are meant to be executed by an execution engine. An example of orchestration can be found in the pizza collaboration example. The example has two pools, that represent separate business processes. The activities performed within a pool can be described as a process in an orchestration language.

Choreography is typically associated with the public message exchanges, rules of interaction, and agreements that occur between multiple business process endpoints, rather than a specific business process that is executed by a single party [27]. CDL[24] can be used for describing choreographies using an XML-based format. CDL allows its users to describe how peer-to-peer participants communicate within the choreography. The communication between the two pools in the pizza collaboration example is an example of a choreography, where two participants with two different business processes collaborate with each other.

### 2.4 Enactment

When the business process models are implemented in the implementation phase, the enactment phase can be started. In this phase the system is used at runtime, so that each initiation of the process is monitored and coordinated by the BPMS. For each initiation of a process, a process instance is
created. The BPMS keeps track of the progress within each of the process instances. The most important tool within the enactment phase is the monitoring tool, since it gives an overview of the running and finished process instances. By keeping track of these instances, problems that occur in a process instance can be easily detected.

2.5 Business Process Management System

Several vendors of Business Process Management software solutions offer complete suites for modelling, managing and monitoring business processes. Inside these systems there is a process execution environment, which is responsible for the enactment phase of the BPM lifecycle [30]. An abstract schema of a typical BPMS is shown in Figure 2.3.

![Figure 2.3: Schematic representation of a business process management system][3]

The tools shown in Figure 2.3 provide the following functionality:

- The Business Process Modelling component consists of tools for creating business process models. It often consists of graphical tools for developing the models.

- Business Process Environment is the main component that triggers the instantiation of process models

- The Business Process Model Repository is a storage facility for storing process models as created by the modelling component

- The Process Engine keeps track of the running instances of process models. It communicates with service providers in order to execute activities or receive status updates.
• Service Providers are the information systems or humans that communicate with the process engine. These entities perform the actual activities and report to the process engine.

2.6 Conclusion

In this chapter we have introduced BPM by considering the phases of the BPM lifecycle. In this section we identify which of the introduced phases, tools and languages are relevant for this work.

The use of standardized languages such as BPMN and BPEL is interesting for this work, since both design and enactment tools can be placed in the cloud and work with these languages.

In addition, the structure of a BPMS is relevant. The different components within a BPMS and the relation between these components might change when a BPMS is moved to the cloud.
Chapter 3

Cloud Computing

Cloud computing is one of the trending topics in Computer Science. Many market influencing players as Microsoft, Google and Amazon offer cloud computing solutions. The goal of this chapter is to introduce cloud computing from both a conceptual level and a more concrete level. At first the general benefits and drawbacks of cloud computing are explained briefly. The three common service models are introduced next and for each of these service models the specific benefits and drawbacks are identified. After that, four different cloud types are discussed. In the end of the chapter, three popular cloud platforms are introduced in terms of their purposes and structure.

3.1 General benefits and drawbacks

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction [25].

The idea of cloud computing is that users are offered computing resources in a pay-per-use manner that are perceived as being unlimited. The cloud provider does not have any expectations or up-front commitments with the user and it offers the user elasticity to scale up or down quickly according to the user’s needs.

Cloud computing gives organizations several benefits:

- **Elasticity**
  Instead of having to buy additional machines, computing resources can be reserved and released as needed. This means that there is no under- or over-provisioning of hardware by the cloud user.
• **Pay-per-use**
  Cloud users are only billed for the resources they use. If a cloud user needs 20 computers once a week for some computation of one hour, it is only billed for these computing hours. After that the computers can be released and can be used by other cloud users.

• **No hardware maintenance**
  The computing resources are maintained by the cloud provider. This means that operational issues such as data redundancy and hardware maintenance are attended by the cloud provider instead of the cloud user.

• **Availability**
  Clouds are accessible over the Internet. This gives cloud users the flexibility to access their resources over the Internet. Cloud users are able to use software or data that is stored in the cloud not only inside their organization but everywhere they are provided with Internet access.

There are also drawbacks and threats in using cloud computing:

• **Security**
  Data is stored inside the cloud and accessible through the Internet. In several situations cloud users deal with confidential information that should be kept inside the cloud user’s organisation. In these situations cloud computing might not be a good solution, although there are solutions with cloud computing in which data is stored inside the cloud user’s organisation but applications are hosted in the cloud. There are also technical solutions for making data unintelligible for unauthorized people, for example, by using encryption algorithms.

• **Availability**
  Clouds are accessible through the Internet. This gives cloud users the freedom to work with the services everywhere they have an Internet connection. The downside is that when the Internet connection fails, for example, on the side of the cloud provider, cloud users are not able to access their services any more. This might lead to business failures, especially when the services are part of a business process.

• **Data transfer bottlenecks**
  Users that use software systems might need to transfer large amounts of data in order to use the system. Data should be transported not only from the user to the system, but also to multiple systems in order to cooperate inside a company. Cloud computing providers do not only bill the computation and storage services, but also data transportation is measured and billed. For companies that deal with a lot of data, cloud computing may be expensive because of the data transportation costs. Another problem can be the time it takes to transfer data to
the cloud. For example, a company needs to upload a huge amount of data in order to perform a complex computation. The data transfer may take more time than the computation itself. In these situations it might be faster and cheaper to perform the computation inside the premisses of the cloud user.

3.2 Service models

The National Institute of Standards and Technology (NIST) defines cloud computing by means of three service models: Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS) and Infrastructure-as-a-Service (IaaS) [25]. The three service models are closely related and can be seen as a layered architecture, as shown in Figure 3.1. Each service model is explained in the sequel. For each of the models, the specific benefits and drawbacks are given for both the user and the provider of the service models.

![Figure 3.1: An overview of the layers in cloud computing based on [32]](image)

3.2.1 Infrastructure as a Service (IaaS)

Infrastructure as a Service is the lowest layer in the cloud computing stack. As shown in Figure 3.1 IaaS combines two layers: the hardware layer and the infrastructure layer. IaaS users are interested in using hardware resources such as CPU power or disk storage. Instead of directly offering these services to the user, IaaS providers provide users with a virtualization platform. Customers need to install and configure a virtual machine, which runs on the hardware of the cloud provider. In this model, cloud users are responsible for their virtual machine and cloud providers are responsible for the actual hardware. Issues such as data replication and hardware maintenance are addressed by the cloud provider, while the management of the virtual machine is performed by the cloud user.
Benefits of IaaS for cloud users are:

- **Scalable infrastructure**
  The biggest advantage of IaaS is the elasticity of the service. Instead of having to buy servers, software and data centre capabilities, users rent these resources on a pay-per-use manner. In situations where the workload of computer resources fluctuates, IaaS might be a good solution. For example, consider a movie company that uses servers for rendering 3D effects. The company has a small data centre on-premise which is used for the rendering, once a week. The rendering of one scene takes 50 hours when calculated on 1 machine. By scaling up to 50 machines, the rendering of the scene would take 1 hour. Scaling up the internal network of the company might be an expensive operation considering the installation and maintenance of the machines, especially when the servers are only used for rendering once a week. Instead of buying these machines, one might consider to rent the machines and only pay for the rendering operation once a week.

- **Portability**
  Since IaaS works with virtual machine images, porting an on-premise system to the cloud or porting a virtual machine from one cloud to another can be easy. This, however, depends on the virtual machine image format that is used by the cloud provider.

Drawbacks of IaaS for cloud users are:

- **Virtual machine management**
  Although cloud users do not have to manage the rented computer hardware, cloud users are still responsible for the installation and configuration of their virtual machine. A cloud user still needs experts inside its company for the management of these virtual servers.

- **Manual scalability**
  IaaS does not offer automated scalability to applications. Users are able to run virtual machines and might boot several instances of virtual machines in order to scale up to their needs. Collaboration between the virtual machines has to be coordinated and programmed by the cloud user.

Benefits for IaaS cloud providers are:

- **Focus on hardware**
  Cloud providers are mainly focused on hardware related issues. Everything that is software related, such as database management, threading and caching needs to be performed by the cloud user.

- **Exploiting internal structure**
  Several providers are able to offer cloud computing services as an ex-
tension to their core business. For example, the Amazon infrastructure stack was built for hosting Amazon’s services. By offering this infrastructure as a service, Amazon is able to exploit infrastructure and offer a new service to its customers at low cost.

Drawbacks for IaaS cloud providers are:

- **Under- and overprovisioning**
  Cloud providers have to offer their resources as if they are unlimited to the cloud user. This means that a cloud provider needs to own enough resources in order to fulfil the needs of a cloud user. These needs, however, may vary every time. Underprovisioning of a data centre causes that a cloud user might not be able to obtain the resources it asks for, since the cloud provider does not have enough machines available. Overprovisioning is extremely expensive, since servers are bought and maintained, but are not used.

### 3.2.2 Platform as a Service (PaaS)

Platform as a Service is a service model in which users are offered a platform on which they can develop and deploy their applications. The platform offers support for using resources from the underlying infrastructure. Platforms are mostly built for a certain domain, e.g., development of web applications, and are programming language-dependent.

There are several cloud platforms available nowadays. Microsoft offers the Windows Azure platform, which can be used for developing (web) applications and services based on the .NET framework. Google’s App Engine is a platform for the development and deployment of Python or Java-based (web) applications.

Benefits of PaaS for cloud users are:

- **Development platform**
  PaaS offers cloud users a platform on which they can manage and deploy their applications. Instead of having to manage important issues such as scalability, load balancing and data management, cloud users handle these issues by using services offered by the platform.

- **No hardware and server management needed**
  Customers can deploy applications relatively easily on the platform, since no network administrators are necessary for installing and maintaining servers or virtual machines.

Drawbacks of PaaS for cloud users are:
• **Forced to solutions in the cloud**

PaaS offers combinations of services. For example, Windows Azure provides users with a .NET environment. The platform offers support for databases in the form of SQL Azure. Application developers can choose to use a different database, but have to perform difficult operations to install these services on the platform, or have to host the database by a third party. PaaS users are more or less forced to use the solutions that are offered by the cloud provider in order to get the full benefits from the platform.

Benefits for PaaS cloud providers are:

• **Focus on infrastructure and platform**

The software that runs on the platform is managed by the cloud user and the cloud provider is responsible for the infrastructure and the platform.

Drawbacks for PaaS cloud providers are:

• **Platform development**

The platform that is offered by the cloud provider is a piece of software. Complex software is needed to offer services such as automatic scalability and data replication. Faults in the platform can lead to failure of customer applications, so the platform has to be fault tolerant and stable.

### 3.2.3 Software as a Service (SaaS)

With Software as a Service, cloud providers offer an application that is deployed on a cloud platform. Users of the application access the application through the Internet, often using a browser. One of the benefits of SaaS is that cloud providers are able to manage their software from inside their company. Software is not installed on the computers of the cloud users, but instead runs on the servers of the cloud provider. When a fault is detected in the software, this can be easily fixed by repairing the software on the server, instead of having to distribute an update to all the users.

There are several examples of Software as a Service. For example, Google offers several web applications, such as Gmail and Google Docs. Both applications are offered as an online service. Another example is SalesForce.com, which offers CRM online solutions as a service.

Benefits of SaaS for cloud users are:

• **Pay-per-use**

Instead of having to purchase a license for each user of an application,
organizations are billed based on the usage of the software. A couple of years ago software was often purchased on a pay-per-license base. Network administrators had to install applications on the workstations of a cloud user’s company and for each application instance the cloud user had to pay for a license, even if the user of a particular workstation did not use the application. With pay-per-use, cloud users pay only for the users and the usage time of the application.

- **Updates**
  Applications in the cloud are managed by a cloud provider. The cloud provider is able to perform updates to the software directly in the cloud. Instead of having to distribute updates to the cloud user, the users always work with the most actual version since they access the application in the cloud.

Drawbacks of SaaS for cloud users are:

- **Data lock-in**
  Data lock-in is one of the typical problems of SaaS. In case cloud users decide to work with another application, offered by a different provider, it might be hard to move the data to this other application. Not every application provider stores data in a standardized way and interfaces for retrieving all the data from an application may not be available.

Benefits for SaaS cloud providers are:

- **Maintenance**
  Maintenance can be directly performed in the cloud application itself. Updates do not have to be distributed to the cloud users but are directly applied upon the software in the cloud.

Drawbacks for SaaS cloud providers are:

- **Infrastructure needed**
  In traditional software deployment, software is shipped to the user. The hardware on which the application is installed is managed by the user. With cloud computing, the software runs on servers of the cloud provider. This means that cloud providers have to perform infrastructure maintenance, or they have to rent infrastructure or a platform for hosting their applications.

- **Responsibility**
  Applications that run in the cloud are managed by the SaaS provider. When the application in the cloud is not accessible or not working any more because of erroneous updates or changes in the software, cloud users are not able to work with the software any more. It is a big
responsibility for cloud providers to make sure the software is kept up and running.

3.3 Cloud types

The cloud types identified in [23][25] are discussed below.

3.3.1 Public Cloud

A public cloud is provisioned for exclusive use by the general public. Cloud users access the cloud through the Internet. Public clouds are widely available nowadays, for example companies as Microsoft, Google and Amazon offer public cloud computing services. The biggest benefit of public clouds is that the management of the servers is provided by the third-party provider. Users just pay for the usage of the cloud and issues as scalability and replication are handled by the cloud provider.

3.3.2 Private Cloud

Private clouds are for exclusive use of a single organization. Private clouds can be hosted inside or outside the cloud user’s organisation and managed by the cloud user’s organisation itself or by a third-party provider. This form of cloud computing can be used when cloud users have to deal with strict security concerns, where data has to be hosted inside the cloud user’s organization itself.

3.3.3 Hybrid Cloud

Hybrid clouds are created by combining a private and a public cloud. With hybrid clouds, organizations can choose to store their critical data inside the company using a private cloud, while the less critical data and services can be stored in the public cloud. The hybrid cloud approach benefits from the advantages of both public and private clouds. Scalability is maintained, since the public cloud is used for offering the services, while data security is maintained by storing critical data in the private cloud.

3.3.4 Community Cloud

A community cloud is available for a specific community. Several companies that deal with the same concerns may decide to host their services together,
in order to collaborate. Community clouds can be managed by one or more organizations within the community, but the cloud may alternatively be hosted by a third-party provider.

3.4 Cloud providers

Several companies offer cloud services to customers. Below we give an overview of the three most important players in the cloud market and their offerings.

3.4.1 Amazon

Amazon was one of the first companies to offer public cloud solutions to organizations. At first, Amazon focused on offering IaaS solutions, but throughout the years several new services were introduced, which made the Amazon Web Services platform richer and more than just an IaaS solution. Below we briefly discuss some of the these services.

Amazon EC2

Amazon Elastic Compute Cloud (EC2) \[2\] is the computational service of Amazon. EC2 offers computational resources to its users, such as CPU power and memory. In order to use EC2, a customer has to install a custom virtual machine in the form of an Amazon Machine Image (AMI) in an Amazon Virtual Server. Amazon offers pre-installed images, with several additional software such as LAMP and Hadoop \[29\] \[13\]. It is also possible for a customer to upload their own virtual machine image. Virtual machines can be instantiated by the user and the user is free to bootup multiple instances of a virtual machine to enlarge the computational power. New instances can be launched within 2 minutes. The billing of EC2 services is based upon the number of instantiated virtual machines per hour.

Amazon S3

Amazon S3 stands for Simple Storage Service \[4\]. S3 offers a storage service to customers in which they can store objects up to 5 Tb. To organize the storage, data needs to be stored in buckets with a unique name. Each user gets 100 buckets. The number of objects inside a bucket is unlimited, the only restriction within a bucket is that each object needs to have a unique name. Each object that is uploaded to the service is directly replicated
Objects can either be private or public. In this way users can share files over the Internet by making the corresponding object public. Amazon offers several interfaces for accessing objects. Currently supported interfaces are REST, SOAP and BitTorrent.

Amazon SQS

Amazon SQS stands for Amazon Simple Queue Service[3]. The goal of the queue service is to provide a mechanism for exchanging messages between computers. Developers are able to move data from one component to another by adding data to the queue of a component, instead of directly communicating with the component.

Amazon SimpleDB

SimpleDB [5] is a flexible and scalable non-relational data store. SimpleDB offers to developers a datastore in which they can store and query structured data by using webservice requests. Data is stored in domains. A domain is a collection of similar data items, comparable to a relational table, although SimpleDB does not offer relational storage. Inside a domain, data is stored in collections of key-value pairs. A key-value pair can contain multiple values and has a maximum of 1024 bytes per value. An entity in a domain can have at maximum 256 key-value pairs. Data stored inside the datastore is automatically indexed. This leads to fast query performance when querying the datastore.

3.4.2 Microsoft Windows Azure

Microsoft’s solution for cloud computing is called Windows Azure [11][12]. Windows Azure is a public cloud platform that consists of a group of cloud technologies, each providing services to application developers. The platform consists of four parts: Windows Azure, SQL Azure, Windows Azure AppFabric and Windows Azure Marketplace. These parts are briefly explained below. An overview of the platform is shown in Figure 3.2

Windows Azure

The main goal of the Windows Azure part of the Windows Azure platform is to run applications and store data in the cloud. As shown in Figure 3.2 The Windows Azure block consists of 5 parts.
The Compute part presents the computing service of the cloud platform. The service is able to run .NET applications using a Windows Server foundation. Applications that are built on the compute service are structured by using one or more roles. There are 3 possible roles inside the compute service: web roles, worker roles and VM roles. A web role can be used for running web applications. Each web role instance is provided with a preconfigured IIS installation on which the applications can be deployed. Worker roles are intended for processing operations in the background. Often web roles delegate tasks to worker roles in order to execute complex operations. VM roles are the most custom roles. Users need to upload a Windows Server 2008 image as a custom VM role. This solution is often useful when an on-premise service is moved to the cloud.

The Storage part contains services for storing data. Three storage services are provided: blobs, tables and queues. Blobs are binary large objects and are stored in containers. Blobs are useful for storing unstructured data, such as images. In case of structured data, tables can be used. Tables are not to be confused with relational tables, as used in relational database systems, such as SQL Azure. Tables are represented as a set of entities with properties. There is no defined schema for a table. Developers can
query data from tables by using a simple query language defined by OData. Queues are not used for the storage of data, but for passing data from one role to another.

The Fabric controller manages the machines that are available within a data centre. The goal of the controller is to divide the jobs and roles over the available machines and to manage data storage and data replication on these machines.

Connect is an integration service for the integration of on-premise services with the Azure platform. For example, a company wants to move an existing application to the cloud but decides to keep their database system on-premise. Connect helps organizations to deal with these situations.

The Content Delivery Network service is used for improving performance on data access. The service keeps track of often used data and caches it at sites closer to the clients that use it.

SQL Azure

SQL Azure is Microsoft’s cloud-based database solution and is based on the Microsoft SQL server. It does not only consist of a database solution, but it also offers services for reporting and data synchronization. SQL Azure Reporting provides users with data reports, based on the stored data in the database. Created reports can be published in a reporting portal, in which users can access the reports. The Data synchronization service provides a mechanism for synchronization of multiple databases, for both on- and off-premise databases. The synchronization service is based on the Microsoft Sync Framework.

Windows Azure AppFabric

The AppFabric can be seen as a layer above the Windows Azure part. AppFabric adds support for application access, access control and caching.

The first component inside the AppFabric is a service bus. The service bus offers clients a registry in which they can publish their services. Other clients are able to find and use services, by querying the registry.

Often, applications make use of digital identities. Examples of these identities are Active Directory, Facebook, Google accounts or Windows Live ID. AppFabric offers support for several different digital identities in such a way that users can use them relatively easily within their applications.
AppFabric also offers a caching service for caching frequently accessed information. The service is used for reducing the number of database queries, which leads to better performance of the overall platform.

Windows Azure Marketplace

In the marketplace, Microsoft offers existing cloud applications and cloud data. The marketplace has been split up into two parts: DataMarket and AppMarket. The DataMarket offers organizations with datasets. Organizations can browse through the offered sets and purchase the data they need for their applications. The AppMarket can be used for companies to share their cloud applications, so that other organizations can use and integrate them into their own applications.

3.4.3 Google App Engine

Google App Engine [19] is Google’s cloud platform. The platform lets customers run web applications on Google’s infrastructure. The platform currently supports three programming languages: Python, Java and Go.

The App Engine was developed with two goals in mind [16]: (1) to reduce the time for deployment for web application developers. Traditionally, developers do not only spend a lot of time programming their application, but deployment and maintenance of applications also takes its share. By providing users with an already configured infrastructure and several programming API’s, developers can reduce their deployment and maintenance costs. (2) to provide developers with a system that scales automatically to the needs of the application. Users are able to let their application scale from a simple experimental project towards a fully mature software application. Scaling is performed automatically by the App Engine, but within definable boundaries.

Sandboxed applications

Web applications run in a sandboxed environment. Users are provided with limited access to the underlying operating system. In this way Google is able to distribute the web requests across multiple servers. This limited access leads to several restrictions for the developer:

- Applications are only able to communicate with other computers using the provided URL fetch or email services.
External applications and computers can only access the web application via HTTP/HTTPS requests on the standard ports.

Web applications can not write to the file system. Instead Google offers a datastore and a memory cache service.

An application can only execute code during a request. As soon as a request comes in, the application is able to start processing data until the response is sent back to the user. The maximum request-response time is 60 seconds.

Datastore

Google offers developers a distributed data storage service called BigTable [10]. BigTable is not a traditional relational database, since it is column-based instead of row-based. The table itself is schemaless and consists of entities of a certain kind and a set of properties.

The datastore can be queried by using Google’s query language GQL [18], which is an advanced query language optimized for retrieving entities from the datastore.

Services

Google App Engine offers several services that can be used by cloud user’s’ applications, like a mail service for sending mails to customers, an URL Fetch service for accessing resources on the internet or communicating with other services, a memory cache that provides the developer with high performance key-value caching.

Applications in the Google App Engine run only during the request-response cycle. However, this can be circumvented by scheduling tasks using the Cron service that is available on the App Engine platform.

Administration console

In order to manage the web application, Google offers an administration console to developers [17]. The console is an online dashboard in which the user can logon to monitor and configure the applications that are running in the cloud. The console gives the user the opportunity to monitor the performance of the application and also to view and edit the data that is stored in the datastore.
3.5 Conclusion

Three popular cloud platforms have been introduced in this chapter. All three the platforms are public clouds, but the strategies of the platforms differ. The services offered by Amazon are mainly on the IaaS level.

Microsoft’s Windows Azure and Google App Engine are both based on the PaaS service model. Google App Engine focuses on web applications and offers a SDK for creating cloud based web applications. Applications can only execute code during a web request, which has a maximum execution time of 60 seconds. Compute-intensive operations that need more time to execute can only be executed by scheduling them as a task in the Cron service.

Windows Azure is more than a web application platform. In addition to Google App Engine, Azure can also host non-web applications and pure computational applications. Windows Azure has support for custom VM roles, in which cloud users can install and manage their own Windows based VM. By offering this service, Microsoft does offer partial IaaS functionality.

From a database point of view, both Google and Amazon offer schemaless column-based storage to the cloud user, whereas Microsoft offers a relational row-based database solution.

Choosing between the presented cloud providers is difficult and depends on the needs of a user. When a cloud user needs full control over the server settings, an IaaS solution is necessary. When a cloud user, wants to deploy software on top of a platform, a platform can be chosen based upon the programming language and database needs of a cloud user.
Chapter 4

BPM and cloud computing

In this chapter we investigate the state-of-the-art of the combination of BPM software and cloud computing, by discussing relevant scientific literature and some existing commercial products.

4.1 General benefits and drawbacks

Cloud-based BPM gives cloud users the opportunity to use cloud software in a pay-per-use manner, instead of having to make upfront investments on BPM software, hardware and maintenance [22]. Systems scale up and down according to the cloud users needs, which means that the user does not have to worry about over-provisioning or under-provisioning.

BPM in the cloud has also several downsides. By putting a BPMS in the cloud, cloud users might lose control over their sensitive data. The efficiency and effectiveness of activities that are not computation-intensive may not increase by placing these activities in the cloud, but on the contrary, these activities may become more expensive. For example, an activity that is not computation-intensive might need to process a certain amount of data. The transfer of the data to the cloud might take longer than when the activity is executed on-premise, besides, the costs of the activity may increase since data transfer is one of the billing factors of cloud computing.

4.2 Service Models

In [6], moving an existing application from on-premise to each of the service models is considered. In the analysis, the application is supposed to be orchestrated by a BPEL specification and executed by a BPEL execution
4.2.1 IaaS

When an application is moved to the IaaS service model, the cloud user is responsible for the operating system, the middleware and the applications running in the virtual machine, as shown in Figure 4.1. Installing BPM software in an IaaS cloud solution is therefore comparable to installing BPM software on-premise, since everything except the hardware is managed by the cloud user. In addition, the cloud user has to take certain security measures to secure the system from intruders. Possible security measures are blocking ports, enforcing access control policies and keeping the software and operating systems up to date.

![Figure 4.1: Responsibilities for cloud users and cloud providers when a workflow based application is moved to the IaaS service model](image)

4.2.2 PaaS

By placing a workflow-based application on the PaaS service model, the responsibilities for both the cloud provider and cloud user change, as shown in Figure 4.2. The execution engine is assumed to be part of the platform in this case and is offered by the provider. Users need to upload their processes in order to run them in the cloud. The engine can be used by multiple users, since the platform is shared. The responsibility of data storage and data management is no longer in hands of the cloud user, which leads to several security issues.

The following three requirements need to be realized in order to offer a secure BPEL engine on PaaS:

![Figure 4.2: Responsibilities for cloud users and cloud providers when an application is moved to the PaaS service model](image)
Concerning travel booking changed or in case the agency providing hotel information and hotel reservation is no longer available and another provider has to be found.

In order to achieve these requirements, process model descriptions have to be encrypted and signed. Encryption ensures that the process models are not readable for intruders. By signing, one can ensure that a file is only valid for one particular execution engine and that porting to another execution engine leads to failure.

Database storage may also be a problem. Often, BPEL engines deploy processes by reading process model description files and creating relational data in tables. Not only the process model itself, but also the process instance information is stored in the database. The data in these databases has also to be encrypted in order to be unreadable for intruders. The problem of data encryption in databases is that it leads to a restriction of query expressiveness with respect to relational operators. For example, performing join operations might be difficult when values in the database are encrypted.

### 4.2.3 SaaS

By moving the application to the SaaS service model, the cloud provider is now also responsible for the application itself. The application is no longer an asset of the cloud user’s enterprise, but is offered by the cloud provider, as shown in Figure 4.3. The application can be offered to multiple cloud users using a single-tenant or multi-tenant architecture. In a single-tenant architecture a new BPEL engine is installed for each process model. In a multi-tenant architecture, multiple cloud users and process models are
served by one BPEL engine. The data that is stored by the cloud provider should be secured to prevent unintended access by the SaaS provider or by other cloud users. The security measures, as explained in the previous section, can be applied to resolve this issue.

<table>
<thead>
<tr>
<th>Customer</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Provider</th>
<th>Applications</th>
<th>BPEL Processes</th>
<th>Process Models</th>
<th>Process Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Middleware</td>
<td>BPEL Engine</td>
<td>DBMS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hardware</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4.3: Responsibilities for cloud users and cloud providers when an workflow based application is moved to the SaaS service model [10].

In a multi-tenant architecture, multiple cloud users use the same BPEL engine. The data used by one cloud user should not be accessible by other cloud users. As a solution, one can choose to create databases for each cloud user, or to add a column to each database table where the identifier that uniquely identifies the user is stored.

4.3 Combining on-premise and cloud

Privacy protection is one of the barriers for performing BPM in the cloud environment. Not all users want to put their sensitive data in the cloud. Another issue is efficiency. Compute-intensive activities can benefit from the cloud because of the scalability of the cloud. Non-compute-intensive activities, however, do not always benefit from cloud computing. The performance of an activity that is running on-premise might be higher than in the cloud because of data that needs to be transferred to the cloud first in order to perform the activity. These activities can also make cloud computing expensive, since data transfer is one of the billing factors of cloud computing.
4.3.1 Architecture

In most BPM solutions nowadays, the process engine, the activities and process data are placed on the same side, this is either on-premise or the cloud. The authors of [22] investigate the distribution possibilities of BPM in the cloud by introducing a PAD model, in which the process engine, the activities involved in a process and the data involved in a process are separately distributed, as shown in Figure 4.4. In this figure, P stands for the process enactment engine, which is responsible for activating and monitoring all the activities, A stands for activities that need to be performed in a business process, and D stands for the storage of data that is involved in the business process. By making the distinction between the process engine, the activities and the data, cloud users gain the flexibility to place activities that are not computation-intensive and sensitive data at the user-end side and all the other activities and non-sensitive data in the cloud.

The PAD model, introduced in [22], defines four possible distribution patterns. The first pattern is the traditional BPM solution where everything is distributed at the user-end. The second pattern is useful when a user already has a BPM system on the user-end, but the computation-intensive activities are placed in the cloud to increase their performance. The third pattern is useful for users who do not have a BPM system yet, so that a cloud-based BPM system can be utilised in a pay-per-use manner and non-computation-intensive activities and sensitive data can be placed at the user-end. The fourth pattern is the cloud-based BPM pattern in which all elements are placed in the cloud.

Figure 4.4: Patterns for BPM placement [22]
Business processes consist of two types of flows, namely a control-flow and a data-flow. The control-flow regulates the activities that are performed and the sequence of these activities, while the data-flow determines how data is transferred from one activity to another. BPM workflow engines have to deal with both control-flows and data-flows. A data-flow might contain sensitive data, therefore, when a BPM workflow engine is deployed on the cloud, data-flows should be protected.

In the architecture proposed in [22], the cloud side engine only deals with data-flow by using reference IDs instead of the actual data. When an activity needs sensitive data, the transfer of the data to the activity is handled under user surveillance through an encrypted tunnel. Sensitive data is stored at the user-end and non-sensitive data is stored in the cloud. An overview of the architecture proposed in [22] is shown in Figure 4.5.

![Figure 4.5: Architecture of a cloud-based BPM system combined with user-end distribution](image)

**4.3.2 Optimal distribution**

The costs for using cloud computing are investigated in several articles [9, 14]. In [22], formulae are given for calculating the optimal distribution of activities, when activities can be placed in the cloud or on-premise. The calculation takes into account the time costs, monetary costs and privacy risk costs. By using these formulae, cloud users can make an estimation of the costs of deploying parts of their application on-premise and in the cloud.
4.4 Social BPM

Social software supports the interaction of human beings and the production of artefacts by combining the input from independent contributors without predetermining how to do this \cite{28}. An example of social software is Wikipedia, where independent contributors are able to adjust the documents in the system. Business processes can benefit from social software, since users involved in the business processes can easily share their knowledge and information by using social software. Not only business processes, but also BPM itself can benefit from social software. By offering BPM software as social software, multiple users can work on the design, operation and improvement of a business process simultaneously, and collaboration is supported since knowledge is shared through the software.

One of the issues in BPM is the Model-Reality divide issue. Abstract process models and executable process models are often separated. In the design phase, processes are modelled in by business designers. When the models are passed onto the business implementers, the implementers might decide to implement the processes differently than defined by the business designer, simply because not all the details are provided in the design or certain design decisions have not been made. This leads to inconsistency between the designed process models and the executed process models. By involving process implementers and process users in the design process, designers can obtain more information from the implementers and process users, and both parties can react on design decisions that are made by the designers.

Process improvement can also benefit from social software. During the process evaluation phase, not all information inside the process might be taken into account because of the information pass-on threshold \cite{28}. The information pass-on threshold means that users might not share all the information with the designer. This might be the case when a user thinks it is too much effort to document all the information, or when the user thinks that the information is not useful for the designer. Social software can provide the users with interfaces for sharing the information and possibilities for discussing the information with a process designer.

Below we give an overview of possibilities that are provided by social software for each of the BPM lifecycle phases, as introduced in Figure 2.1:

- **Design**
  In the design phase, not only process designers but also process users are able to define requirements and comment on process ideas. By collaborating with process users, process designers get better insight into the process, and users agree with the end result, since they have participated in the design of the process, instead of being forced to
commit to an imposed business process.

- **Implementation**
  In the implementation phase, social software may be used for sharing issues with multiple implementers and process users. Time scheduling of deployment can also be broadcasted using social software.

- **Enactment**
  In the enactment phase, social software can be used for optimizing the communication between process users. Tasks can be spread by using an online dashboard. Information can be passed on through the dashboard, and tasks and comments can be exchanged using mobile devices.

- **Evaluation**
  The evaluation phase can be improved by giving process users access to this phase. Process users can share what they want to change in the process and other users can comment and propose solutions for these issues.

Tools that can be used for enabling BPM as social software are shared modelling tools, blogs, wikis, fora, user rating systems and data exchange services, as proposed in [28].

### 4.5 Environments

On the internet, about 15 BPM cloud providers can be found. The products that are offered by the providers differ from complete BPM suites, to simple modelling tools for business processes. In this section we give an overview of available systems and we briefly discuss some of these products, to give an idea of the variety of products available.

#### 4.5.1 Product overview

Table 4.1 gives an overview of some of the BPM cloud products currently available on the market. For each of the cloud products the cloud service models and the cloud types were identified.

#### 4.5.2 Oracle Fusion

Fusion is a middleware solution offered by Oracle [26]. With Fusion, enterprises can create their own private cloud, based on the PaaS and SaaS
<table>
<thead>
<tr>
<th>Product</th>
<th>Service Model</th>
<th>Cloud Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oracle Fusion</td>
<td>PaaS, SaaS</td>
<td>Private</td>
</tr>
<tr>
<td>Appian BPM Suite</td>
<td>PaaS, SaaS</td>
<td>Private or Public</td>
</tr>
<tr>
<td>Fujitsu InterstageBPM.com</td>
<td>PaaS, SaaS</td>
<td>Private or Public</td>
</tr>
<tr>
<td>Barium Live</td>
<td>SaaS</td>
<td>Public</td>
</tr>
<tr>
<td>Pega Business Process Cloud</td>
<td>SaaS, limited PaaS</td>
<td>Public</td>
</tr>
<tr>
<td>PNM Soft Sequence Cloud BPM</td>
<td>SaaS</td>
<td>Public</td>
</tr>
<tr>
<td>Elite BPM Cloud</td>
<td>IaaS, PaaS, SaaS</td>
<td>Public</td>
</tr>
<tr>
<td>Billfish BPM</td>
<td>Paas, SaaS</td>
<td>Private or Public</td>
</tr>
<tr>
<td>Cloud Harbor BOP</td>
<td>Paas, SaaS</td>
<td>Private or Public</td>
</tr>
<tr>
<td>Cordys</td>
<td>PaaS</td>
<td>Private</td>
</tr>
<tr>
<td>Intalio BPMS</td>
<td>IaaS, PaaS, SaaS</td>
<td>Private</td>
</tr>
<tr>
<td>Adeptia On Demand BPM</td>
<td>SaaS</td>
<td>Public</td>
</tr>
</tbody>
</table>

Table 4.1: Overview of BPM cloud products

service model. The middleware has support for running SOA-based applications. In addition, the middleware offers the Oracle BPEL Process Manager in which business processes can be composed and deployed.

4.5.3 Appian

Appian BPM Suite [7] is a complete BPMS running in the cloud. The suite consists of cloud, mobile and social BPM solutions [8]. Appian offers both on-premise and cloud-based BPM, and gives users the freedom to port their solutions between cloud and on-premise. The suite offers, amongst others, a process engine, content sharing, discussion boards, SOA integration tools and sharepoint.

Appian BPM Suite is considered to be a social BPM, since all process users can participate in the design, enactment and optimization phase of the BPM process. Users can discuss their findings in discussion boards and can share content with each other by using the cloud. In addition, Appian offers mobile applications to users. These applications notify users when a new task comes in, or gives users a reminder when a certain activity failed or needs attention.

Appian offers several security guarantees to their customers: 99.5% uptime, SAS-70 Type II infrastructure audit reports, SAML and LDAP/AD integration for secure authentication, SSL encryption for communication between systems and compliance with national data privacy laws through local hosting.
4.5.4 Fujitsu InterstageBPM.com

InterstageBPM.com is the BPM platform of Fujitsu [15]. The BPM platform is available as a cloud service or as a standalone product which can be deployed on-premise. Cloud users are offered a PaaS and SaaS based cloud solution where all of the phases in the BPM lifecycle are fully supported by the platform. The public variant of the cloud platform is hosted in the datacenter of Fujitsu.

4.5.5 Barium Live

Barium Live is an public SaaS based BPM solution with two tools:

1. An online design tool in which multiple users can collaborate. Multiple users can work simultaneously on the same business process. The design tools works with a versioning system. When a user changes a process, the process can be checked in again and a new version of the process is stored in the cloud. Users can browse through the different versions of the process and can merge different versions of the process.

2. A process enactment tool. Process users can logon to the Barium Live software and get an overview of the tasks they have to perform. Barium does not integrate with information systems, instead it can only be used as a workflow engine for managing human-based tasks.

4.6 Conclusion

In this chapter, we reported on the current state of the art with respect to BPM in the cloud. Reduction of upfront investments and scalability have been identified as the benefits of BPM solutions in the cloud. Security of data is the main issue for not placing sensitive data in the cloud. Placement of activities that are non-computation-intensive in the cloud might also lead to high costs when BPM is placed in the cloud. As a solution, an approach has been presented in which data is securely stored on-premise and only computation-intensive activities are placed in the cloud.

Offering BPM as social software has been identified as a promising approach, in which improvement of communication and collection of knowledge have been identified as its main benefits.

We have introduced four available commercial cloud-based BPM products. In addition, we introduced a table with an overview of existing BPM cloud solutions. Unfortunately, there is not a lot of information about how these
products handle security issues to offer a secure environment to the cloud user.

There are currently no open-source public cloud-based BPM solutions available. There are some companies that offer open-source BPM solutions, such as Intalio BPMS, but these products can only be used as a private cloud solution. Since private clouds are only available within the borders of an organization, less strict security measures are necessary when compared to deployment in a public cloud.
Chapter 5

Research directions

We identify research directions by further discussing the results of the literature study.

5.1 BPEL engine deployment in the cloud

The deployment of a BPEL engine in the cloud according to one of the three service models is discussed in [6]. Several problems can occur when such an engine is moved to the cloud. Their research is mainly based on literature and small experiments. Based on [6], it would be interesting to take an open-source BPEL engine and deploy it onto one of the service models and extend it in a way that each of the identified security issues is solved, presenting a solid solution to cloud users who want to benefit from secure workflow engines in the cloud.

A possible approach might be to investigate the structure of an open-source BPEL engine and identify the modules that need to be updated in order to make it suitable for the cloud. Extensions need to be developed to solve security issues. The next step would be to deploy the system to the cloud and perform advanced security tests to demonstrate that the system is secure.

5.2 Social BPM

Social BPM is one of the trends in BPM development, as mentioned in [28]. By offering BPM as a social solution, not only process designers but also process users can be involved in the whole BPM lifecycle. The benefits of social BPM have been investigated in [28]. Social BPM might influence the BPM lifecycle. For example, collection of data for each of the phases
might be optimized since social BPM offers tools for exchanging information. These influences can be researched in more detail. In addition, concerns such as authorisation, stability, versioning, auditing and security need to be investigated, in order to guarantee consistency.

5.3 Process decomposition

A distribution model in which business process activities can be placed on-premise or in the cloud has been proposed in [22]. Four patterns were identified for the distribution of a process engine, activities and data. However, we can generalise this distribution and identify a fifth pattern, in which process engines, activities and data are deployed on both the cloud and the end-user side. This solution has two potential benefits:

1. A process engine regulates both control flow and data flow. An activity receives data from the process engine and after the execution of the activity, the data that is produced by the activity is passed back to the process engine. Now consider that a sequence of activities is placed in the cloud, while the process engine is deployed at the end-user side. Each activity uses the output data of the previous activity as input data. The data is not directly passed from activity to activity, but is sent back to the process engine first. An example of this situation is shown in Figure 5.1. Since data transfer is one of the billing factors of cloud computing, these situations can become expensive when large amounts of data are transferred from activity to activity. To avoid this problem, a process engine can be added to the cloud, which regulates the control flow and data flow of the activities placed in the cloud. When a sequence of activities is placed in the cloud, data is now regulated by the process engine in the cloud, which reduces the amount of data that needs to be transferred between on-premise and the cloud.

2. When the cloud is unavailable, users can run the business process completely on-premise until the cloud is available again.

In order to run a business process on two process engines, the process has to be split up into two individual processes. It would be convenient for BPM users to take a business process and an activity distribution list, which can be transformed automatically into two individual business processes, one for the cloud and one for the end-user process engine. The communication between both systems can be described by using a choreography description language.

The transformation described above is schematically represented in Figure 5.2. In addition, the distribution list can be created automatically according
Monitoring of the business process is now more complicated, since the business process has been split up in two new business processes. As a solution, a business processes monitor tool can be developed for monitoring the original business process, by combining the monitored details of both the individual processes.

A possible approach to handle process decomposition is to identify the structure and the semantics of the processes. When the control dependencies and data dependencies are known, the consequences of moving certain activities to either the end-user side or the cloud side can be investigated.

When the consequences of distributing activities are known, a model transformation can be created in which a business process and a list with markings is used for creating two individual business processes, one for the cloud and one for the end-user side. In addition, a choreography description can be generated to describe the communication between both the business processes.
Figure 5.2: Schematic representation of separating business processes
Chapter 6

Conclusion

In this report, we investigated combination of BPM and cloud computing. We discussed both cloud computing and BPM and gave an overview of literature that discusses their combination.

BPM has been introduced by identifying the four phases of the BPM lifecycle. In addition, three standard languages for designing and implementing business processes have been briefly discussed. We explained the structure of a BPMS and identified concepts within BPM that are relevant for this work.

We explained cloud computing by giving an overview of the three service models, and the specific benefits and drawbacks of each of these service models. Four cloud types were identified, and products of three cloud providers, namely Amazon, Google and Microsoft, have been introduced.

We discussed the most relevant combinations of BPM and cloud computing found in the literature. The deployment of an orchestrated application onto each of the cloud service models was discussed. The distribution of data and activities within a process were also discussed by looking into four deployment patterns. We also investigated the impacts of offering BPM software as social software and gave an overview of existing cloud products.

We identified three possible research directions. The first direction is the deployment of a workflow engine in the cloud, where several security measures need to be taken to offer a business process platform as a service. The second direction is to investigate the impact of offering BPM as social service. In the third direction, we introduced a new distribution pattern in which two process engines were applied, one in the cloud and one on-premise. A business process can be annotated with a distribution schema and the business process can be transformed automatically to a business process for the cloud-based process engine and the on-premise process engine.
It was quite difficult to find scientific papers about the combination of cloud computing and BPM. We looked into several journals and browsed through proceedings of conferences about cloud and computing and BPM, such as CLOUD, BPM and CLOSER. Papers about the subject are scarce, but several conference websites ask for submission of papers about this subject, which indicates that the subject is still interesting for further research.
Bibliography


