Collaborative Data Management for Business

A review of collaborative techniques

Corina Savulescu
University of Pitesti
Pitesti, Romania
corina_savulescu@yahoo.com

Zdzislaw Polkowski
Jan Wyzkowski University
Polkowice, Poland
z.polkowski@ujw.pl

Nitul Dutta
Computer Engineering
Department, MEF Group of
Institutions,
Rajkot, Gujarat, India
nituldutta@gmail.com

Abstract – In this paper, we conducted a review of techniques and methods used for collaborative data management for business purposes on the WEB. The aim was to propose a document flow optimization and to offer an information technology support for all departments of an organization. In a business organization or a firm, professionals involved in solving specific tasks either use the same data (common information) or the activity of one professional depends on the current status of activity of another business colleague. Therefore, it is necessary to find the quickest and most convenient method to access the shared resources, to facilitate the structural data exchange between members of the same organization, as well as a real time update of shared information. Beyond that, a special attention should be given to information security, involving both storage systems and data access protocols.

Keywords - information technology; security; cloud computing; business on WEB; data management.

I. INTRODUCTION

A successful business activity implies, beyond the intrinsic production effort, an adequate data management. Data management is involved in all business components: technical documents, production planning, contracts, financial documents etc. The information contained in these documents need to be shared between company’s departments with different access rights (such as preview or editable mode). Nowadays almost all documents are stored in an electronic format which allows to be easily retrieved, consulted and updated as needed.

The document management system (DMS) has two major components: storage and communication. Besides these, there are other components that offer supplementary performance and features to a DMS.

The systems for storing and indexing documents and their backups can be provided with automation elements on server (before and after processing). The communication protocols can be provided with special procedures for encryption of data needed to be transferred or exchanged (including login elements and the actual data).

Overall, the use of the most adequate solutions for document management has as global effect an important decrease in the time needed for projects’ implementation and a significant decrease in the associated costs. The aim of this paper is to offer to business companies a review of methods used for the collaborative digital data management; these methods can optimize their activity (by decreasing the response time) and reduce costs (by decreasing human efforts).

II. DMS CLASSIFICATION CRITERIA

The main aspects that should be taken into account while choosing a DMS could be: (a) the method used to store data (shared or cloud), (b) the programming model used to access and update data (local or online), (c) the presence of automate repetitive activities, like extracting metadata from files which allows indexing them, that could be carried out by the files’ administration system, (d) the prerequisite start of an automatically processed work flow which directs documents according to the processing results in a predefined way (like meta data extraction, automatic or human – when automation failed – this triggers events consisting in notifying a designated person about these specific results, (e) the necessity of tracking activity of people involved in developing the work flow (automatic tasks are created), (f) the decrease of downtimes that occur between the moment of achieving results of a task and their delivery as input data to another task – starting already parameterized, subsequent data flow, (g) the necessity of ensuring an adequate level of security which implies a supplementary module that manages the encryption and decryption operations when data is transferred over communication channels between users. [7]

III. DATA EXCHANGE CLASSIFICATION

Taken into account the aspects previously presented, the main decision consists in choosing the most adequate method for data exchange between involved actors. The data exchange method can be classified according to storage place and the type of updating, as follows: (a) by sharing a folder on a system, where each user can update the information and from where the user can extract the needed information. The advantage of this method is the reduced cost of the implementation. However, it has
two major disadvantages: (a.1) no possibility to update a file in the same time by two users unless the file is divided in sub-files, and (a.2) the process running on that system affects the overall performance; (b) by using a third party repository system that manage documents access. In this case the document update can be done in two ways: (b.1) the file is locked on repository and it is updated online or (b.2) the file is locked on the repository, it is locally downloaded, locally updated and uploaded back on the repository.

IV. ONLINE DATA STORAGE AND UPDATE MECHANISM

The data repository implies a third party data system where the access is made through a centralized login mechanism. Here, the users can update the documents in two ways: locally or online: (a) The locally update consists in creation of a local copy of the document which is modified by several users and subsequently synchronized with the server after a definite time interval. The disadvantage here is that a file can be modified in the same time by two users resulting in a conflict at the upload (owncloud); (b) The online update is made directly to repository and consists in a mechanism that manages the document in small pieces which allows that any modification made by a user to be immediately updated on the documents of the other online users that have in preview the changed part of the document. As an example, this system is used by Google drive. Here, when a file is opened, the user can both make modifications and see the real time modifications made by the other users with granted access.

V. ONLINE STORAGE ENHANCEMENTS

The most performant document management systems allow a set of additional tasks, such as: (a) creation of workflows, (b) insertion of plugins for automation, (c) automatic extraction of metadata, (d) association of the automatically extracted or human specified metadata with documents and workflows, (e) start of the new workflows when a set of conditions is met (triggers with tripers), (f) retrieve of documents by searching the metadata, (g) use of dedicated servers for specific processing, (e) capability of distributing the documents along few data servers in order to optimize the communication flow etc. This type of system is also known as cloud system.

VI. CLOUD STORAGE

The cloud storage involves the use of servers from the whole world where files are stored. In few words, by cloud storage we can have access to our data, but those data are not stored in a computer or other device with internal memory. Data stored in a cloud can be accessed by computer, tablet or Smartphone (Fig. 1). For example, it is possible to watch movies, without download them in the internal memory of your device. [1] Table I presents the few cloud storage examples and the facilities offered.

<table>
<thead>
<tr>
<th>Cloud Storage Examples [2]</th>
<th>The storage space free</th>
<th>Facilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google Drive</td>
<td>5GB</td>
<td>- The automatic synchronization of files · The application for Android can scan the documents and save like PDF in Drive</td>
</tr>
<tr>
<td>Dropbox</td>
<td>2GB</td>
<td>- The video streaming system · Can edit the .docx, .xlsx și .pptx files in the web interface</td>
</tr>
<tr>
<td>SkyDrive</td>
<td>25GB</td>
<td>- Friendly and familiar interface (similar with Explorer from Windows 7) · A very good file management</td>
</tr>
<tr>
<td>OneDrive</td>
<td>15GB</td>
<td>- Offer the remote access possibilities for the computers which use the same account</td>
</tr>
</tbody>
</table>

If we compare a cloud storage structure with a conventional server structure, we can note a few advantages for the cloud structure:

1. Scalability - This environment can grow if necessary to accommodate the increased amount of work that can occur globally. The virtual environment dedicated to hardware systems will assure the necessary computing power from an efficient modulation of specific instances. In this way we can be sure that the hardware equipment is used at its maximum efficiency.

Various solutions address each type of resource: [3]
- Content Delivery Network – Using a content delivery mechanism, our application will only access the proxy servers when there is a resource request. For this reason, the load level will be reduced.
- Cache – This solution can be used to avoid making countless requests to different external services or to re-compute different parameters. In this way, when scaling using a cache, we do not need to buy a special device or licenses or to do the synchronization between nodes.
- Video stream – This service, totally discharging our instances for other tasks ensures scalability by default.
- Communication between instances – In general, complex applications need involvement of many computers to run different components of these applications. We can try to implement a communication system via a database or through another instance to deal with only this. When scaling is needed, it would be necessary to think about how we can solve problems such as synchronization of instances.

2. Flexibility - This aspect addresses one of the most important problem of traditional information technologies infrastructures: if a change is needed, this can generate the system collapse because the most important part of resources will be used for the most important tasks. Especially for situations like this, the cloud could be very flexible, and adding a new process is realized with a minimal effort.

3. Self-service - This aspect gives the possibility to obtain computing resources without delay. Hence, the
The principal advantages of Cloud Computing are:

- Easy access to data: Regardless of the location, we can access personal data or company data, through Internet, using any device that is endowed with a web browser.
- Costs: Total Cost of Ownership (TCO) is very attractive because is not necessary to buy computers and servers but to rent them, or access them on Internet.
- Several disadvantages should be mentioned: [6]
  - Long term costs: If this service is rented for a long period of time, the costs could become significant, equivalent with a personal acquisition of the system.
  - Servers overloaded or an insufficient bandwidth can change the performances of Cloud Computing.
- Data security: Transferring data through Internet may pose an important security risk while using Cloud Computing. For this reason, it is compulsory to implement suitable measures to prevent security risks when the Cloud Computing is used, such as:
  - Data encryption
  - Interconnecting servers for periodically backup of the same files using different devices.

VIII. THE ELECTRONIC DOCUMENTS MANAGEMENT ON WORKFLOWS

In a business on Web, but not only, a lot of solutions are useful for the management of electronic documents on workflows: [8]

1. Digitizing documents:
   - Simple scanning of documents - digitization of archive invoices, contracts etc.
   - Transferring the scanned documents in word, excel or another format that can be edited.
   - Using the OCR format for quickly finding the documents using a complex mechanism and criteria for searching.

2. Creating a content management:
   - Create documents using templates when these are available.
   - Generate files content according to metadata values.
   - Versioning and control the access to documents.

3. Quick access and easy processing:
   - Ensure a quick access to the all working documents.
   - Establish an easy way to find the documents using a complex mechanism and criteria for searching.

4. Safe documents’ storage
   - Store documents in a centralized way with an embedded backup mechanism.
   - Regular and automatic backup of documents to make sure that documents or current versions of them will be kept if a local computer fails.
   - Store all important documents and make sure that they can only be accessed by the company.
   - Protect the access to documents using different solutions such as: encryption techniques or electronic signature.

5. Collaborative work on documents
   - Ensure an easy access to the documents and prevent any problems related to overwriting or lose the documents versions.
• Use versioning - A way for visualization to the any version of documents from history. The old version can be restored anytime.
• Share Link – providing the access links to the documents.
• Guarantee traceability - A way for keeping all version documents through processing from the first draft until the final version that is approved and delivered.

IX. A WORKFLOW MANAGEMENT SYSTEM

Working collaboratively within a business company by sharing documents is a current practice, but it requires a radical change of approach when making the transition from workflow based solely on paper to that based on documents in electronic format. [9]
As defined by the Workflow Management Coalition (http://www.wfmc.org/), creating a workflow of documents represents the automatization of business processes in a company through which documents, information and / or tasks follow a route predefined by a set of procedural rules, moving from one worker to another, each of them having defined roles in the interaction with the document (information, tasks etc.).

At the elementary level, a document workflow configures the route that a document must follow, each step having defined the activity, participants, the rules by which they operate and how the document moves from one step to another. Advanced workflow systems are hierarchical systems of workflows including integrated sub-processes.

The workflow should also have implemented automatic operations for notification by email of a designated participant about the current stage of the process or an over deadline. In this way, the time is considerably reduced by reducing the steps where manual intervention is needed, representing a real advantage.

Another advantage of using a workflow is the increase of efficiency because the task will reach the appropriate worker in real time and all activities can be followed in a correct way according the company rules. [8]

X. SECURITY

The information security in cloud computing environment involves a security level for each component: the network security, the server security and the application security. These security requests are according with internal infrastructure and with specific workflow's rules. [10]

The encryption is not a complete solution of security in cloud computing because the data must be decrypted in different situations by the cloud service provider. Even though the client uses data encryption, the data security can be affected.

A good solution can be emails filtering (including back-up and spam), the web content filtering and a good management of vulnerable points.

The identification, access management and related policies should be in accordance with the current policies of the company and compatible with the existing applications. [11]

CONCLUSION

In business on web, but not only, a solution for improving the efficiency of the activity could be represented by an integrated platform in cloud. This offers to companies a business tool that integrate and interconnect specific activities as well as provide a complex module of reporting. The users are informed in real time, regardless their location, about their activities, sometimes using only a smartphone, or a tablet.

Companies have a real advantage while using a business application in cloud because it offers them efficiency, scalability, flexibility which have as ultimate result the increase of productivity.

In a complex economic situation, this solution reduces costs offering the above mentioned advantages without the necessity of making additional investments in software licenses.

In addition, there are efficient solutions for ensuring the security in this systems, limiting errors arising from the introduction of flawed data.

REFERENCES
