

What to consider for business document
Cloud Storage



Introduction

Most ERP, CRM, or BPM generate millions of business documents that must be stored somewhere.

Cloud storage has evolved beyond a mere digital filing cabinet hosting documents for legal reasons or back office convenience. Identifying and implementing an effective archival strategy is integral to advanced business practices that bring significant gains to the modern Enterprise.

If you are considering sharing vendor access to digitally signed purchase orders, the ability to track read/receipt of invoices, or accessing packing-lists from mobile devices, you will need an advanced business document repository -- either based On-Premises or in the Cloud.

All of these processes rely on the ability to secure external access to business documents. Digital Transformation, Mobile and Cloud technologies offer significant opportunities for improvement regarding business documents distribution or consultation.

When facing challenges handling large volumes of documents with limited internal IT resources, today's Enterprise must rely on a scalable document storage architecture seamlessly connected to its core business applications.

In this white paper, we will discuss best practices for implementing a PaaS Cloud Storage model for business documents, regardless of the provider you select.

Table of contents

- Introduction 1
- Table of contents 2
- Capturing Pertinent Metadata for Pinpoint Searches 3
- Resilient Handoffs without Document Loss from Process Disruption 3
- Accelerate Optimized Indexing for Large Volumes of Documents..... 4
- Reduce Document Redundancy with Targeted GUID Links 5
- Managing Multiple Document Types in the Same Repository 5
- Conclusion 6
- About the Author 7
 - Kirk Fackre – Director, Enterprise Solutions..... 7
- About Symtrax 7
- Contact – Locations and offices 7

Capturing Pertinent Metadata for Pinpoint Searches

Whatever Cloud providers you use (e.g., Azure or Amazon), document retrieval can be cumbersome if the search returns more documents than you are expecting. This can result from full text indexing or inappropriate metadata to filter the search. Business documents are easy to classify per category, with relevant references such as customer ID, invoice number or date, etc. For better results, develop a set of metadata to pass to the Cloud application along with the PDF document. Metadata should be defined by reading the document content and selecting the right words, or retrieved from an external resource (such as an email address of the contact point).

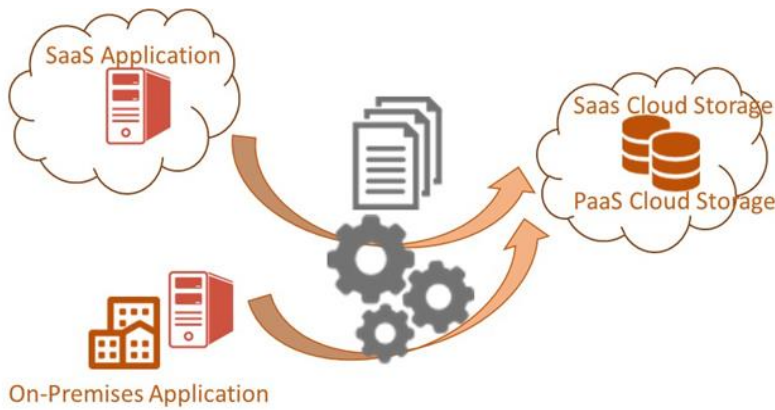
The image shows a PDF invoice from Fly & Smile. A blue box labeled "Identifying metadata within PDF" points to several key pieces of information: the company address (Fly & Smile, 4 Truckee Way, New York, NY 12456-4574), contact details (Our clerk: Herrmannspahn, Telephone: (212) 99-10 99, Telefax: (212) 99-12 99, Customer number: 00000005, Date: 21.06.2017), a QR code for the frequent flyer program, and the recipient's address (Mister Martin, 44 Harolds Road, L07 1AZ Liverpool). Below the invoice text is a table of flight bookings.

Flight No.	Flight Date	From	To	Class	Price
American Airlines Bookings for American Airlines					
AA0064	28.08.2009	SAN FRANCISCO	NEW YORK	Y	422.94 USD
Alitalia Bookings for Alitalia					
AZ0788	22.09.2009	ROME	TOKYO	Y	944.95 USD

Feeding a document cloud storage for easy retrieval requires the ability to identify and pass relevant metadata for each related document.

Resilient Handoffs without Document Loss from Process Disruption

Any process that sends documents from one system to another -- whether on premises or in the Cloud -- requires synchronization and acknowledgments. A robust process should be able to recover from a network failure or processing delays, thus ensuring each document is saved and easily retrievable. Since creating business in an ongoing process, the archiving document processing must operate smoothly without human intervention.



Document processing should rely on a resilient automation system, well designed communication, exception handling, and asynchronous processing.

Once in the Cloud hosted repository, PaaS infrastructure and operations preserve stored documents with high availability architecture and automatic back-up.

Accelerate Optimized Indexing for Large Volumes of Documents

As an example, the maximum file size of SQL Server 2012 is 16 terabytes. Assuming an average PDF document size of 500KB, the capacity would be 320 million documents for just a single database. Further, the number of documents is not the major factor in estimating hosting capabilities. Building indexes and searching through indexes is a time consuming process. This is why full text indexing may give some apparent comfort, but can overwhelm the storage system with huge index tables. Indexing documents with pertinent but limited metadata, or full text indexing confined to a text zone within the document (such as first page or specific paragraph) will ease performance and speed-up research.

<i>Index (one for each word)</i>	<i>Pointer</i>
This	as many references as occurrences, address #1, address #2.... address #n
is	as many references as occurrences, address #1, address #2.... address #m
not	as many references as occurrences, address #1, address #2.... address #p
efficient	as many references as occurrences, address #1, address #2.... address #q



→

<i>index is "Invoice number"</i>	<i>Pointer</i>
INV2017A45DE5147	invoice address
INV2017A45DE5148	invoice address
INV2017A45DE5149	invoice address
INV2017A45DE5150	invoice address

Reduce Document Redundancy with Targeted GUID Links

Once a business document is saved to a PaaS Cloud platform, should the Enterprise keep a local copy? Would your ERP Application be able to generate it again if needed? If the document is to be distributed is there a copy available somewhere? Is the document meant to be accessed internally only?

All these variables can be addressed by associating a unique URL (GUID) referencing the saved document that can be returned to the Enterprise Application. In this case, each business document is accessed by clicking the GUID. This can address customer-facing use-cases (such as invoices, sales orders or POs) as well as internal requirements, allowing end users may locate the relevant document easily.



<http://archivor.symtrax.eu/cmparch/DocumentHandler/ViewPublishDoc.aspx?PubId=6cfdeeca-5592-4d92-a3f4-df5bda06246c>



This feature allows Enterprise Application, ERP CRM, BPM, to keep straightforward reference and access to all business documents they may generate. Handling business documents outside the Applications that created them has multiple benefits especially when using a SaaS Application.

Managing Multiple Document Types in the Same Repository

Is there a common point between an XML invoice, a PDF invoice or a JPEG invoice? Can they be stored in the same repository? As long as metadata are associated with the business document any format is able to share the same Cloud storage space whether Microsoft Azure, Amazon Web Services AWS, etc.



Conclusion

Business Document storage was originally designed as an on-premises application. Given recent technological advances, including but not limited to hybrid Cloud solutions, tremendous efficiencies are now possible. Enterprises that are hosting their Systems and Applications in the Cloud -- whether in IaaS or PaaS mode, or accessing in SaaS mode -- are increasingly hosting their Business Documents in a separated solution for flexibility and platform independence. Business Documents are no longer held hostage by the Application platform that created them, allowing for rapid changes as the business requires it.

In addition of multiple requirements to select a provider, such as broadband, storage volume, concurrent accesses, number of documents, service performance and availability, all previous considerations will directly affect the success of your project and are integral to a successful implementation of a PaaS Cloud Storage.

Want to know more about how to implement a PaaS Cloud Storage for your business documents? Click here to contact us for a one-on-one demo :

<http://www.symtrax.com/who-we-are/request-a-live-demonstration/>

About the Author

Kirk Fackre – Director, Enterprise Solutions

Throughout much of his career, Mr. Fackre has concentrated on how technology helps organizations become more productive by capturing, organizing and sharing critical knowledge, information and expertise.

Prior to joining Symtrax, Mr. Fackre cofounded Airtime, a Mobile SaaS Solution provider and ResearchAgent Corporation where he pioneered a product line used to manage web content.

Upon graduating from the University of Massachusetts at Amherst, Mr. Fackre joined Equitrac Corporation, where he enjoyed a long and successful career in a variety of progressively responsible leadership and sales management positions.

About Symtrax

Symtrax is a worldwide software company specialized in Data and Document Management solutions. Our business intelligence software provides companies with the ability to analyze business data efficiently. Our document management software values & delivers business documents electronically.

Contact – Locations and offices

Symtrax USA

sales@symtrax.com

West Coast

Tel: +1 (310) 216 9536

East Coast

Tel: +1 (617) 391 5800

Symtrax Europe

sales@symtrax.eu

EN: +33 (0)4 66 04 54 24

SP: +33 (0)4 66 04 54 35

Symtrax France

contact@symtrax.fr

Tel: + 33 (0)4 66 04 54 05

Symtrax UK

sales@symtrax.co.uk

Tel: +44 (0)207 533 6616

Symtrax APAC

salesapac@symtrax.com

Tel: +91 22 40 76 55 55

Symtrax Deutschland
verkauf@symtrax.de
Tel: +49-6227-381-155

Symtrax Latin America
ventas@symtrax.com
Tel: +1 (310) 216 9536

All Rights Reserved.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publisher, Symtrax Corporation. Whilst every care is taken to ensure the accuracy of the information contained in this material, the facts, estimates and opinions stated are based on information and sources which, while we believe them to be reliable, are not guaranteed. In particular, it should not be relied upon as the sole source of reference in relation to the subject matter. No liability can be accepted by Symtrax Corporation, its directors or employees for any loss occasioned to any person or entity acting or failing to act as a result of anything contained in or omitted from the content of this material, or our conclusions as stated. The findings are Symtrax's current opinions; they are subject to change without notice. Symtrax has no obligation to update or amend the research or to let anyone know if our opinions change materially.

© Symtrax. Unauthorized reproduction prohibited.

This report is a licensed product and is not to be reproduced without prior permission.