



WHITEPAPER

3-Heights™ Scan to PDF Server Grundlagen und Anwendung



Die 3-Heights™ Produktfamilie der
PDF Tools AG steht für:

High Quality – High Volume – High Performance

Copyright ©2016 PDF Tools AG. Alle Rechte vorbehalten..

Namen und Marken Dritter gelten als rechtlich geschütztes Eigentum. Die Rechte können jederzeit geltend gemacht werden. Die Darstellung von Produkten und Dienstleistungen Dritter dienen ausschliesslich zu Informationszwecken.

PDF Tool AG ist für die Performance und den Support von Produkten von Drittfirmen nicht verantwortlich und übernimmt keine Gewähr bezüglich Qualität, Zuverlässigkeit, Funktionalität und Kompatibilität dieser Produkte und Geräte. .

PDF Tools AG

Kasernenstrasse 1, 8184 Bachenbülach, Schweiz

Tel.: +41 43 411 44 51 , www.pdf-tools.com

Inhaltsverzeichnis

Inhaltsverzeichnis	3
Funktion und Nutzen	4
Warum braucht es mehr als nur einen Scanner?	4
Was ist die Aufgabe eines zentralen Scan Server?	5
Wo wird der Dienst eingesetzt?	7
Was sind die Vorteile eines zentralen Dienstes?	8
Welchen Zusatznutzen bietet der Dienst?	8
Architektur und Leistungsmerkmale	10
Kompression und Bildqualität	10
Texterkennung	10
ISO-Konformität und Konformitätsprüfung	11
Verteilte Architektur und Skalierbarkeit	11
Leistung	13
Schnittstellen zur Applikationsintegration	13
Erweiterbarkeit durch Zusatzfunktionen	14
Übersicht über die implementierten Leistungsmerkmale	15
Über PDF Tools AG	16

Funktion und Nutzen

Warum braucht es mehr als nur einen Scanner?

Das Scannen von Papierdokumenten im Posteingangsbereich einer Unternehmung ist zum Alltag geworden. Je nach Art und Menge der anfallenden Papierdokumente werden dafür Multifunktionsgeräte (MFP) oder Hochleistungsscanner eingesetzt. In den meisten Fällen werden die gescannten Bilder als TIFF-Dateien in Schwarz und Weiss erzeugt, so wie man dies von den FAX-Maschinen gewohnt ist. In speziellen Anwendungen wie dem Scannen von Checks, Fotos für Ausweise usw. wird die Datei in Farbe erzeugt. Allerdings verzichtet man oft auf das Scannen in Farbe, weil die resultierenden TIFF-Dateien entweder zu gross sind oder die verwendete JPEG-Kompression die Bildqualität sichtbar reduziert. Eine gute Bildqualität ist jedoch eine wichtige Voraussetzung für eine gute Texterkennungsrate. Für eine hohe Kompressionsrate bei gleichzeitig guter Bildqualität benötigt man Rechenleistung, die in den dezentralen Multifunktionsgeräten selbst oft nicht zur Verfügung steht. Für diesen Aspekt bietet eine separate Scan-Software entscheidende Vorteile. Der PDF/A-Standard hat sich heute in Posteingangs-Anwendungen weitgehend durchgesetzt. Die wichtigsten Vorteile des PDF/A-Standards gegenüber den klassischen Dokumentenformaten wie TIFF und JPEG sind:

- **Einheitliches Format:** PDF/A ist für die Speicherung sowohl von gescannten als auch von digital erzeugten Dokumenten gleichermaßen geeignet.
- **Hohe Kompressionsrate:** Der PDF/A-Standard unterstützt modernere und leistungsfähige Kompressionsverfahren und somit auch kleine Dateigrößen für Farbbilder.
- **Texterkennung:** Die erzeugten PDF/A-Dokumente können durch das Einbetten von Texten aus einer OCR-Maschine durchsuchbar gemacht werden.
- **Eingebettete Metadaten:** Damit das Dokument und die dazugehörigen Metadaten eine unteilbare Einheit bilden, werden in PDF/A die Metadaten in die Datei eingebettet. PDF/A verwendet für die Speicherung das Extensible Metadata Platform (XMP) Format, welches unabhängig von PDF/A als eigener ISO Standard definiert ist.
- **Digitale Signatur:** Um die Integrität und Authentizität der erzeugten Dokumente zu gewährleisten, kann optional eine digitale Signatur nach dem PAdES-Standard auf das PDF/A-Dokument aufgebracht werden. Die digitale Signatur ist eine Form der elektronischen Signatur, welche dem Erfordernis der handschriftlichen Unterschrift gleich gerecht werden kann, wie die handschriftliche Unterschrift selbst, sofern die gesetzlichen Voraussetzungen (nationale Signaturgesetze) dafür erfüllt sind.

Alle diese Vorteile lassen sich mit TIFF-Dokumenten grundsätzlich auch realisieren, jedoch nur als proprietäre Erweiterungen, da der TIFF-Standard selbst dafür keine Lösungen bereithält.

Anforderung	TIFF	PDF/A
Langfristige Lesbarkeit	+	+
Eindeutige Wiedergabe	+	+
Datenkonsistenz	Proprietäre Tags für Metadaten	+
Authentizität /Integrität	Mit abgesetzten Signaturen	+
Speicherplatzbedarf	Schwarz/Weiss: + Farbe: -	+
Durchsuchbarkeit	Proprietäre Tags für OCR Text	+
Langzeiterfahrung	+	+

Abbildung 1: Die Vorteile von PDF/A gegenüber TIFF

Die einzelnen Bearbeitungsschritte wie Texterkennung, Kompression, PDF/A-Erzeugung und digitale Signatur können in der Regel nicht durch den Scanner alleine ausgeführt werden, weil oft nachträglich Metadaten von einer Index-Station hinzugefügt werden. Dieser Arbeitsschritt bricht jedoch das Siegel der digitalen Signatur und macht sie wertlos. Auch für diesen Aspekt bietet eine separate Software einen entscheidenden Vorteil.

Was ist die Aufgabe eines zentralen Scan Server?

Der 3-Heights™ Scan to PDF Server ist ein zentraler Dienst, welcher in einer Unternehmung dezentral gescannte Dateien und dazugehörige Indexdaten in das standardisierte Dateiformat PDF/A umwandelt. Dafür übernimmt der Dienst alle Aufgaben, welche von den dezentralen Scan-Stationen an ihn delegiert werden können. Besonders geeignet sind Verarbeitungsschritte, welche keine Benutzerinteraktion erfordern oder die Effizienz der lokalen Scan-Stationen mit leistungsintensiven Funktionen (OCR, Kompression) belasten.

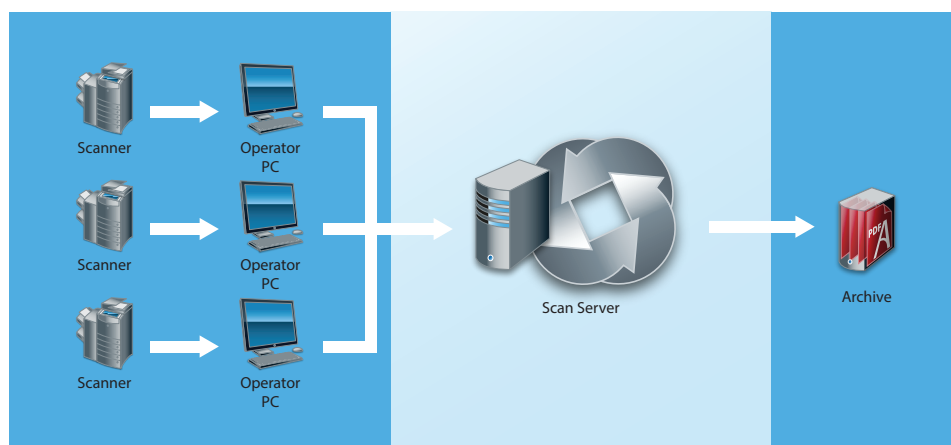


Abbildung 2: Ein zentraler Dienst für die Erzeugung von PDF/A Dateien aus gescannten Dokumenten

Die Hauptfunktionen des Dienstes sind:

- **Text- und Barcodeerkennung:** Gescannte Bilddateien sollen durchsuchbar gemacht werden. Der Dienst kann den Texterkennungsdienst 3-Heights™ OCR Service nutzen, um den Text in der Bilddatei zu erkennen und in die umgewandelte Datei so einzubetten, dass diese durchsuchbar wird. Die erkannten Barcodes können mehrfach verwendet werden, in der Textsuche, als Teil der eingebetteten Metadaten und zur Steuerung der Verarbeitung (Name der Ausgabedatei, Seitentrennung usw.) im Dienst.
- **Kompression:** Farbbilder werden in mehrere Ebenen zerlegt und durch das Mixed Raster Content (MRC) Verfahren stark und ohne sichtbare Verluste komprimiert.
- **Einbettung von Metadaten:** Der PDF/A Standard sieht vor, dass Metadaten in Form von XMP-Paketen in das Dokument eingebettet werden. Der Dienst bietet diese Funktion an.
- **Erzeugen einer PDF/A-Datei:** Der Dienst erzeugt ein- oder mehrseitige Ausgabedokumente entsprechend der ISO-19005 Standardreihe. Alle zurzeit veröffentlichten Normteile PDF/A-1, PDF/A-2 und PDF/A-3 werden unterstützt.
- **Digitale Signatur:** Die Signatur kann fortgeschritten oder qualifiziert, für die Langzeitaufbewahrung oder nur für den Austausch geeignet sein und wahlweise einen Zeitstempel enthalten. Anstelle der persönlichen Signatur kann auch nur ein Zeitstempel aufgebracht werden. Der Dienst kann eine kryptografische Infrastruktur (USB Token, HSM) über eine Standardschnittstelle (PKCS#11) nutzen, um digitale Signaturen zu erzeugen.

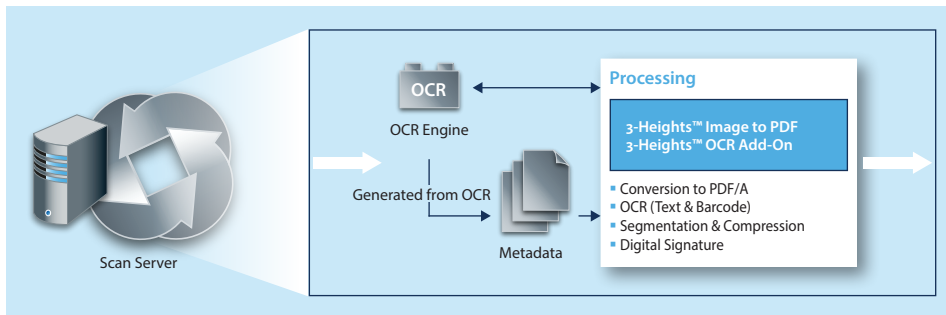


Abbildung 3: Die Hauptfunktionen des 3-Heights™ Scan to PDF Server

Ein typischer Ablauf sieht wie folgt aus:

- **Bildakquisition:** Der Scan-Operator startet den Scanvorgang und erzeugt eine TIFF-Datei in Farbe. Der Scanner legt Dateien in der Regel in einem Dateordner ab. Faksimile-Dokumente werden von der FAX-Maschine empfangen und als TIFF-Dateien in Schwarz und Weiss in einen speziellen Ordner abgelegt.
- **Manuelle Klassifikation:** Wahlweise und je nach Prozess führt der Scan-Operator eine Klassifikation durch. Er steuert den Scanner dabei so, dass die Bilder in verschiedenen Dateordnern abgelegt (z. B. für Rechnungen oder Lieferscheine) oder spezielle Barcode-Blätter eingefügt werden, welche der Trennung und Klassifikation der Dokumente dienen oder er erfasst einen minimalen Satz von Indexdaten.
- **Segmentierung und Kompression:** Das Farbbild jeder Seite wird in seine Bestandteile wie Hintergrund, Text und Bilder zerlegt. Die einzelnen Teile werden durch spezifisch dafür entworfene Kompressionsverfahren in der Grösse reduziert. Dieses MRC-Verfahren ermöglicht Farbdokumenten, konkurrenzfähige Dateigrößen zu erreichen.
- **Text- und Barcodeerkennung:** Die Bilder werden durch eine OCR-Maschine weiterverarbeitet. Als Erstes wird das Bild entfleckt und gerade gerichtet, danach erfolgt die Erkennung der Texte und der Barcodes.
- **Metadaten:** Informationen aus der manuellen Klassifizierung, der erkannten Barcodes und weiteren Quellen werden zu standardisierten XMP-Metadaten zusammengefügt.
- **PDF/A-Erzeugung:** Die aufbereiteten Bilder jeder Seite, der erkannte Text und die Metadaten werden zusammen mit dem ICC-Farbprofil des Scanners zu einem PDF/A-Dokument zusammengefügt. Optional kann eine Index-Datei erzeugt werden, welche nur die Metadaten enthält.
- **Digitale Signatur:** Wahlweise kann eine digitale Signatur aufgebracht werden, damit die Nachvollziehbarkeit und Revisionsfestigkeit des Dokuments sichergestellt ist.
- **Validierung:** Wahlweise können die PDF/A-Konformität des erstellten Dokumentes und die Gültigkeit der digitalen Signatur überprüft werden.

Angaben für aktuellen Batch

Batch Nummer

Klassifizierung

Mandant

Archiv Ablagefach

Der Dienst bietet auch eine Reihe von Zusatzfunktionen an (siehe weiter unten).

Wo wird der Dienst eingesetzt?

Der 3-Heights™ Scan to PDF Server wird für die folgenden Anwendungen eingesetzt:

- **Paper Capture:** Elektronische Archivierung von Papier-Dokumenten, welche im Posteingang einer Unternehmung anfallen.
- **Facsimile Capture:** Elektronische Archivierung des gesamten FAX-Verkehrs zwischen der Unternehmung und seinen Geschäftspartnern.
- **Archive Migration:** Migration von Papier-Archiven in ein elektronisches Archiv mit dem standardisierten PDF/A Format.
- **Web / Mobile Capture:** Nutzung des zentralen Dienstes in Client / Server-Anwendungen über einen Webdienst.
- **Enterprise Application Integration:** Nutzung des zentralen Dienstes für die PDF/A-Dokumentenerzeugung über eine Programmierschnittstelle (API) aus Fachapplikationen heraus, welche TIFF- oder JPEG-Dateien erzeugen.

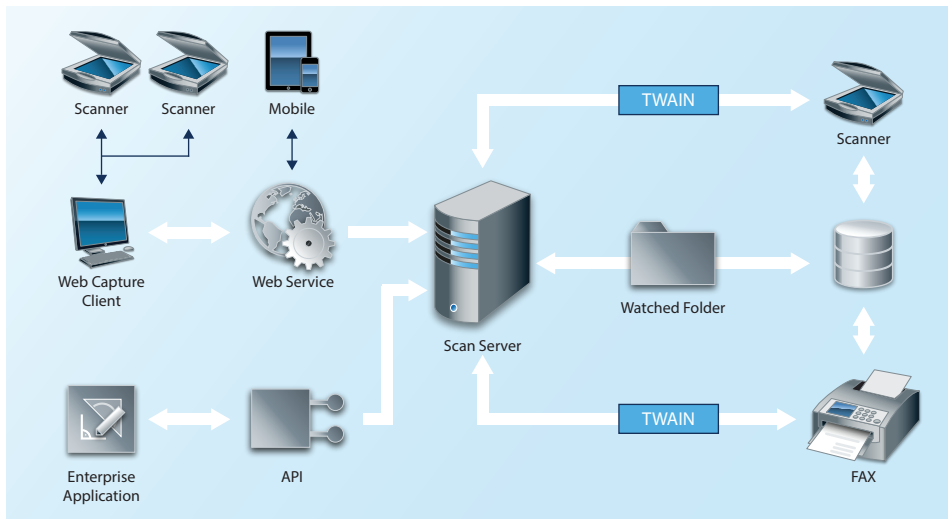


Abbildung 4: Die Anwendungen des 3-Heights™ Scan to PDF Server

Was sind die Vorteile eines zentralen Dienstes?

Der 3-Heights™ Scan to PDF Server kann sich bereits bei einer kleinen Anzahl von Multifunktionsgeräten lohnen, auf jeden Fall beim Einsatz von Hochleistungsscannern. Über die wirtschaftlichen Vorteile hinaus gibt es qualitative Vorteile:

	Client	Server
Leistungskalierung	Die Scan-Leistung ist durch die Leistung der Arbeitsstation begrenzt	Rechenintensive Funktionen (Kompression, OCR) lassen sich an den Server delegieren. Die Leistung lässt sich durch Lastverteilung beliebig skalieren
Qualität	Die Qualität insbesondere der Kompression und der Texterkennung sind durch die Leistungsmerkmale der Scan-Station begrenzt	Die Qualität der Kompression und der Texterkennung ist durch die skalierbare Leistung optimal
Installation und Wartung	Die alle Funktionen umfassende Scan-Software muss auf alle Clients ausgerollt und durch den Benutzer konfiguriert werden	Der grösste Teil der Software kann zentral auf dem Server installiert, konfiguriert und gewartet werden. Auf dem Client muss nur die Operator-Software installiert werden
Unterstützung des Benutzers	Probleme lassen sich durch Hotline und Fernwartung auf der Arbeitsstation beheben	Probleme können auf der Testinfrastruktur nachvollzogen und im produktiven Betrieb nachhaltig gelöst werden
Unterschiede in der Version und Konfiguration	Die installierte Version und deren Konfiguration kann sich von Arbeitsstation zu Arbeitsstation unterscheiden	Die zentrale Installation und Konfiguration der Software garantiert eine einheitliche und gleichbleibende Qualität der erzeugten Dokumente

Abbildung 5: Die Vorteile einer zentralen Scan-Software

Welchen Zusatznutzen bietet der Dienst?

Die Hauptfunktion des 3-Heights™ Scan to PDF Server ist die Umwandlung von gescannten Dokumenten in ein einheitliches und standardisiertes Dateiformat wie PDF/A.

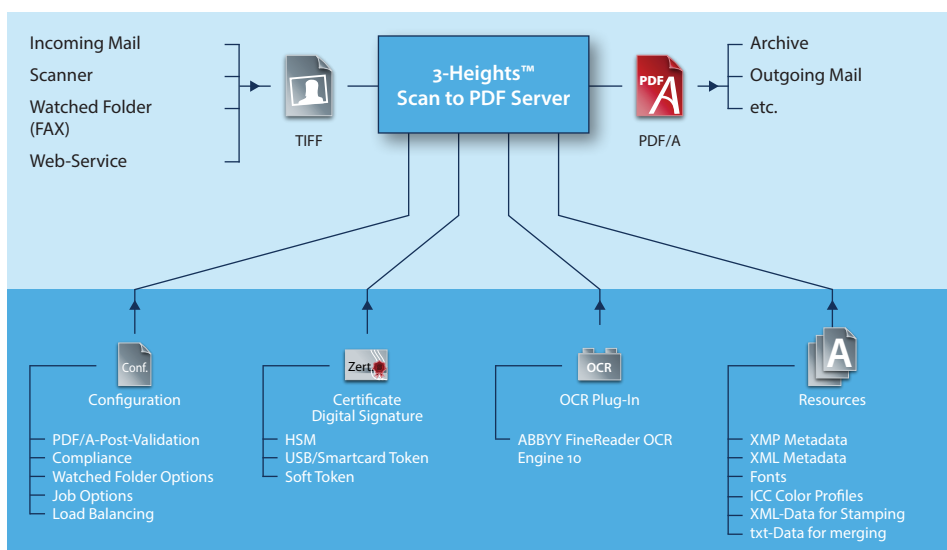


Abbildung 6: Haupt- und Zusatzfunktionen des 3-Heights™ Scan to PDF Server

Der Dienst bietet folgende Zusatzfunktionen an:

- **Einbettung der XML-Daten:** Wenn die TIFF-Dateien aus Fachapplikationen erzeugt werden, dann ist es oft erwünscht, dass XML-Rechnungsdaten eingebettet werden, beispielsweise nach dem ZUGFeRD-Standard. Dazu werden die Möglichkeiten von PDF/A-3 genutzt.
- **PDF/A Validierung:** Zur Sicherung der Qualität bei der Erzeugung von PDF/A Dateien kann wahlweise eine Prüfsoftware eingesetzt werden, um die Konformität mit dem ISO-Standard zu gewährleisten.
- **Dokumente abmischen:** Einseitige Bilder sollen in mehrseitige Dateien abgemischt werden. Oder, Dokumente, welche zu demselben Geschäftsfall gehören, sollen zu einer einzigen Datei oder Dateisammlung, welche beispielsweise einer Akte entspricht, zusammengefügt werden. Für diese Funktion kann der Dienst Textdateien lesen, welche das Abmischen steuern.
- **Stempeln:** Nach Bedarf können die erzeugten Dokumente mit Stempel oder Wasserzeichen versehen werden. Der Dienst verarbeitet dafür eine XML-Datei, welche die Stempel-Daten enthält.

Zusätzliche Funktionen können durch Erweiterung in den Dienst integriert werden (siehe weiter unten).

Architektur und Leistungsmerkmale

Die Leistungsmerkmale des 3-Heights™ Scan to PDF Server sind das Ergebnis der vom Entwicklerteam gewählten Systemarchitektur. Die Ausgangslage für den Entwicklungsprozess bildeten die folgenden Anforderungen:

- Hohe Qualität, insbesondere Konformität mit ISO-Standards und Abbildungstreue bei der Umwandlung der Dokumente
- Robuster und interventionsarmer Betrieb
- Hoher Durchsatz
- Skalierbarkeit der Leistung
- Schnittstellen für die Anwendungsintegration
- Erweiterbarkeit für Zusatzfunktionen

Kompression und Bildqualität

Für die Umwandlung der Rasterbildformate TIFF und JPEG in PDF/A verwendet der 3-Heights™ Scan to PDF Server eingebaute Programme aus der 3-Heights™ TIFF Tool Suite. Diese Sammlung von Bildbearbeitungsprogrammen führt auch die MRC-Kompression durch. Die in den Programmen verwendeten Bildsegmentierungs- und Bildkompressionsverfahren sorgen einerseits für eine hohe Kompressionsrate und andererseits für eine hohe Bildqualität.

Die Bildsegmentierung zerlegt die Farbbilder in Hintergrundbild (z. B. die Farbe des Papiers), in Bildmaske (z. B. Text und Linien von Tabellen) und einem Vordergrundbild, welches die Füllfarbe der Bildmaske enthält. Die Grösse der zerlegten Bilder wird mit unterschiedlichen Kompressionsverfahren reduziert. Das Hinter- und Vordergrundbild kann beispielsweise mit JPEG2000 komprimiert werden und die Bildmaske mit JBIG2. Die Kompressionsverfahren und das Verhältnis zwischen Kompressionsrate und Bildqualität können für den Dienst konfiguriert und den Anforderungen aus der Anwendung angepasst werden.

Die Bildsegmentierung kann durch den Einsatz einer OCR-Maschine erheblich verbessert werden. Die OCR-Maschine liefert die Koordinaten zu Bildbereichen wie Fotos, Text und übrige Inhalte. Diese Koordinaten werden vom Segmentierungsalgorithmus genutzt, um beispielsweise Fotos zu isolieren und getrennt zu komprimieren.

Texterkennung

Für die Texterkennung verwendet der 3-Heights™ Scan to PDF Server den 3-Heights™ OCR Server und das entsprechende OCR-Plugin für die Programme der 3-Heights™ TIFF Tool Suite. Der OCR-Server selbst verwendet die ABBYY FineReader Engine. Diese Engine enthält auch Module für Handschriftenerkennung (ICR/IWR) als Option.

Diese Architektur erlaubt die Verteilung der Kompressions- und Konversionsfunktion einerseits und der OCR-Funktion andererseits auf verschiedene Server und ermöglicht auf diese Weise eine flexiblere Verteilung der Last.

ISO-Konformität und Konformitätsprüfung

Zur Überprüfung der Konformität der erzeugten PDF/A Dateien mit dem ISO-Standard, kann wahlweise die Zusatzfunktion für die PDF/A Validierung eingesetzt werden. Dafür verwendet der 3-Heights™ Scan to PDF Server das eingebaute Programm aus dem 3-Heights™ PDF Validator. Das Programm prüft folgendes:

- Korrekter physischer Aufbau der Datei (Syntax)
- Korrekter logischer Aufbau des Dokuments (Semantik)
- Prüfung der Gebote und Verbote für PDF/A
 - Zugänglichkeit (keine Verschlüsselung)
 - Eindeutigkeit (kalibrierte Farben, keine unsichtbaren, dynamischen oder alternativen Inhalte)
 - Selbstbezug und Konsistenz (eingebettete Schriften und Farbprofile, keine Bezüge auf externe Inhalte)
 - Metadaten (XMP Standard und eingebettete Schemas für Erweiterungen)
 - Durchsuchbarkeit der Texte (nur bei Konformitätsgrad a und u)
- Kundenspezifische Prüfungen
 - Auflösung gescannter Dateien
 - Verwendete Kompressionsverfahren
 - Vorhandensein von OCR-Text
 - Vorhandensein erforderlicher Metadaten
 - Corporate Identity Farben und Schriften (nur digital erzeugte Dokumente)

Die Konformitätsprüfung ist oft eine notwendige Voraussetzung für die Anbindung des Dienstes an ein Archivsystem, welches nicht über diese Prüffunktion verfügt.

Verteilte Architektur und Skalierbarkeit

Der 3-Heights™ Scan to PDF Server ist ein skalierbarer und weitgehend konfigurierbarer Dienst. Der Dienst ruft für jeden Arbeitsschritt wie Kompression, OCR-Erkennung, Umwandlung in PDF/A usw. ein separates Programm auf. Dieses erhält als Eingabe das Resultat des vorangegangenen Arbeitsschritts und stellt die Ausgabe für den nächsten Arbeitsschritt bereit. Die Verknüpfung der Arbeitsschritte erfolgt über eine Konfigurationsdatei in XML. Diese Architektur erlaubt eine sehr flexible Gestaltung der Arbeitsprozesse des Dienstes und fast beliebige Erweiterungsmöglichkeiten (siehe weiter unten), indem zusätzliche Arbeitsschritte hinzugefügt werden.

Um den Grad der parallelen Verarbeitung zu steigern, können die Dokumente in einzelne Seiten zerlegt und parallel durch die Verarbeitungsstufen geschickt werden, an deren Ende sie wieder zu einem einzigen Dokument zusammengefügt werden. Diese Möglichkeit ermöglicht eine stark verbesserte Auslastung der Computerressourcen (Prozessorkerne, Speicher, Ein- und Ausgabe, OCR-Maschine usw.).

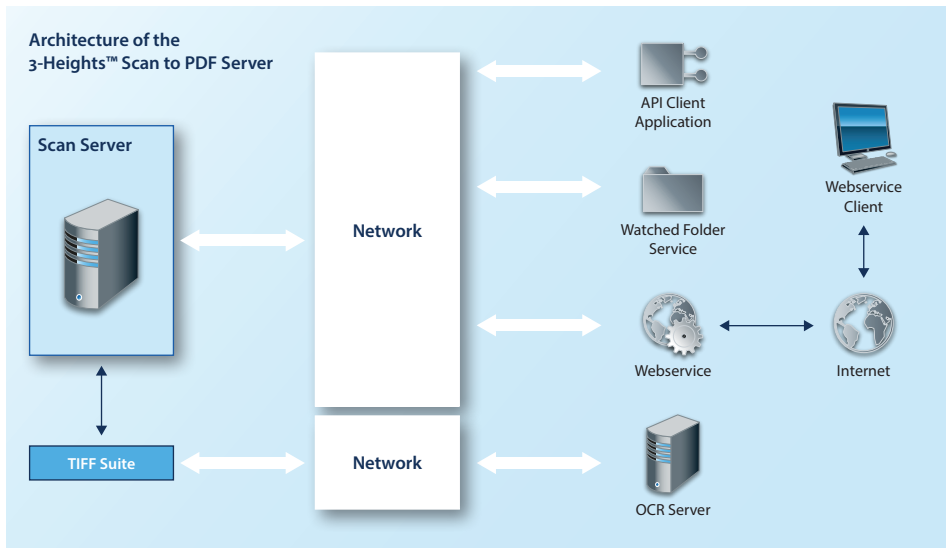


Abbildung 7: Verteilte Architektur des 3-Heights™ Scan to PDF Server

Die Aufgaben des Dienstes werden auf mehrere Subsysteme verteilt. Die wichtigsten Subsysteme sind:

- **Scan Server:** Der Scan Server hat die folgenden Aufgaben:
 - Die Konversionsaufträge werden entgegengenommen, in Arbeitsschritte zerlegt und ausgeführt
 - Führt die Haupt- und Zusatzfunktionen aufgrund der Konfiguration aus
 - Delegiert die Texterkennung an den OCR-Server und verarbeitet die Resultate der OCR-Erkennung weiter
- **OCR Server:** Dieser Dienst hat die folgenden Aufgaben:
 - Nimmt Text- und Barcodeerkennungsaufträge entgegen
 - Führt eine Bildvorverarbeitung (Entflecken, Geraderichten, etc.) durch
 - Segmentiert das Bild in Fotos, Text und übrigen Inhalt
 - Führt die Text- und Barcodeerkennung mittels einer OCR-Maschine durch
 - Übermittelt die erkannten Texte und Barcodes an den Auftraggeber zurück. Wahlweise werden auch das vorverarbeitete Bild und die Segmentierungsdaten zurückgeliefert.
- **Watched Folder Service:** Dieser Dienst überwacht Verzeichnisse auf Dateisystemen
- **Webservice:** Dieser Dienst nimmt Konversionsaufträge über das Netzwerk entgegen und sendet die konvertierte Datei zurück an den Absender.

Durch die Verteilung der Aufgaben auf die Subsysteme lässt sich der 3-Heights™ Scan to PDF Server skalieren. Je nach Anforderung an den Durchsatz sind verschiedene Konfigurationen möglich. Es ist denkbar, dass in der einfachsten Konfiguration die Subsysteme auf einem Rechner und in der komplexesten Konfiguration die Subsysteme auf getrennten Rechnern betrieben werden könnten.

In praktischen Betriebsumgebungen wird jedoch oft nur ein Rechner eingesetzt, welcher durch leistungsfähige Multi-Core Prozessoren über viel Speicher verfügt.

Leistung

Für die Messung der Leistung wurde der 3-Heights™ Scan to PDF Server mit der folgenden Konfiguration betrieben:

- Hardware: HP Server DL 380p Gen 8
- Virtualisierung: ESXi/VMWare
- Betriebssystem: Windows Server 2012
- Converter Service: 4 Worker Sessions
- Client: Windows 7

Mit dieser Konfiguration ergaben sich die folgenden Performance-Zahlen:

Test	Zeit
Konversion von TIFF in PDF/A	0,05–0,07 s/Seite
Overhead OCR	0,5 s/Seite
Overhead Webservice	0,2 s/Seite

Abbildung 8: Leistungswerte des 3-Heights™ Scan to PDF Server

Schnittstellen zur Applikationsintegration

Für die Integration von Arbeitsstationen und Server Computer, die Applikationen betreiben, welche den 3-Heights™ Scan to PDF Server nutzen wollen, gibt es eine Reihe von Schnittstellen. Die Wichtigsten sind:

- **Webservice:** Der Webdienst erlaubt das Scannen via Internet / Intranet aus einem Webclient oder einer Applikation für ein mobiles Gerät.
- **Programmierschnittstelle (API):** Diese Komponente ermöglicht die programmatische Integration des Dienstes in Applikationen. Sie bietet Schnittstellen für Java, C, COM und .NET Technologien an. Die Komponente ist auch für andere Plattformen wie Linux, Sun OS, AIX, HP-UX, Mac OS/X etc. verfügbar.
- **Command Line Tool:** Dieses Werkzeug ist ein eigenständiges Programm, welches ohne weitere Voraussetzungen direkt auf der Befehlszeile ausgeführt werden kann. Damit lassen sich Abläufe mithilfe der Befehlssprache (Shell Command) automatisieren, ohne dass eine Entwicklungsumgebung benötigt wird. Das Befehlszeilenprogramm ist auch für andere Plattformen wie Linux, Sun OS, AIX, HP-UX, Mac OS/X etc. verfügbar.
- **File Explorer Add-on:** Diese Komponente ist eine Erweiterung zum Windows Date Explorer und bietet dem Benutzer die interaktive Konversion einzelner Dateien an.

Erweiterbarkeit durch Zusatzfunktionen

Der 3-Heights™ Scan to PDF Server kann auf einfachste Weise erweitert werden. Dazu ist es nur notwendig, ein ausführbares Programm zu erstellen, welches den gewünschten Arbeitsschritt implementiert. Das Programm kann mittels Konfiguration in den 3-Heights™ Scan to PDF Server integriert werden. Voraussetzung dafür ist, dass das Programm auf der Befehlszeile gestartet werden kann und die Ein- und Ausgabedateien sowie Steueroptionen als Parameter auf der Befehlszeile übergeben werden können.

Hier ein paar Beispiele solcher Zusatzfunktionen:

- **Automatische Klassifikation:** Die automatische Klassifikation von Dokumenten aufgrund ihres Inhalts, das Auslesen von Lieferanten und Kundenadressen, Rechnungsnummern usw. kann die Verarbeitung von hohen Dokumentenvolumen stark beschleunigen. Dieses Verfahren macht die Index-Stationen für einen grossen Teil der gescannten Dokumente überflüssig.
- **Zerlegen und Zusammenfügen von Seiteninhalten:** Der Inhalt einer Seite kann mehrere logische Abschnitte haben, welche beispielsweise durch Barcodes getrennt sind. Eine wünschenswerte Funktion könnte sein, dass diese Abschnitte isoliert und auf getrennte Seiten verteilt werden.
- **Umwandlung von Farbe in Graustufen:** Wenn Farbe in der spezifischen Anwendung nicht erforderlich ist, dann spart dies noch zusätzlich Speicherplatz.
- **Import anderer Dateiformate:** Es gibt Scanner, welche direkt PDF-Dateien liefern, welche durch den 3-Heights™ Scan to PDF Server importiert und optimiert werden können.
- **Automatische Steuerung der Arbeitsschritte:** Aufgrund von Inhalten oder Formaten kann der 3-Heights™ Scan to PDF Server Art und Abfolge der Arbeitsschritte steuern.

Übersicht über die implementierten Leistungsmerkmale

Die nachfolgende Tabelle beschreibt, welche Leistungsmerkmale in der aktuellen Freigabe bereits realisiert sind. Die übrigen Leistungsmerkmale werden in zukünftigen Versionen der Software implementiert.

Leistungsmerkmal	Release 4.6
Eingabeformate: JPEG, TIFF, PDF	<input checked="" type="checkbox"/>
Ausgabeformate: PDF, PFD/A	<input checked="" type="checkbox"/>
Mixed Raster Content (MRC) Kompression	<input checked="" type="checkbox"/>
Optionale OCR Verarbeitung	<input checked="" type="checkbox"/>
Optionale Barcode Verarbeitung	<input checked="" type="checkbox"/>
Einbettung von XMP Metadata	<input checked="" type="checkbox"/>
Optionale Digitale Signatur	<input checked="" type="checkbox"/>
Optionale PDF/A Validierung	<input type="checkbox"/>
Überwachte Verzeichnisse	<input checked="" type="checkbox"/>
TWAIN Schnittstelle	<input type="checkbox"/>
Webdienst	<input type="checkbox"/>
Mail Server Schnittstelle	<input type="checkbox"/>
Einbettung von ZUGFeRD Rechnungsdaten	<input type="checkbox"/>
Zusatzfunktionen: Zusammenfügen von PDF Documenten, Stempeln, usw.c.	<input type="checkbox"/>
Programmierschnittstelle (API)	<input type="checkbox"/>
Befehlszeilenschnittstelle (CLI)	<input type="checkbox"/>

Über PDF Tools AG

Weit über 4000 Unternehmen und Organisationen in 60 Ländern zählen zu den Kunden der PDF Tools AG. Damit ist sie ein weltweit führender Hersteller von Softwarelösungen und Programmierkomponenten für PDF und PDF/A Produkte.

Diese Geschichte begann vor mehr als 15 Jahren als der Gründer und CEO, Dr. Hans Bärffuss, PDF Technologien erstmals in Kundenprojekten einsetzte. Das PDF und PDF/A Format haben sich seither zu einem mächtigen, weit verbreiteten Format und ISO Standard mit nahezu unbeschränkten Anwendungsmöglichkeiten entwickelt. Auch die PDF Tools AG wurde in dieser Zeit zu einem der Unternehmen im Markt für PDF Technologie und hat den PDF/A ISO Standard für die elektronische Langzeitarchivierung massgeblich mitgeprägt.

Mit ihren Produkten lassen sich PDF-Dokumente anzeigen, drucken, in Bildformate konvertieren (oder umgekehrt), analysieren, reparieren, optimieren, validieren, zusammensetzen, zerlegen, schützen, stempeln, digital signieren oder auch erweitern.

Als Schweizer Vertreter im ISO-Komitee für PDF und PDF/A, lässt das Unternehmen sein Wissen direkt in die Produkteentwicklung einfließen. So entstehen qualitativ hochwertige Produkte mit einer effizienten Leistungsfähigkeit – ganz nach dem 3-Heights™ Motto des Entwicklungsteams: dieses steht für hohe Qualität und hohes Volumen bei gleichzeitig hoher Performanz.

Das Unternehmens-Portfolio erstreckt sich von Komponenten und Services bis hin zu Lösungen. Die Produkte unterstützen den gesamten Dokumentenfluss vom Rohmaterial, über Scanningprozesse bis hin zu Signierung und Archivierung. Hier hilft die Funktionalität der einzelnen Softwarekomponenten, Langzeitarchive mit den damit verbundenen Compliance-Vorschriften rechtskonform zu betreiben. Ein Vorteil der Komponenten und Lösungen ist die breite Palette von Schnittstellen, welche eine reibungslose und einfache Integration in bestehende Umgebungen gewährleisten.

Unser Team aus qualifizierten Ingenieuren übernimmt ausserdem den Support selbst. So erkennen sie rasch aktuelle Trends und Kundenbedürfnisse und setzen das Wissen bei der Planung für Erweiterungen und Komponenten ein.

Sämtliche Entwicklungen erfolgen in der Schweiz. So findet der gesamte Entwicklungsprozess zentral an einem Standort statt. Dahinter steht nicht nur der eigene Anspruch an die 3-Heights-Technologie, sondern auch das Vertrauen in die Qualität des Standorts. Kunden wie Bell Canada, Bombardier Inc. und der World Intellectual Property Organization (WIPO) bestätigen den Erfolg dieses Ansatzes und sind gleichzeitig das schönste Kompliment und die beste Motivation für unser Team, weiter einen wichtigen Beitrag in der PDF und PDF/A Welt zu leisten.

Weiterführende Informationen

auf unserer Webiste via www.pdf-tools.com

- Handbuch 3-Height™ Scan to PDF Server
- PDF Expertenblog: blog.pdf-tools.com