Alfresco is an open source Enterprise Content Management (ECM) system for Windows and Linux-like operating systems.

The year-on-year growth of business connections, contacts, and communications is expanding enterprise boundaries more than ever before. Alfresco enables organizations to collaborate more effectively, improve business process efficiency, and ensures information governance.

The basic purpose of Alfresco is to help users to capture and manage information in a better way. It helps you capture, organize, and share binary files.

This book will cover the basic building blocks of an Alfresco system, how the components fit together, and the information required to build a system architecture. This book will also focus on security aspects of Alfresco such as authentication, troubleshooting, managing permissions, and so on. It will also focus on managing content and storage, indexing and searches, setting up clustering for high availability, and so forth.

Who this book is written for

The target audience would be users with a basic knowledge of content management system, and also users who want to understand Alfresco from the administration and high-level design perspectives.

What you will learn from this book

- Understand Alfresco’s architecture and important building blocks
- Learn to install Alfresco on various application servers such as Tomcat and JBoss
- Become familiar with various configurations in Alfresco such as databases, filesystems, e-mail, and audits
- Administrate Alfresco using the Explorer Admin Console, Share Admin Console, and Workflow Admin Console
- Understand how to integrate LDAP and Active Directory with Alfresco for centralized user management
- Learn how Alfresco environments can be clustered for high availability
- Fully understand how Alfresco stores content and easily retrieves any information from Alfresco
- Monitor and manage Alfresco systems in production

In this package, you will find:

- The author biography
- A preview chapter from the book, Chapter 1 'Understanding Alfresco'
- A synopsis of the book’s content
- More information on Alfresco for Administrators
Vandana Pal is a software engineer and author. She currently works as senior consultant at CIGNEX Datamatics.

She has extensive experience working with Enterprise Digital Asset Management and Content Management Systems. She has worked with various deployments of Alfresco in various domains, such as media, finance, and healthcare, for different organizations across the world. She has hands-on experience working with architecture design, performance tuning, security implementation, integration, and the orchestration of complex workflows in Alfresco.

She has more than 7 years of experience in software engineering. Her journey in this field began when she started working with different open source technologies and found them interesting. She holds a bachelor's degree in information technology from Gujarat University, India.

Vandana has also coauthored *Alfresco 4 Enterprise Content Management Implementation*. 
This book focuses on the administration part of Alfresco. It also gives you a high-level understanding of Alfresco and its capabilities from the perspective of its architecture. This book provides you with details of how to administer and troubleshoot problems in Alfresco. It also gives you an in-depth insight into configuration, clustering, backup recovery, and maintenance. You thoroughly understand Alfresco's repository structure and learn how to install, configure, search, and administrate Alfresco.

What this book covers

Chapter 1, Understanding Alfresco, gives you a thorough understanding of Alfresco's architecture and its features.

Chapter 2, Setting Up the Alfresco Environment, explains the various installation processes of Alfresco in different application servers. It also provides details of best practices and troubleshooting for the environment setup.

Chapter 3, Alfresco Configuration, explains the ways in which Alfresco can be configured to suit business needs. It gives you a detailed understanding of the different components of Alfresco and how they can be configured.

Chapter 4, Administration of Alfresco, explains how the Alfresco repository can be administered. It provides you with detailed steps for the administration of the repository and the functions it can perform. The chapter gives you a thorough understanding of the administration console, users and group creation processes, node browsers, and so on.

Chapter 5, Search, focuses on the search component of Alfresco. It will provide you with a detailed insight into the installation, configuration, troubleshooting, and maintenance of the search server Solr.
Chapter 6, Permissions and Security, explores the details of the permissions required in the Alfresco repository. It provides you with an understanding about the different types of permissions and roles in Alfresco and how to integrate with different third-party authentication tools.

Chapter 7, High Availability in Alfresco, explores the different ways in which the Alfresco system can be made highly available. It covers the different methods of clustering Alfresco and its backup and recovery process as well as troubleshooting Alfresco's clustered environment.

Chapter 8, The Basics of the Alfresco Content Store, explains how Alfresco actually stores content. The content life cycle is discussed in detail in this chapter as well as the database structure in Alfresco.

Chapter 9, Maintenance and Troubleshooting, covers how to monitor and manage the Alfresco system in production using JMX and other tools. It also provides details about different ways to troubleshoot Alfresco. You will also learn about different audit trails in Alfresco for the purpose of better administration.

Chapter 10, Upgrade, explains how Alfresco can be upgraded from one version to another. It provides detailed steps in order to understand the process of upgrading.
Understanding Alfresco

Alfresco is one of the leading open source enterprise content management systems (ECM). For more details about ECM refer to the Wiki; https://en.wikipedia.org/wiki/Enterprise_content_management. Alfresco allows you to manage content in a simple and smart way. It provides enterprise solutions based on open standards, and open source technologies for managing business critical content. As it is a very stable player in the market and provides enterprise-level features and support, Alfresco has been named Visionary by Gartner for five years in a row. Gartner is a leading research company, which provides insight into technology; refer to http://www.gartner.com/technology/about.jsp for more details about Gartner.

This chapter provides you with an introduction to Alfresco 5.x, its features, and its benefits. It helps you to understand the main building blocks of Alfresco.

By the end of this chapter, you will have learned about:

- An overview of Alfresco
- Key features of Alfresco
- Alfresco architecture
- Using Alfresco for your ECM requirements

Overview of Alfresco

Alfresco Community Edition is only for small-scale development or research purposes. It is not recommended for production systems as there are certain functional differences. The Community version doesn't support clustering, enterprise application servers such as WebLogic, enterprise databases such as Oracle, encryption of content stores, advanced admin tools, advanced media management, and so on. There is no Alfresco support provided for the Community version. Alfresco Enterprise Edition is production-ready code. It has been load tested and certified for use in production. The Enterprise build is fully supported by Alfresco. Alfresco in Cloud is a SaaS (Software as a Service) version of Alfresco. More details on this are given in later sections.

Refer to the following URL for more details about the differences between Community and Enterprise versions:


Enterprise Edition has various unique features, which distinguish it from other ECM systems.

**Enterprise and open source**

As Alfresco is built upon open source technologies, it reduces the cost of overall software acquisition, development, and maintenance. Due to this open source model, Alfresco can use the best open source technologies on the market and build a strong system at a low cost. Alfresco provides a very cost-effective solution.

**Scalable**

Scalability is a very important aspect for any ECM system. For enterprise organizations in fields such as media, healthcare, finance, and so on, the amount of content grows exponentially, so scalability becomes an important parameter. As Alfresco is built using open source standards and technologies, it provides a very scalable architecture.

Alfresco Enterprise can be deployed on any platform, and supports multiple databases such as MySQL, Oracle, PostgreSQL, and so on. It also supports multiple application servers such as Tomcat, JBoss, WebLogic, and so on. Each tier in an Alfresco application can be deployed on a separate machine, which allows the vertical scalability of the system. Alfresco supports a clustered environment, which allows it to scale horizontally.
Rich media support
ECM systems should support any type of content, regardless of application or organization. Alfresco supports the storage and management of multiple types of electronic content, from normal documents to any multimedia files. It provides automatic extraction of the information from files, associates it as metadata with content, and enables easy searching.

Secured system
Security and content protection is critical for any ECM system. Alfresco has a very strong authentication and authorization model. It provides an out-of-the-box database membership system; it can also be integrated with identity management systems like LDAP and Active Directory (AD), and have centralized security and single sign-on. Alfresco provides full access control on individual content to ensure that security and business integrity is maintained. Access control can be set at the folder level or individual content.

Highly extensible
Because of its open source model, Alfresco can be extended and customized as per requirements. Organizations can have a trained in-house team to maintain and customize Alfresco as per their needs.

External integration
Alfresco supports open standard protocols for integration with external systems. Alfresco can be integrated with any Java-based portal, such as Liferay (https://www.liferay.com/products/liferay-portal/overview) using the CMIS or REST protocols.

CMIS is a standard open source protocol to allow a document management repository to connect with a web application. It defines an abstract layer so the web interface can connect with any repository. For more details, refer to https://en.wikipedia.org/wiki/Content_Management_Interoperability_Services.

The REST protocol allows an external application to access the repository using the HTTP protocol using the same HTTP verbs, such as GET, POST, and so on. For more details, refer to https://en.wikipedia.org/wiki/Representational_state_transfer.
Alfresco provides integration with various scanning solutions, such as Ephesoft or Kofax, which gives a complete end-to-end solution. It allows organizations to perform document capture, extraction, classification, storage, and distribution via a centralized environment.

For more details about Ephesoft and Kofax refer to these URLs:

- http://ephesoft.com/products

**Collaboration**

Nowadays, due to social media, collaboration has become very important for any organization as part of ECM. Alfresco, as well as content management, also provides a platform for collaboration between users internally and externally with full security and control over content. Powerful tools such as blogs, wikis, forums, and so on are provided within the Alfresco system to provide collaboration within teams.

Each project can have its own space for complete collaboration and the sharing of content.

Alfresco supports the publishing of content to various social platforms such as Twitter, Facebook, YouTube, SlideShare, and so on. It also provides Google Doc integration, which allows users to have real-time collaboration.

**Business process management**

Efficient business processes are an integral part of any organization. Automation of this process helps organizations to streamline processes, improve efficiency, and reduce cost. In organizations where the review and approval process of any document is very important, there would always be a need for these documents to be moved and accessed effectively.

Alfresco provides the Java-based, highly configurable BPM engine Activiti (http://www.activiti.org/). It also provides graphical tools so that less technical persons can easily design the process flow, allowing the faster rollout of processes.

As Alfresco can be accessed by any supported browser or mobile device, users get the flexibility to perform their tasks from anywhere.

Alfresco provides easy configurable rules, which can help to trigger and control this business process in a smart way.
Cloud-based ECM

Alfresco provides a fully managed SaaS ECM solution, leveraging the power of a cloud-based environment. Alfresco in Cloud is a ready-to-go Alfresco implementation which requires no installation and minimal configuration by customers. It allows full control over, and collaboration on, documents, similar to what can be achieved by Alfresco deployed on-premises.

Alfresco also supports a hybrid model, where content can be synchronized from your on-premises Alfresco to the cloud. This allows content to be always in sync and easily available from any location. An Alfresco on-premises solution can be used for long-term storage and compliance, and Alfresco in cloud can be used for sharing and collaboration too.

Search

Finding the correct content within a system is very important for any content management system. Alfresco provides searching with Apache Solr (http://lucene.apache.org/solr). It provides full-text indexing of content, and metadata indexing, which allows users to easily search and locate the content in the repository. Alfresco also provides advanced search capabilities.

Alfresco also supports searches for archived content, users, and groups in the system.

Version control

Maintaining all versions of a document is also a critical aspect of an ECM system. Alfresco provides strong version management for documents. It maintains all the version changes of a document and its associated metadata. Alfresco also has a feature that allows you to revert a document to any version.

Auditing

Alfresco provides very strong auditing. Each and every action on content is captured in an audit trail. This audit information can be easily retrieved and generated as a report.
Alfresco architecture overview

Alfresco is the leading open source option for ECM. Alfresco architecture is designed based on open standards JSR-170, JSR-168, and JSR-283. JSRs are industry standards defined by the Java community for uniform repository access, using the Java platform application programming interface. Refer to https://en.wikipedia.org/wiki/Content_repository_API_for_Java for more details.

Alfresco supports pluggable aspect-oriented architecture. It is lightweight, modular, and scalable.

The following is a high-level diagram of the Alfresco architecture:

![Alfresco Architecture Diagram](image)

### Alfresco Share

This is the collaboration content management platform in Alfresco. It is built on the Surf framework. The Surf framework was developed by Alfresco, but in 2009 Alfresco began working with Spring Source and announced the Spring Surf Extension framework. Later on, both Spring Source and Alfresco were collectively developed and are available as plugins in Spring MVC 3.x.
Refer to the following links for more details:

- http://www.springsurf.org/

Alfresco Share simplifies document capturing and sharing, and the retrieval of data for teams, resulting in better collaboration. This in turn increases the productivity of teams and reduces the volume of e-mails.

Alfresco Share also provides advanced administrative tools. It supports module-based extension, which supports the ability to remove, add, or modify any component without changing any out-of-the-box code.

**Alfresco repository**

This is the main core of Alfresco. Alfresco repository is a bundle of service implementations based on the open standards of CMIS and JCR. This service provides cutting edge content management features such as:

- Content storage
- Content retrieval
- Content modeling
- Query interface
- Access control
- Audit
- Versioning

These services provide a public interface based on REST/CMIS or Java JSR-170 protocol standards which allows the client application to communicate with the repository. Alfresco Share communicates with the repository using the REST interface.

The content repository is more than a normal database application, due to the level of control over individual content it provides. Access to content is wrapped by a security layer which prevents any unauthorized access. The fine-grained security control requires a more complex approach than a traditional database application.

In Alfresco, the actual binary stream of content is located in the file system. The file folder structure and reference to this binary stream is maintained in the database.
Filesystem protocol (CIFS/WebDAV/FTP)
Access to content stored in a repository is very important for any ECM. Alfresco supports various protocols, such as CIFS (Common Internet File System), WebDAV, and FTP.

These protocols allow you to support the mapping of the same file folder structure as the repository to a virtual filesystem. With these protocols, any tool that can read and write a filesystem can read and write to an Alfresco repository. Users can still use Alfresco as a locally mapped network filesystem. CIFS provides advanced compatibility with the mapped operation system. With the CIFS protocol, Windows users can use the Windows offline synchronization feature with an Alfresco repository. These virtual filesystem protocols allow users to edit and view content using their locally installed tools.

Database
The database holds all the content related information, such as metadata, content association, content binary stream location reference, and folder structure. The database also stores information related to users, workflow tasks, audits, and so on.

Alfresco supports various database vendors, such as MySQL, PostgreSQL, Oracle, and so on. Oracle is only supported in Alfresco Enterprise Edition. Database schema and more information will be covered in Chapter 8, The Basics of the Alfresco Content Store.

Content store
The content store is a term used for the filesystem location where the actual binary stream of content is stored. In Alfresco, only the reference to the content is stored in the database. The actual content is stored in a filesystem. This filesystem can be any normal NAS or SAN mounted drive. This architecture allows Alfresco storage to grow exponentially and makes Alfresco scalable.

Solr indexes
Searching is a very important aspect of any ECM system. Alfresco supports searches using Apache Solr. All content, metadata, and permissions associated with content in Alfresco are indexed in Solr, which allows fast searches and access to content stored in a repository.

Solr can be bundled with Alfresco on the same machine, or it can be installed as a separate tier. This design allows the horizontal scalability of the search tier. Alfresco and Solr communicate with each other asynchronously.
Business use cases of Alfresco

Alfresco as a true ECM system provides a simple and smart way to manage your content. Alfresco provides various systems as solutions to support document management, record management, collaboration, and so on, in order to solve organizational challenges.

Alfresco as a document management solution

Alfresco can be used as a document management solution for any organization where the documents are business critical, and storing and retrieving them effectively is very important for the business. For example, contracts are very important documents for many firms. All contracts can be stored in a central location within Alfresco. Strong access control can be applied to each contract document, so only authorized users can view/edit the contract.

Metadata information from the contract document can be extracted and indexed in Alfresco, which allows users to search any contract easily. As Alfresco supports full-text searches, users can search the contract document based on its content. The versioning features of Alfresco can be leveraged to ensure that all the versions of the contract are kept.

Alfresco provides a strong audit trail, so all the actions taken on the contract by any user can be captured and an audit report can be generated from them very easily.

Alfresco also supports integration with various scanning and OCR solutions, such as Ephesoft, so any paper contracts can be scanned, classified, and stored in the repository.

For contracts, the review and approval process is very important. Alfresco has strong business process management which can be leveraged to automate this process, reduce the length of the approval cycle, and improve efficiency. As Alfresco can be accessed from the Web, users can view documents and perform operations from any location.

Alfresco as a record management solution

Alfresco record management is a great solution for any organization where compliance is important. It is simple and very user-friendly. Users can adopt this system easily.
It can be extended to create a single centralized repository to manage all kinds of electronic records. Alfresco provides strong access control, so all records are secure. The policies for record use, storage, and disposal can be easily defined with Alfresco record management.

Alfresco record management is designed based on United States Department of Defence 5015.2 record management standards.

With Alfresco, you can easily drag and drop records into the system. Business rules can be defined to classify and mark them as records. A disposition policy can be defined and automated, which includes the transfer of records or their complete destruction after a given period. In addition to this, there is strong auditing that captures all actions on the records.

Alfresco provides different reports that show recent records, records due for expiry, records due for destruction, and records due for transfer.

**Alfresco for collaboration**

Alfresco can be used in collaboration solutions within an organization, along with content management. For example, a marketing team can work on different projects. Alfresco Share can be used as a collaboration platform. Each marketing project can be created as a different space. Only members of that project can have access to that space.

Teams can upload, share, and discuss content within this space. There are dashboards which can be configured as per user needs to see the activity in the project and notifications. Alfresco acts as a central repository to manage all types of marketing documents.

Alfresco provides a feature for publishing content to any social platform, including Twitter, SlideShare, and Facebook, which can be leveraged and content can be published directly.

With Alfresco, content can be shared with external users in a secure and controlled way.

For more case studies on Alfresco, you can refer to [http://www.alfresco.com/customers](http://www.alfresco.com/customers).
Summary

Alfresco is one of the leading open source ECM systems. The key features of Alfresco are security, stability, and a scalable architecture. Due to its open source model, Alfresco can use the best open source technologies on the market and build a strong system at a low cost. Alfresco provides a very cost effective solution.

Alfresco architecture is designed based on JCR open standards. It is lightweight, modular, and scalable.

Alfresco can be used in the cloud, on-premises, or as a hybrid. The next chapter will cover details about the installation of an Alfresco system on various platforms.
Where to buy this book
You can buy Alfresco for Administrators from the Packt Publishing website.
Alternatively, you can buy the book from Amazon, BN.com, Computer Manuals and most internet book retailers.
Click here for ordering and shipping details.