



Alfresco. Принципиальная архитектура и область применения

2012



Аннотация

Alfresco — платформа для управления всем контентом крупных организаций. Какие задачи и каким образом на самом деле решаются при помощи Alfresco? Почему Alfresco, а не ее закрытые аналоги?

Обратная связь

Данный документ подготовлен Бюро Соломатина, подразделением НЦПР.

Сайт: www.bureausolomatina.ru
Электронная почта: bureau@ncpr.su
Телефон: +7 495 988 27 09
Факс: +7 495 745 40 81
Адрес: 125375, Россия, Москва, ул. Тверская, д. 7, под. 7, этаж 2

«Alfresco. Принципиальная архитектура и область применения», Редакция от 05.03.2012

© ООО Национальный центр поддержки и разработки, 2012 г.

Вы имеете право воспроизводить, распространять, перерабатывать документ или использовать его иным образом в соответствии с условиями лицензии Creative Commons Attribution Share Alike 3.0 (текст лицензии доступен по адресу <http://creativecommons.org/licenses/by-sa/3.0/legalcode>).

Автор: Курышева Оксана, Бюро Соломатина

Права на товарные знаки, упомянутые в настоящем документе, принадлежат их законным правообладателям.



Enterprise Content Management

Платформы ECM позволяют хранить и обрабатывать весь корпоративный контент организации в одном месте, независимо от его типа и предназначения. К корпоративному контенту относятся любые документы и записи, обрабатываемые в организации, от бухгалтерских документов и договоров до маркетинговых материалов и руководств пользователя по программному обеспечению, используемому в организации.

Платформы ECM решают следующие задачи:

- упорядоченное хранение всего контента организации в одной системе;
- избежание дублирования документов;
- избежание дублирования функционала систем работы с документами, которое возникает при использовании различных систем в разных отделах в рамках одной организации.

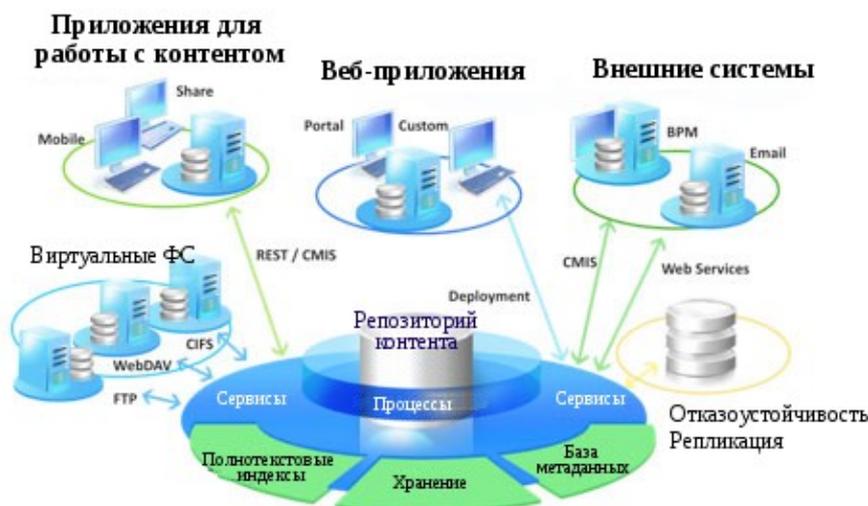
Платформа отвечает за хранение всего контента организации и его описания, а также за реализацию функций, общих для всех этих систем, таких как поиск, разграничение доступа к контенту, правила обработки документов, и позволяет создавать приложения на базе платформы, которые будут решать конкретные пользовательские задачи, такие как согласование договоров, хранение архивных копий документов, выставление счетов и обработка входящей корреспонденции.

Alfresco

Alfresco — одна из таких платформ ECM. Ядром Alfresco является репозиторий контента, который хранит контент, его метаданные (карточки документов), информацию о связях между контентом. Для работы с репозиторием в Alfresco реализованы базовые функции по работе с контентом. К числу таких функций относятся:

- создание своих моделей структуры контента;
- создание, изменение и удаление документов, его метаданных и связей;
- разграничение прав доступа;
- версионирование;
- блокировка;
- использование контента в бизнес-процессах организации;
- импорт/экспорт документов;
- создание правил обработки контента.

Создавая свои приложения, разработчик не создает функции подобного уровня самостоятельно и не думает о том, как и где он должен физически хранить контент, а обращается к репозиторию Alfresco и опирается на уже реализованные функции, обращаясь к ним посредством различных протоколов и интерфейсов API, предоставляемых системой.



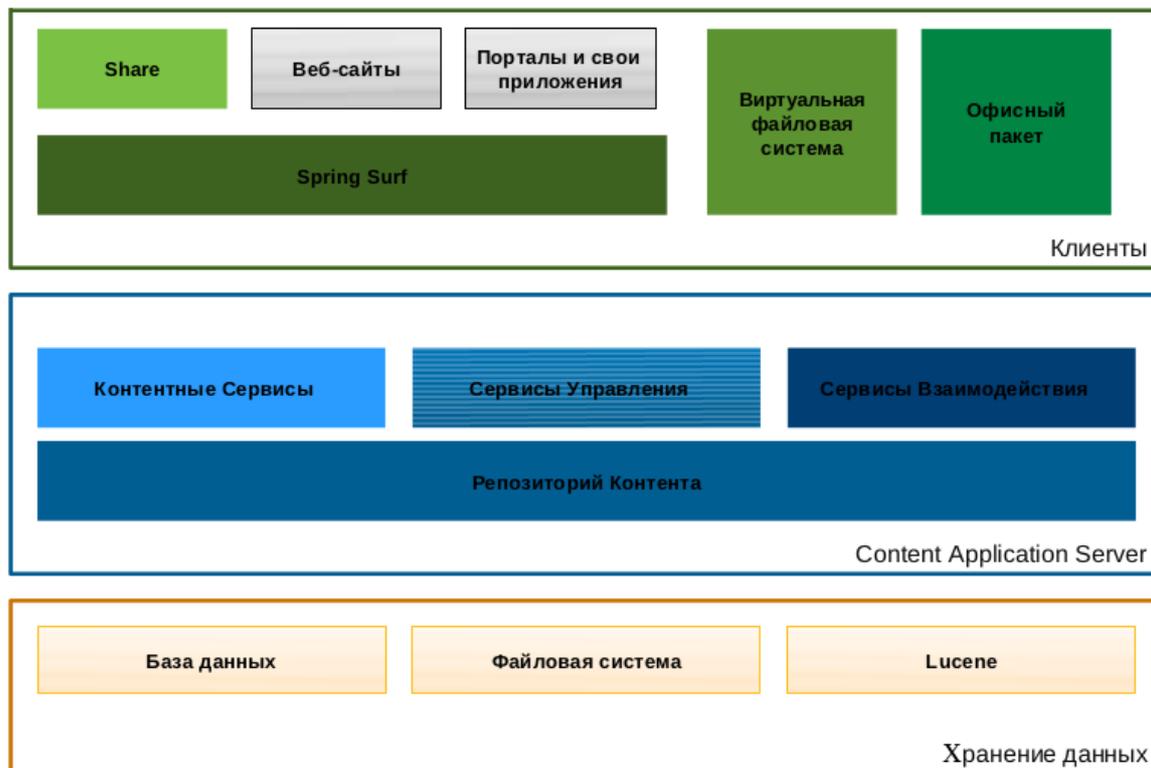
Для того, чтобы показать пример приложений, использующих Alfresco в качестве платформы, в стандартную поставку включены два веб-приложения (Alfresco Explorer, Alfresco Share) и мобильный клиент для iPhone/iPad.



Также доступ к репозиторию контента Alfresco можно получить из Microsoft Office (по протоколу SharePoint), просто из операционной системы пользователя, подключив репозиторий, как внешний сетевой диск, и с порталов, подключив репозиторий Alfresco как портлет.

Архитектура Alfresco

Архитектура Alfresco может быть представлена в виде трех уровней:



- Первый уровень отвечает за физическое размещение данных репозитория на сервере.
- Второй уровень — сервер управления контентом. Он реализован на Java и поддается масштабированию при работе под большими нагрузками. Сервер управления контентом состоит из непосредственно репозитория контента и набора сервисов, через которые пользователь может получать доступ к контенту.
- Третий уровень — клиенты, подключенные к серверу управления контентом через различные протоколы и интерфейсы API.

Таким образом, Alfresco полностью отвечает за реализацию первых двух уровней и позволяет разработчикам подключать свои приложения на третьем уровне.

Хранение контента

Весь контент состоит из двух частей: сами документы и их описание. Документы Alfresco сохраняет на файловой системе сервера. При этом контент хранится в без указания имени файла, владельца и прочих метаданных, по которым можно идентифицировать документ. Так что недобросовестному пользователю, имеющему доступ к ФС сервера, буде нелегко найти конкретный документ на сервере, не имея доступа к репозиторию Alfresco.



Описание документа, включая все метаданные, права на документ и его положение в структуре директорий Alfresco хранятся в базе данных. За связь между контентом и метаданными, а также за структуру файлов и директорий контента отвечает репозиторий, находящийся на втором уровне архитектуры.

Для осуществления быстрого поиска по документам при загрузке контента в репозиторий создаются полнотекстовые индексы. За создание индексов, а также за сам поиск отвечает Lucene. Lucene — приложение, разрабатываемое в рамках проекта Apache, и встроенное в Alfresco.

Сервер управления контентом

Сервер управления контентом отвечает за контроль, доступ и обновление контента в репозитории. Также он позволяет запускать клиентские приложения через веб-скрипты и пакеты расширений. Все приложения на основе Alfresco взаимодействуют напрямую с сервером управления контентом.

Сервер управления контентом состоит из репозитория контента и набора сервисов, через которые пользователь может получать доступ к контенту.

Первая задача репозитория — собрать данные, хранящиеся на файловой системе сервера и в базе данных в понятный пользователю контент с его метаданными и связями между документами. Данная задача выполняется прозрачно для пользователя.

Вторая задача репозитория — предоставить пользователю возможность работы с контентом. Alfresco состоит из нескольких подсистем, работающих в рамках репозитория. Каждая подсистема выделена в отдельный модуль и может быть отключена, перезагружена или перенастроена, не влияя на работоспособность Alfresco в целом. Примерами таких подсистем служит аутентификация пользователей для выдачи им прав доступа к контенту и модуль полнотекстового поиска, подключающий Lucene к Alfresco.

Репозиторий контента Alfresco соответствует стандартам CMIS (Content Management Interoperability Services) и JCR (Java Content Repository / JSR-170/286). Это открытые стандарты по правилам типизации и хранения контента, извлечения контента из репозитория, версионирования и прав доступа. Репозиторий контента Alfresco соответствует обоим стандартам, что позволяет легко мигрировать с одной платформы ЕСМ на другую при необходимости и избежать привязки к одному поставщику.

Вторая часть сервера управления контентом — сервисы взаимодействия с клиентскими приложениями. Задача сервисов — предоставить разработчикам не только одиночные функции, которые можно использовать в своих приложениях, но и готовые блоки функций, которые потребуются при построении различных приложений управления контентом. Все сервисы реализованы поверх репозитория и соответствуют конкретным подсистемам репозитория по тому же принципу модульности, что был описан выше. Сервер управления контентом предоставляет следующие категории сервисов:

- контентные сервисы, например, перевод документов из одного формата в другой, тэгирование и извлечение метаданных контента;
- сервисы управления, например, создание бизнес-процессов, предпросмотр неопубликованных документов, публикация контента в разных внешних



приложениях;

- сервисы взаимодействия, например, ведение блогов, создание вики-статей в Alfresco Share и возможность создания оргструктуры пользователей.

Alfresco продолжает добавлять новые сервисы в каждом новом релизе. На базе доступных API разработчики могут встраивать собственные сервисы в сервер управления контентом Alfresco.

Клиентские приложения

Примерами клиентских приложений, построенных на основе платформы Alfresco, являются два стандартных веб-интерфейса:

- Alfresco Explorer

Alfresco Explorer является первой реализацией клиентского приложения для Alfresco для управления контентом. Приложение написано на Java и позволяет получить доступ к контенту и основным функциям Alfresco через веб-интерфейс. На данный момент приложение уже не дорабатывается и служит готовым примером клиента для работы с Alfresco.

- Alfresco Share

Alfresco Share — веб-приложение на платформе Spring Surf, подключающееся к репозиторию контента Alfresco через веб-скрипты. Приложение реализует функционал системы совместной работы с документами с простым пользовательским интерфейсом, предоставляя пользователям возможность полнотекстового поиска по контенту репозитория, просмотр документов в браузере, использование контента в бизнес-процессах организации и публикация контента в социальных сетях. Для пользователя интерфейс Alfresco Share — набор сайтов (групп) с библиотекой документов по определенной тематике, вики-страницами, ссылками, блогами и форумом. Пользователь может настраивать и конфигурировать Alfresco Share в соответствии со своими предпочтениями, а разработчики могут дорабатывать интерфейс для решения конкретных нужных пользователю задач. Доработка Alfresco Share происходит также, как и любых других приложений на основе Spring Surf. Примером доработки Share для решения конкретных задач является пакет Records Management, сертифицированный по стандартам DOD 5015.2, незначительно изменяющий интерфейс Share.



Область применения Alfresco

Большинство заказчиков Alfresco — крупные организации. Они выбирают Alfresco для решения описанных ранее задач. Как правило, среди других платформ ECM выбором становится именно Alfresco по следующим причинам:

- масштабирование платформы;
- простота миграции с других систем.

Масштабирование платформы

Когда речь идет о крупной организации с большим количеством географически-распределенных офисов, всегда поднимается вопрос создания децентрализованной системы. А иногда даже в централизованной системе невозможно обработать требуемый объем данных, используя один, даже самый лучший, сервер.

Alfresco поддерживает возможность кластеризации как для задач повышения отказоустойчивости, так и для задач улучшения производительности. Благодаря модульности системы все ключевые компоненты Alfresco могут быть разнесены по разным физическим машинам и дублированы. При обработке больших объемов данных используются следующие компоненты системы:

- внешние сервера хранения данных (SAN, NAS, CAS);
- СУБД в режиме кластера (Oracle, DB2, PostgreSQL);
- масштабируемые сервера приложений (Jboss, WebSphere), поверх которых работает распределенный сервер Alfresco;
- сервер индексирования контента SOLR устанавливается на отдельную машину и масштабируется при необходимости.

Простота миграции с других платформ

Благодаря поддержке открытых протоколов и соответствию основным мировым стандартам Alfresco легко встраивается в уже существующую инфраструктуру, которая первоначально может быть ориентирована на отсутствие платформы ECM или заточена под другую платформу. Это также позволяет проводить постепенную миграцию программного обеспечения организации на новую платформу.

Например, при замене Microsoft SharePoint на Alfresco миграция пройдет прозрачно для пользователей Microsoft Office на рабочих станциях. Эти пользователи смогут обращаться к документам в Alfresco также, как раньше — к контенту в SharePoint.

Еще одним примером миграции может служить замена устаревшего IBM Lotus Domino на платформу Alfresco. Во множестве российских организаций Lotus Domino был закуплен в 15-20 лет назад и с тех пор не обновлялся. При этом старые версии Lotus еще не поддерживают возможность масштабирования, но уже не справляются с современным



объемом контента. Решением проблемы является использование Alfresco в качестве репозитория контента для Lotus Domino. Самая большая нагрузка приходится именно на стандартные задачи для репозитория, такие как индексирование, создание и редактирование контента, а не на решение специфичных задач. При использовании Alfresco вся нагрузка будет ложиться на платформу, а Lotus Domino служит интерфейсом для пользователей Lotus Notes и сервером выполнения специфичных для Lotus задач.



Вывод

Alfresco — платформа ECM, предназначенная для решения задач управления всем контентом организации. На базе этой платформы могут быть реализованы различные модули для решения конкретных пользовательских задач, таких как согласование договоров, архив документов, обработка заявок.

Кроссплатформенность, модульная архитектура и открытость используемых компонентов позволяет Alfresco интегрироваться практически в любую инфраструктуру. При необходимости реализации нестандартного функционала разработка новых модулей системы является намного более простой задачей, чем при создании продукта «с нуля» или чем доработка закрытых платформ.