



На платформе Alfresco ECM

**Описание программы Citeck EcoS**

## Оглавление

<b>Общая информация.....</b>	<b>3</b>
Аннотация .....	3
Перечень сокращений .....	3
Основные понятия системы .....	3
<b>Структура системы Alfresco.....</b>	<b>4</b>
Трехзвенная структура .....	4
Структура сервера приложений.....	5
Взаимодействие со смежными системами.....	6
Ресурсы, необходимые для работы Alfresco.....	7
Аппаратные ресурсы .....	7
Программные ресурсы.....	8
Методы расширения и настройки Alfresco .....	9
Лог-файлы Alfresco (журналы событий) .....	10
<b>Модули расширения Citeck EcoS.....</b>	<b>10</b>
Трехуровневая структура модулей .....	10
Состав и функции модулей ядра.....	10
Состав и возможности модулей приложений .....	12
Описание сервиса синхронизации .....	12

## Общая информация

### Аннотация

Настоящий документ содержит техническое описание системы Citeck EcoS, а также системы Alfresco, на которой основана Citeck EcoS. Описываемые версии систем: Citeck EcoS 2.8.1, Alfresco Community 4.2.c.

### Перечень сокращений

Таблица 1. Сокращения

Сокращение	Расшифровка
AD	Microsoft Active Directory (служба каталогов от Microsoft)
ПО	Программное обеспечение
ОС	Операционная система
БД	База данных
СУБД	Система управления базами данных
СЭД	Система электронного документооборота
ИС	Информационная система
SSO	Single-Sign-On (прозрачная аутентификация, однократный ввод пароля)
JDBC	Java DataBase Connectivity (библиотека для доступа к БД в Java)

### Основные понятия системы

**Сервер приложений** – программный компонент Java Enterprise Edition, позволяющий запуск и выполнение приложений Java.

## Структура системы Alfresco

### Трехзвенная структура

Alfresco придерживается трехзвенной архитектуры (Рисунок 1):

- хранилище данных (Physical storage);
- сервер приложений Alfresco (Alfresco Content Application Server);
- клиентские приложения Alfresco (Alfresco Client).

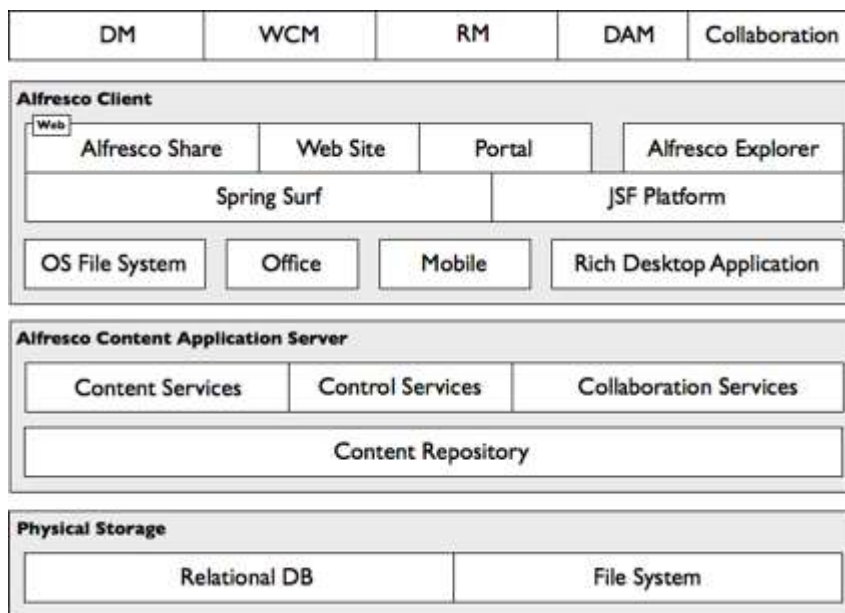


Рисунок 1 – Структура звеньев Alfresco

Звено «Хранилище данных» состоит из хранилища файлов (File System) и базы данных (Relational DB). Хранилище файлов используется для хранения контента документов (файлов) и версий документов. Вся остальная информация (метаданные) хранится в БД. *Схема базы данных Alfresco не является публичным API, часто изменяется от версии к версии Alfresco, и поэтому не должна быть использована при взаимодействии с Alfresco. Вместо этого, для взаимодействия с Alfresco рекомендуется использовать один из множества публичных сервисов (Embedded API, Remote API, см. ниже).*

Звено «Сервер приложений» предоставляет расширенные сервисы для работы с документами. Вся бизнес-логика работы с данными сосредоточена на этом звене. Для работы с различными видами клиентов, сервер приложений предоставляет различные виды внешних протоколов, такие как CMIS, REST API, CIFS, IMAP и другие. С другой стороны с хранилищем данных сервер приложений работает через соединение с СУБД (JDBC) и механизмы файловой системы ОС. Благодаря использованию этих механизмов в качестве хранилища данных можно использовать любые реализации СУБД и файловых систем, единственное существенное требование – наличие необходимых драйверов JDBC и ОС.

Звено «Клиентские приложения» представлено различными клиентскими приложениями Alfresco. Это могут быть как веб-клиенты в браузере, так и приложения для настольных и мобильных платформ и даже доступ через механизмы файловой системы (CIFS, FTP, WebDAV).

Описанная архитектура позволяет реализовывать различные виды документно-ориентированных приложений, таких как управление файлами (Document Management – DM), управление веб-контентом (Web Content Management – WCM), управление документами (Records Management – RM) и другие.

### Структура сервера приложений

Сервер приложений Alfresco является веб-приложением Java и состоит из различных компонентов. На верхнем уровне структура сервера приложений представлена ниже (Рисунок 2).

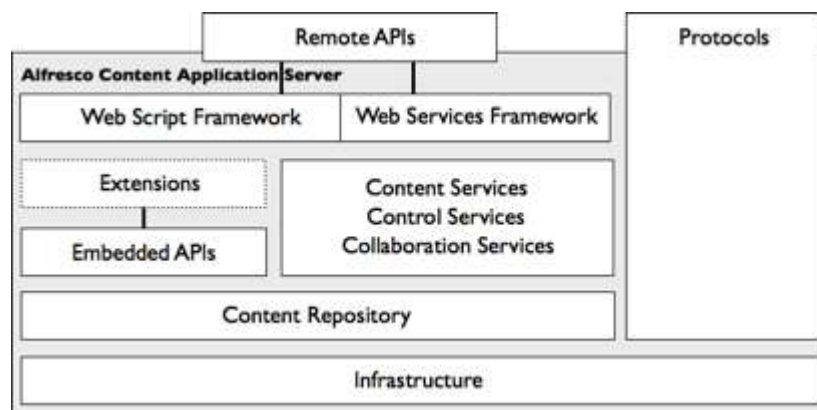


Рисунок 2 – Структура сервера приложений Alfresco

Сервер приложений следует рассматривать как расширенную СУБД, предоставляющую широкий спектр сервисов работы с документами:

- сервисы для работы с контентом (Content Services) – моделирование, поиск, управление версиями, многоязычность, экспорт/импорт, преобразования контента, классификация контента через категории и тэги, выделение полей метаданных из контента;
- сервисы управления (Control Services) – бизнес-процессы, правила и политики, права доступа, аудит доступа, генерация предварительного просмотра, публикация;
- сервисы взаимодействия (Collaboration Services) – избранное, «мне нравится», лента событий, вики, блоги, форумы.

Эти сервисы предоставляются через ряд программных интерфейсов и протоколов:

- внутренние интерфейсы (Embedded APIs) – интерфейс для Java, JavaScript, FreeMarker (шаблоны), языки для моделирования контента и бизнес-процессов;
- внешние интерфейсы (Remote APIs) – интерфейсы веб-сервисов (SOAP), веб-скриптов (REST), а также CMIS API (варианты для SOAP и REST);
- другие протоколы (Protocols) – CIFS, WebDAV, FTP, IMAP, SharePoint.

Архитектура сервера приложений позволяет реализовывать различные виды модулей и расширений к стандартному набору интерфейсов:

- модели контента;
- бизнес-процессы;
- дополнительные сервисы для Java, JavaScript, FreeMarker;
- правила и действия, политики;
- дополнительные веб-скрипты (REST API).

## Взаимодействие со смежными системами

Для реализации своих сервисов сервер приложений пользуется услугами смежных систем. В частности используются сторонние системы для:

- индексации и поиска (Lucene, SOLR);
- аутентификации (LDAP, NTLM, Kerberos, External)
- преобразований контента (LibreOffice, ImageMagick, SWF Tools).

Индексация и поиск – абсолютно необходимые функции в большинстве приложений Alfresco. Доступны две реализации подсистемы поиска: использование встроенной библиотеки индексации Lucene, либо использование внешнего веб-приложения SOLR (который в свою очередь использует Lucene). Эти две реализации различаются своими возможностями, сравнение приведено ниже (Таблица 2).

**Таблица 2. Сравнение реализаций поиска Lucene и SOLR**

Характеристика	Lucene	SOLR
Индексация внутри транзакции	Поддерживается, можно отключить	Не поддерживается
Снижение нагрузки на сервер Alfresco	Нет, работает на том же сервере	Да, можно вынести на другой сервер
Возможность кластеризации для повышения производительности	Кластеризация Alfresco	Независимая кластеризация SOLR

*Индексация внутри транзакции* (или атомарная индексация) позволяет поддерживать индекс в постоянно актуальном состоянии. В некоторых приложениях это может быть необходимо. С другой стороны, если это не является необходимым, то индексация в отложенном времени позволяет снизить время выполнения транзакций и тем самым повысить производительность. Индексация внутри транзакции является практически единственным преимуществом Lucene. В остальных случаях Alfresco рекомендует использовать SOLR.

Какая бы реализация поиска не использовалась, подсистема поиска поддерживает ряд существенных для приложений характеристик:

- проверка прав доступа перед выдачей результатов поиска пользователю;
- поддержка различных языков поиска, в том числе Lucene, CMIS, Full Text Search.

Подсистемы аутентификации позволяют делегировать функцию проверки подлинности пользователей внешним системам. Наиболее часто при интеграции с инфраструктурой предприятия для этих целей используются сервисы службы каталогов, такой как Microsoft Active Directory. При такой интеграции, информация о пользователях импортируется из AD в Alfresco, а аутентификация проходит по одному (или нескольким) из протоколов LDAP, NTLM, Kerberos. Сравнение протоколов аутентификации приведено ниже (

Таблица 3).





Объем БД зависит в основном от примерного объема карточек, например, количество документов  $1000 \times$  количество полей  $20 \times$  средний размер поля  $10$  байт  $\times$  среднее количество версий  $5 \times$  фактор индексации  $2 = 1000 \times 20 \times 10 \times 5 \times 2 = 2\,000\,000$  байт =  $2$  Мб.

Если по документам ставятся процессы, ведется история или хранится другая дополнительная информация, это также следует учитывать.

Объем дискового пространства зависит от размера контента документов, например, количество документов  $1000 \times$  средний размер  $1$  Мб  $\times$  среднее количество версий  $5 = 1000 \times 1 \times 5 = 5\,000$  Мб =  $5$  Гб.

Размер дискового пространства для индекса (lucene, SOLR) вычисляется подобным образом, причем в несколько раз меньше (ориентировочно в  $3$  раза).

### Программные ресурсы

- При стандартной установке Alfresco на Windows-системы создаются следующие службы:
- alfrescoPostgreSQL – для запуска PostgreSQL – СУБД, которая обслуживает базу данных Alfresco;
- alfrescoTomcat – для запуска Apache Tomcat – контейнер сервлетов, который обслуживает веб-приложения Alfresco.

*Если данные имена служб были заняты, то установщик выбирает аналогичные незанятые имена, например, alfrescoPostgreSQL-1 или alfrescoTomcatnum1.*

При установке на Linux-системы создается единственная служба alfresco, которая запускает все необходимые компоненты Alfresco (по умолчанию PostgreSQL и Tomcat).

Компоненты Alfresco занимают по умолчанию следующие сетевые порты TCP (таблица 4).

**Таблица 4. Сравнение различных протоколов аутентификации**

Порт	Компонент	Протокол	Описание
5432	PostgreSQL	PostgreSQL	
8080	Tomcat	HTTP	Порт веб-приложений
8443	Tomcat	HTTPS	Порт веб-приложений
8009	Tomcat	AJP	
8005	Tomcat		Shutdown-порт
8000	Java		Отладочный порт Java (по умолчанию выключен)
7070	VTI	SharePoint	Поддержка онлайн-редактирования
21	Alfresco	FTP	Можно выключить в alfresco-global.properties
445	Alfresco	CIFS	Можно включить в alfresco-global.properties

Порт	Компонент	Протокол	Описание
50500	Alfresco	RMI	Также несколько других портов RMI
8100	LibreOffice	LibreOffice	

*Если данные порты уже заняты другим, то при установке Alfresco можно указать другие порты для PostgreSQL, Tomcat и VTI, а порты Alfresco настроить в файле `alfresco-global.properties`. Если требуется указать другие порты после установки, это также можно сделать, необходимо обратиться к руководствам по настройке компонентов.*

## Методы расширения и настройки Alfresco

Alfresco распространяется в виде WAR-файла (WAR – Web Archive – формат архивов Java) или нескольких WAR-файлов, которые необходимо инсталлировать на сервер(ы) приложений Java (например, Apache Tomcat или JBoss Application Server). Расширения и модули, как правило, запаковываются внутрь веб-приложений, в то время как элементы конфигурации располагаются в отдельном месте – т.н. `shared classpath` (для Tomcat это папка `tomcat/shared`).

Alfresco поддерживает следующие варианты упаковки сторонних расширений:

- не запакованные файлы;
- ZIP-архив;
- JAR-файлы;
- AMP-файлы.

Рекомендуемая техника для упаковки простых расширений – JAR-файлы, рекомендуемая техника для упаковки сложных расширений – AMP-файлы.

JAR-файл (JAR – Java Archive) – это формат архивов Java, он поддерживается всеми серверами приложений Java. JAR-файлы устанавливаются в `shared classpath` или прямо в папку "WEB-INF/lib" веб-приложения. Однако в этом случае не гарантируется их целостность, поэтому рекомендуется эти файлы запаковывать в AMP-архивы.

AMP-файл – это формат модуля Alfresco (AMP – Alfresco Module Package), который также является переименованным ZIP-архивом со специальными соглашениями о внутренней структуре. При установке AMP-файла его содержимое становится частью WAR-файла.

Элементы конфигурации располагаются в `shared classpath`. В частности, основной конфигурационный файл `alfresco` – файл `alfresco-global.properties` располагается в корне `shared classpath` (для Tomcat это папка `tomcat/shared/classes`). Конфигурация других расширений располагается по пути `classpath:alfresco/extension` (для репозитория Alfresco) и `classpath:alfresco/web-extension` (для Alfresco Share). Часто используемые конфигурационные файлы приведены ниже:

- `alfresco/extension/*-log4j.properties` – файлы конфигурации журнала `log4j`;
- `alfresco/extension/subsystems/Authentication` – файлы конфигурации подсистемы аутентификации (например, взаимодействие с MS Active Directory);
- `alfresco/extension/custom-vti*` – файлы конфигурации модуля VTI (поддержка SharePoint Protocol для онлайн-редактирования документа);

- alfresco/web-extension/share-config-custom.xml – custom-конфигурация Share.

### Лог-файлы Alfresco (журналы событий)

Лог-файлы Alfresco в зависимости от версии Alfresco располагаются либо в корневой папке Alfresco, либо в папке tomcat/bin, либо в корне файловой системы (Linux). Alfresco генерирует по одному лог-файлу на веб-приложение. Соответственно, в стандартной инсталляции должны быть лог-файлы:

- alfresco.log – журнал событий репозитория Alfresco
- share.log – журнал событий Alfresco Share (веб-интерфейса)
- solr.log – журнал событий SOLR (сервиса индексации)

Помимо этого, свои журналы ведет Apache Tomcat, они располагаются в папке tomcat/logs:

- catalina.out для Linux, alfrescotomcat-stdout.YYYY-MM-DD.log для Windows – стандартный вывод Apache Tomcat
- localhost\_access\_log.YYYY-MM-DD.txt – журнал обслуженных запросов

Для всех вышеуказанных файлов журналов настроена ротация, каждый день создается и используется новый файл, однако старые файлы не удаляются. Если необходимо удалять и старые файлы, необходимо соответствующим образом скорректировать настройки Alfresco и Tomcat.

## Модули расширения Citeck EcoS

### Трехуровневая структура модулей

Модули расширения предназначены для добавления нового функционала в систему Alfresco. Для того чтобы обеспечить максимальное повторное использование функционала в самых различных областях применения используется трехуровневая структура:

1. модули ядра (core) – содержат базовый функционал, применимый в широком спектре приложений (подобно самой Alfresco);
2. модули приложений (applications) – содержат функционал, применимый только в определенных приложениях управления документами, например, управление договорами (contracts), управление доверенностями (attorneys), управление приказами (orders) и т.п.
3. модули внедрения (custom) – содержат функционал, применимый только в конкретном внедрении системы в конкретной организации. У каждой организации свой набор модулей.

В типовом внедрении системы участвуют модули ядра, один или несколько модулей приложений (можно и без них), и модули внедрения.

Инфраструктура Alfresco позволяет в модулях более специфичных уровней переопределять реализацию и конфигурацию модулей более общих уровней. В частности, в модулях внедрения можно переопределять модули приложений и модули ядра.

### Состав и функции модулей ядра

В состав ядра системы Citeck EcoS входят следующие модули:

- 1st-override-repo;
- 1st-override-share;
- idocs-repo;

- idocs-share.

Модули с суффиксом “-repo” предназначены для установки в веб-приложение alfresco.war (репозиторий Alfresco), модули с суффиксом “-share” предназначены для установки в веб-приложение share.war (Alfresco Share – веб-интерфейс).

Модули с префиксом “1st-override-” предназначены для переопределения файлов Alfresco. Модули с префиксом “idocs-” содержат основной функционал ядра системы Citeck EcoS.

Среди функций, которые добавляет система Citeck EcoS к системе Alfresco, можно отметить следующие основные функции.

**Журналы.** Представляет собой возможность документно-ориентированного просмотра и поиска различных видов документов и других объектов в системе. Отличительная особенность журналов – это учет специфики различных видов контента и отображение только релевантных атрибутов в каждом конкретном случае.

**Оргструктура.** Представляет собой возможность моделирования структуры организации, в которой используется система через встроенный механизм групп Alfresco. Отличительная особенность – возможность пометить группы различными метками, соответствующими различными видам подразделений и должностных лиц. Благодаря использованию механизма групп, имеется возможность выдачи прав на подразделения и должностные лица, а также постановки задач должностным лицам.

**Шаблоны.** Шаблоны *контента* позволяют генерировать по заданному шаблону контент документа. Шаблоны *карточек* позволяют генерировать по заданному шаблону связанные документы, такие как листы согласования, историю доступа и т.п. Шаблоны *уведомлений* позволяют настраивать отправку почтовых уведомлений по заданным событиям. Шаблоны *автонумерации* позволяют генерировать номера документов по заданному шаблону. Также поддерживаются шаблоны в формате docx (MS Word 2007).

**Расширенные возможности процессов.** Возможности по автоматической выдаче прав на время задачи и их отъему после окончания задачи. Возможность приложения документов к задачам. Поддержка заместителей.

**Жизненные циклы.** Возможность простого описания жизненного цикла документов в виде совокупности состояний и переходов между ними. Отличительная особенность компонента жизненных циклов – это простота реализации и расширения жизненных циклов даже после их запуска. Реализованы базовые бизнес-процессы (согласование, подписание и т.п.), из которых можно составлять жизненные циклы документов.

**Поддержка отчетности.** Автоматическая выгрузка информации во внешнюю базу для упрощения построения отчетности с помощью сторонних средств.

**Кейс-менеджмент.** Возможность организовывать кейсы – специальные контейнеры с произвольными типами вложений.

**Интеграция.** Возможность синхронизировать справочники Alfresco с внешними источниками. Поддерживаются такие источники данных как SQL-совместимые БД, XML-файлы,

простые файлы (для импорта контента). Возможна также выгрузка информации во внешние хранилища.

**Карточка документа.** Возможность составлять карточку документа из различных разделов (кардлетов), можно переставлять разделы и указывать произвольные условия их отображения.

**Пользовательский интерфейс.** Различные визуальные компоненты, расширяющие возможности Alfresco Share и повышающие удобство использования системой.

## Состав и возможности модулей приложений

В состав модулей приложений входят следующие модули:

- contracts – управление договорами;
- attorneys – управление доверенностями;
- orders – управление приказами.

Модули приложений разворачивают в системе сайты (сайт договоров, сайт доверенностей и т.п.) и журналы для управления соответствующими видами документов, содержат определение моделей и форм для этих видов документов, определение специальных политик и предустановленных шаблонов, а также некоторые настройки по умолчанию, которые можно переопределять в модулях внедрения.

## Описание сервиса синхронизации

Сервис синхронизации позволяет синхронизировать данные в различных хранилищах данных. Среди таких хранилищ поддерживаются:

- репозиторий Alfresco;
- внешние базы данных;
- папки с XML-файлами;
- папки с произвольными файлами.

Сервис синхронизации оперирует следующими основными абстракциями:

- Object DAO – сервис для доступа к какому-то хранилищу данных, различаются Source DAO (источник данных) и Target DAO (приемник данных);
- Object Type – тип объектов, которым оперирует Object DAO, каждый Object DAO оперирует своим типом объектов, например, объекты репозитория, записи БД, элементы XML и т.п.
- Object Info – информация об объекте типа Object Type, существующем или потенциальном, Object DAO позволяет получить Object Info из Object Type и создать (обновить) Object Type по Object Info;
- Object Converter – преобразует Object Info из формата Source DAO в формат Target DAO;
- Sync Configuration – параметры синхронизации: откуда переносить (Source DAO), как преобразовывать (Object Converter), куда переносить (Target DAO).

Ниже приведена схема потоков данных при синхронизации (Рисунок 3). Преобразование происходит по следующей схеме:

1. Source DAO получает очередной набор объектов для синхронизации (либо все, либо только обновленные с момента последней синхронизации);
2. Source DAO преобразует полученные объекты формат Object Info;

3. Object Converter обеспечивает согласование форматов Object Info между Source DAO и Target DAO; для преобразования отдельных полей могут использоваться дополнительные объекты Object Converter;
4. Target DAO создает или обновляет объекты в соответствии с полученной информацией.

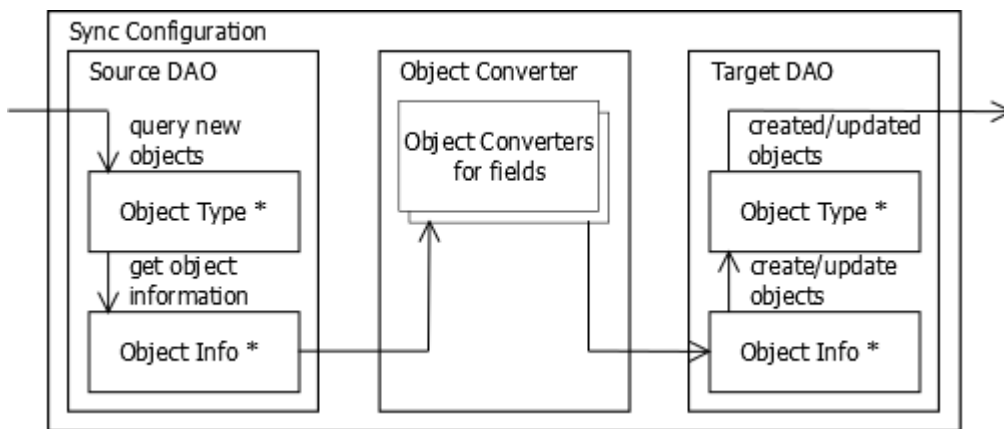


Рисунок 3 – Схема потоков данных при синхронизации

Сервис синхронизации поддерживает загрузку связей объектов благодаря специальным реализациям Object Converter, которые позволяют находить и (или) создавать связанные объекты в соответствии с конфигурацией.

Для повышения скорости импорта/экспорта в сервисе синхронизации поддерживается многопоточная работа и объединение синхронизации нескольких объектов в одной транзакции. Можно регулировать максимальное количество одновременных транзакций (т.е. количество потоков) и максимальное количество объектов в транзакции.