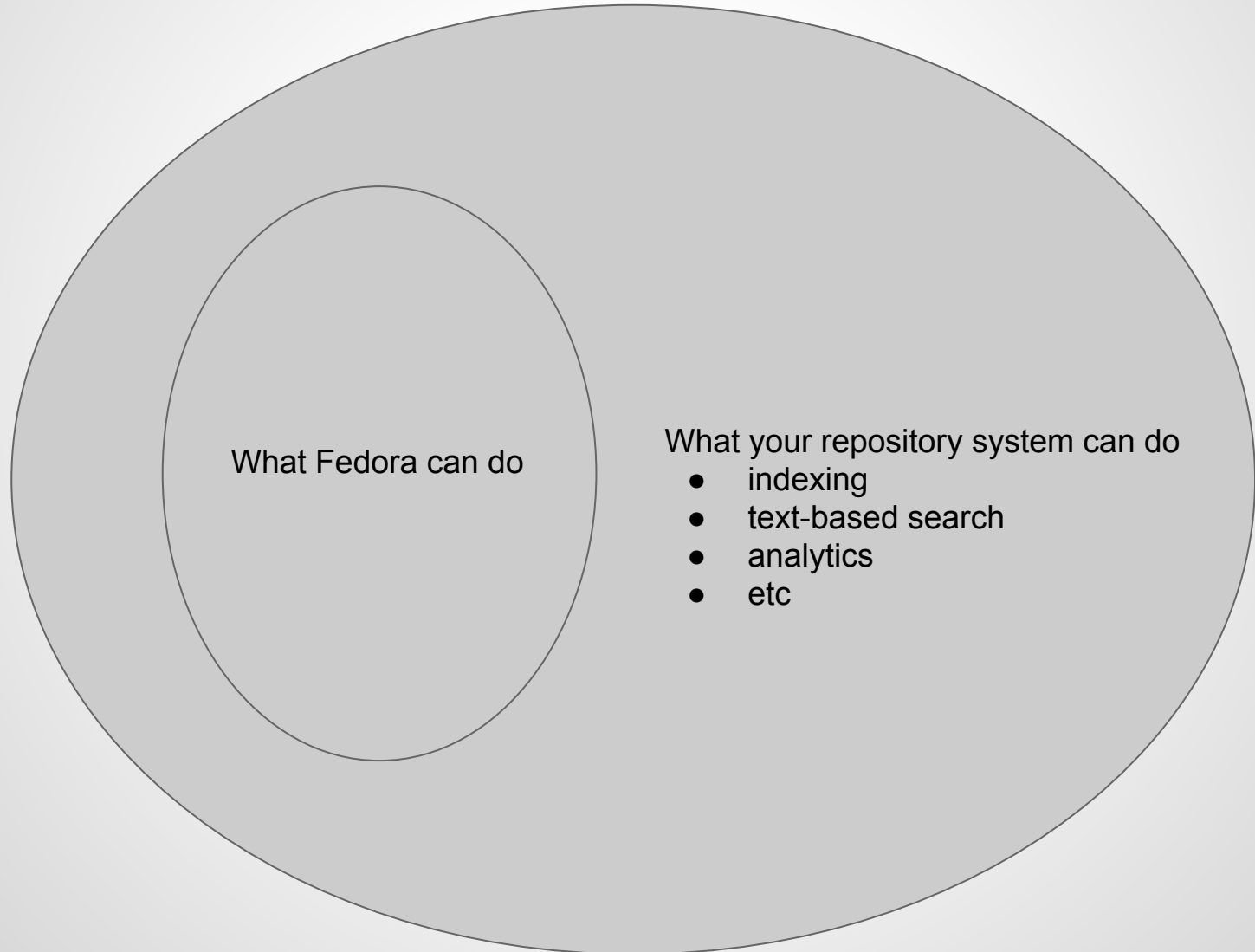


# External Integrations

Introducing Camel

# Repository Scope



# What is Camel?

Good question. See: <http://camel.apache.org/what-is-camel.html>

## **Too many buzzwords - what exactly is Camel?**

Okay, so the description above is technology focused. There's a great discussion about Camel at **Stack Overflow**.

So really you want see this: <http://stackoverflow.com/questions/8845186/what-exactly-is-apache-camel>

# In short...

Camel is a framework for creating small message based applications... and then some.

Camel is middleware for integration with external systems

Camel is all the code you should not have to write in order to work with queues, files, databases, RESTful APIs, common data formats, command line utilities, etc... in a consistent and reliable manner.

# Available Camel Components

<http://camel.apache.org/components.html>

- ActiveMQ
- AWS SQS
- DropBox
- System calls
- Local files
- FTP
- HTTP resources
- LDAP
- SMTP
- SQL
- Twitter
- etc, etc, etc

# Camel can run...

As a stand-alone Java application

In a servlet container like Tomcat or Jetty

In an OSGi runtime such as Karaf

# What is OSGi?

## Open Service Gateway Initiative

Framework for modularizing and deploying Java applications

- Hot deployment
- Automatic reloading of configuration
- Sophisticated dependency resolution
- XML scripting for complex deployments (features)

# Hot Deployment

Bundles can be started, stopped, updated, etc... at runtime!

In other words:

**YOU DO NOT HAVE TO RESTART  
YOUR SERVER TO UPDATE CODE OR  
CONFIGURATION**



# Camel Examples

# Routing

```
from("activemq:topic:fedora")  
    .to("http4:localhost/api");
```

# Transformation

```
from("activemq:topic:fedora")  
  .process(new SparqlUpdater())  
  .to("http4:localhost/api");
```

# Filter

```
from("activemq:topic:fedora")  
  .filter(header("org.fcrepo.jms.identifier").startsWith("/audit"))  
  .to("http4:localhost/api");
```

**Triplestore**

# Fedora 3: internal triplestore

Resource index (mulgara)

Very useful

Very slow

Very resource intensive

# Fedora 4: no internal triplestore

This keeps fedora running faster

There is less code to maintain in the fedora core

The fedora API is simpler

Not everyone uses a triplestore

Decoupled components are easier to scale out

# Mirror Fedora content in a triplestore

Two approaches:

- 1) Synchronous: client(s) operate on content in fedora and then do the same in a triplestore
- 2) Asynchronous: client(s) operate on content only in fedora, a separate process synchronizes changes

[OBJ]

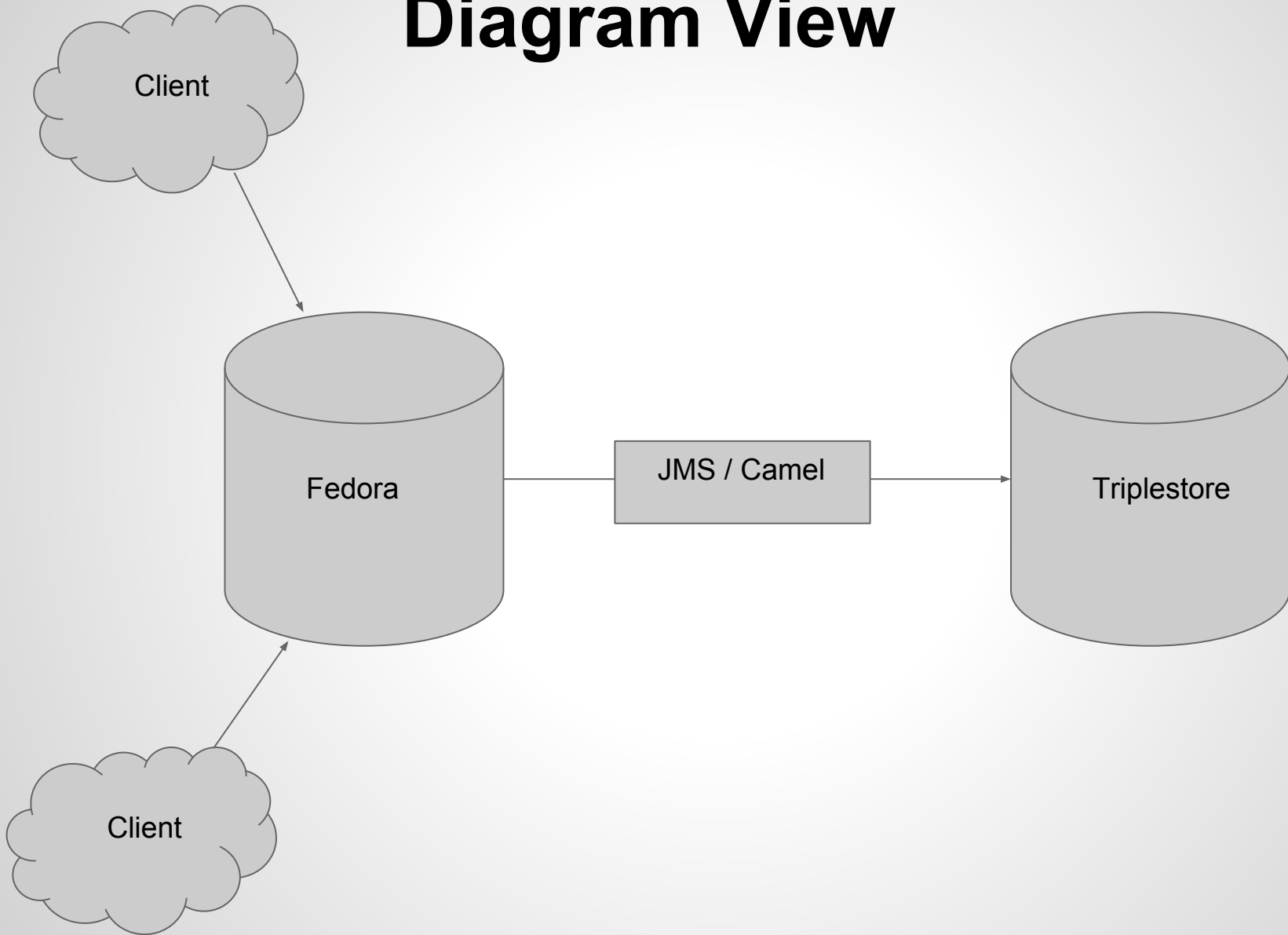


# Asynchronous Patterns

Camel recipe for this (DSL for integration patterns)

`fcrepo-indexing-triplestore` in [github.com/fcrepo4-exts/fcrepo-camel-toolbox](https://github.com/fcrepo4-exts/fcrepo-camel-toolbox)

# Diagram View



# Indexing to a Triplestore

```
from("activemq:topic:fedora")
    .to("fcrepo:{{fcrepo.baseUrl}}?accept=application/n-triples")
    .process(new SparqlUpdateProcessor())
    .to("http4://{{triplestore.baseUrl}}");
```

# Choose your own triplestore

Fuseki comes with the fcrepo-vagrant VM

In production, any will work. These have been tested:

- Sesame/OpenRDF
- Jena/Fuseki
- Blazegraph

...just update the location of the triplestore in the camel configuration

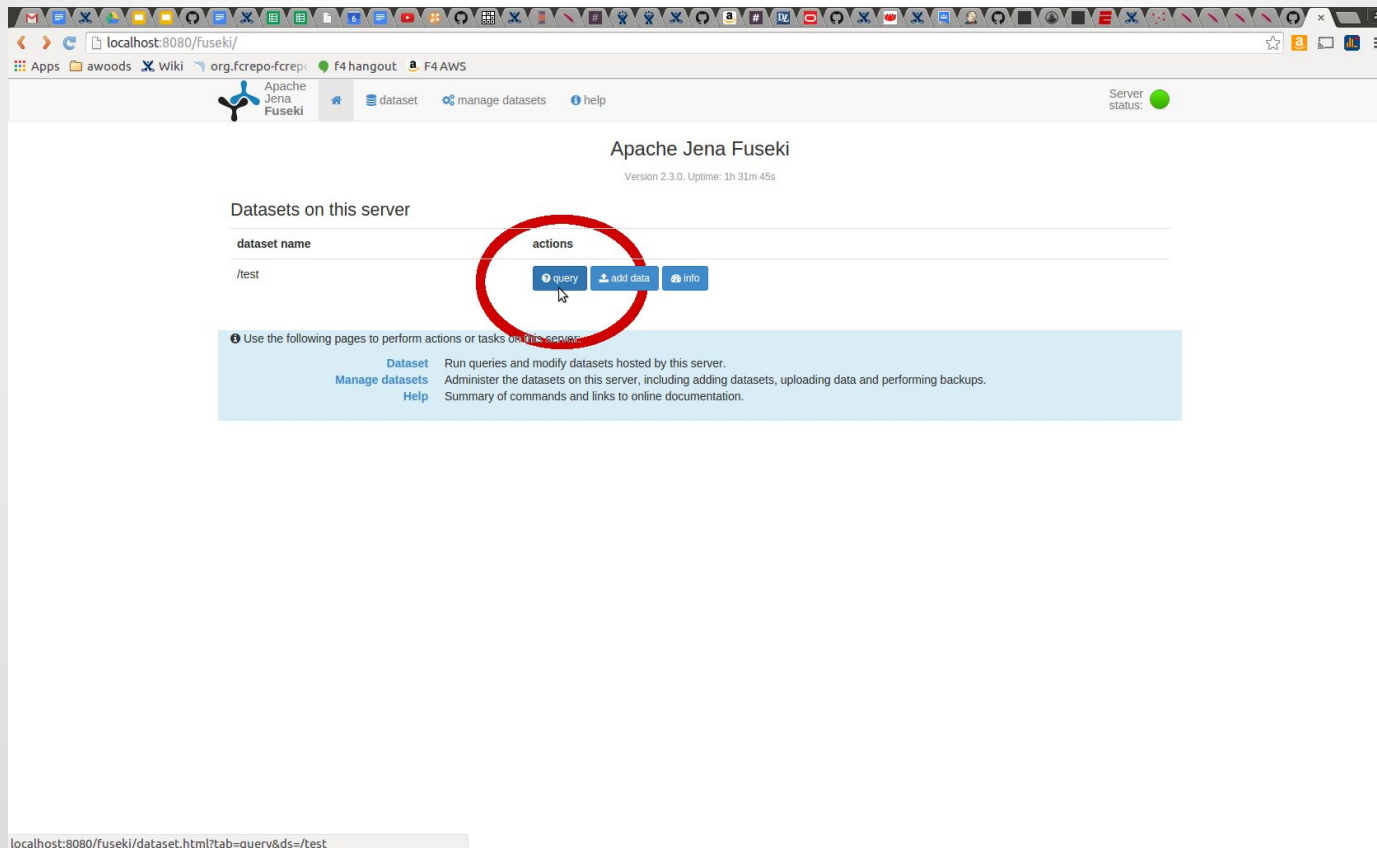
# SPARQL & RDF Triples: S P O

Syntax like English (sort of):

- Statements end with a period
- Clauses end with a semicolon
- Lists are separated with a comma

# Hands-On: Indexing in triplestore

http://localhost:8080/fuseki



The screenshot shows the Apache Jena Fuseki web interface. The browser address bar displays `localhost:8080/fuseki/`. The page header includes the Apache Jena Fuseki logo, navigation links for `dataset`, `manage datasets`, and `help`, and a `Server status: ●` indicator. The main content area is titled `Apache Jena Fuseki` with the version `Version 2.3.0. Uptime: 1h 31m 45s`. Below this, a section titled `Datasets on this server` contains a table with the following structure:

dataset name	actions
/test	<a href="#">query</a> <a href="#">add data</a> <a href="#">info</a>

A red circle highlights the `actions` column header and the `query`, `add data`, and `info` buttons for the `/test` dataset. Below the table, a light blue box contains instructions: `Use the following pages to perform actions or tasks on the server:`

- [Dataset](#) Run queries and modify datasets hosted by this server.
- [Manage datasets](#) Administer the datasets on this server, including adding datasets, uploading data and performing backups.
- [Help](#) Summary of commands and links to online documentation.

The browser's address bar at the bottom shows the URL `localhost:8080/fuseki/dataset.html?tab=query&ds=/test`.

# Hands-On: Indexing in triplestore

```
select * where {  
  <http://localhost:8080/fcrepo/rest/cover> ?p ?o  
}
```

# Hands-On: Indexing in triplestore

PREFIX ldp: <http://www.w3.org/ns/ldp#>

PREFIX ebucore: <http://www.ebu.  
ch/metadata/ontologies/ebucore/ebucore#>

```
select * where {  
  ?s ldp:contains ?o .  
  ?o ebucore:hasMimeType ?m  
}
```



# Hands-On: Indexing in triplestore

## *Audit*

prefix premis: <http://www.loc.gov/premis/rdf/v1#>

prefix xsd: <http://www.w3.org/2001/XMLSchema#>

select ?s ?d where {

  ?s ?p <http://fedora.info/definitions/v4/audit#InternalEvent> .

  ?s premis:hasEventRelatedObject <http://localhost:8080/fcrepo/rest/cover> .

  ?s premis:hasEventDateTime ?d .

  FILTER (?d > "2015-10-06T04:21:14Z"^^xsd:dateTime)

}

**Solr**

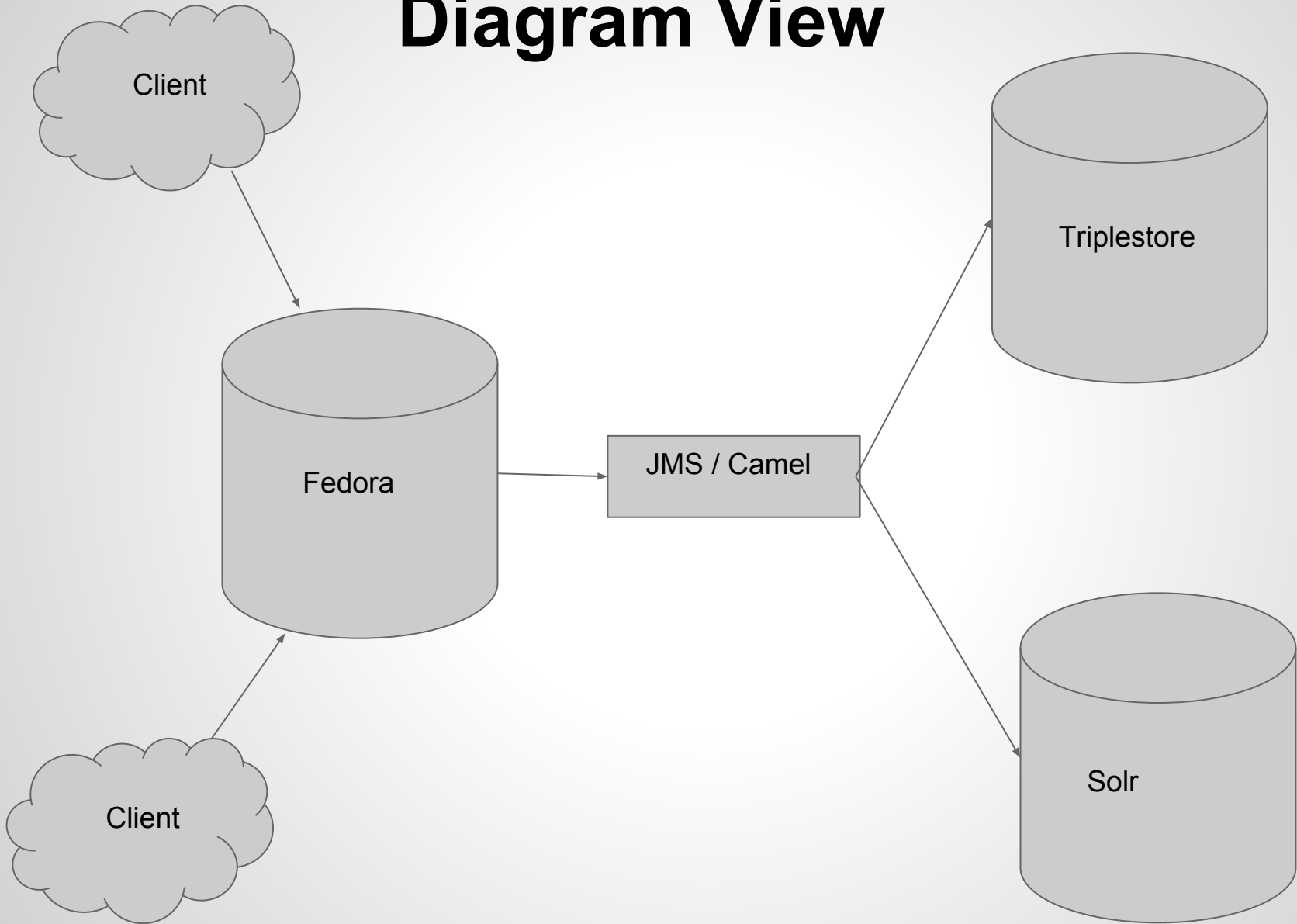
# Indexing to Solr

```
from("activemq:topic:fedora")
    .choice()
        .when(header("org.fcrepo.jms.eventType")
            .isEqualTo("http://fedora.info/definitions/v4/repository#NODE_REMOVAL")
            .to("direct:delete")
        .otherwise()
            .to("direct:index"));

from("direct:delete")
    .process(new SolrDeleteProcessor())
    .to("http4://localhost:8983/solr/collection1/update");

from("direct:index")
    .to("fcrepo:localhost:8080/rest?transform=default")
    .to("http4://localhost:8983/solr/collection1/update");
```

# Diagram View



# Hands-On: Indexing in Solr

<http://localhost:8080/solr>

The screenshot displays the Apache Solr Admin UI. The left sidebar contains navigation links: Dashboard, Logging, Core Admin, Java Properties, Thread Dump, and Core Selector. The Core Selector is highlighted with a red circle and shows a dropdown menu with 'collection1' selected. The main content area is divided into several sections:

- Instance:** Shows the instance started 'about an hour ago'.
- Versions:** Lists installed versions:

Component	Version	Timestamp
solr-spec	4.10.3	
solr-impl	4.10.3	1644336 - mark - 2014-12-10 00:35:44
lucene-spec	4.10.3	
lucene-impl	4.10.3	1644336 - mark - 2014-12-10 00:28:00
- JVM:** Shows runtime details for Oracle Corporation Java HotSpot(TM) 64-Bit Server VM (1.8.0\_60 25.60-b23).

Property	Value
Runtime	Oracle Corporation Java HotSpot(TM) 64-Bit Server VM (1.8.0_60 25.60-b23)
Processors	1
Args	-Djava.io.tmpdir=/tmp/tomcat7-tomcat7-tmp -Dcatalina.home=/usr/share/tomcat7 -Dcatalina.base=/var/lib/tomcat7 -Djava.endorsed.dirs=/usr/share/tomcat7/endorsed -Dfcrepo.audit.container=/audit -XX:+UseConcMarkSweepGC -Xmx128m -Djava.awt.headless=true -Djava.util.logging.manager=org.apache.juli.ClassLoaderLogManager -Djava.util.logging.config.file=/var/lib/tomcat7/conf/logging.properties
- System:** Displays various system metrics with progress bars:
  - Physical Memory: 62.4% (1.22 GB / 1.96 GB)
  - Swap Space: NaN%
  - File Descriptor Count: 3.5% (143 / 4096)
  - JVM-Memory: 65.4% (80.99 MB / 123.75 MB)

At the bottom, there are links for Documentation, Issue Tracker, IRC Channel, Community forum, and Solr Query Syntax.

# Hands-On: Indexing in Solr

The screenshot shows the Apache Solr Admin UI for a collection named 'collection1'. The left sidebar contains a navigation menu with the following items: Dashboard, Logging, Core Admin, Java Properties, Thread Dump, collection1 (selected), Overview, Analysis, Dataimport, Documents, Files, Ping, Plugins / Stats, Query (circled in red), Replication, and Schema Browser. The main content area is divided into several sections:

- Statistics:** Last Modified: about 12 hours ago, Num Docs: 7, Max Doc: 11, Heap Memory Usage: 35136, Deleted Docs: 4, Version: 21, Segment Count: 6. Includes an 'optimize now' button.
- Instance:** CWD: /var/lib/tomcat7, Instance: /var/lib/tomcat7/solr/collection1, Data: /var/lib/tomcat7/solr/collection1/data, Index: /var/lib/tomcat7/solr/collection1/data/index, Impl: org.apache.solr.core.NRTCachingDirectoryFactory.
- Replication (Master):** A table showing replication status for the Master (Searching) and Master (Replicable).
- Healthcheck:** Ping request handler is not configured with a healthcheck file.

At the bottom of the page, there are links for Documentation, Issue Tracker, IRC Channel, Community forum, and Solr Query Syntax. The browser address bar shows localhost:8080/solr/#/collection1 and the status bar shows localhost:8080/solr/#/collection1/query.

**Success!**