JUMPSTART: THE BASICS FOR GETTING STARTED WITH MARKLOGIC

Ruth Stryker, Senior Courseware Developer and Technical Instructor, MarkLogic
So we know that MarkLogic…

- Is an enterprise NoSQL database
- Can be used for storing and searching data in a wide variety of structures and types
- Can be used for all sorts of applications
- Can be used to integrate data from silos
Let’s get started with MarkLogic as we go through…

- Deciding on a project
- Thinking about your application code
- Creating a database
- Bringing your data in
- Querying your data
- Using your data in an application
Yell.com
Small business directory
Aetna
Large health insurer with 35,000 employees, integrating more than 140 data feeds
MLU Training registration system
Organization for Economic Cooperation and Development

34 country organization that promotes policies for improving the economic and social well-being of people around the world
Deciding on a project

- Have a clear vision for it
- Ideal would be to spec it out
- Perhaps start with a small project
News
Source articles from BBC

Uganda's David Sejusa: 'Oppose Museveni's monarchy'
Renegade Ugandan army General David Sejusa has accused President Yoweri Museveni of creating a "political monarchy" to hang on to power. In his first interview since he fled to the UK last month, the four-star general told the BBC that Mr Museveni's plans should be opposed. 2013-06-18T16:20:10

G8 backs urgent Syria peace talks in Geneva
G8 leaders meeting in Northern Ireland have backed calls for Syrian peace talks to be held in Geneva "as soon as possible". After adopting a statement, UK Prime Minister David Cameron said the leaders had managed "to overcome fundamental differences". 2013-06-18T16:34:33

Arrest in death of US man dribbling football to Brazil
An Oregon motorist suspected of running down a US adventure seeker on a quest to dribble a football from Seattle to Brazil for charity has been arrested. Scott Van Hiatt is
News

Source articles from BBC
Steps

✓ Deciding on a project
  • Think about your application code
  • Creating a database
  • Bringing your data in
  • Querying your data
  • Using your data in an application
Think about your application code

- Data will be in a MarkLogic database
- Application code can be inside or outside of MarkLogic
Steps

✓ Deciding on a project

✓ Think about your application code
  - Creating a database
  - Bringing your data in
  - Querying your data
  - Using your data in an application
Install MarkLogic and start it running
Install MarkLogic and start it running
Create a database

- Can do through graphical interface or programmatically
Create a database
Where data is stored

- The data for a database is stored on the file system in a compressed format in one or more forests.

- A database must have at least one forest attached to it, to be able to load data into the database.

- Database is logical.

- Forests are physical (a forest is a directory, or folder, on the file system).
Create a forest
Attaching a forest to a database
Steps

✓ Deciding on a project
✓ Think about your application code
✓ Creating a database
  ▪ Bringing your data in
  ▪ Querying your data
  ▪ Using your data in an application
Bringing your data in

- Data is modeled/stored as a document
- Document types are JSON, XML, text, binary, and RDF triple
- Different types can all be in the same database
- Documents can be loaded into a database through a variety of MarkLogic interfaces
Example: JSON document

```json
{
    "article": {
        "title": "Angelina Jolie mobbed by media on Greece migrant trip",
        "date": "2016-03-16",
        "content": "Angelina Jolie was mobbed by the media while visiting a migrant reception center in Piraeus, near Athens, Greece. The actor was there in her role as Special Envoy of the UN High Commissioner for Refugees, highlighting the migrant crisis."
    }
}
```
Example: XML document

<article>
  <title>Angelina Jolie mobbed by media on Greece migrant trip</title>
  <date>2016-03-16</date>
  <content>Angelina Jolie was mobbed by the media while visiting a migrant reception centre in Piraeus, near Athens, Greece. The actor was there in her role as Special Envoy of the UN High Commissioner for Refugees, highlighting the migrant crisis.</content>
</article>
Text documents

- Have no structure
- But can be searched
Binary documents

- Includes most other types (e.g. images, video, etc.)
- Cannot be searched
- But any properties (e.g. metadata) included with the document can be searched
- Documents in some binary formats (e.g. PDF or Word) can be converted to an XML document while being loaded into a MarkLogic database by using the Content Processing Framework
Example: RDF triple (for semantics)

<sem:triple>

<sem:subject>
  http://dbpedia.org/resource/Greece
</sem:subject>

<sem:predicate>
  http://www.w3.org/1999/02/22-rdf-syntax-ns#type
</sem:predicate>

<sem:object>
  http://dbpedia.org/ontology/Country
</sem:object>

</sem:triple>
Semantic News Application

Media playback is unsupported on your device.

Angelina Jolie mobbed by media on Greece migrant trip

16 March 2016 Last updated at 17.11 GMT

Angelina Jolie was mobbed by the media while visiting a migrant reception centre in Piraeus, near Athens, Greece.
Angelina Jolie mobbed by media on Greece migrant trip

16 March 2016 Last updated at 17.11 GMT

Angelina Jolie was mobbed by the media while visiting a migrant reception centre in Piraeus, near Athens, Greece.
Angelina Jolie mobbed by media on Greece migrant trip

16 March 2016 Last updated at 17:11 GMT

Angelina Jolie was mobbed by the media while visiting a migrant reception centre in Piraeus, near Athens, Greece.

Country
Greece
Currency
Euro
Government
Republic
Description
; Greek: Ελλάς), is a country in Southern Europe and Balkans. According to the 2011 census, Greece's population is around 11 million. Athens is the nation’s capital and largest.
News

Uganda's David Sejusa: 'Oppose Museveni's monarchy'
Renegade Ugandan army General David Sejusa has accused President Yoweri Museveni of creating a "political monarchy" to hang on to power. In his first interview since he fled to the UK last month, the four-star general told the BBC that Mr Museveni's plans should be opposed. 2013-06-18T16:20:10

G8 backs urgent Syria peace talks in Geneva
G8 leaders meeting in Northern Ireland have backed calls for Syrian peace talks to be held in Geneva "as soon as possible". After adopting a statement, UK Prime Minister David Cameron said the leaders had managed "to overcome fundamental differences". 2013-06-18T16:34:33

Arrest in death of US man dribbling football to Brazil
An Oregon motorist suspected of running down a US adventure seeker on a quest to dribble a football from Seattle to Brazil for charity has been arrested. Scott Van Hiatt is
Source documents to load

C:\mls-projects\news\docs_to_load
Loading documents

Can be done through a variety of MarkLogic interfaces, including:

- MarkLogic Content Pump
- Query Console (XQuery or JavaScript)
- REST API
- Node.js API or Java API
- Content Processing Framework
Using MarkLogic Content Pump

- Java-based command line tool
- Requires:
  - JRE 1.6 or later
  - MLCP files (can be downloaded from developer.marklogic.com/products/mlcp)
  - A port for connecting to the database
Creating a REST API Instance

- Basically, setting up a port for being able to connect to a MarkLogic database
- Benefits of a REST API instance:
  - Can load documents through it
  - Can read documents
  - Can search the database
  - And more…
- To create a REST API instance, need to send a configuration file to MarkLogic that specifies a port to set up
Creating a REST API Instance

Example configuration file:

```xml
<rest-api xmlns="http://marklogic.com/rest-api">
  <name>8011-news-rest</name>
  <database>news</database>
  <port>8011</port>
</rest-api>
```
Creating a REST API Instance

Example of sending the configuration file to MarkLogic:

```bash
curl --anyauth --user admin:admin
```
Seeing that a REST API Instance has been set up
Loading documents with MLCP

```
mlcp.bat import -host localhost -port 8011 -username admin -password admin -input_file_path C:\mls-projects\news\docs_to_load -mode local -input_file_type documents -output_collections "http://www.bbc.co.uk/news/content" -output_uri_replace "C:/mls-projects/news/docs_to_load,'content/news'"
```
Viewing database documents
Steps

✓ Deciding on a project
✓ Think about your application code
✓ Creating a database
✓ Bringing your data in
  ▪ Querying your data
  ▪ Using your data in an application
Searching a database

Can be done through any of the programming interfaces, including:

- Query Console (XQuery or JavaScript, or SPARQL for triples)
- REST API
- Node.js API or Java API
- Hosted XQuery or JavaScript application
Using XQuery cts:search()

```xquery
cts:search(fn:doc(), "earthquake")[1 to 10]
```
Using XQuery cts:search() in a for loop

```
declare namespace ns = "http://www.w3.org/1999/xhtml";
for $doc in cts:search(fn:doc(), "earthquake")[1 to 10]
return $doc/ns:title/text()
```
Using XQuery cts:search()

```
declare namespace ns = "http://www.w3.org/1999/xhtml";
for $title in cts:search(fn:doc()//ns:title, "earthquake")[1 to 10]
  return $title/text()
```

BBC News - Predicting earthquakes and saving lives - with smartphones

BBC News - Indonesia Aceh rescue operation after deadly earthquake
Using JavaScript cts.search()
Using JavaScript cts.search()
Using XQuery search:search()
Snippets and highlighting in use

- Predicting earthquakes and saving lives - with smartphones
  - BBC News - Predicting earthquakes and saving lives - with smartphones
  - Earthquakes and saving lives - with smartphones...
  - In 1989, the strongest... 2013-07-09T01:22:25

- Distant quakes 'can trigger wastewater-site temblors'
  - Earthquakes can be triggered at the sites of wastewater injection by quakes on the other side...
  - Natural earthquakes triggering is not new; in hydrothermal and volcanic areas, tremors can be... 2013-07-12T02:38:35
Using JavaScript `search.search()`
JSearch

- Designed for JavaScript developers searching JSON documents
- Features:
  - Input and output is in the form of JavaScript objects
  - A fluent call chain pattern that is natural for JavaScript
  - Powerful convenience methods for operations such as snippet generation and faceting

```javascript
var sr = require('/MarkLogic/jsearch.sjs');
var output = sr.documents()
  .where(cts.jsonPropertyValueQuery('city', 'london'))
  .result();
output;
```
Using the REST API for search

Steps

✓ Deciding on a project
✓ Think about your application code
✓ Creating a database
✓ Bringing your data in
✓ Querying your data
  ▪ Using your data in an application
Using your data in an application

- Understanding the application programming interfaces
- Example: Standing up an XQuery application
- Seeing the full-feature application
Understanding the application programming interfaces

- If you write your application in **XQuery** or **JavaScript**, the application code can be **hosted** by MarkLogic.

- If you write your application in **Java** or **Node.js**, it can communicate with your MarkLogic database through the Java API or Node.js API.

- If you write your application in any **other** language (that can make HTTP requests, such as C#, Python, PHP, etc.), it can communicate with your MarkLogic database through the **REST API**.
Example: Setting up an XQuery application

- Have a location for your XQuery application code
- Create an HTTP App Server that:
  - Has a port
  - References the location of the application code
  - References the database
Have a location for your XQuery application code

Can be on the file system, or in a second MarkLogic database (a.k.a. a ‘modules’ database)
Create an HTTP App Server
See that the application page displays

BBC News - Predicting earthquakes and saving lives - with smartphones
BBC News - Indonesia Aceh rescue operation after deadly earthquake
Add the rest of the application files

News

Predicting earthquakes and saving lives - with smartphones
BBC News - Predicting earthquakes and saving lives - with smartphones Predicting earthquakes and saving lives - with smartphones ... meaning it is prone to earthquakes. In 1989, the strongest... 2013-07-09T01:22:25

Distant quakes 'can trigger wastewater-site temblors'
Earthquakes can be triggered at the sites of wastewater injection by quakes on the other side... ...of natural earthquake triggering is not new; in hydrothermal and volcanic areas, tremors can be... 2013-07-12T02:38:35
Use the application

Distant quakes 'can trigger wastewater-site temblors'

Earthquakes can be triggered at the sites of wastewater injection by quakes on the other side of the world, research suggests.

The injection of wastewater from underground operations such as oil drilling is known to increase local seismic activity.
Understand the application with semantics
Angelina Jolie mobbed by media on Greece migrant trip

16 March 2016 Last updated at 17.11 GMT

Angelina Jolie was mobbed by the media while visiting a migrant reception centre in Piraeus, near Athens, Greece.
Angelina Jolie mobbed by media on Greece migrant trip

16 March 2016 Last updated at 17.11 GMT

Angelina Jolie was mobbed by the media while visiting a migrant reception centre in Piraeus, near Athens, Greece.
Angelina Jolie mobbed by media on Greece migrant trip

16 March 2016 Last updated at 17:11 GMT

Angelina Jolie was mobbed by the media while visiting a migrant reception centre in Piraeus, near Athens, Greece.

http://dbpedia.org/resource/Greece

Country
Greece
Currency
Euro
Government
Republic
Description
Greece (Greek: Ελλάς, romanized: Ellás), is a country in Southern Europe and the Balkans. According to the 2011 census, Greece’s population is around 11 million. Athens is the nation’s capital and largest.
Can access the project files on GitHub

- Simple project:
  
github.com/mlruthstryker/simple-news-app

- Advanced project (with semantics functionality):
  
github.com/marklogic/semantic-news-search
Steps

✓ Deciding on a project
✓ Thinking about your application code
✓ Creating a database
✓ Bringing your data in
✓ Querying your data
✓ Using your data in an application
What to do next

- Attend the MarkLogic Fundamentals course (hands-on)
  marklogic.com/training

- Read the MarkLogic guides
  docs.marklogic.com

- Watch feature-focused training videos
  mlu.marklogic.com/ondemand

- Ask questions on the developer mailing list
  developer.marklogic.com/discuss
Thanks!

ruth.stryker@marklogic.com