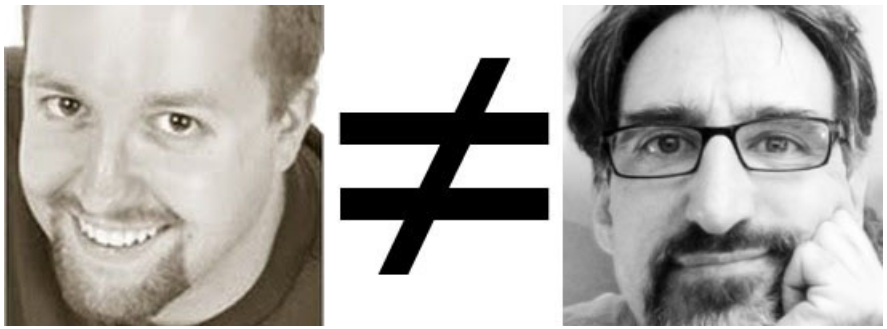
Corona: XML and JSON via REST

Norman Walsh, Lead Engineer
11 Feb 2012

Introduction

Preamble

- The observant among you will have noticed that I am not Jason Hunter.



- Jason and the Corona dev team get the credit.
- Any errors, omissions, gaffes, and general cluelessness are mine.

MarkLogic™

Background

- MarkLogic Server is a database purpose-built for Big Data.
- It's document-centric, transactional, search-centric, structure-aware, schema-agnostic, high performance, and clusters on commodity hardware.
- MarkLogic specializes in a new type of indexing that enables ad hoc queries against documents with widely varying schemas, delivering subsecond answers.

MarkLogic™

I say

It's the best thing ever because…

- It's a native XML repository (with support for text and binary too)
- It has a built in application server
- It supports XQuery and XSLT as native programming languages.

MarkLogic™

Your enterprise dev says

Yeah, but, ...

- XML is hard. Namespaces are stupid. I don't need no steekin' XML.
- JSON is cool. I'm using JSON in the front end. I want to use JSON.
- What's an XQuery?

MarkLogic™

The Challenge

- Expose the core MarkLogic functionality—the important things developers need— as a set of services callable from other languages.
- Let users be successful without knowing XQuery or having special knowledge about MarkLogic's internals.
- Allow users to store XML or JSON without having to worry about it.

MarkLogic™

Corona

- Corona is an attempt to meet that challenge.
- Corona is built as a set of web endpoints exposing the core MarkLogic application services.
- Using REST principles means that the Corona functionality plays nicely with load balancers, proxies, etc.
- Current focus is on the REST layer; native language bindings are expected to come later.

MarkLogic™

Play along at home!

- Corona is available today: https://github.com/marklogic/Corona
- API Docs: https://github.com/marklogic/Corona/wiki
- MarkLogic is available today: http://community.marklogic.com/

MarkLogic™

Exploring Corona

MarkLogic™

Corona User Roles

- Corona Developer: The developer using Corona via the REST API.
- Corona Admin: A developer with more access to the Corona back end (to adjust index settings, update stored transformations, etc.).
- MarkLogic Admin: The classic IT database administrator; they install the server, manage uptime, perform backups, etc. They don't need to know anything about the Corona applications.

MarkLogic™

Storing Documents

What kind of documents? Corona supports

- XML and text documents.
- Binaries, from which metadata may automatically be extracted.
- JSON, mapped to XML behind the scenes in a format designed for efficient queries.

MarkLogic™

Technical aside: JSON→XML

```
{
  "": "Empty key.",
  "list":
          [1,
           false,
           null],

  "1": "one",
  "a+b": "c"
}
```

```
<json type="object"
  xmlns="http://marklogic.com/json
  <_ type="string">Empty key.</_>
  <list type="array">
    <item type="number">1</item>
    <item boolean="false"/>
    <item type="null"/>
  </list>
  <_0031 type="string">one</_0031>
  <a_002Bb type="string">c</a_002B
</json>
```


MarkLogic™

Storing Documents

What's a document?

- A unique name: a URI.
- Permissions: security roles that determine who, and what kind, of access is allowed to the document.
- Properties: arbitrary metadata as key/value pairs.
- Collections: a set of URI-named collections (distinct from the implicit grouping determined by the URI path of the document's name).
- A quality: an integer representing the intrinsic relevance of the document.

MarkLogic™

Retrieving Documents

If you know the name of a document, a simple HTTP GET request will return the whole thing. But sometimes it's useful to retrieve only part of a document:

- For XML documents, a subset of XPath.
- For JSON documents, an ad-hoc notation that looks something like JavaScript object traversal.

Transforming Documents

- Documents (or portions of documents) can be transformed automatically on retrieval with named XSLT or XQuery transformations.
- Documents can also be transformed on insert.
- For security reasons, only a Corona Admin can add new transformations.

MarkLogic™

Powerful searching

- Search in traditional database fashion with value, range, and geospatial constraints.
- Search in "search engine" fashion with free-text, language-aware constraints.
- Present results in relevance order, or (soon) ordered by a scalar such as date or price.
- Paginate results with snippeting and highlighting.

MarkLogic™

Search queries (1/2)

- Key/value queries: find documents that have a key (JSON key, XML element) equal to a specific value.
- String queries: user-friendly search syntax such as provided by search engines:

```
winter storm
```

or

```
winter NEAR storm
```

or

```
winter NEAR storm
(title:Lebowski OR title:Country OR title:Fargo)
AND (cast:Buscemi OR cast:Jones)
-director:Ethan
```


MarkLogic™

Search queries (2/2)

- Structured queries: programmer-friendly search syntax using hierarchical constraints expressed in XML or JSON:

```
{"and": [
  { "element": "author",
    "equals": "Noam Chomsky"  // Atomic or array
  },
  { "range": "price", "from": 10.00, "to": 14.99 },
  {
    "geo": "location", // The name of the geo index
    "region": { "polygon": [
      {"point": {"latitude": 1, "longitude": -1}},
      {"point": {"latitude": 1, "longitude": 1}},
      {"point": {"latitude": -1, "longitude": 1}},
      {"point": {"latitude": -1, "longitude": -1}},
      {"point": {"latitude": 1, "longitude": -1}}
    ]} } ]}
```


MarkLogic™

Search configuration

The Corona Admin can define:

- Places: names assigned to sets of locations in a document (grouping, for example, `title`, `rdf:title`, and `atom:title` into a single "title" place).
- Ranges: names assigned to a JSON key or XML node that should be treated as a scalar value.
- Named queries: for frequently used queries. Eventually, it will be possible to use named queries for "reverse queries".
- Facets: analytics about search results: a range index plus an optional query returns all the distinct values (or buckets) for documents matching the query as well as a count of the frequency.

MarkLogic™

Transaction support

Transactions allow a Corona developer to group several API calls into a single, atomic unit:

- Transaction start
- *`Other API calls here`*
- Transaction commit

  OR

  Transaction rollback

MarkLogic™

Demo

If there's time…

Q&A/Discussion

Does Corona meet the challenge?