



# MarkLogic Enterprise NoSQL

Do more with less!

*Mats Stellwall – Principal Sales Engineer*

*Rickard Fernström – Nordic Sales Manager*

© COPYRIGHT 2015 MARKLOGIC CORPORATION. ALL RIGHTS RESERVED

Not only SQL

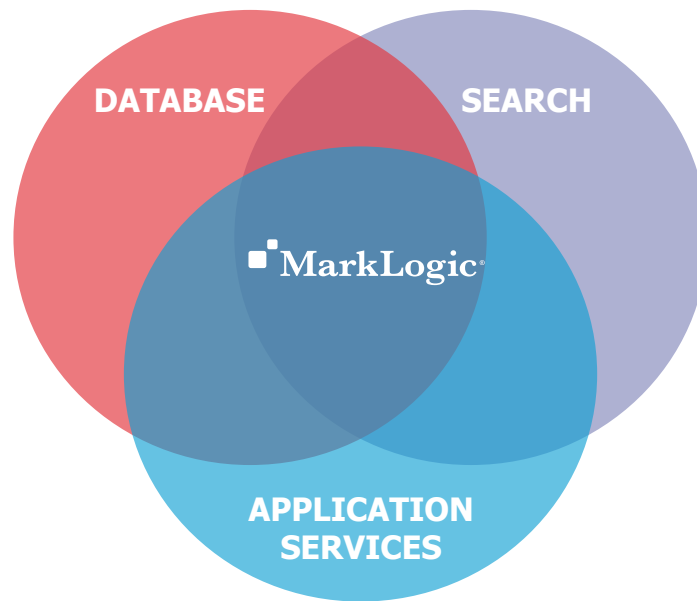
- **Key-Value**
- **Graph/Triple Store**
- **Document Oriented**
- **Column**

# NoSQL Databases

- **Schema Flexibility:** Ability to query multiple schema at run
- **Free of Complex Joins:** Generally-designed so joins aren't necessary or recommended
- **Horizontally Scalable:** Easy to incorporate new machines (of comparable power) into cluster
- **Compatible with Commodity Hardware:** Most NoSQL database systems run on commodity hardware
- **Self-contained:** Nothing is shared – each node is a stand-alone with respect to storage and processing power.
- **Rapid Application Development:** Data is always ready – whenever you are

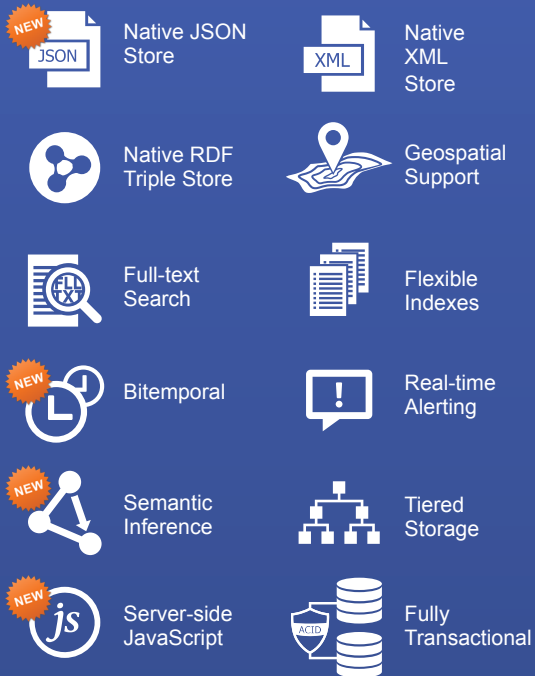
# The Only Enterprise NoSQL Database

- Search & Query
- ACID Transactions
- High Availability / Disaster Recovery
- Replication
- Government-grade Security
- Scalability & Elasticity
- On-premise or Cloud Deployment
- Hadoop for Storage & Compute
- Semantics/triple store
- Faster Time-to-Results



# MarkLogic / Enterprise NoSQL Database Platform

## POWERFUL



## AGILE



## TRUSTED



# MarkLogic Product evolution for over a decade

## Cerisent XQE Server 1

- ACID Transactions
- Text Based Search
- Backup & Restore
- Linux Support
- Web-based Protocols
- HTTP & XDBC
- XQuery

## MarkLogic Server 3

- 11 Advanced Search Features
- Content processing (including PDF, Word, Excel, PPT)
- HTTP calls
- Failover
- Support for Linux, Windows Server, .NET

## MarkLogic Server 4

- Alerting
- Entity Enrichment
- Geospatial
- Analytics (co-occur, value lexicons, bucketing)
- Modular documents
- Security auditing
- HA: forest-level failover

## MarkLogic Server 5

- Complete Enterprise Roadmap
- Database Replication
- Multi-statement and distributed transactions
- Point-in-time recovery
- Start Hadoop Roadmap
- Hadoop Connector

## MarkLogic 7

- Semantics Foundation
- Next-gen Infrastructure Support
- Elasticity
- Tiered Storage
- Continue Hadoop Roadmap
- Run on HDFS

## MarkLogic Server 2

- Clustering
- Role-based security w/ BASIC authentication
- Document Collections
- Enhanced Search (stemming, thes., wildcard)
- WebDAV support
- Document locking
- Enhanced XDBC support

## MarkLogic Server 3.1

- 15 Advanced Search Features
- Wildcard queries
- Directories
- Forward Compatibility
- Support for Sun Solaris
- XCC

## MarkLogic Server 4.2

- Replication
- Failover
- Database Rollback
- Compartment Security
- Search Optimizations
- Search API
- Information Studio
- Application Builder

## MarkLogic 6

- Accessibility
- SQL/BI
- Java/REST/JSON
- UDFs/Analytics
- mlcp
- Hadoop Distributions
- HDFS Tech Preview

## MarkLogic 8

- Developer Ease of Use
- JSON Storage
- Native JavaScript
- SPARQL 1.1
- Inferencing
- Bitemporal
- Prescriptive Sizing
- Ref. Architecture

2003

2004

2005

2006

2008

2010

2011

2012

2013

2014

# How Does it Work?



Schema-agnostic design



Real-time indexing for search and query



Reverse queries for event processing and alerting



Scale-out shared-nothing cluster topology

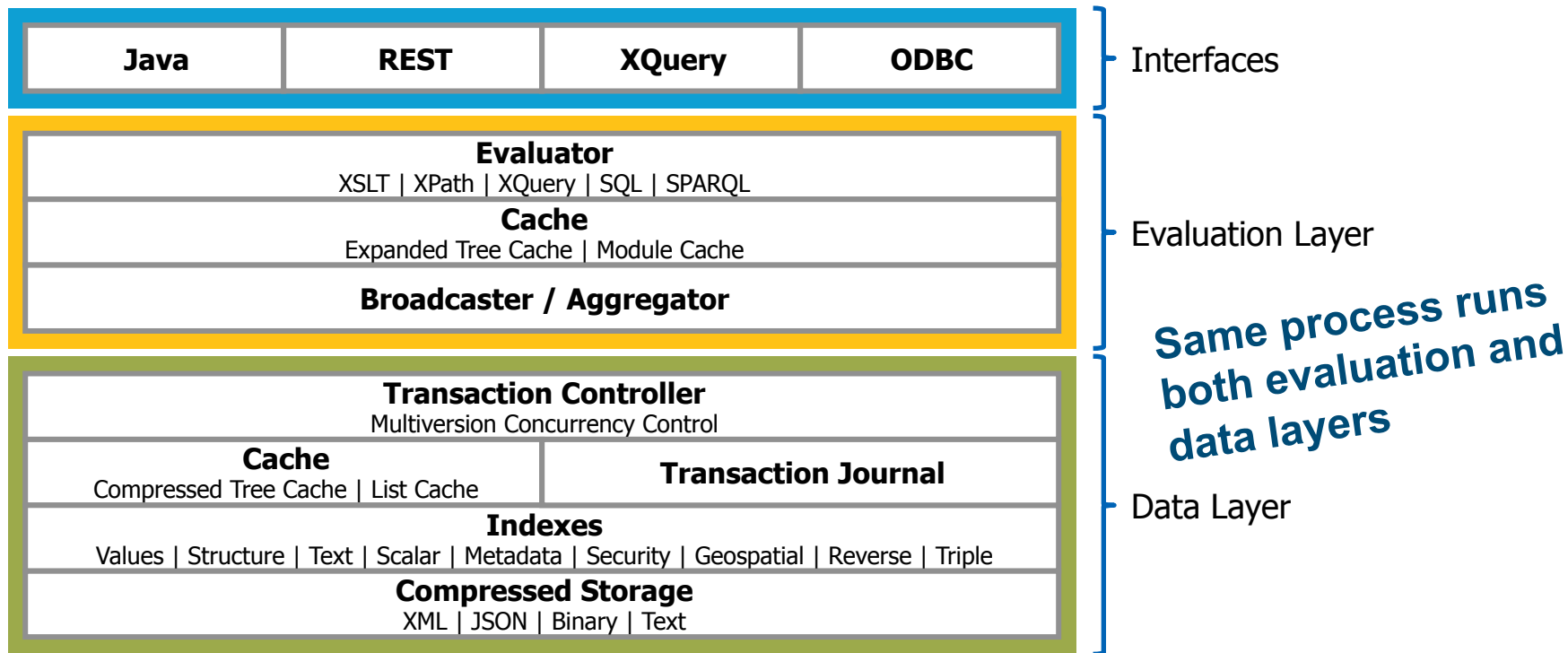


Analytics and Visualization



Transactions

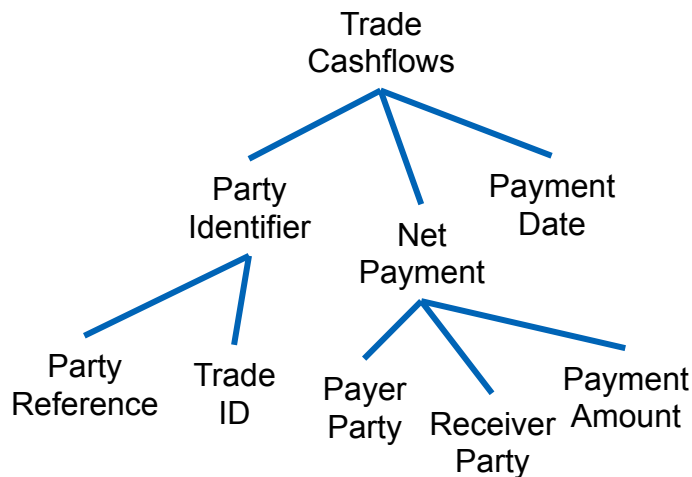
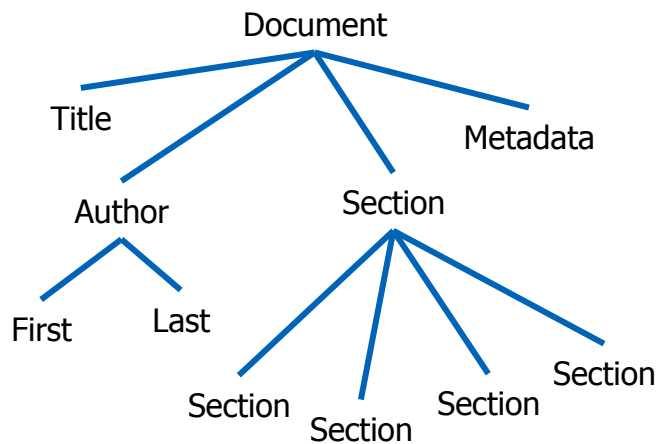
# MarkLogic Architecture





# Hierarchical Data Model

- MarkLogic Server is a document-centric database
  - Supports any-structured data via hierarchical data model
  - Stores compressed binary trees



# MARKLOGIC IN ACTION

# What we will do

- Create a database
- Loading data
- Modifying data
- Searching and querying
- Fun with Indexes
- Building an application



LET'S START!

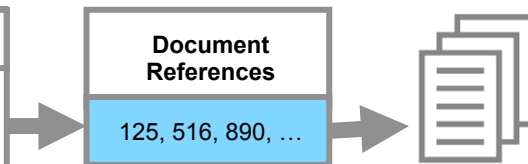
A faint, light gray background graphic on the left side of the slide. It depicts a document with several horizontal lines representing text, and a magnifying glass is superimposed over the center of the document, with its handle pointing towards the bottom right.

# SOME WORDS ABOUT INDEXES

# Universal Index

*Which vetted SAR reports contain the phrase **blue van**?*

| Term                             | Term List                         |
|----------------------------------|-----------------------------------|
| "blue"                           | 123, 125, 129, 130, 152, 344, ... |
| "van"                            | 123, 125, 126, 129, 130, 152, ... |
| "observed"                       | 125, 152, 516, 522, 765, 890, ... |
| "blue van"                       | 123, 125, 129, 130, 152, 486, ... |
| STEM "observe"                   | 125, 152, 516, 522, 765, 890, ... |
| <SAR>                            | ...                               |
| <SAR>/<location>                 | ...                               |
| <threat>/<category>              | ...                               |
| <type>suspicious activity</type> | ...                               |
| <date>2012-11-12Z</date>         | ...                               |
| Collection:Vetted                | ...                               |
| Role:Analyst + Action:Read       | ...                               |
| ...                              | ...                               |
| ...                              | ...                               |
| ...                              | ...                               |



## MarkLogic indexes...

- Words
- Phrases
- Stemmed words and phrases
- Structure
- Words and phrases in the context of structure
- Values
- Collections
- Security Permissions

# Range Index

*Which vetted SAR reports containing the phrase **blue van** were submitted before 2013?*

| Term                             | Term List                         |
|----------------------------------|-----------------------------------|
| "blue"                           | 123, 125, 129, 130, 152, 344, ... |
| "van"                            | 123, 125, 126, 129, 130, 152, ... |
| "observed"                       | 125, 152, 516, 522, 765, 890, ... |
| "blue van"                       | 123, 125, 129, 130, 152, 486, ... |
| STEM "observe"                   | 125, 152, 516, 522, 765, 890, ... |
| <SAR>                            | ...                               |
| <SAR>/<location>                 | ...                               |
| <threat>/<category>              | ...                               |
| <type>suspicious activity</type> | ...                               |
| <date>2012-11-12Z</date>         | ...                               |
| Collection:vetted                | ...                               |
| Role:Analyst + Action:Read       | ...                               |
| ...                              | ...                               |
| ...                              | ...                               |
| ...                              | ...                               |

| Document References |
|---------------------|
| 125, 516, 890, ...  |

| DocID | <date>      |
|-------|-------------|
| 1     | <date>      |
| 3     | <date>      |
| 4     | 1999-12-31Z |
| 5     | 2003-02-16Z |
| 8     | 2007-06-01Z |
| 10    | 2008-01-01Z |
| 11    | 2009-04-28Z |
| 17    | 2011-05-12Z |
| ...   | 2012-11-12Z |
| ...   | 2013-09-30Z |
| ...   | ...         |
| ...   | ...         |
| ...   | ...         |

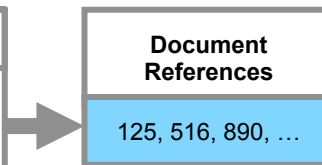


Range indexes map document IDs to values, and vice-versa in a compact in-memory representation.

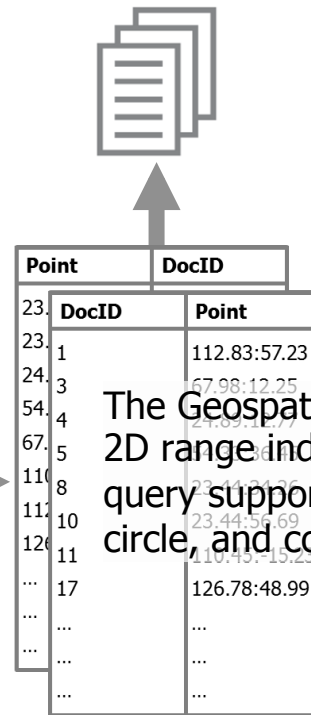
# Geospatial Index

*Which vetted SAR reports about a blue van from before 2013 refer to a location near the airport?*

| Term                             | Term List                         |
|----------------------------------|-----------------------------------|
| "blue"                           | 123, 125, 129, 130, 152, 344, ... |
| "van"                            | 123, 125, 126, 129, 130, 152, ... |
| "observed"                       | 125, 152, 516, 522, 765, 890, ... |
| "blue van"                       | 123, 125, 129, 130, 152, 486, ... |
| STEM "observe"                   | 125, 152, 516, 522, 765, 890, ... |
| <SAR>                            | ...                               |
| <SAR>/<location>                 | ...                               |
| <threat>/<category>              | ...                               |
| <type>suspicious activity</type> | ...                               |
| <date>2012-11-12Z</date>         | ...                               |
| Collection:Vetted                | ...                               |
| Role:Analyst + Action:Read       | ...                               |
| ...                              | ...                               |
| ...                              | ...                               |
| ...                              | ...                               |



| DocID | <date>      |
|-------|-------------|
| 1     | <date>      |
| 3     | 1999-12-31Z |
| 4     | 2003-02-16Z |
| 5     | 2007-06-01Z |
| 8     | 2008-01-01Z |
| 10    | 2009-04-28Z |
| 11    | 2011-05-12Z |
| 17    | 2012-11-12Z |
| ...   | 2013-09-30Z |
| ...   | ...         |
| ...   | ...         |
| ...   | ...         |
| ...   | ...         |



| Point  | DocID        |
|--------|--------------|
| 23.1   | DocID        |
| 23.1   | 112.83:57.23 |
| 24.3   | 67.98:12.25  |
| 54.4   | 24.89:12.77  |
| 67.5   | 23.44:24.26  |
| 110.8  | 23.44:56.69  |
| 112.10 | 110.45:15.23 |
| 126.11 | 126.78:48.99 |
| ...    | ...          |
| ...    | ...          |
| ...    | ...          |
| ...    | ...          |
| ...    | ...          |

The Geospatial index is like a 2D range index, with built-in query support for point, box, circle, and complex polygons.



*Which vetted SARs about a blue van from before 2013 with this location refer to partial plate **ABC**?*

**Document  
References**

125, 516, 890, ...

**All indexes fully composable  
for sub-second response to  
very complex queries**

| DocID | <date>         |
|-------|----------------|
| 1     | <date>         |
| 3     | DocID          |
| 4     | 1999-12-31Z 8  |
| 5     | 2003-02-16Z 10 |
| 8     | 2007-06-01Z 4  |
| 10    | 2008-01-01Z 5  |
| 11    | 2009-04-28Z 3  |
| 17    | 2011-05-12Z 11 |
| ...   | 2012-11-12Z 1  |
| ...   | 2013-09-30Z 17 |
| ...   | ...            |
| ...   | ...            |
| ...   | ...            |

| Point | DocID |
|-------|-------|
| 23.   | DocID |
| 23.   | Point |
| 23.   | 1     |
| 24.   | 3     |
| 54.   | 4     |
| 67.   | 5     |
| 110.  | 8     |
| 117.  | 10    |
| 126.  | 11    |
| ...   | 17    |
| ...   | ...   |
| ...   | ...   |
| ...   | ...   |

[illegible]

The Triple index is an index of “facts” expressed as Semantic triples. It can efficiently query and join billions of “linked data” triples.

# Triple Index

| DocID | Subject | Predicate | Object    |
|-------|---------|-----------|-----------|
| 1     | Object  | Subject   | Predicate |
| 3     | -122    | Predicate | Subject   |
| 4     | 37.4    | Object    | DocID     |
| 5     | fnar    | Subject   | Object    |
| 8     | ABC     | isa       | Entity1   |
| 10    | Doe     | isa       | Entity1   |
| 11    | John    | isa       | Entity1   |
| 17    | licen   | lat       | Entity2   |
| ...   | locat   | lat       | Entity2   |
| ...   | pers    | long      | Entity2   |
| ...   | valu    | value     | Entity3   |
| ...   | ...     | ...       | ...       |
| ...   | ...     | ...       | ...       |
| ...   | ...     | ...       | ...       |

With the new Triple index, you can:

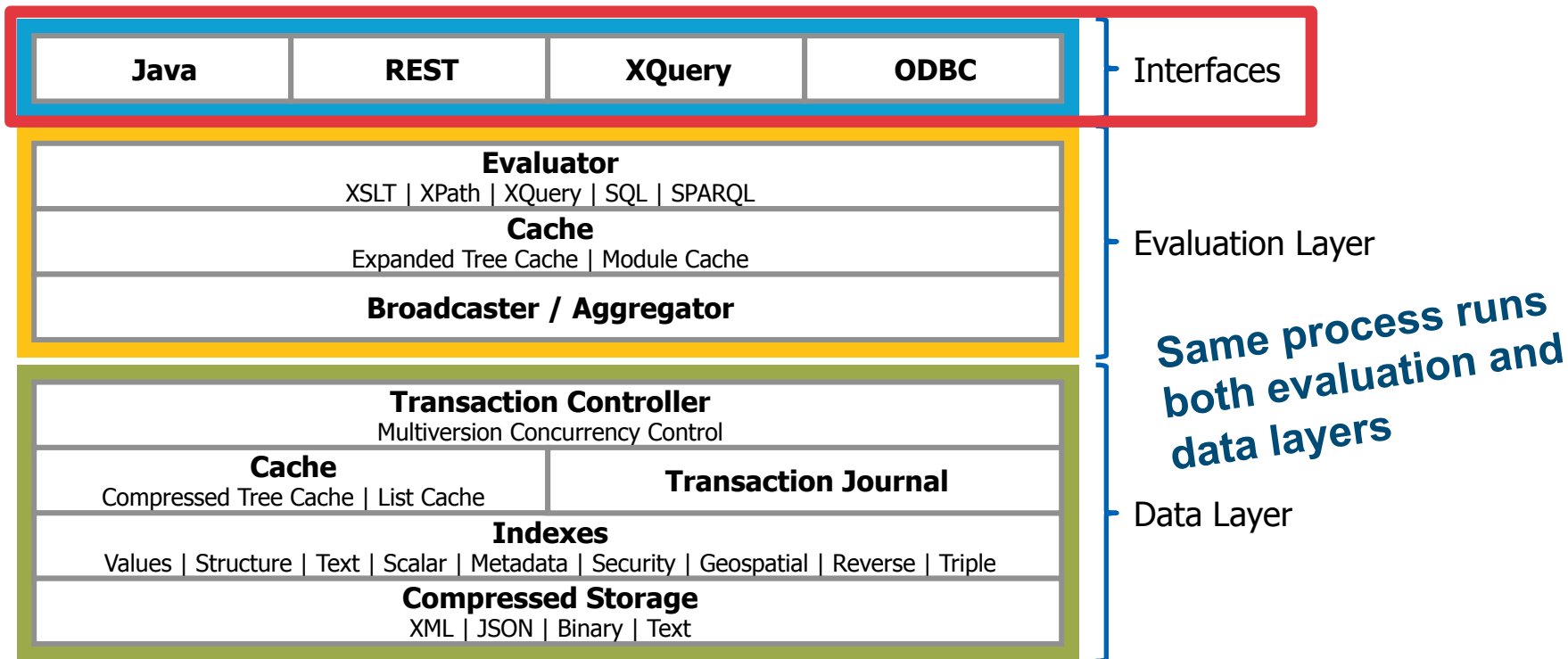
- Ingest triples from public linked data sources
- Combine triples in your documents with triples from other sources to answer questions like
  - What drivers' licenses (and names, addresses, etc) are linked to this license plate?
  - Show me everything you know about this license plate (traffic cam info, registration info, etc)



LET'S CONTINUE!

# DEVELOPMENT

# Remember this?





# REST

MarkLogic is REST Server

## **REST Interface to MarkLogic Server**

Foundation for language-specific API's

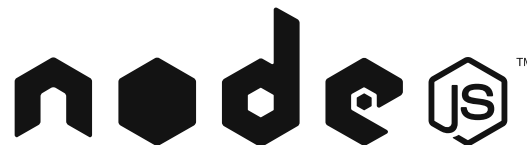
- Full-text and facet search
- Access to value indexes and aggregates
- Document management for XML, text, binary, and JSON
- Separated read, write and administrative access
- Administrative services for managing API configuration
- Extensibility with custom URL endpoints
- Scriptable and UI bootstrapping

# Node.js Client API

## New in version 8

A fluent JavaScript interface for key MarkLogic capabilities such as search, transactions, aggregates, alerting, and geospatial—the Node.js Client API allows developers to move faster to production with proven reliability at scale.

The Node client is developed in the open on GitHub,  
<https://github.com/marklogic/node-client-api>,  
and is licensed under the open-source Apache 2.0 license.



```
var marklogic = require('marklogic');
var conn = require('./env.js').connection; // Host and auth details

var db = marklogic.createDatabaseClient(conn);
var q = marklogic.queryBuilder;
db.documents.query(
  q.where(
    q.collection('countries'),
    q.value('region', 'Africa'),
    q.or(
      q.word('background', 'France'),
      q.word('Legal system', 'French')
    )
  )
)
.result(function(documents) {
  documents.forEach(function(document) {
    console.log(JSON.stringify(document))
  })
});
```



# Java Client API

## New in version 8

The MarkLogic Java Client API is an open-source library that allows developers to quickly, easily, and reliably access MarkLogic from their Java applications.

- Faster development and less custom code with out-of-the-box data management, search, and alerting
- Pure Java query builder and conveniences for POJOs, JSON, XML, and binary I/O
- Built-in extensibility for moving performance-critical code to the database
- Always open-source and developed on GitHub, <https://github.com/marklogic/java-client-api>

```
package com.marklogic.examples;
```

```
import com.marklogic.client.DatabaseClient;  
import com.marklogic.client.DatabaseClientFactory;
```

```
public class Configuration {  
    private static DatabaseClient client =  
        DatabaseClientFactory.newClient(  
            "localhost", 8000, // Every instance comes with a REST Client API  
                               // pre-installed on port 8000  
            "Documents", // Each connection can specify its database at runtime  
            "admin", "*****", DatabaseClientFactory.Authentication.DIGEST);  
  
    public static DatabaseClient exampleClient() {  
        return client;  
    }  
}
```





BUILDING AN APPLICATION!

# MLJS

## Developed by an MarkLogic employee

MLJS is a MarkLogic JavaScript wrapper API. It comprises several components:-

- Core API that wraps the MarkLogic REST API and provides utility objects. Works in Node.js and the Browser
- Browser Widgets API – provides over 30 JavaScript widgets linked to MarkLogic
- MLJS Workplace – A drag/drop application that enables visual creation and configuration of a MarkLogic web application

More information at  
<https://adamfowlerml.wordpress.com/mljs/>

### {mljs} Ecosystem

#### Yeoman MLJS Workplace Generator

- Just 6 commands to create, deploy, and run an app!

#### MLJS Workplace

- Drag & Drop app creation

#### ML Node Tools

- Deploy tool & webserver

#### MLJS Widgets API

- JavaScript UI Widgets

#### MLJS Core API – Node.js & Browser

- REST API wrappers, utility objects

# Getting started!

- Download your free developer license
  - <http://developer.marklogic.com/products>
  - Or test the latest release (early access) <http://ea.marklogic.com>
- Learn
  - By tutorials/examples <http://developer.marklogic.com/learn>
  - By free instructor lead online training <http://mlu.marklogic.com/registration/>
- Get involved
  - <http://stackoverflow.com/questions/tagged/marklogic>
  - MarkLogic Mailing list <http://developer.marklogic.com/discuss>

Contact me on [mats.stellwall@marklogic.com](mailto:mats.stellwall@marklogic.com)

Follow me on twitter [@MatsStellwall](https://twitter.com/MatsStellwall)



ANY QUESTIONS?