

# Semantics

Semantics describes MarkLogic's ability to store graph data as RDF Triples—a new approach to modeling data that focuses on *relationships*. Storing RDF enhances MarkLogic's document model by providing a smart way to *connect and enhance* the JSON and XML documents that MarkLogic stores. This is important for data integration and better querying. Semantics also provides *context* for data by providing a smart way to store metadata. For example, consider a database that has information about parts, and one part is listed with a size of "42." But, where is the contextual information: *What are the units of "42"?* *What is the tolerance?* *Who measured it?* *When was it measured?* *Who can see this data?* That contextual information is the *semantics* of your data, and is easily stored in MarkLogic as RDF triples.

**DATA** is stored in Triples, expressed as:

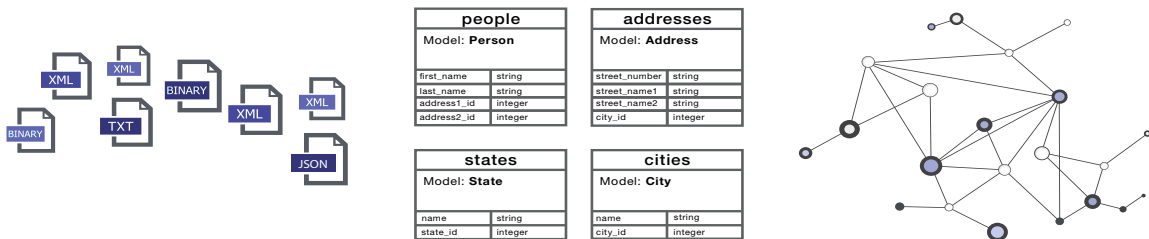
Subject	Predicate	Object
John Smith	livesIn	London
London	isIn	England

**QUERY** with SPARQL, gives us a simple look up... and more! Find people who live in (a place that's in) England

**Based on W3C standards, triples are written using RDF (Resource Description Framework), and queried using SPARQL.**

## RDF and SPARQL

The standard way to represent semantic data is with RDF Triples (Resource Description Framework), and the standard query language is SPARQL. Triples are derived from subject-predicate-object constructions based on entities (people, places, or things) and their relationships. One example is, "John lives in London." The database can also infer new facts, such as "John lives in England." These simple facts can all be linked together to form a graph of hundreds of billions of facts and relationships. This simple model is ideal for describing the *nature* and *meaning* of relationships.

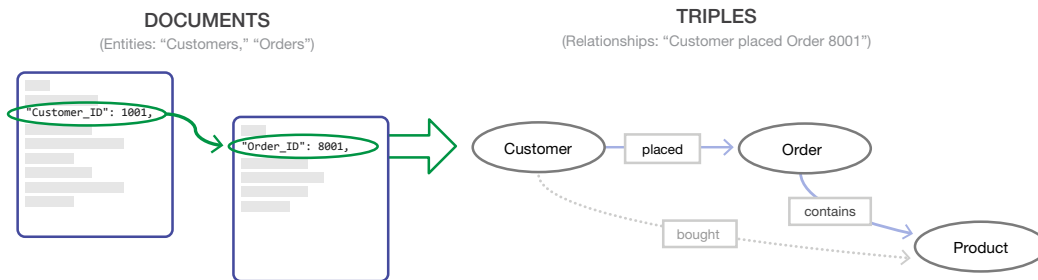


## Multi-Model Database: Documents, Data, and Triples

- **Documents + Data + Triples** – MarkLogic natively stores documents, data, and triples together
- **Fully Composable** – MarkLogic is truly multi-model—you can write fully composable queries across all data types
- **Enterprise Ready** – Full set of enterprise features features, including ACID transactions, security, disaster recovery, etc.
- **Sophisticated Indexing** – Specialized triple index to improve query power and performance
- **Enormous Scale** – MarkLogic stores hundreds of billions of triples

## Supported Features

- Store and manage hundreds of billions of RDF triples
- Query across documents, data, and triples
- Triple index for sub-second search results
- Triple cache for high performance across large clusters
- Bulk-load triples via MarkLogic Content Pump (mlcp)
- Provenance and reification by adding metadata
- XQuery helper modules for serializations and transitive closures
- Updates, aggregates via MarkLogic APIs
- Graph traversal with property paths and transitive closures
- Semantic inference using rule sets at query time
  - Supplied rule sets for RDFS, RDFS+, and OWL Horst
  - Support for user-defined rule sets
- Full support for SPARQL 1.1
- SPARQL endpoint and graph store protocol support
- SPARQL from server-side JavaScript, Node.js
- Support for Jena and Sesame APIs
- Full integration with semantics technology partners (Smartlogic, Pool Party, Cambridge Semantics)
- MarkLogic enterprise features: ACID transactions, certified security (at document/triple level), high availability and disaster recovery, scalability and elasticity



## How Semantics Helps With Data Integration

In a relational database, relationships are weak. They are defined only by primary and foreign key relationships or obfuscated in application code. Relational databases also make it hard to distinguish the real relationships from the spurious ones that exist only to allow reconstruction of the entities that were broken apart across tables in the first place.

With triples, you can define the *meaning* of relationships. Consider the example above, which shows documents for “customers” and “orders.” With triples, you can define what it means to “place” an order, what it means for an order to “contain” something, and new inferences to say that a customer “bought” a certain product. Semantics makes such relationships all have meaning, and that meaning will always stay intact, and can be leveraged, as you integrate data from different systems.

## About MarkLogic

MarkLogic is the world’s best database for integrating data from silos, providing an operational and transactional Enterprise NoSQL database platform that integrates data better, faster, with less cost. Visit [www.marklogic.com](http://www.marklogic.com) for more information.