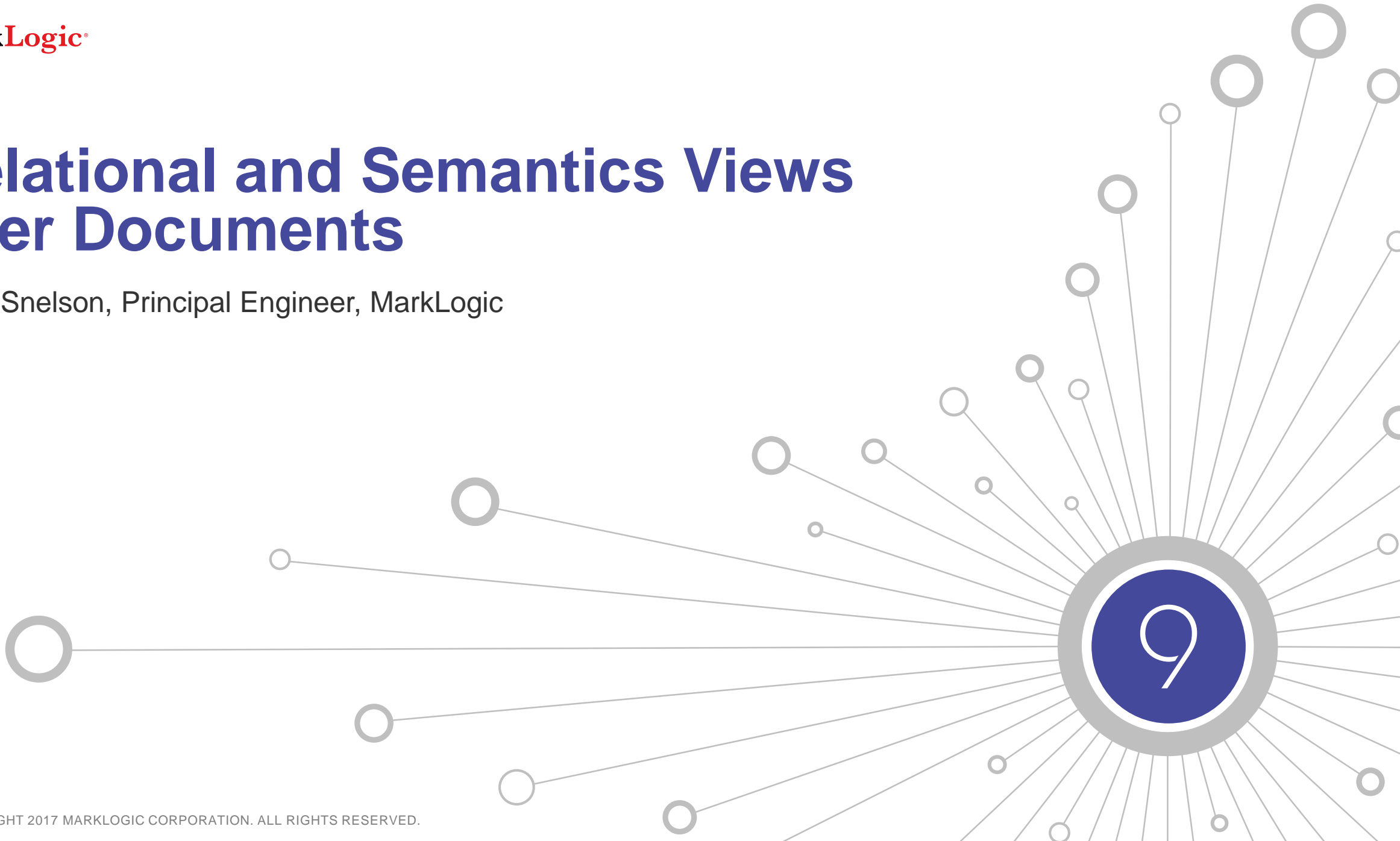


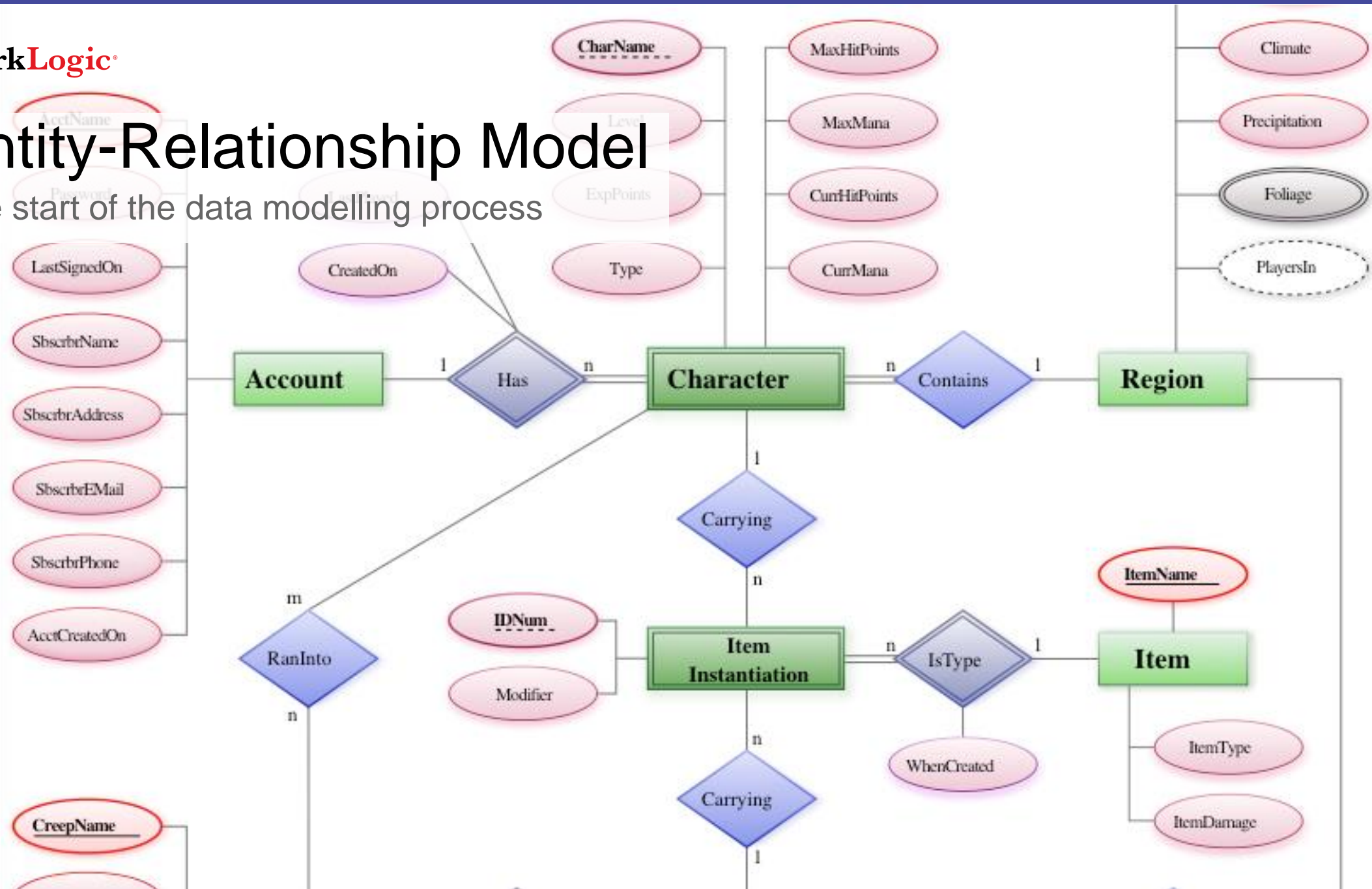
# Relational and Semantics Views over Documents

John Snelson, Principal Engineer, MarkLogic



# Entity-Relationship Model

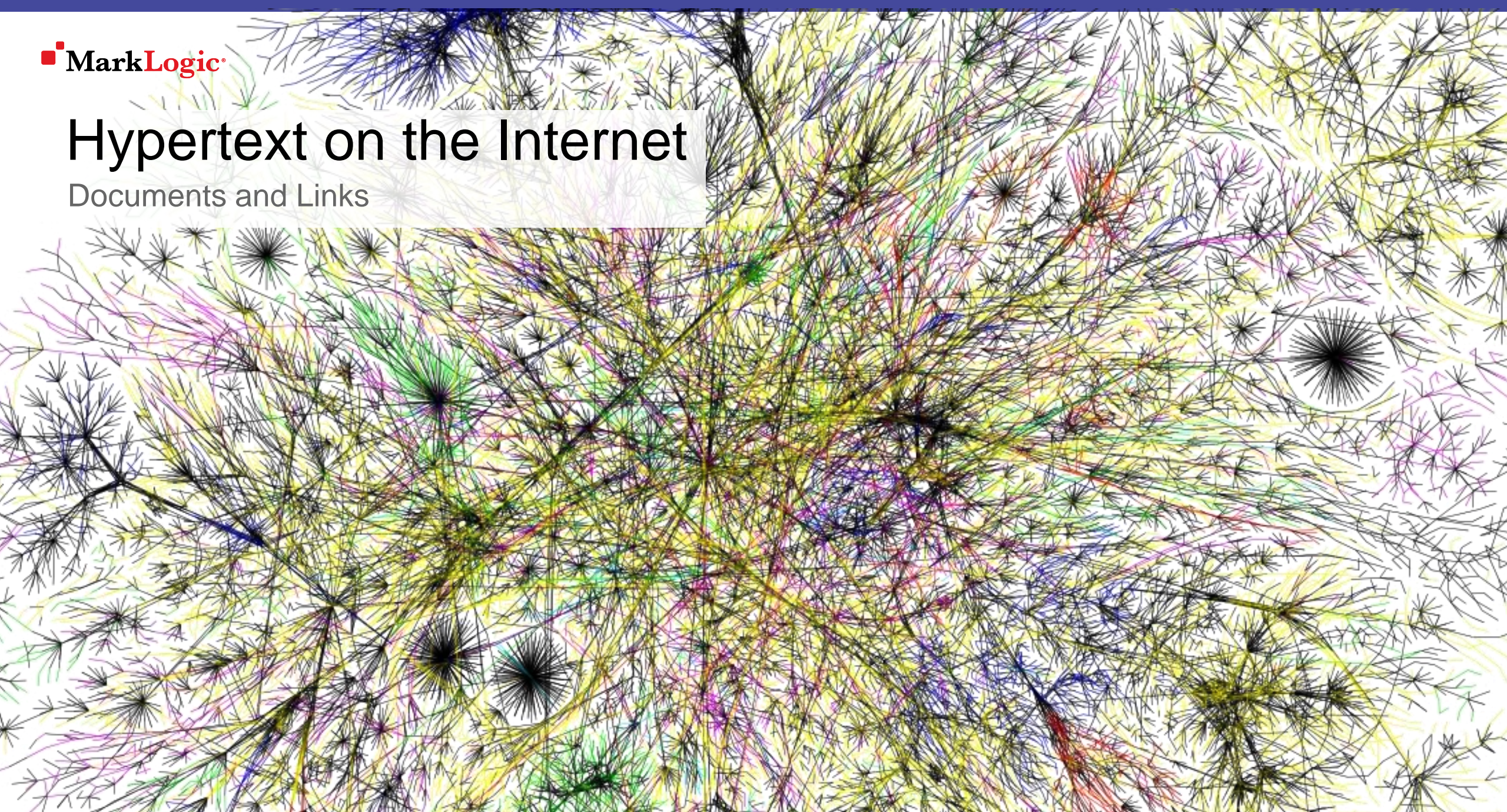
The start of the data modelling process





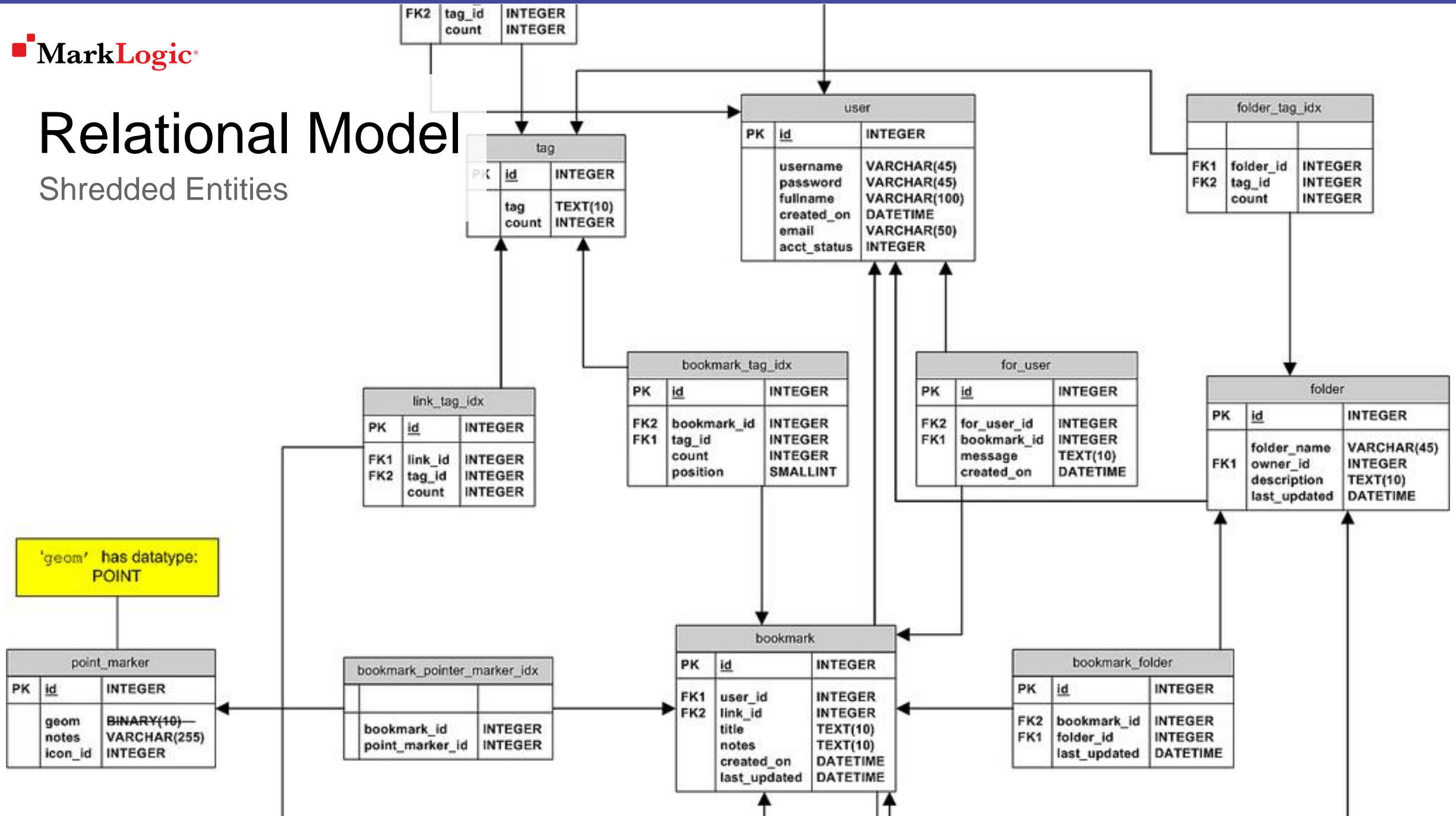
# Hypertext on the Internet

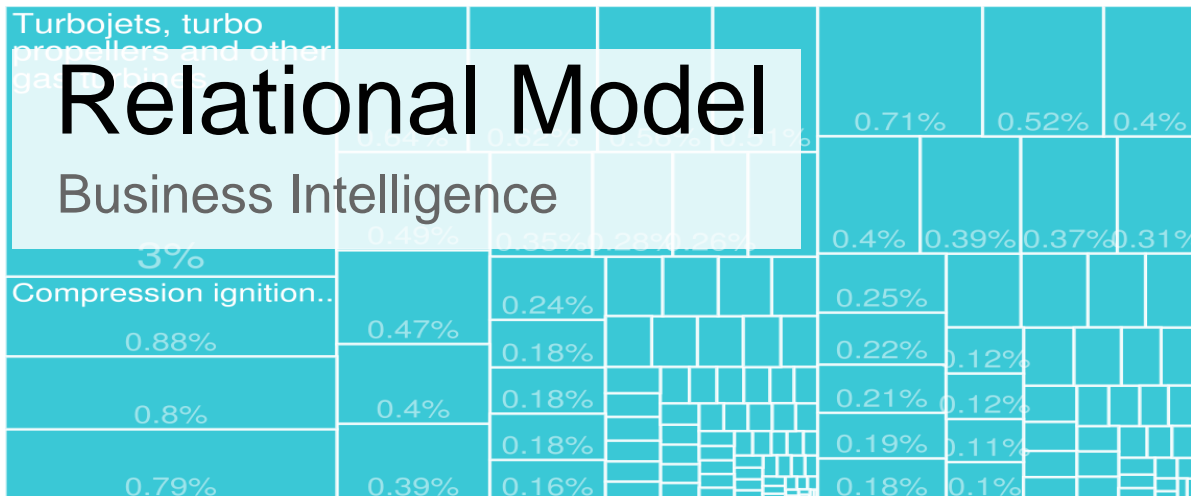
Documents and Links





# Shredded Entities





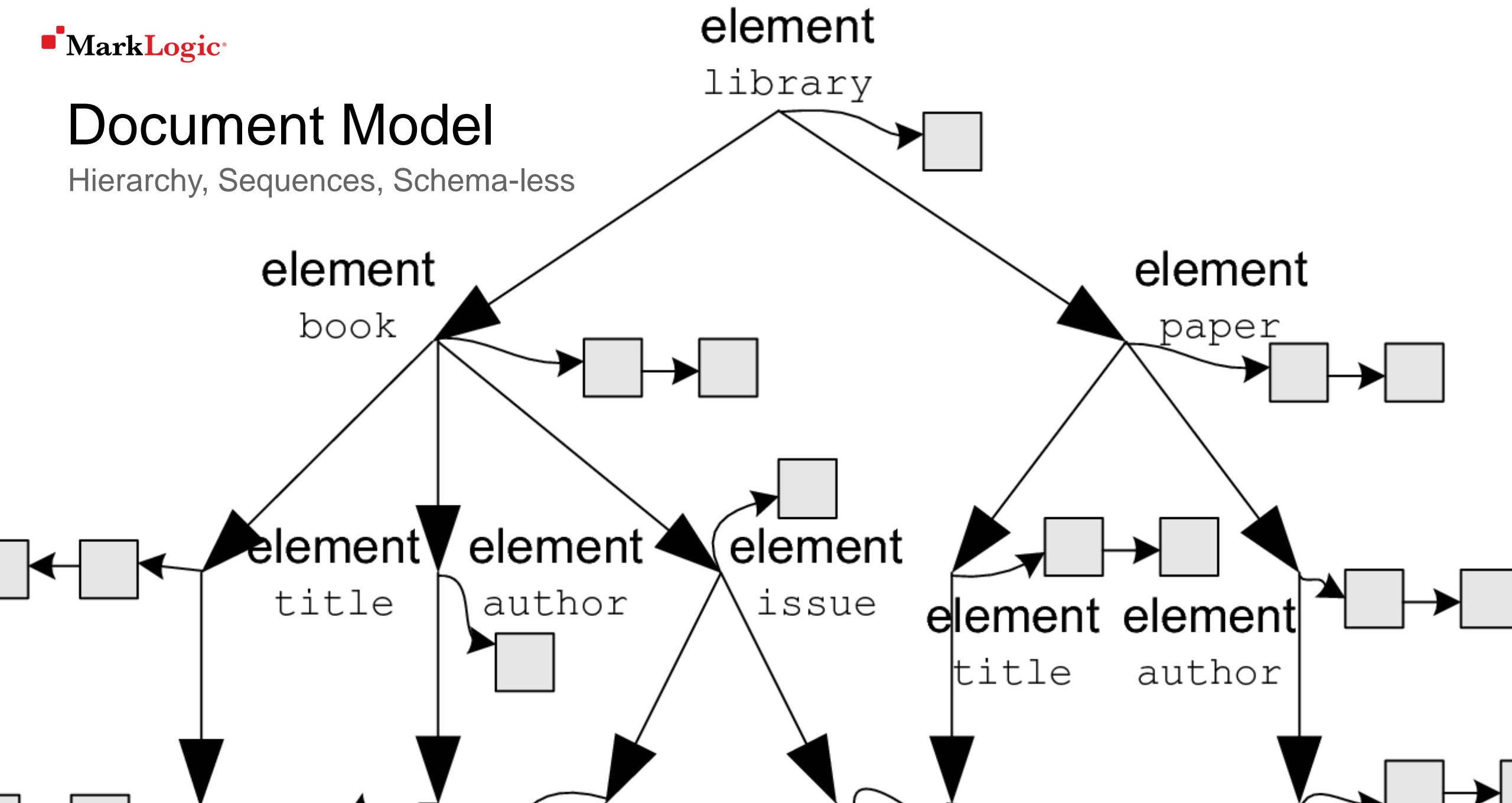




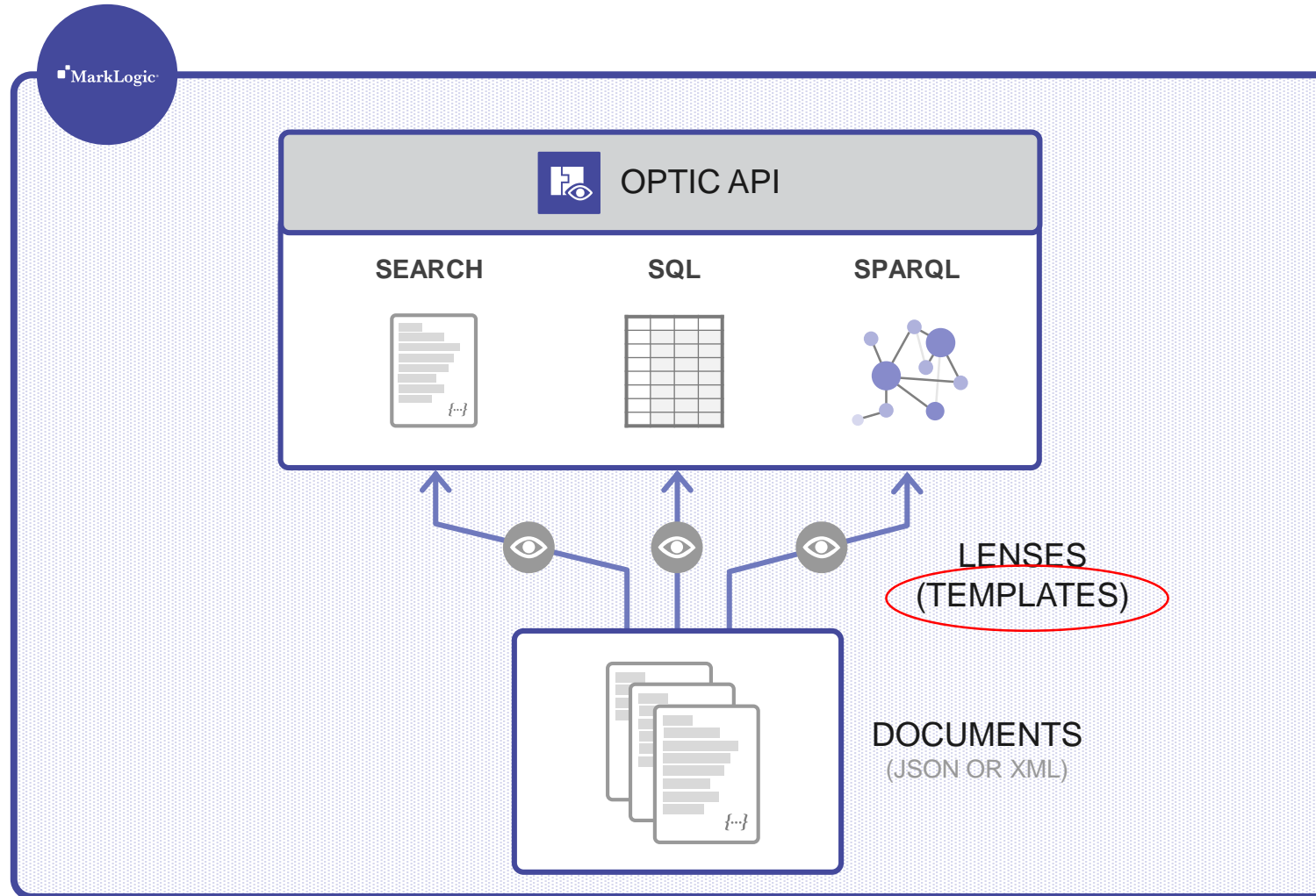


# Document Model

Hierarchy, Sequences, Schema-less

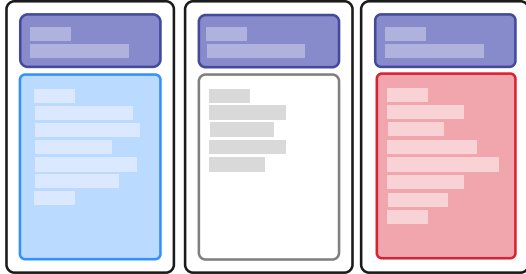


# MarkLogic 9 – Easier Data Query



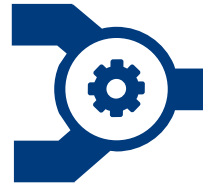


# Documents



```
{
  "name": "10K",
  "runs": [
    {
      "name": "John",
      "age": 36
    }
  ]
}
```

# Template-Driven Extraction



```
<template>
...
<context>/runs/name</context>
...
<schema-name>Race</schema-name>
<view-name>Runs</view-name>
<columns>
  <column>
    <name>age</name>
    <scalar-type>integer</scalar-type>
    <value>../age</value>
  </column>
  <column>
    ...
  
```

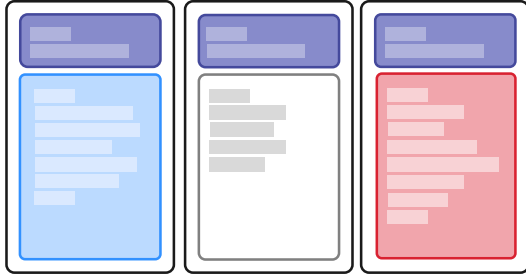
## Row Index

	Self Defense	Protection of Others	Protection of Property
Justified	85%	75%	65%
Not Justified	10%	20%	30%
Don't Know	5%	5%	5%

```
select age
from Race.Runs
```

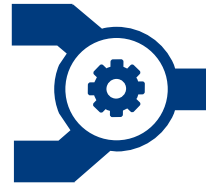
```
op:from-view("Race","Runs")
  =>op:select("age")
  =>op:result()
```

## Documents



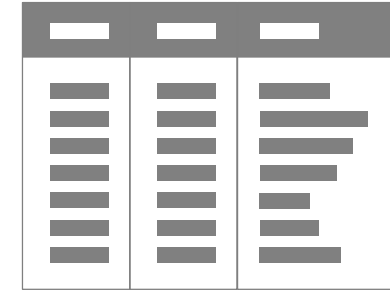
```
{
  "name": "10K",
  "runs": [
    {
      "name": "John",
      "age": 36
    }
  ]
}
```

## Template-Driven Extraction



```
<template>
  <context>/runs/name</>
  <schema-name>Race</>
  <view-name>Runs</>
  <triples>
    <triple>
      <subject>
        <val>sem:iri( xdmp:node-uri(.) )</val>
      </subject>
      <predicate>
        <val>sem:iri($EX||"hasAge" )</val>
      </predicate>
      <object>
        <val>xs:integer(./age)</val>
      </object>
    </triple>    ...
```

## Triple Index

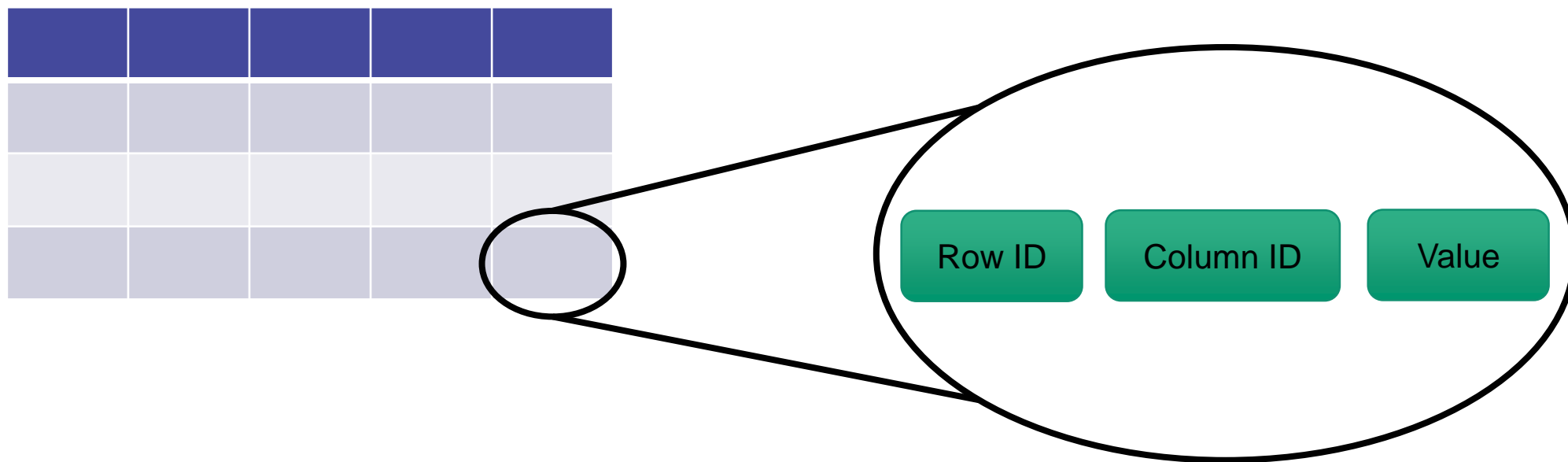


```
select distinct ?p {
  ?s ?p ?o
}
```

```
op:from-triples((
  op:pattern(op:col("s"),
    op:col("p"), op:col("o"))
))
=>op:select("p")
=>op:where-distinct()
=>op:result()
```

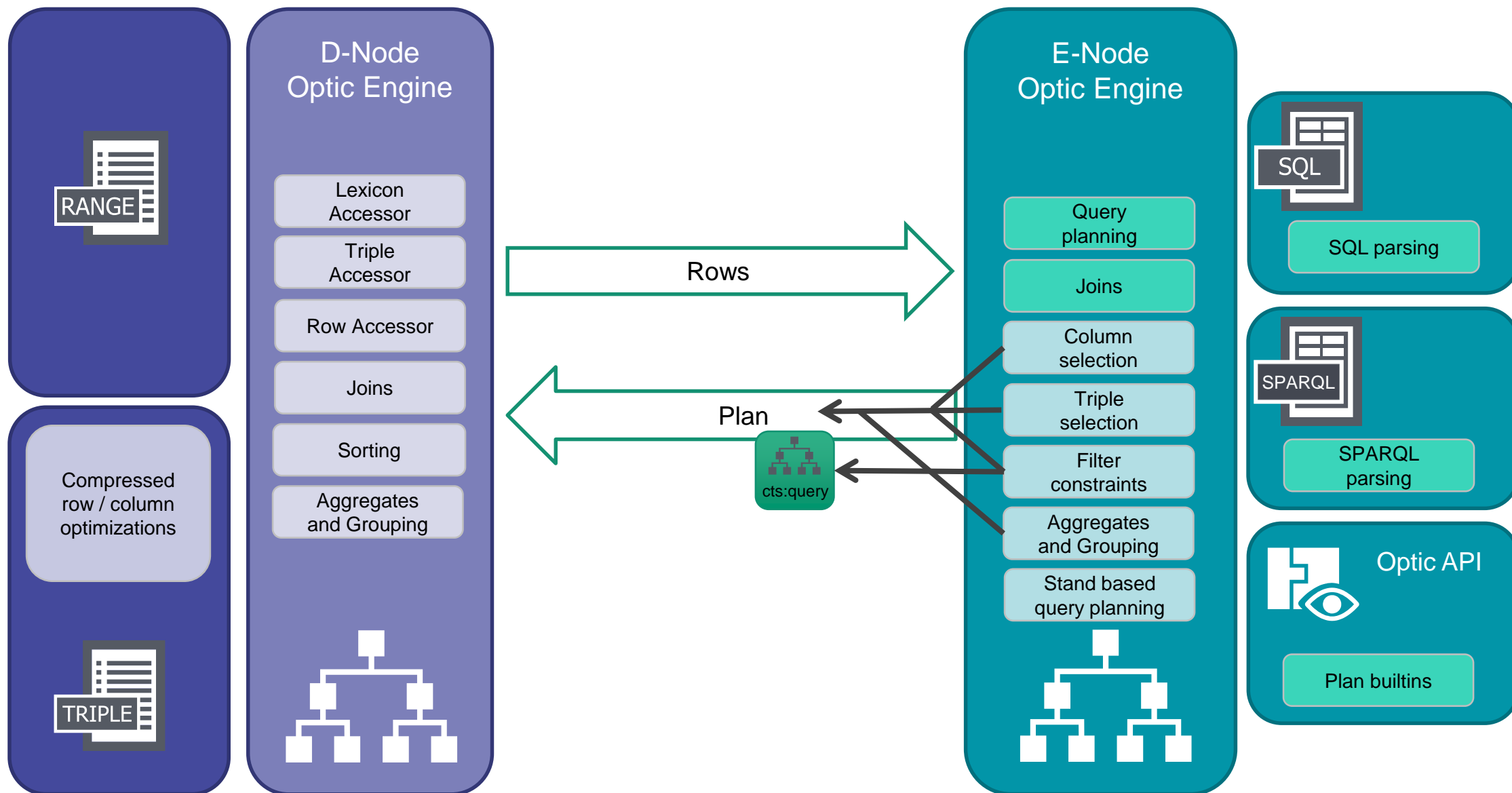


# Tables in the Triple Index



Triple Index Permutation	Major Order	Minor Order	Use Case
PSO	Column ID	Row ID	Column access ordered by row ID
SOP	Row ID		Row access
POS <sup>*new*</sup>	Column ID	Value	Column access ordered by value

# Optic Engine Architecture – MarkLogic 9





# Templates - Demo



# Benefits

Multiple Views, One Source

- Happy Business Analysts, Ontologists, etc.
- One document to update or delete
- One database to query
- Transactional
- Backups
- Replication



# Benefits

## Multiple Sources, One View

- Not just for different views of a single document type
- Create one view over different documents
- Unify linking styles
- Same semantics extracted from multiple document formats

# Benefits

## NoETL

- Keep the original content (no extract)
- Active rather than one time only transform
- No load-reload cycle
- Progressive enhancement
- Lossless modelling
  - No extract means nothing is thrown away
- Data provenance



# Benefits

## Data in Context

- Relational rows and triples maintain a link to their source document
- Fetch source and display or query
- Pre-filter a SQL/SPARQL/Optic query based on document context
- Provenance – check source, verify confidence, etc.
- Inherits (bi-)temporal validity of source document
- Inherits access control of source document


# BI Tools: connect and extract

Reports and Analytics over relational data (tables)

- Interactive SQL/ODBC
  - Ad hoc SQL via drag and drop
  - Most general: requires only ad hoc SQL
- Data Extract
  - Batch extract of data into a relational dump
  - Most efficient, powerful; requires some programming
- Web Data Connector
  - Web page / form to define a real-time extract
  - Compromise: efficient/powerful; requires less programming


# Early Access 4 – Publically available now!

[MARKLOGIC.COM](#) [DEVELOPERS](#) [MARKLOGIC WORLD](#) [UNIVERSITY](#) [RESOURCES](#) [EVENTS](#) [CONTACT US](#) : 1 877 992 8885

 [HOME](#) [FEATURES](#) [FORUMS](#) [MY ACCOUNT](#) [DOCUMENTATION](#) [DOWNLOAD](#) [Q](#)

## MarkLogic 9

Early Access Now Available




### Data Integration

Creating more value by combining information from multiple sources.

**INCLUDES:** Entity Services, Data Movement SDK, Template Driven Extraction, SQL Enhancements, Optic API

[LEARN MORE](#)




### Security

Enterprise-grade security, now stronger and more granular.

**INCLUDES:** Encryption, Redaction, Element-level Security

[LEARN MORE](#)




### Manageability

Providing one clear picture of your MarkLogic system.

**INCLUDES:** Ops Director

[LEARN MORE](#)



### Additional Enhancements

**INCLUDES:** Geospatial, Search, Enhanced Tiered Storage, JavaScript, Query Console

[LEARN MORE](#)



*NoSQL Joins*

*Telemetry*

**Encryption**

*Entity Services*

**Ops Director**

*Rolling Upgrade*

*NoSQL Aggregates*

**Cloud**

*JavaScript Enhancements*

*Template-Driven Extraction*

**Enhanced SQL**

**Double Precision Geospatial**

*Data Movement SDK*

*Geospatial Region Search*

*Search Enhancements*

*Stemming & Tokenization Plugin*

**Bitemporal Enhancements**

*Inside MarkLogic Book*

*Enhanced ODBC*

**MarkLogic Optic API**

*Element Level Security*

*Secure HDFS*

*Redaction on Export*

*Policy-Driven Tiered Storage*

*Common Criteria Security Certification*

