MarkLogic WORLD 2017

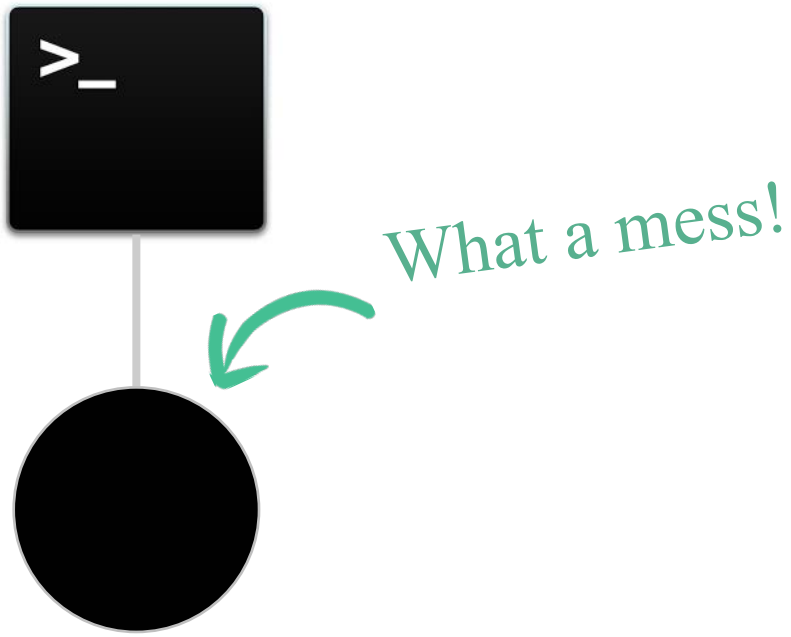
Mastering Data Access with the Optic API & Template-Driven Extraction

Erik Hennum, Principal Engineer, MarkLogic

Fayez Saliba, Staff Engineer, MarkLogic

© COPYRIGHT 2017MARKLOGIC CORPORATION. ALL RIGHTS RESERVED.

The Customer Support Experience



The costly, risky black hole of information

Automatic problem resolution

© COPYRIGHT 2017 MARKLOGIC CORPORATION. ALL RIGHTS RESERVED.

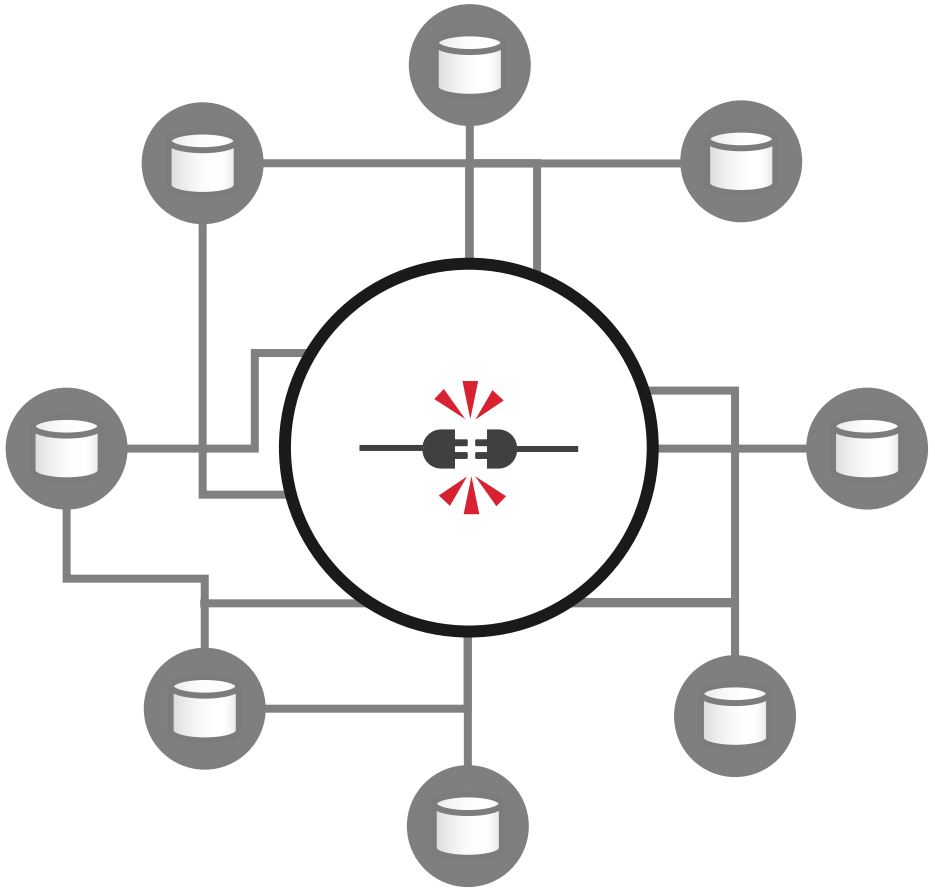
Introductions

- Erik Hennum
  - MarkLogic Principal Engineer (6 years)
  - Experience: Client APIs – Java, REST, Node.js
- Fayez Saliba
  - MarkLogic Staff Engineer (2 years)
  - Experience: Template-driven extraction

© COPYRIGHT 2017 MARKLOGIC CORPORATION. ALL RIGHTS RESERVED.

MarkLogic

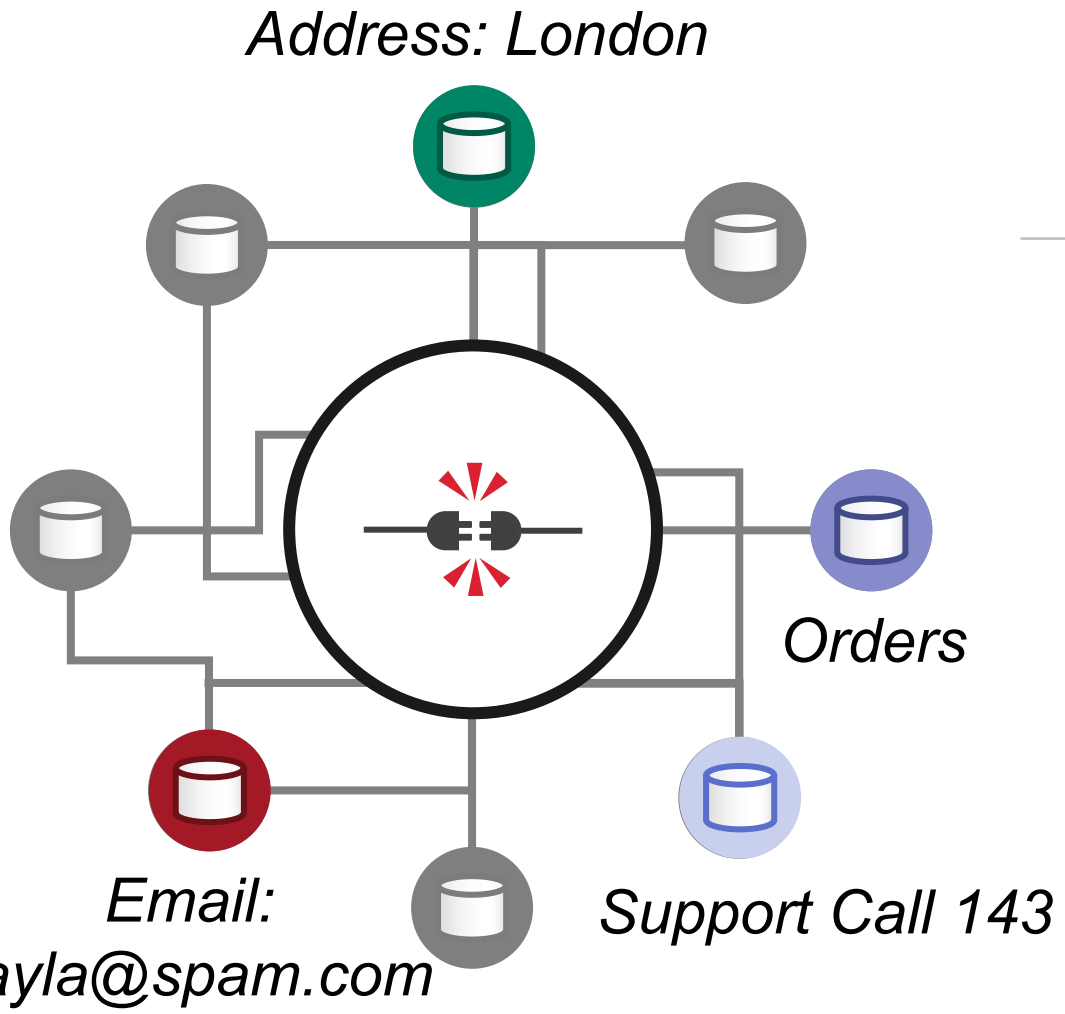
Why Is This Hard Today?



Kayla the Customer

- Who is Kayla? What is our relationship with her?
- Is ”Kayla” on the phone the same “Kayla” that ordered 3 products from us?
- What has she purchased? Returned? Received support on?
- Where does she live? How do I contact her?
- Is she a “high-value” customer? What risk does she represent?

© COPYRIGHT 2017 MARKLOGIC CORPORATION. ALL RIGHTS RESERVED.



Kayla the Customer

- ***CRM 1*** – Where does Kayla live?
- ***CRM 2 Marketing*** – How do we contact Kayla?
- ***ERP*** – Is this the same “Kayla” that ordered 3 products from us?
- ***Support*** – What has Kayla received support on?

© COPYRIGHT 2017 MARKLOGIC CORPORATION. ALL RIGHTS RESERVED.

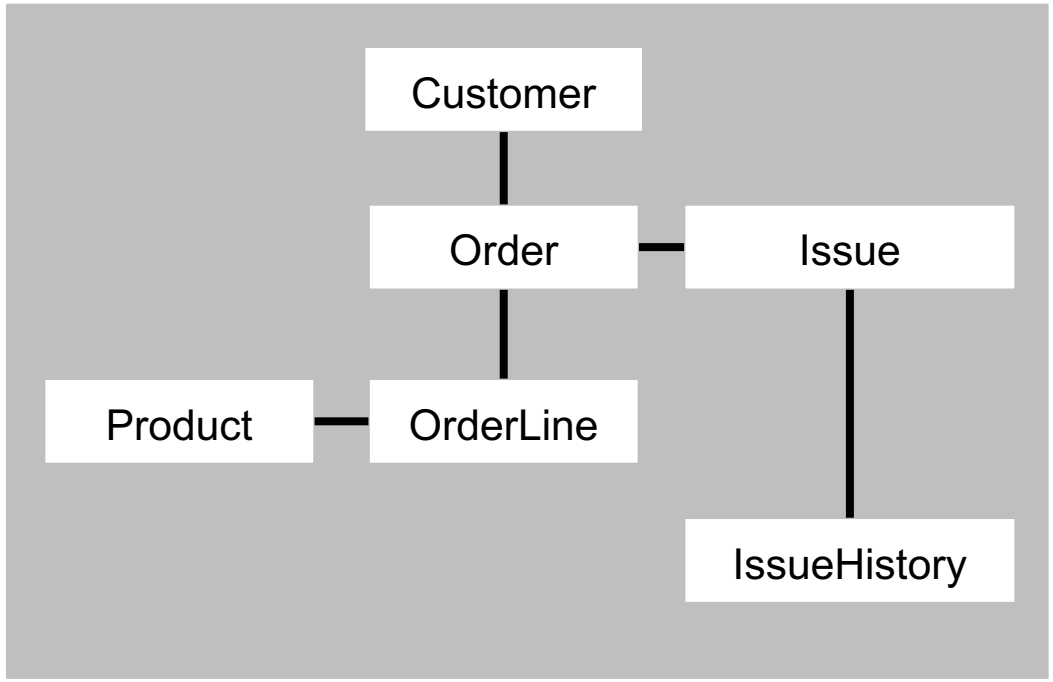
Building on the MarkLogic Foundation

- Load from any source in its natural form
- No pre-defined schema required
- Entities, messages, hierarchical data naturally as documents
- Integrated search enables discovery

NEW!

  - Expose data using lenses
  - Perform flexible data query

© COPYRIGHT 2017 MARKLOGIC CORPORATION. ALL RIGHTS RESERVED.



The 360° Call Center

- Data from disparate relational data sources
- Data cleansing in MarkLogic after ingestion
  - Easy to model with Entity Services
- Data set for board games – but could be
  - Health care – Patient, Procedure, …
  - Financial services – Trade, Counterparty, …

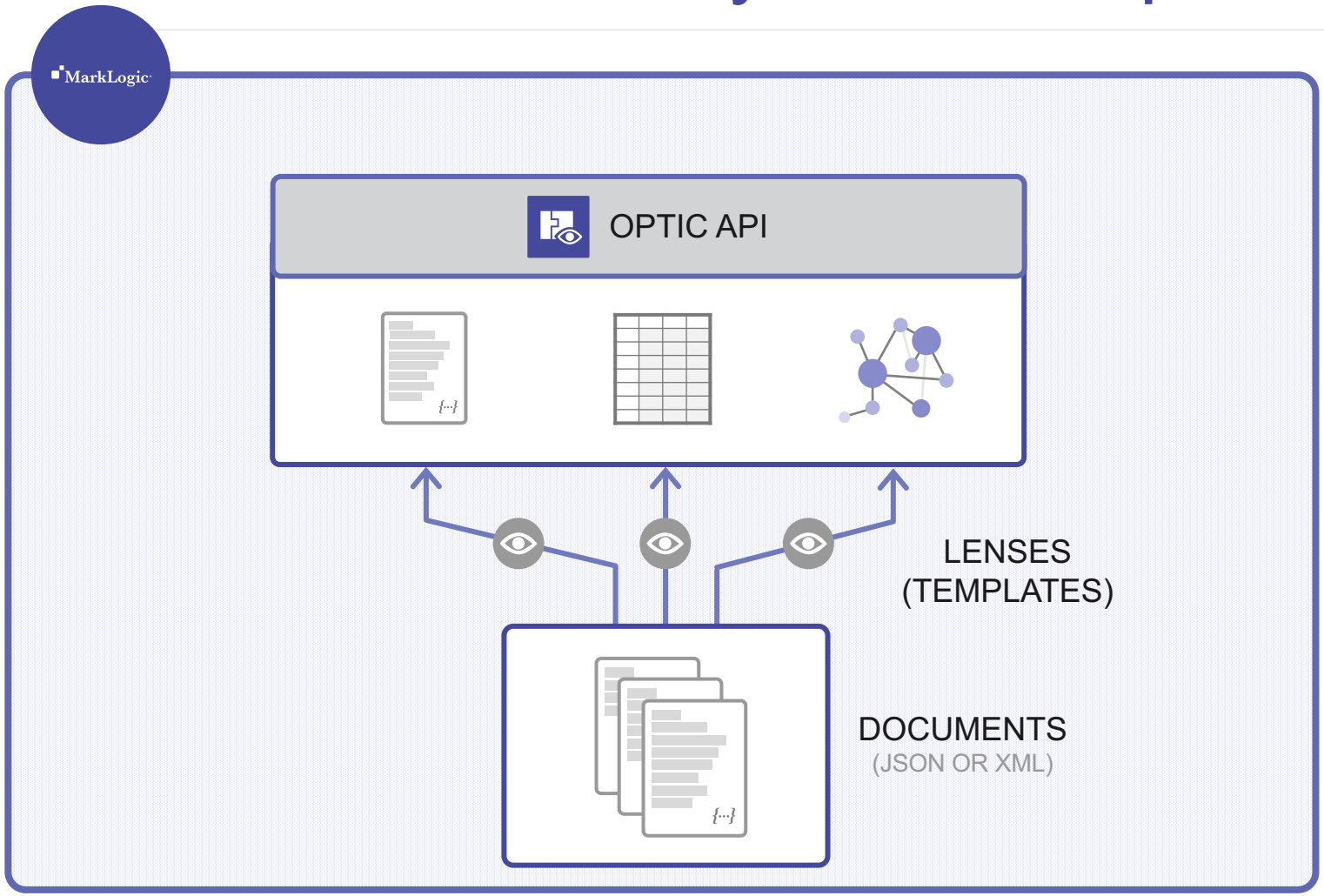
© COPYRIGHT 2017 MARKLOGIC CORPORATION. ALL RIGHTS RESERVED.

Customer 360 Challenge

- Support Team needs to understand the **Customer 360** – to take the appropriate next step:
  - Route the incoming customer call to the appropriate Support Personnel – based on order history and customer status
  - Provide supplier-specific detail about products to improve engagement
  - Review the support ticket transcript to evaluate customer service interaction

© COPYRIGHT 2017 MARKLOGIC CORPORATION. ALL RIGHTS RESERVED.

Easier Data Query With Templates and the Optic API



BENEFITS

- Single, integrated platform – single, unified query interface
- Access any shape of data, or use all together
- Combine strengths of each underlying query mechanism

© COPYRIGHT 2017 MARKLOGIC CORPORATION. ALL RIGHTS RESERVED.

MarkLogic

Optic API

- Language-integrated, multi-model joins and aggregates over rows and triples
- Fluent JavaScript, XQuery, and Java
  - leverage your language skills
  - avoid string concatenation or injection
  - modularize with variables or functions
- Extract from or construct documents
- Same engine, indexes as SQL/SPARQL
  - same d-node pushdown optimizations

```
op.fromView(…)
  .where(…)
  .select(…)
  .joinInner(
      op.fromView(…), op.on(…, …)
      )
  .groupBy(…, op.count(…, …))
  .where(…)
  .orderBy(…)
  .limit(…)
  .result();
```

Mix ‘n match

© COPYRIGHT 2017 MARKLOGIC CORPORATION. ALL RIGHTS RESERVED.

MarkLogic

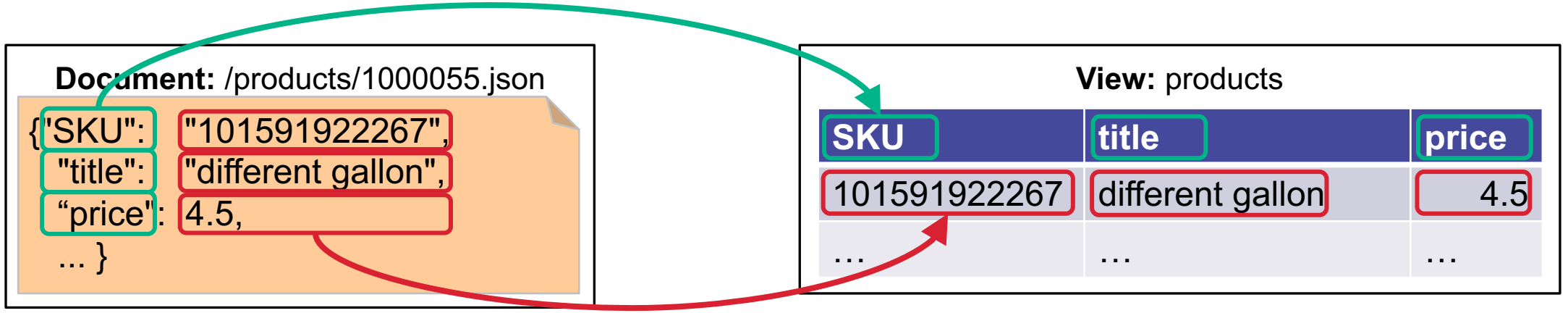
Translate Challenges Into Optic API Queries

Customer 360 Challenge	Optic API Capability
Which customers ordered the most product (by count and price)?	Joins & aggregates
Which supplier-specific product detail matches these terms?	Document query
Review the support ticket transcript to confirm that top customers are treated well	Document joins

© COPYRIGHT 2017 MARKLOGIC CORPORATION. ALL RIGHTS RESERVED.

From Source Documents To Indexed Rows

The row structures projected from a document are stored in the index



- Project the names and data types of the columns
- Project the values of columns in the rows

© COPYRIGHT 2017 MARKLOGIC CORPORATION. ALL RIGHTS RESERVED.

MarkLogic

Join and Aggregate on the Rows

Declarative joins and aggregates – find the top customers by order volume

customers

id	last_name	sales_region
1	BURDETTE	Nevada
2		
…		
92		

orders

order_date	customer	title	price
3/21/2017	1	gothic cuckoo	15.00
			9
			6

joined customers and orders

id	last_name	sales_region	order_date	customer	title	price
1	BURDETTE				gothic cuckoo	15.00
1	BURDETTE				reliable nephew	14.36
…	…				…	…

top customers

last_name	order_count	sum_price
SAYERS	11	430.81
KING	8	464.78
…	…	…

© COPYRIGHT 2017 MARKLOGIC CORPORATION. ALL RIGHTS RESERVED.

MarkLogic

Query on the Documents Providing the Rows

Query on the source document for rows– find products based on a variant detail substructure

products

SKU	title	price
101591922267	different gallon	4.5
131841112651	gothic cuckoo	15.00
159929577929	reliable nephew	14.36
…	…	…

/products/1000055.json


```
{"SKU":"101591922267",
  "title":"different gallon",
  “price":4.5,
  ...
  “detail":{
    "features": "Humorous",
    "endorsements": "Games …"}}
```


© COPYRIGHT 2017 MARKLOGIC CORPORATION. ALL RIGHTS RESERVED.

Joins on Document URIs

Join documents based on uri – retrieve chat log for issue to assess support for top customer

Issue history

title	chat
conceptual cobweb not functional	/issue-chats/40.xml

/issue-chats/40.xml

```
<support-chat>
  <chat-transcript>
  <staff>
```

customer and most recent issue with transcript

first_name	last_name	title	transcript
Emma	SCHMIDT	conceptual cobweb not functional	<support-chat> <chat-transcript> <staff> <timestamp>2017-03-27T14:36:28</timestamp> <message>Hello, how can I help you?</message> …

© COPYRIGHT 2017 MARKLOGIC CORPORATION. ALL RIGHTS RESERVED.

MarkLogic

Where Does the Data Come From?

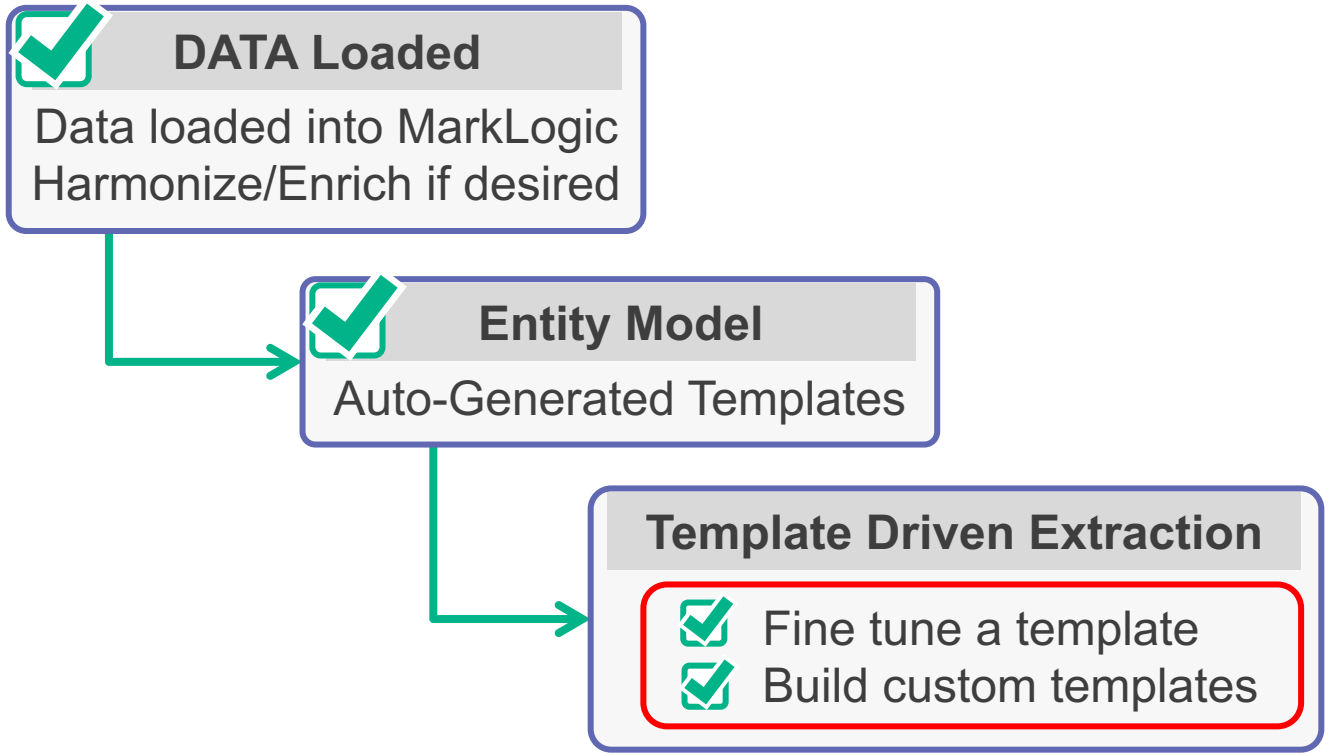
- Introducing Template-Driven Extraction (TDE)

Customer 360 Challenge	TDE Capability
The Customer Director has different needs from the customer support agent. Queries have different pieces of data.	Support for multiple views
Answering detailed questions requires a granular view of the data – but each order contains many lines, each with its own product and price.	Handling repeating rows
Two entity types (apparel & eyewear) describe product accessories, but need efficient aggregation across both.	Limited Transformation during Indexing

© COPYRIGHT 2017 MARKLOGIC CORPORATION. ALL RIGHTS RESERVED.

Simple Workflow

- Where in the lifecycle does Template generation fit?



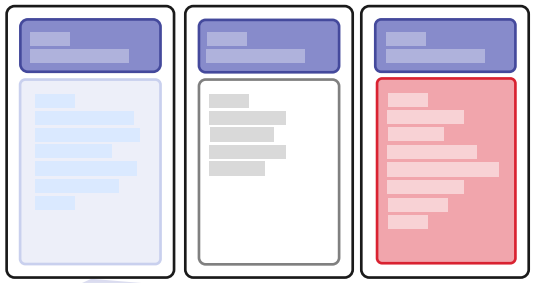
Template-Driven Extraction

- Cover advanced way of building a custom template from scratch
- Show how templates can index views and triples from documents
- Demo some powerful features like transformation in templates

© COPYRIGHT 2017 MARKLOGIC CORPORATION. ALL RIGHTS RESERVED.

MarkLogic

Documents for Orders



```
{
  id : 166
  "customer": 962
  "order_date": 2017-04-22
  "ship_date": 2017-04-30
  "lines":[
    {
      "product_id":"1000213",
      "price":35,
      "quantity":2,
      "discounted_price":35,
      "title":"meaningful
wedding"}
  ],
}
```

Template-Driven Extraction

```
{
"template":{
  "context": /id ,
  ...
    "schemaName":"customer360",
    "viewName":"orders",
    "columns":[
      {"name":"id",
       "scalarType":"integer",
       "val":"../id"},
      {
        "name":"customer",
        "scalarType":"integer",
        "val": "../customer"},
      ...
```

Indexed “Orders” View

id	customer	order_date	ship_date
…	…	…	…
…	…	…	…
…	…	…	…
…	…	…	…
…	…	…	…

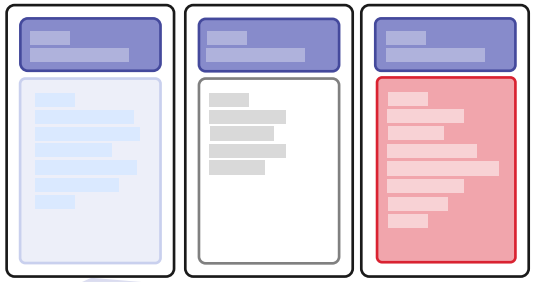
```
SELECT *
FROM customer360.orders
WHERE …

op.fromView("customer360",
"orders")
  .result()
```

© COPYRIGHT 2017 MARKLOGIC CORPORATION. ALL RIGHTS RESERVED.

MarkLogic

Repeating Order Lines



```
{
  "lines":[
   { product_id  1000213 ,
     "price" 35
     "quantity" 2
     "discounted_price":33.5
     "title":'golden clam' },
    { product_id  1000109 ,
     "price":8.99
     "quantity" 3
     "discounted_price" 6.29,
     "title":'White blade' }
  ],
  "id": 166
  ...
```

Template-Driven Extraction

```
{
"template":{
  "context": /lines/product_id ,
  ...
    "schemaName":"customer360",
    "viewName":"orderlines",
    "columns":[
      {"name":"order_id",
       "scalarType":"integer",
       "val":"../../../id"},
      {"name":"product_id",
       "scalarType":"integer",
       "val":"../product_id"},
       ...
       ]
      }
```

Indexed View

order_id	Product_id	price	quantity	discounted_price
…	…	…	…	…
…	…	…	…	…
…	…	…	…	…
…	…	…	…	…
…	…	…	…	…

```
SELECT …
FROM orders
WHERE …
JOIN
  orderlines ON…
GROUP BY…
```

© COPYRIGHT 2017 MARKLOGIC CORPORATION. ALL RIGHTS RESERVED.

MarkLogic

Template-Driven Extraction

Flexible Data Projection From Documents Into Indexes

Data Projection

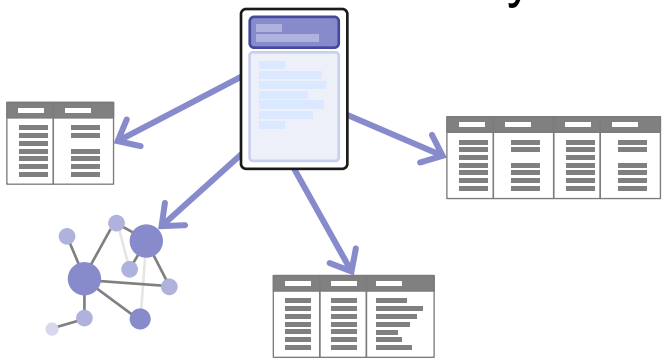
- Project data from documents into the index
- Create SQL Views or semantics Triples
- Query from SQL, SPARQL, ODBC, or Optic
- Simple language with context-based projection

Transactional & Data Provenance

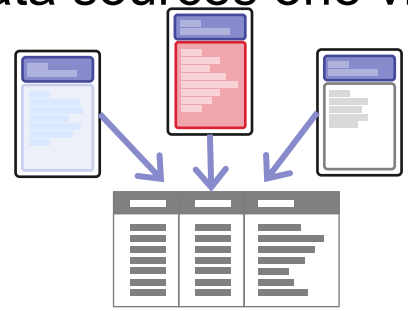
- Rows and Triples are updated with a document update
- Document is intact, not transformed nor modified

Multiple Projections and Transformation

- Transforming data on index
- One document many Views



Many data sources one view



© COPYRIGHT 2016 MARKLOGIC CORPORATION. ALL RIGHTS RESERVED.

MarkLogic

Different Documents

Board Game

```
{
  "game_id": "1000085",
  "SKU": "101897294018",
  "title": "wonderful
footnote",
  "price": 25,
  "years_active": 0,
  ...
}
```

Game Accessory

```
{
  "id": "42",
  "sku": "416739731622",
  "title": "Cards",
  "price": 10
}
```

Template-Driven Extraction

```
{
"template":{
  "context":"/game_id",
  ...
}
```

```
{
"template":{
  "context":"/id",
  ...
}
```

Single Products View

id	sku	title	price
...	...	...	...
...	...	...	...
...	...	...	...
...	...	...	...
...	...	...	...

© COPYRIGHT 2017 MARKLOGIC CORPORATION. ALL RIGHTS RESERVED.

MarkLogic

Documents

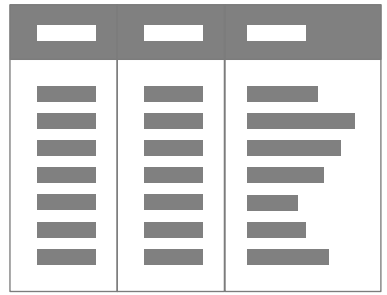
Game Accessory

```
{
  "id": "42",
  "sku": "416739731622",
  "title": "Cards",
  "price": 10,
  "game_ID": "1000146"
}
```

Template-Driven Extraction

```
<template>
  ...
  <context>/id</>
  ...
  <triple>
    <subject><val>
     sem:iri($prefix-product||../id)</>
    <predicate><val>
     sem:iri($prefix-pred||"isAccessoryFor")</>
    <object><val>
     sem:iri($prefix-product||../game_ID)</>
  </triple>
  ...
</>
```

Triple Index



```
PREFIX prefix-pred: <http…
SELECT ?product ?accessory
WHERE { ?accessory ... }
```

```
op.fromTriples([
  op.pattern(...),
  op.pattern(...)
])
```


© COPYRIGHT 2017 MARKLOGIC CORPORATION. ALL RIGHTS RESERVED.

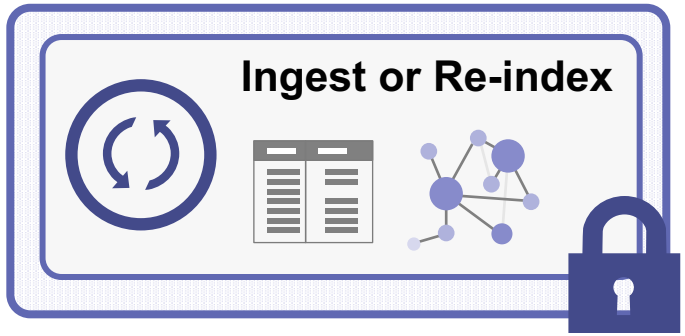
MarkLogic

DEMO

MarkLogic

Template-Driven Extraction Highlights / Recap

- Indexed views and triples:
  - Updated with documents. **ACID**
  - Inherent **Backup**, **Replication, Failover**, and so on
  - Document level **Security** inherited
- No physical extraction, copying, changes to the underlying document
- One source with multiple views
- Multiple sources with one view
- Update your template with Schema changes
- Single document management
  - Templates actively keep the projections up-to-date



© COPYRIGHT 2017 MARKLOGIC CORPORATION. ALL RIGHTS RESERVED.

MarkLogic

Optic API Highlights / Recap

- The MarkLogic-idiomatic, language-integrated interface to the SQL / SPARQL engine
  - Access to rows, triples, and range indexes as well as documents
- Relational operations
  - Row joins and grouping with aggregates
- Document-oriented operations
  - Constraining queries, document joins, XPath extraction, node and sequence construction

© COPYRIGHT 2017 MARKLOGIC CORPORATION. ALL RIGHTS RESERVED.

MarkLogic

Integrated Queries Over Your Data

- Transformative integration of multiple data sources and structures by query
  - In the call center scenario, breaking the barrier to a new level of customer engagement
- What are the barriers for your organization?
  - How could integrated query transform your enterprise?

© COPYRIGHT 2017 MARKLOGIC CORPORATION. ALL RIGHTS RESERVED.

MarkLogic

Q & A