

# MASTERING ENTERPRISE DATA FOR CONSISTENCY & ACCURACY

Christy Haragan, Principal Sales Engineer, MarkLogic  
Justin Makeig, Director, Product Management, MarkLogic



Hello, my name is Justin

- Product Manager for 8+ years at MarkLogic
- Background in consulting and web development
- Focus on software architecture, APIs, and data integration

[jmakeig@marklogic.com](mailto:jmakeig@marklogic.com)

<https://github.com/jmakeig>

Hello, my name is Christy

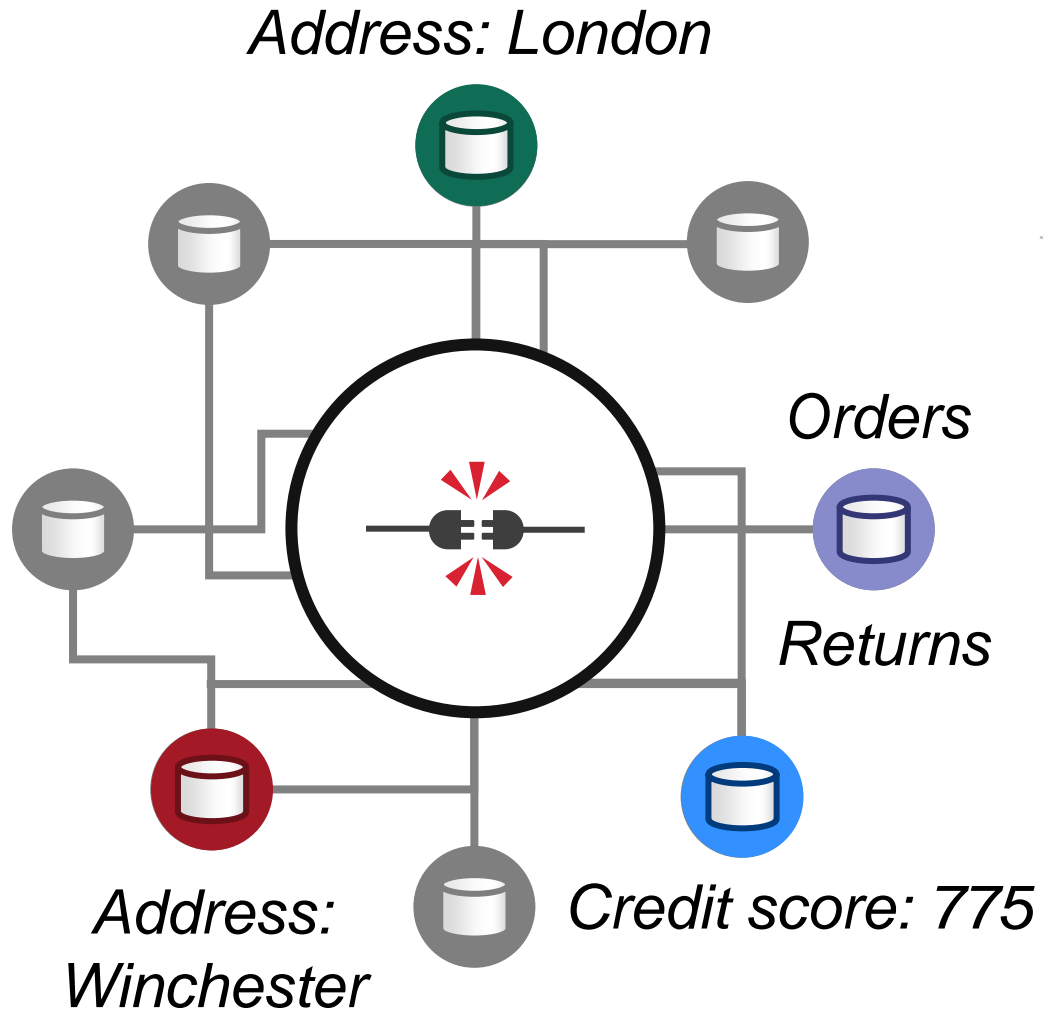
- 8 years experience in the data management space
- Previously a MDM Software Engineer, and Software Architect

[christy.haragan@marklogic.com](mailto:christy.haragan@marklogic.com)

<https://github.com/christyharagan>

# Agenda

- What is Master Data and why should I care?
- Traditional approaches to Master Data Management
- How MarkLogic changes MDM
- Use cases: Techniques and tips
- Q&A

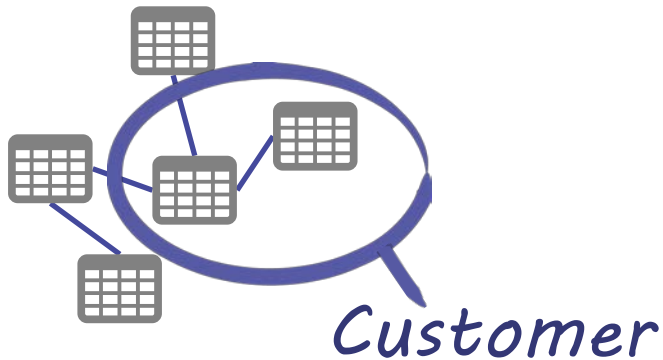


## “Christy Haragan”

- What is “Christy Haragan”? Who decided that?
- What is our relationship with her?
- What has she purchased? Returned? Received support on?
- Where does she live? With whom? How do I contact her? When have we contacted her?
- Is she a “high-value” customer? What risk does she represent?
- Who is authorized to see Christy’s data? Who can change it? Under what conditions?

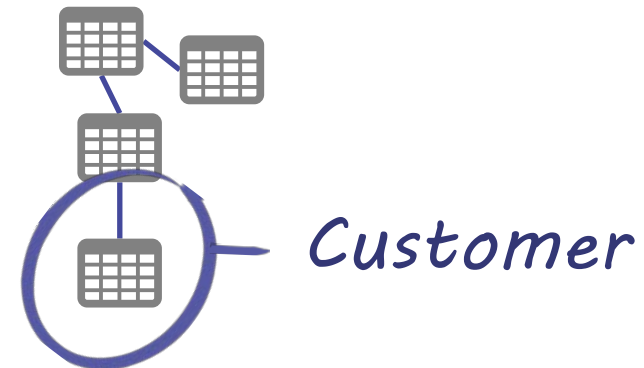
# Relational strips important context from data

System A



```
CREATE TABLE Customers (  
  CustID          NUMBER(6) PRIMARY KEY,  
  StartDate      DATE NOT NULL  
  PartyID        NUMBER(6)  
  ...  
)
```

System B



```
CREATE TABLE CUST_MASTER (  
  CID            VARCHAR2(40) PRIMARY KEY,  
  Evt_Dt        TIMESTAMP  
  Cst_Type      VARCHAR2(120)  
  ...  
)
```

# TRADITIONAL APPROACHES TO MDM



# Master Data Registry



**Customer: Christy Haragan**






←  ERP • PARTY\_MASTER • HARA\_C

←  CRM • customers\_1 • 100067

←  Credit • http://... • 9930-221

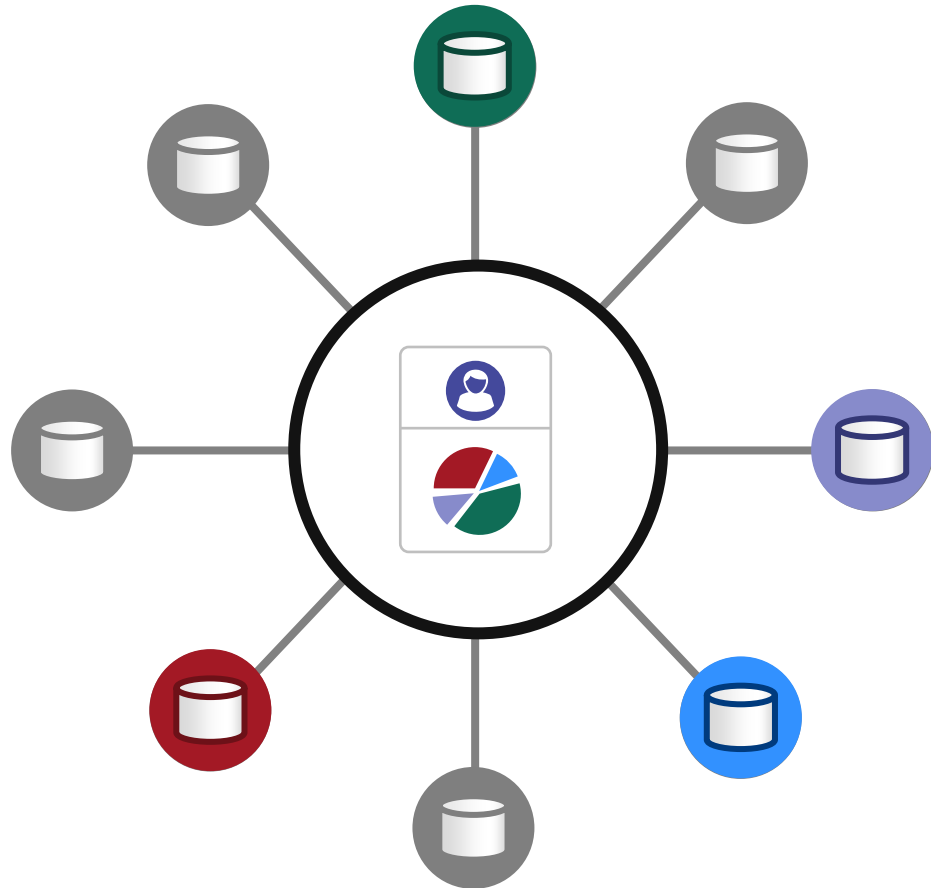
←  CRM2 • refcustview • H\_3325

## Repository of *references* to records in source systems

-  Easy to start
-  Clear separation from live systems
-  Difficult to query, update
-  Separation of context and data
-  Dependence on source systems for quality, governance, security

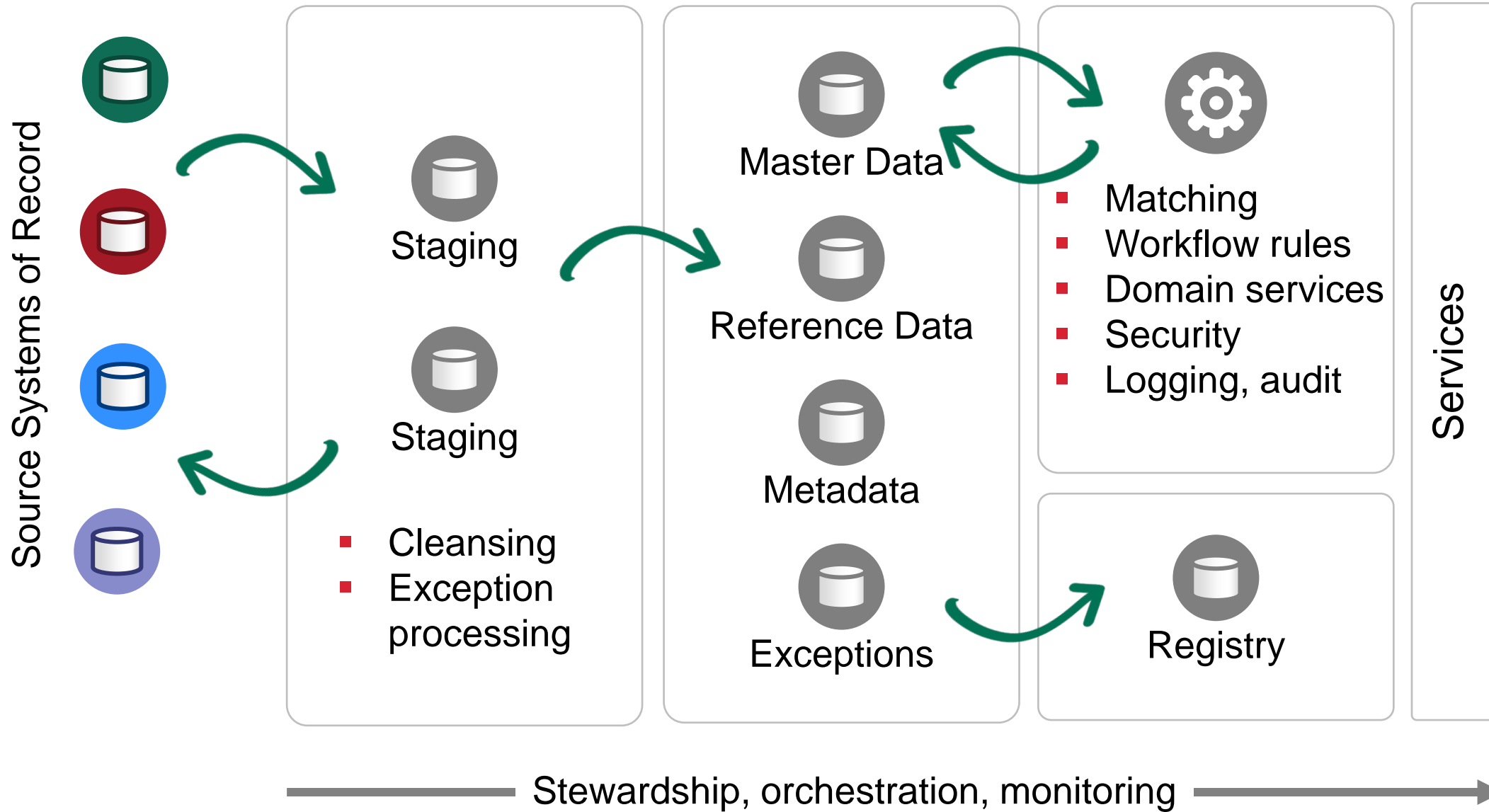


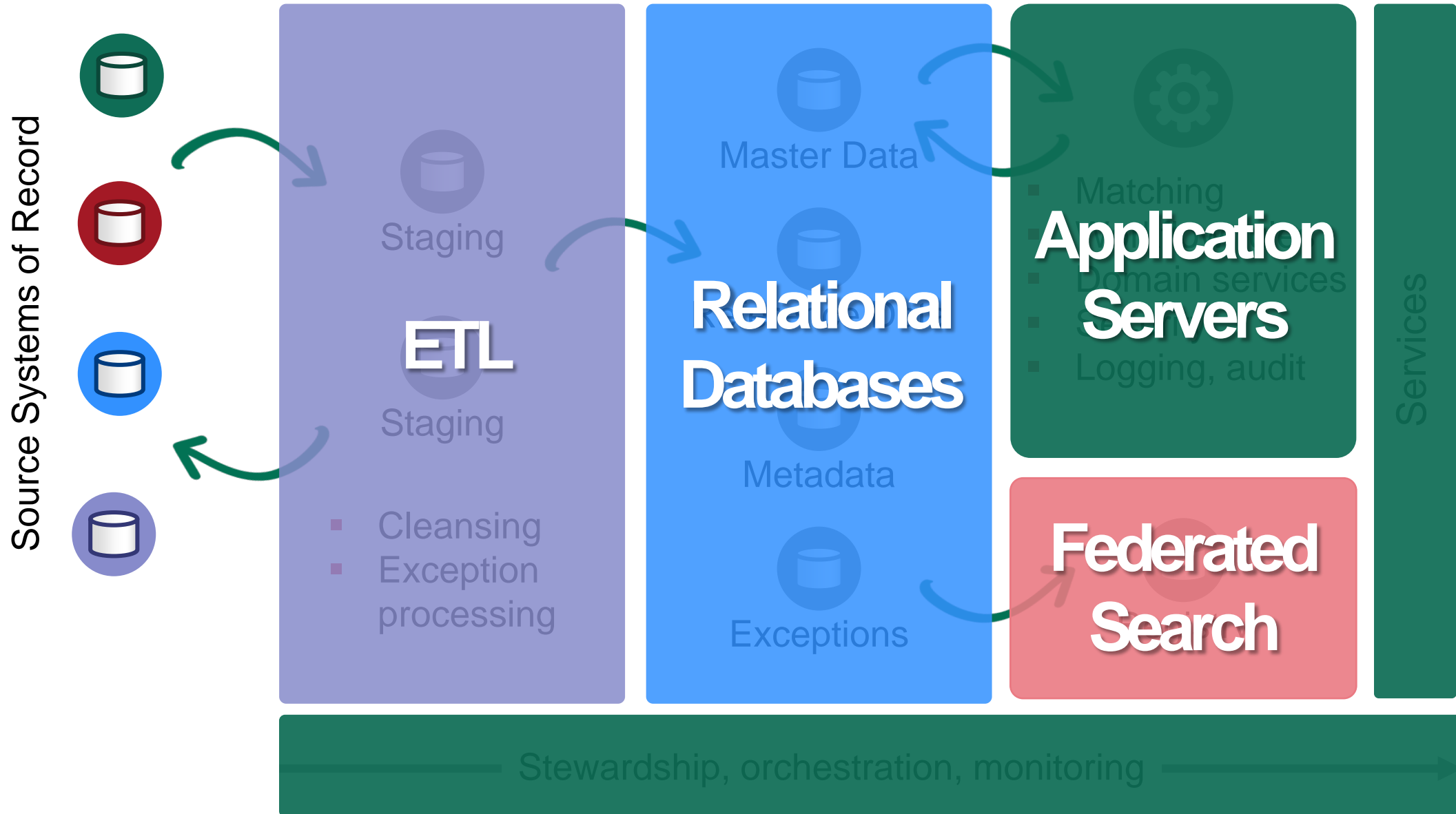
# Master Data Hub



## Aggregation of data from source systems

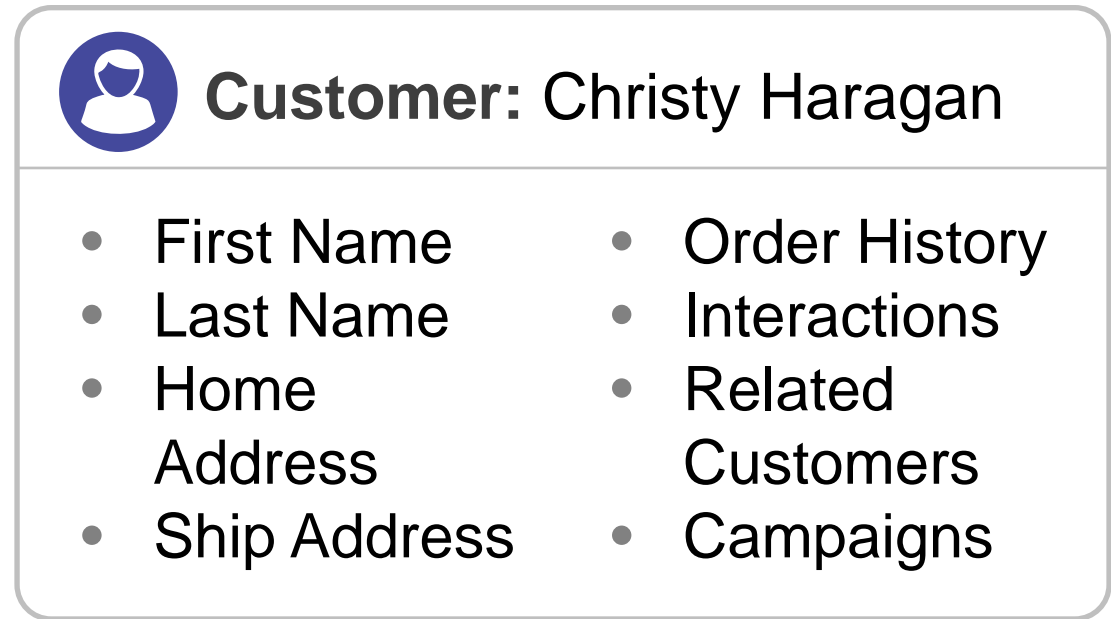
- ✓ Authoritative, bi-directional
- ✓ Centralized quality, governance, security
- ✓ Rich queries, aggregates, enrichment, and transactions
- ✗ Technical canonicalization
- ✗ Political collaboration





# Shortcomings of Traditional MDM Approaches

- Canonical models require political and technical alignment
- Single version of the truth, but which version?
- Provenance: show your work
- Changing requirements, changing data
- “Unstructured”, unknown data



A customer profile card for Christy Haragan. The card has a header with a person icon and the name. Below the header is a list of attributes and related items, organized into two columns.

**Customer: Christy Haragan**

- First Name
- Last Name
- Home Address
- Ship Address
- Order History
- Interactions
- Related Customers
- Campaigns

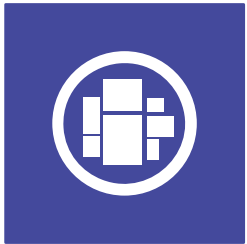


# Implications of Traditional MDM Approaches

- **Time to Value:** Average ROI for an MDM implementation is *3 years*
- **Inflexibility:** Snapshot is unable to support multiple Business Units, domains, and changing requirements
- **Cost:** ETL is brittle, expensive and time-consuming to build and maintain
- **Accountability:** Difficult or impossible to explain past decisions on current data
- **Complexity:** Orchestration, governance, and security across many moving parts

# A NEW APPROACH TO MASTER DATA MANAGEMENT





FLEXIBLE  
**DATA MODEL**

---

Capture data and context  
without having to rigorously  
model upfront



UNIVERSAL  
**INDEXING**

---

Discover data and project  
views of business entities in  
real-time



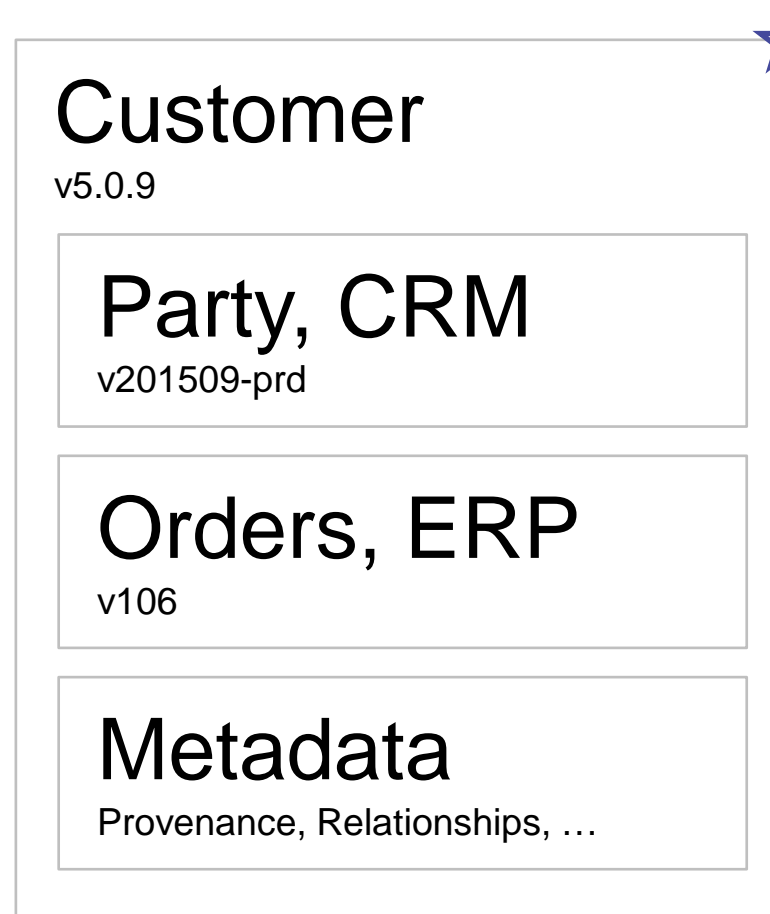
TRUSTED  
**MANAGEMENT**

---

Enforce governance, security,  
quality across the entire  
lifecycle

# Thinking in Entities

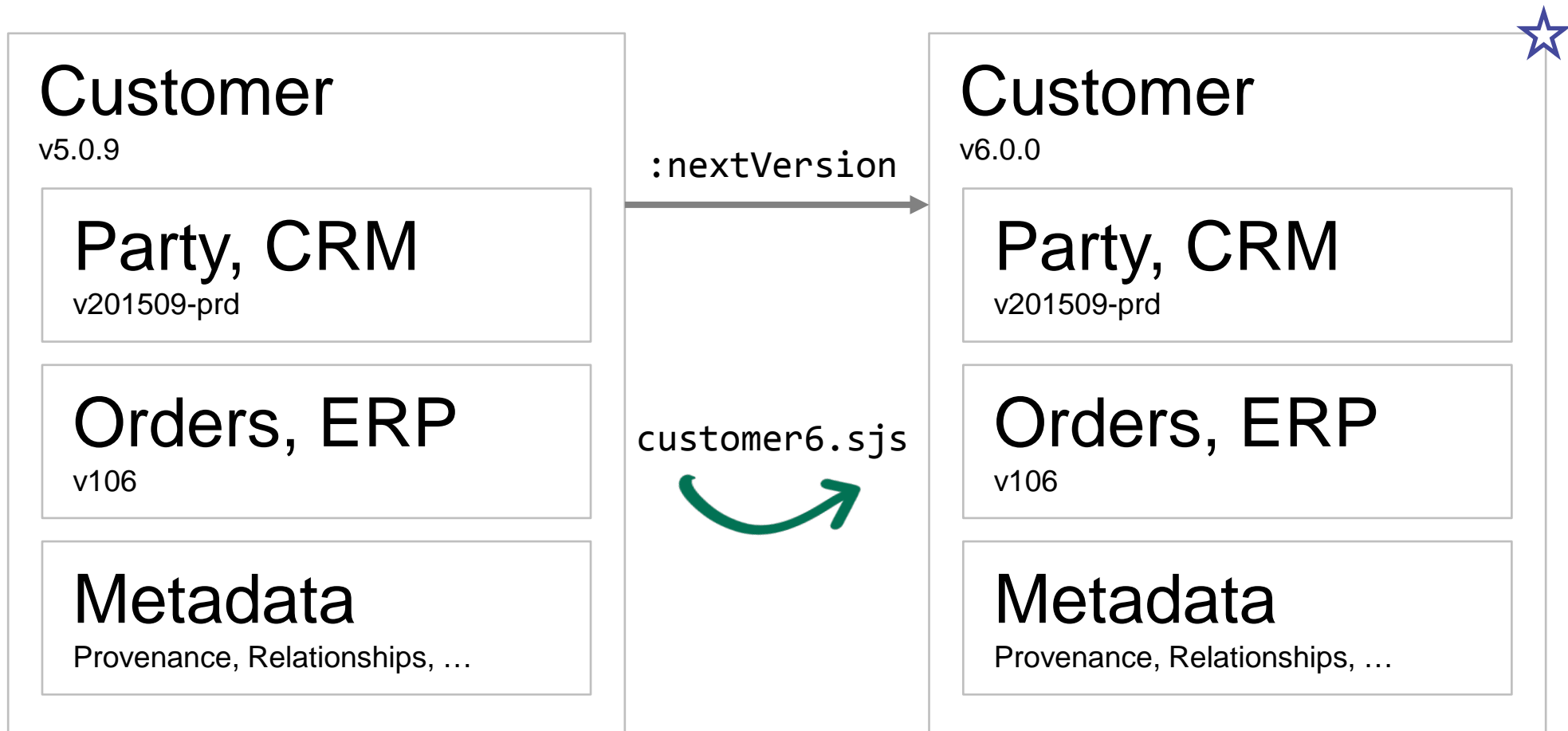
*Source  
systems  
"as-is"*

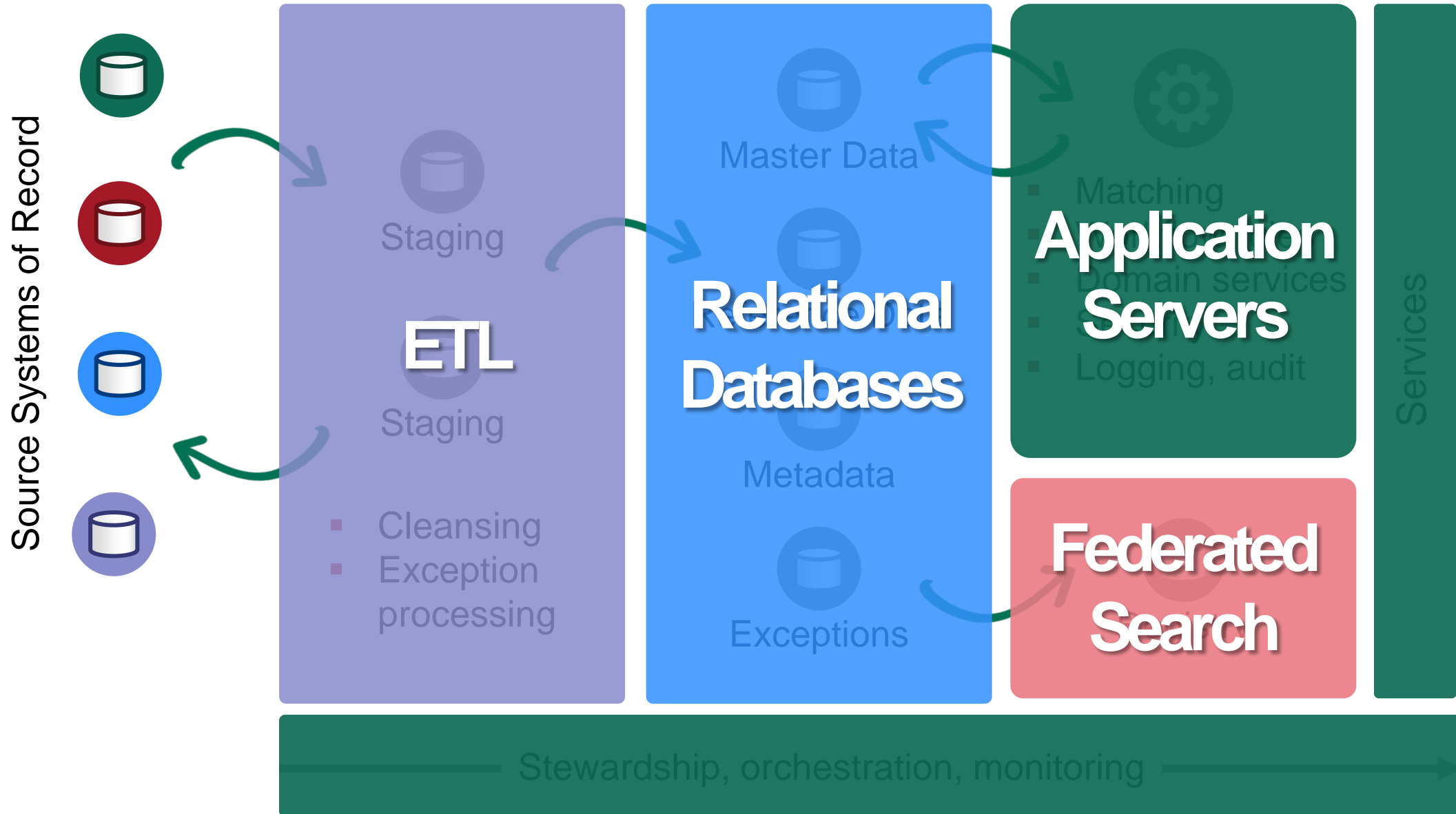


- Transactional updates
- Granular security
- Indexes
- Bitemporal history

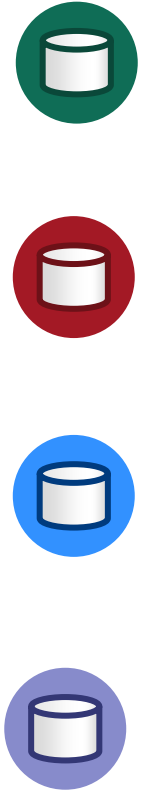


# Thinking in Entities: Versioning

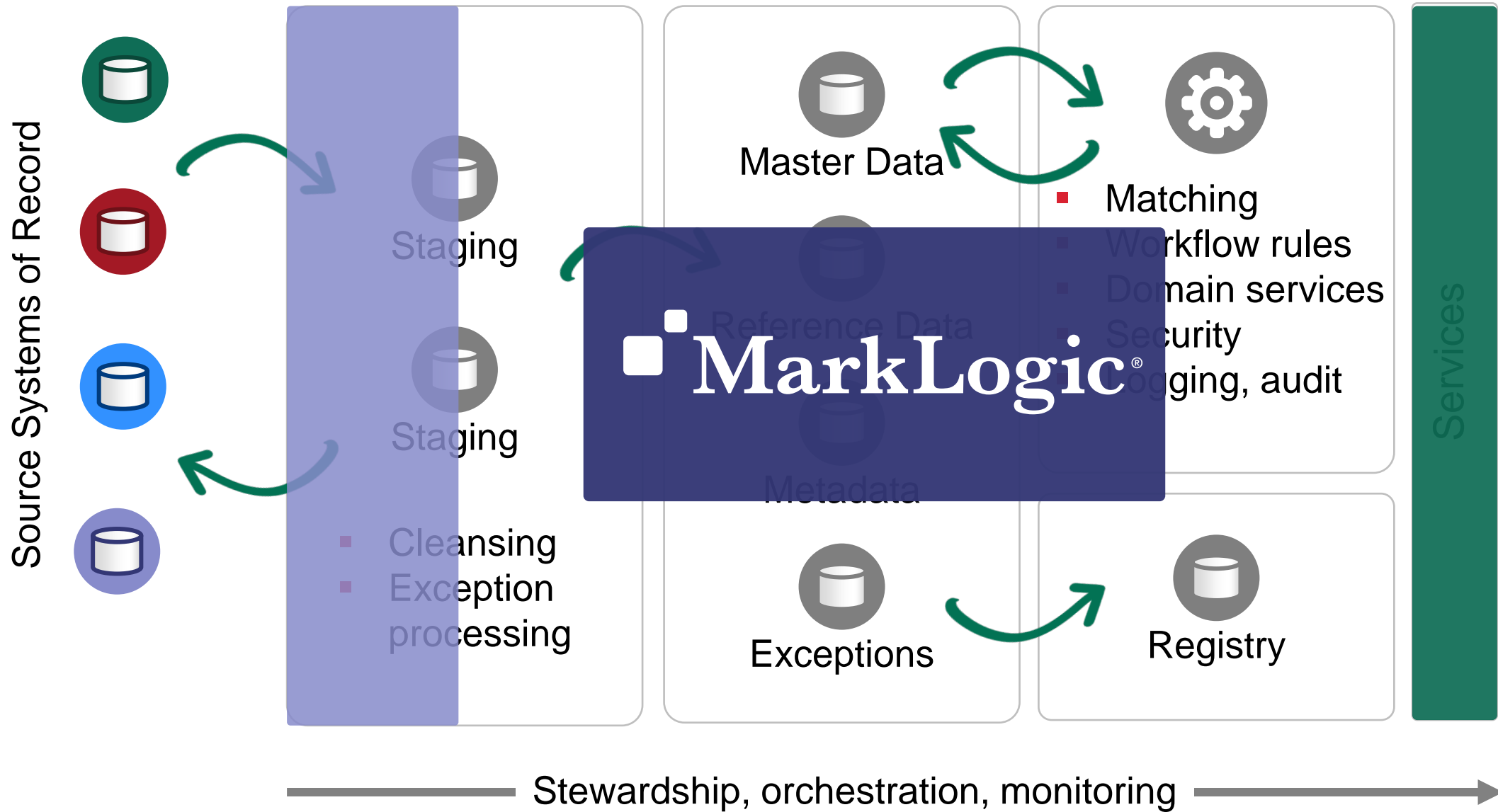


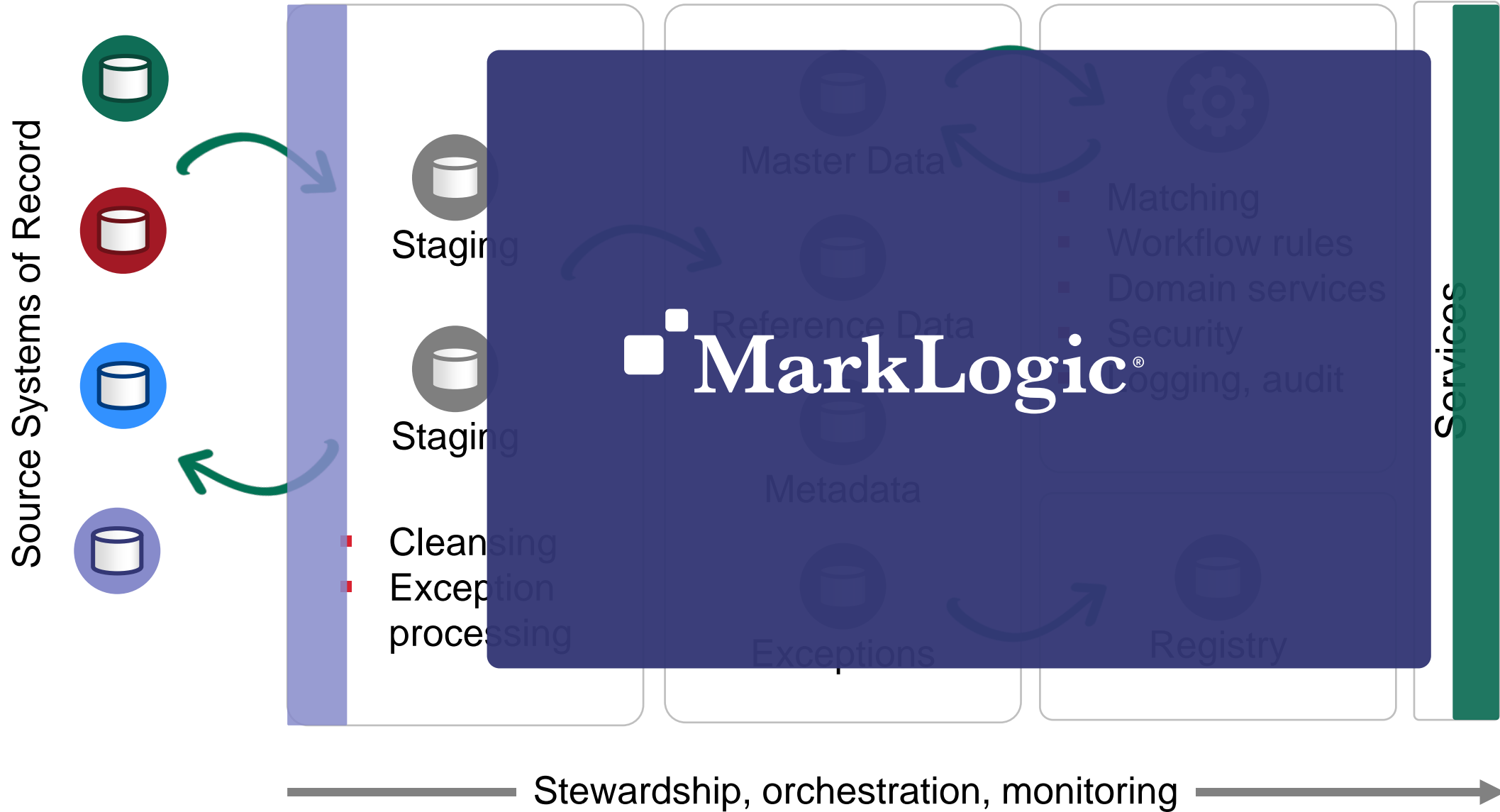


Source Systems of Record



Stewardship, orchestration, monitoring →





# MASTER DATA MANAGEMENT IN PRACTICE



# Customers Leveraging MarkLogic for MDM

## Capability

---

- Time to Value
- Evolving Business Requirements
- 5 year Replacement Cycle
- Legacy Systems Retirement
- Decision Accountability

## Customers

---



HealthCare.gov

Mitchell1



KPMG

aetna®

# Our customers are already doing this today...



## **Aetna**

Hub for sharing master data among hundreds of source systems and hundreds of subscribers.



## **US Combatant Command**

Secure sharing, exploitation, and analysis from dozens of sources at HQ, theater operations centers, and detached users.



## **Top Entertainment Company**

Apply corporate standards for creation and use of digital assets from production through distribution across the global enterprise.



## **Healthcare.gov**

Insurance marketplace and exchange hub for millions of consumers, thousands of providers, dozens of stakeholder agencies.



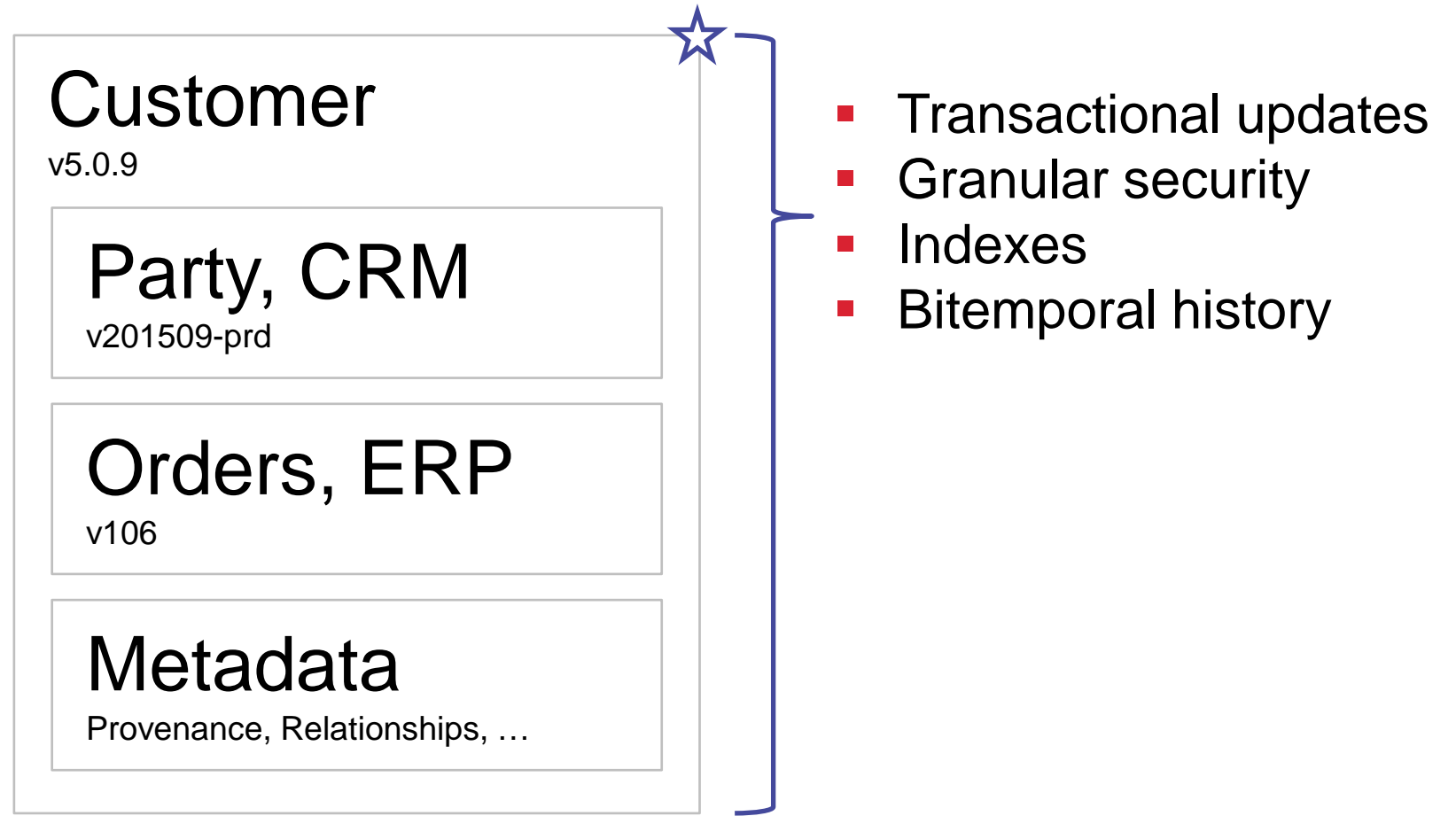
# Techniques and Tips

- **Envelope Pattern:** Model what you need when you need it
- **Common Services Layer:** The distributed Hub
- **Semantics:** Relating to your Data
- **Decision Accountability:** Bi-temporal and Tiered Storage

# Techniques and Tips – Envelope Pattern

- Data Modeling can easily take up 30% of a project's resources and is commonly cited as a key reason for project failure
- Need to be able to maintain the original data in context so systems can continue to use it
- Have to be able to expand the data model to meet existing and new requirements
- Solution:
  - The envelop pattern: Leave you data as is and wrap it with the information that you need

# Managing Entities in Envelopes



# Techniques and Tips – Envelope Pattern

```
<env:customer-envelope>
```

```
  <customer xmlns="...">
```

```
    <uuid>123abc...</uuid>
```

```
    <first-name>Christy</first-name>
```

```
    <last-name>Haragan</last-name>
```


```
  </customer>
```

```
<env:metadata>...</env:metadata>
```

```
<crm:party version="v201509-prd">...</crm:party>
```

```
<erp:orders ...>...</erp:orders>
```

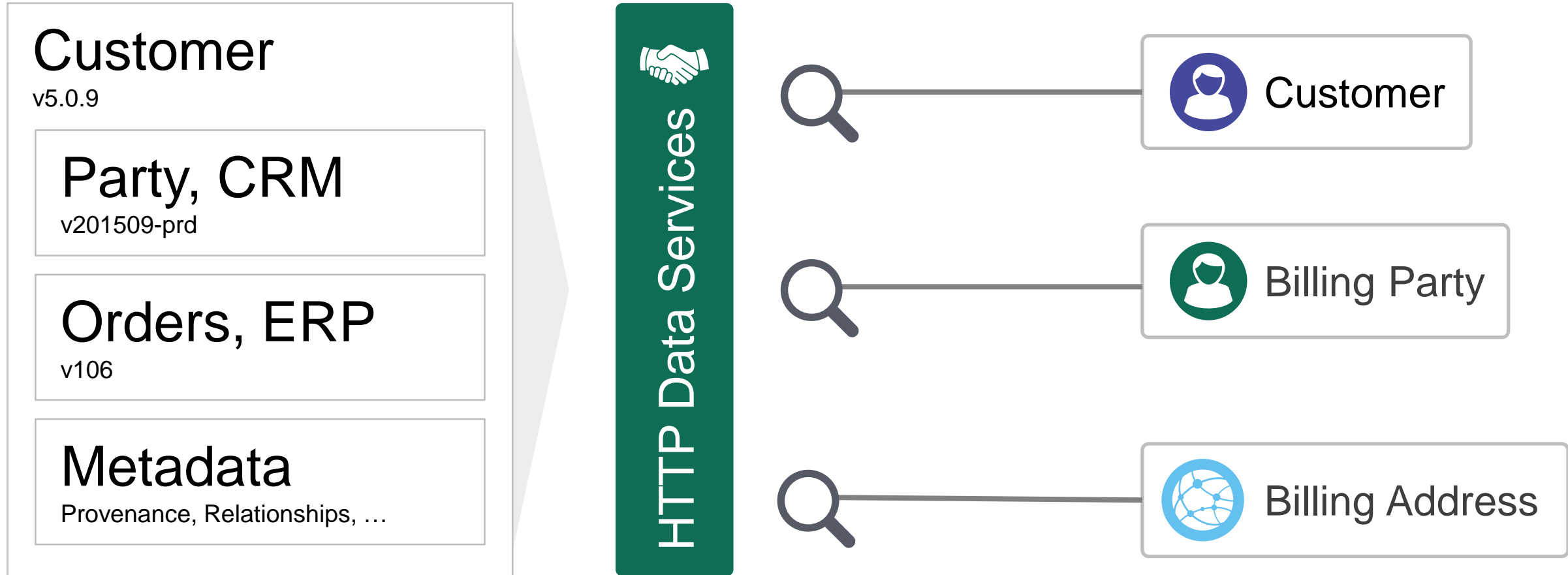
```
</env:customer-envelope>
```

- 
- Zero to many
  - Indexes
  - Bitemporal history

# Techniques and Tips – Common Services Layer

- Business Units need to own and manage data that's important to them in the way that works best for their requirements
- The enterprise needs to be able to enforce consistent access, visibility and quality requirements on key pieces of data

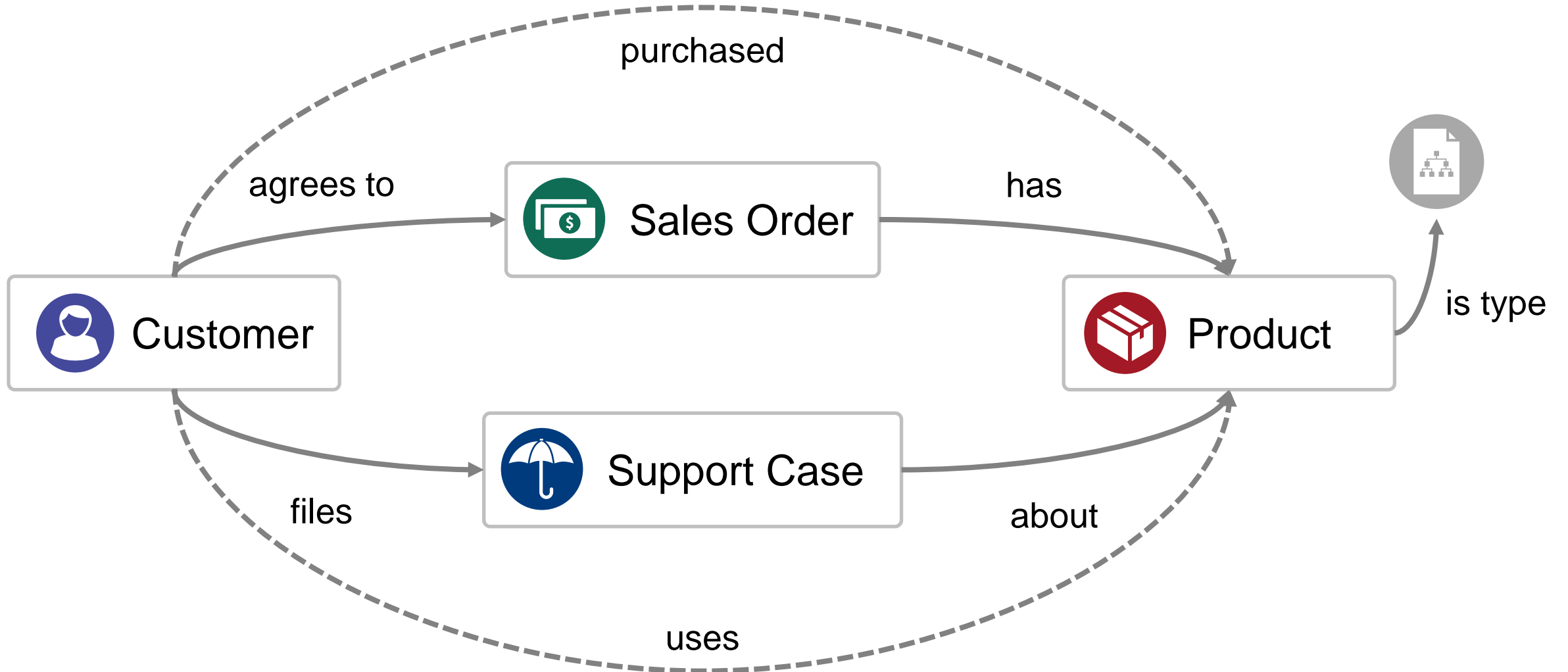
# Techniques and Tips – Common Services Layer



# Techniques and Tips - Semantics

- Not every organizational unit within the enterprise will have the same naming conventions, cardinality rules, or perspective on master data
- The enterprise and each business unit may have one or more taxonomies or ontologies that apply to their data
- Solution:
  - Semantics can be used to model relationships between data across the enterprise and allow OUs to relate their data to others and vice-versa
  - Key relationships: sameAs, belongsTo, parentOf, childOf

# Semantic Relationships

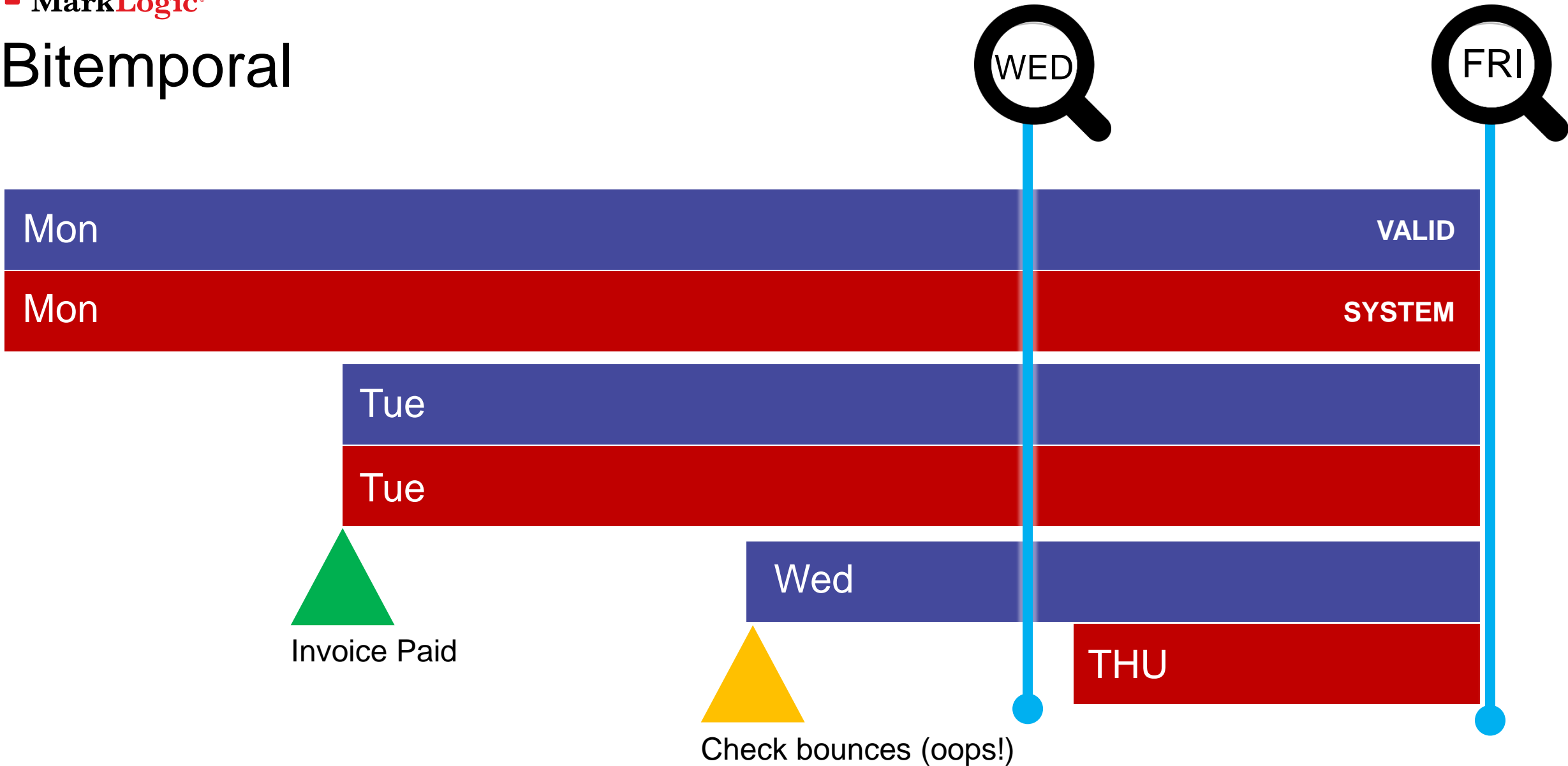




# Techniques and Tips – Decision Accountability

- Businesses Need to
  - Understand why a decision was made
    - Based on what data
    - For regulatory and compliance reasons
    - For back testing
  - Manage the costs of doing so

# Bitemporal



# Tiered Storage

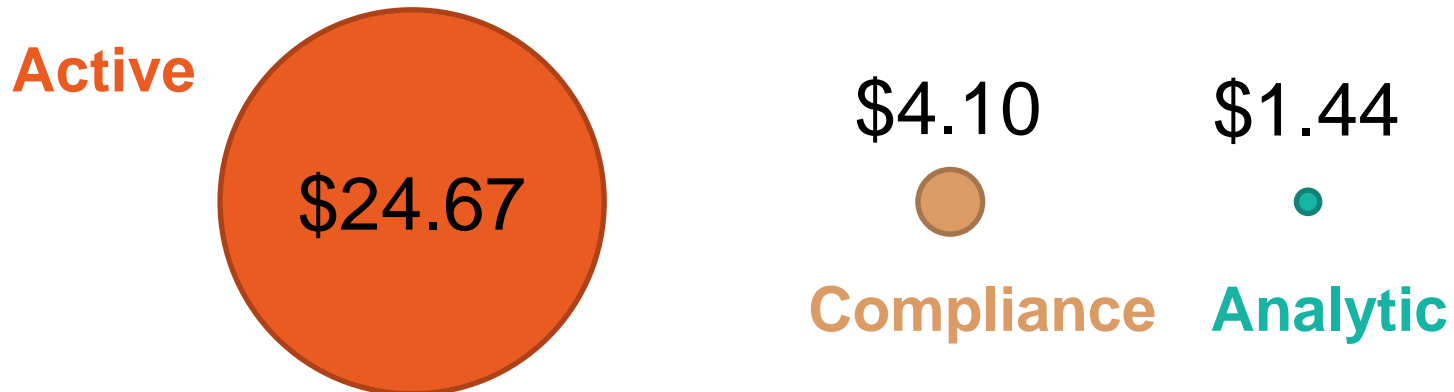
Total Size (TB)



Total Cost (\$000)



Effective Unit Cost (\$/GB)



# Conclusions

# Additional Resources

- MarkLogic University Training - <http://www.marklogic.com/training/>
- MarkLogic Developer Site – <http://developer.marklogic.com>
- MarkLogic Data Modeling - <https://developer.marklogic.com/learn/data-modeling>
- MarkLogic Data Hub Framework - <http://marklogic.github.io/marklogic-data-hub/>

QUESTIONS?

