



#MongoDBWorld

Creating a single view: Overview and data analysis

Brian D. Goodman

Enterprise Architect, MongoDB

What Is He Going To Talk About?

Creating A Single View

Part
1

Overview &
Data Analysis

Part
2

Data Design &
Loading
Strategies

Part
3

Securing Your
Deployment

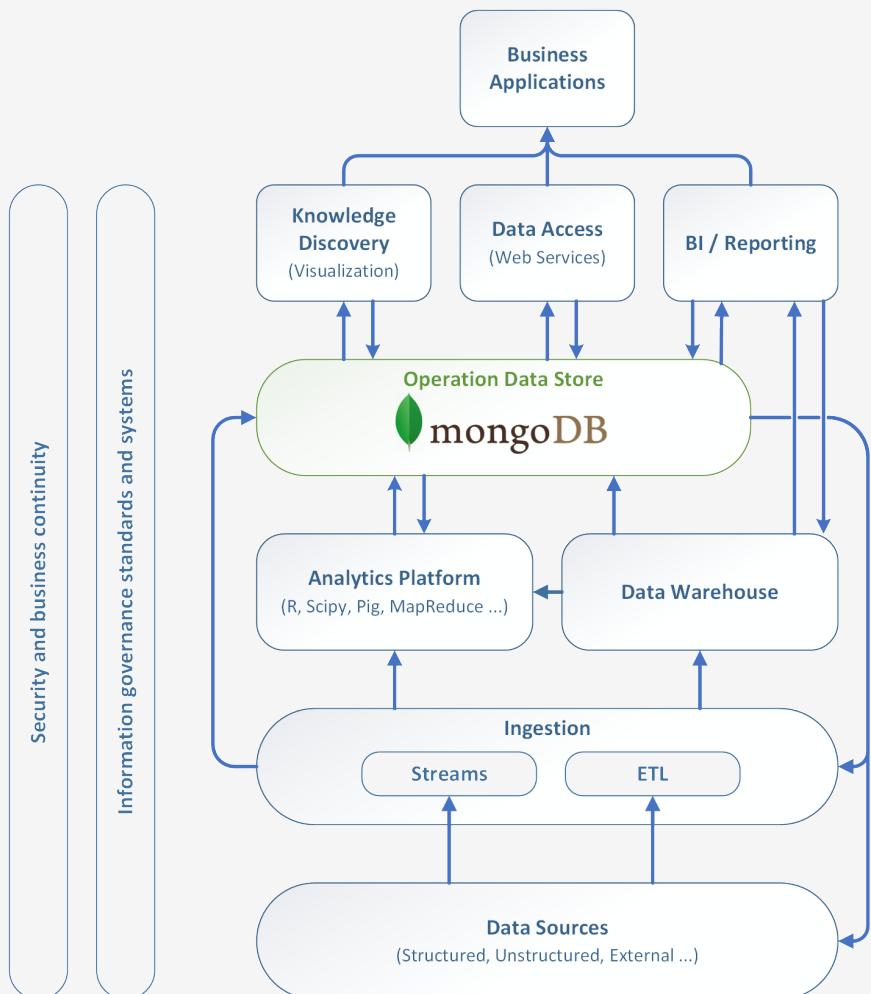


What are we trying to do?

How do approach doing it?

How do you navigate to success?

A data platform is common process, tooling and management from ingestion to presentation



- **Integrating technologies**
- **Streamlining ingestion**
- **Centralizing access**
- **Supporting analytics**

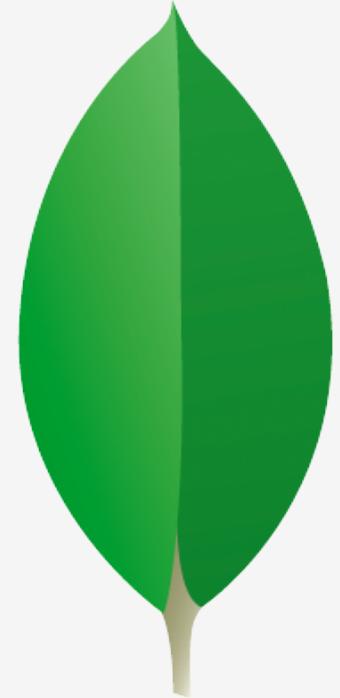
I have sat in your chair looking for answers



13+ years

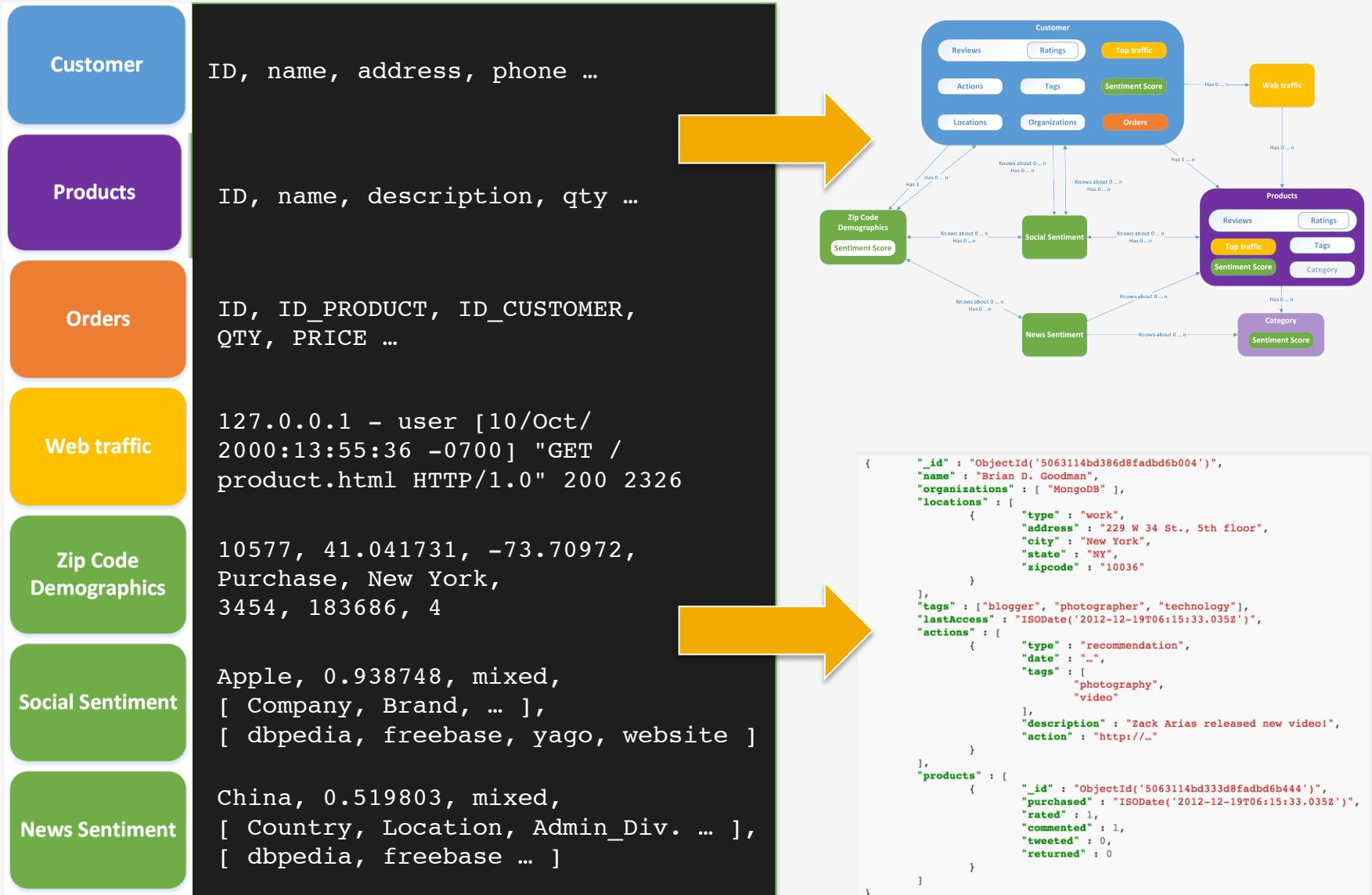


1.5 years



Present

Source: Elephant by James Fujii @ <http://jamesfujiihistoryartist.blogspot.com/2010/05/dancing-elephant-april-2010.html>
Source: Gorilla by Luigi Lucarelli @ <http://loaduniverse.blogspot.com/2012/04/gorilla-sketch.html>



Single view use cases abound



Financial Services

Creating a single view of a product, trader, client

Generating recommendations to sales people

Identifying anomalous patterns

Data sources: client profile, trade data, market data, service catalog



Retail

Creating a single view of a product or customer

Generating recommendations to customers

Proactive, personal marketing to customers

Data sources: customer profile, loyalty program, inventory, product catalog



Healthcare

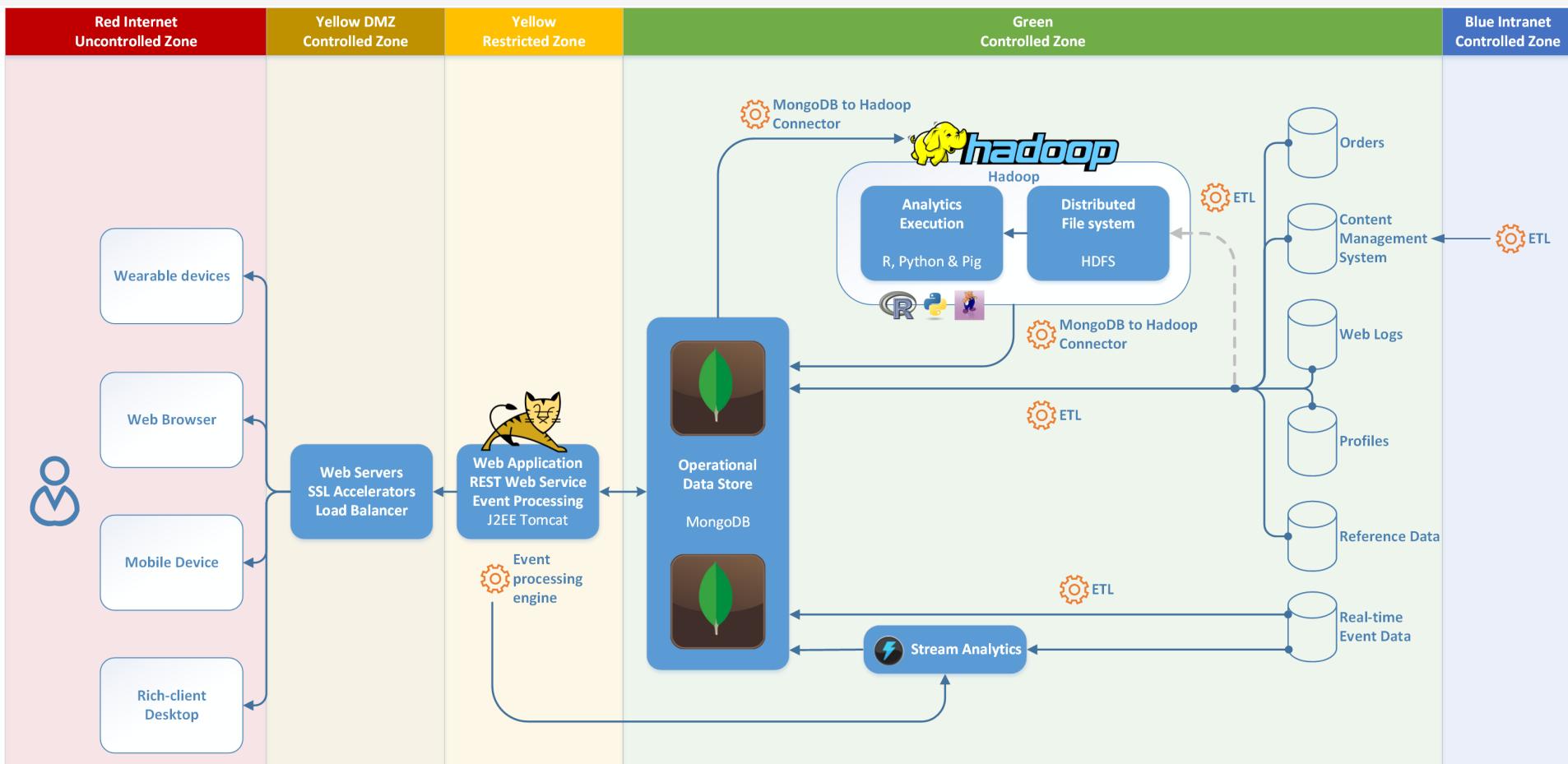
Creating 360-degree patient view

Managing electronic health-care records for patients

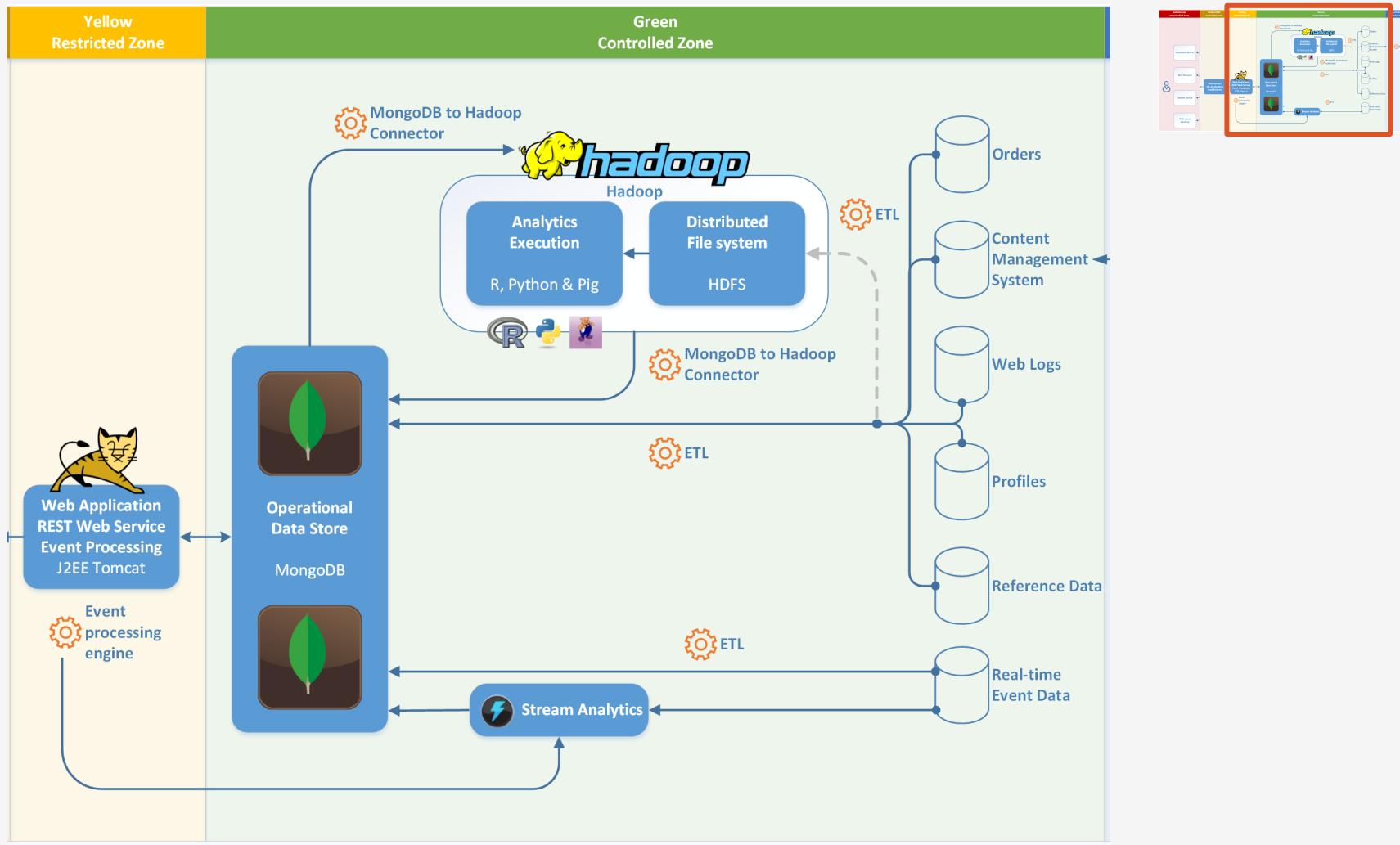
Population management for at-risk demographics

Data sources: patient records, sensor data, care center metrics

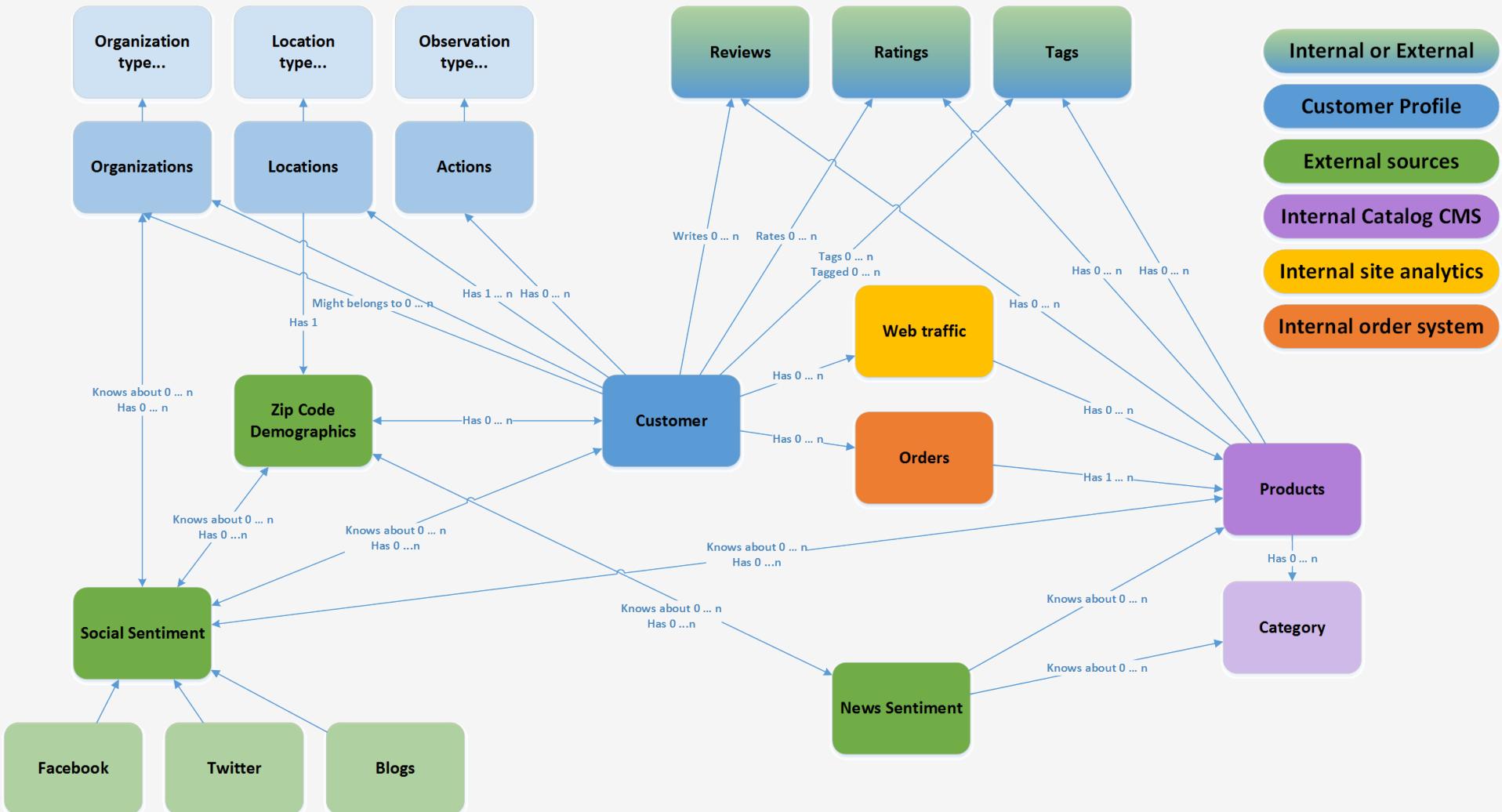
High-level architecture for a Single View



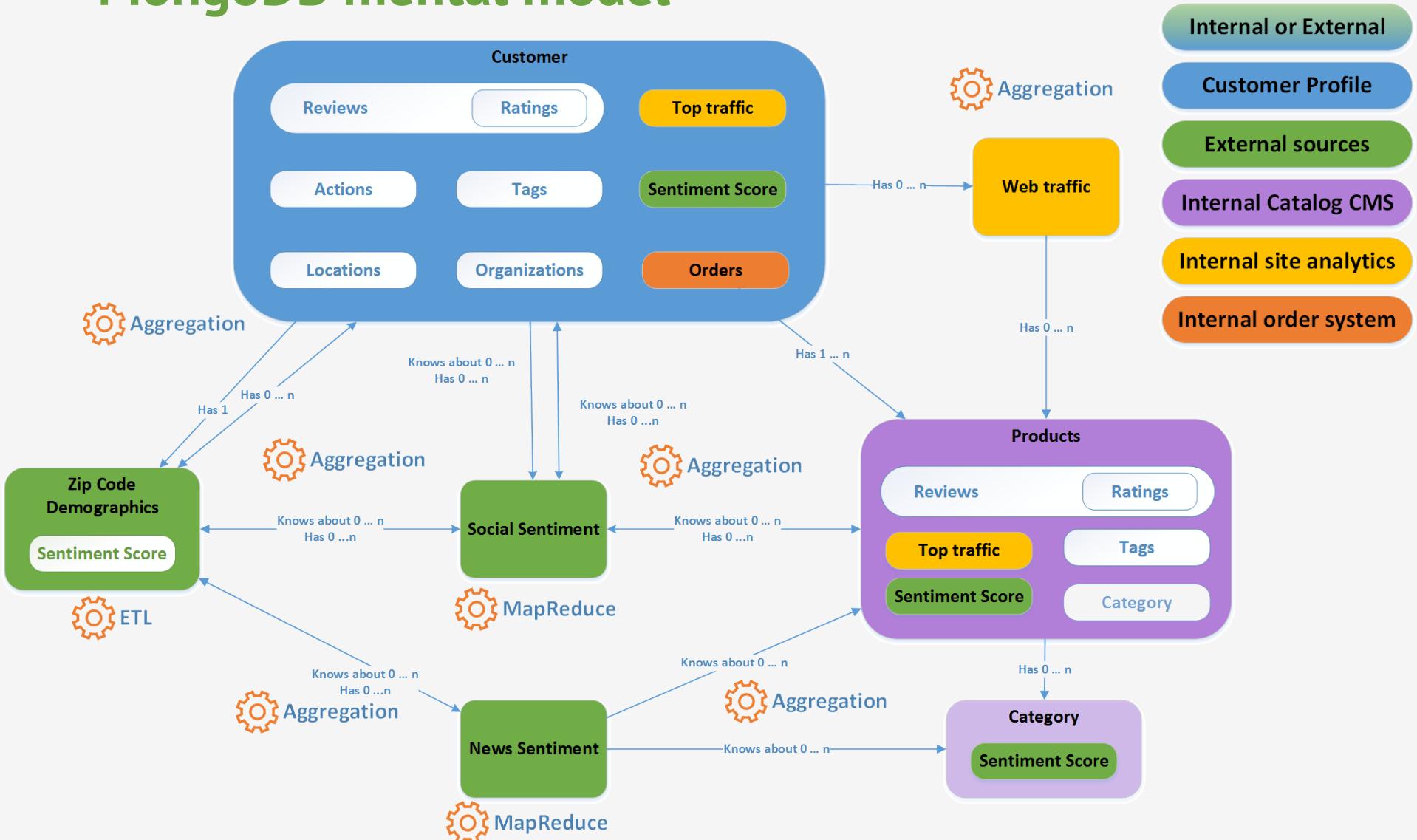
Zooming in to Single View backend



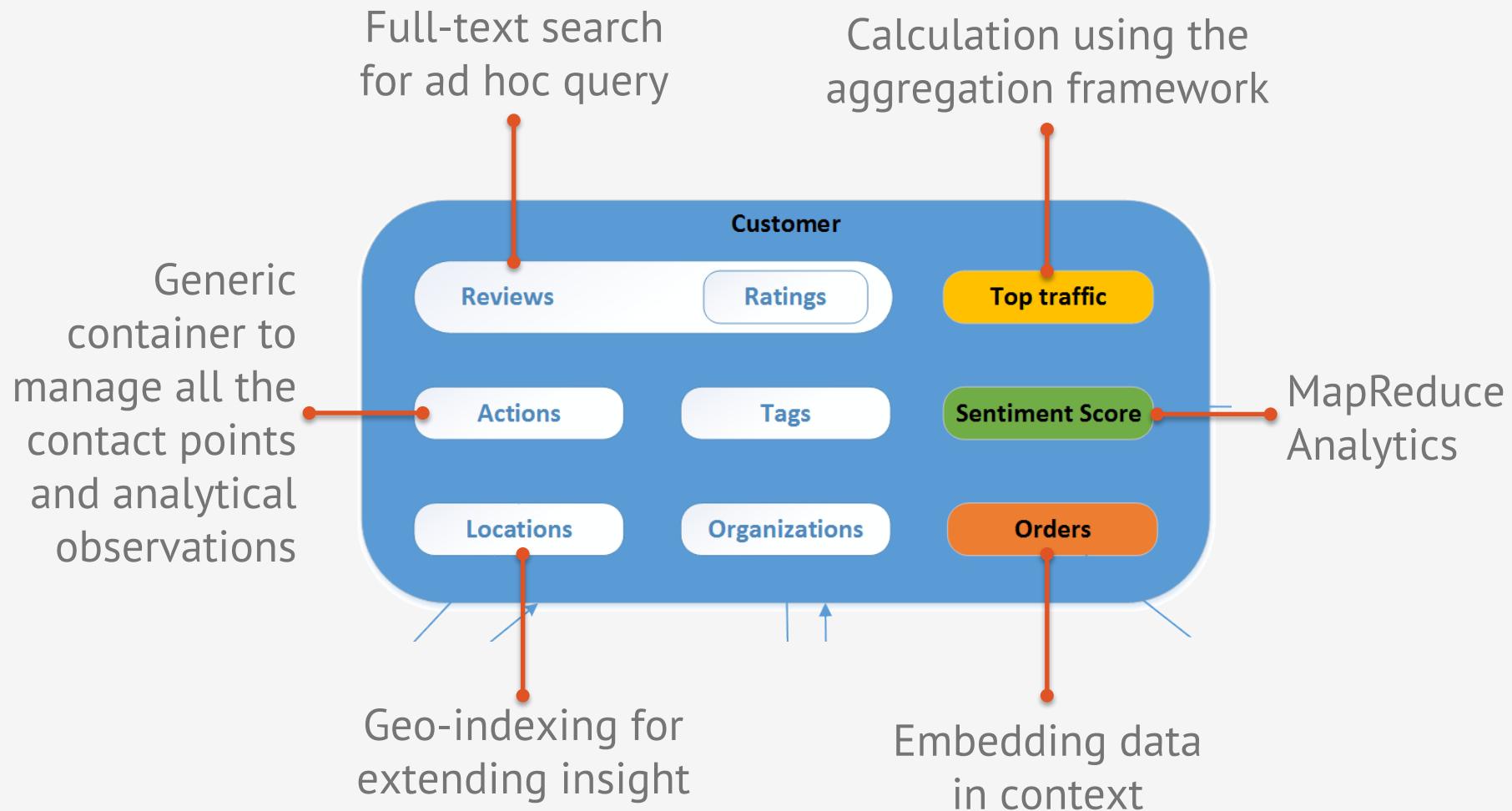
Traditional mental model



MongoDB mental model



Getting to a Single View of the customer

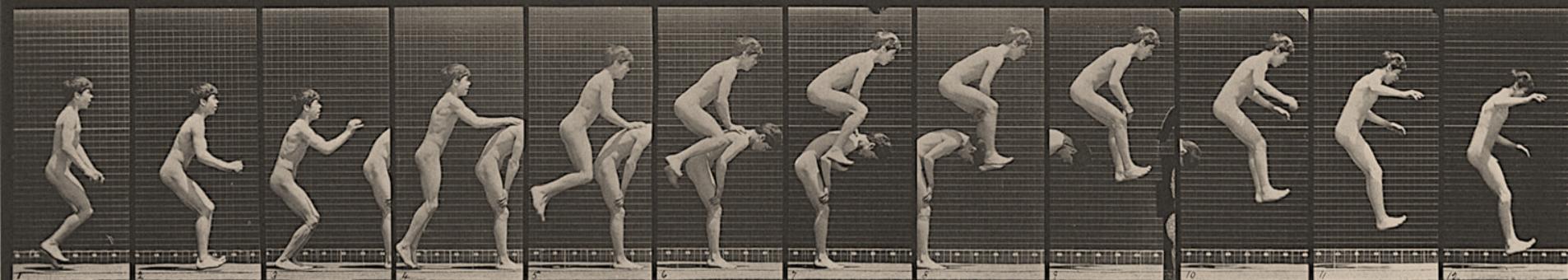


Questions we can now ask ...

- **What kinds of products has this person purchased?**
e.g. `distinct("orders.category", { "id" : 12345678 })`
- **How close are they from a point of service? Now.**
e.g. `find("location" : { $near : [40.8768,-73.787] })`
- **Who is most dissatisfied with our service?**
e.g. `find().sort({ sentiment : 1 }).limit(100)`
- **What should my customer representative mention in the next conversation?**
e.g. `find({ "action.topic" : "talkingpoint" }).sort(createdOn : -1)`

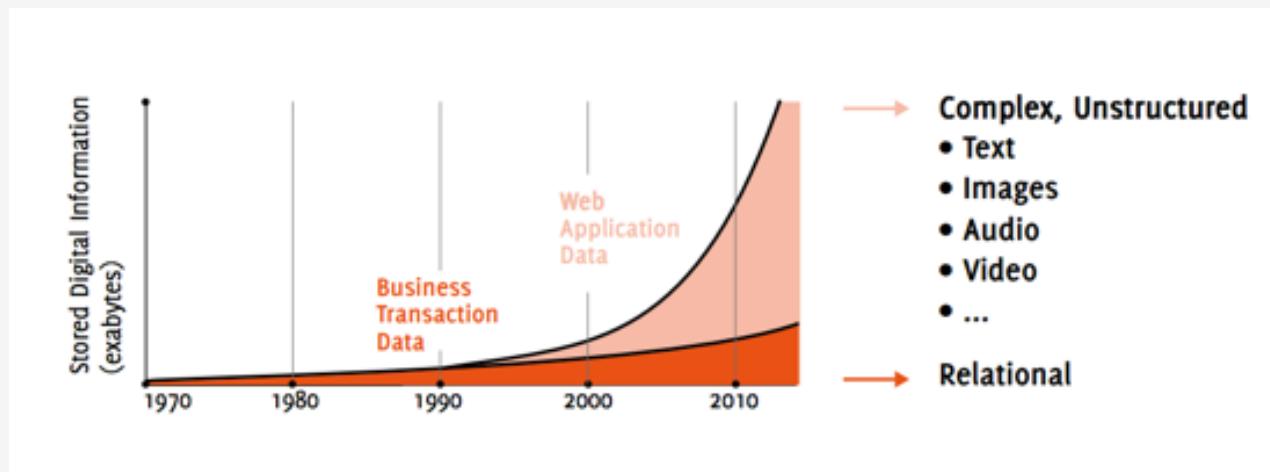
Start focused and iterate quickly

1. Focus on a single effort, a couple of data sources, and a familiar ‘actor’ to avoid being overwhelmed
2. Consider what the data is and a few questions you might ask of it – this informs the initial data model
3. Prototype, prove then iterate



Dynamic schemas enables flexibility and agility

- Data is coming in many forms
- New data sources surface constantly
- Single-views facilitate analytics and breed more data
- Flexible schemas make it easy to evolve vs. reinvent
- Schemas reflect the access patterns of the data



Modeling a single view of the customer begins with the types of queries we think we will want to make

- What is the **customer (individual or company)** doing that is not normal?
- What are the **customer's peers** doing, that they would do, but haven't?
- What **products has the customer purchased** and what is likelihood that they will buy something new in the next week | thirty days?
- What **will the customer** do this week?
- What are **customers telling us** about them, **us**, the **world**?
- What **content / products** are effective at generating sales?
- Who is influential – both **us and customers**?
- What does the world look like from **our proprietary point of view** vs. the what is happening in the **public**?
- Which **customers** should I focus on, what should I show them and why is it smart?

What do all these crazy formats look like?

```
    <!--Indicates version of XML encoding-->
<?xml version="1.0" encoding="UTF-8"?>
<findProductsRequest xmlns="http://www.ebay.com/Marketplace/MarketplaceCatalog/v1/services">
    <!--Call specific input Fields -->
    <!--Category ID -->
    <categoryString>string</categoryString>
    <!--Dataset -->
    <datasetName>string</datasetName>
    <!--More dataset values allowed here ... -->
    <!--Dataset Property Name -->
    <datasetPropertyName>string</datasetPropertyName>
    <!--More dataset values allowed here ... -->
    <!--Invocation ID -->
    <invocationId>string</invocationId>
    <!--Key words -->
    <keyWords>string</keyWords>
    <!--Pagination Input -->
    <paginationInput>
        <entriesPerPage>int</entriesPerPage>
        <pageNumber>int</pageNumber>
    </paginationInput>
    <!--Product Status Filter -->
    <productStatusFilter>
        <!--Exclude For BayReviews -->
        <excludeForBayReviews>boolean</excludeForBayReviews>
        <!--Exclude For BaySelling -->
        <excludeForBaySelling>boolean</excludeForBaySelling>
        <!--Exclude For HalfSelling -->
        <excludeForHalfSelling>boolean</excludeForHalfSelling>
    </productStatusFilter>
    <!--Property Filter -->
    <propertyFilter>
        <!--Property Name -->
        <propertyName>string</propertyName>
        <!--Value -->
        <value>string</value>
        <!--Comparison Operator -->
        <comparator>string</comparator>
        <!--Unit of Measurement -->
        <unitOfMeasurement>token</unitOfMeasurement>
        <!--Value -->
        <value>string</value>
        <!--Name -->
        <name>string</name>
        <!--Type -->
        <type>string</type>
        <!--URL Value -->
        <urlValue>anyURI</urlValue>
        <!--URL -->
        <url>string</url>
    </propertyFilter>
    <!--More value nodes allowed here ... -->
    <!--Sort Order -->
    <sortOrder>
        <!--Order by OrderType -->
        <order>OrderType</order>
        <!--Property Name -->
        <propertyName>string</propertyName>
    </sortOrder>
    <!--Product Search -->
    <productSearch>
        <!--More productSearch nodes allowed here ... -->
    </productSearch>
</findProductsRequest>
```

Twitter timeline

Comments

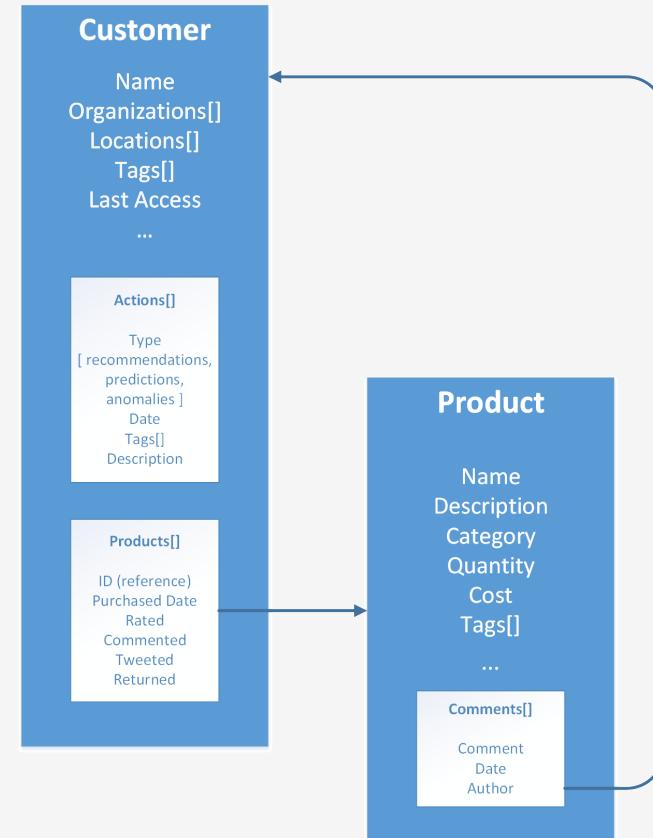
0000	FF	D8	FF	E1	1D	FE	45	78	69	66	00	00	49	49	2A	00		
0010	08	00	00	00	09	00	0F	01	02	00	00	00	00	00	7A	00		
0020	00	00	10	01	02	00	14	00	00	00	80	00	00	00	12	01		
0030	03	00	01	00	00	00	01	00	00	00	1A	01	05	00	01	00		
0040	00	00	A0	00	00	00	1B	01	05	00	01	00	00	00	A8	00		
0050	00	00	28	01	03	00	01	00	00	00	02	00	00	00	32	01		
0060	02	00	14	00	00	00	B0	00	00	00	13	02	03	00	01	00		
0070	00	00	01	00	00	00	69	87	04	00	01	00	00	00	C4	00		
0080	00	00	3A	06	00	00	43	61	6E	6F	6E	00	43	61	6E	6F		
0090	DE	20	50	6F	77	65	72	53	68	6F	74	20	41	36	30	00		
00A0	00	00	00	00	00	00	00	00	00	00	00	00	B4	00	00	00		
00B0	01	00	00	00	B4	00	00	00	01	00	00	00	32	30	30	34		
00C0	3A	30	36	3A									3	30	3A	32	35	00
00D0	1F	00	9A	82									6	03	00	00	9D	82
00E0	05	00	01	00									0	90	07	00	04	00

Binary

```json
{
 "facets": {
 "per\_facet": [
 {
 "count": 13,
 "term": "OBAMA, BARACK"
 },
 {
 "count": 15,
 "term": "MCCHRISTAL, STANLEY A"
 },
 {
 "count": 14,
 "term": "DAVIS, ROBERT N"
 },
 {
 "count": 14,
 "term": "PARSAI, HAMID"
 },
 {
 "count": 14,
 "term": "MULLEN, MICHAEL G"
 },
 ...
 {
 "count": 11,
 "term": "VILLAJOS, PAT"
 }
 ],
 "application\_year": [
 {
 "count": 23,
 "term": "2009"
 },
 {
 "count": 19,
 "term": "2008"
 },
 {
 "count": 11,
 "term": "2007"
 },
 {
 "count": 11,
 "term": "2006"
 }
 ],
 "offset": "0",
 "total": 127
 },
 "body": "The most important line in President Obama's Afghan speech was about Afpak policy (no named by the White House) but about the U.S. domestic situation in Afghanistan that could not be openly stated - because the nation that it is most interested in building is our own." An military strategy for winning a war."
},
{
 "defect": {
 "per\_facet": [
 {
 "term": "UNITED STATES DEFENSE AND MILITARY FORCES",
 "value": "UNITED STATES ECONOMY",
 "label": "AFGHANISTAN WAR (2001- )"
 }
 ],
 "title": "OP-ED COLUMNIST: Afghanistan on Main Street",
 "url": "http://www.nytimes.com/2009/12/05/opinion/05hicks-edcolumnist.html"
 }
},
{
 "body": "After the dramatic three-month buildup, you'd think that Barack Obama's speech announcing his policy for Afghanistan would be the most significant news story of the year. History may take a different view. When we look back on this point in America's longest war, we may discover that a relatively trivial White House speech."
},
{
 "defect": {
 "per\_facet": [
 {
 "term": "UNITED STATES DEFENSE AND MILITARY FORCES",
 "value": "UNITED STATES ECONOMY",
 "label": "AFGHANISTAN WAR (2001- )",
 "label2": "SPEECHES AND STATEMENTS"
 }
 ],
 "title": "OP-ED COLUMNIST: Obama's Logic Is No Match for Afghanistan",
 "url": "http://www.nytimes.com/2009/12/05/opinion/05hicks-edcolumnist.html"
 }
},
{
 "body": "WASHINGTON - Defense Secretary Robert M. Gates , Secretary of State Hillary Rodham Clinton and the nation's top military officer on Wednesday laid out a muscular defense of America's war in Afghanistan, insisting on additional troops to Afghanistan, but they made clear that plan to begin withdrawal will start as early as July 1.
   
 "data": "20091202",
 "defect": {
 "per\_facet": [
 {
 "term": "UNITED STATES DEFENSE AND MILITARY FORCES",
 "value": "AFGHANISTAN WAR (2001- )",
 "label": "SPEECHES AND STATEMENTS"
 }
 ],
 "title": "Obama Tries to Defend Policy on Afghanistan",
 "url": "http://www.nytimes.com/2009/12/03/world/asia/03policy.html"
 }
},
{
 "body": "Americans have reason to be pessimistic, if not despairing, about the war in Afghanistan. After eight years of fighting, more than 800 American soldiers have died there, and the conflict has become a drain on the U.S. economy. The war is barely legitimate and barely hanging on in the face of an increasingly determined insurgency. In his speech",
 "data": "20091202",
 "defect": {
 "per\_facet": [
 {
 "term": "UNITED STATES DEFENSE AND MILITARY FORCES",
 "value": "AFGHANISTAN WAR (2001- )",
 "label": "SPEECHES AND STATEMENTS"
 }
 ],
 "title": "EDITORIAL: The Afghanistan Speech",
 "url": "http://www.nytimes.com/2009/12/02/opinion/02edit.html"
 }
},
{
 "body": "The top military commander in Afghanistan war is in a confidential meeting with the president and other top officials at Camp David within the next year or else the conflict "will likely result in failure." The grim prediction comes from a 45-page report by top military commander, Gen. Stanley A. McChrystal, submitted to Defense Secretary Robert M. Gates.
   
 "data": "20090924",
 "defect": {
 "per\_facet": [
 {
 "term": "UNITED STATES DEFENSE AND MILITARY FORCES",
 "value": "AFGHANISTAN WAR (2001- )",
 "label": "SPEECHES AND STATEMENTS"
 }
 ],
 "title": "Gen. McChrystal Calls for More Troops To Avoid Afghanistan Failure",
 "url": "http://www.nytimes.com/2009/09/21/world/asia/21afghan.html"
 }
},
{
 "body": "The top military commander in Afghanistan war is in a confidential meeting with the president and other top officials at Camp David within the next year or else the conflict "will likely result in failure." The grim prediction comes from a 45-page report by top military commander, Gen. Stanley A. McChrystal, submitted to Defense Secretary Robert M. Gates.
   
 "data": "20090924",
 "defect": {
 "per\_facet": [
 {
 "term": "UNITED STATES DEFENSE AND MILITARY FORCES",
 "value": "AFGHANISTAN WAR (2001- )",
 "label": "SPEECHES AND STATEMENTS"
 }
 ],
 "title": "Gen. McChrystal Calls for More Troops To Avoid Afghanistan Failure",
 "url": "http://www.nytimes.com/2009/09/21/world/asia/21afghan.html"
 }
},
{
 "body": "The top military commander in Afghanistan war is in a confidential meeting with the president and other top officials at Camp David within the next year or else the conflict "will likely result in failure." The grim prediction comes from a 45-page report by top military commander, Gen. Stanley A. McChrystal, submitted to Defense Secretary Robert M. Gates.
   
 "data": "20090924",
 "defect": {
 "per\_facet": [
 {
 "term": "UNITED STATES DEFENSE AND MILITARY FORCES",
 "value": "AFGHANISTAN WAR (2001- )",
 "label": "SPEECHES AND STATEMENTS"
 }
 ],
 "title": "Gen. McChrystal Calls for More Troops To Avoid Afghanistan Failure",
 "url": "http://www.nytimes.com/2009/09/21/world/asia/21afghan.html"
 }
}
]
}]]

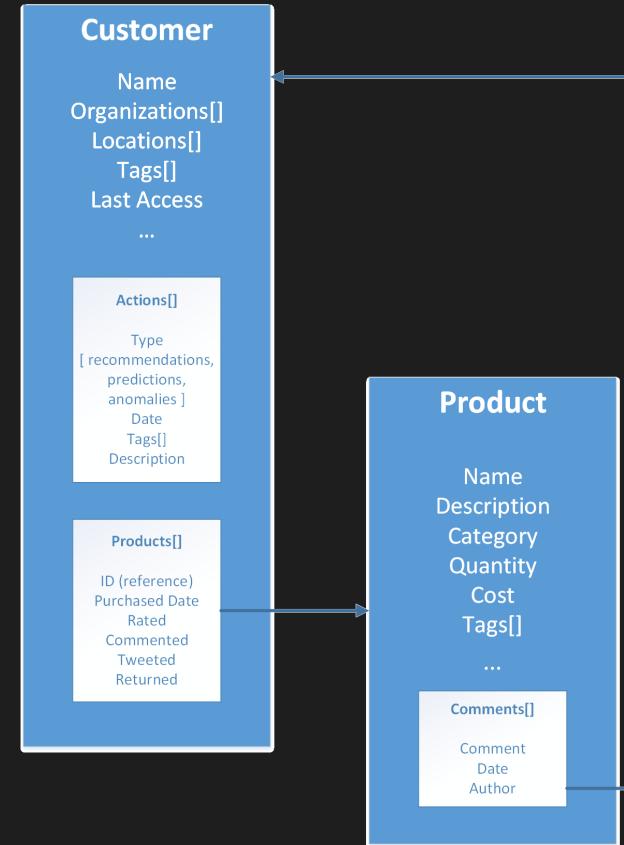
# Centering around customers enables single access pattern for fast rich queries and iterative extension

- Using arrays to allow for multiple relationships (organizations & locations)
- Tags as a means to flexibly profile e.g. top\_10\_percent, badge\_of\_honor, flight\_risk
- Actions array to store analytical outputs that describe the customer to sales channels and people e.g. anomalies, recommendations, predictions
- Products array to store products they customer has purchased and if they publicly engaged e.g. rating, commenting, tweeting returned etc.
- Products collection to store details
- Product comments reference Customers



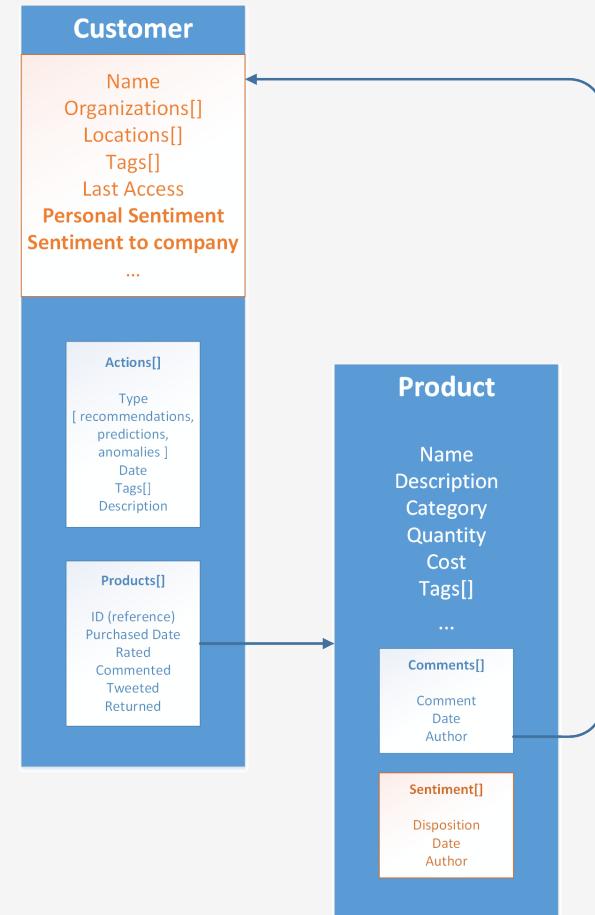
# A single-view of the customer enables fast, rich queries and iterative extension

```
{
 "_id" : ObjectId("5063114bd386d8fadbd6b004"),
 "name" : "Brian D. Goodman",
 "organizations" : ["MongoDB"],
 "locations" : [
 {
 "type" : "work",
 "address" : "229 W 34 St., 5th floor",
 "city" : "New York",
 "state" : "NY",
 "zipcode" : "10036"
 }
],
 "tags" : ["blogger", "photographer", "technology"],
 "lastAccess" : ISODate("2012-12-19T06:15:33.035Z"),
 "actions" : [
 {
 "type" : "recommendation",
 "date" : "...",
 "tags" : [
 "photography",
 "video"
],
 "description" : "Zack Arias released new video!",
 "action" : "http://..."
 }
],
 "products" : [
 {
 "_id" : ObjectId("5063114bd333d8fadbd6b444"),
 "purchased" : ISODate("2012-12-19T06:15:33.035Z"),
 "rated" : 1,
 "commented" : 1,
 "tweeted" : 0,
 "returned" : 0
 }
]
}
```



# Adding streaming customer sentiment to base model enables new analytics and customer insight

- Sentiment analytics could be performed
  - Across all customer expressions
  - Towards the company
  - Towards specific products
- An aggregate sentiment score might place the customer on a positive / negative scale as a means to identify predisposition and can be contrasted with sentiment towards the company
- Sentiment towards products might be stored in an array, similar to comments, enabling context specific dispositions



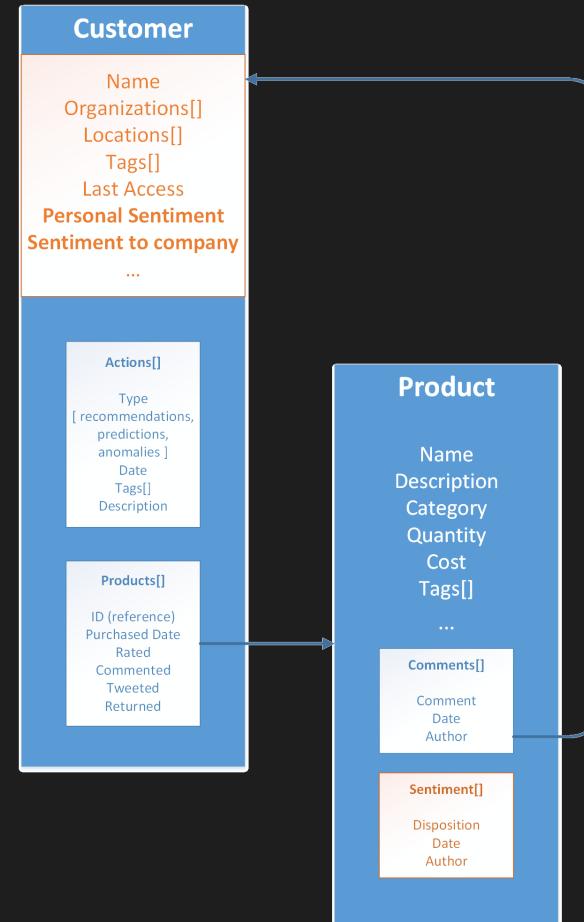
# Adding streaming customer sentiment to base model enables new analytics and customer insight

```
{
 "_id" : ObjectId("5063114bd386d8fadbd6b004") ,
 "name" : "Brian D. Goodman",
 "organizations" : ["MongoDB"] ,
 "locations" : [
 {
 "type" : "work",
 "address" : "229 W 34 St., 5th floor",
 "city" : "New York",
 "state" : "NY",
 "zipcode" : "10036"
 }
],

 "sentiment" : {
 "personal" : 1,
 "towardsCompany" : 0
 },

 "tags" : ["blogger", "photographer", "technology"],
 "lastAccess" : ISODate("2012-12-19T06:15:33.035Z"),
 "actions" : [
 . . .
],

 "products" : [
 {
 "_id" : ObjectId("5063114bd333d8fadbd6b444"),
 "purchased" : ISODate("2012-12-19T06:15:33.035Z"),
 "rated" : 1,
 "commented" : 1,
 "retweeted" : 0,
 "returned" : 0
 }
]
}
```



# So, we have a model, but what if I need more than one way to look at my data?

- **Aggregate instead of recalculate**

Examples: *Creating an activity score based on a few variables on the fly or... Storing a variety of counters, but not the individual records*

- **Store the original granularity for a detailed query and embed aggregations**

*Example: Total number of reviews vs. counting the reviews themselves*

- **Embedding and linking gives you the best of performance and management**

*Example: a master collection for addresses, but an embedded copy for performance*

# MongoDB shifts responsibilities to developers

- Enforcing security, auditing, redaction, schema
- Assuming change
- Managing the data holistically
- Balancing the decision to embed vs. link and  
the extra management of data that may be required

# What does your single-view look like?

- What is at the center of your single-view?
- What questions are you asking of your data?
- How do you plan to get the data in there?
- How do you make sure it doesn't exit out the backdoor?



#MongoDBWorld

# Thank You

Brian D. Goodman

*Enterprise Architect, MongoDB*

