



WOWODC '011

MONTREAL 1/3 JULY 2011



WebObjects Under Stress

Dan Beatty, Ph.D. Candidate
Texas Tech University



WOWODC

MONTREAL 1/3 JULY 2011



Q&A

Even with legacy metadata catalogs containing 1000 attribute tables, ER ModernLook can be made to service these tables, that leads to mapping solutions to better structures. In some cases, D2W's Display Batch Group can make the requests efficient. In other cases, the faulting of the display batch group and plain blank queries cause a cache miss for accessing the data. This study is how to determine which tables are what, and things one can do to recover data in these conditions.

Topics

- Background
 - Cloud Computing
 - History of the Sloan Digital Sky Survey (SDSS) Data Release 1 (DR1)
Mirror at Texas Tech University
 - Motivation
- My Approach
 - Server Architectures
 - Transition to through D2W and RESTful Architecture
- Highlight of the contributions

Best Definition of Cloud Computing

“Clouds are large pools of easily usable and accessible virtualized resources (such as hardware, development platforms, and/ or services). These resources can be dynamically reconfigured to adjust to a variable load (scale), allowing also for an optimum resource utilization. This pool of resources is typically exploited by a pay-per-use model that are offered by the Infrastructure Provider by means of customized Service Level Agreements.”

Forms of Cloud Computing?

- Software as a Service (SaaS):
 - “Cloud computing refers to both, the applications delivered as services over the Internet, and the hardware and systems software in the data-centers that provide those services.” (The Berkeley View of Cloud Computing [5])
- Infrastructure as a Service (IaaS)
 - Epitomized by the Traveling Executive with a Swiss Army Knife and Credit Card [37]
- Platform as a Service (PaaS):
 - Cloudlets are typically virtual machines sometimes presented as the OS itself. [69]

[5] Michael Armbrust, et. al. Above the clouds: A Berkeley view of cloud computing. Technical report, UC Berkeley Reliable Adaptive Distributed Systems Laboratory, February 2009.

[37] Brian Hayes. Cloud computing. Communications. ACM, 51(7):9–11, 2008.

[69] Mahadev Satyanarayanan, Paramvir Bahl, Ramon Caceres, and Nigel Davies. The case for vm-based cloudlets in mobile computing. IEEE Pervasive Computing, 8:14–23, 2009.

Background of the Sloan Digital Sky Survey

The primary objective of the Sloan Digital Sky Survey (SDSS) catalog is to provide data to “support detailed investigations of the distributions of luminous and non-luminous matter in the Universe.” (D.G. York “The Sloan Digital Sky Survey” Technical Summary, 2000).

Background System Diagram (Old)

REPLICATION OF THE DATA

SLOAN DIGITAL SKY SURVEY

by

SHARATH HEGDE

A THESIS

IN

COMPUTER SCIENCE

Submitted to the Graduate School
of Texas Tech University
Partial Fulfillment
of the Requirements
for the Degree

MASTER OF SCIENCE



ls



ve

The tragedy of many catalogs

- There are what could be called legacy catalogs where the original interface has been lost or destroyed for one reason or another.
- In some other cases, the means to retrieve the data may still exist, but those means represent a problem legally or logistically.
- Regardless of how well that catalog was assembled, the data is considered valuable enough to be retrieved and compared with data today.

Original Motivation of Revised Image Query Service

- The original design contained flaws that interfered with usage of the system. Consequently, neither owner or clients were truly satisfied.
- Furthermore, Sharath's thesis offered three recommendations for improvement
 - Restructure of the data
 - Implementation using other choices of database (such as Frontbase, Oracle, etc)
 - Implementation using WebObjects

The Overall Challenge

- Build a new catalog that can fulfill the overall SDSS objective:
- Develop a structure that
 - is well organized.
 - Fits the astronomical community needs
- Ensure catalog can address remote user's needs
 - Contributors at the telescope
 - Analyst at the labs

Solution Approach

- Content Management (WOWODC 2010, WWDC 2007)
- Mashup (client based Web Browser) Augmentation (WWDC 2008, WOWODC 2010)
- Content - Rule Based Factories (Today!)
- RESTful Web Services (In progress)

Data Release I Tables

- Target Selection Segment
- Target Selection Fields
- Resource Description Framework
- Z Emission and Cross Correlation
- Line and Measurement Ratios
- Spectral Inventory
- Spectral Object Family of Tables (4 of them)
- Plate Object Table
- Quasar Sample Object

Data Release 1 Tables

The screenshot displays the Eclipse IDE interface with the Entity Modeler tool. The Outline view on the left shows a tree structure for the DR1Scheme, listing various entities. The Properties view at the bottom left shows configuration options for the selected entity, including Attribute Prefix, Attribute Suffix, Attribute Case, Attribute Separator, Entity Prefix, Entity Suffix, Entity Case, Entity Separator, and a checked 'Reverse Engineered' checkbox. The main Entity Modeler window displays a table mapping entity names to table names and class names.

Name	Table Name	Class Name
BestQSO	bestQSO	com.webobjects.eocontrol.EOGenericRecord
BestRDF	bestRdf	com.webobjects.eocontrol.EOGenericRecord
BestTSField	bestTsField	edu.ttu.cs.dcglab.dr1.model.BestTSField
BestTSSeg	bestTsSeg	edu.ttu.cs.dcglab.dr1.model.BestTSSeg
FoundLines	foundLines	edu.ttu.cs.dcglab.dr1.model.FoundLines
LineRatios	lineRatios	edu.ttu.cs.dcglab.dr1.model.LineRatios
MeasuredLines	measuredLines	edu.ttu.cs.dcglab.dr1.model.MeasuredLines
Plateobj	plateObj	edu.ttu.cs.dcglab.dr1.model.Plateobj
Specbesttsobj	specBestTsObj	edu.ttu.cs.dcglab.dr1.model.Specbesttsobj
Specbesttsobjdistinct	specBestTsObjDistinct	edu.ttu.cs.dcglab.dr1.model.specBestTsObjDistinct
Specobj	specObj	edu.ttu.cs.dcglab.dr1.model.Specobj
Spectargettsobj	specTargetTsObj	edu.ttu.cs.dcglab.dr1.model.Spectargettsobjdistinct
Spectargettsobjdistinct	specTargetTsObjDistinct	edu.ttu.cs.dcglab.dr1.model.Spectargettsobjdistinct
Spinventory	spInventory	com.webobjects.eocontrol.EOGenericRecord
Targetclose	targetClose	com.webobjects.eocontrol.EOGenericRecord
Targettsseg	targetTsSeg	com.webobjects.eocontrol.EOGenericRecord
Zcrosscorrelation	zCrossCorrelation	com.webobjects.eocontrol.EOGenericRecord
Zemission	zEmission	com.webobjects.eocontrol.EOGenericRecord

Data Release I Tables

BestTSField
aaerrG
aaerrI
aaerrR
aaerrU
aaerrZ
aaG
aaI
aaR
aaU
aaZ
acceptablemask
aG
aI
airmassG
airmassI
airmassR
airmassU
airmassZ
aR
aU
aZ
badmask
bberrG
bberrI
bberrR
bberrU
bberrZ
bbG

FoundLines
chisq
continuum
ew
ewerr
ewmin
height
heighterr
linemask
nsigma
nu
restwave
sigma
sigmaerr
sigmamax
sigmamin
specindex
specobjid
wave
waveerr
wavemax
wavemin
weight
z
zerr

LineRatios
ew
ewerr
mag
magerr
name
sn
specobjid
wavemax
wavemin
z

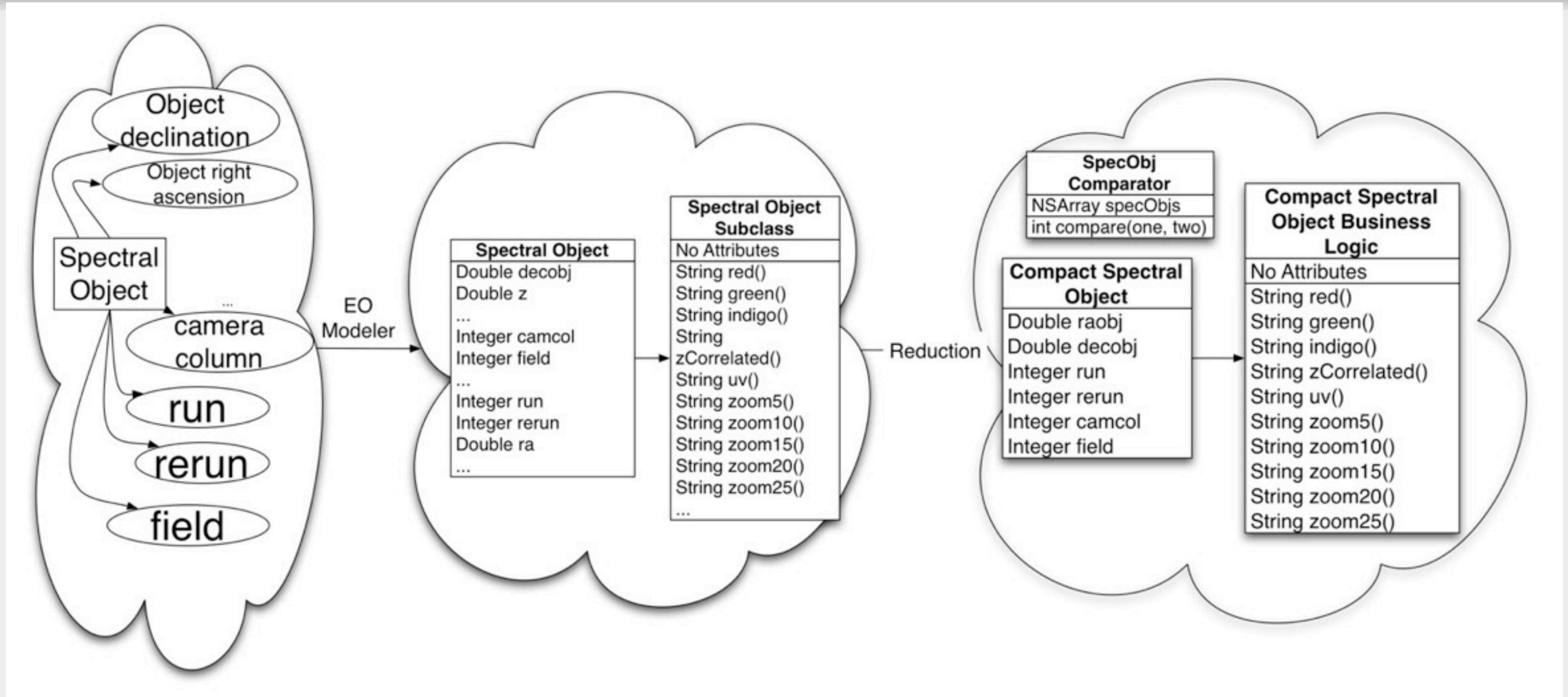
BestTSSeg
camcol
chunkendmu
chunkid
chunkstartmu
chunkstripe
endmu
endmudeg
field0
inclination
nfields
node
rerun
run
startmu
startmudeg
status
strip
stripenumber

MeasuredLines
chisq
continuum
ew
ewerr
ewmin
height
heighterr
linemask
nsigma
nu
restwave
sigma
sigmaerr
sigmamax
sigmamin
specindex
specobjid
wave
waveerr
wavemax
wavemin
weight
z
zerr

Plateobj
airtemp
cartid
decdeg
dewdep
dewpoint
dusta
dustb
dustc
dustd
exptime
gustd
gusts
helioRv
humidity
humidout
inspectdate
mapid
mjd
name
nexp
observer
plateid
pressure
radeg
spec1G
spec1I
spec1R
spec2G
spec2I

Specbesttsobj
abDevrrG
abDevrrI
abDevrrR
abDevrrU
abDevrrZ
abDevG
abDevI
abDevR
abDevU
abDevZ
abExperrG
abExperrI
abExperrR
abExperrU
abExperrZ
abExpG
abExpI
abExpR
abExpU
abExpZ
b
camcol
catid
colcerrG
colcerrI
colcerrR
colcerrU
colcerrZ
colcG

Data Model of Revised Image Query Service (RIQS)

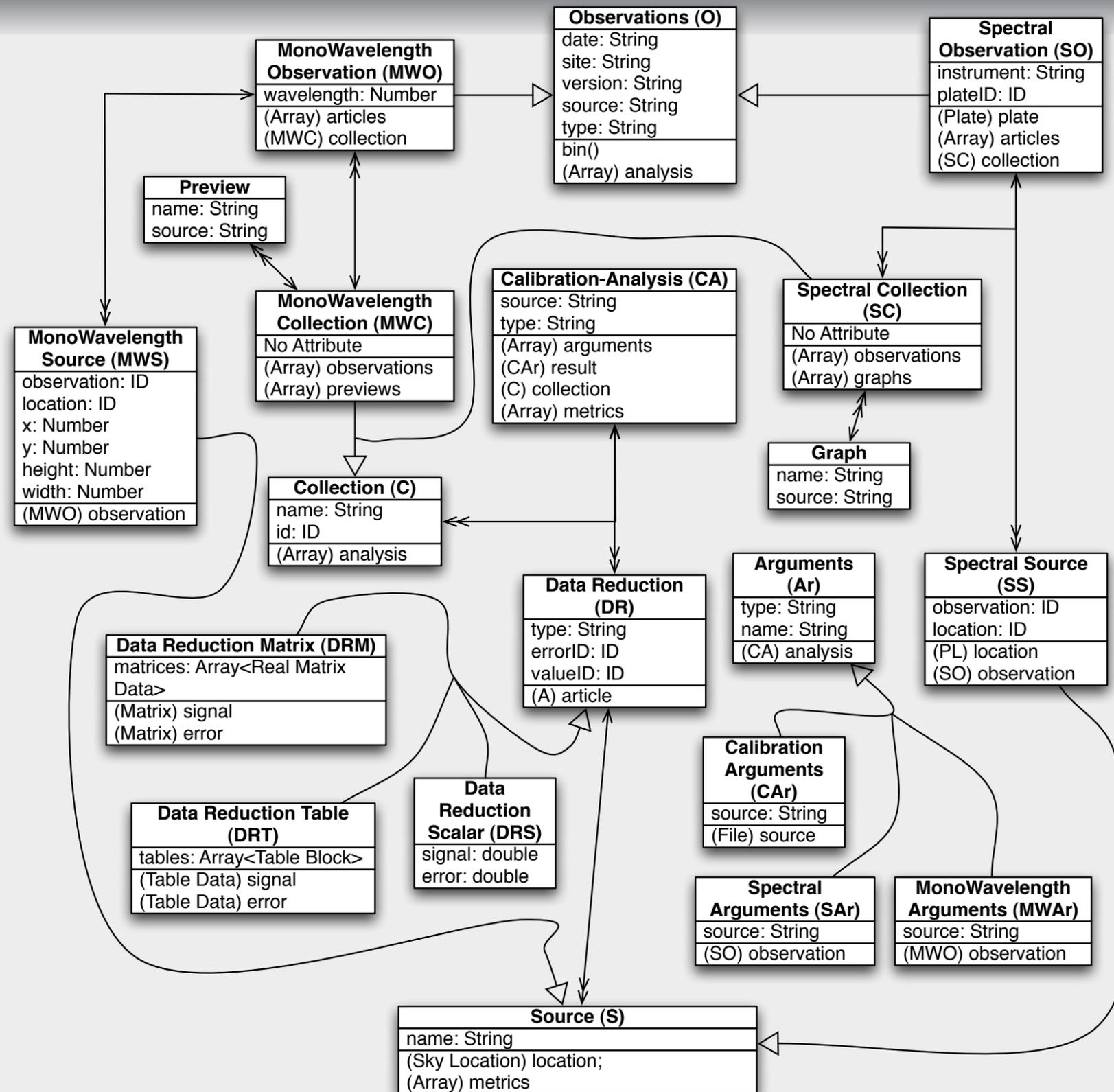


src: "Image Query Service Using Content Management Techniques" by Daniel Beatty and Noe Lopez Benitez (Texas Tech Univ.) 6th International Conference Information Technology Next Generation 2009

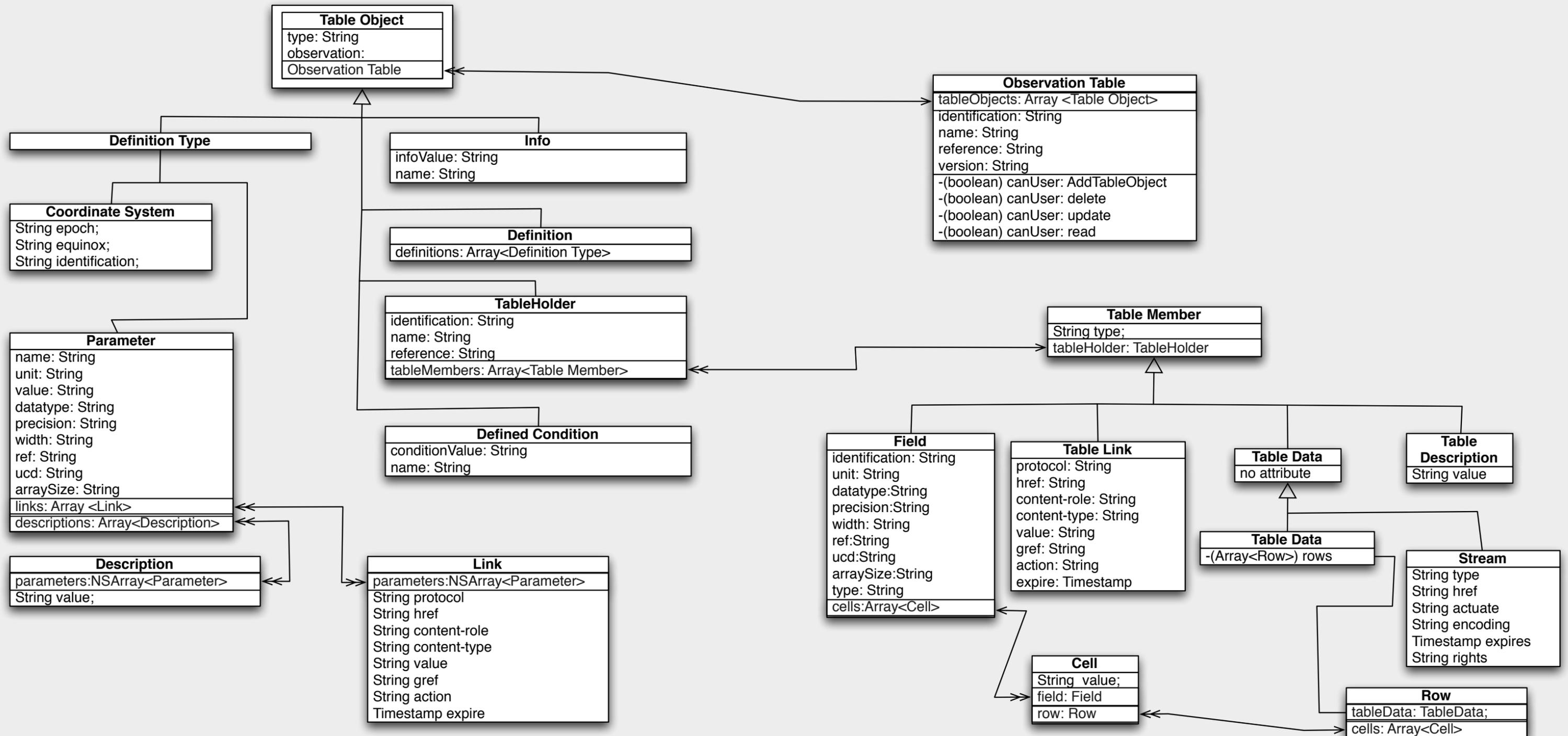
Another Example Practically from Outer Space

- Enhancements to the internet Flexible Image Transport System (iFITS)
 - Spectral Observations can be done
 - Virtual Observation Tables can be made ORM friendly
- D2W provides the means to refine this model and make it less volatile

iFITS Photos and Spectrographs

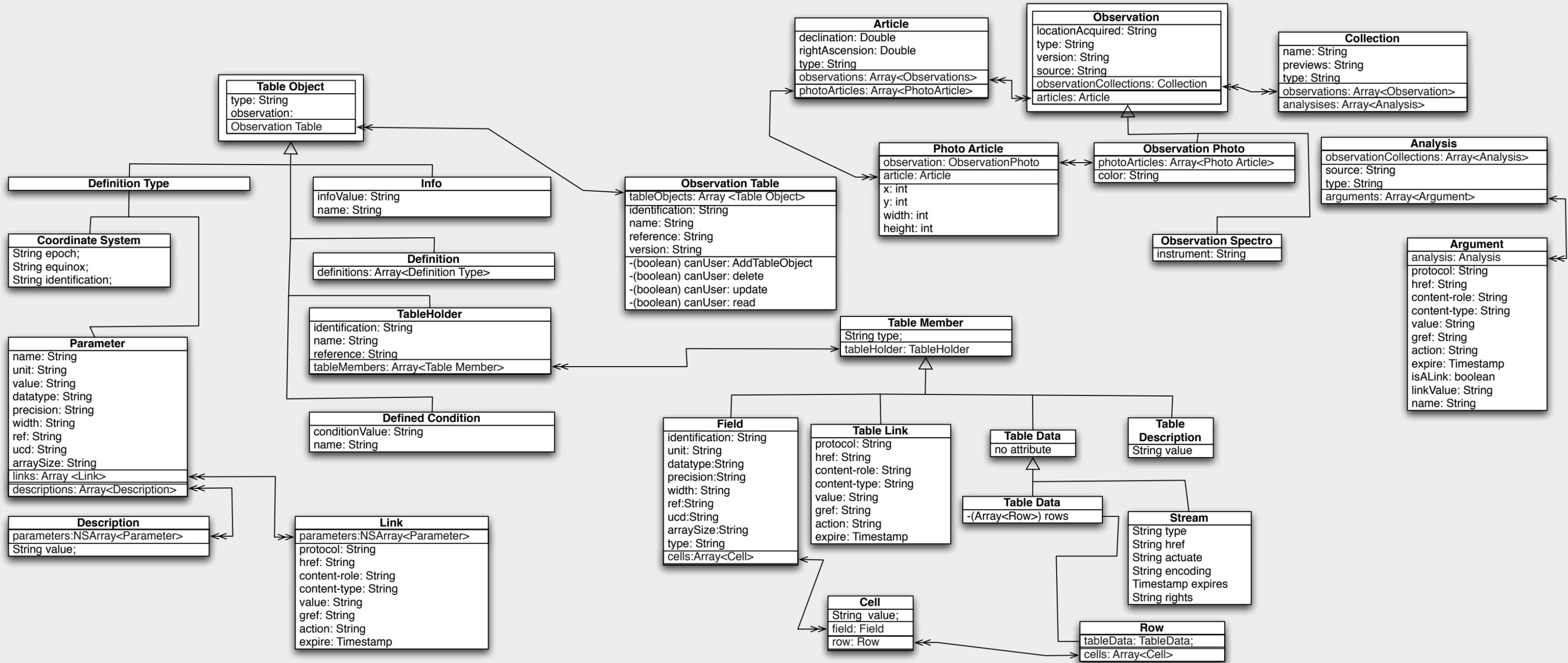


What about Virtual Observation Tables?



Equi-satisfiable model to the VO Table Standard.

(I) New Data Model

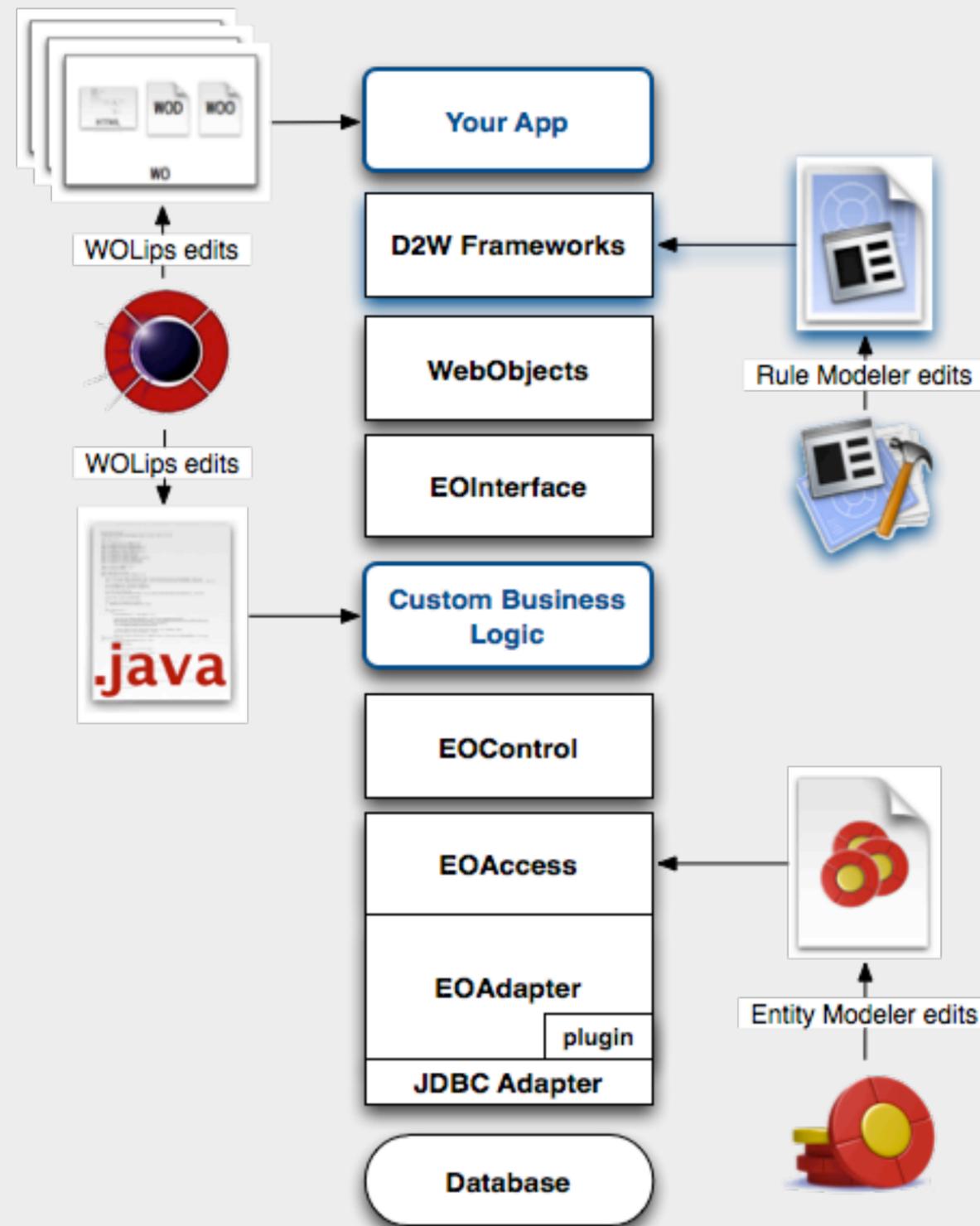


Rule-based Management of Large Unorganized Data Sets

- Rule-based systems have been studied for nearly two decades in applications such as
 - geographical information systems (GIS) and
 - metadata catalog systems.
- Recovering large data sets that are not well organized is a challenge that imposes constraints on applications.
- These constraints include
 - utilizing huge amounts in memory,
 - consuming excessive amounts of time, and
 - risk exceeding these resources causing instability.

Direct To Web Is A Rule-Based System

*This slide was taken from David's WOWODC 2009 session

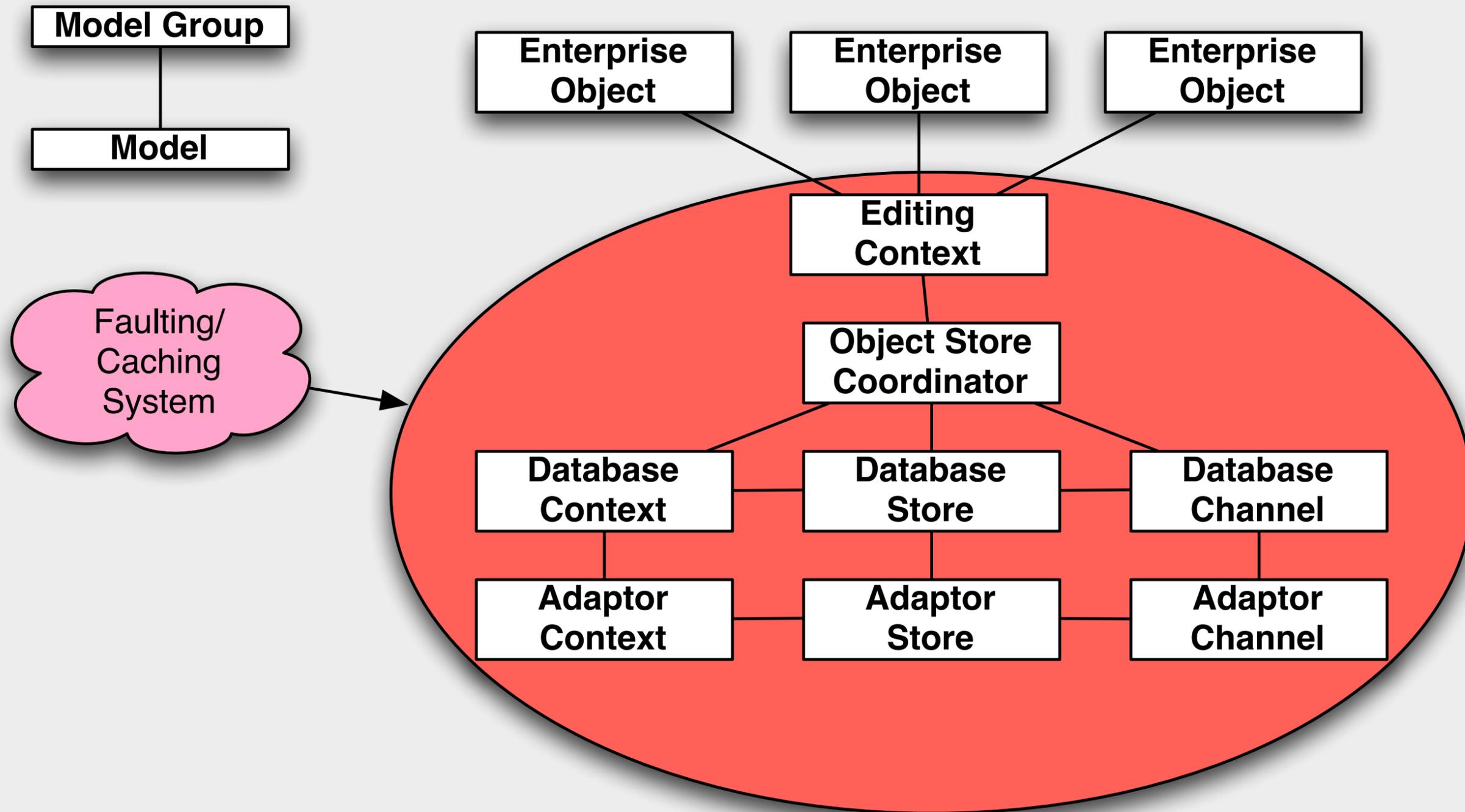


Request and Response

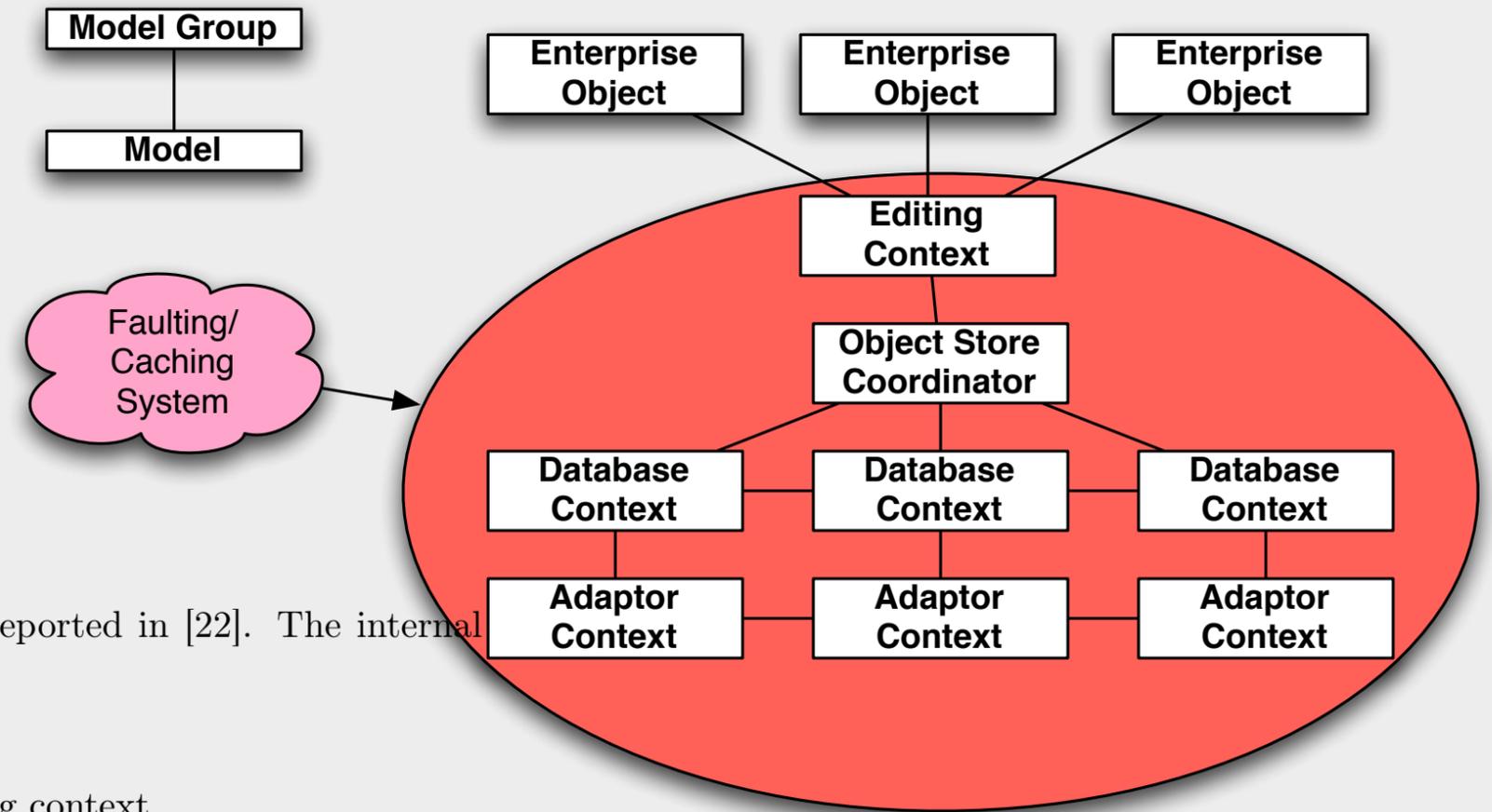
- The caching system of the Object Relational Machine (ORM) of Enterprise Objects can be overwhelmed by the size of Sloan Digital Digital Sky Survey (SDSS) Data Release I (DR I) metadata catalog.
- A consequence of this limitation is that the responses from must be throttled down,
 - One solution can be found by limiting the fetch size for each table in the metadata catalog.
 - Another solution is to supply a sufficient amount of memory.
- Direct to Web (D2W) has the capability of doing so in granular way and thereby providing DR I in its entirety.

Object Relational Map Engine

Concept presented by Mark Ritchie (WOWODC 2009)



Object Relational Map Engine



These layers work together to fulfill an internally consistent view constraint, as reported in [22]. The internal consistent view constraint is defined as a state where the following:

- each enterprise object, e_o , only has references to other e_o within the same editing context,
- any subset of managed objects \bar{e}_o has a consistent graph mapping regardless of a start point,
- If an e_o is not presently in an editing context, E , then it is brought into E when referenced,
- and changes to e_o are reflected throughout the engine's layers.

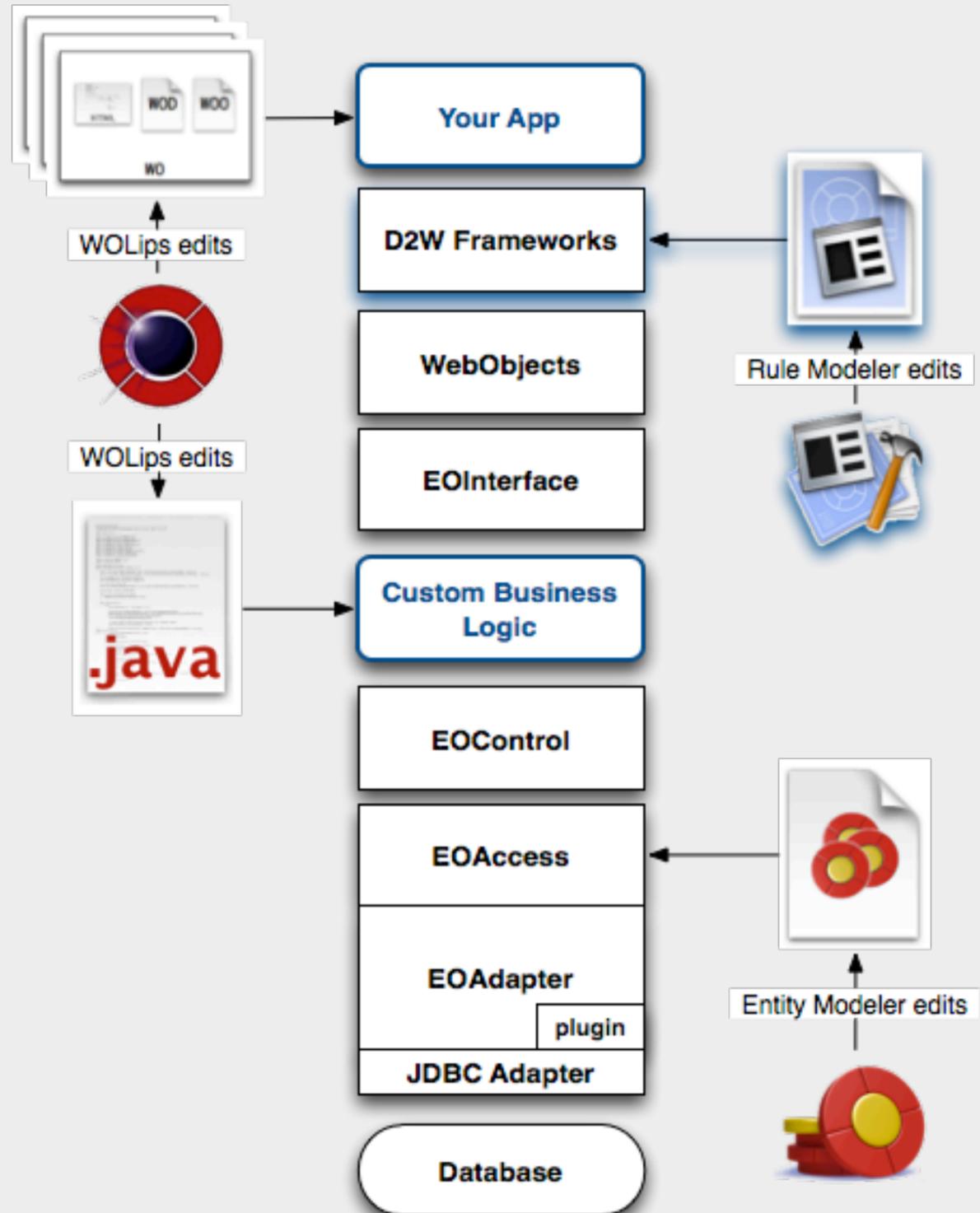
[22] Charles Hill and Steven Anglin. *Practical WebObjects*. Springer-Verlag New York, Inc (Apress) 175 Fifth Avenue New York, NY 10010, 2004. 2.3

Managing Fault States

- Managing fault states is a classic problem.
- However, there is not a known solution that applies to all possible distinct instances of this problem.

What is a rule?

*This slide was taken from David's WOWODC 2009 session

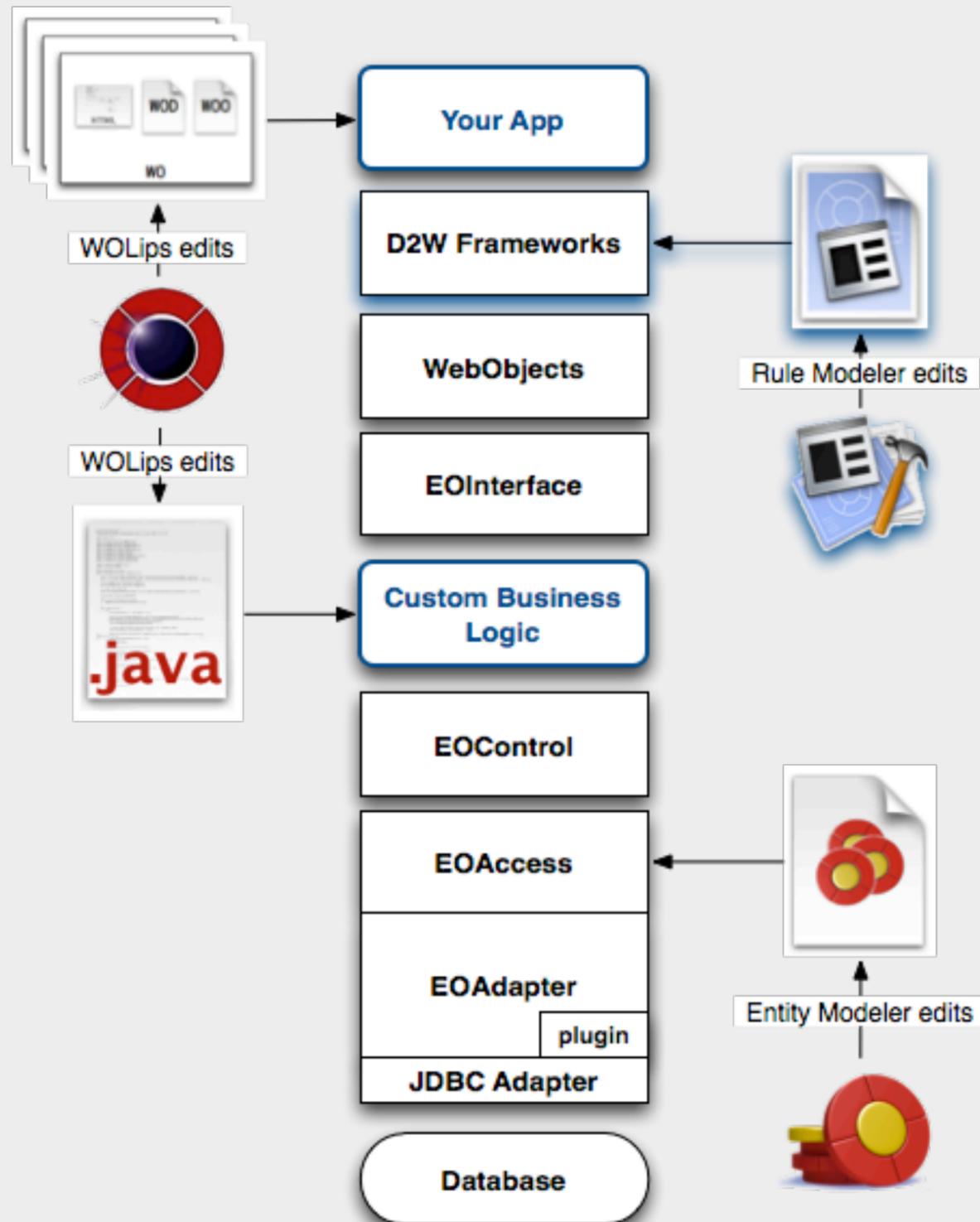


Listing 1: 'Syntax of a Rule'

```
priority : leftHand => rightHandSide [type]
```

What is a rule?

*Adding to this work

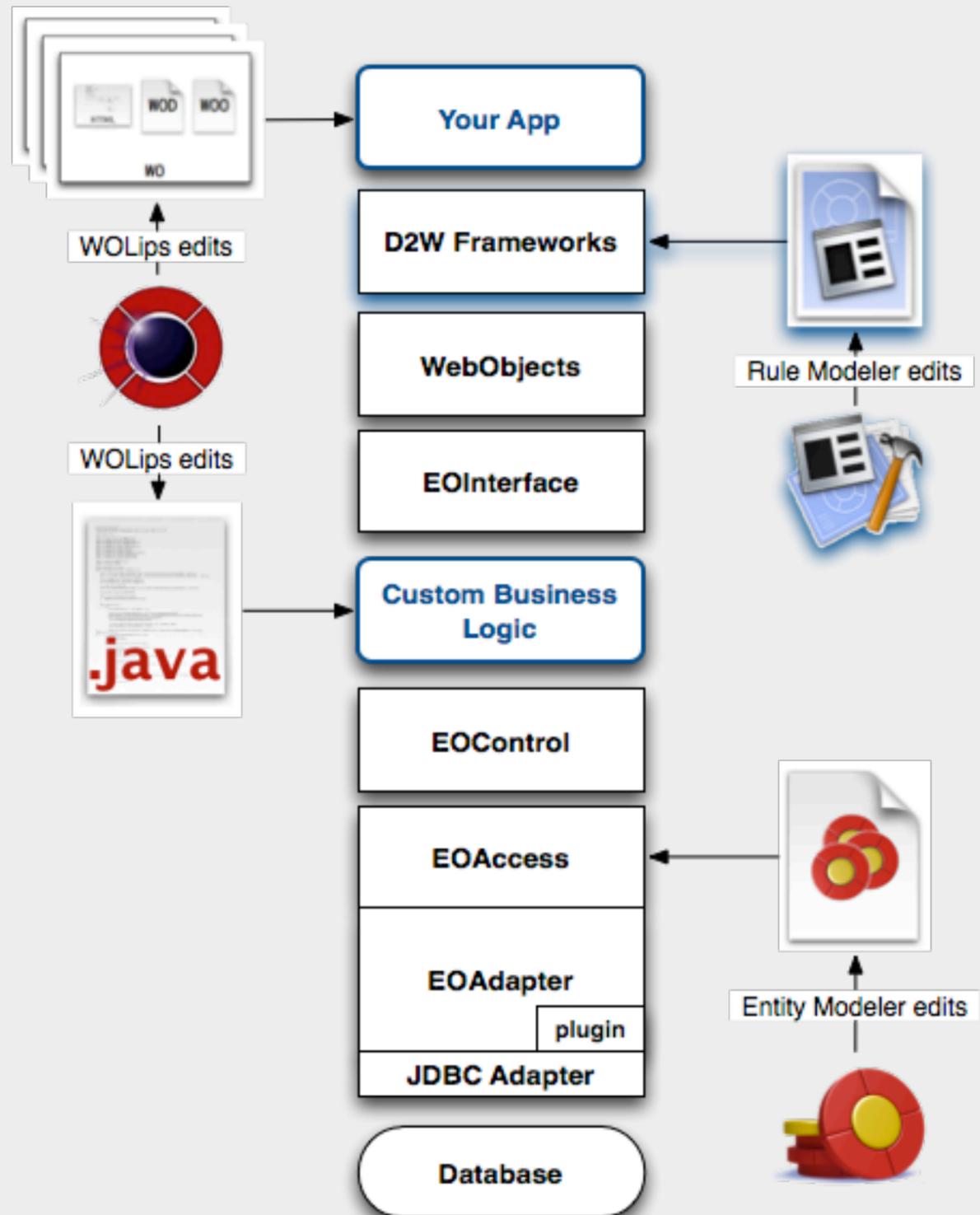


Listing 1: 'Syntax of a Rule'

```
priority : leftHand => rightHandSide [type]
```

The rule engine maintains a state for dynamic pages, caching, rule files, collections of rules, regular and debug mode, rule order, and client configuration.

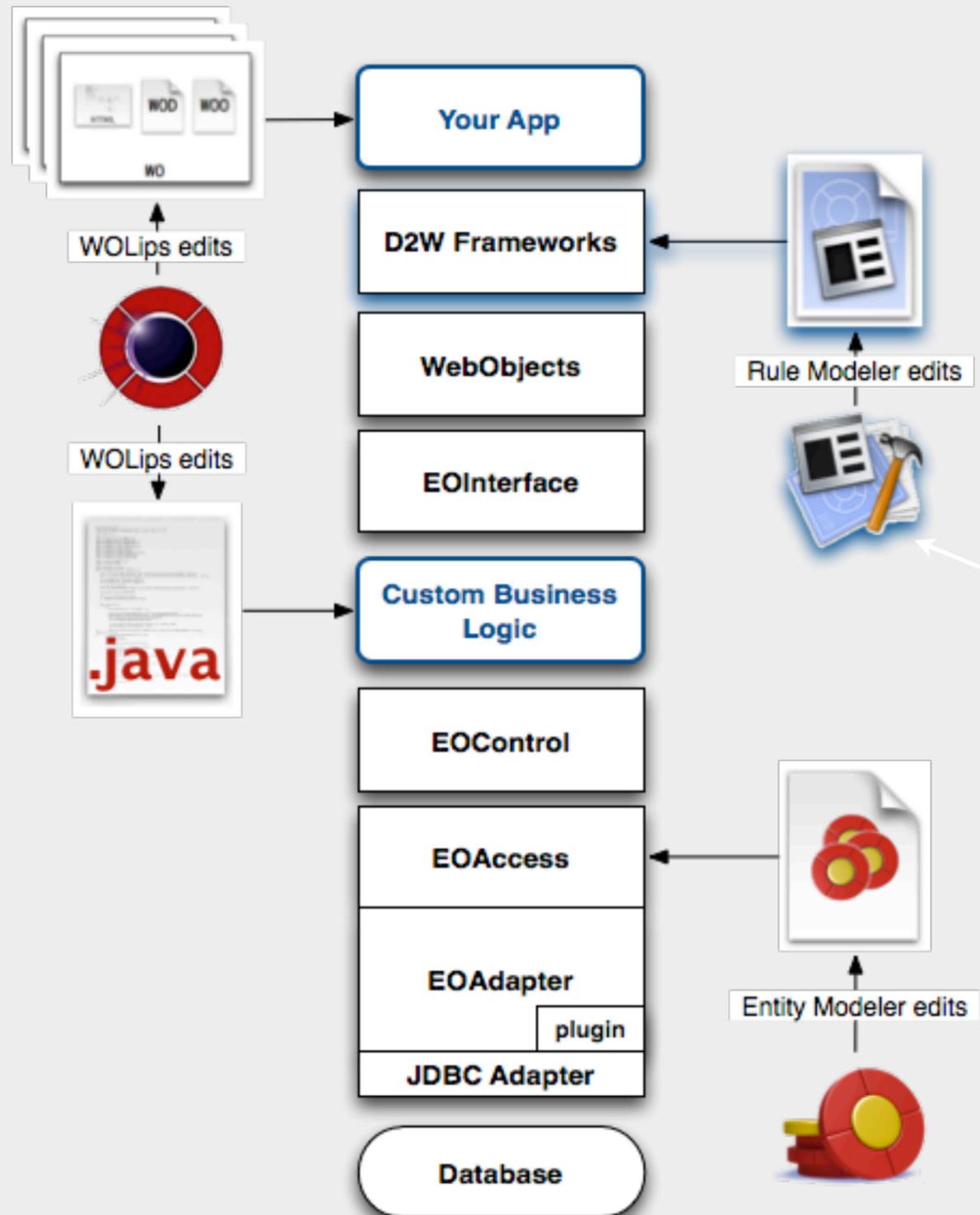
What is a rule?



Listing 1: 'Syntax of a Rule'

```
priority : leftHand => rightHandSide [type]
```

What is a rule?



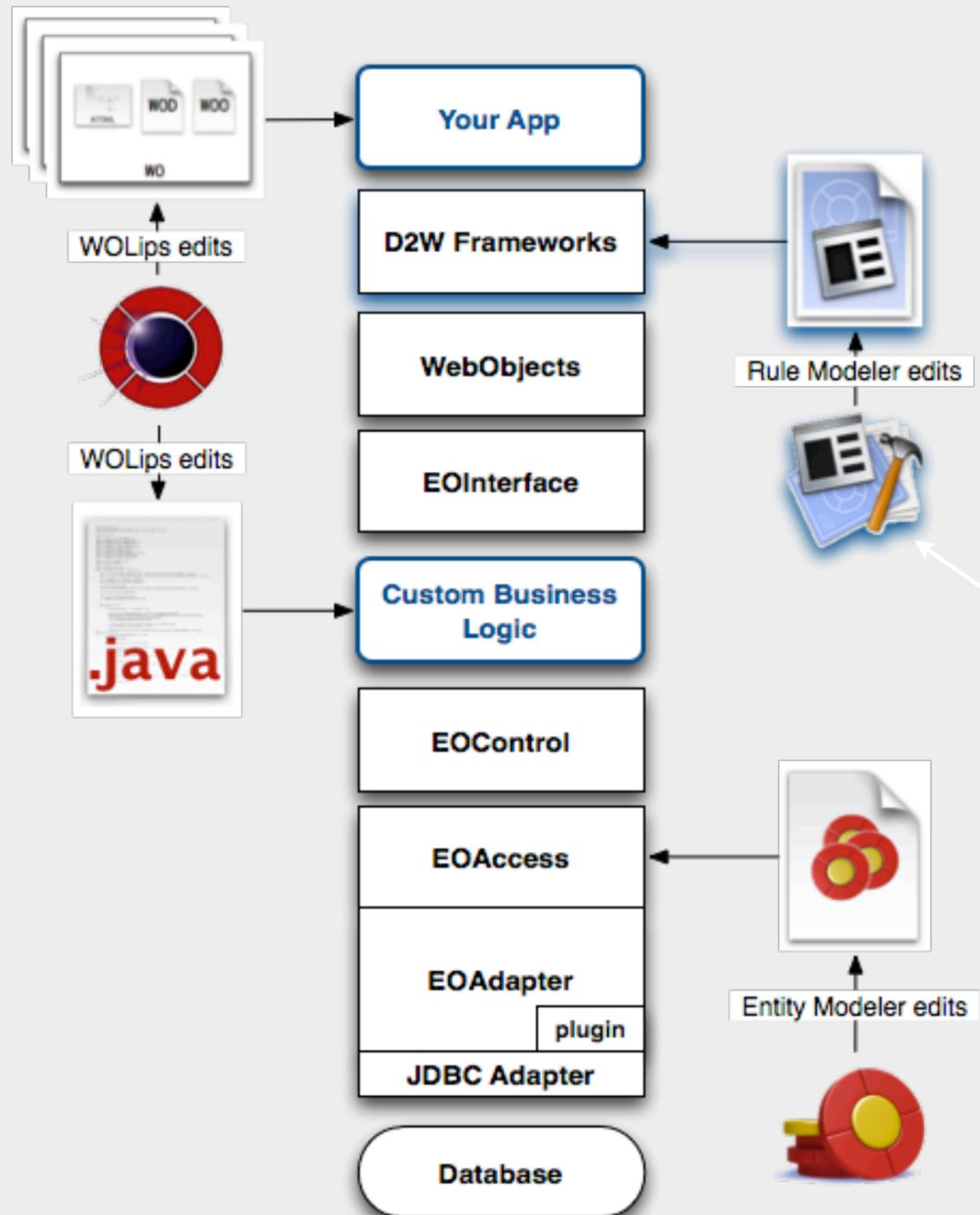
Listing 1: 'Syntax of a Rule'

```
priority : leftHand => rightHandSide [type]
```

Listing 2: Brute Force Memory Loading Rules Group

```
1 : *true* => useBatchingDisplayGroup = false  
  [BooleanAssignment],  
1 : *true* => fetchLimit = 20  
  [Assignment],
```

What is a rule?



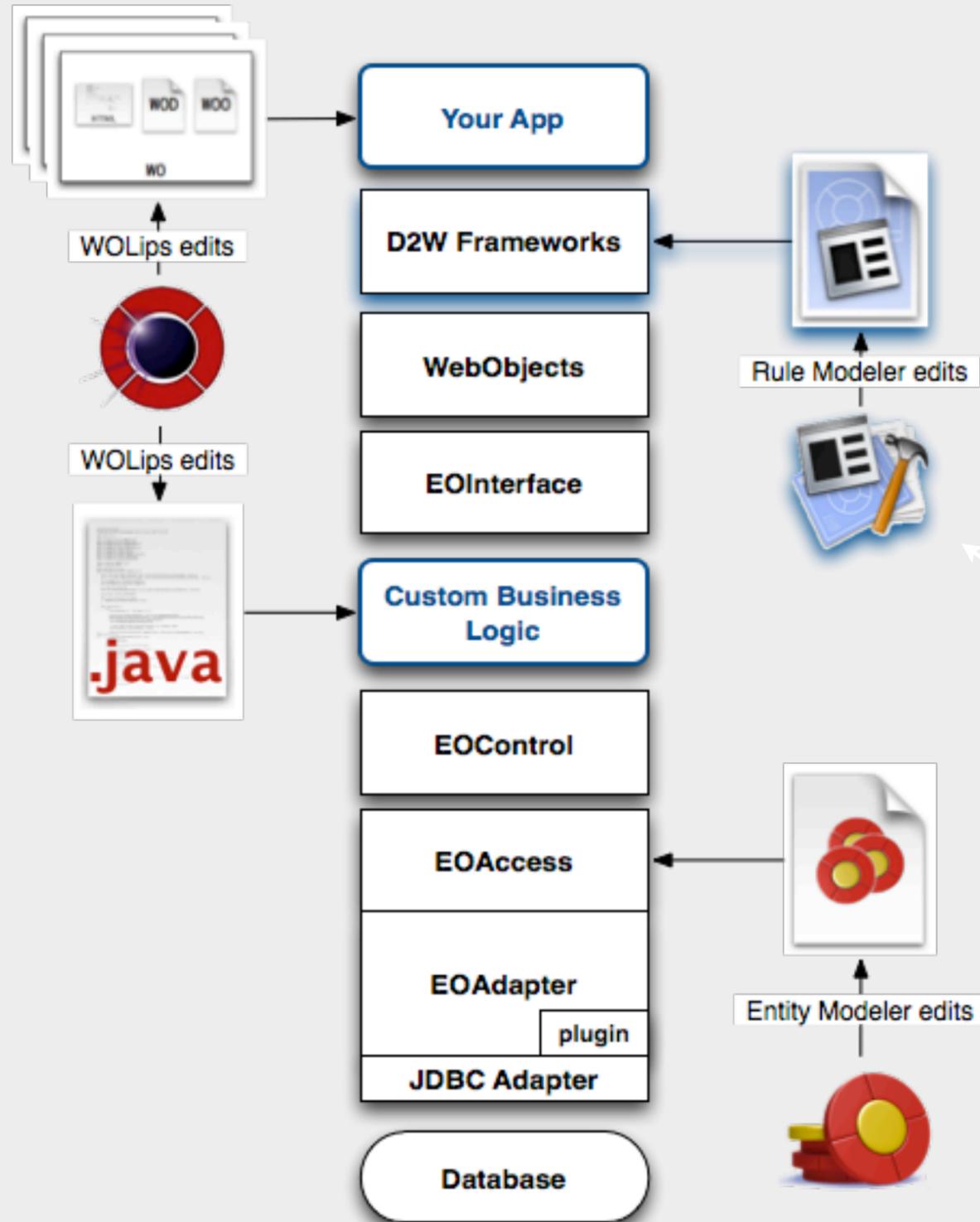
Listing 1: 'Syntax of a Rule'

```
priority : leftHand => rightHandSide [type]
```

Listing 3: Batching Display Group

```
150 : *true* =>
        useBatchingDisplayGroup = true
    [ BooleanAssignment ],
160 : *true* => batchSize = 5
    [ Assignment ]
```

What is a rule?



Listing 1: 'Syntax of a Rule'

```
priority: leftHand => rightHandSide [type]
```

Listing 4: Batching Display Group

```
160:  entity.name='BestTSField' &&
      (task='list' or task='query') =>
      displayPropertyKeys = (
run ,
rerun ,
camcol ,
field
);
```

Best Target Selection Field

Browser: http://elmer.allnightstarparty.org/cgi-bin/WebObjects/RiqsModernMySQL.woa/1/wo/

RiqsModernMySQL

Logged In User · Homepage · Log Out

Home Tab1 Tab2

List BestTS Field

77068 items : Show 10 items

	Run	Rerun	Camcol	Field
Inspect	1,339	21	1	11
Inspect	1,339	21	1	12
Inspect	1,339	21	1	13
Inspect	1,339	21	1	14
Inspect	1,339	21	1	15
Inspect	1,339	21	1	16
Inspect	1,339	21	1	17
Inspect	1,339	21	1	18
Inspect	1,339	21	1	19
Inspect	1,339	21	1	20

Prev 1 2 3 4 5 6 7 ... 7706 7707 Next

[Return](#)

Browser: http://elmer.allnightstarparty.org/cgi-bin/WebObjects/RiqsModernMySQL.woa/1/wo/

RiqsModernMySQL

Logged In User · Homepage · Log Out

Home Tab1 Tab2

Inspect BestTS Field

AG	208.32
AI	208.32
AR	208.32
AU	208.32
AZ	208.32
AaG	-24.68
AaI	-23.68
AaR	-24.11
AaU	-24.23
AaZ	-22.01
AaerrG	.00
AaerrI	.00
AaerrR	.00
AaerrU	.00
AaerrZ	.00
Acceptablemask	2
AirmassG	1.17
AirmassI	1.17
AirmassR	1.17

Testing Concepts

- a server node (Mac Pro) with 6 to 8 GB of RAM
- a client node (MacBook Pro)

Extend Memory

Monitor Applications Hosts Site Preferences Help Log Out

Configuring Instance "1" on elmer.allnightstarparty.org

The Instance must be restarted for changes to take effect.
(Changes to these settings only affect this instance; for all instances, use the [Application Configure Page](#))

Instance Settings

Port:	2003	(The instance must be off to change this setting)
ID:	1	
Path:	/Volumes/wabbit/Hudson/jobs/RigsModernMySQL/workspace/	Pick...
Minimum Active Sessions:	0	
Caching enabled:	<input checked="" type="checkbox"/>	
Output Path:	/Library/WebObjects/Logs/RigsModernMySQL-1	Pick...
Auto Open In Browser:	<input type="checkbox"/>	
Debugging enabled:	<input type="checkbox"/>	
Lifebeat Interval:	30	
** Additional Arguments:	-Xmx3072m	

* The path of the executable (inside the '.woa' directory). ** Setting differs from the Application level setting.

Update Instance Settings

With the current settings, this instance will be started with the following arguments:

```
-WOPort 2003 -WOCachingEnabled YES -WODEbuggingEnabled NO -WOOutputPath /Library/WebObjects/Logs/RigsModernMySQL-1 -WOAutoOpenInBrowser NO -WOAutoOpenClientApplication NO -WOLifebeatInterval 30 -WOLifebeatEnabled YES -WOLifebeatDestinationPort 1085 -WOWorkerThreadCount 8 -WOListenQueueSize 128 -WOWorkerThreadCountMin 16 -WOWorkerThreadCountMax 256 -NSProjectSearchPath {} -WOSessionTimeout 3600 -WOApplicationName RigsModernMySQL -WOMonitorEnabled YES -WONoPause YES -Xmx3072m
```

Adaptor Settings

Send timeout:	<input type="text"/>
Receive timeout:	<input type="text"/>
Connect timeout:	<input type="text"/>
Send Buffer Size:	<input type="text"/>
Receive Buffer Size:	<input type="text"/>

Update Adaptor Settings

Force Quit

- Xms3072m
- Xms5072m
- Xms7072m

RIQS in D2W

Search All

http://elmer.allnightstarparty.org/cgi-bin/WebObjects/RiqsModernMySQL.woa/wa/log

Logged In User · Homepage · Log Out

Home Tab1 Tab2

Search All

BestQSO	where <input type="text" value="Primtarget"/> is = <input type="text"/>	Find	Advanced...
BestRDF	where <input type="text" value="Decl"/> is = <input type="text"/>	Find	Advanced...
BestTS Field	where <input type="text" value="AaerrG"/> is = <input type="text"/>	Find	Advanced...
BestTS Seg	where <input type="text" value="Camcol"/> is = <input type="text"/>	Find	Advanced...
Found Lines	where <input type="text" value="Chisq"/> is = <input type="text"/>	Find	Advanced...
Line Ratios	where <input type="text" value="Name"/> is = <input type="text"/>	Find	Advanced...
Measured Lines	where <input type="text" value="Chisq"/> is = <input type="text"/>	Find	Advanced...
Plateobj	where <input type="text" value="Name"/> is = <input type="text"/>	Find	Advanced...
Specbesttsobj	where <input type="text" value="Ab DeverrG"/> is = <input type="text"/>	Find	Advanced...
Specbesttsobjdistinct	where <input type="text" value="Ab DeverrG"/> is = <input type="text"/>	Find	Advanced...
Specobj	where <input type="text" value="Bestobjid"/> is = <input type="text"/>	Find	Advanced...
Spectargettsobj	where <input type="text" value="Ab DeverrG"/> is = <input type="text"/>	Find	Advanced...
Spectargettsobjdistinct	where <input type="text" value="Ab DeverrG"/> is = <input type="text"/>	Find	Advanced...
Spinventory	where <input type="text" value="Chunkrun"/> is = <input type="text"/>	Find	Advanced...
Targetclose	where <input type="text" value="Closeobjid"/> is = <input type="text"/>	Find	Advanced...
Targettsseg	where <input type="text" value="Camcol"/> is = <input type="text"/>	Find	Advanced...
Zcrosscorrelation	where <input type="text" value="Con"/> is = <input type="text"/>	Find	Advanced...
Zemission	where <input type="text" value="Conf"/> is = <input type="text"/>	Find	Advanced...

Best Target Selection Field

Browser: http://elmer.allnightstarparty.org/cgi-bin/WebObjects/RiqsModernMySQL.woa/1/wo/

RiqsModernMySQL

Logged In User · Homepage · Log Out

Home Tab1 Tab2

List BestTS Field

77068 items : Show 10 items

	Run	Rerun	Camcol	Field
Inspect	1,339	21	1	11
Inspect	1,339	21	1	12
Inspect	1,339	21	1	13
Inspect	1,339	21	1	14
Inspect	1,339	21	1	15
Inspect	1,339	21	1	16
Inspect	1,339	21	1	17
Inspect	1,339	21	1	18
Inspect	1,339	21	1	19
Inspect	1,339	21	1	20

Prev 1 2 3 4 5 6 7 ... 7706 7707 Next

Return

Browser: http://elmer.allnightstarparty.org/cgi-bin/WebObjects/RiqsModernMySQL.woa/1/wo/

RiqsModernMySQL

Logged In User · Homepage · Log Out

Home Tab1 Tab2

Inspect BestTS Field

AG	208.32
AI	208.32
AR	208.32
AU	208.32
AZ	208.32
AaG	-24.68
AaI	-23.68
AaR	-24.11
AaU	-24.23
AaZ	-22.01
AaerrG	.00
AaerrI	.00
AaerrR	.00
AaerrU	.00
AaerrZ	.00
Acceptablemask	2
AirmassG	1.17
AirmassI	1.17
AirmassR	1.17

Results

- Categories of results:
 - nearly no-cache hit index error
 - miss-and-hit
 - nearly-always-hit
- The tables in each category then get assigned rules appropriately set the faulting system to optimize performance.

Nearly No-Cache Hit Index Error

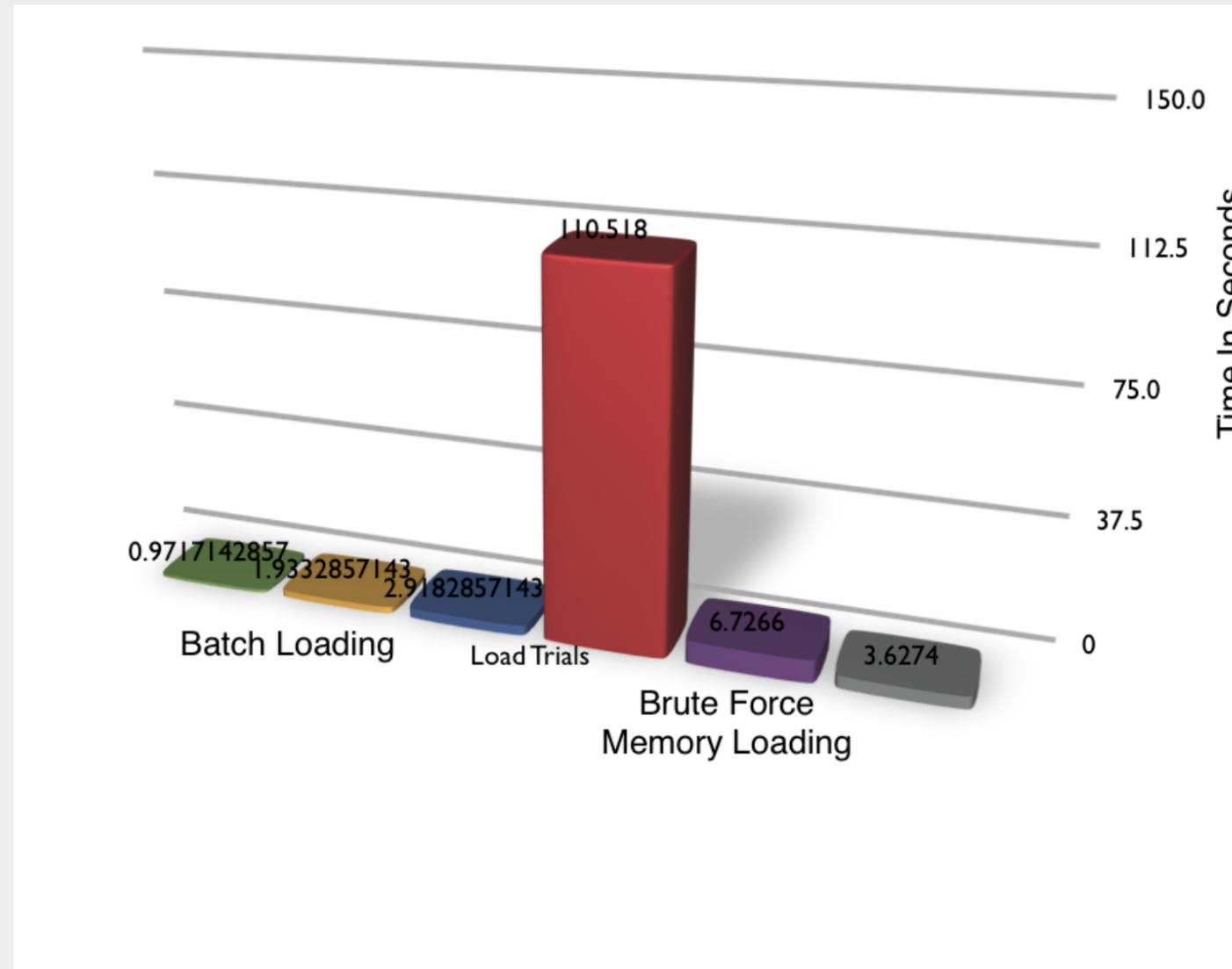


Figure 3: Brute Force Memory Load Delivery Times.

In these cases, less than 1 out 28 attempts actually allow for the system to obtain a cache hit and present the data.

Miss-And-Hit

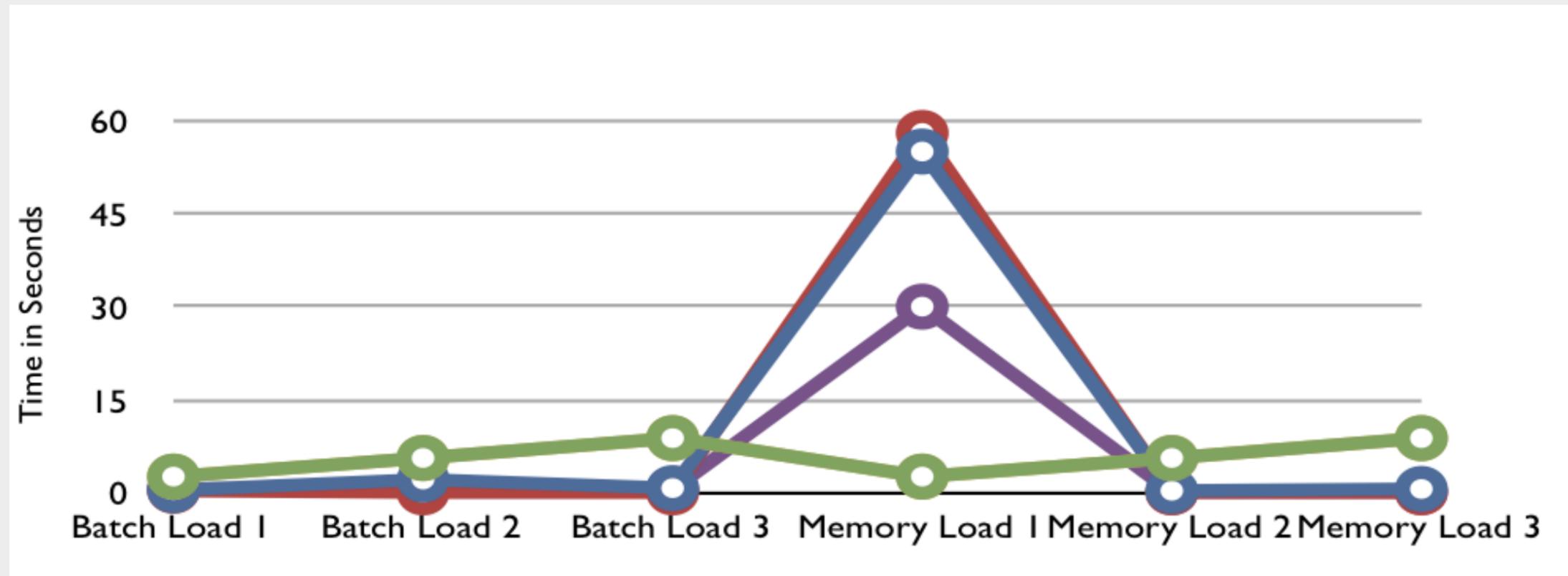


Figure 4: Load Times for the Table Line Ratios

These tables will fail when queried for the first time, but are often successful on the second try.

Miss-And-Hit

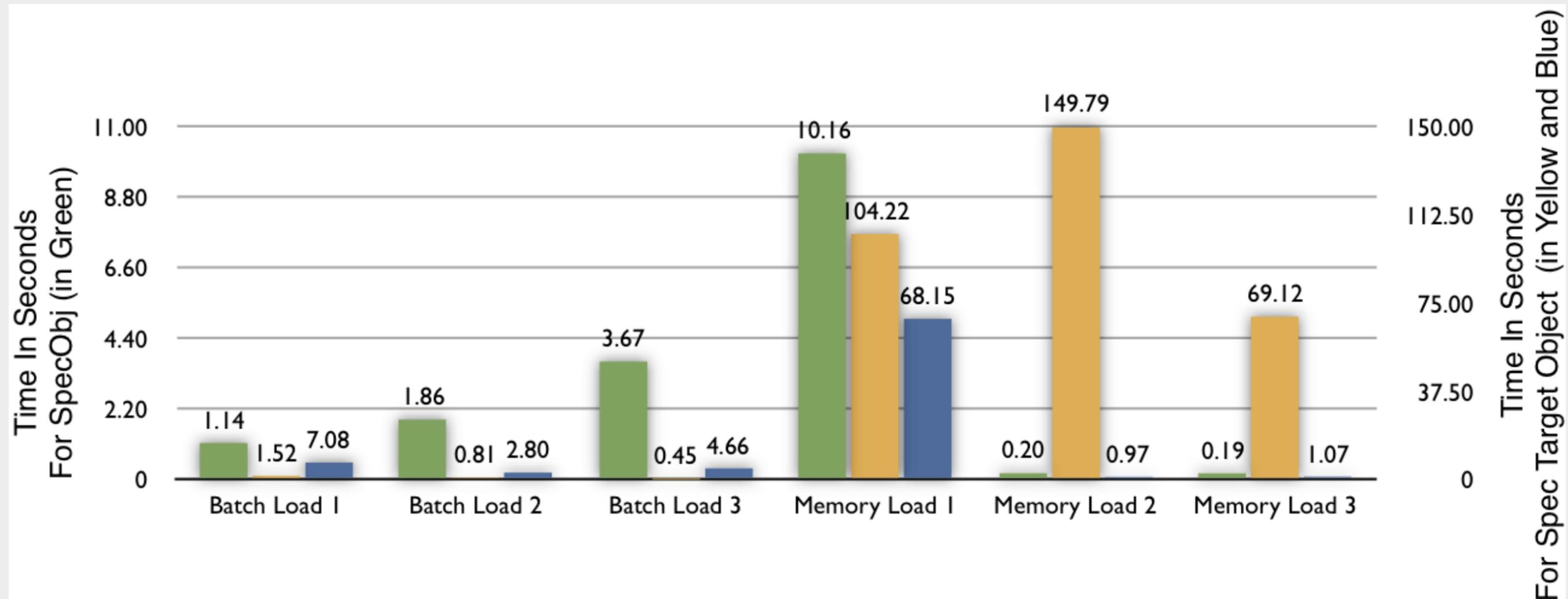


Figure 5: Times for a Spectral Object Family Table for both Batch Fault and Brute Force Memory

These tables will fail when queried for the first time, but are often succeed on the second try.

Memory Consumption

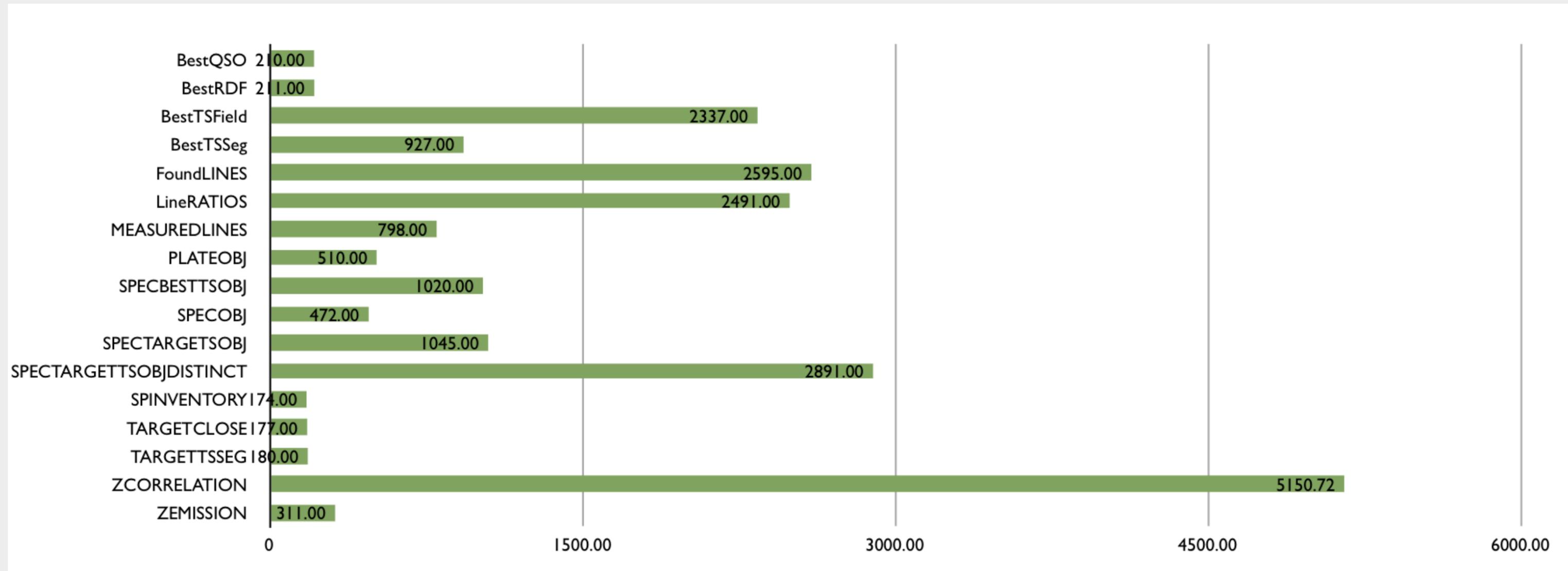


Figure 6: Memory Footprint in MegaBytes (MB) for Data Release 1 tables.

Thus the contributions

- Previous catalogs can be rescued
- With characteristics revealed by D2W it may be possible
 - to map on the fly DRI's data to a better structure
 - migrate this data to a better structure that is also provided by D2W and ERRest
- In either case, WebObjects allow us to exploit and exposing powerful architectures for their use with high volume catalogs



WOWODC

MONTREAL 1/3 JULY 2011



Q&A

Even with legacy metadata catalogs containing 1000 attribute tables, ER ModernLook can be made to service these tables, that leads to mapping solutions to better structures. In some cases, D2W's Display Batch Group can make the requests efficient. In other cases, the faulting of the display batch group and plain blank queries cause a cache miss for accessing the data. This study is how to determine which tables are what, and things one can do to recover data in these conditions.