



Unified Analytics through Apache Lens @Inmobi

Amareshwari Sriramadasu

Agenda

Introduction
to Apache
Lens

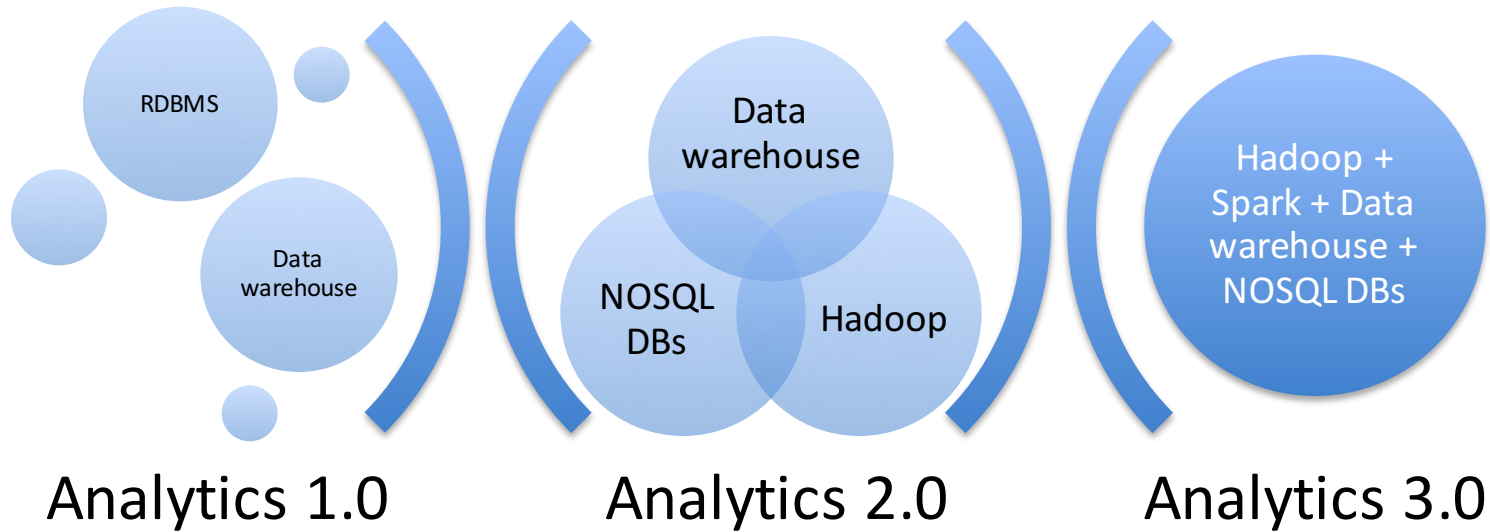
Architecture

How is it
used at
Inmobi?

Challenges
faced

Roadmap

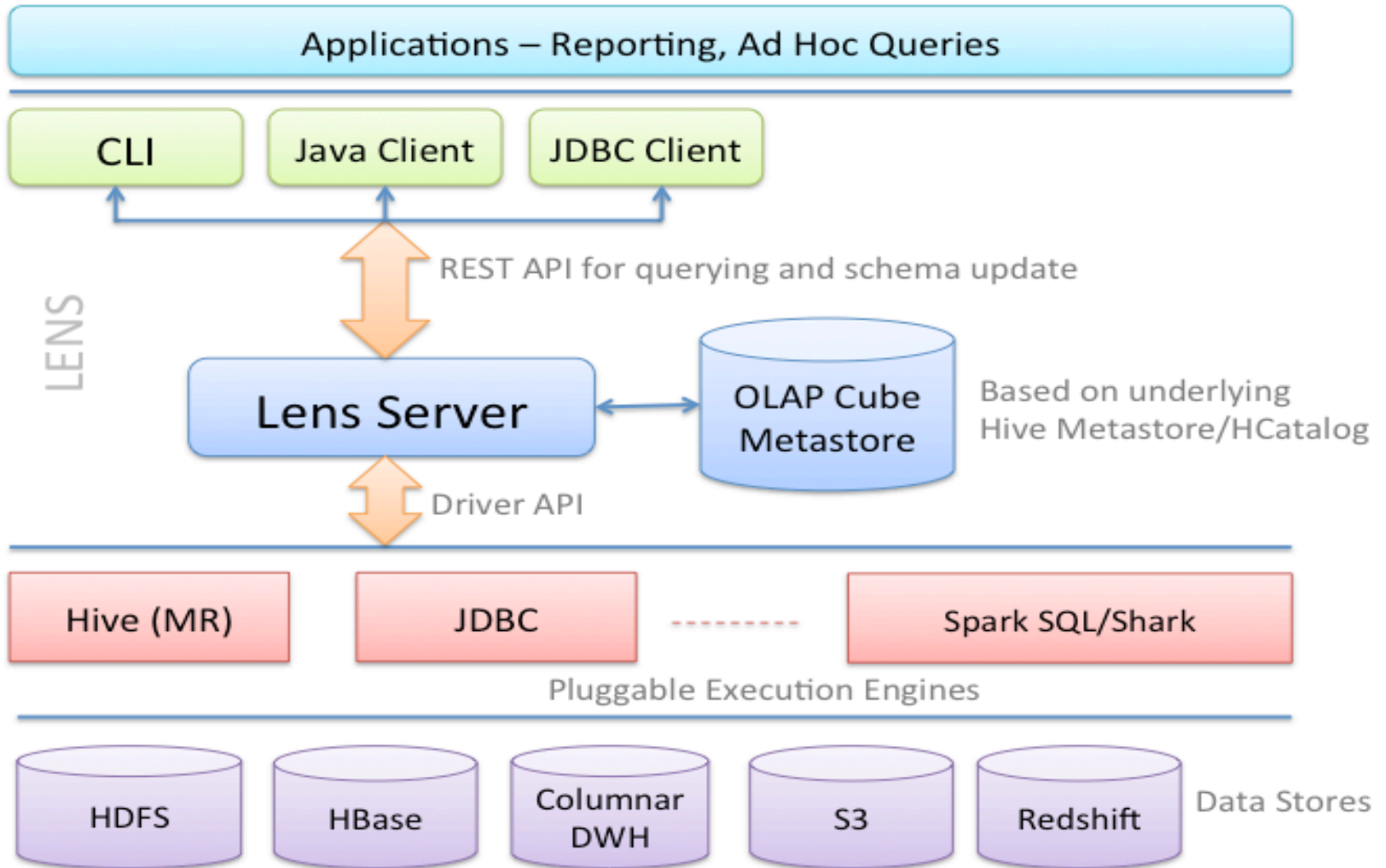
Analytics Evolution



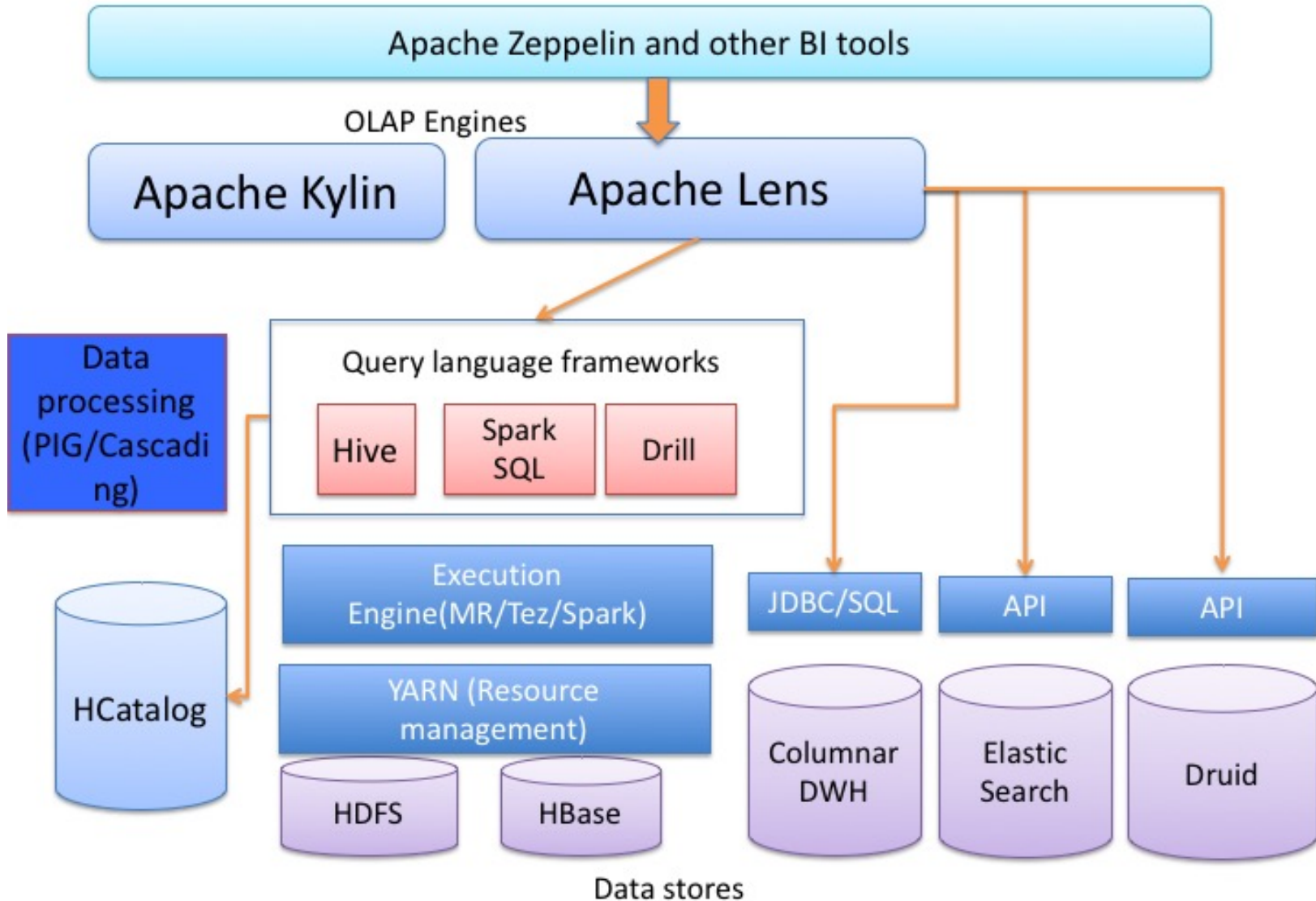
Apache Lens

- *Platform to enable multi-dimensional queries in a unified way over datasets stored in multiple warehouses*
 - OLAP Cube abstraction
 - Data discovery by providing single metadata layer
 - Unified access to data by integrating Hive with other traditional warehouses
 - Queries get pushed to where data resides

Lens Architecture



Where does it fit in ecosystem?



Use cases @Inmobi

- Operational
 - Reporting
 - Close of books
- Decision making
 - Real time decisions
 - Recommendations
- Explorative
 - Adhocanalysis
 - Debugging

How is Lens used at Inmobi?

Unified schema across HDFS and columnar DB through OLAP

Adhoc querying on unified data

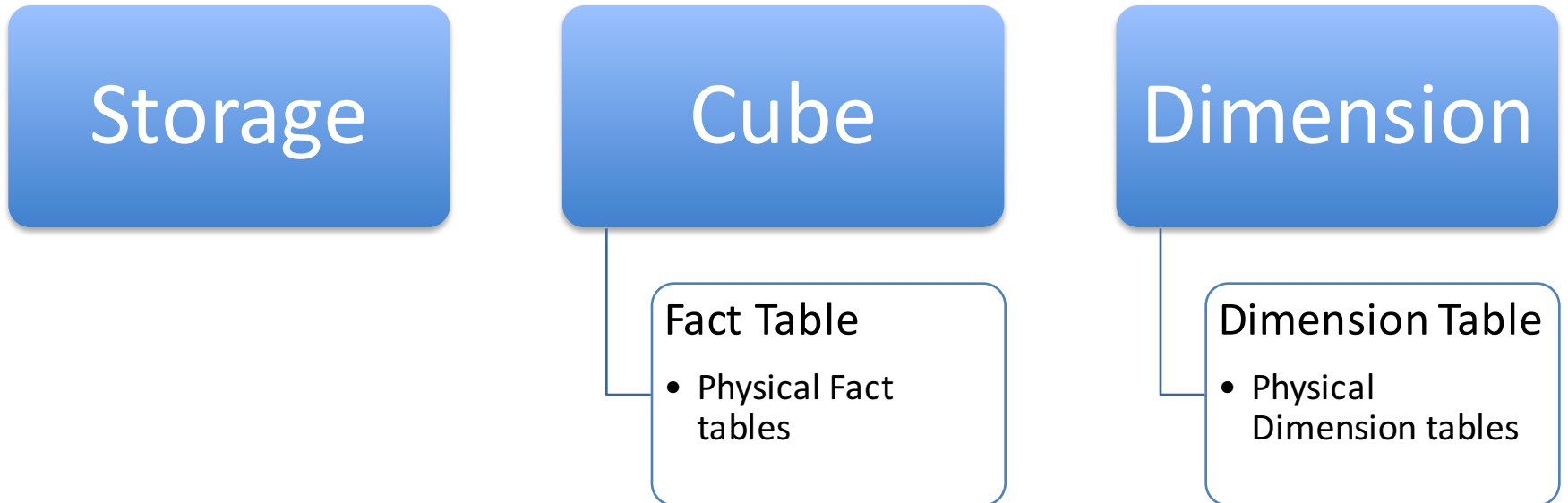
Real time data queryable with Druid driver (In progress)

Experiment with Hive and SparkSQL in production (In progress)

Scheduling reports (Soon)

Defacto data access layer at Inmobi (Soon)

OLAP Data model



Example query

cube select
cityname,
revenue from
testcube where
timerange_in(dt,
'2015-03-10-03',
'2015-03-12-03')

```
SELECT (citytable.name), sum((testcube.revenue)) FROM c2_testfact testcube  
INNER JOIN c1_citytable citytable ON ((testcube.cityid)= (citytable.id)) WHERE ((  
testcube.dt='2015-03-10-03') OR (testcube.dt='2015-03-10-04') OR  
(testcube.dt='2015-03-10-05') OR (testcube.dt='2015-03-10-06') OR  
(testcube.dt='2015-03-10-07') OR (testcube.dt='2015-03-10-08') OR  
(testcube.dt='2015-03-10-09') OR (testcube.dt='2015-03-10-10') OR  
(testcube.dt='2015-03-10-11') OR (testcube.dt='2015-03-10-12') OR  
(testcube.dt='2015-03-10-13') OR (testcube.dt='2015-03-10-14') OR  
(testcube.dt='2015-03-10-15') OR (testcube.dt='2015-03-10-16') OR  
(testcube.dt='2015-03-10-17') OR (testcube.dt='2015-03-10-18') OR  
(testcube.dt='2015-03-10-19') OR (testcube.dt='2015-03-10-20') OR  
(testcube.dt='2015-03-10-21') OR (testcube.dt='2015-03-10-22') OR  
(testcube.dt='2015-03-10-23') OR (testcube.dt='2015-03-11') OR  
(testcube.dt='2015-03-12-00') OR (testcube.dt='2015-03-12 -01') OR  
(testcube.dt='2015-03-12-02') )AND (citytable.dt = 'latest')
```

```
GROUP BY(citytable.name)
```

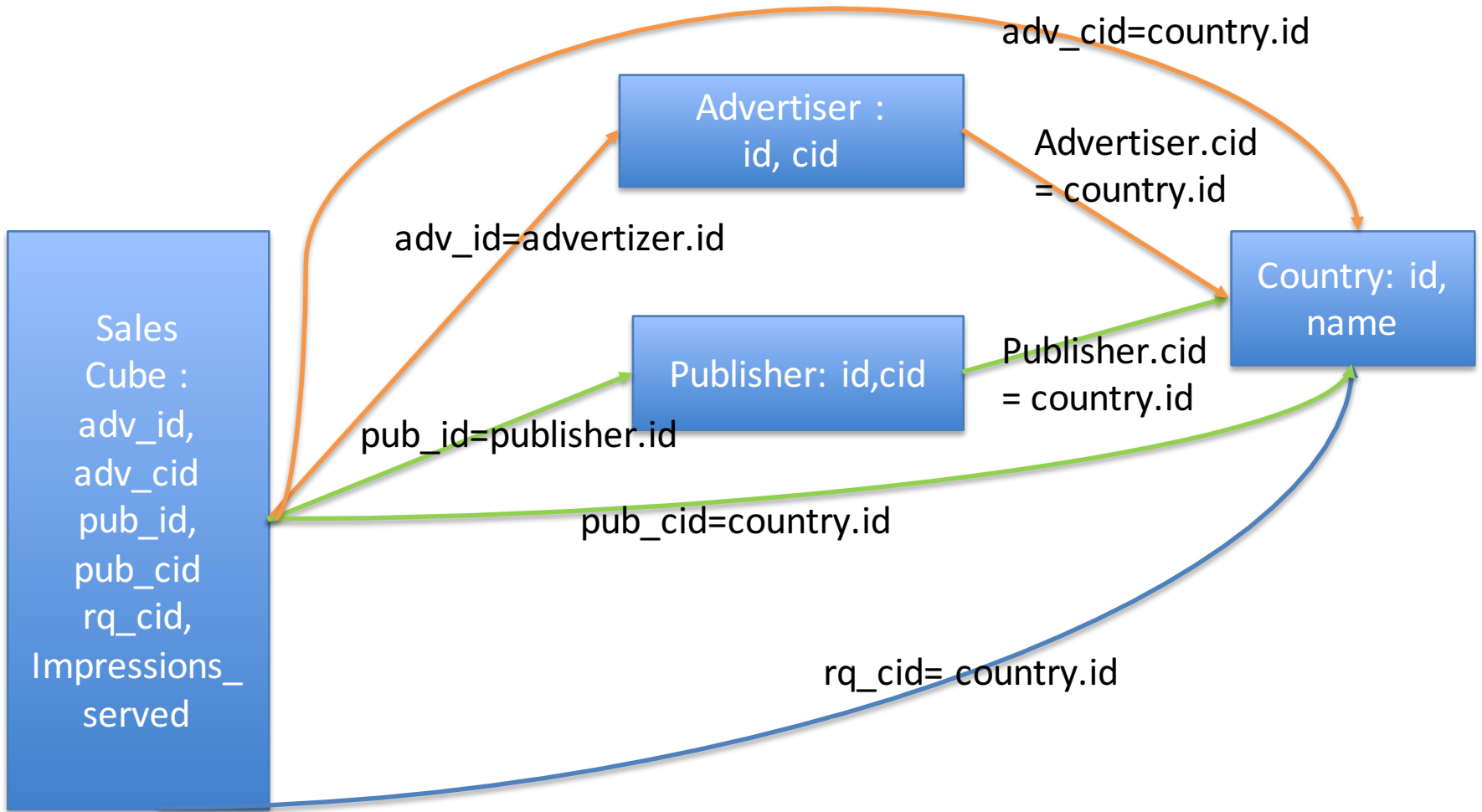
Unifying schema

Multiple expressions

De-normalized variables

Join chains

JoinChains explained



Other challenges

Partition timeline

Query submission throttling

DB Resources

Query statistics

Bridge tables

Bridge Tables explained

User		
ID	Name	Gender
1	Ram	M
2	John	M
3	Suma	F

Fact Table	
User Id	Revenue
1	100
2	50

Report1	
User's sport interest	Revenue
Football	150
Cricket	150
BasketBall	50

Sports	
ID	Name
1	Football
2	Cricket
3	Basketball

User Interests	
User ID	Sport ID
1	1
1	2
2	1
2	2
2	3

Report2	
User's sport interest	Revenue
Football, Cricket	100
Football, Cricket, BasketBall	50

Lens features added in previous releases

- Zeppelin integration
- Pluggable user rewriter
- Elastic search driver
- Query submission throttling
- Saved queries and parameterization
- Instrumentation for all api

Roadmap

Immediate

Move to Apache Hive dependency
Ability to load multiple instances of same driver class
Druid driver

Medium Term

Distributed server deployment
Estimate query execution times
Authorization across all services and storages
Add scheduler service
Make it suitable to integrate with BI tools

Long term

Query caching
Metastore UI
Administrator console
Automatic roll up suggestions on hot datasets

Stay Involved

Its a TLP now!

Web site

- <http://lens.apache.org/>

Source repo

- <https://git-wip-us.apache.org/repos/asf/lens.git>

Mailing lists

- dev@lens.apache.org
- user@lens.apache.org

Thank You!

- Questions?