Mondriaan update

Pentaho community meetup
  Amsterdam
September 2012

@julianhyde

Agenda

Mondrian 4 – beta

Other new stuff

(Yahoo)

Mondrian 4 – What's new?

Attributes

Measure groups

Physical schema

Internals

Richer semantic model

Physical schema:

- Only define attributes and relationships once

- Compound keys

Attribute hierarchies

Hierarchies & attributes grouped into dimensions

- E.g. Customers dimension contains Customer hierarchy (State-City-Customer) and Age, Gender, Salary attribute hierarchies

Measure groups

In Mondrian 3.x, if you want a cube with multiple fact tables, you build a virtual cube:

```
<Cube name="Sales">
    <Table name="sales_fact"/>
</Cube>

<Cube name="Warehouse">
    <Table name="warehouse_fact"/>
</Cube>

<VirtualCube name="Warehouse and Sales">
    <Cube name="Sales"/>
    <Cube name="Warehouse"/>
</VirtualCube>
```

Measure groups (2)

In Mondrian 4, cubes can contain multiple measure groups

Virtual cubes are obsolete

Many-to-many association between measure groups and dimensions

```
<Cube name="Warehouse and Sales">
    <MeasureGroups>
        <MeasureGroup name="Sales">
            <Table name="sales_fact"/>
            <Measure name="unit_sales"/>
        </MeasureGroup>
        <MeasureGroup name="Warehouse">
            <Table name="warehousee_fact"/>
            <Measure name="inventory_units"/>
        </MeasureGroup>
    </MeasureGroups>
</Cube>
```

Different ways to link dimensions to fact tables

Aggregate tables are measure groups

	Sales	Warehouse
Time	X	X
Product	X	X
Customer	X	
Warehouse		X

Gone / Replacements

Mondrian 3 schema

Mondrian 4 Schema
Schema upgrader

Aggregate recognizer

Aggregate table API
(define / enable /
disable)

Schema workbench

Pentaho modeler?

XMLA server

olap4j-xmlaserver
@github

Hierarchy syntax
- `[Time.Weekly].[Day]`
- `[Time].[Month]`

SSAS-style syntax
- `[Time].[Weekly].[Day]`
- `[Time].[Time].[Month]`

Done / Remaining

The important things work!

Ragged hierarchies

Schema converter

Analyzer upgrade

2511 of 2770 tests pass

Aggregate table API

Complex schema mappings

Beta

1. Download from CI

`http://ci.pentaho.com/view/Analysis/job/mondrian-git-4.0/`

2. Run Mondrian-4 on your current schema

- Auto-upgrade

- Schema converter tool TBA

- MDX syntax differences

  `mondrian.olap.SsasCompatibleNaming=true`

3. Write a new-style schema

4. Log bugs!

Futures

Open-source business analytics

Mondrian
IN ACTION

William D. Back
Nicholas A Goodman
Julian Hyde

"Mondrian in Action" book

Publish date: Spring 2013

Join the early-access program:
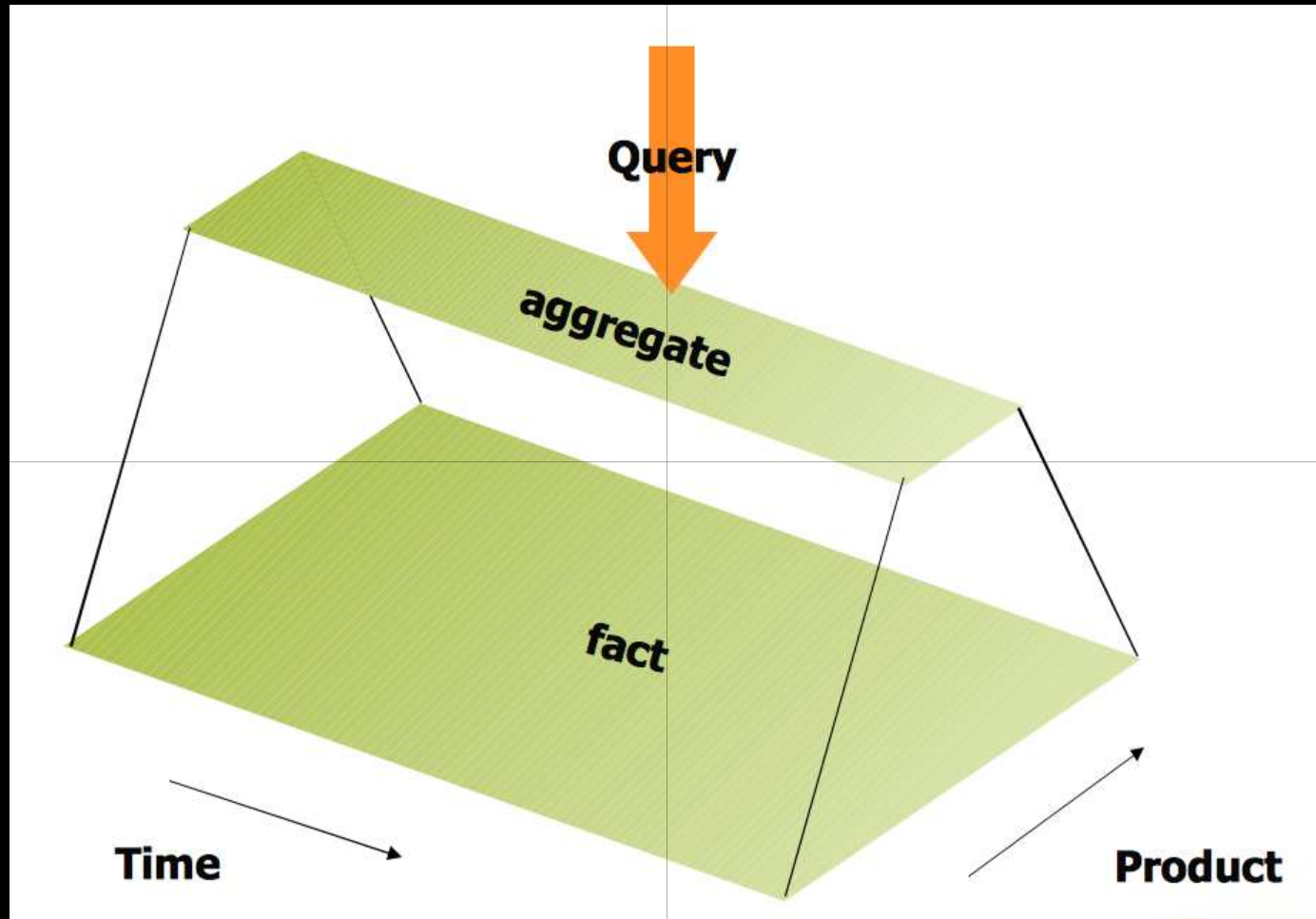 http://www.manning.com/back/

Future features

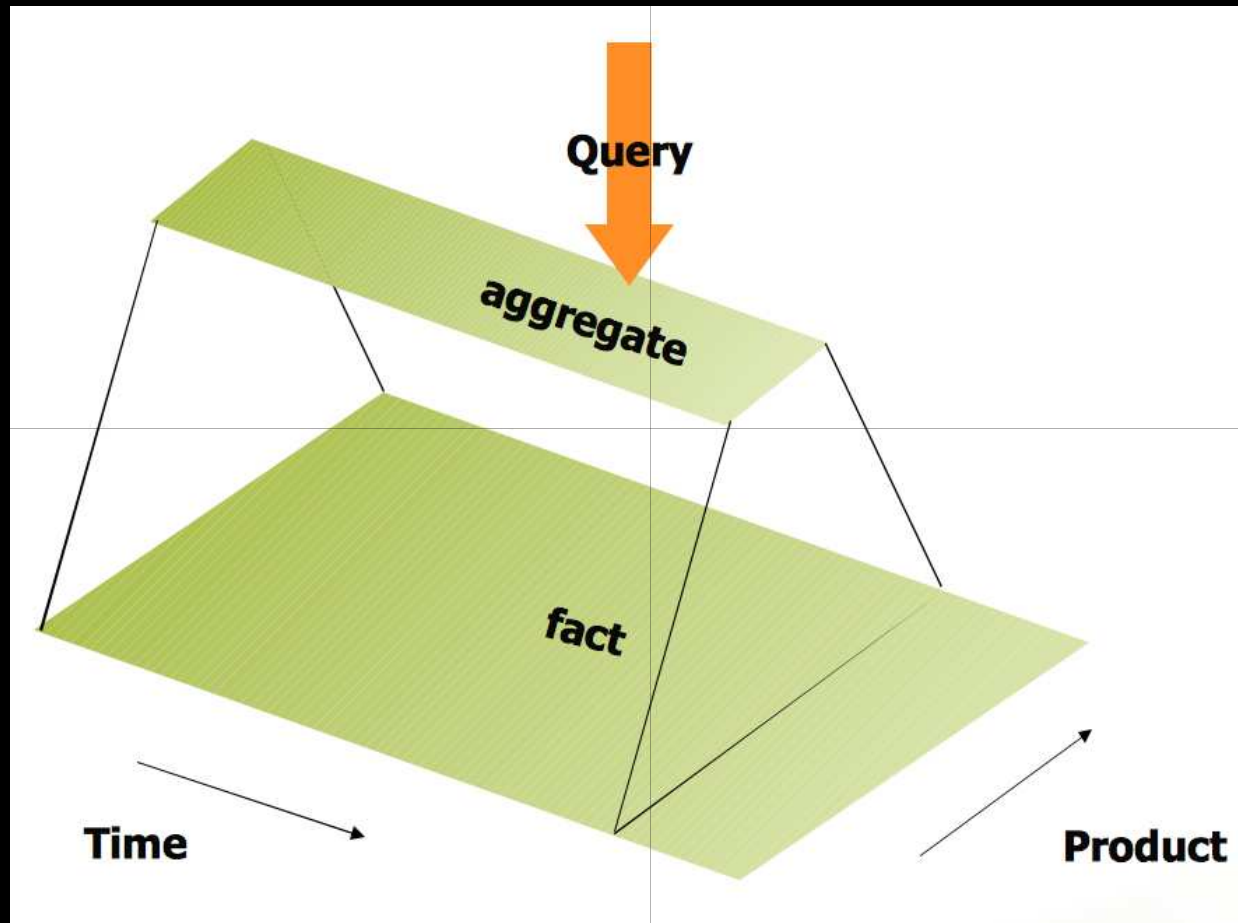Shelved aggregate tables

Connections

- Defined in schema

- Multiple connections

- Non-JDBC databases

Advanced SQL generation

Regular aggregate table

Shelved aggregate table

Aggregate table API – some ideas

- Define

- Enable

- Disable

- Specify beginning/end of valid range

- Kettle can tell Mondrian that aggregate table is no longer valid

- Kettle can ask Mondrian to tell it when it has finished using an aggregate table

Multiple connections in schema

```
<Schema name='FoodMart'>
    <Connections>
      <Connection name='default' default='true' uuid='abcd-1234'>

  <Jdbc>jdbc:mysql://localhost/foodmart?characterEncoding=latin1&l
  t;/Jdbc>
        <JdbcUser>foodmart</JdbcUser>
        <JdbcPassword>foodmart</JdbcPassword>
      </Connection>
      <Connection name='aggs' default='false' uuid='abcd-2345'>

  <Jdbc>jdbc:mysql://localhost/foodmartAggs?characterEncoding=lati
  n1</Jdbc>
        <JdbcUser>foodmartAggs</JdbcUser>
        <Properties>
          <Property name='prop1'>value1</Property>
          <Property name='prop2'>value2</Property>
        </Properties>
      </Connection>
    </Connections>
```

- Cannot join tables from different connections
- Also: non-JDBC connection (via SPI or Optiq)

Advanced SQL generation

- Access control

- Killing big IN lists

- Push down aggregates (esp. time ranges)

- Need a new strategy... TBD

Summary

Mondrian 4 – A major improvement to Mondrian model & engine

As compatible as possible

Will enable further improvements in performance / flexibility in upcoming releases

Help us test it, and get it to production quality faster

Questions?

@julianhyde

jhyde@pentaho.com

http://julianhyde.blogspot.com

https://github.com/julianhyde/