Bay Area Pentaho User Group

San Francisco . USA

pentaho

April 3rd, 2014
@julianhyde

What's new in Mondrian 4?

San Francisco . USA

pentaho

Mondrian 4 – What's new?

- Attributes & dimensions
- Physical schema
- Measure groups & aggregate tables
- Internals
- Delivery
- Futures

Richer schema

Richer semantic model

Goals

Easier schema evolution

More flexible for user

- "Attribute-oriented analysis"

Wider range of schemas

Non-SQL databases

Features

Attributes and Measures

Dimensions as buckets of attributes

Hierarchies to make navigation easier

Physical schema:

- Only define tables, columns, relationships once
- Composite keys

Measure groups representing fact and aggregate tables

Attributes

Mondrian 3: Dimension, Hierarchy, Level

Mondrian 4: Dimension, Attribute

Dimension: Customer

Attributes: Nation, Gender, Zipcode, State, City, Age, Name

Hierarchy: Customers {Nation, State, City, Name}

Attribute hierarchies: Nation, Gender, Zipcode, State, City, Age, Name (created automatically)

Hierarchies

Hierarchies are much less important in Mondrian 4 than Mondrian 3.

Build dimensions and attributes first, hierarchies & levels later.

Hierarchies optimize the user experience — double-click to drill

Defining an attribute

<Attribute> attributes:

- Key column (or columns if composite)
- Name (default key)
- Caption (default name)
- Order column (default name)
- Properties

All of this stuff was in <Level> in Mondrian-3.

Attribute key must be unique:

- `[Customer].[City].&[San Francisco]&[CA]`
- `[Time].[Month].&[4]&[2014]`

Physical schema

- Define joins, data types, SQL expressions only once
- Portability across different kinds of database (including non-SQL), and multiple databases

```
<PhysicalSchema>
  <Table name='product' keyColumn='product_id'/>
  <Table name='product_class' keyColumn='product_class_id'/>
  <Link target='product' source='product_class'>
    <ForeignKey>
      <Column name='product_class_id'/>
    </ForeignKey>
  </Link>
```

Calculated columns

```
<PhysicalSchema>
  <Table name='customer'>
    <Key>
      <Column name='customer_id'/>
    </Key>
    <ColumnDefs>
      <CalculatedColumnDef name='full_name' type='String'>
        <ExpressionView>
          <SQL dialect='oracle'>
           <Column name='fname'/>  || ' ' || <Column name='lname'/>
          </SQL>
          <SQL dialect='access'>
            <Column name='fname'/>  + ' ' + <Column name='lname'/>
          </SQL>
```

- <Column/> element allows Mondrian to use correct table alias

Localization

```
<Cube name='Sales' defaultMeasure='Unit Sales'>
  <Annotations>
    <Annotation name='caption.de_DE'>Verkaufen</Annotation>
    <Annotation name='caption.fr_FR'>Ventes</Annotation>
    <Annotation name='caption+fr_FR'>Ventes</Annotation>
    <Annotation name='description.fr_FR'>Cube des ventes</Annotation>
    <Annotation name='description.de'>Cube Verkaufen</Annotation>
    <Annotation name='description.de_AT'>Cube den Verkaufen</Annotation>
  </Annotations>
```

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>
```

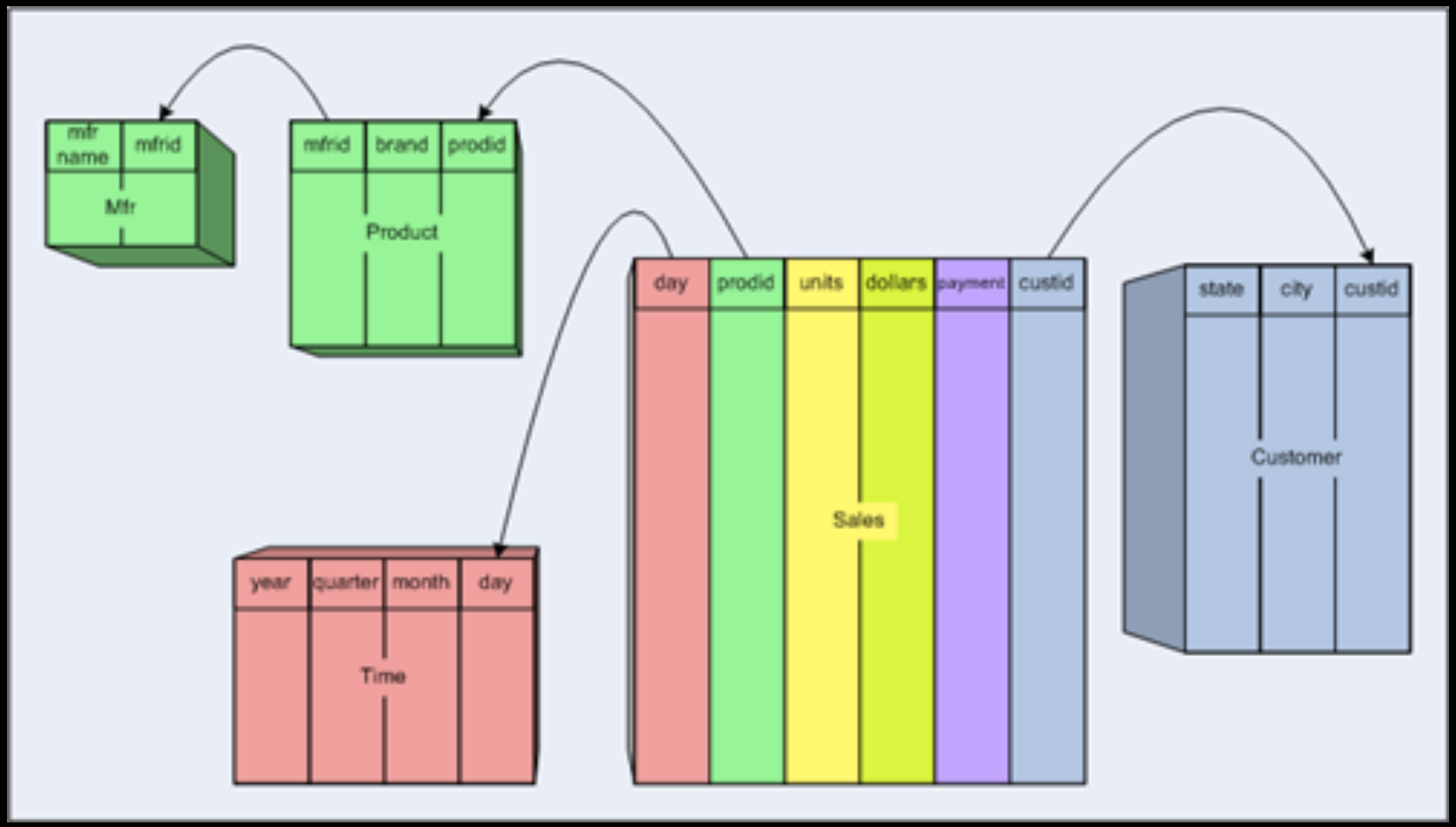
Measure groups (2)

- In Mondrian 4, cubes can contain multiple measure groups
- Virtual cubes are obsolete
- Many-to-many association between measure groups and dimensions
- Different ways to link dimensions to fact tables
- Aggregate tables are measure groups

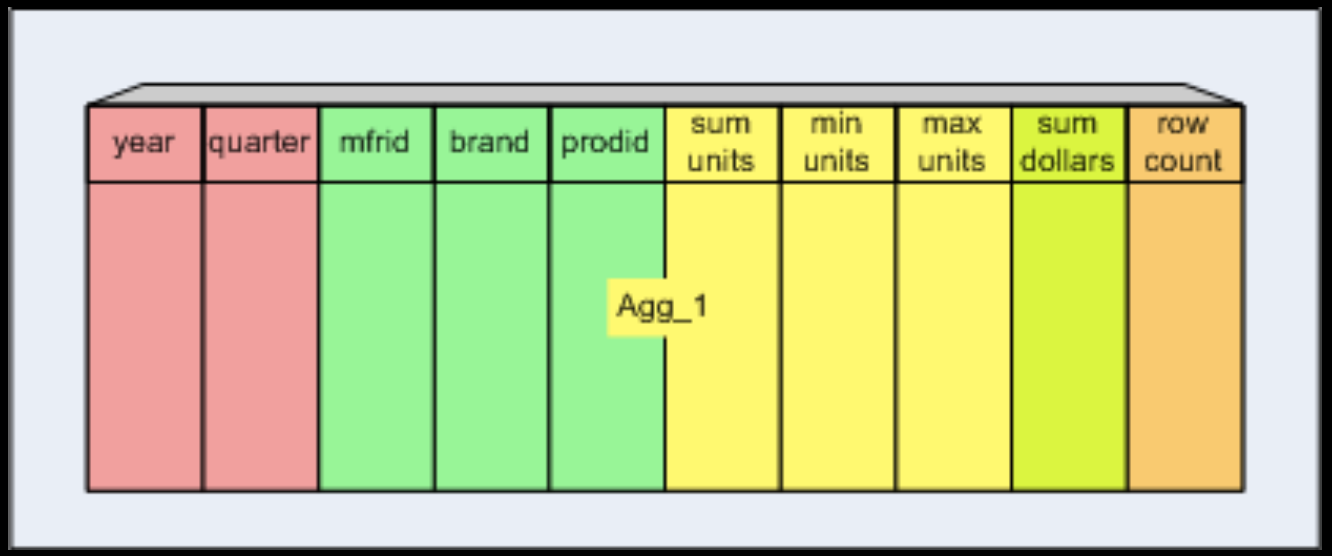
```
<Cube name="Warehouse and Sales">
  <MeasureGroups>
    <MeasureGroup name="Sales">
      <Table name="sales_fact"/>
      <Measure name="unit_sales"/>
    </MeasureGroup>
    <MeasureGroup name="Warehouse">
      <Table name="warehouse_fact"/>
      <Measure name="inventory_units"/>
    </MeasureGroup>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

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

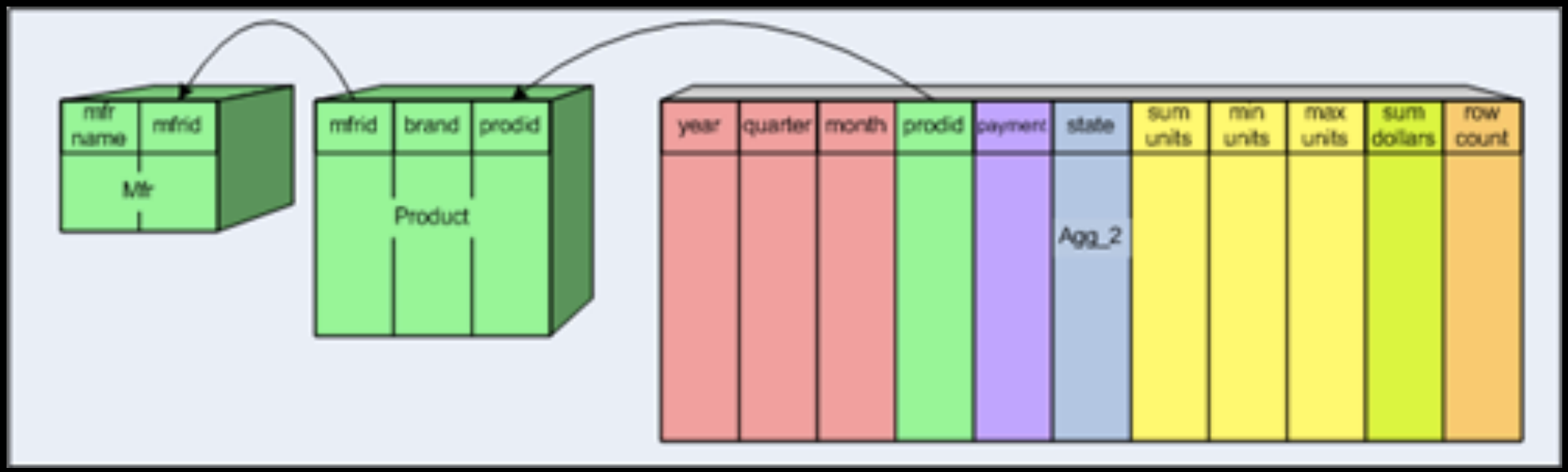
Fact tables


mfr name
mfrid
Mfr
mfrid
brand
prodid
Product
day
prodid
units
dollars
payment
custid
Sales
state
city
custid
Customer
year
quarter
month
day
Time

Simple aggregate table



More complex aggregate table



Hybrid fact/aggregate tables

Measure groups (fact tables):

- Have varying dimensionality, granularity;
- Reference dimension tables;
- Store un-aggregated measures.

Aggregate tables:

- Have various dimensionality, granularity;
- Sometimes reference dimension tables;
- Store aggregated measures.

Let’s store measures in aggregate tables!

Better… let’s make aggregate tables and measure groups the same thing.

Delivering Mondrian 4

Gone	*Replacement*
Mondrian 3 schema	Mondrian 4 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]`
JPivot no longer in distro	Get your own UI: Analyzer, Saiku, Pivot4J

Done	*Remaining*
The important things work!	Ragged hierarchies
Schema converter	Aggregate table API
3,224 of 3,277 tests pass	Pentaho modeler
Analyzer, Saiku, Pivot4J integration	Complex schema mappings
	BA server compliance testing

Beta

1. Binaries at http://repo.pentaho.org/artifactory/repo under pentaho:mondrian:4.0.0
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 http://jira.pentaho.com/browse/MONDRIAN or send a pull request: https://github.com/pentaho/mondrian/tree/lagunitas
5. Getting started: http://mondrian.pentaho.com/documentation/developers_guide.php
6. Running Mondrian as an XMLA service: https://github.com/ThoughtWorksInc/mondrian-xmla-spike

Mondrian 4 in Pentaho BA server

- Mondrian 4 is an OSGI compatible module. It can be used alongside Mondrian 3.
- Both versions of Mondrian are available.
- Create olap4j connections to either one:
  “jdbc:mondrian:Catalog=…” or
  “jdbc:mondrian4:Catalog=…”
- Behind the scenes, we inject our OSGI driver into the top-level class loader and we register with the java.sql.DriverManager
- *Caveat*: All classpath resources must also be OSGI compatible and bundled as modules. (No loose files in WEB-INF/classes.)
- Third party plugins: Most of the more popular analysis UIs of the Pentaho ecosystem already support Mondrian 4.

Future features

Connections

- Defined in schema
- Multiple connections
- Non-SQL databases

Advanced SQL generation

Multiple connections in schema

```
<Schema name='FoodMart'>
  <Connections>
    <Connection name='default' default='true' uuid='abcd-1234'>
      <Jdbc>jdbc:mysql://localhost/foodmart</Jdbc>
      <JdbcUser>foodmart</JdbcUser>
      <JdbcPassword>foodmart</JdbcPassword>
    </Connection>
    <Connection name='aggs' default='false' uuid='abcd-2345'>
      <Jdbc>jdbc:mysql://localhost/foodmartAggs</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)
- Evaluate as much as possible of the MDX query in SQL
- Pivot & decompose
- MDX query planning
- Cache segments as in-memory tables

Open-source business analytics

Mondrian IN ACTION

William D. Back
Nicholas A Goodman
Julian Hyde

MANNING

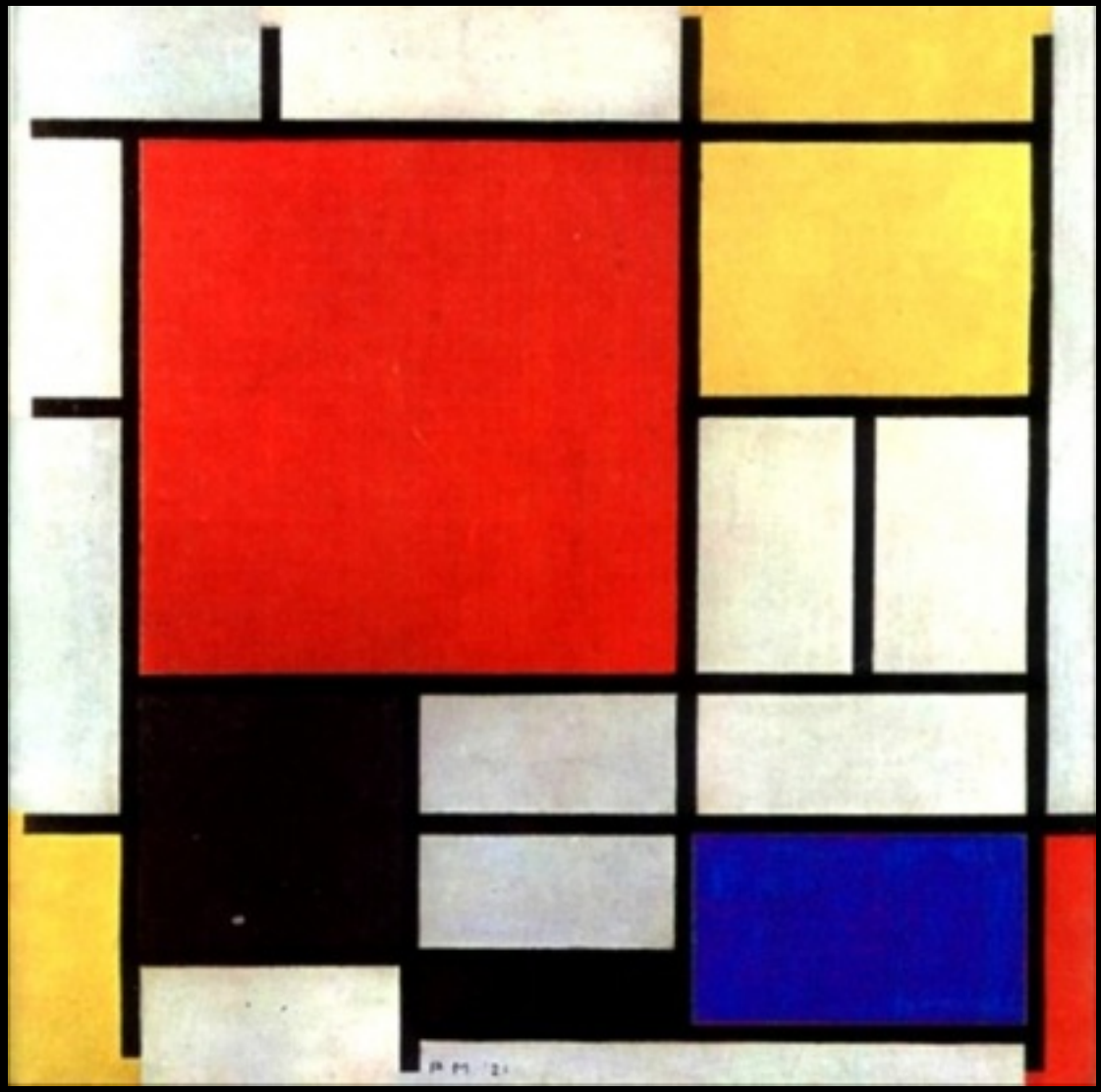
Summary

Mondrian 4 – A major improvement to Mondrian model & engine

As compatible as possible

Will enable further improvements in performance / flexibility in upcoming releases

Questions?



@julianhyde

http://julianhyde.blogspot.com

http://github.com/julianhyde

http://community.pentaho.com/projects/mondrian/

Bay Area Pentaho User Group

San Francisco . USA

pentaho