



2017
PentahoDay

Curitiba . Brasil

11 e 12 de Maio



Design Patterns para Tuning Pentaho com Ctools

Eduardo Alves de Carvalho

Objetivo

Em primeiro lugar, o título do curso é ambicioso, mas não tem a intenção de elencar <todas> as melhorias possíveis e/ou necessárias para uma solução Pentaho. Mas sim, em ajudar a lembrar de pontos importantes, que às vezes podem passar despercebidos .

Eduardo Alves de Carvalho

Consultor Sênior na e-
Setorial Business
Analytics

www.e-Setorial.com.br



Colocando o Pentaho para funcionar

- Para nossos estudos, vamos criar um plugin Pentaho simples, com apenas um filtro e um gráfico;
- Pelo menu Tools/AppBuider, crie um plugin com o nome “DesignPatternsPentaho” e um Dashboard com o nome “pentahoday”;
- Aplique um template, por exemplo o “Filter Template”;
- Crie um Simple Parameter com o nome “anoParam” e valor “All Years”;

Crie dois data sources do tipo mdx over mondrianJndi

Name dsFiltroAno

Jndi SampleData

Mondrian Schema SteelWheels

Output Columns 0

```
Query      WITH
           SET [~COLUMNS] AS
             {[Product].[All Products]}
           SET [~ROWS] AS
             Hierarchize({[Time].[All Years]}, {[Time].
[Years].Members}))
           SELECT
           NON EMPTY [~COLUMNS] ON COLUMNS,
           NON EMPTY [~ROWS] ON ROWS
           FROM [SteelWheelsSales]
```

Crie dois data sources do tipo mdx over mondrianJndi

Name dsVendasPaisLinhaAno

Jndi SampleData

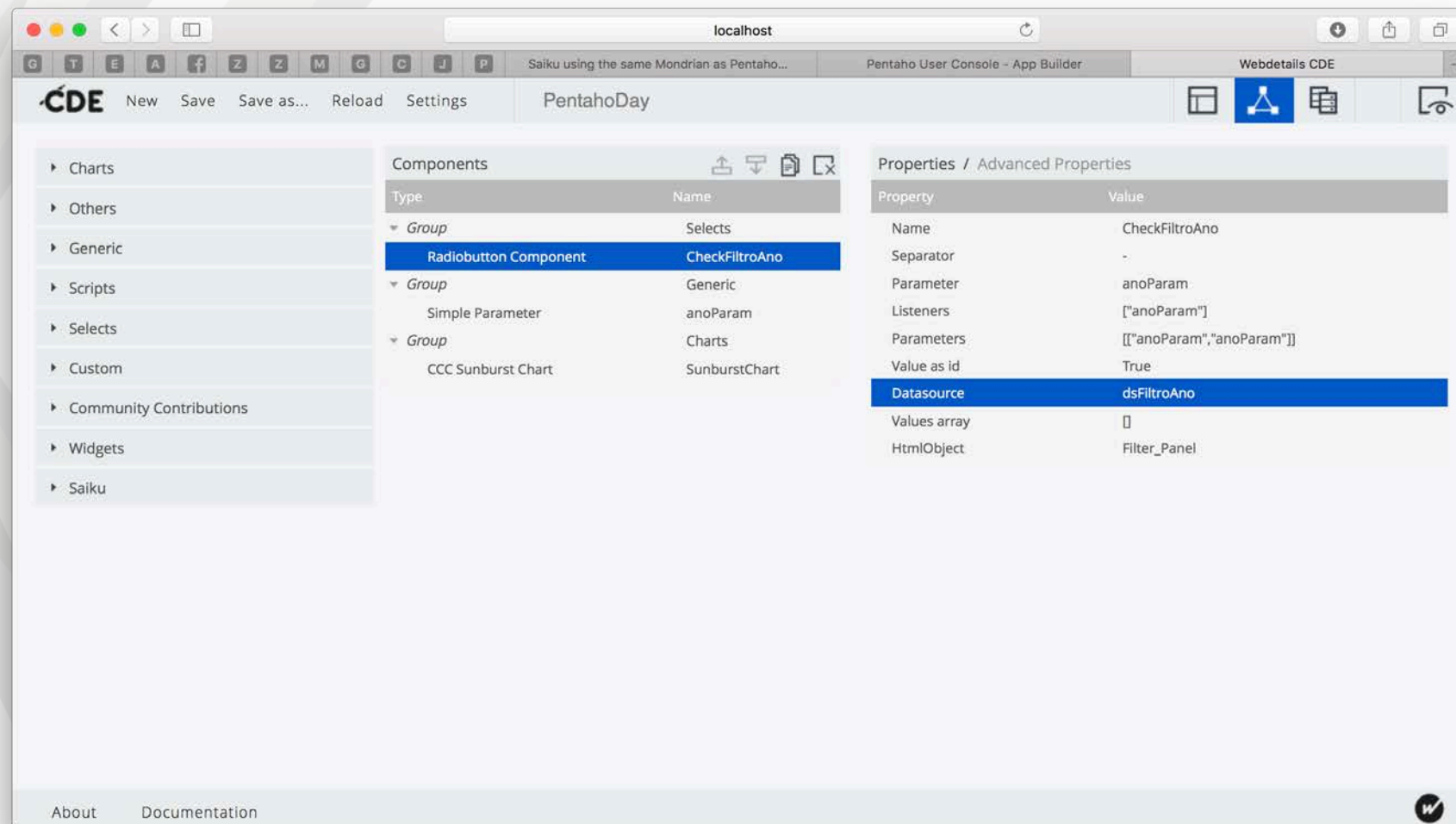
Mondrian Schema SteelWheels

Query

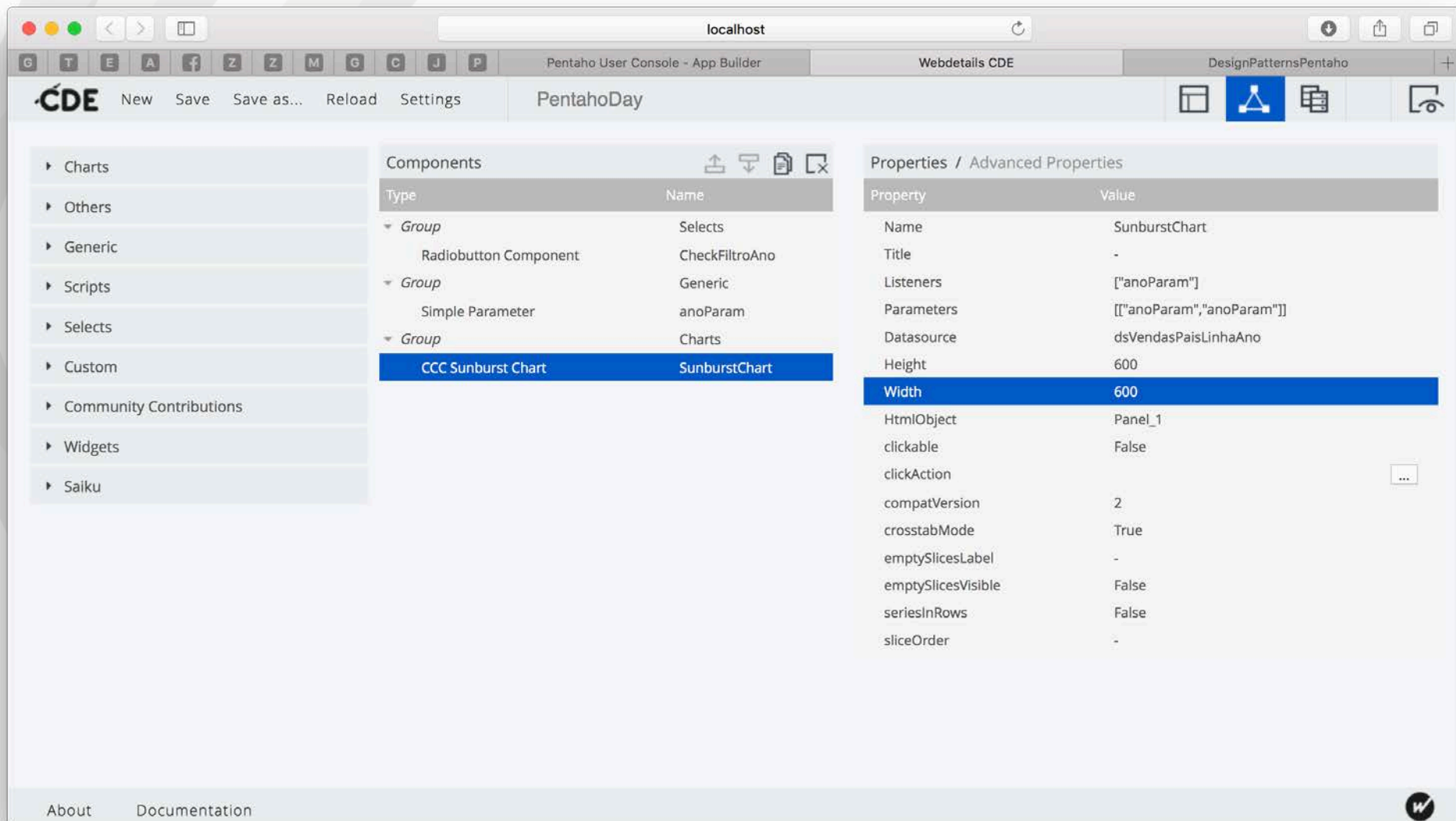
```
with set [~FILTER] as '{[Time].[${anoParam}]}'  
  set [~COLUMNS] as '{[Product].[Line].Members}'  
  set [~ROWS] as '{[Markets].[Country].Members}'  
select NON EMPTY Crossjoin([~COLUMNS], {[Measures].  
[Sales]}) ON COLUMNS,  
  NON EMPTY [~ROWS] ON ROWS  
from [SteelWheelsSales]  
where [~FILTER]
```

Parameters anoParam anoParam

Inclua um componente do tipo Radiobutton Component



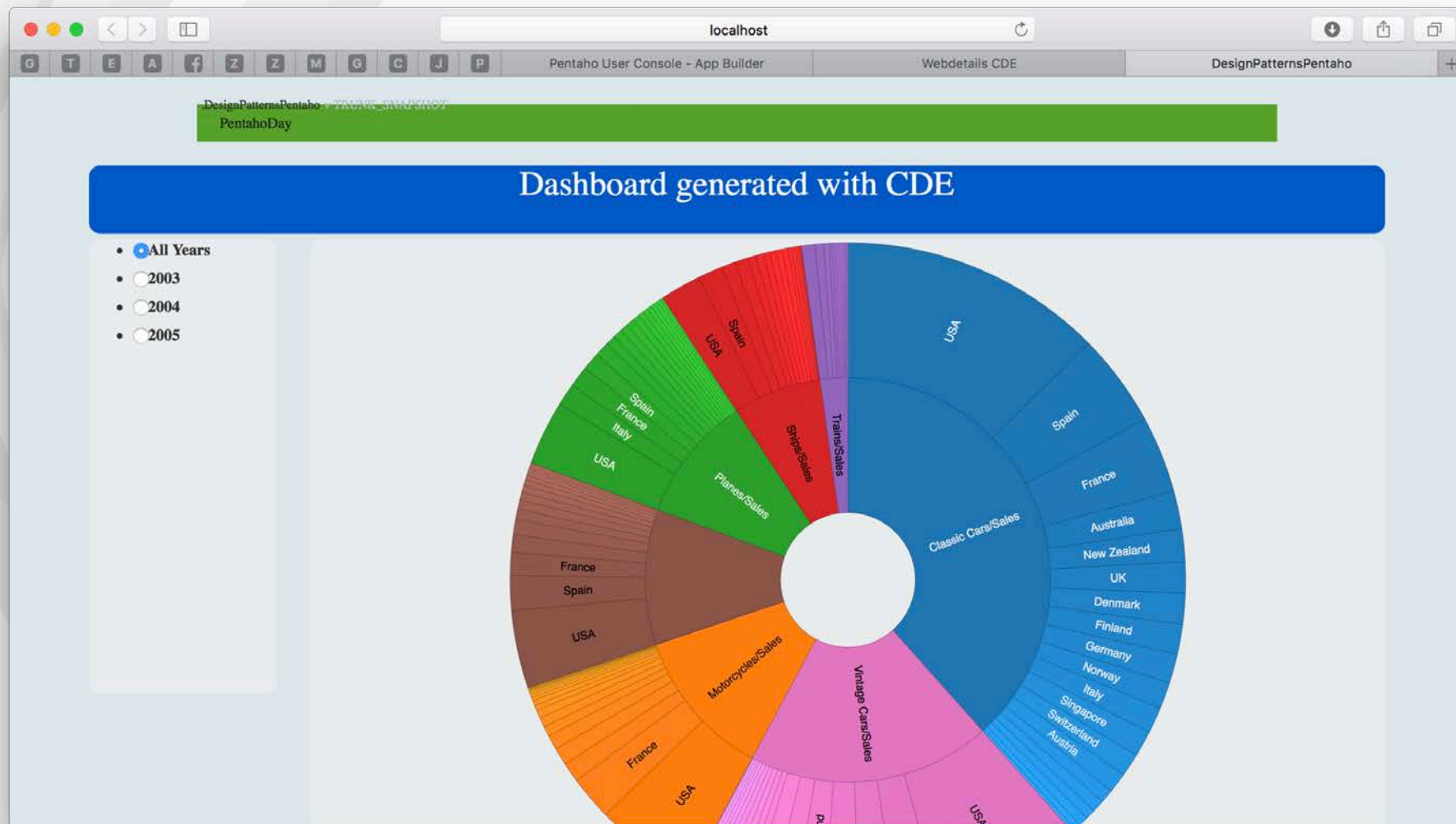
Inclua um componente do tipo CCC Sunburst Chart



The screenshot displays the Pentaho CDE interface for configuring a component. The left sidebar shows a tree view of components, with 'Charts' expanded. The 'Components' panel in the center lists various components, with 'CCC Sunburst Chart' selected. The 'Properties / Advanced Properties' panel on the right shows the configuration for the selected component.

Property	Value
Name	SunburstChart
Title	-
Listeners	["anoParam"]
Parameters	[["anoParam", "anoParam"]]
Datasource	dsVendasPaisLinhaAno
Height	600
Width	600
HtmlObject	Panel_1
clickable	False
clickAction	...
compatVersion	2
crosstabMode	True
emptySlicesLabel	-
emptySlicesVisible	False
seriesInRows	False
sliceOrder	-

Dashboard



Pentaho Design Patterns

1 - Ajuste a Inicialização do Pentaho Server

- **No Linux, edite o arquivo:**

```
./pentaho-server/start-pentaho.sh;
```

ou, no windows

```
.\pentaho-server\start-pentaho.bat
```

- **Servidor Headless**

Se o seu servidor não utiliza interface gráfica, adicione a opção

```
-Djava.awt.headless = true
```

- **Memória**

Edite os parâmetros que limitam o uso de memória pela JVM. O primeiro parâmetro indica a quantidade de memória inicial e o segundo o máximo de memória disponível para o java.

-Djava.awt.headless=true -Xms1024m -Xmx16G -
XX:MaxPermSize=256m

- **Encode**

Adicione/Verifique -> no pentaho 7, finalmente, já vem como default

-Dfile.encoding=utf8

- **Timezone**

-Duser.timezone=America/Sao_Paulo

- **Exemplo de configuração**

```
CATALINA_OPTS="-Djava.awt.headless = true -  
Xms2048m -Xmx16G -XX:MaxPermSize=256m -  
Dsun.rmi.dgc.client.gcInterval=3600000 -  
Dsun.rmi.dgc.server.gcInterval=3600000 -  
Dfile.encoding=utf8 -Duser.timezone=America/  
Sao_Paulo -DDI_HOME=\"$DI_HOME\""
```

2 - Utilize os Logs Pentaho

- Os logs do BI Server são gerados por uma API Java chamada Log4J, do projeto Apache e armazena informações importantes da plataforma na pasta

`./pentaho-server/tomcat/logs.`

Mude o Delimitador para “tab” ou para “;”

Para facilitar a visualização dos dados Mude o Delimitador para “tab” ou para “;”

Tab ajuda a leitura no console e editores de texto. O ponto e vírgula facilita leitura em planilhas de calculo

1. Abra o arquivo

tomcat/webapps/pentaho/WEB-INF/classes/log4j.xml

2. Procure a linha que formata a saída das informações

<param name="ConversionPattern" (linha 38)

3. Troque o separador, de espaço, para “tab” (\t) ou para “;” alterando a string

“%d %-5p [%c] %m%n” para “%d\t%-5p\t[%c]\t %m%n”

4. Salve. Se o BI Server estiver rodando, pare-o, apague os arquivos do log e suba o serviço novamente.

Habilite mais logs somente para testes e debug

1. Para configurar o nível de alerta para os logs do pentaho, ajuste o valor da tag "Threshold" (linha 53) para um dos possíveis valores: ALL, DEBUG, ERROR, FATAL, INFO, OFF, TRACCE ou WARN <https://logging.apache.org/log4j/1.2/apidocs/org/apache/log4j/Level.html>

```
<param name="Threshold" value="DEBUG"/>
```

2. Conforme a necessidade, descomente o intervalo de linhas abaixo, para habilitar os logs do Mondrian, das Consultas MDX e o das Consultas SQL. Em cada categoria pode-se ajustar o nível do log, ajustando a tag <priority value="DEBUG"/>

- mondrian.log: linhas 145 e 166;
- mondrian_mdx.log: linhas 172 e 187;
- mondrian_sql.log: linhas 193 e 208.

Outros arquivos possuem configuração para seus logs

- pentaho-solutions/system/cda/log4j.xml
- pentaho-solutions/system/karaf/etc/org.ops4j.pax.logging.cfg
- pentaho-solutions/system/osgi/log4j.xml

Analise os Logs de Auditoria

`./pentaho-server/pentaho-solutions/system/logs/
audit/PentahoAuditLog.log`

- Quem é o maior usuário em quantidade de acessos?
- Quem é o maior em tempo de execução?
- Quais comandos geraram erros durante a última semana?
- Quais as queries mais demoradas?

Como utilizar estes logs

Logs do Tomcat não são afetados pelas configurações tratadas aqui

- catalina.aaaa-mm-dd.log
- catalina.out log do dia atual

Log do Pentaho

- pentaho.log

Logs do Mondrian

- mondrian.log
- mondrian_sql.log
- mondrian_mdx.log

Logs de Auditoria

- pentaho-solutions/system/logs/audit/PentahoAuditLog.log

Leia os logs e encontre as informações relevantes

- Leia no terminal Linux `tail -f pentaho.log`
- Importe para o Excel ou LibreOffice Calc utilizando os delimitadores configurados
- Abra com um editor de Textos com realce de sintaxe. Procure por Verilog, Python ou alguma sintaxe que facilite a leitura dos diferentes tipos de dados em um log.
- Plugin Pentaho CE Audit, do nosso amigo Caio, se utiliza dessas informações para alimentar um cubo de estatísticas dos usuários. Vale verificar se roda no pentaho 7 e também se evoluiu e funciona com todos os bancos de dados (além do Hsqldb e PostgreSQL)
<https://github.com/it4biz/pentaho-ce-audit/releases/tag/2016.05.19>

Verifique:

- a) Se as queries SQL são coerentes
- b) Se as queries estão usando o índice que você espera
- c) Se o seu database é coerente
- d) Se os metadados do cubo são coerentes
- e) Se as consultas MDX estão coerentes

Query Showplan

SGBDR	Exemplo de Comando	Referência
MySQL	EXPLAIN {Select ...}	https://dev.mysql.com/doc/workbench/en/wb-tutorial-visual-explain-dbt3.html
PostgreSQL	Explain Select xpto from dual;	https://www.postgresql.org/docs/9.5/static/using-explain.html
SQL Server	USE AdventureWorks2012; GO SET SHOWPLAN_TEXT ON; GO SELECT * FROM Production.Product WHERE ProductID = 905; GO SET SHOWPLAN_TEXT OFF;	https://technet.microsoft.com/en-us/library/ms176005(v=sql.105).aspx https://docs.microsoft.com/en-us/sql/t-sql/statements/set-showplan-text-transact-sql
Oracle	EXPLAIN PLAN FOR SELECT last_name FROM employees;	https://docs.oracle.com/cd/B10501_01/server.920/a96533/ex_plan.htm
Sybase	set showplan on go exec proc A go	http://infocenter.sybase.com/help/index.jsp?topic=/com.sybase.infocenter.dc00743.1502/html/queryprocessing/queryprocessing45.htm
Vertica	EXPLAIN { SELECT... INSERT... UPDATE... COPY... DELETE... MERGE... EXPORT... }	https://my.vertica.com/docs/7.1.x/HTML/Content/Authoring/SQLReferenceManual/Statements/EXPLAIN.htm

3 - Apague arquivos desnecessários

- Apague, Mova ou Grave no Banco de Dados periodicamente os arquivos de auditoria, logs e temporários

São eles:

- /pentaho-solutions/system/logs/audit/
- /tomcat/logs
- /tomcat/temp
- Plugin Pentaho Log Manager ajuda a apagar os logs
<http://localhost:8080/pentaho/plugin/logManager/api>

4 - Ajuste os Parâmetros Mondrian

As configurações do Mondrian ficam no arquivo `./pentaho-server/pentaho-solutions/system/mondrian/mondrian.properties`

- Ativar o uso de agregadas:

`mondrian.rolap.aggregates.Use = true`

`mondrian.rolap.aggregates.Read = true`

- Limitar o número de linhas retornadas: `mondrian.result.limit = 10000`
- Limitar a espera pelo fim da consulta: `mondrian.rolap.queryTimeout = 600` (Segundos)
- Não altere: `mondrian.trace.level = 1`

5 - Verifique o cubo com o Pentaho Schema Workbench

- colunas como inteiro, sempre que possível;
- hierarquias em colunas com índice no banco de dados, que pode ser composto (mais de uma coluna);
- habilite o uso do cache em todos os cubos.

6 - Use o Pentaho Aggregation Designer

- Abra o PAD;
- Crie a conexão para seu DW;
- Selecione o arquivo XML com os metadados do Cubo a ser analisado;

Se o cubo tiver restrições por roles, crie uma cópia do xml e apague todas as roles

- Selecione o Cubo desejado, dentre os publicados no xml escolhido;
- Observe a validação do seu banco e dados

Caso necessário, faça os ajustes na estrutura do banco até que passe pela validação do PAD

- Indique o limite:

De tempo, em segundos, para que as queries sejam executadas ou;

Quantidade de tabelas agregadas a serem criadas.

- Analise o resultado e faça os ajustes, caso necessários

Perceba se os filtros mais utilizados pela aplicação (os dashboards) têm uma tabela agregada que os contemple

- Exporte:

O SQL para criar as tabelas agregadas;

O SQL para popular as tabelas agregadas

Os metadados destas tabelas para dentro do xml do cubo

Ficarão indentadas dentro da tabela fato do cubo analisado

- (Re)publique o cubo no pentaho server

Lembre-se de limpar os caches do Mondrian, do CDA e do Saiku (caso não utilize o mesmo mondrian que o Pentaho)

- No job que faz a integração dos dados, logo após popular a tabela fato, inclua dois steps

Truncate table - selecione todas as tabelas agregadas;

SQL Script - inclua as queries de insert nas agregadas.

7 -Habilite no CDE o uso do Cache em seus data-sources

Em cada datasouce do seu dashboard, deixe marcada a opção de Cache.

8 -Gerencie o CDA com o CDA Cache Manager

Permite limpar o cache somente do que for necessário oferecendo todas as urls, oq eu premite automatizar a limpeza dos caches com precisão.

- Tools/ CDA Cache Manager ou
- <http://localhost:8080/pentaho/plugin/cda/api/manageCache>

9 - Use o Utilize o CDA – O Pulo do gato

CDA é chamado por HTTP requests para `$BASE_URL/pentaho/plugin/cda/api/`.

Documentação oficial disponível em: <http://community.pentaho.com/ctools/cda/#section=web-api-reference>

Método suportados:

- `doQuery;`
- `listQueries;`
- `getCdaList;`
- `listParameters;`
- `clearCache;`
- `listDataAccessTypes.`

Exemplo: <http://localhost:8080/pentaho/plugin/cda/api/getCdaList>

Abra o arquivo CDA

- Se for um dashboards normal, fora de um plugin pentaho, é só clicar no .cda e abrir.
- Em nosso caso, que criamos um plugin Sparkl, vamos precisar do método previewQuery do CDA

<http://localhost:8080/pentaho/plugin/cda/api/previewQuery?path=/system/DesignPatternsPentaho/dashboards/PentahoDay.cda>

- Veja as informações sobre cada query;
- Agende execução de queries;
- Utilize iteradores em seus parâmetros quando for conveniente.

Easter Egg

Praticamente um Easter egg, escondido no cda, no campo dos params, no mouse hover é apresentada uma interrogação, que ao ser clicada, mostra esse texto:

Parameter formulas

You can use formulas instead of fixed values for CDA parameters. Here are some useful examples:

`${TODAY()}` → Today's date

`${YESTERDAY()}` → Yesterday's date

`$FOREACH(dataAccess,colIdx,param1=val1,...,paramn=valn)`

FOREACH

Iterate over the values for colIdx in query dataAccess with the given parameters. Queries using this construct return no values, and its main purpose is to provide parameter ranges for cache warming.

Please refer to the Pentaho Reporting Engine documentation for a full description of the syntax. <http://wiki.pentaho.com/display/Reporting/9.+Report+Designer+Formula+Expressions>

- Ex.: \$FOREACH(dsFiltroAno,0)

10 - Debugue seu código no CDE

Adicione na URL do Dashboard, os parâmetros abaixo e seja feliz com o inspector de seu navegador.

“?debug=true”, no final da url, ele nao compacta as coisas. Nao usa cache.

“&debugLevel=5” ajusta o nível do log CDE

<http://localhost:8080/pentaho/plugin/DesignPatternsPentaho/api/pentahoday?debug=true&debugLevel=5>

11 - Conheça o CDC - Community Distributed Cache

Ainda não disponível na versão 7 do Pentaho.
Referências em:

- <http://community.pentaho.com/ctools/cdc/>

e

- <http://pedroalves-bi.blogspot.com.br/2012/05/cdc-community-distributed-cache.html>

12 - Saiku - share Mondrian

Faz com que o Saiku CE utilize o mesmo Mondrian que o Pentaho e, assim, se beneficie de todos os caches criados

Pentaho até versão 5

```
cd /pentaho/server/5.01/biserver-ce/pentaho-  
solutions/system/saiku
```

```
./saiku-shareMondrian.sh -w /pentaho/server/  
5.01/biserver-ce/tomcat/webapps/pentaho
```

Pentaho 6 ou mais recente

In the plugin, if you remove the mondrian jar from the plugin, it'll fall back to using the same jar as the server. You can also follow that shell script to remove the other jars even if its not quite automatic any more, to run using the bi server cache.

Tom

<https://groups.google.com/a/saiku.meteorite.bi/forum/#!topic/user/wFQstt5Cd7o>

Contribuição para a comunidade Pentaho

Artigo completo sobre BI com Pentaho
publicado na SQL Magazine, com dados e
fontes:

[https://br.groups.yahoo.com/neo/groups/
pentahobr/conversations/messages/16965](https://br.groups.yahoo.com/neo/groups/pentahobr/conversations/messages/16965)

Obrigado



e-SETORIAL
Business Analytics

Eduardo Alves de Carvalho
eduardo@e-Setorial.com.br
(48) 3334-2307 | (48) 99953-1260
www.e-Setorial.com.br