IBM

IBM Cognos Forum

Ignite knowledge, ideas, connections

3260 Advanced Transformer Case Studies

Mark Morton

Information Management

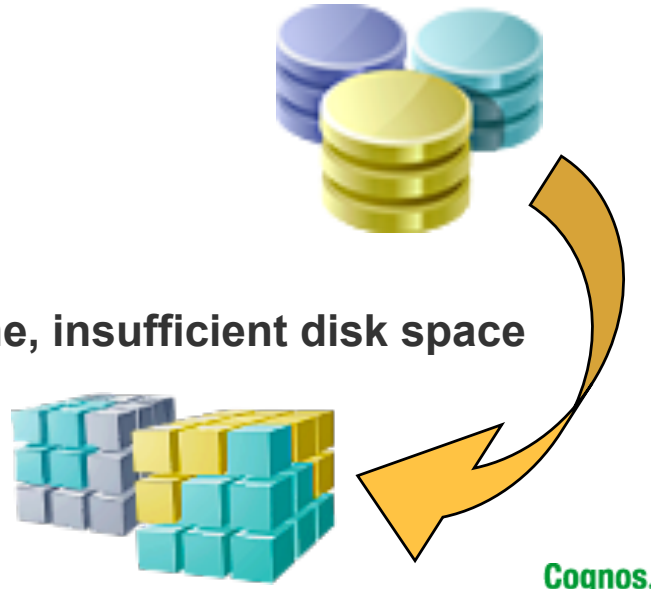
Cognos. software

Agenda

- Cube Build Phases
- Improving Build Performance - Tips
- Using the Transformer Log File for Phase Timing
- Transformer Log Key Words
- Auto-Partitioning
- Exam Time!
  - Scenario #1 - Cube build takes much longer on one server versus another
  - Scenario #2 - Cube build fails at same location every time
  - Scenario #3 - Long cube build with limited data set
  - Scenario #4 - Which pass is most successful at partitioning
- Bonus Round
- Questions

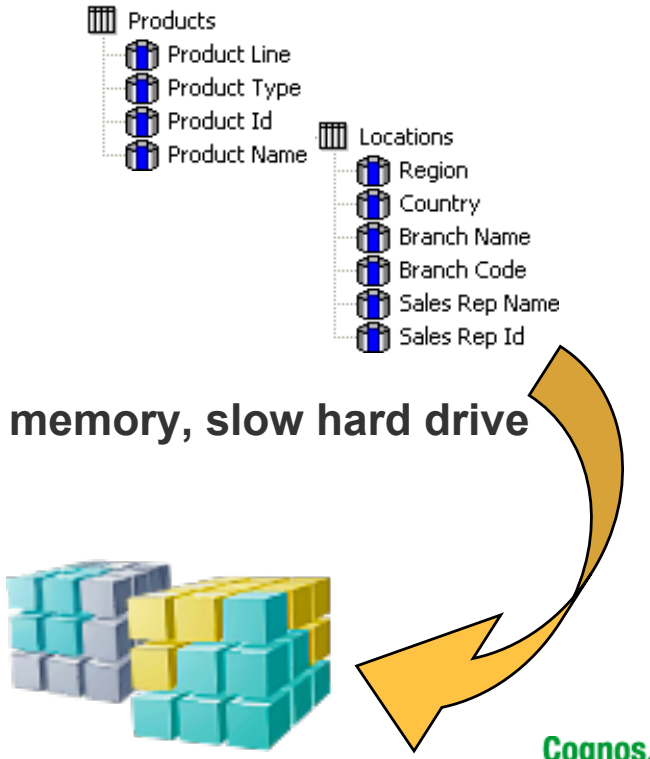
Cube Build Phases – Data Read Phase

- Transformer reads the source data and creates a temporary work file based on the structure of the Transformer model
- Memory use climbs rapidly when categories are being generated during this phase. The more categories in the model, the more memory required
- Reviewing this phase will tell you how long it took Transformer to read and process all of the data sources and how many records were read
- Key Words:
  - INITIALIZING CATEGORIES
  - OPEN DATA SOURCE
  - READ DATA SOURCE
  - MARKING CATEGORIES USED
- Associated Problems:
  - Database connectivity problems, slow read time, insufficient disk space



Cube Build Phases – Metadata Update Phase

- The contents of the temporary work files are compared to the categories in the Transformer model to determine which categories will be put in the PowerCube
- When the list of eligible categories is complete the categories are inserted into the PowerCube
- Key Words:
  - SORTING
  - UPDATE CATEGORY AND PROCESS WORK FILE
  - METADATA
- Associated Problems:
  - Long update times which can be due to lack of memory, slow hard drive and write cache settings



Cube Build Phases – Data Update Phase

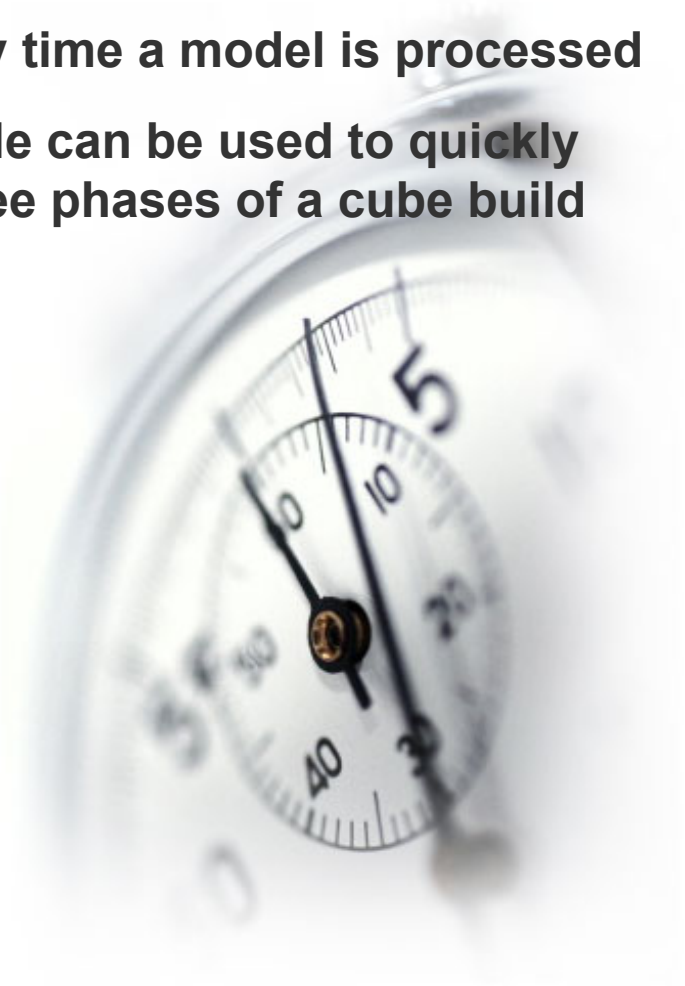
- The data values in the temporary work files are inserted into the PowerCube
- Each record inserted into the cube is a 'data point' that consists of a category reference from each dimension in the model along with the measure values for the intersection of those categories
- Key Words:
  - CUBE UPDATE
  - CUBE COMMIT
- Associated Problems:
  - Long commit times which could be due to the UDA sortmemory setting, slow hard drives, inefficient auto-partition settings and/or low system memory

Improving Build Performance - Tips

- **Processor**
  - Choosing the fastest available processor speed should be considered as the addition of a second CPU can result in a significant reduction in the data read phase
- **Memory**
  - Optimally there should be enough memory on the build computer to handle all running application requests for memory and allow the operating system disk cache to grow as required
- **RAID Level**
  - Level 0 provides the fastest performance
- **System Configuration**
  - The type, speed and configuration of the drive subsystem can cause a significant increase in the time it takes to build a PowerCube
  - 1st Controller: Operating System and applications
  - 2nd Controller: Transformer Data Work directory
  - 3rd Controller: Sort directory and PowerCube directory

Using the Transformer Log File for Phase Timing

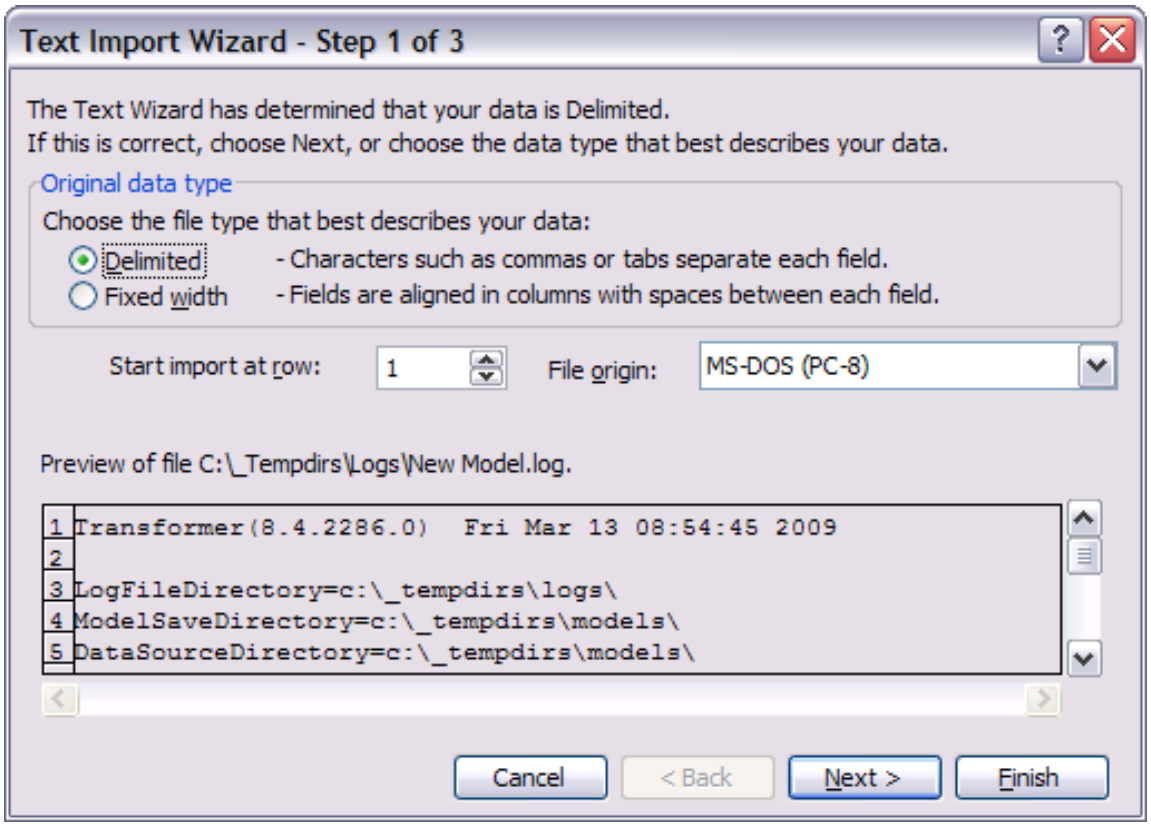
- A Transformer log file is generated every time a model is processed
- Using a spreadsheet program, this log file can be used to quickly understand the length of each of the three phases of a cube build



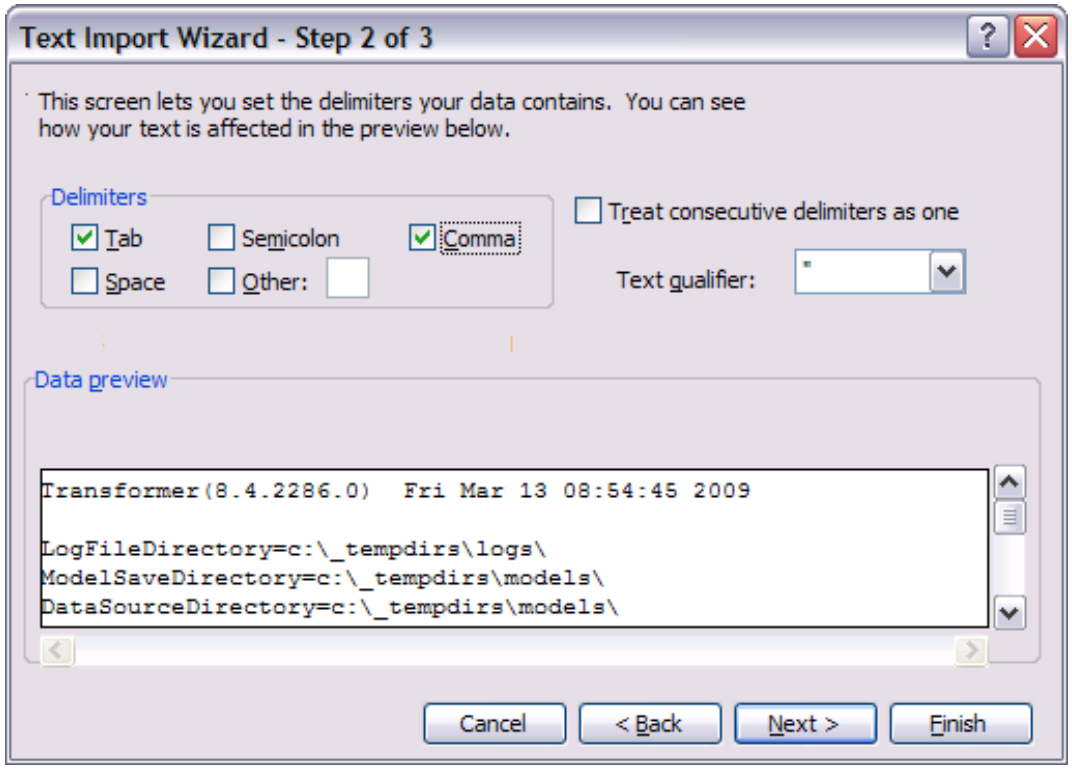
```
Transformer(8.4.2286.0)  Fri Mar 13 08:54:45 2009

LogFileDirectory=c:\_tempdirs\logs\
ModelSaveDirectory=c:\_tempdirs\models\
DataSourceDirectory=c:\_tempdirs\models\
CubeSaveDirectory=c:\_tempdirs\cubes\
DataWorkDirectory=c:\_tempdirs\temp\
ModelWorkDirectory=c:\_tempdirs\modelwork\
MaxTransactionNum=500000
Product locale: English (en)
Run locale: English (Canada) (en-ca)
...
```

Launch a spreadsheet program (Excel was used for this example) and select File Open for files of type **'All Files (*.*)'**.

The dialog for Step 1 of the Text Import Wizard appears:

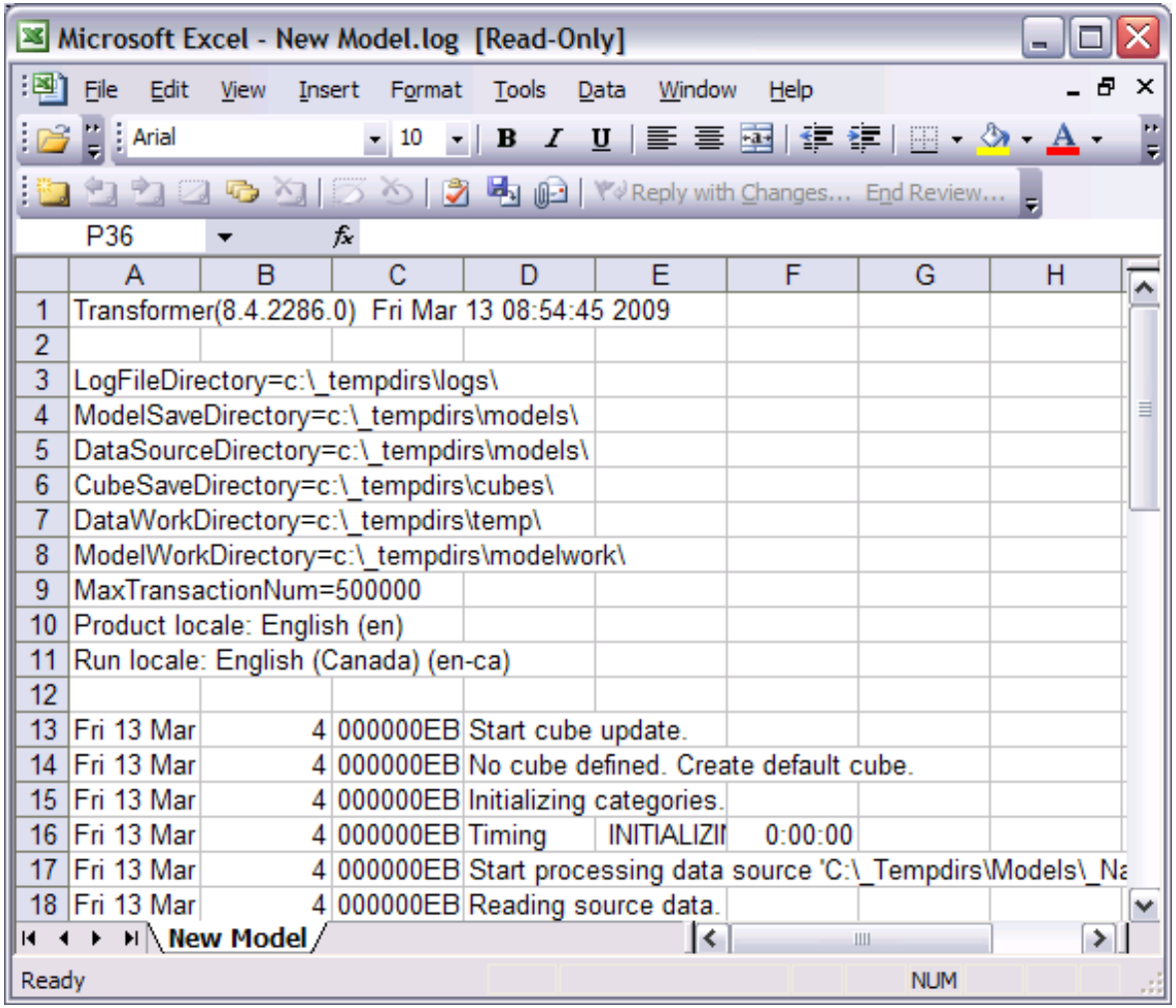


Keep the type as 'Delimited' and select 'Next'. The dialog for Step 2 of the Text Import Wizard appears:

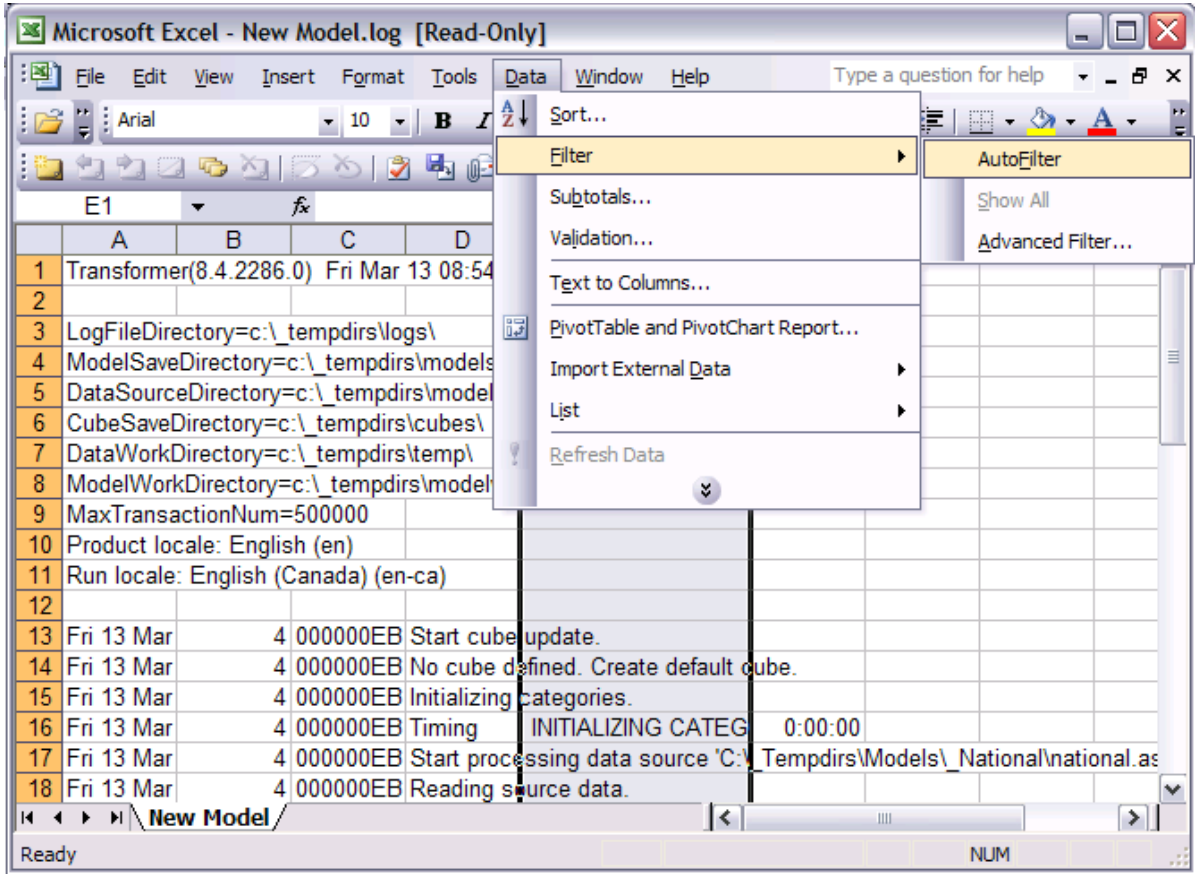


Ensure that 'Tab' and 'Comma' are selected, then click 'Finish'.

The log file is loaded into Excel and appears as follows:



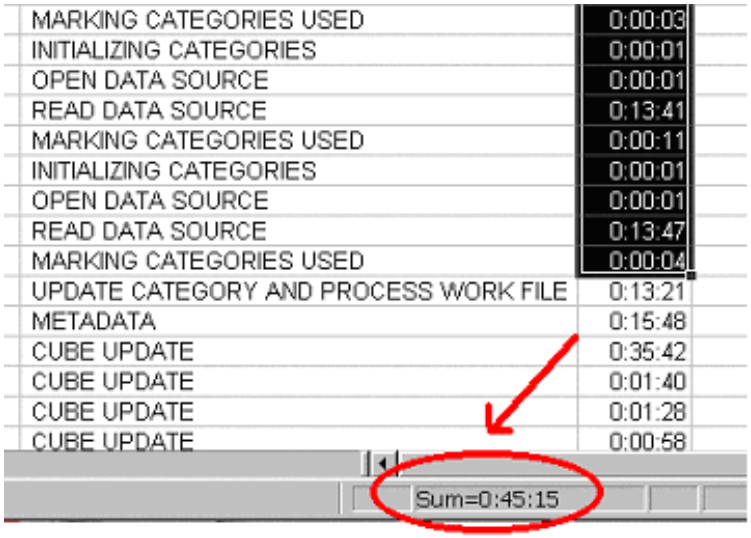
Select the entire ‘E’ column and choose the ‘Data’ menu followed by the ‘Filter’ item and finally select the ‘AutoFilter’ option:



From the drop down list that appears in the ‘E’ column select ‘(NonBlanks)’ or a specific phase such as Read Data Source or Metadata.

The spreadsheet now shows only the lines that contain timing information.

Once the spreadsheet is in this format, select a range of cells in the ‘F’ column and look at the bottom of the Excel window to see the sum of the timing values:



Transformer Log Key Words

> As discussed the following displays the keywords that relate to each of the phases of a PowerCube build:

Data Read

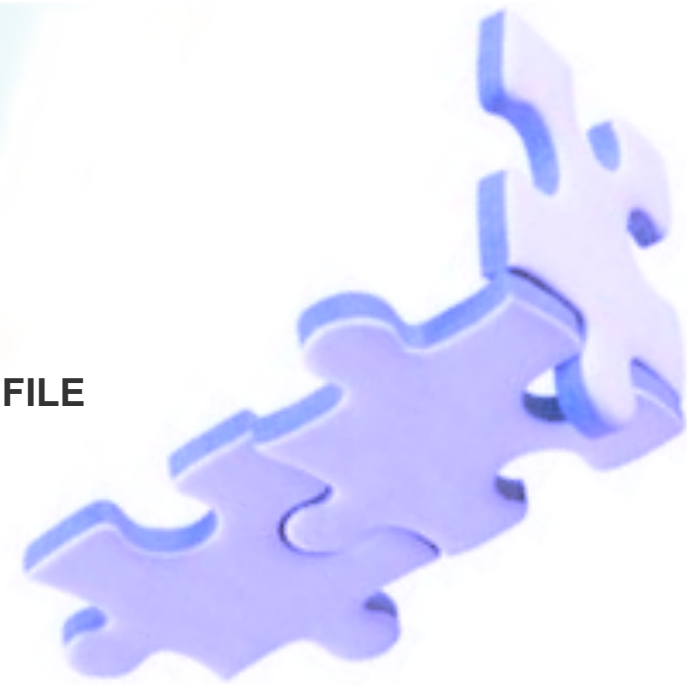
- INITIALIZING CATEGORIES
- OPEN DATA SOURCE
- READ DATA SOURCE
- MARKING CATEGORIES USED

Metadata Update

- SORTING
- UPDATE CATEGORY AND PROCESS WORK FILE
- METADATA

Data Update

- CUBE UPDATE
- CUBE COMMIT



Partitioning

- Partitioning will pre-summarize the data in the PowerCube and group it into several subordinate partitions so retrieval will be significantly faster
- As the number of partitions increase, the longer it will take to create the cube
- The first partition pass will always pick the dimension with the highest number of categories, the next pass will pick the dimension with the second highest number of categories, etc
- The first dimension in the model is always referred to as Dimension “0”

Date (923) (22)	Geography (2567) (2)	Pass 0 Invoice Location (38404) (2)	Pass 1 Ship Location (24412) (2)	Products (13163) (2)
Year (53)	Country (5)	Invoice Country (145)	Ship Country (156)	Performance Center (324)
Quarter (212)	BU Name (10)	Invoice State (821)	Ship State (948)	Product (2538)
Month (636)	Region (227)	Invoice City (10836)	Ship City (23306)	Code (10299)
	Country Name (523)	Invoice Name (26600)		
	Company (1800)			

Partitioning – Pass 0

--- Performing Pass 0 with 22770681 rows and 8708 categories remaining.

Selected Dimension 2 for next pass of partitioning.

Sorting the work file.

Counting category hits.

End sorting 22770681 records.

Start Count and Consolidation with 22770681 rows and 8708 categories remaining.

End Count and Consolidation with 22770681 rows and 8708 categories remaining.

Start Write leaving 8708 categories remaining.

Updating the PowerCube data.

Performing DataBase Commit at record number 2000000…

Performing DataBase Commit at record number 22770682.

End Write leaving 8708 categories remaining..

Pass 0

Date (923) (22)	Geography (2567) (2)	Invoice Location (38404) (2)	Ship Location (24412) (2)	Products (13163) (2)
Year (53)	Country (5)	Invoice Country (145)	Ship Country (156)	Performance Center (324)
Quarter (212)	BU Name (10)	Invoice State (821)	Ship State (948)	Product (2538)
Month (636)	Region (227)	Invoice City (10836)	Ship City (23306)	Code (10299)
	Country Name (523)	Invoice Name (26600)		
	Company (1800)			

Partitioning – Pass 1

--- Performing Pass 1 with 22770681 rows and 8708 categories remaining.

Selected Dimension 3 for next pass of partitioning.

Counting category hits.

End sorting 22770681 records.

Start Count and Consolidation with 22770681 rows and 8708 categories remaining.

End Count and Consolidation with 15522151 rows and 8708 categories remaining.

Start Write leaving 8708 categories remaining.

Updating the PowerCube data.

Performing DataBase Commit at record number 2000000…

Performing DataBase Commit at record number 15522152.

End Write leaving 8708 categories remaining..

Pass 1

Date (923) (22)	Geography (2567) (2)	Invoice Location (38404) (2)	Ship Location (24412) (2)	Products (13163) (2)
Year (53)	Country (5)	Invoice Country (145)	Ship Country (156)	Performance Center (324)
Quarter (212)	BU Name (10)	Invoice State (821)	Ship State (948)	Product (2538)
Month (636)	Region (227)	Invoice City (10836)	Ship City (23306)	Code (10299)
	Country Name (523)	Invoice Name (26600)		
	Company (1800)			

Partitioning – Pass 2

--- Performing Pass 2 with 15522151 rows and 8708 categories remaining.

Selected Dimension 4 for next pass of partitioning.

Counting category hits.

End sorting 15522151 records.

Start Count and Consolidation with 15522151 rows and 8708 categories remaining.

End Count and Consolidation with 14848450 rows and 8708 categories remaining.

Start Write leaving 8708 categories remaining.

Updating the PowerCube data.

Performing DataBase Commit at record number 2000000…

Performing DataBase Commit at record number 14848451.

End Write leaving 8708 categories remaining..

Pass 2

Date (923) (22)	Geography (2567) (2)	Invoice Location (38404) (2)	Ship Location (24412) (2)	Products (13163) (2)
Year (53)	Country (5)	Invoice Country (145)	Ship Country (156)	Performance Center (324)
Quarter (212)	BU Name (10)	Invoice State (821)	Ship State (948)	Product (2538)
Month (636)	Region (227)	Invoice City (10836)	Ship City (23306)	Code (10299)
	Country Name (523)	Invoice Name (26600)		
	Company (1800)			

Partitioning – Pass 3

--- Performing Pass 3 with 14848450 rows and 8708 categories remaining.

Selected Dimension 4 for next pass of partitioning.

Counting category hits.

End sorting 14848450 records.

Start Count and Consolidation with 14848450 rows and 8708 categories remaining.

End Count and Consolidation with 11465545 rows and 8708 categories remaining.

Start Write leaving 8708 categories remaining.

Updating the PowerCube data.

Performing DataBase Commit at record number 2000000…

Performing DataBase Commit at record number 11399138.

End Write leaving 8708 categories remaining..

Pass 3

Date (923) (22)	Geography (2567) (2)	Invoice Location (38404) (2)	Ship Location (24412) (2)	Products (13163) (2)
Year (53)	Country (5)	Invoice Country (145)	Ship Country (156)	Performance Center (324)
Quarter (212)	BU Name (10)	Invoice State (821)	Ship State (948)	Product (2538)
Month (636)	Region (227)	Invoice City (10836)	Ship City (23306)	Code (10299)
	Country Name (523)	Invoice Name (26600)		
	Company (1800)			

Partitioning – Pass 4

--- Performing Pass 4 with 11465545 rows and 8542 categories remaining.

Selected Dimension 1 for next pass of partitioning.

Counting category hits.

End sorting 11465545 records.

Start Count and Consolidation with 11465545 rows and 8708 categories remaining.

End Count and Consolidation with 7312012 rows and 8708 categories remaining.

Start Write leaving 8708 categories remaining.

Updating the PowerCube data.

Performing DataBase Commit at record number 7312013.

End Write leaving 8542 categories remaining..

Pass 4

Date (923) (22)	Geography (2567) (2)	Invoice Location (38404) (2)	Ship Location (24412) (2)	Products (13163) (2)
Year (53)	Country (5)	Invoice Country (145)	Ship Country (156)	Performance Center (324)
Quarter (212)	BU Name (10)	Invoice State (821)	Ship State (948)	Product (2538)
Month (636)	Region (227)	Invoice City (10836)	Ship City (23306)	Code (10299)
	Country Name (523)	Invoice Name (26600)		
	Company (1800)			

Partitioning – Pass 5

--- Performing Pass 5 with 7312012 rows and 8542 categories remaining.

Start Write leaving 8542 categories remaining.

Updating the PowerCube data.

Updating the PowerCube data.

Performing DataBase Commit at record number 2000000.

Performing DataBase Commit at record number 4000000.

Performing DataBase Commit at record number 6000000.

Performing DataBase Commit at record number 7312013.

End Write leaving 8542 categories remaining..

Partitioning – Passes 0 through 5

--- Performing Pass 0 with 22770681 rows and 8708 categories remaining.

Selected Dimension 2 for next pass of partitioning.

--- Performing Pass 1 with 22770681 rows and 8708 categories remaining.

Selected Dimension 3 for next pass of partitioning…

--- Performing Pass 2 with 15522151 rows and 8708 categories remaining.

Selected Dimension 4 for next pass of partitioning…

--- Performing Pass 3 with 14848450 rows and 8708 categories remaining.

Selected Dimension 4 for next pass of partitioning…

--- Performing Pass 4 with 11465545 rows and 8542 categories remaining.

Selected Dimension 1 for next pass of partitioning…

--- Performing Pass 5 with 7312012 rows and 8542 categories remaining.

Auto-Partitioning - Important Points

Pass 0:

Start Count and Consolidation with 22770681 rows and 8708 categories remaining.

End Count and Consolidation with 22770681 rows and 8708 categories remaining.

Pass 2:

Performing Pass 2 with 15522151 rows and 8708 categories remaining.

Selected Dimension 0 for next pass of partitioning…

Start Count and Consolidation with 15522151 rows and 8708 categories remaining.

End Count and Consolidation with 14848450 rows and 8708 categories remaining.

Pass 3:

Performing Pass 3 with 14848450 rows and 8708 categories remaining.

Selected Dimension 0 for next pass of partitioning…

Start Count and Consolidation with 14848450 rows and 8708 categories remaining.

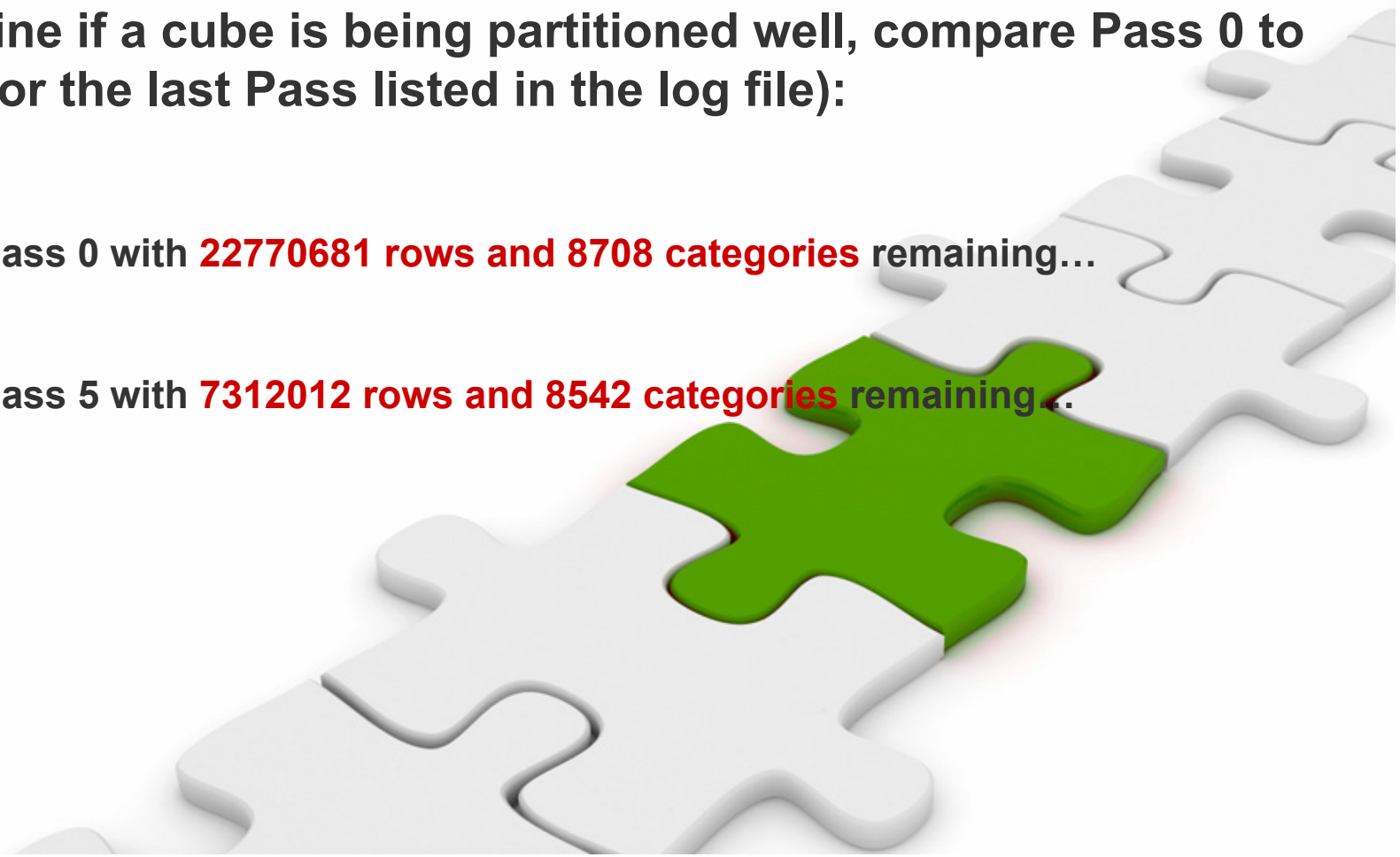
End Count and Consolidation with 11465545 rows and 8708 categories remaining.

Partitioning – Start and End Counts

To determine if a cube is being partitioned well, compare Pass 0 to Pass 5 (or the last Pass listed in the log file):

Performing Pass 0 with 22770681 rows and 8708 categories remaining…

Performing Pass 5 with 7312012 rows and 8542 categories remaining…



Exam Time!!

Cube build takes much longer on one server versus another

- Unix build takes 33 hours
- Windows build takes 100 hours
- Same model
- Same database

Question: Why does it take so much longer on Windows?

Question: Why does it take so much longer on Windows?

Unix

Timing, TOTAL TIME (CREATE CUBE),33:18:07

Windows

Timing, TOTAL TIME (CREATE CUBE),100:39:23

A: Differences in hardware

B: Database performance

C: Multiprocessing wasn't enabled

D: Category count is sufficiently higher

The correct answer is D!

Unix

Performing Pass 0 with 127245055 rows and 42900 categories remaining

Windows

Performing Pass 0 with 66194774 rows and 313485 categories remaining

Comparing the log files showed that the Windows build had over 270 thousand more categories than the Unix build

Cube build fails at same location every time

- Unix build failing during the Data Read Phase
- All previous cube builds work
- Same model
- Same database

Question: Why is the cube build now failing?

Question: Why is the cube build now failing?

(TR0112) There isn't enough memory available

Closing model file /apps/cognos/tmp/pp10626.qyi

Transformer Server exiting - operation failed

REPOS-E-NOMEM, Insufficient memory on machine for operation

Closing model file /apps/cognos/tmp/pp26979.qyi

Transformer Server exiting - operation failed

Question: Why is the cube build now failing?

A: Computer needs a break at the same point every time

B: A limitation has been reached

C: Multiprocessing wasn't enabled

D: Database connectivity problems

The correct answer is B!

Performing Pass 0 with 4057850 rows and 2420919 categories remaining

This model has over **2 million** categories. The current supported limit for the number of categories in a model is 2 million **categories.**

We redesigned model and no longer exceeded the category count limitation

Long cube build with limited data set

- Windows build completing successfully
- PowerCube build taking too long for the number of input records and categories

Question: Why is the cube taking almost 8 hours to complete?

Question: Why is the cube taking almost 8 hours to complete?

The log file confirmed:

- Under 3 million records
- 550 thousand categories

However it was taking over 7.5 hours to complete the cube build…

A: Low system memory

B: Inefficient preference settings

C: Auto-partitioning isn't being utilized

D: Database read is slow

The correct answer is C!

The reason it took so long?

(TR2757) This model contains one or more cubes that use a dimension view in which the primary drilldown is cloaked. Auto-Partitioning is not possible when a primary drilldown is cloaked

We changed the primary drilldown and the cube now built in 40 minutes

Which pass is MOST successful at partitioning?

A:

Start Count and Consolidation with 7444543 rows and 110227 categories remaining.

End Count and Consolidation with 5675377 rows and 110227 categories

B:

Start Count and Consolidation with 5675377 rows and 110227 categories remaining.

End Count and Consolidation with 5675377 rows and 110227 categories remaining.

C:

Start Count and Consolidation with 5675377 rows and 110227 categories remaining.

End Count and Consolidation with 12446 rows and 110227 categories remaining.

BONUS ROUND – Based on partition results from the last slide, what can you do to reduce the PowerCube build time without degrading end user performance?

Selected Dimension 3 for next pass of partitioning…

Start Count and Consolidation with 7444543 rows and 110227 categories remaining.

End Count and Consolidation with 5675377 rows and 110227 categories

Selected Dimension 0 for next pass of partitioning…

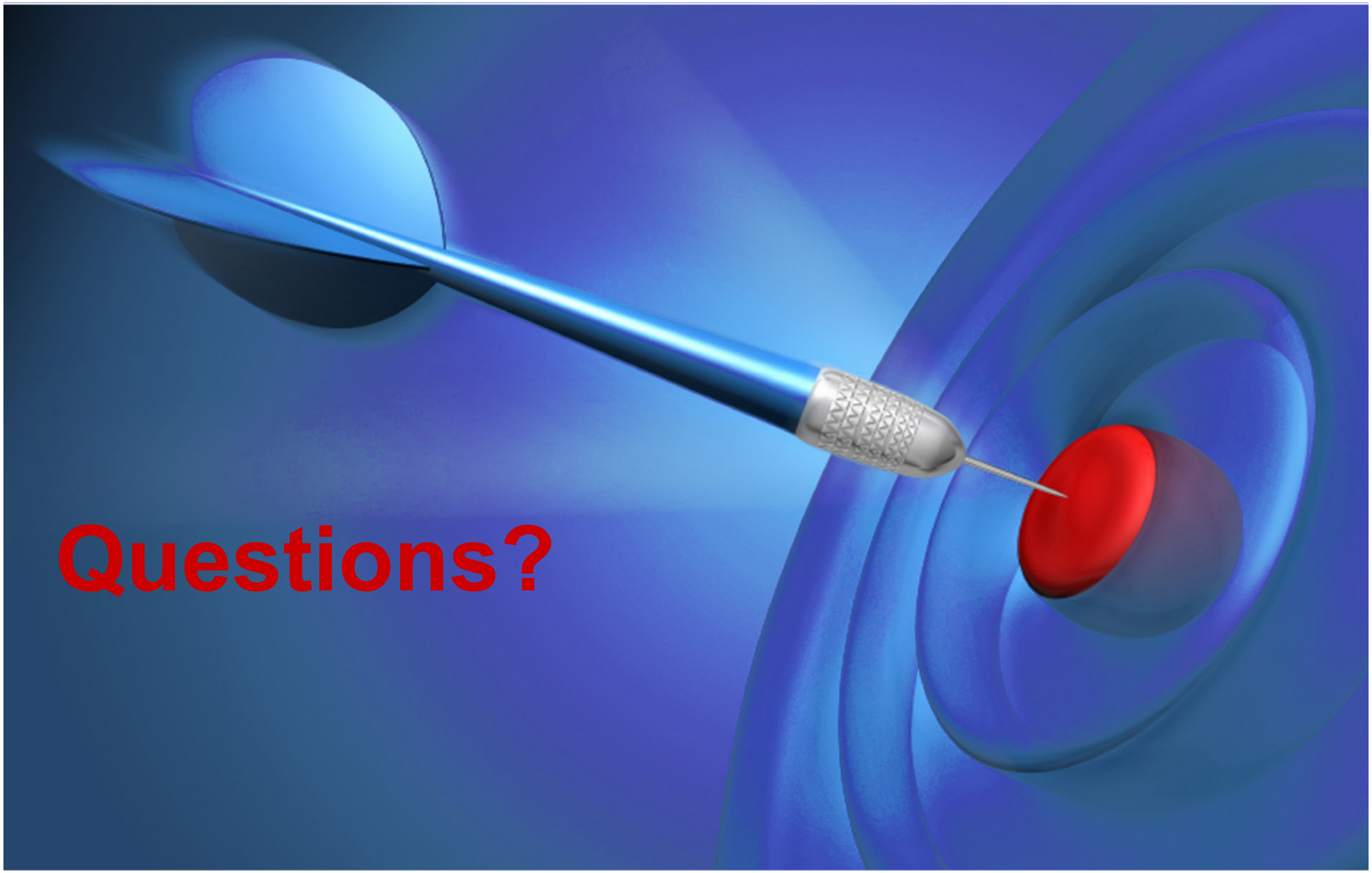
Start Count and Consolidation with 5675377 rows and 110227 categories remaining.

End Count and Consolidation with 5675377 rows and 110227 categories remaining.

Selected Dimension 5 for next pass of partitioning…

Start Count and Consolidation with 5675377 rows and 110227 categories remaining.

End Count and Consolidation with 12446 rows and 110227 categories remaining.


Questions?

Evaluation Forms

SHARE YOUR FEEDBACK AND WIN

Visit IBM Cognos Central to fill out your session evaluations online.

- **Each completed evaluation qualifies you to win one of five $100 daily prizes.**
- **Complete evaluations for every session you attend and qualify to win an additional $500! An overall conference survey will be available at IBM Cognos Central on Friday morning, and will also be emailed to you.**