

Automated Partition Management for Analysis Services Tabular Models

Microsoft BI Technical Article

Writer: Christian Wade

Contributors and Technical Reviewers: Owen Duncan, Marco Russo (SQLBI), Bill Anton (Opifex Solutions), Josh Caplan, Anand Bheemarajaiah, Akshai Mirchandani, Marius Dumitru

Published: January 2017

Applies to: Microsoft SQL Server 2016 Analysis Services, Microsoft Azure Analysis Services

Summary: This whitepaper and associated samples describe partition management automation by using the Tabular Object Model (TOM).

Copyright

This document and associated samples are provided as-is. Information and views expressed in this document, including URL and other Internet Web site references, may change without notice. You bear the risk of using it.

Some examples depicted herein are provided for illustration only and are fictitious. No real association or connection is intended or should be inferred.

This document does not provide you with any legal rights to any intellectual property in any Microsoft product. You may copy and use this document for your internal, reference purposes.

© 2016 Microsoft. All rights reserved

Contents

Introduction	3
Partitioning Strategy & Assumptions.....	4
Rolling-window pattern	4
Partition granularity.....	4
Mixed granularity.....	4
Parallelization.....	4
Online & offline processing.....	4
Non-partitioned table processing.....	4
Table omission	4
Configuration & logging database	5
Date key format & type	5
Getting Started.....	5
Requirements.....	5
AsPartitionProcessing solution	5
AdventureWorks.....	5
SampleClient	5
Configuration & Logging Database	8
Data model.....	8
ModelConfiguration.....	9
TableConfiguration	9
PartitioningConfiguration	10
ProcessingLog.....	10
Sample Configuration	11
Database deployment.....	12
Database connection info	13
Test Different Configurations.....	13
Incremental mode.....	13
Increment partition range.....	14
Offline processing	15
Non-partitioned table processing & table omission.....	16
Merging partitions	18
Mixed-granularity configurations	19

Validation of date ranges for mixed granularity configurations.....	21
Other Considerations.....	22
Cloud architecture	22
Command-line execution.....	23
Syntax.....	23
Arguments.....	23
Examples	24
AsPerfMon	25
Custom logging	25
Fragmentation	26
Locking and blocking.....	26
Auto retry.....	27
Parallelization of incremental processing.....	27
Model deployment	28

Introduction

Analysis Services tabular models can store data in a highly-compressed, in-memory cache for optimized query performance. This provides fast user interactivity over large data sets.

Large datasets normally require table partitioning to accelerate and optimize the data-load process. Partitioning enables incremental loads, increases parallelization, and reduces memory consumption. The [Tabular Object Model](#) (TOM) serves as an API to create and manage partitions. TOM was released with SQL Server 2016 and is discussed in more detail [here](#). Model Compatibility Level 1200 or above is required.

This document describes how to use the [AsPartitionProcessing](#) TOM code sample for automated partition management.

The sample,

- Is intended to be generic and configuration driven.
- Works for both Azure Analysis Services and SQL Server Analysis Services tabular models.
- Can be leveraged in many ways including from an SSIS script task, Azure Functions and others.

Note: Loading data into the in-memory cache is often referred to as *processing*. This terminology is used by this document.

Partitioning Strategy & Assumptions

Rolling-window pattern

AsPartitionProcessing follows the rolling-window pattern, which is common in traditional Analysis Services implementations. The data is kept within a predefined date range and incremented as necessary. This maintains memory usage within a predictable range over time.

Partition granularity

Yearly, monthly and daily partition granularities can be configured. Choice of granularity is influenced by various factors including how much data is required to be incrementally refreshed and how much processing time is acceptable. For example, if only the last 3 days need to be refreshed daily, it may be beneficial to use daily granularity.

Mixed granularity

Mixed granularity for a table can also be configured for scenarios such as near-real time refresh at low grain coupled with historical, static partitions at higher granularity. This results in fewer partitions for a table, but also increases management overhead to ensure partition ranges are defined correctly. Unless there are hundreds of partitions or more, there is normally no significant query-performance penalty resulting from keeping the partitions at the lowest grain.

Parallelization

Initial setup processing is sequential. Incremental processing can be performed in parallel.

Initial setup will create and process the partitions for the first time based on the configuration. This is performed one partition at a time to limit memory consumption (data is not fully compressed during processing). For a large data set, the initial load may typically take a few hours depending on factors such as the query performance of the source system.

Incremental processing is submitted as a parallelized operation for all tables and partitions within a model.

Online & offline processing

Incremental processing can be performed as an online operation, or offline for less memory usage; it is configuration driven. Online incremental processing requires a copy of the data to be maintained in memory for queries until the new data is ready, and then switches to the new data. When processing multiple tables, keeping the model online can be less efficient because it often requires recalculation of the same calculated columns, relationships and indexes multiple times. Offline processing has the benefit of performing this recalculation just once at the end of the processing window.

Non-partitioned table processing

The sample can be configured to process non-partitioned tables in addition to partitioned ones. This avoids having to set up a separate process to refresh non-partitioned tables.

Table omission

It is possible to configure that some tables in the model are not refreshed at all during normal incremental processing. Tables that may not require frequent processing often include the date dimension, categorical dimensions, and facts that may be defined annually such as budget.

Configuration & logging database

Traditional Analysis Services implementations that require partitioning often use a configuration and logging database. AsPartitionProcessing is meant to work in this way, although this is optional. It can be set up to log messages to other targets. This enables easy partition configuration, and diagnosis of issues resulting from automated processing operations.

Date key format & type

Date keys in source tables used for partitioning can be either integers formatted as yyyyymmdd, or based on one of the date datatype.

Getting Started

Requirements

Before you get started, you'll need these tools:

SQL Server 2017 - Install the database engine and SSAS in tabular mode. You can download and install CTP from here [here](#).

SQL Server Data Tools – Download and install the latest version [here](#).

SQL Server Management Studio - Download and install the latest version [here](#).

Visual Studio 2017 – Download and install the free Community Edition [here](#).

AsPartitionProcessing solution

Get the AsPartitionProcessing solution [here](#).

1. Open the solution in Visual Studio and build the project. The hint path for the client library DLLs is the following (assuming installation is on C:\ drive):
C:\Program Files (x86)\Microsoft SQL Server\130\SDK\Assemblies
2. Ensure AsPartitionProcessing.SampleClient is set as the startup project.

AdventureWorks

The quickest way to understand the code sample is to run it on the AdventureWorksDW sample database. The backup file, AdventureWorksDW.bak, is included in the solution.

The tabular project, AsPartitionProcessing.AdventureWorks, is also provided in the solution. It should be used instead of the version from CodePlex because partitioning has been removed from the Internet Sales and Reseller Sales tables. Instead, these tables each have a single partition with the same name as the table, which is the default when you create a new table in SSDT. This partition acts as the *template partition* used by the AsPartitionProcessing sample.

Deploy and process the AdventureWorks tabular model.

SampleClient

AsPartitionProcessing.SampleClient is a console application with a reference to the AsPartitionProcessing class library. It can easily be converted to work for customer projects.

Alternatively, it provides sample client code to execute from an SSIS package, Azure Function, or other mechanism.

In Program.cs, the `_executionMode` variable is set to `InitializeInline`.

```
private static ExecutionMode _executionMode = ExecutionMode.InitializeInline;
```

This ensures the `InitializeAdventureWorksInline` method will be executed to initialize parameters. This method contains inline definition of the partitioning configuration. It shows that Internet Sales is partitioned by the `OrderDateKey` column, which is an integer formatted as `yyyymmdd`. Reseller Sales is partitioned by the `OrderDate` column, which is of type `datetime`. Both formats are supported.

Internet Sales is partitioned by month and holds 12 months of data. Reseller Sales is partitioned by year and holds 3 years of data.

Execute the `AsPartitionProcessing.SampleClient` project. The console output should be displayed like this:

```
file:///C:/Users/wadecb/Source/Repos/AsPartitionProcessing/AsPartitionProces...
Server: localhost
Database: AdventureWorks

Rolling-window partitioning for table Internet Sales
-----

=>Table not yet partitioned

=>New partition range (Monthly):
  MIN partition: 2012-01
  MAX partition: 2012-12
  Partition count: 12

=>Actions & progress:
  Create new partition      2012-01
  Sequentially process      2012-01 /DataOnly
  Create new partition      2012-02
  Sequentially process      2012-02 /DataOnly
  Create new partition      2012-03
  Sequentially process      2012-03 /DataOnly
  Create new partition      2012-04
  Sequentially process      2012-04 /DataOnly
  Create new partition      2012-05
  Sequentially process      2012-05 /DataOnly
  Create new partition      2012-06
  Sequentially process      2012-06 /DataOnly
  Create new partition      2012-07
  Sequentially process      2012-07 /DataOnly
  Create new partition      2012-08
  Sequentially process      2012-08 /DataOnly
  Create new partition      2012-09
  Sequentially process      2012-09 /DataOnly
  Create new partition      2012-10
  Sequentially process      2012-10 /DataOnly
  Create new partition      2012-11
  Sequentially process      2012-11 /DataOnly
  Create new partition      2012-12
  Sequentially process      2012-12 /DataOnly

Rolling-window partitioning for table Reseller Sales
-----

=>Table not yet partitioned

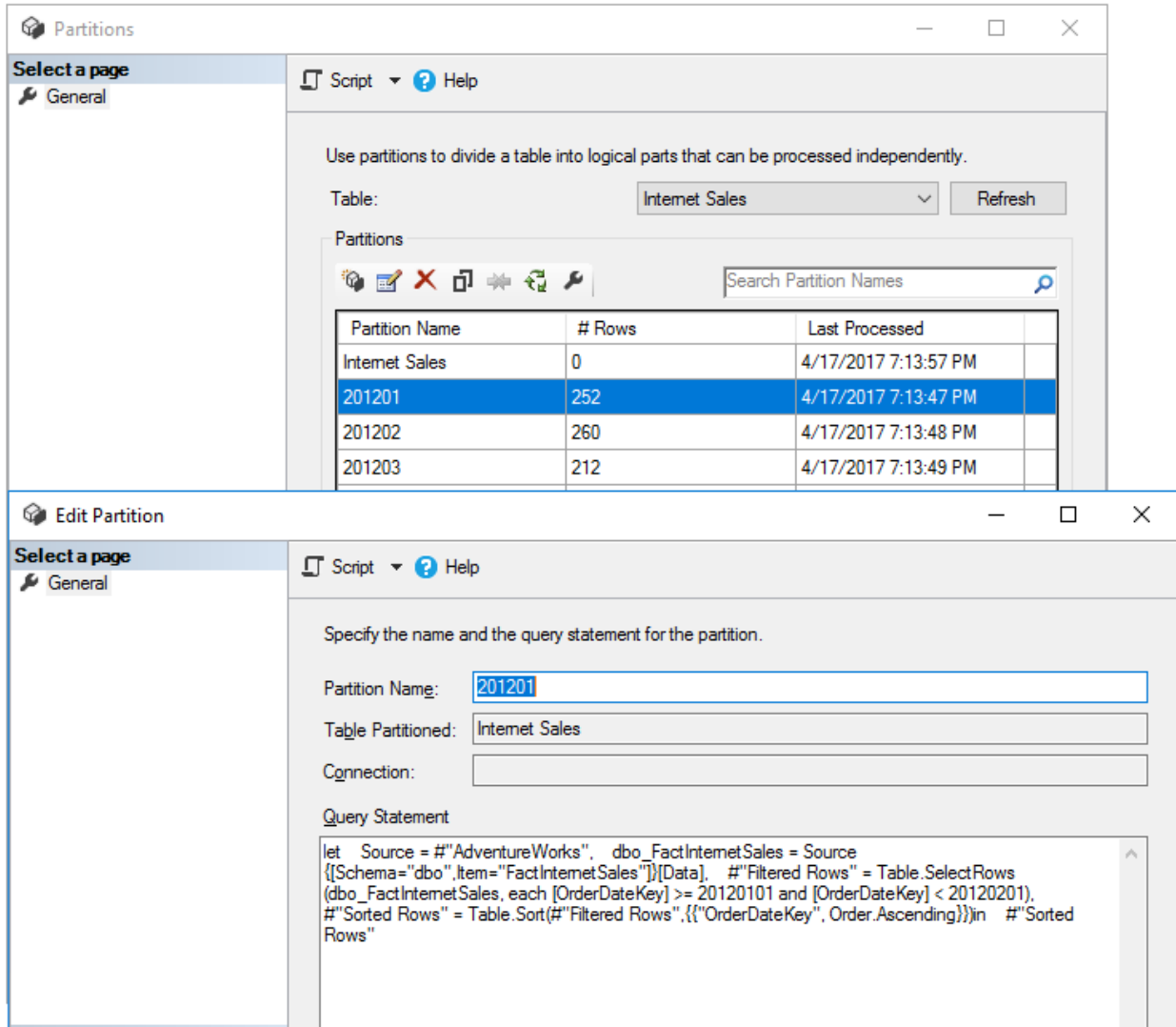
=>New partition range (Yearly):
  MIN partition: 2010
  MAX partition: 2012
  Partition count: 3

=>Actions & progress:
  Create new partition      2010
  Sequentially process      2010 /DataOnly
  Create new partition      2011
  Sequentially process      2011 /DataOnly
  Create new partition      2012
  Sequentially process      2012 /DataOnly

Final operations
-----
  Save changes ...
  Recalc model to bring back online ...

Finish: 09:03:42 AM
Press any key to exit.
```

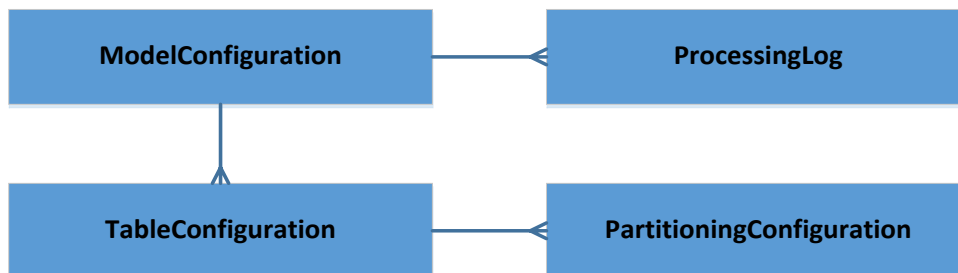
Use SSMS to inspect the partitions created. Note that the date ranges have been specified for each M expression.



Configuration & Logging Database

The AsPartitionProcessing.ConfigurationLogging project is a SQL Server Database project containing the necessary database schema. AsPartitionProcessing contains the methods for reading and writing to the database.

Data model



ModelConfiguration

Configuration information for an AS tabular model:

Column	Description
ModelConfigurationID	Primary key.
AnalysisServicesServer	Name of the Analysis Services instance. Can be SSAS or an Azure AS URL.
AnalysisServicesDatabase	Name of the Analysis Services database.
InitialSetUp	True for initial set up to create partitions and process them sequentially. False for incremental processing. See Partitioning Strategy & Assumptions section above for more information.
IncrementalOnline	When initialSetUp=false, determines if processing is performed as an online operation, which can require more memory, but allows users to query the model during processing. True to keep the model online (process Full). See Partitioning Strategy & Assumptions section above for more information.
IntegratedAuth	Should always be set to true for SSAS implementations that will run under the current process account. For Azure AS, use this option if the current Windows process account is synchronized with Azure AD.
UserName	Only applies when integratedAuth=false. Can be used for Azure AD UPNs to connect to Azure AS.
Password	Only applies when integratedAuth=false. Can be used for Azure AD UPNs to connect to Azure AS.
MaxParallelism	Sets the maximum number of threads on which to run processing commands in parallel. -1 will not set the value.
CommitTimeout	Set to override of CommitTimeout server property value for the connection. -1 will not override; the server value will be used.
RetryAttempts	Number of times a retry of the processing operation will be performed if an error occurs. Use for near-real time scenarios and environments with network reliability issues.
RetryWaitTimeSeconds	Number of seconds to wait before a retry attempt.

TableConfiguration

Configuration information for a table within an AS tabular model:

Column	Description
TableConfigurationID	Primary key.
ModelConfigurationID	Foreign key to ModelConfiguration table.

AnalysisServicesTable	Name of the partitioned table in the tabular model.
DoNotProcess	Set to true to exclude the table from processing. This can be used to dynamically include/exclude tables. For example, near-realtime processing during the day requires only a few tables to be processed; overnight processing may process all tables.

PartitioningConfiguration

Configuration information for partitioning of a table within an AS tabular model.:

Column	Description
PartitioningConfigurationID	Primary key.
TableConfigurationID	Foreign key to TableConfiguration table.
Granularity	Partition granularity, which can be Yearly, Monthly or Daily. Daily = 0, Monthly = 1, Yearly = 2
NumberOfPartitionsFull	Count of all partitions in the rolling window. For example, a rolling window of 10 years partitioned by month would require 120 partitions.
NumberOfPartitionsForIncrementalProcess	Count of <i>hot partitions</i> where the data can change. For example, it may be necessary to refresh the most recent 3 months of data every day. This only applies to the most recent partitions.
MaxDatelsNow	Assumes maximum date to be accounted for in the partitioned table is the date that the code sample is run. Typically, data is loaded up to previous day, or the current day. If the data fits this profile, this setting allows running the code sample every day without having to update the MaxDate value in the configuration and logging database.
MaxDate	If MaxDatelsNow=false, the maximum date that needs to be accounted for in the partitioning configuration.
IntegerDateKey	Assumes date keys in the source database are integers of the format yyyyymmdd. If false assumes dates.
TemplateSourceQuery	Template M expression for M partitions, or source-database native query used for query partitions. Requires placeholders for date key parameterization.

ProcessingLog

Log of partitioning execution:

Column	Description
ProcessingLogID	Primary key.
ModelConfigurationID	Foreign key to ModelConfiguration table.

ExecutionID	GUID generated for the execution run.
LogDateTime	Date and time the message was logged.
Message	The log message.

Sample Configuration

The AsPartitionProcessing.ConfigurationLogging project contains the SampleConfiguration.sql script to initialize the configuration for AdventureWorks. It can be modified for use in customer implementations. The script is executed automatically upon publishing the database.

```

INSERT INTO [dbo].[ModelConfiguration]
VALUES(
    1                --[ModelConfigurationID]
    , 'localhost'    --[AnalysisServicesServer]
    , 'AdventureWorks' --[AnalysisServicesDatabase]
    , 1              --[InitialSetUp]
    , 1              --[IncrementalOnline]
    , 1              --[IntegratedAuth]
    , -1             --[MaxParallelism]
    , -1             --[CommitTimeout]
    , 0              --[RetryAttempts]
    , 0              --[RetryWaitTimeSeconds]
);

INSERT INTO [dbo].[TableConfiguration]
VALUES(
    1                --[TableConfigurationID]
    , 1              --[ModelConfigurationID]
    , 'Internet Sales' --[AnalysisServicesTable]
    , 0              --[DoNotProcess]
),
(
    2                --[TableConfigurationID]
    , 1              --[ModelConfigurationID]
    , 'Reseller Sales' --[AnalysisServicesTable]
    , 0              --[DoNotProcess]
);

INSERT INTO [dbo].[PartitioningConfiguration]
VALUES(
    1                --[PartitioningConfigurationID]
    , 1              --[TableConfigurationID]
    , 1              --[Granularity] 1=Monthly
    , 12             --[NumberOfPartitionsFull]
    , 3              --[NumberOfPartitionsForIncrementalProcess]
    , 0              --[MaxDateIsNow]
    , '2012-12-01'  --[MaxDate]
    , 1              --[IntegerDateKey]
    ,
    'let
    Source = #"AdventureWorks",
    dbo_FactInternetSales = Source{[Schema="dbo",Item="FactInternetSales"]}[Data],
    #"Filtered Rows" = Table.SelectRows(dbo_FactInternetSales, each [OrderDateKey] >=
    {0} and [OrderDateKey] < {1}),
    #"Sorted Rows" = Table.Sort(#"Filtered Rows",{ "OrderDateKey", Order.Ascending})
in

```

```

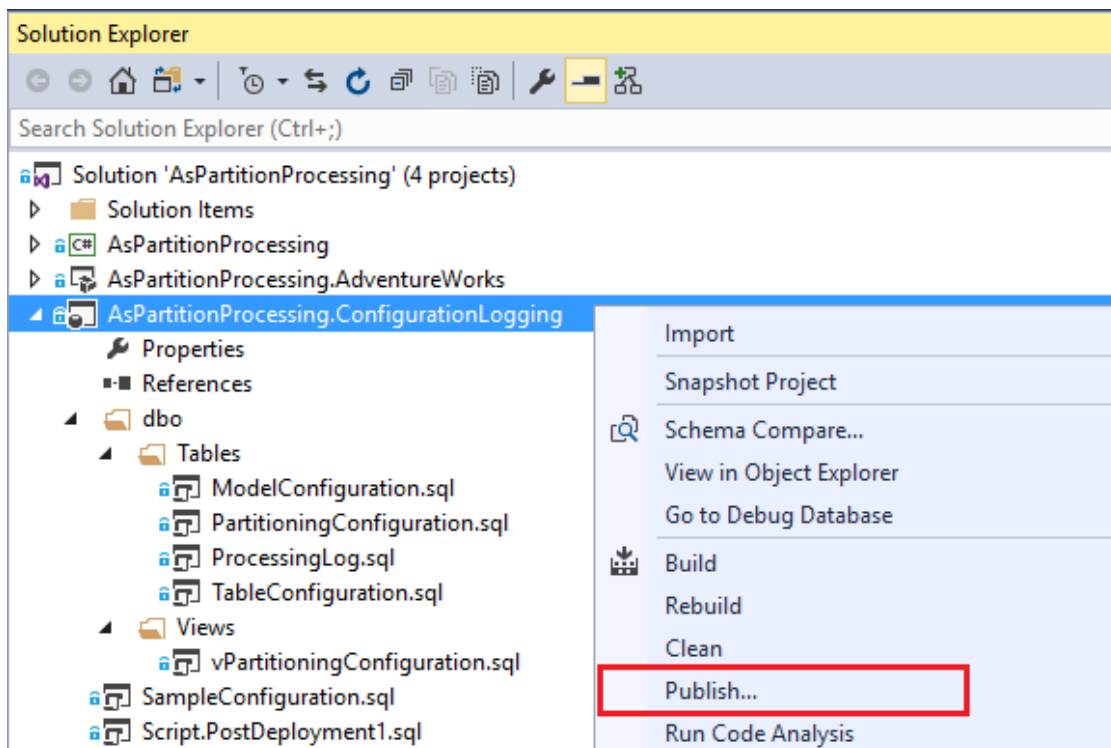
        #"Sorted Rows"'          --[TemplateSourceQuery]
    ),
    (
        2                        --[PartitioningConfigurationID]
        ,2                        --[TableConfigurationID]
        ,2                        --[Granularity] 2=Yearly
        ,3                        --[NumberOfPartitionsFull]
        ,1                        --[NumberOfPartitionsForIncrementalProcess]
        ,0                        --[MaxDateIsNow]
        , '2012-12-01'          --[MaxDate]
        ,0                        --[IntegerDateKey]
    )
'let
    Source = #"AdventureWorks",
    dbo_FactResellerSales = Source{[Schema="dbo",Item="FactResellerSales"]}[Data],
    #"Filtered Rows" = Table.SelectRows(dbo_FactResellerSales, each [OrderDate] >= {0}
and [OrderDate] < {1}),
    #"Sorted Rows" = Table.Sort(#"Filtered Rows",{ "OrderDate", Order.Ascending})
in
    #"Sorted Rows"'          --[TemplateSourceQuery]
);

```

Database deployment

To publish the database, right click on the AsPartitionProcessing.ConfigurationLogging project and select Publish. To match the app.config settings below, the database name should be AsPartitionProcessing.

One of the advantages of using a SQL Server Database project is that new versions can be easily schema compared with an existing version that is already set up with customer data.



Database connection info

Connection information to the configuration and logging database can be set in App.config in the userSettings section.

```
<userSettings>
  <AsPartitionProcessing.SampleClient.Settings>
    <setting name="ConfigServer" serializeAs="String">
      <value>localhost</value>
    </setting>
    <setting name="ConfigDatabase" serializeAs="String">
      <value>AsPartitionProcessing</value>
    </setting>
    <setting name="ConfigDatabaseIntegratedAuth" serializeAs="String">
      <value>True</value>
    </setting>
  </AsPartitionProcessing.SampleClient.Settings>
</userSettings>
```

Test Different Configurations

In this section, we will update the configuration, execute the sample, and view the log messages.

In Program.cs, change the `_executionMode` variable to be assigned `InitializeFromDatabase`.

```
private static ExecutionMode _executionMode = ExecutionMode.InitializeFromDatabase;
```

Execute the SampleClient application. If the SampleClient application was previously run from the Getting Started section, messages will show the partitions already exist and are processed.

In addition to the console output, the `usp_LastProcessingLogs` stored procedure in the configuration and logging database can be used to show the execution results.

```
EXEC [dbo].[usp_LastProcessingLogs]
```

The `usp_LastProcessingLogs` stored procedure has 2 optional parameters; one to specify the number of executions, and another to display only error messages. The following example shows error messages for the last 3 executions.

```
EXEC [dbo].[usp_LastProcessingLogs] @ExecutionCount=3, @ErrorsOnly=1
```

Incremental mode

Execute the following UPDATE statement to switch to incremental mode:

```
UPDATE [dbo].[ModelConfiguration] SET [InitialSetUp] = 0
```

Execute SampleClient application, and the log query. The following results should be shown. Only the specified number of most recent partitions are processed as an online operation.

```
Start: 12:41:02 PM
Server: localhost
Database: AdventureWorks

Rolling-window partitioning for table Internet Sales
-----
```

```

=>Current partition range (Monthly):
  MIN partition: 2012-01
  MAX partition: 2012-12
  Partition count: 12

=>New partition range (Monthly):
  MIN partition: 2012-01
  MAX partition: 2012-12
  Partition count: 12

=>Actions & progress:
  Parallel process partition 2012-10 /Full
  Parallel process partition 2012-11 /Full
  Parallel process partition 2012-12 /Full

Rolling-window partitioning for table Reseller Sales
-----

=>Current partition range (Yearly):
  MIN partition: 2010
  MAX partition: 2012
  Partition count: 3

=>New partition range (Yearly):
  MIN partition: 2010
  MAX partition: 2012
  Partition count: 3

=>Actions & progress:
  Parallel process partition 2012 /Full

Final operations
-----
  Save changes ...

Finish: 12:41:07 PM

```

Increment partition range

Execute the following UPDATE statement to increment the partition range by one period.

```
UPDATE [dbo].[PartitioningConfiguration] SET [MaxDate] = '2013-01-01'
```

Execute SampleClient application, and the log query. The following results should be shown. The oldest partition is removed from both tables, a new one is added and the most recent partitions are processed.

```

Start: 12:47:26 PM
Server: localhost
Database: AdventureWorks

Rolling-window partitioning for table Internet Sales
-----

=>Current partition range (Monthly):
  MIN partition: 2012-01
  MAX partition: 2012-12
  Partition count: 12

```

```

=>New partition range (Monthly):
  MIN partition: 2012-02
  MAX partition: 2013-01
  Partition count: 12

=>Actions & progress:
  Remove old partition 2012-01
  Parallel process partition 2012-11 /Full
  Parallel process partition 2012-12 /Full
  Create new partition 2013-01
  Parallel process partition 2013-01 /Full

Rolling-window partitioning for table Reseller Sales
-----

=>Current partition range (Yearly):
  MIN partition: 2010
  MAX partition: 2012
  Partition count: 3

=>New partition range (Yearly):
  MIN partition: 2011
  MAX partition: 2013
  Partition count: 3

=>Actions & progress:
  Remove old partition 2010
  Create new partition 2013
  Parallel process partition 2013 /Full

Final operations
-----
  Save changes ...

Finish: 12:47:34 PM

```

Offline processing

Execute the following UPDATE statement to perform offline processing and potentially use less memory.

```
UPDATE [dbo].[ModelConfiguration] SET [IncrementalOnline] = 0
```

Execute SampleClient application, and the log query. The following results should be shown. The partitions are processed using RefreshType of DataOnly, and a Calculate operation is performed to bring the model back online. The model may or may not be online during this time. These operations correspond to the processing actions available in SSMS and are documented [here](#).

```

Start: 12:53:19 PM
Server: localhost
Database: AdventureWorks

Rolling-window partitioning for table Internet Sales
-----

=>Current partition range (Monthly):
  MIN partition: 2012-02
  MAX partition: 2013-01

```

```

Partition count: 12

=>New partition range (Monthly):
  MIN partition: 2012-02
  MAX partition: 2013-01
  Partition count: 12

=>Actions & progress:
  Parallel process partition 2012-11 /DataOnly
  Parallel process partition 2012-12 /DataOnly
  Parallel process partition 2013-01 /DataOnly

Rolling-window partitioning for table Reseller Sales
-----

=>Current partition range (Yearly):
  MIN partition: 2011
  MAX partition: 2013
  Partition count: 3

=>New partition range (Yearly):
  MIN partition: 2011
  MAX partition: 2013
  Partition count: 3

=>Actions & progress:
  Parallel process partition 2013 /DataOnly

Final operations
-----
  Save changes ...
  Recalc model to bring back online ...

Finish: 12:53:27 PM

```

Non-partitioned table processing & table omission

Execute the following INSERT statement to create table configurations for the Customer, Product and Sales Quota tables.

```

INSERT INTO [dbo].[TableConfiguration]
VALUES(
  3          --[TableConfigurationID]
  ,1        --[ModelConfigurationID]
  , 'Customer' --[AnalysisServicesTable]
  ,0        --[DoNotProcess]
),
(
  4          --[TableConfigurationID]
  ,1        --[ModelConfigurationID]
  , 'Product' --[AnalysisServicesTable]
  ,0        --[DoNotProcess]
),
(
  5          --[TableConfigurationID]
  ,1        --[ModelConfigurationID]
  , 'Sales Quota' --[AnalysisServicesTable]
  ,1        --[DoNotProcess]
)

```



```
);
```

Sales Quota has DoNotProcess equal to 1, so it will be excluded from processing. This flag can be used to dynamically include and exclude tables. For example, certain tables can be processed during the day for near-real time requirements and other tables processed overnight.

The Customer and Product tables will be processed. They do not have entries in the PartitioningConfiguration table, so they will be treated as non-partitioned tables and processed at the table level.

Other tables in the model that do not have table configuration entries are omitted from processing. Some tables may have no ongoing processing requirements. For example, the date-dimension table and categorical dimension tables typically have no need to be processed daily, so they can be left out altogether.

Execute SampleClient application, and the log query. The following results should be shown. The Customer and Product non-partitioned tables are processed.

```
Start: 08:55:05 PM
Server: localhost
Database: AdventureWorks

Rolling-window partitioning for table Internet Sales
-----

=>Current partition range (Monthly):
  MIN partition: 2012-02
  MAX partition: 2013-01
  Partition count: 12

=>New partition range (Monthly):
  MIN partition: 2012-02
  MAX partition: 2013-01
  Partition count: 12

=>Actions & progress:
  Parallel process partition 2012-11 /DataOnly
  Parallel process partition 2012-12 /DataOnly
  Parallel process partition 2013-01 /DataOnly
  Save changes for table Internet Sales ...

Rolling-window partitioning for table Reseller Sales
-----

=>Current partition range (Yearly):
  MIN partition: 2011
  MAX partition: 2013
  Partition count: 3

=>New partition range (Yearly):
  MIN partition: 2011
  MAX partition: 2013
  Partition count: 3

=>Actions & progress:
```

```
Parallel process partition 2013 /DataOnly
Save changes for table Reseller Sales ...
```

```
Non-partitioned processing for table Customer
-----
```

```
Process table Customer /DataOnly
```

```
Non-partitioned processing for table Product
-----
```

```
Process table Product /DataOnly
```

```
Final operations
-----
```

```
Save changes ...
Recalc model to bring back online ...
```

```
Finish: 08:55:17 PM
```

Merging partitions

Merging of partitions may be useful in mixed-granularity scenarios. For example, merging historical days into a month, or merging historical months into a year. Care must be taken when merging partitions to ensure it is done correctly. Once merged, it is not possible to unmerge.

In Program.cs, change the `_executionMode` variable to be assigned `MergePartitions`.

```
private static ExecutionMode _executionMode = ExecutionMode.MergePartitions;
```

Execute SampleClient application, and the log query. The following results should be shown. The months in 2012 are merged into the year 2012.

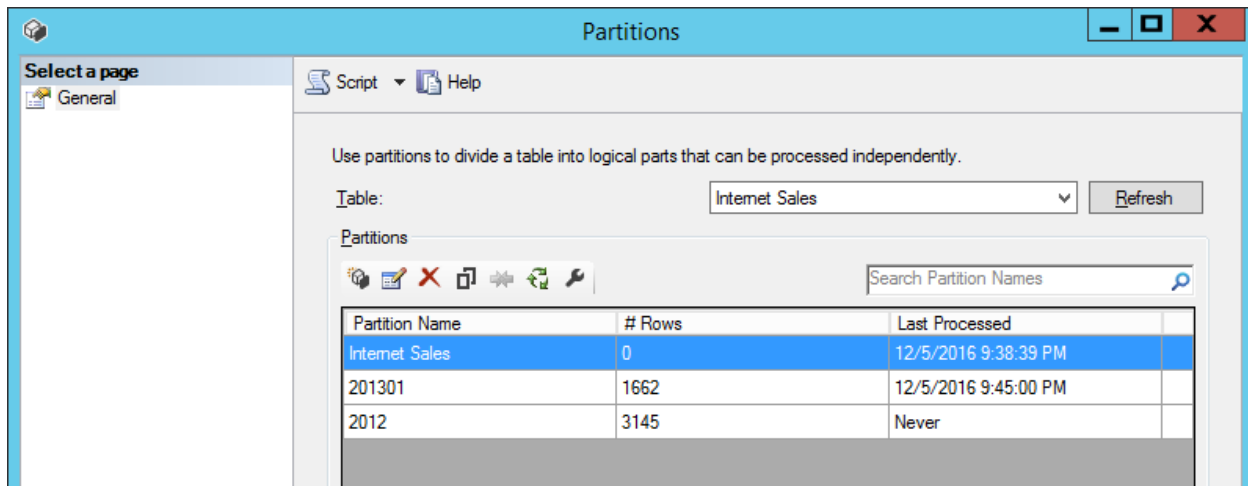
```
Merge partitions into 2012 for table Internet Sales
-----
```

```
=>Actions & progress:
```

```
Create new merged partition 2012 for table Internet Sales
Partition 201202 to be merged into 2012
Partition 201203 to be merged into 2012
Partition 201204 to be merged into 2012
Partition 201205 to be merged into 2012
Partition 201206 to be merged into 2012
Partition 201207 to be merged into 2012
Partition 201208 to be merged into 2012
Partition 201209 to be merged into 2012
Partition 201210 to be merged into 2012
Partition 201211 to be merged into 2012
Partition 201212 to be merged into 2012
Save changes for table Internet Sales ...
```

```
Finish: 10:03:38 PM
```

Inspect the new partition structure in SSMS.



Mixed-granularity configurations

For mixed granularity scenarios, it may be necessary to set up multiple configurations for a single table. This allows automated removal of old partitions – at different granularities – that fall out of range. If automated removal is not required, it is not necessary to set up mixed-granularity configurations. Care must be taken to ensure they are maintained correctly to avoid integrity issues.

In Program.cs, change the `_executionMode` variable back to `InitializeFromDatabase`.

```
private static ExecutionMode _executionMode = ExecutionMode.InitializeFromDatabase;
```

Execute the following INSERT and UPDATE statements to configure mixed granularity for the Internet Sales table. The yearly configuration covers 2012; the monthly one covers January through March 2013. There is no overlap between the date ranges at different granularities.

```
--Insert yearly configuration for Internet Sales
INSERT INTO [dbo].[PartitioningConfiguration]
VALUES
(
    3                --[PartitioningConfigurationID]
    ,1               --[TableConfigurationID]
    ,2               --[Granularity] 2=Yearly
    ,1               --[NumberOfPartitionsFull]
    ,1               --[NumberOfPartitionsForIncrementalProcess]
    ,0               --[MaxDateIsNow]
    , '2012-12-31'  --[MaxDate]
    ,1               --[IntegerDateKey]
)
'let
    Source = #"AdventureWorks",
    dbo_FactInternetSales = Source{[Schema="dbo",Item="FactInternetSales"]}[Data],
    #"Filtered Rows" = Table.SelectRows(dbo_FactInternetSales, each [OrderDateKey] >=
{0} and [OrderDateKey] < {1}),
    #"Sorted Rows" = Table.Sort(#"Filtered Rows",{ "OrderDateKey", Order.Ascending})
in
    #"Sorted Rows" --[TemplateSourceQuery]
);
--Update monthly configuration for Internet Sales
```

```

UPDATE dbo.[PartitioningConfiguration] SET
    NumberOfPartitionsFull = 3,
    NumberOfPartitionsForIncrementalProcess = 3,
    MaxDate = '2013-03-01'
WHERE PartitioningConfigurationID = 1;

```

Execute SampleClient application, and the log query. The following results should be shown. Internet Sales has 3 partitions at the month level, and 1 at the year level.

```

Start: 11:24:05 PM
Server: localhost
Database: AdventureWorks

Rolling-window partitioning for table Internet Sales
-----

=>Current partition range (Yearly):
    MIN partition: 2012
    MAX partition: 2012
    Partition count: 1

=>New partition range (Yearly):
    MIN partition: 2012
    MAX partition: 2012
    Partition count: 1

=>Actions & progress:
    Parallel process partition 2012 /DataOnly

Rolling-window partitioning for table Internet Sales
-----

=>Current partition range (Monthly):
    MIN partition: 2013-01
    MAX partition: 2013-01
    Partition count: 1

=>New partition range (Monthly):
    MIN partition: 2013-01
    MAX partition: 2013-03
    Partition count: 3

=>Actions & progress:
    Parallel process partition 2013-01 /DataOnly
    Create new partition      2013-02
    Parallel process partition 2013-02 /DataOnly
    Create new partition      2013-03
    Parallel process partition 2013-03 /DataOnly

Rolling-window partitioning for table Reseller Sales
-----

=>Current partition range (Yearly):
    MIN partition: 2011
    MAX partition: 2013
    Partition count: 3

```

```

=>New partition range (Yearly):
  MIN partition: 2011
  MAX partition: 2013
  Partition count: 3

=>Actions & progress:
  Parallel process partition 2013 /DataOnly

Non-partitioned processing for table Customer
-----
  Process table Customer /DataOnly

Non-partitioned processing for table Product
-----
  Process table Product /DataOnly

Final operations
-----
  Save changes ...
  Recalc model to bring back online ...

Finish: 11:24:15 PM

```

Validation of date ranges for mixed granularity configurations

It is important to ensure there are no overlapping date ranges at different granularities for the same table. AsPartitionProcessing performs some validation to avoid this.

Execute the following UPDATE statement to extend the yearly configuration range to include 2012 and 2013. There is now an overlap with the monthly configuration range, which covers January through March of 2013.

```

--Update yearly configuration for Internet Sales
UPDATE dbo.[PartitioningConfiguration] SET
  NumberOfPartitionsFull = 2,
  MaxDate = '2013-01-01'
WHERE PartitioningConfigurationID = 3;

```

Execute SampleClient application, and the log query. The following results should be shown. An exception was raised due to overlapping date ranges for the same table.

```

Start: 11:51:22 PM
Server: localhost
Database: AdventureWorks

Exception occurred: 11:51:48 PM
Exception message: Table Internet Sales contains partitioning configurations with
overlapping date ranges, which is not allowed. Yearly upper boundary is 2013-12-31;
Monthly lower boundary is 2013-01-01.

```

Execute the following UPDATE statement to remove the overlap.

```

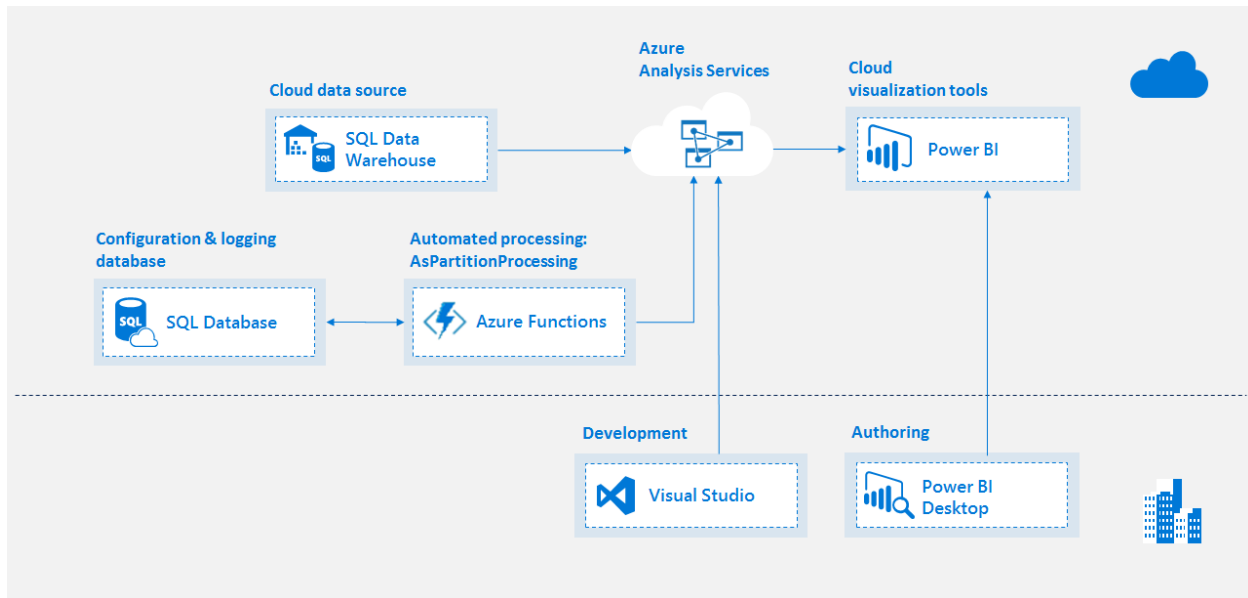
--Update yearly configuration for Internet Sales
UPDATE dbo.[PartitioningConfiguration] SET
  NumberOfPartitionsFull = 1,
  MaxDate = '2012-01-01'
WHERE PartitioningConfigurationID = 3;

```

Other Considerations

Cloud architecture

AsPartitionProcessing can be used in cloud PaaS/SaaS architectures. The following diagram shows an example of such an architecture.



[Azure SQL Database](#) is used for the configuration and logging database.

[Azure Functions](#) is used with a reference to the AsPartitionProcessing class library. Execution can be triggered in various ways. The following list is just a sample of the options available with Azure Functions. Detailed discussion on the pros and cons of each option is outside the scope of this document.

- Scheduled using a Timer function CRON expression. In this case, it is not necessary to set up a separate scheduling system.
- Using a webhook request for a WebHook function, or an HTTP request for an HttpTrigger function. This allows integration with existing scheduling systems that can call a URL.
- Triggered from Azure Queue using built-in integration points in Azure Functions.

To create an Azure Function, search for Function App in the Azure Marketplace. Specify a Hosting Plan of App Service Plan (not Consumption Plan) to allow long-running function calls. Note that the costs are different depending on the hosting plan.

It is necessary to upload the TOM and AsPartitionProcessing DLLs to Azure Functions. See [here](#) for more information on how to do this. The TOM DLLs must be the minimum version to work with Azure Analysis Services, and are available by installing the client libraries from [here](#). The following screenshot shows the Azure Functions editor with the DLLs uploaded to the bin folder and referenced at the top.

Code (run.csx)

Save

Run

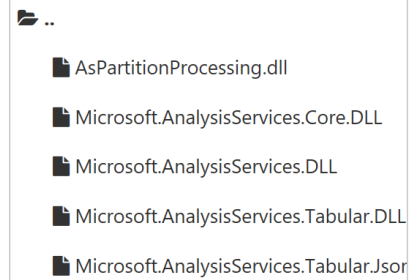
Logs

View Files

Test

Keys

```
1 #r "Microsoft.WindowsAzure.Storage"
2 #r "System.Configuration"
3 #r "Microsoft.AnalysisServices.DLL"
4 #r "Microsoft.AnalysisServices.Core.DLL"
5 #r "Microsoft.AnalysisServices.Tabular.DLL"
6 #r "AsPartitionProcessing.DLL"
7
8 using System;
9 using System.Configuration;
10 using AsPartitionProcessing;
11
12 public static void Run(TimerInfo myTimer, TraceWriter log)
13 {
```



With the necessary class-library references in place, the function body can be written using `SampleClient` as a template. Again, `PerformProcessing` is the key method.

```
PartitionProcessor.PerformProcessing(modelConfig, ConfigDatabaseHelper.LogMessage);
```

The `modelConfig` parameter can be initialized in a similar way to the `SampleClient` application. Information that `SampleClient` stores in `App.config` for the configuration and logging database can be stored using built-in Azure Functions application settings.

It is not necessary for the `LogMessage` delegate parameter to refer to a method defined in Azure Functions if there are no custom logging requirements. Instead the `ConfigDatabaseHelper.LogMessage` method can be referenced directly as it has the required signature.

Command-line execution

In the examples above, the `_executionMode` variable was changed in the code. It can also be set using a command-line argument from the Command Prompt or PowerShell. This reduces the cases where code changes are required.

Syntax

```
AsPartitionProcessing.SampleClient.exe
  [--ExecutionMode ExecutionMode]
  [--MergeTable MergeTable]
  [--MergeTargetGranularity MergeTargetGranularity]
  [--MergePartitionKey MergePartitionKey]
  [--ModelConfigurationIDs ModelConfigurationIDs]
```

Arguments

`--ExecutionMode`

Execution mode of `SampleClient`. Possible values are:

InitializeInline: Initialize configuration inline using sample values.

InitializeFromDatabase: Initialize from configuration and logging database.

MergePartitions: Merge partitions in a table based on other arguments.

DefragPartitionedTables: Defragment partitioned tables in the model. List of partitioned tables defined in the configuration and logging database.

--MergeTable

When ExecutionMode=MergePartitions, name of the partitioned table in the tabular model.

--TargetGranularity

When ExecutionMode=MergePartitions, granularity of the newly created partition. Possible values are Yearly or Monthly.

--MergePartitionKey

When ExecutionMode=MergePartitions, target partition key. If year, follow yyyy; if month follow yyyyymm.

--ModelConfigurationIDs

Comma-delimited list of ModelConfigurationIDs to filter on when getting worklist from the configuration and logging database.

--Help

Display the help screen.

Examples

The following command initializes from the configuration and logging database. All models listed in the ModelConfiguration table are considered for processing.

```
AsPartitionProcessing.SampleClient.exe --ExecutionMode InitializeFromDatabase
```

The following command initializes from the configuration and logging database. Only models with ModelConfigurationID equal to 1, 2 or 3 are considered for processing.

```
AsPartitionProcessing.SampleClient.exe ^  
  --ExecutionMode InitializeFromDatabase ^  
  --ModelConfigurationIDs 1,2,3
```

The following command defragments partitioned tables defined in the configuration and logging database.

```
AsPartitionProcessing.SampleClient.exe --ExecutionMode DefragPartitionedTables
```

The following PowerShell example merges monthly partitions from the previous year in the Internet Sales table. It could be scheduled for execution on the earliest day of a new year where business rules dictate the previous year's data is no longer subject to change.

```
$Today = Get-Date  
$LastYear = $Today.AddYears(-1) | Get-Date -Format yyyy  
.\AsPartitionProcessing.SampleClient.exe  
  --ExecutionMode MergePartitions  
  --MergeTable 'Internet Sales'  
  --TargetGranularity Yearly  
  --MergePartitionKey $LastYear
```

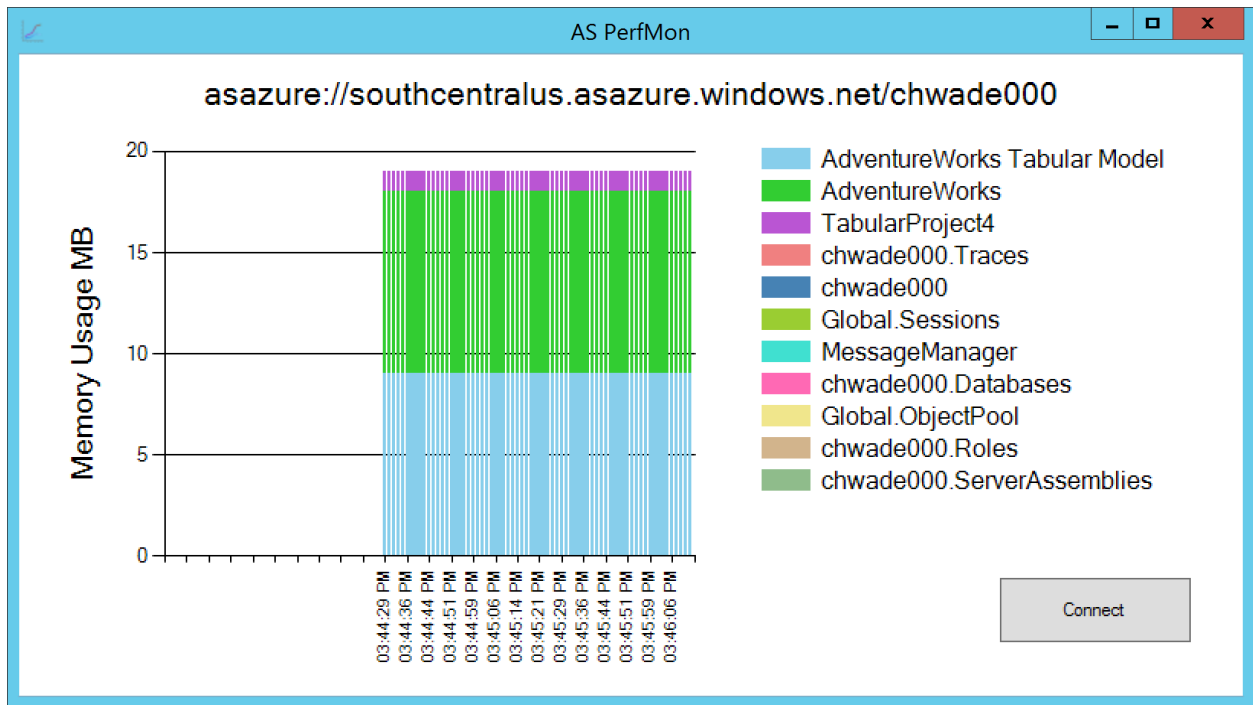

The following PowerShell example merges daily partitions from the previous month in the Internet Sales table. It could be scheduled for execution on the earliest day of a new month where business rules dictate the previous month's data is no longer subject to change.

```
$Today = Get-Date
$LastMonth = $Today.AddMonths(-1) | Get-Date -Format yyyyMM
.\AsPartitionProcessing.SampleClient.exe
--ExecutionMode MergePartitions
--MergeTable 'Internet Sales'
--TargetGranularity Monthly
--MergePartitionKey $LastMonth
```

AsPerfMon

Another code sample that may be useful in conjunction with AsPartitionProcessing is the AsPerfMon tool, which is available [here](#). AsPerfMon can be used to check real-time memory usage during processing. It splits memory usage by database, which is informative when multiple databases share the same server.

This is useful for Azure AS since you can't use Task Manager or create Performance Monitor counters. Similar functionality is provided by the Metrics section in the control blade for an Azure AS server in the Azure Portal. By using Metrics, you can check usage for the past day or week. AsPerfMon is for real-time monitoring during processing.



Custom logging

The LogMessage method in SampleClient Program.cs is passed as a delegate into the PerformProcessing method, so it can easily be changed for custom logging requirements.

```
private static void LogMessage(string message , MessageType messageType,
                             ModelConfiguration partitionedModel)
{
    //Can provide custom logging code here
```

...

Fragmentation

Partitioned tables may suffer from fragmentation over time. When a partition is removed from a table, hash dictionary entries for the columns are retained even if all the rows referring to those values have been removed. Defragmentation removes the unused dictionary entries. It is not necessary to perform defragmentation on non-partitioned tables because they are processed at the table level.

Defragmentation of large tables can be an expensive, sometimes time-consuming, operation. Further information on defragmentation is available [here](#). Dictionary and table size can be monitored using community tools such as [VertiPaq Analyzer](#) and [SSAS Memory Usage Report](#).

In Program.cs, change the `_executionMode` variable to be assigned `DefragPartitionedTables`.

```
private static ExecutionMode _executionMode = ExecutionMode.DefragPartitionedTables;
```

Execute SampleClient application, and the log query. The following results should be shown. The tables with partitioning configurations are defragmented.

```
Start: 12:30:15 PM
Server: localhost
Database: AdventureWorks

Defrag partitioned tables in database AdventureWorks
-----

=>Actions & progress:
    Defrag table Internet Sales ...
    Defrag table Reseller Sales ...

Finish: 12:30:17 PM
```

Locking and blocking

Detailed discussion around locking causes and diagnosis is out of the scope of this document. However, it is worth pointing out that queries take `Commit_Read` locks, which can block `Commit_Write` locks. `Commit_Write` locks are required by processing operations. Therefore, long-running queries can block processing operations. Normally, processing operations will wait for the period set in the `ForceCommitTimeout` server property; the default is 30 seconds. If the query still hasn't finished, the query will be cancelled by the server and the processing operation will then continue. The client tool executing the query will see an error such as the following.

```
Executing the query ...
The operation was cancelled because of locking conflicts.
```

In some cases, it may be beneficial to prioritize queries over processing operations. For example, when performing small, near-real time refreshes at regular intervals during the day. If many users are using the system, it is often preferable to allow queries to run to completion and fail the processing operation instead, especially if it will run again after a short interval anyway.

This can be achieved using the `CommitTimeout` property. `CommitTimeout` is a server property to specify how long the server will wait to acquire write locks (typically for processing operations). The default value is zero (0), meaning the server will wait indefinitely. In practice, it does not normally wait more

than 30 seconds because ForceCommitTimeout will kick in and fail long-running queries (as explained above).

CommitTimeout can be overridden for a connection used by processing operations. With ForceCommitTimeout set to the default of 30 seconds and CommitTimeout set to 20 seconds, processing operations will wait 20 seconds to acquire write locks. If still blocked by a long-running query, the processing operation will fail instead, allowing the query to run to completion.

In Program.cs, change the `_executionMode` variable to be assigned `InitializeFromDatabase`.

```
private static ExecutionMode _executionMode = ExecutionMode.InitializeFromDatabase;
```

Execute the following UPDATE statement to set CommitTimeout to 20 seconds (20,000 milliseconds).

```
UPDATE [dbo].[ModelConfiguration] SET [CommitTimeout] = 20000
```

If you now start execution of a long-running query (over 20 seconds) just before execution of the SampleClient application, the processing operation should fail with the following error message.

```
Exception message: Failed to save modifications to the server. Error returned: 'The lock operation timed out'.
```

Note that commit operations have been optimized considerably for tabular models in SQL Server 2016. This has caused noticeable improvements in locking and blocking for some customers with near-real time processing requirements. Database write-commit locks are required to safely complete tasks such as merging pending changes, persisting files to disk, clearing some cached state, deletion of old files, etc. In previous versions of Analysis Services, a server-level write commit lock was taken while most of these tasks were performed. With SQL Server 2016, the server-level locks are far more limited; they are only taken while producing the delta of transaction updates, and are then immediately released.

Auto retry

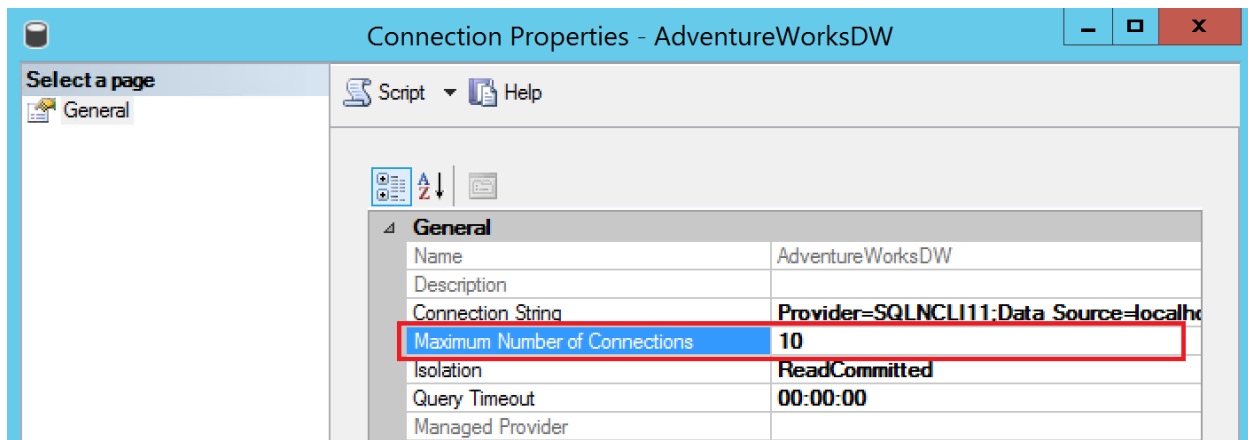
In the near-real time scenario described above, and in cases where the environment is subject to reliability problems such as network connectivity issues, auto retry can be configured.

The following UPDATE statement instructs the code sample to retry 3 times with an interval of 20 seconds between each retry attempt.

```
UPDATE [dbo].[ModelConfiguration] SET [RetryAttempts] = 3, [RetryWaitTimeSeconds] = 20
```

Parallelization of incremental processing

Incremental processing is submitted by `AsPartitionProcessing` as a parallelized operation for all tables and partitions within a model. Parallelization can be reduced by setting the `MaxConnections` property on a data source in SSMS. The default of 10 means that no more than 10 queries will be submitted to that data source at a time. This can be used to limit stress on source systems during processing.



Analysis Services will determine the number of concurrent processing jobs based on the number of cores and available threads. These threads are shared across the server instance, and therefore one command may not receive all the available threads. The threads that do launch processing jobs may then be stalled to stay within the MaxConnections limit.

Resource-constrained environments can further limit parallelization by setting the MaxParallelism property for a connection used to submit processing requests. The following example UPDATE statement limits AsPartitionProcessing to 2 concurrent processing operations.

```
UPDATE [dbo].[ModelConfiguration] SET [MaxParallelism] = 2
```

The impact of setting MaxConnections and MaxParallelism is visible using the [SSAS performance counters](#) in the Threads section such as *Processing pool busy non-I/O threads*.

Model deployment

When deploying new versions of partitioned tabular models that already exist on the target environment, it is necessary to be aware of the partitioning process in-place. As shown by the code sample, partitions are normally created and managed separately from the deployment process. This means the version of the tabular model from source control typically does not contain the partitions. A simple deployment process such as right-click, Deploy from SSDT will lose the partitions and all the data within them. Tools that support deployment retaining partitions include [BISM Normalizer](#) and the [Analysis Services Deployment Wizard](#). Both these tools support command-line execution for automated deployment. Detailed discussion on this topic including the pros and cons of these tools is outside the scope of this document.