

# Engine Settings For System-Defined Aggregates Only

AtScale provides a number of settings for configuring system-defined aggregates.

For descriptions of the different types of aggregates, see [Types of Aggregate Tables in AtScale](#).

## Settings For Enabling And Changing The Compression Factor

Use these settings to specify a compression factor. This factor is a measure of the quality of a proposed aggregate. It is calculated as the number of rows in the fact dataset divided by the estimated number of rows in a proposed aggregate.

### **AGGREGATES.CREATE.THRESHOLD.ENABLED**

Set to **True** to turn on the setting `AGGREGATES.CREATE.COMPRESSION.THRESHOLD`.

### **AGGREGATES.CREATE.COMPRESSION.THRESHOLD**

Specify the compression factor that aggregates proposed by the engine must meet or exceed.

## Settings For Narrowing And Widening

Use these settings to enable the narrowing and widening of aggregates.

### **AGGREGATES.CREATE.NARROWING.ENABLED**

Set to **True** to allow the engine to define new aggregates as narrower versions of existing aggregates when the compression factor is met or exceeded by the new aggregates. Narrower aggregates contain fewer dimensions than their predecessors.

You can set the compression factor with the setting `AGGREGATES.CREATE.COMPRESSION.THRESHOLD`.

### **AGGREGATES.CREATE.WIDENING.ENABLED**

Set to **True** to allow the engine to define new aggregates as wider versions of existing aggregates. Wider aggregates contain more measures than their predecessors.

### **AGGREGATES.WITHDISTINCTCOUNTS.WIDENING.ENABLED**

Allows widening of distinct count aggregates with other measures, including distinct counts. This setting requires `AGGREGATES.CREATE.WIDENING.ENABLED` to be set to **True** (see above).

Note that this setting is applied per model.

### **AGGREGATES.CREATE.WIDENING.MEASURE.LIMIT**

Specify the maximum number of measures that can be added when widening. This setting requires `AGGREGATES.CREATE.WIDENING.ENABLED` to be set to **True** (see above).

## **Settings For Prediction-Defined Aggregates**

These settings enable the AtScale engine to create prediction-defined aggregates.

### **AGGREGATE.SPECULATIVE.ENABLED**

Set to **True** to activate the following settings that are related to prediction-defined aggregates.

The default is **True**.

### **AGGREGATE.SPECULATIVE.ALLMEMBER.ENABLED**

Set to **True** to enable the AtScale engine to create all-member aggregates when new catalogs are deployed for the first time, as well as when they are redeployed.

The default is **True**.

### **AGGREGATE.SPECULATIVE.DIMENSIONAL.ENABLED**

Set to **True** to enable the AtScale engine to create dimension-only aggregates when new catalogs are deployed for the first time and when they are redeployed. This type of aggregate is used for populating filters in client BI applications, such as Tableau and Microsoft Excel.

The default is **True**.

### **AGGREGATE.SPECULATIVE.DIMENSIONAL.MINCOMPRESSIONRATIO**

Specify the ratio as the number of rows in the full dimension dataset divided by the number of rows in the proposed aggregate for a level in the dimensional hierarchy.

For example, for a Date dimension, the lowest level in the hierarchy might be Day. The number of rows in an aggregate defined on Day would be the same number of rows in the dataset overall. There would be no aggregation. However, an aggregate defined on a higher level in the hierarchy, such as Quarter, would aggregate the data and therefore have a compression ratio. The level Year would aggregate further and have a higher compression ratio.

The default compression ratio is 10.

### **AGGREGATE.SPECULATIVE.SUPERAGGREGATE.ENABLED**

Set to **True** to allow the AtScale engine to define and create instances of super aggregate tables. Super aggregates, a type of prediction-defined aggregate, contain all keys in a fact dataset. They also contain all degenerate dimensions for which a fact dataset contains values at all levels in the dimensional hierarchies.

Whether or not a super aggregate is defined from a fact table is determined by calculations that are based on `AGGREGATE.SPECULATIVE.SUPERAGGREGATE.COMPRESSION`.

### **AGGREGATE.SPECULATIVE.SUPERAGGREGATE.COMPRESSION**

Specify a value of the Double data type. The AtScale engine, when it considers whether to define a super aggregate table for a fact dataset, divides the number of rows in the dataset by the value that you specify for this setting. If the estimated number of rows in the super aggregate table is less than or equal to the resulting quotient, then the engine defines the super aggregate table.

The default value is 2.0.

## **Settings For Specifying Whether And How To Use Joins**

These settings enable the AtScale engine to define aggregates that use joins and affect how it determines whether a join is appropriate for a given aggregate.

### **AGGREGATES.CREATE.JOINS.ENABLED**

Set to **True** to allow the AtScale engine to use joins when defining aggregates.

This setting must be set to **True** for the following three settings to have an effect.

The default value of this setting is **True**.

### **AGGREGATES.CREATE.JOINS.COMPRESSION**

Specify the minimum compression ratio for any proposed join. This ratio is calculated as the cardinality of the join key to the cardinality of the grouped dimension values (i.e.  $\#(\text{Join Key}) / \#(\text{Dim Table grouped by Dim Value})$ ). Joins for which the compression ratio is lower than this minimum will not be used.

The default is 100.

### **AGGREGATES.CREATE.JOINS.MAXIMUMDEPTH**

Specify the maximum number of dimensions that can be traversed in a join path.

The default is 3.

### **AGGREGATES.CREATE.JOINS.MAXIMUMKEYCARDINALITY**

Specify the maximum cardinality that the AtScale engine will allow in join keys when the engine is determining whether to use a join in the definition of an aggregate. Higher cardinalities will cause the engine not to use a join.

The default value is 10,000,000.

## **Settings That Control Partitioning Of System-Defined Aggregate Tables**

One of these settings enables the AtScale engine to partition instances of aggregate tables that it defines. The other setting lets you set a threshold for determining when to partition them. Both of these settings can also be applied at the model level.



**Note:** AtScale supports aggregate table partitioning for Google BigQuery using columns of type Date, DateTime, and Integer for partition columns.

### **AGGREGATES.CREATE.PARTITION.SYSTEMDEFINEDAGGREGATE.ENABLED**

Set to **True** to enable the AtScale engine to partition system-defined aggregates. For this setting to have an effect, the setting TABLES.CREATE.PARTITIONS.ENABLED must be set to **True**.

### **AGGREGATES.CREATE.PARTITION.SYSTEMDEFINEDAGGREGATE.THRESHOLD**

Specify minimum number of rows per partition. The AtScale engine divides the estimated cardinality of a proposed system-defined aggregate table by the estimated number of partitions. If the estimated number of rows per partition does not meet or exceed this threshold, the engine will not partition the aggregate table. This value prevents the engine from creating too many partitions per aggregate table, as query processing times can increase if the number of

partitions becomes too high. This value also prevents the engine from creating not enough partitions per aggregate table, as a small number of very large partitions can also cause query processing times to increase. In both cases, the advantages of partitioning are negated. The default value is 50000.0.

## Settings For Automatic Removal Of Unused Aggregates

Use these settings to purge system-generated demand-defined aggregates with zero utilizations after a configurable time period, independent of the set retention limit. This improves resource utilization within the target data warehouse. When system-generated demand-defined aggregates meet the criteria for removal, they are first deactivated and then later deleted from the target data warehouse.

### REMOVE UNUSED AGGREGATES

Enable or disable the setting for the organization. Then, set a time period ranging from 10-365 days, to track the daily usage of aggregate definitions during the aggregate purging time window. Only system-generated demand-defined aggregate definitions with zero utilizations identified during the time window (default of 45 days) will be purged by the AtScale engine.

## Other Settings For System-Defined Aggregates Only

A number of features of system-defined aggregates are enabled or affected by particular settings.

### **Aggregates.Systemgenerated.Activeinstance.Retentionlimit**

The system default value for the maximum number of system-defined aggregates retained per model. Setting this value too high will cause long aggregate batch build times and may impact data warehouse workloads.

### **Aggregates.Create.Allowexactdistinctcountmeasures.Enabled**

Used for creating System-Defined Aggregates that contain Distinct Count measures. It is disabled by default. To enable it, set its value to **True**.

See also `aggregates.systemgenerated.withdistinctcountmeasure.retentionpercentage` and `aggregates.systemGenerated.activeInstance.retentionLimit`.

### **Aggregates.Systemgenerated.Withdistinctcountmeasure.Retentionpercentage**

To prevent Distinct Count aggregates from dominating the Aggregate retention limit, AtScale defines a separate pool of Distinct Count aggregates as a percentage of the overall model-scoped System-Defined aggregate retention limit. The percentage is controlled by this setting, at either the Engine or model level. It accepts values between 0 and 100,

with a default of 40. The size of the Distinct Count aggregate pool is calculated as a percentage of the effective model value of the "retentionLimit" setting (see above).

For example if the model's effective retention limit is 100, and `aggregates.systemGenerated.withDistinctCountMeasure.retentionPercentage` is 40, then the system will allow the creation of up to 40 System-Defined aggregates that use Distinct Count measures.

### **Aggregates.Systemgenerated.Activeinstance.Extraallowance**

The system default value for the maximum number of additional system-defined aggregates temporarily permitted per model when the retention limit is reached. Setting this value too high will cause long aggregate batch build times and may impact data warehouse workloads.

### **Aggregates.Completeness.Degeneratedimensionswithmeasures.Enabled**

Set to **True** to allow the AtScale engine to direct dimension-only queries to aggregate instances that contain complete degenerate dimensions.

### **Aggregates.Dimensional.Build**

Set to **True** to allow the engine to create aggregates that contain dimensional attributes only. Such aggregates can be useful in Tableau for queries against fact datasets that contain degenerate dimensions.

### **Aggregates.Create.Includehigherlevels.Enabled**

Set to **True** to enable the aggregation process to evaluate higher levels of a hierarchy when building an aggregate. If a higher level meets the needed criteria, it will be built in. For example: if this setting is enabled and a hierarchy contains Year and YearMonth, when an aggregate build against YearMonth \* myMeasure is requested, the build process evaluates Year as a possible candidate too.

### **Aggregates.Prediction.CheckForNonAdditiveMeasuresAndConstraints.Enabled**

AtScale checks for aggregate candidates to queries selecting distinct count measures and with a non-equals WHERE constraint. When this option is set to **True** (default), such aggregate candidates are rejected, because they attempt to re-aggregate the data.