# Aggregate Definitions And Aggregate Instances

AtScale separates the concept aggregate table into two separate concepts: the definition of an aggregate table and an actual instance (or materialization) of that definition. To summarize, each aggregate table in AtScale is an instance of a definition.

- An aggregate-table definition is the list of measures and dimensions that are included in instances of an aggregate table. You can think of this definition as a virtual dataset that is defined on a cube and that usually includes one or more measures from one fact dataset together with a number of dimensional attributes and their values. The measures are aggregated to the granularity of the dimensional attributes.
- An aggregate-table instance is a materialized definition. This is what is commonly known as an aggregate table (or aggregate), because each instance contains data. For every definition, the AtScale engine maintains one or more instances. Instances are stored in the data warehouse where the unaggregated data is located.

**If Hadoop is your data warehouse:** The instances are stored in their own directory in HDFS and their metadata is stored in the Hive metastore.

Most often, instances are stored in Parquet files, which use a highly compressed format to keep file sizes small and also use columnar storage, so that compression is applied per column. However, AtScale can also store instances in RC files, ORC files, and sequence files. AtScale can even use Hive's SerDe interface.

The choice of file format depends on an algorithm and on the given scenario, as Parquet does not support all possible scenarios in which aggregate tables can be used. For example, some column types do not support Parquet.