

Analytics Data Model

This document describes the data model for the Jive SBS analytics feature. You'll find the specifics of the schema -- table and column names, for example -- in the [Analytics Schema Reference](#). For more on the optional analytics module, see the [Jive web site](#).

Overview of Design

The analytics data model is a star schema form. The schema is comprised of a fact table that represents the events in Jive SBS, along with corresponding dimension tables that represent the actors and objects that take part in those events. Each column in the fact table contains a key that relates to an entry in the corresponding dimension table. For example, the `user_id` column contains IDs that can be found in the `jivedw_user` table. A basic query against the analytics schema will be a `SELECT` from the fact table, optionally performing `INNER JOINS` against the dimension tables. The dimension tables are used both constrain the results of the query based on their attributes and also to provide attribute data, such as the usernames of users.

Activity Fact Table

At the heart of the model is the activity fact table. Each record in the fact table represents a unique event that occurred. Essentially each record conveys the following:

At <TIME>, <USER> performed <ACTION> on <OBJECT> (optionally with <INDIRECT OBJECT>) in <CONTAINER>

Each of the variables in <> is represented as an ID in one of the columns of the fact table.

The `activity_ts` column provides the timestamp when the activity occurred in the system. Since timestamps represent times down to the second or nanosecond level, this column provides the ability to analyze activity down to a very fine granularity. However, a foreign key to the `JIVEDW_DAY` table has also been provided to allow for selecting records using a time dimension. The `JIVEDW_DAY` dimension is the smallest granularity of the time dimensions available.

The `user_id` column contains the ID of the user that initiated the action, sometimes referred to as the actor. This is a foreign key to the `JIVEDW_USER` table.

The `activity_type` column contains a code indicating the action or event that occurred. Refer to the table on event codes. Most common activity codes will be `READ(10)`, or `CREATE(20)`, but there are many more.

The `dw_direct_object_id` contains the ID of the action's object and its entry can be found in the `JIVEDW_OBJECT` dimension table. Objects represent all the types of data entities in the application. Most of these types are content types such as documents, blog posts, discussions, comments and any custom content type implemented by a plug-in. However there are other types of entity that can act as an object of an event such as a user or container. For example, an event may represent a connection being made between two users, or another event may describe a user viewing another user's profile.

The type of the object in the `direct_dw_object_id` column will depend on the activity type of the event. For example, a `WATCH` event will contain any type of object that represents watchable content, but will not contain a user object. For convenience, the object type is also stored in the fact record, although it can also be determined from the `DW_OBJECT` dimension table.

The `indirect_dw_object_id` is an optional field that provides the ID of a secondary object in the event. Not all events require a secondary object. An example of an event that has a secondary object is the `MOVE` event. With the move event the direct object is the ID of the object being moved and the indirect object is the ID of the container it is being moved to.

The container ID column contains the ID of the container of the object. This is not applicable to all events.

Aggregate Fact Tables

The activity fact table provides the finest level of granularity of events – each event is recorded as a record in the table. Reporting does not often require this level of granularity, especially when performing analysis over a large time frame. Additionally there may be multiple events that are the same -- for example, a user may read the same document multiple times in the same day. This leads to an explosion in the size of the table that ultimately has a detrimental effect on query performance.

Aggregate tables provide summarized data across all or specific dimensions. Instead of representing each unique event, each record provides a count of the number of events that occurred for the combination of dimension values. For example, if the aggregate is representing the day granularity, and a user viewed a document multiple times on that day,

there will be a single record recording that activity but with a count indicating the number of times it occurred. Aggregate tables will contain significantly fewer records than the base fact table, providing much better query performance.

The analytics data model provides aggregate tables that summarize the unique events at the Day, Week and Month periods. Additionally, some aggregate tables represent unique events for “All Users”, such as by removing the user dimension from the fact table and summarizing the events. For example, if three users view the same document, it will be represented as a single record but with a count of three. This aggregate is also provided at the Day, Week and Month periods.

Note: When aggregating up to a particular time period, it is not possible to retain the timestamp in the aggregate table. Instead of the time key of the period can be retained. Fortunately, each of the time dimensions contains beginning and end timestamps representing the period's range. Therefore you can still use timestamps in queries.

Table	Description
JIVEDW_ACTIVITY_AGG_DAY	Activity aggregated for each day period
JIVEDW_ACTIVITY_AGG_WEEK	Activity aggregated for each week period
JIVEDW_ACTIVITY_AGG_MONTH	Activity aggregated for each month period
JIVEDW_ACTIVITY_AGG_USER_DAY	Activity aggregated for all users, for each day period
JIVEDW_ACTIVITY_AGG_USER_WEEK	Activity aggregated for all users, for each week period
JIVEDW_ACTIVITY_AGG_USER_MONTH	Activity aggregated for all users, for each month period

Jive Object Dimension

The Jive object dimension, JIVEDW_OBJECT, contains a record representing each Jive object entity in the system. A Jive object can be any content type, such as a document, discussion message, blog post, comment, and so on. It can also be a container, such as a group, space, or user container. Even a user can be considered an object. Each type of object has a unique object type ID, and for each object type the object ID is unique.

The Jive object dimension provides these attributes, and also provides a generic “Name” attribute. The name will contain either the title or subject of the content, if it’s available, for any content type that has a specific data load process. For all other content types in the system the name will simply be “CONTENT”.

For certain content types there are some additional dimension tables that can be joined to the object table - so-called "snowflake" dimensions. These

provide additional attributes for those types of objects. The following snowflake dimensions exist:

Table	Description
JIVEDW_DOCUMENT	Contains attributes for documents. NOTE: Object types DOCUMENT (102) and DOCUMENT_VERSION (120) refer to the same documents and thus share the same id. The document dimension will contain the most recent name of the document - which is the name of the DOCUMENT_VERSION with the greater version number.
JIVEDW_MESSAGE	Contains attributes for discussion messages.
JIVEDW_BLOGPOST	Contains attributes for blog posts.

User Dimension

The user dimension contains a record for each user defined in the system and provides attributes such as username, name, firstname and last name. The user ID is unique. The user_id column can be joined to the user_id in the activity fact and aggregate tables to query activity for that user.

User Profile Fields

The JIVEDW_USER_PROFILEFIELD dimension is a snowflake from the user dimension. It contains the values for profile fields set for the user and can be used to constrain analysis queries for certain types of users.

Profile Fields

The profile fields table contains the names and IDs of all profile field and custom field types defined in the application. It can be joined to the JIVEDW_USER_PROFILEFIELD table to provide the field name if necessary.

Time Dimensions

The Time dimensions provide unique keys that represent a time period. The time dimension also provide attributes to assist in selecting those keys, such as the day of the month, the month of the year and the year. They also provide beginning and ending timestamps that

represent the boundaries of the time period. This allows queries to select a period using a particular timestamp.

The primary keys of the time dimensions are used as foreign keys in their corresponding aggregate tables. For example, the MONTH_ID of TIME_MONTH is used in JIVE_ACTIVITY__AGG_MONTH and JIVEDW_ACTIVITY__AGG_USER_MONTH.

Note: For convenience, the base fact table contains a DAY_ID column in addition to the actual timestamp of the event.

Analytics Event Code Reference

Event Code	Value
VIEW	10
CREATE	20
DELETE	30
MODIFY	40
HIDE	50
COPY	60
MOVE	70
UNDELETE	80
EXPIRE	90
RATE_OR_VOTE	100
MODERATE	110
SEND	120
APPROVE	130
ONLINE	140
OFFLINE	150
LOG_IN	160
LOG_OUT	161
SEARCH_CONTENT	170
SEARCH_USERS	180
SPOTLIGHT_SEARCH	190
ASSOCIATE	200

DISSOCIATE	210
REGISTER	220
VALIDATE	230
REJECT	240
WATCH	250
REMOVE_WATCH	260
FOLLOW	270
UNFOLLOW	280
AIDED	290
RESOLVED	300

Object Type Reference

Object Type	Value
DISCUSSION THREAD	1
DISCUSSION MESSAGE	2
USER	3
ATTACHMENT	13
COMMUNITY (Space)	14
POLL	18
PRIVATE_MESSAGE	20
ANNOUNCEMENT	22
BLOG	37
BLOGPOST	38
TRACKBACK	40
TAG	41
TAG SET	42
USER STATUS	48
USER RELATIONSHIP	49
DOCUMENT	102
COMMENT	103
RATING	107
DOCUMENT VERSION	120
DOCUMENT VERSION COMMENT	121

PROJECT	600
SOCIAL GROUP	700
BOOKMARK	800
VIDEO	1100
USER CONTAINER	2020

Special User Reference

User Type	Value
Anonymous (Guest) User	-1
System User	1