Perceptive Core

Reference Guide

Version: 2.0.x

Written by: Product Knowledge, R&D
Date: November 2018

Hyland™

Table of Contents

PerceptiveDataWarehouseCore Documentation

The PerceptiveDataWarehouseCore DDL provides the foundational elements of the data warehouse upon which other supplementary solutions may be added. The Core DDL should be run first to create the Perceptive Data Warehouse. When running the script, the following tables will be created in the warehouse.

Table Name	Purpose
CORE_DATE_D*	Date dimension
CORE_DOC_D*	Document dimension (SCD)
CORE_DOC_TYPE_D*	Document type dimension (SCD)
CORE_ERROR_LOG	Error log for warehouse processing
CORE_TIME_D	Time dimension
CORE_USER_D*	User dimension

* In these dimension tables, the first two rows in the table are reserved for the Not Applicable and Not Found items. Not Applicable items are those which are processed by the ETL job but do not apply to the dimension. In contrast, the Not Found items are those that are processed but the ETL job is unable to find a match when looking up the values. Both of these instances should be rarities rather than the norm.

CORE_DATE_D

The date dimension provides date attribute information for all fact and throughput tables. When the PerceptiveDataWarehouseCore DDL script is run, it populates the CORE_DATE_D table with dates to the end of 2033, although the script can be modified to specify the start and end dates. This table is not updated by the ETL job because all dates are accounted for in the table.

One date is broken out into several different columns to allow the consumer of the data to choose the date type they prefer. For example, given a date of 2014-06-18, several key pieces of obvious and not so obvious information can be pulled from this date:

- DAY_NAME = Wednesday
- DAY_OF_WEEK = 4
- WEEK_OF_QUARTER = 12
- LAST_DAY_OF_WEEK = 2014-06-21

This can then be used within a query or report to filter down the data based on the preferred date criteria.

Column definitions for the CORE_DATE_D table

Column Name	Purpose	Example
CORE_DATE_D_KEY	Primary key used to identify the unique date record	20150502
FULL_DATE	The full date of row in normal date format	2015-05-02
DAY_OF_MONTH	Integer value of day number in month, 1-31	2
DAY_NAME	The day name in string format	Saturday
DAY_OF_WEEK	Integer value of day number in week, 1-7	7
DAY_OF_YEAR	Integer value of day number in year, 1-366	122
WEEK_OF_MONTH	Integer value of week number in month, 1-6	1
WEEK_OF_QUARTER	Integer value of week number in quarter, 1-14	5
WEEK_OF_YEAR	Integer value of week number in year, 1-54	18
MONTH	Integer value of month number in year, 1-12	5
MONTH_NAME	The month name in string format	May
MONTH_OF_QUARTER	Integer value of month number in quarter, 1-3	2
CALENDAR_QUARTER	Integer value of quarter number in calendar year, 1-4. Calendar year begins January 1.	2
CALENDAR_QUARTER_NAME	The calendar quarter name in string format	SECOND
YEAR	Integer value of year	2015
YEAR_NAME	The calendar year in string format	CY 2015
MONTH_YEAR	The calendar month and year in string format	May-2015
MM_YYYY	The calendar month and year in a concatenated string format	052015
FIRST_DAY_OF_MONTH	The first day of the month, based on the full date value	2015-05-01
LAST_DAY_OF_MONTH	The last day of the month, based on the full date value	2015-05-31
FIRST_DAY_OF_QUARTER	The first day of the quarter, based on the full date value	2015-04-01
LAST_DAY_OF_QUARTER	The last day of the quarter, based on the full date value	2015-06-30
FIRST_DAY_OF_YEAR	The first day of the year, based on the full date value	2015-01-01

Column Name	Purpose	Example
LAST_DAY_OF_YEAR	The last day of the year, based on the full date value	2015-12-31
FIRST_DAY_OF_WEEK	The first day of the week, based on the full date value	2015-04-26
LAST_DAY_OF_WEEK	The last day of the week, based on the full date value	2015-05-02
IS_WEEKDAY	A flag indicating if the day is a weekday (1) or not (0)	0

CORE_DOC_D

The document dimension stores attribute information on all document-based data that comes from Perceptive Experience to the warehouse. This is a slowly changing dimension (SCD) meaning that it will retain historical records of previous versions of the document while creating a new record of the document when updated data streams from Perceptive Experience. For example, we have a document with the following values.

Column Name	**Value**
CORE_DOC_D_KEY	4
DOC_ID	321YZ7Z_0002WLSJF00004V
FIELD_1_VALUE	RED
IS_LATEST	1
EFFECTIVE_DATE_UTC	2015-08-03 10:24:49.250
EXPIRED_DATE_UTC	9999-01-01 12:00:00.000

This document is reindexed so the Field 1 value is now Purple instead of Red. The first change that happens is the above table will be modified as follows.

Column Name	**Value**
CORE_DOC_D_KEY	4
DOC_ID	321YZ7Z_0002WLSJF00004V
FIELD_1_VALUE	RED
IS_LATEST	0
EFFECTIVE_DATE_UTC	2015-08-03 10:24:49.250

Column Name	Value
EXPIRED_DATE_UC	2015-08-04 13:15:11.321 **<-indicates the date of change**

The second change that happens is a new row will be inserted with the following information.

Column Name	Value
CORE_DOC_D_KEY	5
DOC_ID	321YZ7Z_0002WLSJF00004V
FIELD_1_VALUE	PURPLE
IS_LATEST	1
EFFECTIVE_DATE_UTC	2015-08-04 13:15:11.321
EXPIRED_DATE_UTC	9999-01-01 12:00:00.000

Column definitions for the CORE_DOC_D table

Column Name	Purpose	Example
CORE_DOC_D_KEY	Primary key used to identify the unique document record	3
DOC_ID	The Document ID in string format. If a document has been modified, there can be multiple rows with the same DOC_ID	321YZ7Z_0002W1SJF000004
DOC_NAME	The Document Name in string format. This string comes from the push service and may not be a readable text value	98806e7e-01db-45ca-8d3f-d7eb7d0a6377
FIELD_1_VALUE	The Field 1 value in string format	P664Q
FIELD_2_VALUE	The Field 2 value in string format	J4839X3
FIELD_3_VALUE	The Field 3 value in string format	E59889
FIELD_4_VALUE	The Field 4 value in string format	Jerry Mouse
FIELD_5_VALUE	The Field 5 value in string format	Small
IS_LATEST	A flag indicating if the row is the most recent (1) or not (0)	1
EFFECTIVE_DATE_UTC	A datetime stamp indicating when the row was initially inserted into the warehouse	2015-07-31 10:37:21.597

Column Name	Purpose	Example
EXPIRED_DATE_UTC	A datetime stamp indicating when the IS_LATEST flag in the row was set to 0. If the datetime stamp = 9999-01-01 12:00:00.000 that indicates the row is still active	9999-01-01 12:00:00.000

CORE_DOC_TYPE_D

The document type dimension stores attribute information on all document type based data that comes from Perceptive Experience to the warehouse. This is a slowly changing dimension (SCD) meaning that it will retain historical records of previous versions of the document type while creating a new record of the document when updated data streams from Perceptive Experience. For example, we have a document type with “CUSTOMER” as a Field 2 name value. It displays in the table as the following:

Column Name	Value
CORE_DOC_TYPE_D_KEY	8
DOC_TYPE_ID	321YZ3C_0000JJJX400001G
FIELD_2_NAME	CUSTOMER
IS_LATEST	1
EFFECTIVE_DATE_UTC	2015-07-15 18:31:56.991
EXPIRED_DATE_UTC	9999-01-01 12:00:00.000

The field name value changes from “CUSTOMER” to “PROSPECT” for this document type. The first change that happens is the above table will be modified as follows:

Column Name	Value
CORE_DOC_TYPE_D_KEY	8
DOC_TYPE_ID	321YZ3C_0000JJJX400001G
FIELD_2_NAME	CUSTOMER
IS_LATEST	0
EFFECTIVE_DATE_UTC	2015-07-15 18:31:56.991
EXPIRED_DATE_UTC	2015-08-09 08:13:12.989 **<-indicates the date of change**

The second change that happens is a new row will be inserted with the following information:

Column Name	Value
CORE_DOC_TYPE_D_KEY	27
DOC_TYPE_ID	321YZ3C_0000JJJX400001G
FIELD_2_NAME	PROSPECT
IS_LATEST	1
EFFECTIVE_DATE_UTC	2015-08-09 08:13:12.989
EXPIRED_DATE_UTC	9999-01-01 12:00:00.000

Column definitions for the CORE_DOC_TYPE_D table

Column Name	Purpose	Example
CORE_DOC_TYPE_D_KEY	Primary key used to identify the unique document type record	10
DOC_TYPE_ID	The Document Type ID in string format. If a document has been modified, there can be multiple rows with the same DOC_ID	321YZ3C_0000JJJX400001G
DOC_TYPE_NAME	The Document Type Name in string format.	Consultation
FIELD_1_NAME	The Field 1 name in string format	MRN
FIELD_2_NAME	The Field 2 name in string format	field2
FIELD_3_NAME	The Field 3 name in string format	CSN
FIELD_4_NAME	The Field 4 name in string format	field4
FIELD_5_NAME	The Field 5 name in string format	field5
IS_LATEST	A flag indicating if the row is the most recent (1) or not (0)	1
EFFECTIVE_DATE_UTC	A datetime stamp indicating when the row was initially inserted into the warehouse	2015-07-31 10:37:21.597
EXPIRED_DATE_UTC	A datetime stamp indicating when the IS_LATEST flag in the row was set to 0. If the datetime stamp = 9999-01-01 12:00:00.000 that indicates the row is still active	9999-01-01 12:00:00.000

CORE_ERROR_LOG

The CORE_ERROR_LOG table is not a dimension table, but an error logging table as the name suggests. Its purpose is to record any problems that occur during the processing of the ETL jobs so that an administrator can work with a support technician to troubleshoot potential issues. Logging is turned on by default within each of the ETL jobs so any errors that may occur will flow into this table.

Column definitions for the CORE_ERROR_LOG table

Column Name	Purpose	Example
CORE_ERROR_LOG_KEY	Primary key used to identify the error log record	14
TIMESTAMP	Date timestamp of when the error was recorded in the warehouse	2015-08-13 14:49:33.247
PID	A process identifier	wXSbyi
PROJECT	The project name for the error which was logged	HCCI_ETL
JOB	The job name for the error which was logged	HCCI_BC_Process
LANGUAGE	The language used by the job (java)	java
ORIGIN	Status evaluation origin	NULL
STATUS	Evaluation of job results	Failed
SUBSTATUS	Substatus: Detailed explanation for failed execution. The explanation can be: • Test logically failed: the investigated Job does not produce the expected result. • Execution error: an execution error occurred at runtime.	Job execution error
DESCRIPTION	Descriptive message indicating cause of error message	net.minidev.json.parser.ParseException: Unexpected token pageId" at position 62.

CORE_TIME_D

The time dimension provides time attribute information for all fact and throughput tables. When the PerceptiveDataWarehouseCore DDL script is run, it populates the CORE_TIME_D table with each second, minute and hour of the day. This table is not updated by the ETL job because all times for the day are accounted for in the table.

One 24-hour time span is broken out into 4 different columns to allow the consumer of the data to choose the time type they prefer. This can then be used within a query or report to filter down the data based on the preferred time criteria.

Column definitions for the CORE_TIME_D table

Column Name	Purpose	Example
CORE_TIME_D_KEY	Primary key used to identify the unique time record	63221
FULL_TIME	The full time of row in normal time format	06:32:21.0000000
HOUR	Integer value of hour number in a day, 0-23	6
MINUTE	Integer value of minute number in a day, 0-59	32
SECOND	Integer value of hour number in a day, 0-59	21

CORE_USER_D

The user dimension provides user attribute information for all fact and throughput tables. As new users are detected in the ETL process job, new rows will be inserted.

Column definitions for the CORE_USER_D table

Column Name	Purpose	Example
CORE_USER_D_KEY	Primary key used to identify the unique user record	7
USER_ID	The user ID for the user in string format	321YH1B_0002WLLBF00003M
USER_NAME	The full user name in string format	John Doe
USER_LAST_NAME	The last name of the user in string format	Doe
USER_FIRST_NAME	The first name of the user in string format	John
IS_DELETED	A flag indicating if the user has been deleted from the system (1) or not (0)	0

Column Name	Purpose	Example
IS_ACTIVE	A flag indicating if the user is active in the system (1) or not (0)	1
NEEDS_PROCESSING	A flag indicating that the row needs further processing to gather more metadata from a different system such as LDAP or Perceptive Content.	1