Are there schema changes for student data in the RAE?

There are multiple possible answers to this, depending on what meaning of "schema" is intended.

Schema (definition 1) = what columns are in tables, and what are their datatypes

Schema (definition 2) = a namespace in a database (e.g. SYSADM, data_student_brz)

For definition 1:

There are some minor changes to table schemas that are unlikely to affect most use cases.

Data types:

Whenever possible, we have preserved column datatypes. In cases where an exact mapping was not possible between the two systems, we selected a slightly more permissive datatype. For example, the Oracle ODS uses a single datatype named "DATE" for both date and date+time fields which stores a timestamp down to the second. Redshift doesn't have an equivalent datatype with seconds-level precision, so we used the redshift TIMESTAMP datatype, which is precise down to the microsecond. It is important to note that the Oracle ODS stores all dates with a time component, if the time component is 00:00:00, the output is often displayed as a date only with no time. Redshift does not automatically truncate the time component in these cases, so applications that are assuming a date without a time may need to be adjusted.

Similar conversions:

NUMBER -> INT/BIGINT (if no decimal)

NUMBER -> DOUBLE PRECISION (if decimal)

VARCHAR2/NVARCHAR2 -> VARCHAR

CHAR/CHAR2/NCHAR -> VARCHAR

CLOB -> VARCHAR(4000)

Added/Removed Fields

While no data fields were removed, we have removed 5 metadata fields from many tables. These fields were populated by the old ODS batch load process and have been replaced with two new fields (see below).

Fields being removed:

SRC_SYS_ID: Set by ODS batch load to “HCM1” in every table

LOAD_ERROR: always set to ‘N’

DATA_ORIGIN: always set to ‘S’

CREATED_EW_DTTM: set to datetime of batch run when row was added

LASTUPD_EW_DTTM: set to datetime of batch run when row was last updated

BATCH_SID: always set to 0

New metadata fields:

ODS_PULL_DTTM: actual datetime the record was pulled from the source system

SIS_SYSTEM_DT: Logical date of the record in the source system, i.e. “as of end of day on”

Example:

A record pulled from the SIS into the RAE at 1:00a on 10-16-2024 would have an ods_pull_dttm = 2024-10-16 01:00:00 with SIS_SYSTEM_DT = 2024-10-15 00:00:00, indicating that for reporting purposes the data represents the state of the SIS at end of day on 10-15-2024. The current ODS load works the same way, but this “as of” date isn’t explicitly exposed anywhere.

For definition 2:

In the Oracle ODS, you may be aware that tables can be accessed from both the PUBLIC and SYSADM schemas. In redshift, you will need to use the schema “data_student_brz” to access the regular ODS tables.

In the Oracle ODS, a table could be referenced in three different ways:

SYSADM.PS_TERM_TBL -> data_student_brz.ps_term_tbl (note that in redshift unquoted table names are case-insensitive)

PUBLIC.PS_TERM_TBL -> data_student_brz.ps_term_tbl

PS_TERM_TBL (implicitly uses PUBLIC schema) -> data_student_brz.ps_term_tbl

If you would like reference the table without giving a schema name, as in example 3 you will need to run the following SQL immediately after logging on to redshift:

set search_path to data_student_brz;

The full list of schemas for student ODS data in the RAE:

data_student_brz: contains tables that used to be found in ODS/DWHCRPT, e.g PS_ACAD_PROG

data_student_hst: contains daily change records of most ODS tables, e.g. PS_ACAD_PROG_HST

data_student_slv: will contain pre-created data models to simplify common queries (not yet populated)

data_student_gld: contains enterprise level cumulative datasets, e.g. OUR_OSR_CUMCENSUS

data_student_ss: contains snapshot tables from DWDMOSU, e.g. SS_PS_ACAD_PROG

data_student_ld: RAE load staging tables (OTDI use only)

Note: Unlike the ODS, the RAE allows you to see the full list of tables in every schema, whether or not you have permissions to query the table.