Appendix: ReclamationHDB Data Store

Overview

The ReclamationHDB data store corresponds to Reclamation's HDB database (version 2), which stores real-time and historical data related to Reclamation operations. Time series in HDB are organized into "real" tables (observations) and "model" tables (from simulations or other external processes).

ReclamationHDB and Standard Time Series Properties

The standard time series identifier for ReclamationHDB time series is of the form:

```
Location.DataSource.DataType.Interval.Scenario~DataStoreName
```

More specifically, the identifier follows the following convention for "real" data:

```
Real:SiteCommonName.HDB.DataTypeCommonName.Interval~DataStoreName
```

and the following for "model" data:

```
Model:SiteCommonName.HDB.DataTypeCommonName.Interval.ModelName-
ModelRunName-HydrologicIndicator-ModelRunDate~DataStoreName
```

where identifier parts are described as follows:

- A location prefix of `Real:` indicates that data from the `R_` data tables will be read. A location prefix of `Model:` indicates that data from the `M_` data tables will be read. The model identifier is stored in the `Scenario` time series identifier field, described below.
- `SiteCommonName` is the identifier for a location and is taken from the `HDB_SITE.SITE_COMMON_NAME` database column.
- A data source of `HDB` is used in all cases. Although it may be possible to use the `R_BASE.AGEN_ID` value to determine an agency, these values vary by data record and may be confusing to use in the general identifier. The general guideline for HDB is that the last data in the table overwrites records. The final data for a site that receives data from multiple agencies is controlled by the Reclamation data load processes. A future enhancement might allow querying only the data records for a specific agency.
- `DataTypeCommonName` is the time series data type and is taken from the `HDB_DATATYPE.DATATYPE_COMMON_NAME` column. The site common name and the data type common name are used to look up the `HDB_SITE_DATATYPE.SITE_DATATYPE_ID` value, which is used internally to access real and model time series data.
- `Interval` is the data interval using TSTool standards (e.g., `Hour`, `Day`, `Month`, `Year`, `Irregular`). HDB does not include metadata for the time series interval. Instead, the data tables (e.g., `R_DAY`) must be queried to determine if a time series for a specific interval exists. This join occurs when interactively querying time series lists in TSTool to assist the user in forming valid time series identifiers (however, this does result in a slight performance penalty when listing time series).

- The scenario part of the time series identifier is omitted for "real" time series and for "model" time series contains dash-separated human-readable model metadata necessary to form a unique identifier, including:
  - o `ModelName` (from `HDB_MODEL.MODEL_NAME`)
  - o `ModelRunName` (from `REF_MODEL_RUN.MODEL_RUN_NAME`)
  - o `HydrologicIndicator` (from `REF_MODEL_RUN.HYDROLOGIC_INDICATOR`)
  - o `ModelRunDate` (from `REF_MODEL_RUN.RUN_DATE`), to the full second.

  This information is used to look up the `REF_MODEL_RUN.MODEL_RUN_ID`, which is used internally to access model time series data.
- `DataStoreName` is the user-defined data store name from the configuration information.

Limitations

ReclamationHDB data store limitations relative to TSTool standard features are as follows:

- Some database string values that are used in time series identifiers may contain periods, which conflict with the TSID conventions. These characters are converted to "?" in the TSID representation and are converted back to periods when querying the database. An alternative is to use the `ReadReclamationHDB()` command and assign an alias to time series, rather than relying on the TSID in commands.
- Although data flags are available with data records, they are not currently set in time series data during reads. How should these be represented since TSTool only allows one string for flags (not separate strings). Separate each flag component by a dash or some other character?
- HDB time series data tables have a `START_DATE_TIME` and `END_DATE_TIME` date/time for each data value, indicating the time span over which data are collected/averaged/summed. However, TSTool uses only a single date/time for each value, which is the same value shown in tables and used for graphing. The TSTool convention is that instantaneous values have a date/time corresponding to the date/time of the observation, and mean and accumulated values have a date/time corresponding to the interval end. For example, a 1-hour accumulated value recorded at `YYYY-MM-DD HH` would be for the hour ending `HH`. The following table illustrates the date/time conversions that are applied, using example data for site `ADATUNCO` flow volume (`HOUR`, `DAY`) and diversion volume (`MONTH`, `YEAR`). In summary, for `HOUR` interval, the `END_DATE_TIME` is used and for others the `START_DATE_TIME` is used. (Are the dates really as precise as shown? I'm using Squirrel to query Oracle and it may be using a generic representation that is more detailed than the database contents):

Data Interval	HDB START_DATE_TIME	HDB END_DATE_TIME	HDB POET date/time	TSTool date/time
HOUR	2007-04-04 13:00:00.0	2007-04-04 14:00:00.0	4/4/2007 2:00 PM	2007-04-04 14
DAY	2007-04-05 00:00:00.0	2007-04-06 00:00:00.0	4/5/2007	2007-04-05
MONTH	1978-01-01 00:00:00.0	1978-02-01 00:00:00.0	1/1/1978	1978-01
YEAR	1978-01-01 00:00:00.0	1979-01-01 00:00:00.0	1/1/1978	1978

  The above conversion works for regular interval mean and accumulated values but may not be correct for instantaneous values. Is this an issue or are the R_HOUR, R_DAY, R_MONTH, R_YEAR tables always mean/accumulated values?.
- TSTool supports 6Hour and other multiples of each base interval. However, HDB only stored 1Hour data. Therefore, 6Hour data may be represented in HDB as a value with 5 missing values for the other hours. TSTool can be used to convert the 1Hour data to 6Hour or values can be filled with interpolation, etc.

- Although HDB does have INSTANT data tables for instantaneous values, these time series are not yet supported in TSTool. If supported in the future, they will likely be implemented as irregular time series unless logic can be implemented to treat as minute interval data.
- TSTool does currently support reading HDB water year time series.
- TSTool does not currently support writing to the database. Need to understand more about this process, in particular whether TSTool will define time series metadata or just write data values.
- Performance for reading hourly time series is slow. This appears to be in the conversion of Oracle date/times to internal representations and is being evaluated.
- Database queries are case-sensitive – input filter criteria are not currently provided as selectable choices but this could be enabled.

Data Store Configuration File

A data store is configured by enabling ReclamationHDB data stores in the main *TSTool.cfg* configuration file, and creating a data store configuration file for each data store connection. Configurations are processed at software startup. An example of the TSTool configuration file is shown below. Multiple data stores can be defined using the [DataStore:DataStoreName] syntax. Properties for each data store are specified in an accompanying configuration file described below.

```
# Configuration file for TSTool

[TSTool]

ReclamationHDBEnabled = true

# Startup data stores (note that data store name in config file takes precedence)

[DataStore:HDB]

ConfigFile = "HDB.cfg"
```

TSTool Configuration File with ReclamationHDB Data Store Properties

The following illustrates the ReclamationHDB data store configuration file format, which in this example is located in the same folder as the TSTool configuration file and configures the "HDB" data store. A service account is used for authentication and allows read-only access to the database (need to complete authentication implementation to allow writing data).

```
# Configuration information for "HDB" data store (connection).
#
# The user will see the following when interacting with the data store:
#
# Name - database identifier for use in applications, for example as the
#     input type/name information for time series identifiers (usually a short string)
# Description - database description for reports and user interfaces (a sentence)
#
# The following are needed to make the low-level data connection:
#
# Type – the data store type (must be specified exactly as shown)
# DatabaseEngine - the database software (default to Oracle since not specified)
# DatabaseServer - IP or string address for database server
# DatabaseName - database name used by the server
# SystemLogin – service account login
# SystemPassword – service account password
#
# Use the syntax Env:EnvVarName to retrieve values from the environment.
# Use the syntax SysProp:SysPropName to retrieve values from the JRE system environment.

Type = "ReclamationHDBDataStore"
Name = "HDB"
Description = "Reclamation Test Database"
DatabaseServer = "xxx"
DatabaseName = "xxx"
SystemLogin = "xxx" (for example app_user)
SystemPassword = "xxx"
```

ReclamationHDB Data Store Configuration File