
Appendix: ReclamationHDB Datastore

2013-04-13

Overview

The ReclamationHDB datastore 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. The HDB design does include agency in several tables, but the information is not integrated enough to be used as part of the time series identifier:
 - The `HDB_AGEN` table lists agencies that are known in the database. `AGEN_NAME` is populated; however, `AGEN_ABBREV` often is null and cannot be relied upon for use in time series identifiers.
 - The `HDB_DATATYPE` table includes `AGEN_ID`, which relates to the `HDB_AGEN` table. However, many values are null and the relationship is not enforced.
 - The `R_BASE.AGEN_ID` value indicates the agency for which values are measured. The general guideline for HDB is that the last data in the table overwrites records. The final data record stored for a site that receives data from multiple agencies is controlled by the Reclamation data load processes. `TSTool` allows the agency to be specified when writing records, but there is currently no way to request records that were reported by an agency. The agency in this design is simply an annotation to explain the source of the data measurements.
- `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 time zone for hourly time series is set to the `TIME_ZONE` property from the HDB `REF_DB_PARAMETER` table.
- 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:
 - `ModelName` (from `HDB_MODEL.MODEL_NAME`)
 - `ModelRunName` (from `REF_MODEL_RUN.MODEL_RUN_NAME`)
 - `HydrologicIndicator` (from `REF_MODEL_RUN.HYDROLOGIC_INDICATOR`)
 - `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 datastore name from the configuration information.
- Missing numerical values are internally indicated with NaN.

Limitations

ReclamationHDB datastore 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 currently are not set in time series data during reads. A future enhancement may transfer the flags.
- 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.

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
IRREGULAR	1978-01-01 01:15:00.0	1978-01-01 01:15:00.0		1978-01-01 01:15

- 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 (or no hourly records for those hours). TSTool can be used to convert the 1Hour data to 6Hour or values can be filled with interpolation, etc.
- TSTool currently does not support reading HDB water year time series.
- 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.

Datastore Configuration File

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

```
# Configuration file for TSTool

[TSTool]

ReclamationHDBEnabled = true

# Startup datastores (note that datastore name in config file takes precedence)

[DataStore:HDB]

ConfigFile = "HDB.cfg"
```

TSTool Configuration File with ReclamationHDB Datastore Properties

The following illustrates the ReclamationHDB datastore configuration file format, which in this example is located in the same folder as the TSTool configuration file and configures the “HDB” datastore. Authentication for writing data to the datagbase is checked based on the account login and password.

```
# Configuration information for "HDB" datastore (connection).
#
# The user will see the following when interacting with the datastore:
#
# Type - ReclamationHDBDataStore (required as indicated)
# 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)
# Enabled - whether the datastore is enabled (default=True)
#
# The following are needed to make the low-level data connection:
#
# Type - the datastore 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
# Enabled - indicates whether datastore is enabled (True or False, default=True)
# KeepAliveSQL - SQL string to periodically run to keep database connection open
#               (useful when remotely accessing the system, default=no keep alive query)
# KeepAliveFrequency - number of seconds between execution of KeepAliveSQL
# 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 Datastore Configuration File