National Park Service

Inventory and Monitoring Division

Climate Database Schema, Table, Field Definitions and System Logic

Revision Date: 4 February 2013
Author: Brent Frakes; Simon Kingston

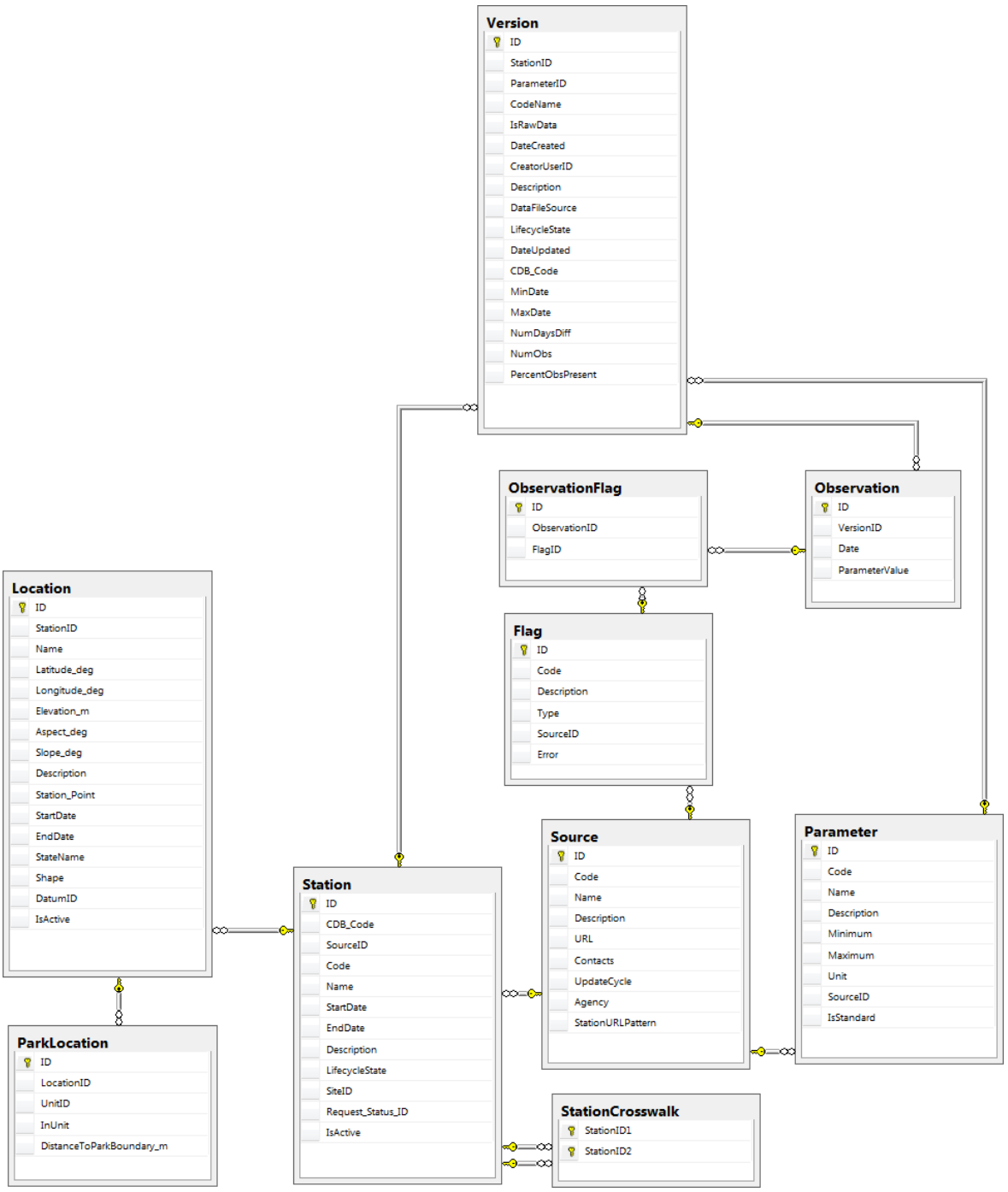
Table of Contents

1. Introduction

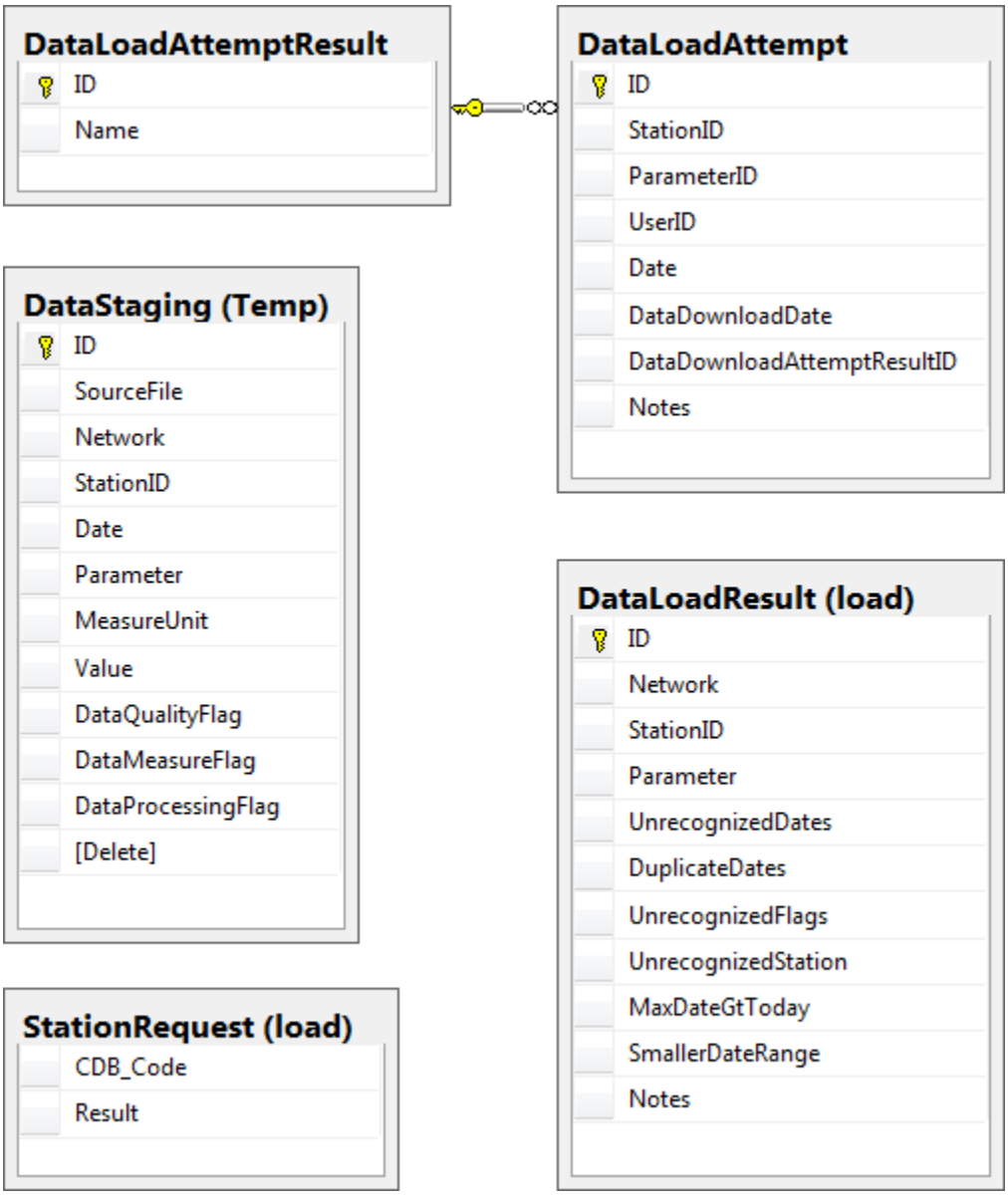
This document provides an overview of the database schema and includes both table and field definitions.

2. Database Schema

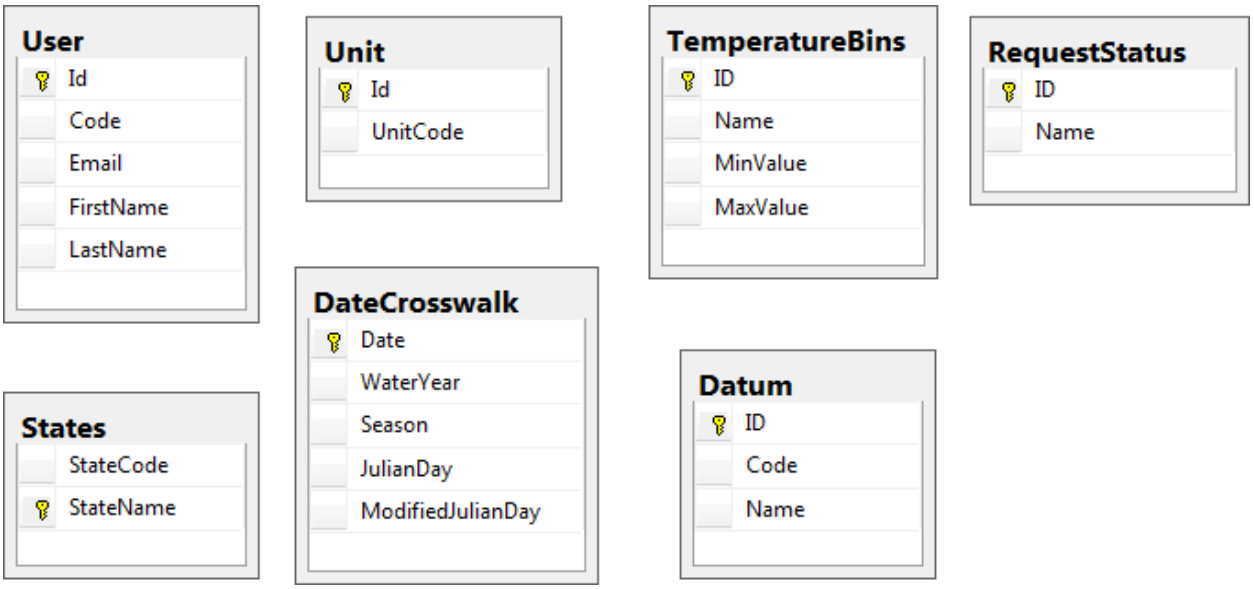
2.1. Main Tables



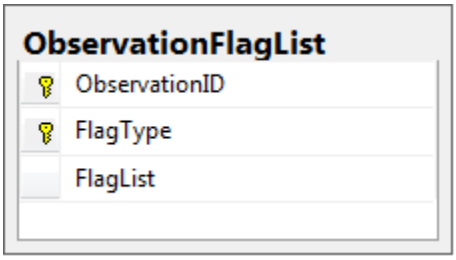
2.2. Data Load Tables



2.3. General Lookup Tables



2.4. Reporting Tables



3. Table and Field Definitions

3.1. Main Tables

These tables contain the primary information to compose climate observation records.

3.1.1. Source

Most weather observations are coordinated by a data provider who is the ultimate source of the weather/climate data. This table houses all of the source information.

Field	Type	Required?	Domain	Definition
ID	tinyint	Yes. PK		Unique identity
Code	varchar(50)	Yes. CK	N/A	Unique source code
Name	varchar(50)	Yes	N/A	The name of the data provider. In many cases, a data provider may have more than one station. An example is "COOP".
Description	varchar(50)	Yes	N/A	Description of the data provider. Indicate who they are beyond what the name would suggest.
URL	varchar(100)	No	N/A	If there is a URL for the data provider's web site, indicate that here.
Contacts	varchar(50)	No	N/A	Many data providers will likely have a point of contact, either as a contact name and phone number or perhaps as a URL.
UpdateCycle	varchar(50)	No	N/A	This indicates how often the observed data is updated by the provider. For example, some providers do annual updates while others update the observations monthly. You can approximate the times if you are not sure.
Agency	varchar(50)	Yes	N/A	In a few cases, the data provider may be identified within a larger agency. For example, the data provider may be the climate group of a larger entity known as the agency. An example is "NWS" for the National Weather Service.
StationURLPattern	varchar(255)	No	N/A	Pattern for getting to station information from the source using a URL, substituting station properties into the address.

3.1.2. Station

This table contains all of the unique information for a particular weather station.

Field	Type	Required?	Domain	Definition
ID	int	Yes. PK		Unique identity
CDB_Code	varchar(20)	(Yes). CK	N/A	This is the unique code given to station when loaded to the database. It is filled on record creation.
SourceID	tinyint	Yes. FK		Code for entity who is the data provider for the climate data.
Code	varchar(50)	Yes		Code given to the station by the data provider.
Name	varchar(255)	No	N/A	This is the full station name given by the data provider. In cases where there is not a name, staff submitting the data are encouraged to create a meaningful name of no more than 255 characters (spaces are allowed).
StartDate	date	No	Between 1/1/1700 and	Known start date when station began recording (this may be different than the available data).

			today. Times are ignored.	
EndDate	date	No	Between 1/1/1700 and today. Times are ignored.	Known end date when station ended recording (this may be different than the available data). Between 1/1/1700 and today. Use 12/31/9999 if still active.
Description	nvarchar(max)	No	N/A	This is a more thorough description of the station. When possible, use the description used by the data provider or create your own.
LifecycleState	varchar(50)	No	Active, Inactive	Indicates whether a station is active, as of this current date, actively collecting climate data
SiteID	varchar(10)	No		Alternate identifier given to the station by the data provider.
Request_Status_ID	tinyint	Yes. FK	RequestStatus.ID	Indicates whether it is possible for a user to request the IMD automate the pulling and loading of the data to the database. See RequestStatus table.
IsActive	bit	No		1 if the station is still active at this location, 0 if it is not, null if indeterminate.

3.1.3. Station Crosswalk

This table is used to indicate when two Station table records are actually the same station.

Field	Type	Required?	Domain	Definition
StationID1	int	Yes. PK, FK	Station.ID	Unique identifier from Station table ID column.
StationID2	int	Yes. PK, FK	Station.ID	Unique identifier from Station table ID column.

3.1.4. Location

A single station can, over time, be located at one or more locations. This table provides information on each of those locations.

Field	Type	Required?	Domain	Definition
ID	int	Yes. PK.		Unique identity
StationID	int	Yes. FK	Station.ID	The station that this location describes.
Name	varchar(50)	Yes	N/A	Name of the station location. If there is no name provided by the data provider, then create a simple descriptive name.
Latitude_deg	float	Yes	-90 to 90	Latitude in decimal degrees. Points in the northern hemisphere are positive; those in the southern hemisphere are negative(e.g., 40.75).
Longitude_deg	float	Yes	-180 to 180	Longitude in decimal degrees. Points in the western hemisphere are negative; those in the eastern hemisphere are positive (e.g., -100.82).
Elevation_m	float	No	-86.0 to 6194.0	Elevation of the station measured in meters. If the station is not measured in meters, you will need to

				convert it.
Aspect_deg	smallint	No	0 to 359	Aspect is the direction of the slope where the station is located. North facing slopes have an aspect of 0. If the aspect is not measured in degrees, you will need to convert it.
Slope_deg	tinyint	No	0 to 90	Slope is the gradient of the slope where the station is located. A zero degree slope indicates a completely flat surface. If the slope is not measured in degrees, you will need to convert it.
Description	varchar(max)	No	N/A	Description of the specific station location. Ideally, location description distributed by the data provider. Otherwise, create your own description.
Station_Point	geometry	No		The location of the station as a geographic feature
StartDate	date	No	Between 1/1/1700 and today	This is the date the station began recording data at this particular location. If station didn't move this is the length of record from "StationInfo"
EndDate	date	No	Between 1/1/1700 and today	Use 12/31/9999 if still active location.
StateName	varchar(50)	No	US State Names or other national political boundaries	All full state names or canadian province names
Shape	geography	No		
DatumID	tinyint	No, FK	Datum.ID	The datum for the coordinates.
IsActive	bit	No		1 if the station is still active at this location, 0 if it is not, null if indeterminate.

3.1.5. ParkLocation

This table describes how a particular station relates to a specific park.

Field	Type	Required?	Domain	Definition
ID	bigint	Yes. PK		Unique identity
LocationID	int	Yes. FK	Location.ID	Location of the station
UnitID	smallint	Yes. FK	Unit.ID	Distinct NPS Unitcodes for which this station should be affiliated. Generally, this is a 4-letter code (e.g., ROMO).
InUnit	bit	No	True or False	Indicate whether the stationlocation is within the park boundary.
DistanceToParkBoundary_m	smallint	No	Between 0 and 32,767	If not within the park, indicates the distance to the park boundary in meters.

3.1.6. Parameter

Each observation pertains to a particular parameter, whether it is minimum temperature, precipitation or river flow. Each parameter is identified by a parameter code, which includes the units of measure.

Field	Type	Required?	Domain	Definition

ID	tinyint	Yes. PK		Unique identity
Code	varchar(50)	Yes	N/A	The parameter code used by the data provider. In general, we recommend that you do not modify the parameter values if you are getting information from a third party. If the provider does not use codes, you will need to create one.
Name	varchar(50)	Yes	N/A	Name of the parameter being measured
Description	varchar(500)	Yes	N/A	Text description of what parameter is being measured
Minimum	real	Yes		Minimum range of acceptable values for this particular parameter
Maximum	real	Yes		Maximum range of acceptable values for this particular parameter
Unit	varchar(50)	Yes	N/A	The measurement unit described as precisely as possible
SourceID	tinyint	Yes. FK	Source.ID	The data source which provided this parameter
IsStandard	bit	No		

3.1.7. Version

A version is defined as a grouping for a series of observations for a given parameter and station.

Field	Type	Required?	Domain	Definition
ID	int	Yes		Unique identity
StationID	int	Yes, FK	Station.ID	The station having observations.
ParameterID	tinyint	Yes, FK	Parameter.ID	The parameter (e.g., minimum temperature F, maximum temperature F, etc.) observed at the station.
CodeName	varchar(50)	Yes		Name created from data file used to load data and parameter.
IsRawData	bit	Yes	0/1	Indicates whether the data was loaded directly from the authoritative data source or if it was provided by a third party. A value of 1 indicates a direct load.
DateCreated	datetime	Yes		The date and time the record was created
CreatorUserID	tinyint	Yes, FK	User.ID	The individual creating the record.
Description	varchar(max)	Yes		For non-authoritative source data, the description provides a complete lineage of how the version was obtained and/or processed. For authoritative data, it indicates file used to create data and who processed data.
DataSourceFile	varchar(255)	No		Provides the name of the original text file used to load the database.
LifecycleState	varchar(50)	Yes	Provisional, Final	Indicates the lifecycle state of the version
DateUpdated	datetime	No		Date the version was loaded to the database
CDB_Code	varchar(20)	No		Auto-generated code based on the ID value for new records.
MinDate	date	No		Calculated value that indicates the earliest observation date in the Observation table for this version.
MaxDate	date	No		Calculated value that indicates the latest observation date in the Observation table for this version.
NumDaysDiff	int	No		Calculated value of the number of days between the first observation date and last observation date for this version.
NumObs	int	No		Calculated value of the actual number of observations for this version.
PercentObsPresent	float	No		Calculated percent of possible observations that are

				actually present in the Observation table.

3.1.8. Flag

This table is a lookup of all of the possible flags for an observation. Flags can be related to measurements, quality and or processing.

Field	Type	Required?	Domain	Definition
ID	smallint	Yes. PK		Unique identity
Code	varchar(50)	Yes		A code representing the flag. Flags should generally not be more than 4 characters.
Description	varchar(max)	Yes		Full description of the flag
Type	varchar(50)	Yes	Data Measure Flag, Data Quality Flag, Data Processing Flag	Measure – Pertaining to the actual collection of the data by the source provider. Includes whether there were issues with the instrumentation or other associated atmospheric events. Examples include there was a severe storm which damaged some of the instrumentation or caused them to go offline. Quality – Pertaining to any quality control applied to the data following collection. Quality control may include modifications to the original observations by the source provider Processing – Pertaining to the processing of the data by the NPS Inventory and Monitoring Division during loading of the data.
SourceID	tinyint	Yes. FK	Source.ID	ID from Source table. Indicates which source defined the flag.
Error	bit	No	1,0	Indicates whether the source considers the flag an error.

3.1.9. Observation

This table houses all of the individual weather/climate observations.

Field	Type	Required?	Domain	Definition
ID	bigint	Yes. PK	N/A	Unique identity
VersionID	int	Yes. FK	Version.ID	VersionID from the Version table
Date	date	Yes	N/A	The date of the weather/climate observation
ParameterValue	real	No		The observed value for a given parameter

3.1.10. ObservationFlag

Contains all flags specific to an single observation

Field	Type	Required?	Domain	Definition
ID	bigint	Yes. PK		Unique identity
ObservationID	bigint	Yes. FK	Observation.ID	The observation to which the flag pertains.
FlagID	smallint	Yes. FK	Flag.ID	The specific flag

3.2. Data Load Tables

These tables are used when records are loaded to the database.

3.2.1. DataLoadAttempt

This table contains all information about attempted data loads.

Field	Type	Required?	Domain	Definition
ID	int	Yes. PK		Unique identity
StationID	int	No. FK	Station.ID	Station for which data was trying to be loaded.
ParameterID	tinyint	No. FK	Parameter.iD	Parameter for the observations that were trying to be loaded.
UserID	tinyint	Yes.FK	User.ID	ID of the User requesting the data load
Date	datetime	Yes		Date/time the user requested the data be loaded
DataDownloadDate	date	No		Date the attempt was made to download the data
DataDownloadAttemptResultID	tinyint	Yes. FK	DataLoadAttemptResult.ID	Result of the attempt to load data.
Notes	varchar(255)	No		Notes specific to the download attempt

3.2.2. DataLoadAttemptResult

Lookup table of data load attempt result values that describe success or failure to load data into the climate database.

Field	Type	Required?	Domain	Definition
ID	tinyint	Yes. PK		Unique identity
Name	varchar(50)	Yes	Success Failure – Data downloaded, but has errors Failure – No daily surface data available Failure – No data available online Failure – Station not found	Short description of the reason for success or failure at loading data.

3.2.1. DataLoadResult

Audit table for loading climate data to production tables.

Field	Type	Required?	Domain	Definition
ID	int	Yes. PK		Unique identity
Network	varchar(50)	Yes		Combination of Agency-Code from source.
StationID	varchar(50)	Yes		Code for the station from the source.
Parameter	varchar(50)	Yes		Code from Parameter table.
UnrecognizedDates	bit	No	0,1	1 if observation records have date column values that aren't

				recognizable as legitimate dates, 0 or null otherwise.
DuplicateDates	bit	No	0,1	1 if the observation set has duplicate dates, 0 or null otherwise.
UnrecognizedFlags	bit	No	0,1	1 if there are unrecognized flags, 0 or null otherwise.
UnrecognizedStation	bit	No	0,1	1 if there is an unrecognized station code, 0 or null otherwise.
MaxDateGtToday	bit	No	0,1	1 if the latest observation date is greater than today's date, 0 or null otherwise.
SmallerDateRange	bit	No	0,1	1 if the date range of the observations is smaller than the currently loaded data for that station and parameter, 0 or null otherwise.
Notes	varchar(max)	No	0,1	Additional notes about the attempt to load the data.

3.2.1. DataStaging

Main staging table for loading climate data to production tables. Data are first loaded to this table from text files and then quality assurance procedures are completed before loading to production tables.

Field	Type	Required?	Domain	Definition
ID	int	Yes. PK		Unique identity
SourceFile	varchar(255)	Yes	N/A	Name of the text file from which data is being loaded.
Network	varchar(50)	Yes	N/A	Combination of Agency-Code from source.
StationID	varchar(50)	Yes	N/A	Code for the station from the source.
Date	varchar(50)	No	N/A	Observation date from the source.
Parameter	varchar(50)	Yes	N/A	Parameter code from the source.
MeasureUnit	varchar(50)	Yes	N/A	Unit of measure from the source.
Value	varchar(50)	No	N/A	Measured value for the observation from the source.
DataQualityFlag	varchar(50)	No	N/A	Data quality flag(s) from the source.
DataMeasureFlag	varchar(50)	No	N/A	Data measure flag(s) from the source.
DataProcessingFlag	varchar(50)	No	N/A	Data processing flag(s) from NRSS data processing code.
Delete	char(1)	No	N/A	Used during data loading; default = 'n'

3.2.1. StationRequest

Staging table for loading station request lists to update which stations users have asked NRSS to acquire observations for.

Field	Type	Required?	Domain	Definition
CDB_Code	varchar(20)	No	Station.CDB_Code	Indicates which station is requested.
Result	varchar(50)	No	RequestStatus.Name	Indicates whether a station is requested or not.

3.3. General Lookup Tables

These tables serve as general lookups.

3.3.1. Request Status

Lookup table containing the request status codes.

Field	Type	Required?	Domain	Definition
ID	tinyint	Yes. PK		Unique identity.
Name	varchar(50)	Yes. CK		'Can Request' – Indicates that a network can request data for a given status. By default, all national-level stations are given a 'Can Request' status. 'Is Requested' – Indicates that one or more networks have requested that IMD actively synhcronize data aqusition for this station 'Data Not Available' – Indicates that it is not possible for IMD to actively obtain data for this station. This status does not preclude a network from obtaining the data from alternate sources. 'No Longer Supported' – Indicates that IMD no longer supports the acquisition of data from this source

3.3.2. TemperatureBins

This table is a lookup that defines the temperature bins (groups) used by functions that group temperature values.

Field	Type	Required?	Domain	Definition
ID	tinyint	Yes. PK		Unique identity.
Name	varchar(50)	Yes		Name of the range of the bin.
MinValue	real	No		Minimum value included in the bin.
MaxValue	real	No		Maximum value included in the bin.

3.3.3. Unit

This table contains a list of all NPS park units.

Field	Type	Required?	Domain	Definition
ID	smallint	Yes, PK		Unique identity.
UnitCode	varchar(10)	Yes	NPS Park Unit Codes	The accepted NPS unit code as defined by NPS NRInfo Unit Service (See https://nrinfo.nps.gov/Unit.mvc/Search).

3.3.4. User

This table records all users of the database. Users generally pertain to the technical support staff (i.e., those doing the data loading and management) as opposed to those using the database.

Field	Type	Required?	Domain	Definition
ID	tinyint	Yes, PK		Unique identity.
Code	varchar(50)	Yes	NPS AD name OR email address for non-NPS users	Unique code for each staff member.

Email	varchar(50)	Yes		Valid email address at time the staff was working.
FirstName	varchar(20)	Yes		First name of staff member
LastName	varchar(20)	Yes		Last name of staff member

3.3.5. States

This table contains a list of all states in the United States.

Field	Type	Required?	Domain	Definition
StateCode	varchar(50)	No		Code for each state or territory.
StateName	varchar(50)	Yes, PK		Name of each state or territory.

3.3.6. DateCrosswalk

Lookup of all dates covered by the range of data in the climate database. Used for many calculations.

Field	Type	Required?	Domain	Definition
Date	date	Yes, PK	Valid dates	Date of an observation in the climate database.
WaterYear	smallint	computed		Water year of the Date.
Season	varchar(10)	computed		Season of the Date.
JulianDay	int	computed		Julian Day of the Date.
ModifiedJulianDay	smallint	computed		Julian Day starting on Mar.1, so that in leap years, each date has the same number as in a non-leap year.

3.3.7. Datum

Lookup table of standardized datums.

Field	Type	Required?	Domain	Definition
ID	tinyint	Yes, PK		Unique identity
Code	varchar(5)	Yes		Standard code for datum.
Name	varchar(50)	Yes		Standard name for datum.

3.4. Reporting Tables

These tables were created to make reporting more efficient. They are denormalized in that they repeat information from other tables.

3.4.1. ObservationFlagList

Not currently in use. This table is designed to contain comma-delimited lists of flags by flag type for each observation.

Field	Type	Required?	Domain	Definition
ObservationID	bigint	Yes. PK	Observation.ID	Unique identifier from Observation table.
FlagType	varchar(50)	Yes, PK	Flag.FlagType	Flag Type from Flag table.
FlagList	varchar(50)	Yes		Comma-delimited list of flags from ObservationFlag table.

4. System Logic

4.1. Quality Control of Loaded Station Data

When station data is loaded to the database, the following logical checks occur:

- Records with unrecognized dates are removed
- Records with non-numeric values have a Data Processing flag of “N” added and the value is cleared
- Records with values that exceed thresholds (except for 99999 or -999999) for a particular parameter have a Data Processing flag of “V” added
- Duplicates on station, parameter, and date
- If values are different, clear values and give flags and give a Data Processing flag of “Z”
- If values are same, but flags are different, keep one record with all flags and add a Data Processing flag of “Z”
- If complete duplicates, keep one and add a Data Processing flag of “D”
- Warning if unrecognized flags – those records not loaded
- Warning if unrecognized stations – those records not loaded
- Warning if date greater than today’s date – those records not loaded
- Warning if date range of replacement data is smaller than date range of data being replaced
- Warning if unrecognized parameters