Star Catalog File Ingestion into MySQL Databases

Version 1.1

5 October 2023

Author:

Cole Odegard

Description

Ingest star catalog files into SQL databases and update any null values or formatting issues.

Deliverables

- MySQL Databases
  - SAO
  - UCAC
  - GSC
- Catalog Descriptions including SQL Schema
  - SAO
  - UCAC
  - GSC
- *Plots showing sky coverage (*supplemental not necessary to project success).

SAO Catalog

Catalog Specifications

The SAO2000 catalog will be ingested into two separate tables, sao1950 and sao2000. Sao1950 contains supplemental details such as flags and 1950's data while the sao2000 table contains data from 2000's measurements. In addition, the following requirements will be followed for database ingestion.

- Keep consistent UNITS.
- Confirm proper motion in "/yr.
- RA and DEC in both sexagesimal and degrees (6 sig figures)
  - this is to avoid the need for translation later.
- watch for nulls, often defined as a number but referenced elsewhere.
  - Set null floats for values like magnitude and proper motion.
  - Remove any null coordinates.
- Add precessed coordinates column to J2020 or J2023
- Look at how to read the file and ingest into DB.
- Look into providing search indexes to improve performance.

Tables

sao2000			
ColumnName	**Datatype**	**Units**	**NullValues**
SaoNumber	INT	id	-
RA	VARCHAR(10)	sexag	-
PMRA	FLOAT	mas/yr	Blank
Decl	VARCHAR(11)	sexag	-
PMDec	FLOAT	mas/yr	Blank
RA_rad	FLOAT	rad	-
Dec_rad	FLOAT	rad	-
RA_deg	FLOAT	deg	-
Dec_deg	FLOAT	deg	-
PhotMag	FLOAT	mag	99.9
VMag	FLOAT	mag	99.9
SpectralType	VARCHAR(3)	-	-
VMag_delta	INT	-	-
PhotMag_delta	INT	-	-

sao1950			
ColumnName	**Datatype**	**Units**	**NullValues**
SaoNumber	INT	id	-
Dup	VARCHAR(1)	-	Blank
RA1950	VARCHAR(10)	sexag	-
PMRA_1950	FLOAT	mas/yr	Blank
PMRA_1950mu	FLOAT	µas/yr	-
RA2m_flag	VARCHAR(1)	-	Blank
RA1950_precessed	FLOAT	-	-
RA1950_precessed_sd	FLOAT	-	-
Original_Epoch	FLOAT	yr	-
Dec-50	VARCHAR(11)	sexag	-
PMDec_1950	FLOAT	mas/yr	Blank
PMDec_1950mu	FLOAT	µas/yr	-
D2m_Flag	VARCHAR(1)	-	Blank
DE2s	FLOAT	-	-
e_DE2	FLOAT	-	-
Dec_orig_epoch	FLOAT	yr	-
e_Pos	FLOAT	-	-
VMag_src	INT	-	99.9
StarNum_src	INT	-	-
PhotMag_src	INT	-	99.9
PM_src	INT	-	-
SpecType_src	INT	-	-
Rem	INT	-	-
SrcCatCode	INT	-	-
SrcCatNum	INT	-	-
DurchmusterungID	VARCHAR(14)	-	-
HenryDraperCatNum	VARCHAR(6)	-	-
HDDuplicateID	VARCHAR(1)	-	-
GeneralCatalogNumber1950	VARCHAR(5)	-	-
RA1950_rad	FLOAT	rad	-
Dec1950_rad	FLOAT	rad	-

Figures 1.1 & 1.2: Name, SQL Datatype, Units and Values to be replaced with NULL for the sao2000 and sao1950 tables.

Database Implementation

- The sao.dat file contains all the columns necessary to implement the catalog specific changes and create the database.
- First, each line in the file was split by byte into the defined columns in the readme document.
- Next, each column was assigned a SQL datatype (defined in the sao1950 and sao2000 tables above), and proper units for each column were defined.
- For consistency, proper motion, RA, and Declination units were not changed.
- Finally, the catalog was updated to include NULL instead of blank values for magnitudes, and the three stars that did not have a proper motion value defined were given proper motion values of zero.
- The final specification of the Sao catalog was to explore adding precession and creating an RA and Declination coordinate for 2020. This was not implemented due to the variance in Proper Motion over time.

Data Cleaning

- RA: Right Ascension in sexagesimal constructed from degree values
- Decl: Declination in sexagesimal constructed from degree values.
- Dec_deg: Converted Decl to degrees
- RA_deg: Converted RA to degrees
- RA_1950: Right Ascension in sexagesimal constructed from degree values.
- Dec_1950: Declination in sexagesimal constructed from degree values.
- Columns where NULL values were added:

- PMDec
- PMRA
- Dup
- RA2m_Flag
- D2m_Flag
- HDDuplicateID
- PMRA_1950
- PMDec_1950

UCAC4 Catalog

Catalog Specifications

- Keep consistent units.
- Confirm proper motion in "/yr.
- RA and DEC in both sexagesimal, radians, and degrees (6 sig figures)
  - this is to avoid the need for translation later.
- watch for nulls, often defined as a number but referenced elsewhere.
  - Set null floats for values like magnitude and proper motion.
  - Remove any null coordinates.
- Look at how to read the file and ingest into DB.
  - How is the file indexed, data split between files, etc.
- Determine the Schema for the columns from the readme file.
- Adapt Sao2000 code to be more flexible.
- Four tables: ucac4 and ucac4_errors_flags & ucac4_not_visible and ucac4_errors_flags_not_visible
  - _not_visible: Not visible from Keck Observatory Declination < -70°
- Check database vs existing file query structure.

Tables

ucac4

ColumnName	Datatype	Units	NullValues
UCAC_ID	INT	-	-
2MASS_ID	INT	-	-
RA	VARCHAR	sexag	-
Decl	VARCHAR	sexag	-
RA_deg	FLOAT	deg	-
Decl_deg	FLOAT	deg	-
RA_orig	INT	-	-
Decl_orig	INT	-	-
MagModel	FLOAT	mag	-
MagApperature	FLOAT	mag	-
Objt	INT	-	-
Cdf	INT	-	-
SigRA	INT	-	-
SigDec	INT	-	-
CepRA	INT	-	-
CepDec	INT	-	-
PmRA	FLOAT	mas/yr	-
PmDec	FLOAT	mas/yr	-
SigPmRA	INT	-	-
SigPmDec	INT	-	-
2MASS_J	FLOAT	mag	20
2MASS_H	FLOAT	mag	20
2MASS_K	FLOAT	mag	20
APASS_B	FLOAT	mag	20
APASS_V	FLOAT	mag	20
APASS_g	FLOAT	mag	20
APASS_r	FLOAT	mag	20
APASS_i	FLOAT	mag	20

ucac4_errors_flags

ColumnName	Datatype	Units	NullValues
UCAC_ID	INT	id	-
SigMag	INT	-	-
Na1	INT	-	-
Nu1	INT	-	-
Cu1	INT	-	-
icqflg_J	INT	-	-
icqflg_H	INT	-	-
icqflg_K	INT	-	-
e2mpho_J	INT	-	-
e2mpho_H	INT	-	-
e2mpho_K	INT	-	-
APASS_B_err	INT	-	99
APASS_V_err	INT	-	99
APASS_g_err	INT	-	99
APASS_r_err	INT	-	99
APASS_i_err	INT	-	99
gcflg	INT	-	-
icf	INT	-	-
leda	INT	-	-
x2m	INT	-	-
zn2	INT	-	-
rn2	INT	-	-

Figures 2.1 & 2.2: Name, SQL Datatype, Units and Values to be replaced with NULL for the ucac4 and ucac4_errors_flags tables.

Database Implementation

- The UCAC4 database was constructed from nine hundred zone files, each corresponding to a 0.2-degree wide declination zone in the sky.
- Each file had anywhere between a few hundred and a few hundred thousand stars, with the data for each star stored in byte format and manipulated to meet certain storage criteria.
- UCAC4 requires more cleaning to acquire the desired database column format.
- These longer formats can be implemented, for example, changing RA from an integer to sexagesimal format, due to the increased performance of MySQL when compared to a simple file search.

Data Cleaning

- RA (sexagesimal) was constructed from the RA_deg column.
- RA_deg (degrees) was constructed from the original RA column.
- Decl (sexagesimal) was constructed from the Decl_deg column.
- Decl_deg (degrees) was constructed from the original Decl column.
- All magnitudes were divided by one thousand to get units of magnitude and not milimag and updated with NULL values.
    - MagModel
    - MagAperture
    - 2MASS_J
    - 2MASS_H
    - 2MASS_K
    - APASS_B
    - APASS_V
    - APASS_g
    - APASS_r
    - APASS_i
- CepRA and CepDec, the epoch years for RA and Dec, were divided by one thousand and added to 1900 to get the epoch in years. Epoch was originally stored as a fraction of years before or after 1900.
- APASS and 2MASS values were changed to NULL if the star did not have a reference in either one of those catalogs.

GSC240 Catalog

Catalog Specifications

- Keep consistent units.
- Confirm proper motion in "/yr.
- RA and DEC in both sexagesimal, radians, and degrees (6 sig figures)
    - this is to avoid the need for translation later.
- watch for nulls, often defined as a number but referenced elsewhere.
    - Set null floats for values like magnitude and proper motion.
    - Remove any null coordinates.
- Look at how to read the file and ingest into DB.
    - How is the file indexed, data split between files, etc.
- Determine the Schema for the columns from the readme file.
- Four tables: gsc240 and gsc240_errors_flags & gsc240_not_visible and gsc240_errors_flags_not_visible
    - _not_visible: Not visible from Keck Observatory Declination < -70°
- Check database vs existing file query structure.

Tables

gsc240			
ColumnName	Datatype	Units	NullValues
GSCid	INT	id	-
GSC1id	VARCHAR(11)	id	-
HSTid	VARCHAR(11)	id	-
RA	VARCHAR(13)	sexag	-
DECL	VARCHAR(13)	sexag	-
RA_rad	DOUBLE	rad	-
Decl_rad	DOUBLE	rad	-
RA_deg	DOUBLE	deg	-
Decl_deg	DOUBLE	deg	-
Original_Epoch	REAL	yr	-
RA_eps	REAL	arcsec	-
Decl_eps	REAL	arcsec	-
PmRA	REAL	mas/yr	99.9
PmDec	REAL	mas/yr	99.9
Delta_Epoch	REAL	yr	-
FpgMag	REAL	mag	99.9
JpgMag	REAL	mag	99.9
VMag	REAL	mag	99.9
NpgMag	REAL	mag	99.9
UMag	REAL	mag	99.9
BMag	REAL	mag	99.9
RMag	REAL	mag	99.9
IMag	REAL	mag	99.9
JMag	REAL	mag	99.9
HMag	REAL	mag	99.9
KMag	REAL	mag	99.9
Classification	INT	-	-
SemiMajorAxis	REAL	-	-
Eccentricity	REAL	-	-
PositionAngle	REAL	deg	-
SourceStatus	INT	-	-

gsc240_errors_flags			
ColumnName	Datatype	Units	NullValues
GSCid	INT	id	-
GSC1id	VARCHAR(11)	id	-
HSTid	VARCHAR(11)	id	-
PmRA_mu	REAL	mas/yr	99.9
PmDec_mu	REAL	mas/yr	99.9
FpgMag_err	REAL	mag	99.9
FpgMag_code	INT	-	99
JpgMag_err	REAL	mag	99.9
JpgMag_code	INT	-	99
VMag_err	REAL	mag	99.9
VMag_code	INT	-	99
NpgMag_err	REAL	mag	99.9
NpgMag_code	INT	-	99
UMag_err	REAL	mag	99.9
UMag_code	INT	-	99
BMag_err	REAL	mag	99.9
BMag_code	INT	-	99
RMag_err	REAL	mag	99.9
RMag_code	INT	-	99
IMag_err	REAL	mag	99.9
IMag_code	INT	-	99
JMag_err	REAL	mag	99.9
JMag_code	INT	-	99
HMag_err	REAL	mag	99.9
HMag_code	INT	-	99
KMag_err	REAL	mag	99.9
KMag_code	INT	-	99
VariableFlag	INT	-	-
MultipleFlag	INT	-	-

Figures 3.1 & 3.2: Name, SQL Datatype, Units and Values to be replaced with NULL for the gsc240 and gsc240_errors_flags tables.

Database Implementation

- The GSC240 database was constructed from ~648,000 zone files, each corresponding to a 0.1 by 1-degree zone in the sky.
- Some files have zero stars and are not included in the catalog's csv file.
- Files are stored in nested folders with the following folder structure:
    - *Degree of Dec*/*Decimal Degree of Dec*/*Degree of RA*
    - Ex: 000/0000/001 is the file corresponding to the -90 to -89.9° Dec by 0° to 1° RA.

Data Cleaning

- RA (sexagesimal) was constructed from the RA_deg column.
- RA_deg (degrees) was constructed from the original RA column.
- Decl (sexagesimal) was constructed from the Decl_deg column.
- Decl_deg (degrees) was constructed from the original Decl column.
- All magnitudes were divided by one thousand to get units of magnitude and not milimag and updated with NULL values.
    - MagModel
    - MagAperture
    - 2MASS_J
    - 2MASS_H
    - 2MASS_K
    - APASS_B

    - APASS_V
    - APASS_g
    - APASS_r
    - APASS_i
- CepRA and CepDec, the epoch years for RA and Dec, were divided by one thousand and added to 1900 to get the epoch in years. Epoch was originally stored as a fraction of years before or after 1900.
- APASS and 2MASS values were changed to NULL if the star did not have a reference in either one of those catalogs.

Project Terms & Calculations

Degrees to Sexagesimal

Whole units of degrees ° 60* remaining decimal *minutes* ‘ 60* remaining decimal *seconds* “

Ex: 123.219 degrees becomes 123°13’8.4” 123 (.219*60) (.14*60)

Sexagesimal to Degrees

First number is your degree whole number. Second number / 60 gets added onto the end and third number / 3600 gets added onto that.

Ex: 23°10’5” becomes 23.168056 (23*1) + (10/60) + (5/3600)

Celestial coordinate system

The Horizon System

- Local to the observer
- Azimuth: The longitudinal equivalent (0 to 360 degrees w due north at 0)
- Altitude: The latitudinal equivalent (-90 to 90 degrees w horizon at 0)

The Equatorial System

- Declination: The Latitudinal equivalent (-90 to 90 degrees with each being celestial poles)
- Right ascension: The longitudinal equivalent (0 to 24 hours with 24 hours being 360 degrees) Each hour is 15 degrees; each minute is $1/60^{th}$ of an hour and each second is $1/60^{th}$ of a minute.

Axial Precession

- The term used to describe the earth’s axis rotating around an axis itself.

Perihelion

- The point when earth is closest to the sun.
- When Perihelion lines up with earth’s axial precession towards the sun the northern hemisphere gets more extreme seasons. The distance between earth from the sun at perihelion and at aphelion is only 3%.
- Extreme seasons are much more apparent in the northern hemisphere because the northern hemisphere is only 60% water while the southern hemisphere is 80% water. Water makes climate much more temperate and thus less affected by distance to the sun or orientation to or from the sun.

Eccentricity

- Amount something deviates from a perfect circle.
- Earth’s orbit is 0.005-0.057 and currently 0.0167.

Obliquity

- Angular difference between earth’s rotational axis and the plane of its orbit around the sun
- Varies between 22.1 and 24.5 degrees and is currently 23.44 degrees.