

Managing Large Data Sets to Support GIS

Paul Klebe, NDPSC.

The NOAA Dataset. What is in it?

- Contains a 19th & 20th Century Version
- Contains a 21st Century Version
- Has a lot of built in functionality; but it was not used to its full potential
- Is not a relational database
- Extends far beyond the time frame of modern computers
- Even when limited to North Dakota, this is a big set when compared to what the ND PSC is used to using

19th & 20th Century Version

- Has Y2K issues
- Conceptual table logic is 31 days of a METEOROLOGICAL_ELEMENT per record.
- ALL fields are character
- Uses a ELEMENT_UNIT_CODE to manage 10th and 100th Inch vs. Inch, Fahrenheit degrees, miles
- Has Null vs. Blank issues.

21st Century Version

- Conceptual table logic is 1 day of a 9 METEOROLOGICAL_ELEMENTS per record
- Original data is all character
- YEAR, MONTH, and DAY are stored in 3 separate character fields
- Monthly statistical data is stored on the “last day of the month” record
- Has Null vs. Blank issues

Support Data

- First attempt at making a relational database
- Cooperative_Station. The place where the data was collected
- Meteorological_Element
- Flag1_Data_Measurement
- Flag2_Data_Quality

PSC is new at this.

- MS SQL Since 2005
- SYBASE Since 1997
- PowerBuilder
- PowerDesigner
- In the past, I used Excel mostly for data transfer
- Limited Access experience
- Work on NOAA is done on a priority 2 basis
- There is a lot of priority 1 work to do

Tools we used

- Sybase/DataDesigner
- Sybase/Database
- Sybase/PowerBuilder data windows
- Moved to MS SQL
 - MS SQL Enterprise Manager
 - Import/Export Wizard
- Excel
- TINN or other Text Editor

Sybase/DataDesigner

- Helps lay out the physical design of the database in a graphical context
- Built in design validation
- Output is high quality SQL Creation script for Tables, Indexes, Keys, Triggers, Stored procedures, Views
- Supports most database vendors

Sybase/PowerBuilder Data Windows

- Easy to use query builder
- Query output can be easily formatted as a report when needed
- Report data can be outputted to many common formats including, .xls, csv, multiple flavors of txt, and many others

Sybase/Database

- This worked for me because I am comfortable using it
- You should first work with your data in whatever tool you are most comfortable with and then develop whatever skills are necessary to get it to where it needs to go

MS SQL Enterprise Manager

- Diagrams
- Tables
 - Viewing Data. Also see Excel
 - Constructing SQL statements for the Import/Export Wizard
 - Duplicating a table using Copy/Paste
- Stored Procedures. Concept is great! The tools are not so good. **USE IT WHEN NEEDED!**

Import/Export Wizard

- From Sybase to MS SQL (NOAA)
 - Data Migration
- From NOAA TO NOAA_COPY
 - Online Live BACKUP Functions
- In NOAA TO **NOAA!**
 - **Data Manipulation**
 - **Use the built in power of SQL**

Excel

- Used it to analyze my results. Copy/Paste from the results of Table → Open table →
- Use “Add Sheet” to help document the results of the various tests you make on the data
- Use “Add Sheet” to bring other data sets
- Use “Copy Sheet” on your documentation and adjust the results as you perform similar test on the “new” data set

TINN or other text viewers.

- Editing stored procedures.
- Right click on a table or tables and copy.
 - Open a text editor and paste.
 - Change the table name and the name of the primary key.
 - Add additional fields if necessary.
 - Change any field name as required.
 - Open SQL Query Analyzer and run.
 - Results: New Table

Example

- CREATE TABLE [METERLOGICAL_ELEMENT] (
 - [DATA_ORIG] [char] (4) COLLATE Latin1_General_CI_AS NOT NULL ,
 - [ELEMENT] [char] (4) COLLATE Latin1_General_CI_AS NOT NULL ,
 - [DEFINITION_ELEMENT] [varchar] (255) COLLATE Latin1_General_CI_AS NULL ,
 - CONSTRAINT [PK_METERLOGICAL_ELEMENT] PRIMARY KEY CLUSTERED
 - ([DATA_ORIG],
 - [ELEMENT]
 -) ON [PRIMARY]
 -) ON [PRIMARY]
 - GO

Tools I would like: Embarcadero DBArtisan

- We use it for SYBASE plan to get for MS SQL Servers in our next budget
 - It is a want if you have a lot of time and little money.
 - It's a need if you have money and not much time.
 - Today it has to be a want. As we do more, I suspect it will become a need.

Embarcadero DBArtisan. Why?

- This will work with both SYBASE and MS SQL in the same package.
- For SYBASE, it is almost a DBO/DBA/SA in a box.
- Without it, I would have to recommend hiring an additional high skill employee.
- I find it hard to find the time to work on a project I like and have an interest.

NOAA: A work in progress.

Current Issues

- Can you believe it?
- Does the data make sense?
- Data Normalization. How far do you go?
- Where can you use this data?
- How do you use this data?
- What form does the data need to be in to be useful?

The Current Situation

- 42 User tables with supporting Indexes and Keys
- 9 user written Stored Procedures
- Over 100 years of weather related data for North Dakota.
- The normalized 20th Century table has 16,630,625 rows. Return all rows from the server to my PC took less than 2 minutes.
- Weather reports from 274 separate sites in North Dakota
- The potential to add data from surrounding states.
- **Ability to transform data to many different formats**

- # Questions and Comments

THANK YOU!