Mi^3 Climatological Station History Database: Architectural Techniques

Jeff Arnfield
National Climatic Data Center
Asheville, NC
Jeff.Arnfield@noaa.gov

Gary Shears
Sherian Corporation
Morgantown, WV

Station History Evolution at NCDC

- Paper forms
  - Cumbersome, inaccessible, slow search
- Operational lists
  - Application-specific, limited info, redundancy
- S2K database
  - Few queries, limited info, limited access
- SHIPS (Station History Production System)
  - Slow, few queries, most access via a few reports, no query-only, design not normalized
- CliServ
  - Queryable, integrates limited station info & inventories from other systems

Mi^3 Background & Goals

- Metadata Integration & Improvement Initiative
- Integrate, enhance & increase access
- Initial focus: system to manage NCDC's climatological station history
- Accommodate broad variety of station details
  - Identity
  - Location
  - Location Mapping
  - Instrumentation
  - Observing practices
  - Data Programs
  - Management
  - Photos & Document images
- Track information source & log changes
- Adapt to new observing systems, data programs and phenomena without recoding

Mi^3 System Overview

- Normalized Oracle relational database
  - More than 140 tables
  - Accommodates new networks, programs and phenomena
  - Easily accessible by other apps & systems
- Web-based delivery
  - ColdFusion, Java and PL/SQL
  - Geographically distributed access
  - Minimizes administration and training
- Dedicated database and application servers
  - Excellent performance, even with ad hoc queries
  - Readily scaleable
  - Layered security and access

Temporal Trepidations

- The Issues
  - All station details have a period of validity
  - Any station detail may change independently
  - Various station details often have different periods of validity
  - Relational design spreads details across many tables
  - Some tables may contain no details for a station for a given period
- These factors can conspire to make a system an almost query-proof “metadata motel”
- Managing and querying temporal information becomes critical to usability

Temporal Data Management

- Possible Solutions
  - Create a new record in all tables whenever a station's data changes in any table
  - Straight programmatic approach to combining data for each desired view
  - Develop a generalized, data-driven approach at the database level

So you want some station information….

STATION NAME, COOP ID, TIME OF TEMP AND PRECIP OBSERVATION					
Begin Date	**End Date**	**Station Name**	**Coop ID**	**Temp Obs**	**Prcp Obs**
07/01/1930	09/15/1940	HAINES	503495	?	?
10/01/1940	09/15/1953	HAINES CAA	503495	?	?
09/15/1953	09/30/1956	HAINES 1 S	503492	?	?
09/06/1963	05/15/1968	HAINES		?	?
06/21/1973	01/01/1984	HAINES	503490	1600	1600
01/01/1984	04/01/1984	HAINES	503490	0800	0800
04/01/1984	04/15/1986	HAINES	503490	0830	0830
04/15/1986	07/01/1986	HAINES	503490	1500	1500
07/01/1986	06/05/1998	HAINES	503490	0900	1500
06/05/1998	12/31/9999	HAINES AIRPORT	503490	0900	1500

Three tables must be combined

STATION NAME			
Station Key	**Begin Date**	**End Date**	**Name**
20021913	07/01/1930	10/01/1940	HAINES
20021913	10/01/1940	09/15/1953	HAINES CAA
20021913	09/15/1953	09/30/1956	HAINES 1 S
20021913	09/06/1963	05/15/1968	HAINES
20021913	06/21/1973	06/05/1998	HAINES
20021913	06/05/1998	12/31/9999	HAINES AIRPORT

- Join on Station Key column
- Must also match by date

- Data change independently
- No table contains all dates

STATION ID				
Station Key	**Begin Date**	**End Date**	**ID Type**	**ID**
20021913	07/01/1930	09/15/1953	Coop number	503495
20021913	09/15/1953	09/30/1956	Coop number	503492
20021913	06/21/1973	12/31/9999	Coop number	503490
20021913	07/01/1930	09/30/1953	WBAN number	25323
20021913	09/06/1963	05/15/1968	WBAN number	25323
20021913	06/21/1973	12/31/9999	WBAN number	25323

Station Phenomenon Observing Protocol					
Station Key	**Begin Date**	**End Date**	**Phenomenon**	**Frequency**	**Obs Time**
20021913	06/21/1973	01/01/1984	TEMP	Daily	1600
20021913	01/01/1984	04/01/1984	TEMP	Daily	0800
20021913	04/01/1984	04/15/1986	TEMP	Daily	0830
20021913	04/15/1986	07/01/1986	TEMP	Daily	1500
20021913	07/01/1986	12/31/9999	TEMP	Daily	0900
20021913	06/21/1973	01/01/1984	PRCP	Daily	1600
20021913	01/01/1984	04/01/1984	PRCP	Daily	0800
20021913	04/01/1984	04/15/1986	PRCP	Daily	0830
20021913	04/15/1986	12/31/9999	PRCP	Daily	1500

- A table may not contain data for a given period

How do we derive the date pairs?

01/01/1925
09/15/1940
09/15/1953
09/30/1956
09/06/1963
05/15/1968
06/21/1973
01/01/1984
04/01/1984
04/15/1986
07/01/1986
06/05/1998
12/31/9999

STATION NAME

Station Key	Begin Date	End Date	Name
20021913	07/01/1930	10/01/1940	HAINES
20021913	10/01/1940	09/15/1953	HAINES CAA
20021913	09/15/1953	09/30/1956	HAINES 1 S
20021913	09/06/1963	05/15/1968	HAINES
20021913	06/21/1973	06/05/1998	HAINES
20021913	06/05/1998	12/31/9999	HAINES AIRPORT

STATION ID

Station Key	Begin Date	End Date	ID Type	ID
20021913	07/01/1930	09/15/1953	Coop number	503495
20021913	09/15/1953	09/30/1956	Coop number	503492
20021913	06/21/1973	12/31/9999	Coop number	503490
20021913	07/01/1930	09/30/1953	WBAN number	25323
20021913	09/06/1963	05/15/1968	WBAN number	25323
20021913	06/21/1973	12/31/9999	WBAN number	25323

Station Phenomenon Observing Protocol

Station Key	Begin Date	End Date	Phenomenon	Frequency	Obs Time
20021913	07/01/1930	09/30/1956	TEMP	Daily	
20021913	06/21/1973	01/01/1984	TEMP	Daily	1600
20021913	01/01/1984	04/01/1984	TEMP	Daily	0800
20021913	04/01/1984	04/15/1986	TEMP	Daily	0830
20021913	04/15/1986	07/01/1986	TEMP	Daily	1500
20021913	07/01/1986	12/31/9999	TEMP	Daily	0900
20021913	07/01/1930	09/30/1956	PRCP	Daily	
20021913	06/21/1973	01/01/1984	PRCP	Daily	1600
20021913	01/01/1984	04/01/1984	PRCP	Daily	0800
20021913	04/01/1984	04/15/1986	PRCP	Daily	0830
20021913	04/15/1986	12/31/9999	PRCP	Daily	1500

Build a list of dates from all tables in the view, discarding duplicates

The Context Date Pair Algorithm

All Dates
07/01/1930
09/15/1940
09/15/1953
09/30/1956
09/06/1963
05/15/1968
06/21/1973
01/01/1984
04/01/1984
04/15/1986
07/01/1986
06/05/1998
12/31/9999

Begin Date	End Date
07/01/1930	10/01/1940
10/01/1940	09/15/1953
09/15/1953	09/30/1956
09/30/1956	09/06/1963
09/06/1963	05/15/1968
05/15/1968	06/21/1973
06/21/1973	01/01/1984
01/01/1984	04/01/1984
04/01/1984	04/15/1986
04/15/1986	07/01/1986
07/01/1986	06/05/1998
06/05/1998	12/31/9999
12/31/9999	

Begin Date	End Date
07/01/1930	10/01/1940
10/01/1940	09/15/1953
09/15/1953	09/30/1956
09/06/1963	05/15/1968
06/21/1973	01/01/1984
01/01/1984	04/01/1984
04/01/1984	04/15/1986
04/15/1986	07/01/1986
07/01/1986	06/05/1998
06/05/1998	12/31/9999

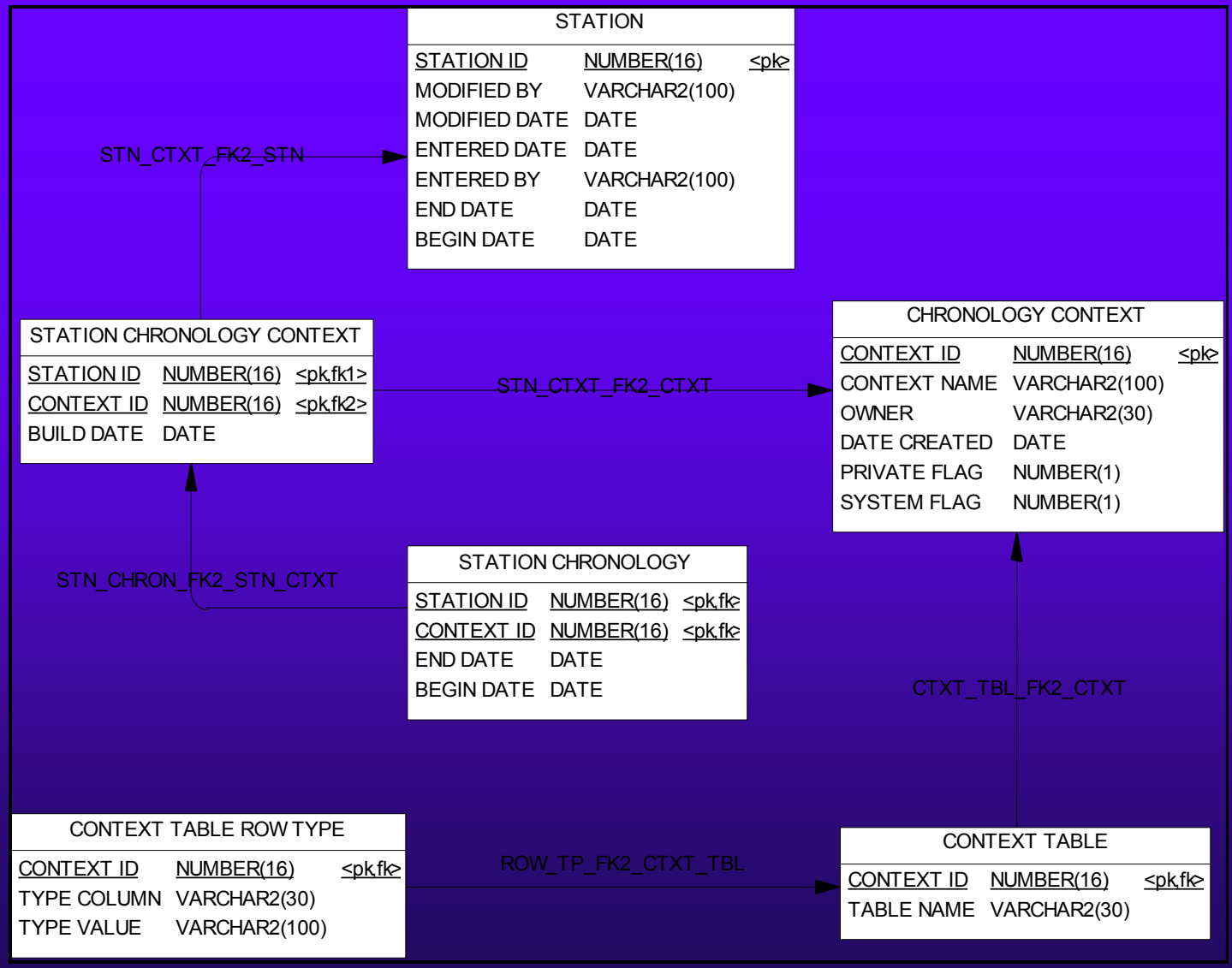
Use list to build date pairs, removing those pairs for which no table in the view contains data

The final date pairs are used to drive the joins in our output view query

Context Date Pairs

- "Context", because a set of date pairs is valid in a specific context… a given view or report
- Original 3-D implementation via Oracle views
  - Dang, this works!
  - Dang, this is slow
  - Dang, this is going to mean a lot of maintenance
- Production implementation
  - Supporting functions implemented via PL/SQL stored procedures
  - Five table schema
  - Triggers ensure date pairs refreshed

The Context Date Pair Schema



Context Date Schema in Action

Chronology Context				
Context ID	**Context Name**	**Owner**	**Private**	**System**
1	VU_NameCoopTempPrcpTOBS	JARNFIEL	N	N
2	VU_StationHistoryOverview	MI3	N	Y

Context Table	
Context ID	**Table Name**
1	Station Name
1	Station ID
1	Station Phenomenon Observing Protocol

Context Table Row Type			
Context ID	**Table Name**	**Type Column**	**Value**
1	Station ID	ID TYPE	COOP Number
1	Station Phenomenon Observing Protocol	Phenomenon	TEMP
1	Station Phenomenon Observing Protocol	Phenomenon	PRCP

Station Chronology Context		
Context ID	**Station Key**	**Build Date**
1	20021913	1/9/2002
1	20000233	1/2/2002
2	20021913	1/2/2002

Station Chronology			
Context ID	**Station Key**	**Begin Date**	**End Date**
1	20021913	07/01/1930	10/01/1940
1	20021913	10/01/1940	09/15/1953
1	20021913	09/15/1953	09/30/1956
1	20021913	09/06/1963	05/15/1968
1	20021913	06/21/1973	01/01/1984
1	20021913	01/01/1984	04/01/1984
1	20021913	04/01/1984	04/15/1986
1	20021913	04/15/1986	07/01/1986
1	20021913	07/01/1986	06/05/1998
1	20021913	06/05/1998	12/31/9999

Context Date System Implementation

- Supporting Functions
  - Define & maintain a context via user interface
  - Rebuild a context for one station
    - Core logic, called by all rebuild functions
  - Rebuild a context for all stations
  - Rebuild all contexts for one station
  - Rebuild all contexts for all stations
- Maintenance Performance
  - Building a single context containing 6 tables for 35,000 stations takes less than 15 seconds, before tuning
  - Single station rebuild nearly instantaneous

Context Date Pair Benefits

- Common schema / body of code
  - supports all queries, views and reports
  - minimizes maintenance
  - permits future enhancements
- Critical in developing this system
- Useful to power users for ad hoc queries
- End-users benefit because special data requests are easier to accommodate

Mi3: Architectural Techniques

- Credit where it's due
  - David Bowman, of NCDC, laid logical groundwork for this algorithm while developing the Master Station History Report
  - Robin Ilardi, of Sherian Corp, developed code to implement the algorithm

- For more information
  - Jeff.Arnfield@noaa.gov
  - Come to the NCDC portion of NOAA's booth
  - NCDC's web site: http://www.ncdc.noaa.gov