Papers 6, 13

Solution The primary database has a detailed structure that need concern us only in so far as we must identify the access keys. Basic data are time series recorded by originating observatories; (these are either uniquely identified points on rivers within a water authority, or uniquely named weather stations. We assume that data for warehousing is identified by specifying the unique identifier of the observatory together with a time interval. Each inland weather station has an attribute identifying the associated water authority; a null value is entered for weather ships.

The data warehouse will record primary data in the same schema as in the primary database. Often the data to be saved (we assume) will be indicated by such high-level judgements as: "keep the records of the Severn Water Authority area from 6/11/2000 to 8/11/2000, with index terms *flood, hurricane, cloudburst.*"

Papers 6, 13

Solution ctd) It's vital that we generate a unique identifier for such incidents, so that when a match of weather patterns is recognized we can retrieve all the records that were saved. So we probably should have a directory structure recording:

incident-ID | area, start-date, keyword*

which will be the primary retrieval structure.

We also need to specify exactly what has been saved associated with this incident: we hope that the relational DBMS supports {foreign keys} referencing attributes

incident_ID, {observatory} start_time, end_time

so that the unique observatory identifier can refer to the keys of a number of primary relations. That's a matter of detail. This relation describes the primary data to be retrieved from the named observatory for each incident.

Papers 6, 13

Solution (d) What we've designed so far will serve as the keyword index for the warehouse together with the reference relation to the saved primary data.

In addition meteorologists may wish to add comments, maybe structured, maybe not. These comments will be at incident level, and we separate them off into a new relation.

incident_ID | comment, nick-name, etc.

I've suggested that particularly cataclysmic incidents might get a nick-name; it would be a big mistake to use such a name to identify the incident ("the great storm of October 1987"). MAYBE could include one optionally in the directory structure...

Papers 6, 13

Solution ctd) ... which is a cop-out at present, because I've left the repeating groups keyword* as part of the schema. Probably the best bet is to have a separate many-many relation

incident_ID, keyword |

though relevant search criteria are likely to include area and maybe time-of-year as well.

I've generated (after one glitch) a schema in 3NF, but they may go through a bit of a song and dance. That's fine.

I think this is OK. There are two technical problems (repeating groups and foreign keys), and plenty of scope for imagination.