2001 Paper 6 Question 8

Databases

The environmental agency is setting up an SQL database to monitor long-term trends in the climate. Data are collected from observatories of a number of different kinds.

Flood risk is of particular concern. Each water authority measures river levels and rates of flow hourly at major points, and records reservoir levels daily.

In addition, the agency maintains weather stations both inland and at sea. These record precipitation (rainfall etc.), temperature, sunshine, air pressure and wind. Values of new precipitation, temperature, pressure, and wind speed and direction are taken hourly; gusts of over 60 m.p.h. are noted whenever they occur.

Maximum and minimum temperature and pressure, the total number of hours of sunshine and the total precipitation are recorded daily. Inland stations can be grouped by water authority.

By default these primary data will be relegated to archive after 2 years. Selected information is retained permanently in a data warehouse. This serves two purposes. First, it holds monthly summary data consisting of the maximum (and minimum as appropriate) day value for each statistic, together with the monthly totals of sunshine and precipitation. The warehouse also keeps detailed information relating to periods of extreme weather from the relevant observatories, with one or more keywords describing the nature of the incident (*flood*, *blizzard*, *hurricane* etc.) and an optional comment.

Write notes to assist in the design of the schema for the relational data warehouse, including any diagrams that you find helpful. Explain how your design will enable meteorologists to find relevant past records, noting any assumptions that you make about the nature of the data.

[You should not go into unnecessary detail about the structure of the primary database. You may assume that expert meteorologists will select the data for the warehouse.]

[20 marks]