

Introduction

The NNSA Knowledge Base consists of Oracle database tables, flat-file information, and tools. The bulk of the KB is in the database tables. The core database tables hold seismic event information -- origins, arrivals, associations, and magnitudes that come from various global, regional and local bulletins. Other database tables include custom parameter information for location, event identification, and coda magnitudes that is generated from research. There are database tables that contain waveform metadata that point to flat-file waveforms on the system. Finally, there are database tables that contain information about the seismic stations -- locations, channel names, instrument responses, etc.

To efficiently accomplish the integration of database tables from multiple sources such that no incorrect information is included, no information is duplicated, and proper links are maintained between tables, the GNEM R&E program developed a specialized metadata schema (the **schema schema**), and two software packages - **QCTool** and **DBTools**. In this presentation, we will discuss each of these and show how they are used in an overall database integration pipeline.

QCTool

QCTool (Quality Control Tool) is an automated quality assurance and quality control tool for database tables. The first check makes sure the tables are in the appropriate schema, i.e. the correct columns in the correct order with the correct format. Then the tool executes three kinds of QC checks - single-table, multi-table and complex joins - and outputs the results to files.

Single-table checks:
These checks are defined by the schema schema tables. They verify that the table matches the documented table structure described in the schema schema tables (primary keys, unique keys, ranges, NA values)

Multi-table checks (one-column, two-column, indirect cross-reference):

These checks are defined in the qc parameter file. One-column cross reference checks require that every value from the specified column in the first table is also found in the specified column of the comparison table. Two-column checks require that each unique pair of values from two specified columns in the first table are also found in the specified columns of the comparison table. The indirect cross-reference check handles cases where the reference for a column in the first column is also found in the specified column which is another column in the same table. (example = wtag table)

Complexjoin checks:

Specified in an auxiliary database table - **complexjoin**. Provides a mechanism to specify any consistency requirements or expectations across as many as three generalized tables. There is always a target column for validation, and additional tables can be defined.

The QCTool checks are defined by the user, and depend on the specific schema defined in the schema schema.

KB Database Integration Process

