SAND REPORT

SAND2002-3055
Unlimited Release
Printed October 2007
Supersedes SAND2002-3055 dated October 2006

National Nuclear Security Administration Knowledge Base Core Table Schema Document

Dorthe Carr

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000.

Approved for public release; further dissemination unlimited.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865)576-8401
Facsimile: (865)576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.doe.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

Telephone: (800)553-6847
Facsimile: (703)605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/ordering.htm

SAND2002-3055
Unlimited Release
Printed October 2007
Supersedes SAND2002-3055
dated October 2006

National Nuclear Security Administration Knowledge Base Core Table Schema Document

Dorthe Carr
Next Generation Monitoring Systems Department
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-0401

Abstract

The National Nuclear Security Administration is creating a Knowledge Base to store technical information to support the United States nuclear explosion monitoring mission. This document defines the core database tables that are used in the Knowledge Base.

Acknowledgements

Much of this document was based on work done by SAIC, who have made revisions to the CSS 3.0 schema developed in 1990. Those people include Bonnie MacRitchie, Don Irons, Tammy Lyons, Craig Morency, Brian Pope, Henry Swanger and N. Putvinskaya.

Chris Young, Mark Harris, Julio Aguilar-Chang, Richard Stead and David Gallegos provided helpful reviews to the document.

This work was supported by the United States Department of Energy under Contract DE-AC04-94AL8500. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy.

Table of Contents

1.0 Introduction 6
1.1 What's New 6
2.0 Table Descriptions 7
2.1 Keys and Data Categories 7
3.0 Table Relationships 8
3.1 Primary Tables 8
3.2 Lookup tables 10
4.0 Column Descriptions 11
5.0 References 13

Figures

Figure 1. Relationships between Primary Tables 8
Figure 2. Lookup Table Relationships 10

1.0 Introduction

The purpose of this document is to present the ORACLE database tables in the NNSA Knowledge Base that are modifications to the CSS3.0 Database Schema developed in 1990. (Anderson et al., 1990). It is intended to be used as a reference by researchers inside and outside of NNSA/DOE as they compile information to submit to the NNSA Knowledge Base. These “core” tables are separated into two groups. The Primary tables are dynamic and consist of information that can be used in automatic and interactive processing (e.g. arrivals, locations). The Lookup tables change infrequently and are used for auxiliary information used by the processing. In general, the information stored in the core tables consists of:

- arrivals
- events, origins, associations of arrivals
- magnitude information
- station information (networks, site descriptions, instrument responses)
- pointers to waveform data
- comments pertaining to the information

This document is divided into four sections, the first being this introduction. Section two defines the twenty tables that make up the core tables of the NNSA Knowledge Base database. Both internal (ORACLE) and external formats for the attributes are defined, along with a short description of each attribute. In addition, the primary, unique and foreign keys are defined. Section three of the document shows the relationships between the different tables by using entity-relationship diagrams. The last section, defines the columns or attributes of the various tables. Information that is included is the Not Applicable (NA) value, the format of the data and the applicable range for the attribute.

1.1 What’s New

We have changed the unique key for the table **instrument** to include the *samprate* column.

We changed the *auth* column from a varchar2(15) to a varchar2(20). This affected the tables **amplitude**, **arrival**, **event**, **netmag**, **network**, **origin** and **stamag**.

2.0 Table Descriptions

The database table descriptions can be found in the linked html document index.html. The name of the table is first, followed by a description of the purpose and use of the table. Below the description is a listing of the columns, in the order which they are defined in the tables. The storage column gives the actual ORACLE datatype for the column in question. The external format and character positions columns are provided for the convenience of database users who wish to transfer data between the ORACLE database tables and flat files.

2.1 Keys and Data Categories

The columns of the database table are categorized as keys and data. Key columns link database tables. Data columns, the reason that database tables exist, are split into three categories: Descriptive, Measurement, and Administrative. The following explains the format used in the entries:

Keys:	Primary	The columns which, when taken together, uniquely identify a row in the database table.
	Unique	Other columns that also uniquely identify a row and may be used as surrogates for primary keys.
	Foreign	Primary keys in another database table.
Data:	Descriptive	Qualitative columns not in primary or unique key
	Measurement	Quantitative columns not in primary or unique key
	Administration	Columns used for database administration not in primary or unique key

Keys provide the links by which database tables are joined. The following definitions explain the types of keys:

- Primary key: Uniquely identifies a row in a database table. We try to use an id as the primary key. For example, *orid* is the primary key for the **origin** table.
- Unique key: Also uniquely identifies a row in a database table and may be used as a surrogate for the primary key. Often it is a concatenation of several columns. For example, every **origin** record is unique by *lat*, *lon*, *depth*, *time* and *auth.*
- Foreign key: The column or combination of columns in a table used to join to a primary key in another table. For example, *evid* is a foreign key in the **origin** database table but is the primary key in the **event** database table.

3.0 Table Relationships

3.1 Primary Tables

Figure 1 shows the summary of the Primary tables and keys. Each of these tables is involved in preserving origin hypothesis and events.

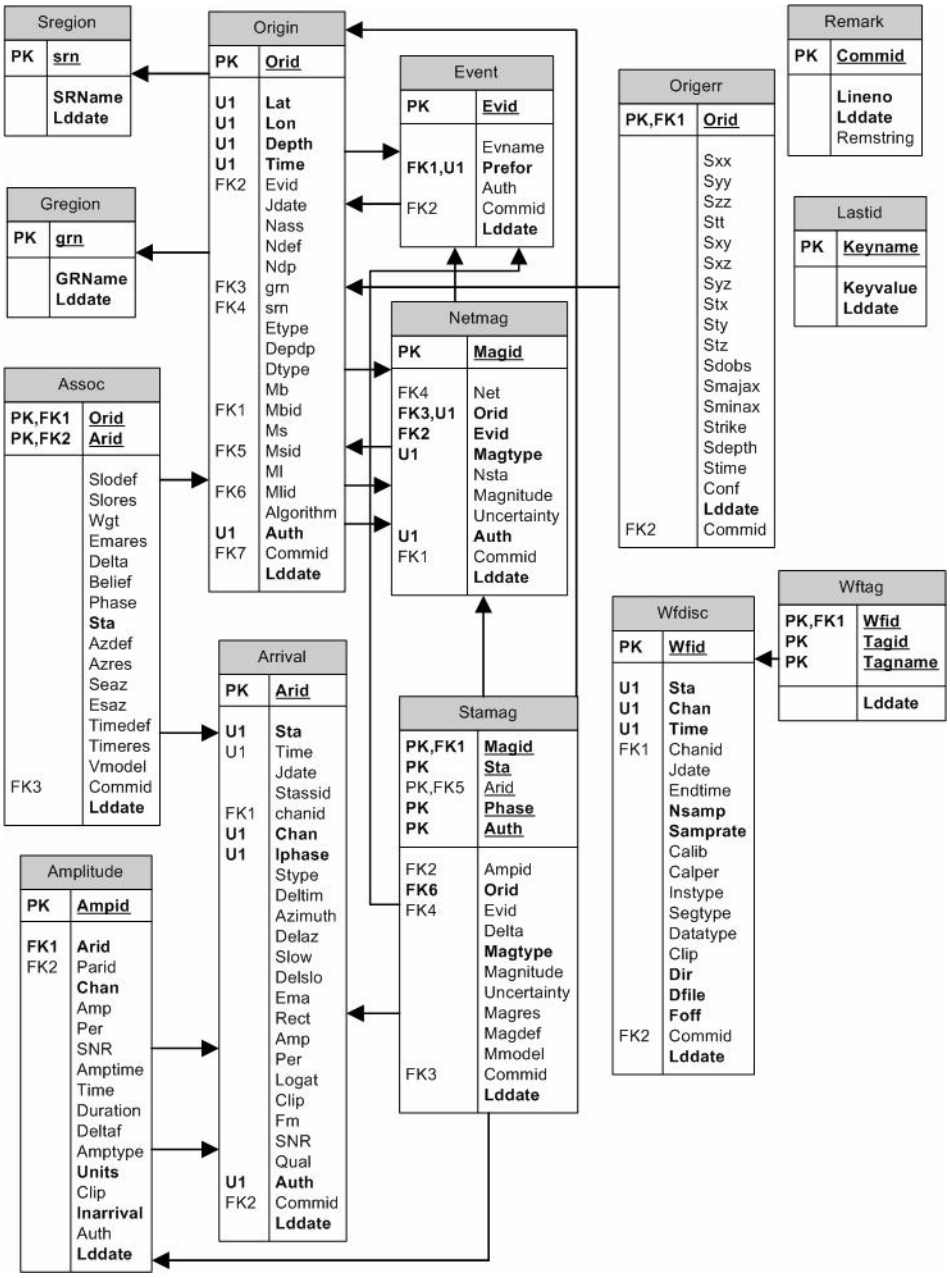

Figure 1. Relationships between Primary Tables

Arrivals in the **arrival** table are associated with different origin hypotheses from the **origin** table through the **assoc** table. Based on seismic signal measurements, groups of arrivals are associated with presumed events. Each event may have many different origin hypotheses, each with a different event location estimate. The preferred origin hypothesis is specified in the **event** table as a *prefor*.

The **stamag** table contains station magnitude estimates based upon measurements made on specific seismic phases. Values in the **stamag** table are used to calculate network magnitudes stored in the **netmag** table. The **netmag** table contains estimates of network magnitudes of different types for an event in the **event** table. Each network magnitude has a unique *magid* that relates to either a body wave magnitude (m_b), surface wave magnitude (M_s) or a local magnitude (M_L) through the columns *mbid*, *msid* or *mlid*.

The waveform tables are **wfdisc** and **wftag**. The waveforms themselves are stored in the flat files on the disk. They are usually called ".w" files and are a sequence of a sample values (usually in a binary representation). The descriptive information on the waveforms is stored in the **wfdisc** table, which provides a pointer (or index) to the waveforms on the disk. The **wfdisc** table is linked to the **arrival** and **origin** tables through *sta*, *chan*, and *time*. The **wftag** table specifies which table the **wfdisc** record is linked to, **origin** or **arrival**.

3.2 Lookup tables

The Lookup tables are fairly static and primarily contain look-up information. Figure 2 is an overview of these tables.

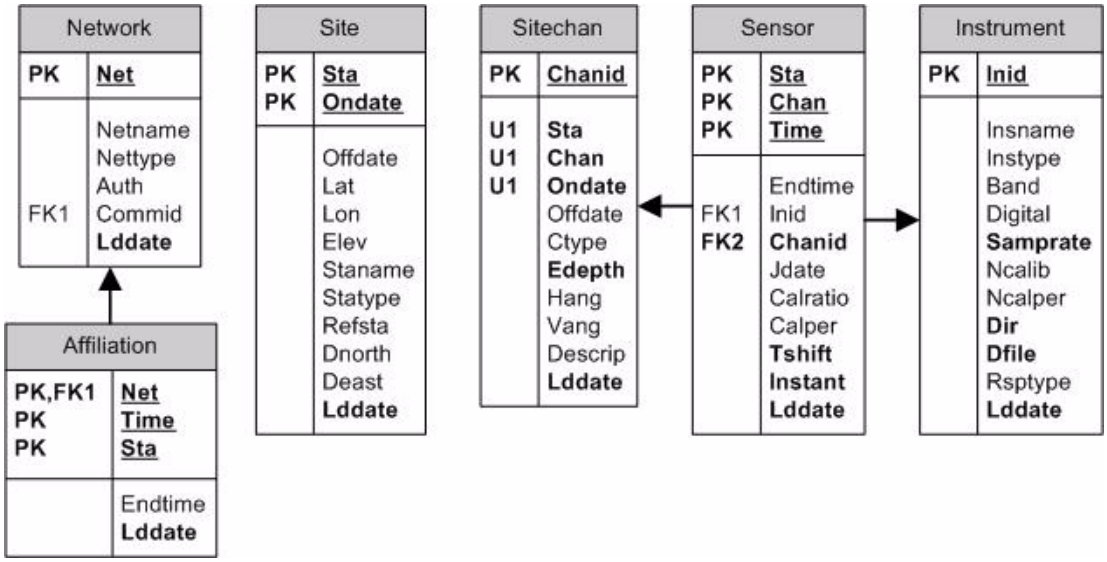

Figure 2. Lookup Table Relationships

The **site** table contains station location information. It describes the geographic location of a station. The **site** table also contains fields that describe the offset of a station relative to an array reference location. The **affiliation** table groups stations across wide geographic areas as networks and to group the elements of arrays to the name of the array. The general information about the seismic networks is stored in the **network** table.

Detailed calibration information is stored in flat files, in a variety of formats. The **instrument** table holds complete instrument response information, including ancillary calibration information and pointers to the flat files with detailed instrument responses. The instrument identifier *inid* links the **instrument** table to the **sensor** table. The **sensor** table contains calibration information for specific sensor channels and is linked to the **wfdisc** and **arrival** tables through *sta*/*chan*/*time*. It provides instrument update records, using the calibration period column *calper*, thus linking a *sta*/*chan*/*time* to a complete instrument response.

4.0 Column Descriptions

The column descriptions can be found in the linked html document index.html.
Each column is described with the following attributes:

Name: name of the column (shown in *italics*)

Table: table(s) which contain the column (shown in **bold**)

Description: column description

Format: ORACLE data type

We are using four of the available data types in ORACLE.

VARCHAR2: All character data in the database are defined to be VARCHAR2(n) where “n” is the number of characters in the string

NUMBER: All integer fields in the database are defined to be Number(n) where “n” is the number of digits allowed in the number

FLOAT: All real number fields in the database are defined to be FLOAT(24) or FLOAT(53). The ORACLE FLOAT(n) type, where “n” is the number of binary digits allowed in the number, allows the approximation of single and double precision floats commonly used in scientific programming.

DATE: The only field in the database that is declared to be the ORACLE DATE data type is lddate, which stores the day and time a record was inserted into the database.

NA Value: A value that is used to indicate that information is not available for this column. The NA Value should be used when the actual value is not known and alerts users and applications that the desired column was not available when the record was created. Essential attributes that must always be given a value are documented with an NA Value of NOT ALLOWED. The following table lists the convention we tried to follow when assigning NA Values.

Column type/range	NA Value	Examples
characters (varchar2)	- (hyphen)	*auth*
non-negative integers and real numbers	-1	*chanid, arid cfreq, deltim*
real numbers > -999.0	-999	*azres*
large real numbers	+/-9999999999	*endtime, time*

An NA Value should not be confused with an ORACLE NULL. NA Values are supplied by users, while ORACLE inserts the database value NULL

when no value is specified. A column containing a database value of NULL appears blank when selected within SQL*Plus. When creating a table, a column may be constrained as NOT NULL to require the user to supply a value. The ORACLE describe command will identify such columns as NOT NULL. No correlation is intended between ORACLE NOT NULL requirements and the requirements that a column must be specified.

Units: Units of measure, if applicable

Range: The range of permissible or recommended values for this column. Whenever possible an explicit range is defined. The range is important for data integrity and database management systems that automatically check ranges. When the range consists of a relatively small number of discreet values, the following notation is used (values in courier font):

column ∈ {`value-1, value-2, ...., value-n`}

No range is documented for columns whose value may be any character string.

The *time* attribute throughout the database is stored as epochal time, the number of seconds since January 1, 1970. Epochal time has a precision of 1 millisecond. Often *time* is matched by the more readable jdate. This so-called “Julian date” represents a day in the form, for example 1981231 where 1981 is the year (YYYY) and 231 is the day of year (DOY).

5.0 References

Anderson, J., W. E. Farrell, K. Garcia, J. Given, H. Swanger, 1990, **Center for Seismic Studies Version 3 Database Schema Reference Manual**, Technical Report C90-01, SAIC-90/1235, 64pp, Science Applications International Corporation, San Diego, California.

Distribution

Sandia National Laboratories Internal:

5	MS1138	Dorthe Carr, 5533
1	MS9018	Central Tech Files, 8945-1
2	MS0899	Technical Library, 9616
1	MS0612	Review & Approval Desk, 9612