

Schema Scanner

Reference Manual
Version 1.0

TNS Software Inc
54 Brock St South
Perth, Ontario
K7H 1Z8
Canada
Email: info@tns-soft.com
<http://www.tns-soft.com>

Copyright 1993-2015 TNS Software Inc., 54 Brock St South, Perth, Ontario Canada. All rights reserved.

This product or document is protected by copyright and distributed under licenses restricting its use, copying, distribution, and decompilation. No part of this product or document may be reproduced in any form by any means without prior written authorization of TNS Software and its licensors, if any. Third-party software is copyrighted and licensed from TNS Software suppliers.

RESTRICTED RIGHTS:

DOCUMENTATION IS PROVIDED "AS IS" AND ALL EXPRESS OR IMPLIED CONDITIONS, REPRESENTATIONS AND WARRANTIES, INCLUDING ANY IMPLIED WARRANTY OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NON-INFRINGEMENT, ARE DISCLAIMED, EXCEPT TO THE EXTENT THAT SUCH DISCLAIMERS ARE HELD TO BE LEGALLY INVALID.



Contents

INTRODUCTION	1
INSTALLATION	3
USING THE PROPERTIES FILE	4
RUNNING SCHEMASCANNER	8
Required Files	8
Classpath	8
Command Line Options	8
SchemaScanner Sample Execution	8
Using XSL Templates	9
Sample xslReport Execution	9
XSL Template	11
Generated Trigger	12
USING THE GENERATED XML OUTPUT	14
Entity Information	14
Attribute Information	14
Generated XML Output	16
Schema Report	22
Attribute Report	23
SCHEMA ERRORS	29
TNSGEN OUTPUT	35
Legacy Conversion	35
Strict Privacy Requirements	35
Load Testing / Benchmarking	36
New Application	36
Summary	36
Example	37
APPENDIX A - SCHEMASCANNER DTD	41

APPENDIX B – ORACLE KEYWORDS **43**

APPENDIX C – ORACLE RESERVED WORDS **45**



Introduction

During the past twenty years, Relational Database Management Systems (RDBMS) have become the de facto repository for information within almost every business and government organization. Database technologies have matured to readily support distributed queries and update, all the while ensuring secure access and both referential and data integrity.

At the same time, tools to model, implement, manage and maintain have not kept pace. Consider tools like Sql*Plus from Oracle Corporation, a lowly command-line SQL interpreter that is still considered a primary development and management tool by most clients and yet has remained virtually unchanged since its introduction over twenty years ago.

Into this gaping void, TNS Software introduces *SchemaScanner*, our facility to help document, manage, maintain and understand your Oracle database. With *SchemaScanner* you can:

- Document selected entities or an entire Oracle schema and produces an XML output file containing all schema components, including entities, attributes, indexes, constraints, comments and triggers.
- Scan each table to produces statistics for each attribute including min/max values, distinct values, percentage null, min/max length and more.
- Examine a schema to identify and document entity and attribute discrepancies, missing foreign key relationships, nested indexes, inconsistent constraints using a rule base of over 50 best-practice modeling rules.
- Generate control files to simplify the generation of random test data that closely resembles that of the source schema, including conformance to date and numeric ranges, enumerated character values, percentage of nulls and more using TNS Software's test data generator, *tnsgen*.

- Snapshot a schema for a current state repository check in during application development
- Create reverse-engineered schemas or other reports using supplied or user-defined XSL templates.
- Incorporate into make, ant or other batch project procedures

SchemaScanner supports any Oracle RDBMS Version 8.0.5 and above is written in Java using JDBC for database connectivity. SchemaScanner will operate on any OS platform which supports JDK 1.4.2 or above without requiring Oracle client software.

Potential users of SchemaScanner include developers who need to create and manage schema objects during the development cycle, QA testers who need to understand the nature of the schema and create test data and test cases, DBA's who need to manage production databases and need to understand data volumes, indexes and constraints.

SchemaScanner is also an invaluable tool for conducting system audits or creating documentation for production systems where little usable documentation exists.

Installation

SchemaScanner is delivered as a single zip file that contains the SchemaScanner JAR file, XSL formatting scripts, shell and batch scripts and late breaking news. The file may be installed by executing the following steps:

1 - cd to a target directory, creating the directory if necessary:

```
mkdir /var/SchemaScanner  
cd /var/SchemaScanner
```

2 - copy the downloaded distribution to the target directory

3 - uncompress and unzip the distribution:

```
unzip SchemaScanner_v1.zip
```

Follow the instructions in the next two sections, **Using the Properties File** and **Running SchemaScanner** to test your installation.

Using the Properties File

SchemaScanner uses a properties file to definition the target database, database account, schema owner and entities. By default, the name of the properties file is *SchemaScanner.props*. This file should be located in *SchemaScanner's* working directory.

The format of this file follows common conventions for properties files:

An entry has the basic format:

key value

The *key*, which is the name for a particular property, may not contain any white space, e.g. space or tab. The *value* is everything from the first non-white space character after the key until the end of the line.

Comment lines must begin with a # character.

All key names are case sensitive.

The following is an explanation of the various properties you may set in this properties file.

DatabaseConnectionString

The database connection string defines the target Oracle database. The format is `jdbc:oracle:thin:@<hostname>:<port>:<database instance>`. If you are unsure of the value for these variables, ask your DBA for assistance. This parameter is mandatory.

DBAccount

This parameter defines the username that you would like to use to connect to the database. This username does not need to be the same as the schema owner but must be able to view the target tables using `all_tables`, `all_indexes`, etc. views. This parameter is mandatory.

DBPassword

This parameter defines the password for the user named as *DBAccount*. This parameter is mandatory.

SchemaOwner

This parameter defines the Oracle owner of the *Entities* to be examined. This parameter is mandatory.**Entities**

This parameter defines the *Entities* to be examined. You may list the entities you would like to examine, each separated by a comma. If the optional *Entities* property is omitted, all entities owned by the schema owner are examined.

ScanFileName

This parameter defines the file name for the SchemaScanner XML output file. This parameter is mandatory.

ExtendedStatistics

This parameter defines whether SchemaScanner will produce data distribution statistics for the *Entities* to be examined. Note that enabling *ExtendedStatistics* for large tables may require a great deal of time and consume a great deal of system resources. The value for *ExtendedStatistics* must be true or false.

AnalyzeSchema

This parameter defines whether SchemaScanner will analyze the schema and provide database structure errors for the *Entities* to be examined. This option requires little extra time and will provide valuable insight into the quality of the underlying schema. The value for *AnalyzeSchema* must be true or false.

GenerateLoaderScript

This parameter defines whether SchemaScanner will generate a tnsген script for the *Entities* examined. If *ExtendedStatistics* is true and reasonable data exists within the *Entities* examined, the tnsген script will be able to create test data which confirms to the data type, limits and data distribution of the source schema. The value for *GenerateLoaderScript* must be true or false.

DictionaryCase

This parameter defines the case for all object names in the generated XML file. The value for DictionaryCase must be one of upper, lower or initcap.

DmlDate

This parameter defines the common name used for the DmlDate attribute, which is recommended for all tables. This parameter is optional.

DmlUser

This parameter defines the common name used for the DmlUser attribute, which is recommended for all tables. This parameter is optional.

CommonAttributes

This parameter defines the common names used for system attributes within the schema. The use of these attributes in non-related entities will not generate functional dependency errors.

Sample Properties File

The following is the properties file that was used to generate the examples for the *Donor* table.

```
DatabaseConnectionString jdbc:oracle:thin:@tnsu2:1521:SYSDEV
DBAccount scott
DBPassword tiger
SchemaOwner scott
Entities donor
ScanFileName scan.xml
ExtendedStatistics true
AnalyzeSchema true
GenerateLoaderScript true
LoaderFileName tnsngen.ini
DictionaryCase initcap
DmlDate dml_date
DmlUser dml_user
CommonAttributes record_status_code,status_code,
```

Running SchemaScanner

Required Files

SchemaScanner requires JDK1.4.2 and the Oracle JDBC class files, contained in classes12.zip. The JDK may be downloaded from the Sun Microsystems web site at <http://www.sun.com> and the Oracle JDBC drivers may be downloaded from <http://www.oracle.com>.

Classpath

The CLASSPATH must be defined to include the SchemaScanner jar file, distributed as ss.jar and the Oracle JDBC classes, distributed as class12.zip. The following Unix shell script is an example of how to set the PATH and CLASSPATH to execute SchemaScanner:

```
#!/bin/sh
JAVA_HOME=/u1/j2sdk1.4.2; export JAVA_HOME
SS_ROOT=/u1/SchemaScanner; export SS_ROOT
CLASSPATH=$CLASSPATH:$SS_ROOT/ss.jar:$SS_ROOT/classes12.zip; export
CLASSPATH
PATH=$PATH:/$JAVA_HOME/bin; export PATH
echo $CLASSPATH
java SchemaScanner -pSchemaScanner.properties
```

Command Line Options

SchemaScanner recognizes a single command-line option (-p) which may be used to override the default properties file location as shown in the previous shell script example.

SchemaScanner Sample Execution

The following is an example of SchemaScanner using the previously defined properties file against a single, populated table.

```
Schema-Scanner: Non-Commercial and Educational Release 1.0 for jdk 1.4.2
Copyright (c) tns software inc., 1993-2003. All Rights Reserved
visit http://www.tns-soft.com for updates or commercial version
```

```

Schema-Scanner-2003-10-23 21:35:36-I> initializing...
Schema-Scanner-2003-10-23 21:35:36-I> Options In Effect:
Schema-Scanner-2003-10-23 21:35:36-I> =====
Schema-Scanner-2003-10-23 21:35:36-I> DatabaseConnectionString: jdbc:oracle:thin:@tnsu2:1521:SYSDEV
Schema-Scanner-2003-10-23 21:35:36-I> DBAccount: webdemo
Schema-Scanner-2003-10-23 21:35:36-I> SchemaOwner: 'WEBDEMO'
Schema-Scanner-2003-10-23 21:35:36-I> Entities: 'DONOR'
Schema-Scanner-2003-10-23 21:35:36-I> ScanFileName: scan.xml
Schema-Scanner-2003-10-23 21:35:36-I> LoaderFileName: tnsgen.ini
Schema-Scanner-2003-10-23 21:35:36-I> ExtendedStatistics: true
Schema-Scanner-2003-10-23 21:35:36-I> GenerateLoaderScript: true
Schema-Scanner-2003-10-23 21:35:36-I> AnalyzeSchema: true
Schema-Scanner-2003-10-23 21:35:36-I> DmlDate: dml_date
Schema-Scanner-2003-10-23 21:35:36-I> DmlUser: dml_user
Schema-Scanner-2003-10-23 21:35:36-I> DictionaryCase: initcap
Schema-Scanner-2003-10-23 21:35:36-I> CommonAttributes:
CREATED_BY_USER,CREATE_DATE,RECORD_STATUS_CODE,STATUS_CODE,CREATE_USER

Schema-Scanner-2003-10-23 21:35:36-I> connecting to jdbc:oracle:thin:@tnsu2:1521:SYSDEV as webdemo...
Schema-Scanner-2003-10-23 21:35:39-I> connected
Schema-Scanner-2003-10-23 21:35:39-I> fetching comments.
Schema-Scanner-2003-10-23 21:35:39-I> fetched 1 comments.
Schema-Scanner-2003-10-23 21:35:39-I> fetching constraints .....
Schema-Scanner-2003-10-23 21:35:40-I> fetched 3 constraints.
Schema-Scanner-2003-10-23 21:35:40-I> fetching tables.
Schema-Scanner-2003-10-23 21:35:40-I> fetched 1 tables.
Schema-Scanner-2003-10-23 21:35:40-I> fetching columns .....
Schema-Scanner-2003-10-23 21:36:15-I> fetched 20 columns.
Schema-Scanner-2003-10-23 21:36:15-I> fetching indexes..
Schema-Scanner-2003-10-23 21:36:16-I> fetched 2 indexes.
Schema-Scanner-2003-10-23 21:36:16-I> fetching triggers
Schema-Scanner-2003-10-23 21:36:16-I> disconnected.
Schema-Scanner-2003-10-23 21:36:16-I> creating SchemaScanner output file scan.xml...
Schema-Scanner-2003-10-23 21:36:16-I> creating tnsgen output file ...
Schema-Scanner-2003-10-23 21:36:16-I> tnsgen output file created.
Schema-Scanner-2003-10-23 21:36:16-I> SchemaScanner scan complete.
Schema-Scanner-2003-10-23 21:36:16-I> Schema-Scanner complete. Elapsed time - 00:40.916

```

Using XSL Templates

SchemaScanner comes with a few XSL report formatting scripts to get you started using SchemaScanner XML output. A simple utility, xslReport has also been provided to process the XSL template against the SchemaScanner XML file. This utility requires two command line parameters:

```
-x[Name of XML file]
-s[Name of XSL file]
```

Sample xslReport Execution

The following is an example of xslReport using an XSL template to create an audit trigger for the sample table.

```

java xslRep -xxscan.ml -saudit_trigger.xml > donor.trg

xslRep : Non-Commercial and Educational Release 1.0 for jdk 1.4.2

Copyright (c) tns software inc., 1993-2003. All Rights Reserved

visit http://www.tns-soft.com for updates or commercial version

xslRep-2003-10-24 09:38:26-I> processing scan.xml using schema.xml...
xslRep-2003-10-24 09:38:32-I> processing complete

```


XSL Template

```
<?xml version="1.0"?>
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
xmlns:DateFormatter="DateFormatter">

<!-- timestamp variable uses java class -->
<xsl:variable name="date_formatter" select="DateFormatter:new()"/>

<xsl:output method="text" omit-xml-declaration="yes" indent="no"/>
<xsl:template match="/">
<xsl:apply-templates/>
</xsl:template>

<xsl:template match="Options">
-- Schema Scanner Version: <xsl:value-of select="SchemaScannerVersion"/>
-- Date Run: <xsl:value-of select="SchemaScannerRunDate"/>
-- Report Run: <xsl:value-of select="DateFormatter:Sysdate()"/>
-- Database Owners: <xsl:value-of select="SchemaScannerOwners"/>
-- Database Tables: <xsl:value-of select="SchemaScannerTables"/>
</xsl:template>

<xsl:template match="Entity">
<xsl:text>
--
-- Audit Trigger for </xsl:text><xsl:value-of select="@Name"/><xsl:text>
--
create or replace trigger </xsl:text><xsl:value-of select="@Name"/><xsl:text>_Trg
before insert or update or delete on </xsl:text><xsl:value-of select="@Name"/><xsl:text>
for each row
declare
Audit_Status_Code varchar(12);
procedure Insert_Audit is
begin
insert
into </xsl:text><xsl:value-of select="@Name"/><xsl:text>_audit
(
</xsl:text>
<xsl:for-each select="Attributes">
<xsl:for-each select="Attribute">
<!-- leading comma ( or not ) -->
<xsl:if test="position() != 1">
<xsl:text>
,</xsl:text>
</xsl:if>
<xsl:if test="position() = 1">
<xsl:text> </xsl:text>
</xsl:if>
<xsl:value-of select="@Name"/>
</xsl:for-each>
</xsl:for-each>
<xsl:text>
,Audit_Id
,Audit_Status_Code
,Audit_Date
,Audit_User
)
values (
</xsl:text>
<xsl:for-each select="Attributes">
<xsl:for-each select="Attribute">
<!-- leading comma ( or not ) -->
<xsl:if test="position() != 1">
<xsl:text>
```



```

        ,</xsl:text>
</xsl:if>
<xsl:if test="position() = 1">
  <xsl:text>          </xsl:text>
</xsl:if>
<xsl:text>decode(Audit_Status_Code,'INSERT',:New.</xsl:text>
<xsl:value-of select="@Name"/>
<xsl:text>,:Old.</xsl:text>
<xsl:value-of select="@Name"/>
<xsl:text>)</xsl:text>
</xsl:for-each>
</xsl:for-each>
<xsl:text>
      ,Audit_Id.NextVal
      ,Audit_Status_Code
      ,Sysdate
      ,User );
  end insert_audit;

begin
  if inserting then
    Audit_Status_Code := 'INSERT';
  elsif updating then
    Audit_Status_Code := 'UPDATE';
  elsif Deleting Then
    Audit_Status_Code := 'DELETE';
  end If;
  Insert_Audit;
end;
.
/
show errors
</xsl:text>
</xsl:template>
</xsl:stylesheet>

```

Generated Trigger

```

-- Schema Scanner Version: 1.0
-- Date Run: Tue Oct 21 23:01:49 EDT 2003
-- Report Run: 2003-10-22 08:31:33
-- Database Owners: 'WEBDEMO'
-- Database Tables: 'DONOR'
--
-- Audit Trigger for Donor
--
create or replace trigger Donor_Trigger
  before insert or update or delete on Donor
  for each row
declare
  Audit_Status_Code varchar(12);
  procedure Insert_Audit is
  begin
    insert
    into Donor_audit
    (
      Donor_Id
      ,First_Name
      ,Last_Name
      ,Last_Name_Upper
      ,Sex
      ,Record_Status_Code
      ,Address_1
      ,Address_2
      ,Address_3

```

```

, City
, Province
, Postal_Code
, Home_Phone
, Work_Phone
, Sin
, Date_Of_Birth
, Year_To_Date_Contributions
, Comments
, Dml_Date
, Dml_User
, Audit_Id
, Audit_Status_Code
, Audit_Date
, Audit_User
)
values (
    decode(Audit_Status_Code,'INSERT',:New.Donor_Id,:Old.Donor_Id)
, decode(Audit_Status_Code,'INSERT',:New.First_Name,:Old.First_Name)
, decode(Audit_Status_Code,'INSERT',:New.Last_Name,:Old.Last_Name)
, decode(Audit_Status_Code,'INSERT',:New.Last_Name_Upper,:Old.Last_Name_Upper)
, decode(Audit_Status_Code,'INSERT',:New.Sex,:Old.Sex)
, decode(Audit_Status_Code,'INSERT',:New.Record_Status_Code,:Old.Record_Status_Code)
, decode(Audit_Status_Code,'INSERT',:New.Address_1,:Old.Address_1)
, decode(Audit_Status_Code,'INSERT',:New.Address_2,:Old.Address_2)
, decode(Audit_Status_Code,'INSERT',:New.Address_3,:Old.Address_3)
, decode(Audit_Status_Code,'INSERT',:New.City,:Old.City)
, decode(Audit_Status_Code,'INSERT',:New.Province,:Old.Province)
, decode(Audit_Status_Code,'INSERT',:New.Postal_Code,:Old.Postal_Code)
, decode(Audit_Status_Code,'INSERT',:New.Home_Phone,:Old.Home_Phone)
, decode(Audit_Status_Code,'INSERT',:New.Work_Phone,:Old.Work_Phone)
, decode(Audit_Status_Code,'INSERT',:New.Sin,:Old.Sin)
, decode(Audit_Status_Code,'INSERT',:New.Date_Of_Birth,:Old.Date_Of_Birth)

, decode(Audit_Status_Code,'INSERT',:New.Year_To_Date_Contributions,:Old.Year_To_Date_Contributions)
, decode(Audit_Status_Code,'INSERT',:New.Comments,:Old.Comments)
, decode(Audit_Status_Code,'INSERT',:New.Dml_Date,:Old.Dml_Date)
, decode(Audit_Status_Code,'INSERT',:New.Dml_User,:Old.Dml_User)
, Audit_Id.NextVal
, Audit_Status_Code
, Sysdate
, User );
end insert_audit;

begin
if inserting then
    Audit_Status_Code := 'INSERT';
elseif updating then
    Audit_Status_Code := 'UPDATE';
elseif Deleting Then
    Audit_Status_Code := 'DELETE';
end If;
Insert_Audit;
end;
/
show errors

```

Using the Generated XML Output

SchemaScanner creates a well-formed XML output file that contains information extracted from Oracle Data Dictionary regarding each scanned entity including attributes, comments, triggers, and constraints. The level of detail provided is determined by the properties file and the amount of data present in the source schema.

Entity Information

The following table shows the properties that may be collected by *SchemaScanner* for each entity. Shaded properties are only collected where sufficient data exists and *ExtendedStatistics* is enabled in the properties file.

Property	Comment
Comment	Entity comment
Owner	Entity owner
Row Count	Row count
Primary Key Constraint	Primary Key Constraint
Foreign Key Constraint(s)	Foreign Key constraints
Related Tables	Related tables (based upon Foreign Key constraints)
Indexes	Indexes, excluding system generated indexes to support constraints
Triggers	Triggers for this table
SchemaCheck	Errors and warning associated with this entity and entity attributes

Attribute Information

The following table shows the properties that may be collected by *SchemaScanner* for each attribute by data type. Shaded properties are only collected where sufficient data exists and *ExtendedStatistics* is enabled in the properties file.

Property	Data Type			
	Char*	Number	Date	Other

	*			
Sequence	✓	✓	✓	✓
Comment	✓	✓	✓	✓
Constraints	✓	✓	✓	✓
Precision	✓	✓		
Scale		✓		
Nullable	✓	✓	✓	✓
Null Values	✓	✓	✓	✓
Percent Null	✓	✓	✓	✓
Min. Value	✓	✓	✓	
Max Value	✓	✓	✓	
Avg. Value		✓		
Sum Value		✓		
Distinct Values	✓	✓	✓	
Max. Duplicates	✓	✓	✓	
Min. Length	✓			
Max. Length	✓			
Avg. Length	✓			
Enum. Values	✓			
Histogram		✓		

* Char includes char, varchar and varchar2

* Number includes number, float and int

Example

The following is a sample DDL script that was manually defined and created in the database. Subsequently, *tnsgen* was used to generate 10,000 sample records and *tnsloader* was used to load the records into the Oracle database.

```
drop table Donor cascade constraints;

create table Donor
(
  Donor_Id          Number(10,0) not null
,First_Name        Varchar2(30) not null
,Last_Name         Varchar2(30) not null
,Last_Name_Upper   Varchar2(30) not null
,Sex               Varchar2(1)  not null
  check ( Sex in ( 'M','F' ))
,Record_Status_Code Varchar2(15) not null
  check ( Record_Status_Code in ( 'A', 'T', 'P' ))
,Address_1         Varchar2(70)
,Address_2         Varchar2(60)
,Address_3         Varchar2(60)
,City              Varchar2(30)
,Province          Varchar2(4)
,Postal_Code       Varchar2(11)
```

```

,Home_Phone          Varchar2(12)
,Work_Phone          Varchar2(12)
,Sin                 Varchar2(9)
,Date_Of_Birth       Date not null
,Year_To_Date_Contributions  Number(10,2) not null
,Comments            Varchar2(255)
,Dml_Date            Date default sysdate not null
,Dml_User            Varchar2(30) default user not null
,primary key ( Donor_Id )
);

create index Donor_Ie1 on Donor (Last_Name_Upper);

comment on column Donor.Record_Status_Code is 'Status code is (A)ctive, (T)erminated or (P)rospective';

```

Generated XML Output

The following is the SchemaScanner XML output generated for the preceding table. ExtendedStatistics was enabled in the properties file to enable the extended output which follows. The DTD for SchemaScanner XML output may be found in Appendix A.

```

<SchemaScanner>
  <Options>
    <SchemaScannerVersion>1.0</SchemaScannerVersion>
    <SchemaScannerRunDate>Thu Oct 23 16:15:01 EDT 2003</SchemaScannerRunDate>
    <SchemaScannerOwners>'WEBDEMO'</SchemaScannerOwners>
    <SchemaScannerTables>'DONOR'</SchemaScannerTables>
    <SchemaScannerMode>reverse-engineer</SchemaScannerMode>
    <SchemaScannerAnalyze>true</SchemaScannerAnalyze>
    <SchemaScannerStatistics>true</SchemaScannerStatistics>
  </Options>
  <Entities>
    <Entity Name="Donor">
      <Owner>Webdemo</Owner>
      <RowCount>10000</RowCount>
      <PrimaryKeyConstraint>
        <Constraint Type="Primary Key" Name="Sys_C0014029">
          <ConstraintAttribute>Donor_Id</ConstraintAttribute>
        </Constraint>
      </PrimaryKeyConstraint>
      <Indexes>
        <Index Name="Donor_Ie1" Unique="N">
          <IndexAttribute>Last_Name_Upper</IndexAttribute>
        </Index>
      </Indexes>
      <SchemaCheck>
        <Message Type="W">SS-W022: Table Donor has no defined relationships</Message>
        <Message Type="C">SS-C053: Attribute Sex: identifier may be too short</Message>
        <Message Type="E">SS-E020: Default value " for attribute Sex not found in constraint value list Sex in ( 'M','F' )</Message>
        <Message Type="E">SS-E020: Default value " for attribute Record_Status_Code not found in constraint value list Record_Status_Code in ( 'A', 'T', 'P' )</Message>
        <Message Type="W">SS-W042: Attribute Address_1 appears to part of repeating group</Message>
        <Message Type="W">SS-W042: Attribute Address_2 appears to part of repeating group</Message>
        <Message Type="W">SS-W042: Attribute Address_3 appears to part of repeating group</Message>
        <Message Type="C">SS-C053: Attribute Sin: identifier may be too short</Message>
      </SchemaCheck>
      <Attributes>
        <Attribute Name="Donor_Id">
          <Sequence>1</Sequence>

```

```

<DataType>Number</DataType>
<Precision>10</Precision>
<Scale>0</Scale>
<Nullable>N</Nullable>
<MinValue>1</MinValue>
<MaxValue>10000</MaxValue>
<DistinctValues>10000</DistinctValues>
</Attribute>
<Attribute Name="First_Name">
<Sequence>2</Sequence>
<DataType>Varchar2</DataType>
<Precision>30</Precision>
<Nullable>N</Nullable>
<MinValue>Aaron</MinValue>
<MaxValue>Zelma</MaxValue>
<DistinctValues>1828</DistinctValues>
<MaxDuplicates>15</MaxDuplicates>
<MinLength>2</MinLength>
<MaxLength>11</MaxLength>
<AvgLength>5</AvgLength>
</Attribute>
<Attribute Name="Last_Name">
<Sequence>3</Sequence>
<DataType>Varchar2</DataType>
<Precision>30</Precision>
<Nullable>N</Nullable>
<MinValue>Aaron</MinValue>
<MaxValue>Zuniga</MaxValue>
<DistinctValues>2926</DistinctValues>
<MaxDuplicates>6</MaxDuplicates>
<MinLength>2</MinLength>
<MaxLength>12</MaxLength>
<AvgLength>6</AvgLength>
</Attribute>
<Attribute Name="Last_Name_Upper">
<Sequence>4</Sequence>
<DataType>Varchar2</DataType>
<Precision>30</Precision>
<Nullable>N</Nullable>
<MinValue>AARON</MinValue>
<MaxValue>ZUNIGA</MaxValue>
<DistinctValues>2926</DistinctValues>
<MaxDuplicates>6</MaxDuplicates>
<MinLength>2</MinLength>
<MaxLength>12</MaxLength>
<AvgLength>6</AvgLength>
</Attribute>
<Attribute Name="Sex">
<Constraints>Sex in ( 'M', 'F' )</Constraints>
<Sequence>5</Sequence>
<DataType>Varchar2</DataType>
<Precision>1</Precision>
<Nullable>N</Nullable>
<MinValue>F</MinValue>
<MaxValue>M</MaxValue>
<DistinctValues>2</DistinctValues>
<MaxDuplicates>5215</MaxDuplicates>
<MinLength>1</MinLength>
<MaxLength>1</MaxLength>
<AvgLength>1</AvgLength>
<EnumValues>
<EnumValue Count="5215" PercentOfTotal="52">F</EnumValue>
<EnumValue Count="4785" PercentOfTotal="47">M</EnumValue>
</EnumValues>
</Attribute>
<Attribute Name="Record_Status_Code">
<Comment>Status code is (A)ctive, (T)erminated or (P)rospective</Comment>

```

```

<Constraints>Record_Status_Code in ( 'A', 'T', 'P' )</Constraints>
<Sequence>6</Sequence>
<DataType>Varchar2</DataType>
<Precision>15</Precision>
<Nullable>N</Nullable>
<MinValue>A</MinValue>
<MaxValue>T</MaxValue>
<DistinctValues>3</DistinctValues>
<MaxDuplicates>3342</MaxDuplicates>
<MinLength>1</MinLength>
<MaxLength>1</MaxLength>
<AvgLength>1</AvgLength>
<EnumValues>
  <EnumValue Count="3342" PercentOfTotal="33">A</EnumValue>
  <EnumValue Count="3324" PercentOfTotal="33">P</EnumValue>
  <EnumValue Count="3334" PercentOfTotal="33">T</EnumValue>
</EnumValues>
</Attribute>
<Attribute Name="Address_1">
  <Sequence>7</Sequence>
  <DataType>Varchar2</DataType>
  <Precision>70</Precision>
  <Nullable>Y</Nullable>
  <NullValues>3417</NullValues>
  <PctNull>34</PctNull>
  <DistinctValues>6583</DistinctValues>
  <MaxDuplicates>1</MaxDuplicates>
  <MinLength>9</MinLength>
  <MaxLength>29</MaxLength>
  <AvgLength>14</AvgLength>
</Attribute>
<Attribute Name="Address_2">
  <Sequence>8</Sequence>
  <DataType>Varchar2</DataType>
  <Precision>60</Precision>
  <Nullable>Y</Nullable>
  <NullValues>10000</NullValues>
  <DistinctValues>0</DistinctValues>
  <MaxDuplicates>0</MaxDuplicates>
  <MinLength>0</MinLength>
  <MaxLength>0</MaxLength>
  <AvgLength>0</AvgLength>
</Attribute>
<Attribute Name="Address_3">
  <Sequence>9</Sequence>
  <DataType>Varchar2</DataType>
  <Precision>60</Precision>
  <Nullable>Y</Nullable>
  <NullValues>10000</NullValues>
  <DistinctValues>0</DistinctValues>
  <MaxDuplicates>0</MaxDuplicates>
  <MinLength>0</MinLength>
  <MaxLength>0</MaxLength>
  <AvgLength>0</AvgLength>
</Attribute>
<Attribute Name="City">
  <Sequence>10</Sequence>
  <DataType>Varchar2</DataType>
  <Precision>30</Precision>
  <Nullable>Y</Nullable>
  <NullValues>3417</NullValues>
  <PctNull>34</PctNull>
  <MinValue>Abbotsford</MinValue>
  <MaxValue>Yorkton</MaxValue>
  <DistinctValues>852</DistinctValues>
  <MaxDuplicates>24</MaxDuplicates>
  <MinLength>3</MinLength>

```

```

    <MaxLength>30</MaxLength>
    <AvgLength>9</AvgLength>
  </Attribute>
  <Attribute Name="Province">
    <Sequence>11</Sequence>
    <DataType>Varchar2</DataType>
    <Precision>4</Precision>
    <Nullable>Y</Nullable>
    <NullValues>3417</NullValues>
    <PctNull>34</PctNull>
    <MinValue>AB</MinValue>
    <MaxValue>YT</MaxValue>
    <DistinctValues>13</DistinctValues>
    <MaxDuplicates>1965</MaxDuplicates>
    <MinLength>2</MinLength>
    <MaxLength>2</MaxLength>
    <AvgLength>2</AvgLength>
  </Attribute>
  <Attribute Name="Postal_Code">
    <Sequence>12</Sequence>
    <DataType>Varchar2</DataType>
    <Precision>11</Precision>
    <Nullable>Y</Nullable>
    <NullValues>3417</NullValues>
    <PctNull>34</PctNull>
    <MinValue>A0A8H4</MinValue>
    <MaxValue>Z9Y5K3</MaxValue>
    <DistinctValues>6583</DistinctValues>
    <MaxDuplicates>1</MaxDuplicates>
    <MinLength>6</MinLength>
    <MaxLength>6</MaxLength>
    <AvgLength>6</AvgLength>
  </Attribute>
  <Attribute Name="Home_Phone">
    <Sequence>13</Sequence>
    <DataType>Varchar2</DataType>
    <Precision>12</Precision>
    <Nullable>Y</Nullable>
    <NullValues>3540</NullValues>
    <PctNull>35</PctNull>
    <MinValue>200-1731</MinValue>
    <MaxValue>998-9912</MaxValue>
    <DistinctValues>6460</DistinctValues>
    <MaxDuplicates>1</MaxDuplicates>
    <MinLength>8</MinLength>
    <MaxLength>8</MaxLength>
    <AvgLength>8</AvgLength>
  </Attribute>
  <Attribute Name="Work_Phone">
    <Sequence>14</Sequence>
    <DataType>Varchar2</DataType>
    <Precision>12</Precision>
    <Nullable>Y</Nullable>
    <NullValues>5154</NullValues>
    <PctNull>51</PctNull>
    <MinValue>200-1368</MinValue>
    <MaxValue>998-7609</MaxValue>
    <DistinctValues>4846</DistinctValues>
    <MaxDuplicates>1</MaxDuplicates>
    <MinLength>8</MinLength>
    <MaxLength>8</MaxLength>
    <AvgLength>8</AvgLength>
  </Attribute>
  <Attribute Name="Sin">
    <Sequence>15</Sequence>
    <DataType>Varchar2</DataType>
    <Precision>9</Precision>

```



```

<Nullable>Y</Nullable>
<NullValues>3681</NullValues>
<PctNull>36</PctNull>
<MinValue>400012670</MinValue>
<MaxValue>998923585</MaxValue>
<DistinctValues>6304</DistinctValues>
<MaxDuplicates>2</MaxDuplicates>
<MinLength>9</MinLength>
<MaxLength>9</MaxLength>
<AvgLength>9</AvgLength>
</Attribute>
<Attribute Name="Date_Of_Birth">
<Sequence>16</Sequence>
<DataType>Date</DataType>
<Nullable>N</Nullable>
<MinValue>1925-01-01 00:00:00.0</MinValue>
<MaxValue>2002-11-29 00:00:00.0</MaxValue>
<DistinctValues>7620</DistinctValues>
<MaxDuplicates>4</MaxDuplicates>
</Attribute>
<Attribute Name="Year_To_Date_Contributions">
<Sequence>17</Sequence>
<DataType>Number</DataType>
<Precision>10</Precision>
<Scale>2</Scale>
<Nullable>N</Nullable>
<MinValue>0.0</MinValue>
<MaxValue>9997.0</MaxValue>
<AvgValue>5003.0</AvgValue>
<SumValue>5.003637894E7</SumValue>
<DistinctValues>9786</DistinctValues>
<MaxDuplicates>2</MaxDuplicates>
<Histogram>
<DataPoint>
<DataRange>0 to 999</DataRange>
<Count>984</Count>
<PercentOfTotal>9</PercentOfTotal>
</DataPoint>
<DataPoint>
<DataRange>1000 to 1998</DataRange>
<Count>989</Count>
<PercentOfTotal>9</PercentOfTotal>
</DataPoint>
<DataPoint>
<DataRange>1999 to 2997</DataRange>
<Count>1018</Count>
<PercentOfTotal>10</PercentOfTotal>
</DataPoint>
<DataPoint>
<DataRange>2998 to 3996</DataRange>
<Count>1027</Count>
<PercentOfTotal>10</PercentOfTotal>
</DataPoint>
<DataPoint>
<DataRange>3997 to 4995</DataRange>
<Count>973</Count>
<PercentOfTotal>9</PercentOfTotal>
</DataPoint>
<DataPoint>
<DataRange>4996 to 5994</DataRange>
<Count>999</Count>
<PercentOfTotal>9</PercentOfTotal>
</DataPoint>
<DataPoint>
<DataRange>5995 to 6993</DataRange>
<Count>1015</Count>
<PercentOfTotal>10</PercentOfTotal>

```

```

</DataPoint>
<DataPoint>
  <DataRange>6994 to 7992</DataRange>
  <Count>967</Count>
  <PercentOfTotal>9</PercentOfTotal>
</DataPoint>
<DataPoint>
  <DataRange>7993 to 8991</DataRange>
  <Count>995</Count>
  <PercentOfTotal>9</PercentOfTotal>
</DataPoint>
<DataPoint>
  <DataRange>8992 to 9990</DataRange>
  <Count>1025</Count>
  <PercentOfTotal>10</PercentOfTotal>
</DataPoint>
<DataPoint>
  <DataRange>9991 to 10989</DataRange>
  <Count>8</Count>
  <PercentOfTotal>0</PercentOfTotal>
</DataPoint>
</Histogram>
</Attribute>
<Attribute Name="Comments">
  <Sequence>18</Sequence>
  <DataType>Varchar2</DataType>
  <Precision>255</Precision>
  <Nullable>Y</Nullable>
  <NullValues>9244</NullValues>
  <PctNull>92</PctNull>
  <DistinctValues>752</DistinctValues>
  <MaxDuplicates>2</MaxDuplicates>
  <MinLength>1</MinLength>
  <MaxLength>253</MaxLength>
  <AvgLength>124</AvgLength>
</Attribute>
<Attribute Name="Dml_Date">
  <Sequence>19</Sequence>
  <DataType>Date</DataType>
  <Nullable>N</Nullable>
  <Default>sysdate</Default>
  <MinValue>2003-10-21 12:11:14.0</MinValue>
  <MaxValue>2003-10-21 15:51:01.0</MaxValue>
  <DistinctValues>89</DistinctValues>
  <MaxDuplicates>142</MaxDuplicates>
</Attribute>
<Attribute Name="Dml_User">
  <Sequence>20</Sequence>
  <DataType>Varchar2</DataType>
  <Precision>30</Precision>
  <Nullable>N</Nullable>
  <Default>user</Default>
  <MinValue>SYSADMIN</MinValue>
  <MaxValue>WEBDEMO</MaxValue>
  <DistinctValues>2</DistinctValues>
  <MaxDuplicates>9999</MaxDuplicates>
  <MinLength>7</MinLength>
  <MaxLength>8</MaxLength>
  <AvgLength>7</AvgLength>
  <EnumValues>
    <EnumValue Count="1" PercentOfTotal="0">SYSADMIN</EnumValue>
    <EnumValue Count="9999" PercentOfTotal="99">WEBDEMO</EnumValue>
  </EnumValues>
</Attribute>
</Attributes>
</Entity>
</Entities>

```

Schema Report

The following is a reverse-engineered DDL script created from the preceding XML output using a supplied XSL formatting script. *AnalyzeSchema* was enabled in the properties file to enable the schema check messages embedded in the output. You can edit the XSL template file to create the output schema in any format you choose, or even use the file to create HTML schema documentation!

```
-- Schema Scanner Version: 1.0
-- Date Run: Thu Oct 23 16:15:01 EDT 2003
-- Report Run: 2003-10-23 16:15:21
-- Database Owners: 'WEBDEMO'
-- Database Tables: 'DONOR'

--
-- Entity Donor
--
-- SS-W022: Table Donor has no defined relationships
-- SS-C053: Attribute Sex: identifier may be too short
-- SS-E020: Default value "" for attribute Sex not found in constraint value list Sex in ( 'M','F' )
-- SS-E020: Default value "" for attribute Record_Status_Code not found in constraint value list
Record_Status_Code in ( 'A', 'T', 'P' )
-- SS-W042: Attribute Address_1 appears to part of repeating group
-- SS-W042: Attribute Address_2 appears to part of repeating group
-- SS-W042: Attribute Address_3 appears to part of repeating group
-- SS-C053: Attribute Sin: identifier may be too short

drop table Donor cascade constraints;

create table Donor
(
  Donor_Id          Number(10,0) not null
  ,First_Name       Varchar2(30) not null
  ,Last_Name        Varchar2(30) not null
  ,Last_Name_Upper  Varchar2(30) not null
  ,Sex              Varchar2(1) not null
  check ( Sex in ( 'M','F' ) )
  ,Record_Status_Code Varchar2(15) not null
  check ( Record_Status_Code in ( 'A', 'T', 'P' ) )
  ,Address_1        Varchar2(70)
  ,Address_2        Varchar2(60)
  ,Address_3        Varchar2(60)
  ,City             Varchar2(30)
  ,Province         Varchar2(4)
  ,Postal_Code      Varchar2(11)
  ,Home_Phone       Varchar2(12)
  ,Work_Phone       Varchar2(12)
  ,Sin              Varchar2(9)
  ,Date_Of_Birth    Date not null
  ,Year_To_Date_Contributions Number(10,2) not null
  ,Comments         Varchar2(255)
  ,Dml_Date         Date default sysdate not null
  ,Dml_User         Varchar2(30) default user not null
  ,primary key ( Donor_Id )
);

create index Donor_1e1 on Donor (Last_Name_Upper);
```

comment on column Donor.Record_Status_Code is 'Status code is (A)ctive, (T)erminated or (P)rospective';

Attribute Report

The following is an Attribute Report created from the preceding XML output using a supplied XSL formatting script. *ExtendedStatistics* was enabled in the properties file to enable the statistical collection of data characteristics. The format has been kept purposely simple, but as before you can edit the supplied XSL template to create the attribute report in any format you require.

Schema Scanner Version: 1.0
Date Run: Thu Oct 23 16:15:01 EDT 2003
Report Run: 2003-10-23 19:43:31
Database Owners: 'WEBDEMO'
Database Tables: 'DONOR'

Database Statistics for Entity Donor

Row Count:10000

Attribute: Donor_Id

Sequence: 1
Data Type: Number
Precision: 10
Scale: 0
Nullable: N
Minimum Value: 1
Maximum Value: 10000
Distinct Values: 10000

Attribute: First_Name

Sequence: 2
Data Type: Varchar2
Precision: 30
Nullable: N
Minimum Value: Aaron
Maximum Value: Zelma
Distinct Values: 1828
Maximum Duplicates: 15
Minimum Length: 2
Maximum Length: 11
Average Length: 5

Attribute: Last_Name

Sequence: 3
Data Type: Varchar2
Precision: 30
Nullable: N
Minimum Value: Aaron
Maximum Value: Zuniga
Distinct Values: 2926
Maximum Duplicates: 6
Minimum Length: 2
Maximum Length: 12
Average Length: 6

Attribute: Last_Name_Upper

Sequence: 4
Data Type: Varchar2
Precision: 30
Nullable: N
Minimum Value: AARON
Maximum Value: ZUNIGA
Distinct Values: 2926
Maximum Duplicates: 6
Minimum Length: 2
Maximum Length: 12
Average Length: 6

Attribute: Sex

Sequence: 5
Data Type: Varchar2
Precision: 1
Nullable: N
Constraints: Sex in ('M','F')
Minimum Value: F
Maximum Value: M
Distinct Values: 2
Maximum Duplicates: 5215
Minimum Length: 1
Maximum Length: 1
Average Length: 1

Count	Percent	Value
-------	---------	-------

5215	52	F
4785	47	M

Attribute: Record_Status_Code

Sequence: 6
Data Type: Varchar2
Precision: 15
Nullable: N
Constraints: Record_Status_Code in ('A', 'T', 'P')
Minimum Value: A
Maximum Value: T
Distinct Values: 3
Maximum Duplicates: 3342
Minimum Length: 1
Maximum Length: 1
Average Length: 1

Count	Percent	Value
-------	---------	-------

3342	33	A
3324	33	P
3334	33	T

Attribute: Address_1

Sequence: 7
Data Type: Varchar2
Precision: 70
Nullable: Y
Null Count: 3417
Percent Null: 34

Distinct Values: 6583
Maximum Duplicates: 1
Minimum Length: 9
Maximum Length: 29
Average Length: 14

Attribute: Address_2

Sequence: 8
Data Type: Varchar2
Precision: 60
Nullable: Y
Null Count: 10000
Distinct Values: 0
Maximum Duplicates: 0
Minimum Length: 0
Maximum Length: 0
Average Length: 0

Attribute: Address_3

Sequence: 9
Data Type: Varchar2
Precision: 60
Nullable: Y
Null Count: 10000
Distinct Values: 0
Maximum Duplicates: 0
Minimum Length: 0
Maximum Length: 0
Average Length: 0

Attribute: City

Sequence: 10
Data Type: Varchar2
Precision: 30
Nullable: Y
Null Count: 3417
Percent Null: 34
Minimum Value: Abbotsford
Maximum Value: Yorkton
Distinct Values: 852
Maximum Duplicates: 24
Minimum Length: 3
Maximum Length: 30
Average Length: 9

Attribute: Province

Sequence: 11
Data Type: Varchar2
Precision: 4
Nullable: Y
Null Count: 3417
Percent Null: 34
Minimum Value: AB
Maximum Value: YT
Distinct Values: 13
Maximum Duplicates: 1965
Minimum Length: 2
Maximum Length: 2
Average Length: 2

Attribute: Postal_Code

Sequence: 12

Data Type: Varchar2
Precision: 11
Nullable: Y
Null Count: 3417
Percent Null: 34
Minimum Value: A0A8H4
Maximum Value: Z9Y5K3
Distinct Values: 6583
Maximum Duplicates: 1
Minimum Length: 6
Maximum Length: 6
Average Length: 6

Attribute: Home_Phone

Sequence: 13
Data Type: Varchar2
Precision: 12
Nullable: Y
Null Count: 3540
Percent Null: 35
Minimum Value: 200-1731
Maximum Value: 998-9912
Distinct Values: 6460
Maximum Duplicates: 1
Minimum Length: 8
Maximum Length: 8
Average Length: 8

Attribute: Work_Phone

Sequence: 14
Data Type: Varchar2
Precision: 12
Nullable: Y
Null Count: 5154
Percent Null: 51
Minimum Value: 200-1368
Maximum Value: 998-7609
Distinct Values: 4846
Maximum Duplicates: 1
Minimum Length: 8
Maximum Length: 8
Average Length: 8

Attribute: Sin

Sequence: 15
Data Type: Varchar2
Precision: 9
Nullable: Y
Null Count: 3681
Percent Null: 36
Minimum Value: 400012670
Maximum Value: 998923585
Distinct Values: 6304
Maximum Duplicates: 2
Minimum Length: 9
Maximum Length: 9
Average Length: 9

Attribute: Date_Of_Birth

Sequence: 16
Data Type: Date
Precision:
Nullable: N

Minimum Value: 1925-01-01 00:00:00.0
Maximum Value: 2002-11-29 00:00:00.0
Distinct Values: 7620
Maximum Duplicates: 4

Attribute: Year_To_Date_Contributions

Sequence: 17
Data Type: Number
Precision: 10
Scale: 2
Nullable: N
Minimum Value: 0.0
Maximum Value: 9997.0
Distinct Values: 9786
Maximum Duplicates: 2

Count	Percent	Range
984	9	0 to 999
989	9	1000 to 1998
1018	10	1999 to 2997
1027	10	2998 to 3996
973	9	3997 to 4995
999	9	4996 to 5994
1015	10	5995 to 6993
967	9	6994 to 7992
995	9	7993 to 8991
1025	10	8992 to 9990
8	0	9991 to 10989

Attribute: Comments

Sequence: 18
Data Type: Varchar2
Precision: 255
Nullable: Y
Null Count: 9244
Percent Null: 92
Distinct Values: 752
Maximum Duplicates: 2
Minimum Length: 1
Maximum Length: 253
Average Length: 124

Attribute: Dml_Date

Sequence: 19
Data Type: Date
Precision:
Nullable: N
Minimum Value: 2003-10-21 12:11:14.0
Maximum Value: 2003-10-21 15:51:01.0
Distinct Values: 89
Maximum Duplicates: 142

Attribute: Dml_User

Sequence: 20
Data Type: Varchar2
Precision: 30
Nullable: N
Minimum Value: SYSADMIN
Maximum Value: WEBDEMO
Distinct Values: 2

Maximum Duplicates: 9999
Minimum Length: 7
Maximum Length: 8
Average Length: 7

Count	Percent	Value
1	0	SYSADMIN
9999	99	WEBDEMO

Schema Errors

SchemaScanner examines the entities specified in the properties file and generates error messages for all discovered schema discrepancies. The messages relate to variances from generally recognized modeling practice and are pre-fixed by a severity indicator as follows:

- E – Error which should be corrected
- W – Warning which should be investigated
- C – Caution which should be considered
- I – Information which should be considered

Keyword and Reserved Words used for error generation may be found in Appendix B and C. The messages that may be generated and a brief explanation as to the nature of the error immediately follow.

<i>Msg</i>	<i>Error Text</i>	<i>Error Description</i>
E001	Attribute <Attribute Name> is a reserved word	No object identifiers should be reserved words since this may cause incorrect operation
W002	Attribute <Attribute Name> is a keyword	No object identifiers should be keywords since this may cause incorrect operation
W003	Table <Entity Name> contains more than <Max Recommended Attributes> attributes	Tables with too many attributes are slower to query and update and usually indicates poor data modeling or excessive database de-normalization
E004	Table <Entity Name> has no primary key defined	Every table should have a primary key defined using a formal constraint and not merely a unique index
E005	Attribute <Attribute Name> has inconsistent domain. Defined as <Data Type> in <Entity Name> and <Data Type> in <Entity Name>	Attributes should be defined consistently throughout a schema. This is particularly important for middleware based systems using CORBA or Tuxedo.
E006	Attribute <Attribute Name> has inconsistent precision. Defined as	Attributes should be defined consistently throughout a schema. This is particularly

	<Precision> <Entity Name> and <Precision> in <Entity Name>	important for middleware based systems using CORBA or Tuxedo.
E007	Attribute <Attribute Name> has inconsistent scale. Defined as <Scale> in <Entity Name> and <Scale> in <Entity Name>	Attributes should be defined consistently throughout a schema. This is particularly important for middleware based systems using CORBA or Tuxedo.
W008	Attribute <Attribute Name> has inconsistent default. Defined as '<Default Value>' in <Entity Name> and <Default Value> in <Entity Name>	Attributes should be defined consistently throughout a schema. If a default value cannot be consistently defined for all instances of the attribute, omit the default value entirely to avoid confusion.
W009	Attribute <Attribute Name> has inconsistent constraints. Defined as '<Attribute Constraint>' in <Entity Name> and '<Attribute Constraint>' in <Entity Name>	Attributes should be defined consistently throughout a schema. If a constraint cannot be consistently defined, omit the constraint entirely to avoid confusion.
W010	Table <Entity Name> has a compound primary key consisting of <Attribute Count> attributes	Compound primary keys often introduce problems for queries, particularly where queries are executed in a "pseudo-conversational manner", where context must be established for queries. Consider the use of a system generated (surrogate) key.
W011	Dml-date attribute <Attribute Name> is not mandatory	Where used, a value for dml-date must be mandatory.
W012	Dml-date attribute <Attribute Name> has no default value	Where used, a default for dml-date must be present and is typically valued as sysdate.
W013	Dml-user attribute <Attribute Name> is not mandatory	Where used, a value for dml-user must be mandatory.
W014	Dml-user attribute <Attribute Name> has no default value	Where used, a default for dml-user must be present and us typically valued as user.
W015	Table <Entity Name> has no dml-user attribute	Every entity must have a dml-user attribute.
W016	Table <Entity Name> has no dml-date attribute	Every entity must have a dml-date attribute.
W017	Dml-date attribute <Attribute Name> has invalid data type <Data Type>	Where used, the data type for dml-date must be date.
W018	Dml-user attribute <Attribute Name> has invalid data type <Data Type>	Where used, the data type for dml-user must be varchar2(30).

E019	Constraint for attribute <Attribute Name> is single-valued constraining data to a constant	If a constraint permits only a single value then the attribute is a constant and should not be present in the entity.
E020	Default value '<Default Value>' for attribute <Attribute Name> not found in constraint value list <Attribute Constraint List>	If a default value is defined for a constrained attribute, the value must be present in the constraint, otherwise insert operations will fail where the attribute is undefined.
W021	Primary key attribute (<Attribute Name>) has a date data type	The data type is a poor choice for a primary key, since it is easy to insert duplicate values and difficult to query. Consider the use of a system generated (surrogate) key.
W022	Table <Entity Name> has no defined relationships	Where no foreign keys are present and no other entity references the table no relationships are present to the table. This indicates either a single table schema or the absence of foreign key constraints.
W023	Table <Entity Name> is primary entity but has no indexes defined	Where an entity has dependent tables as defined by foreign key constraints, there almost always a requirement to access the table through a variety of attributes. The absence of indexes indicates that no other access paths are supported.
W024	Table <Entity Name> has index <Index Name> contained within index <Index Name>	When an index is contained within an index there is the possibility to confuse the optimizer or even cause queries to be executed using a full table scan. The contained indexes should be dropped.
W025	Table <Entity Name> has too many (<Total Count>) user-defined indexes	When an entity has an excessive number of indexes, insert and update performance will suffer due to index balancing operations and the query optimizer may become confused as to which index to use. Carefully consider all indexing and avoid the temptation to "shotgun" index creation to save time.
W026	Table <Entity Name> has too little data to warrant user-defined indexes	Where a table has only a small number of rows, indexes slow query performance and are unnecessary, unless used to constrain data to unique values.
W027	Table <Entity Name> has the same primary key definition as	Where two tables have the same primary key, the schema is either excessively

	<Entity Name>	generalized, for example where every table is identified by a primary key of obj_id yielding a valueless schema, or where errors in normalization have been made.
E028	Table name <Entity Name> is a reserved word	No object identifiers should be reserved words since this may cause incorrect operation
W029	Table name <Entity Name> is a keyword	No object identifiers should be keywords words since this may cause incorrect operation
E030	Primary key constraint name <Constraint Name> is a reserved word	No object identifiers should be reserved words since this may cause incorrect operation
W031	Primary key constraint name <Constraint Name> is a keyword	No object identifiers should be keywords words since this may cause incorrect operation
E032	Attribute <Attribute Name> has both constraint and foreign key reference	A constrained attribute should have a foreign key constraint and vice-versa.
E033	Attribute <Attribute Name> has both default and foreign key reference	An attribute should not have a foreign key and a default value, since the default value may not necessarily be found in the related table
W034	Attribute <Attribute Name> has incorrect functional dependency – also present in unrelated table <Entity Name>	Where attribute names are used in unrelated tables. The schema is either excessively generalized or incorrectly normalized. Exceptions are surrogate attributes, such as dml_date, dml_user, etc, which are used universally. SchemaScanner permits the definition of these surrogate attributes to avoid spurious occurrences of this message.
W035	Primary key attribute (<Attribute Name>) should be constrained to integer values	Where a numeric primary key attribute is not constrained to integer values, it is possible to insert primary keys values which would be difficult to access.
W036	Attribute <Attribute Name> should be present in foreign key constraint to <Entity Name>	Based upon the attribute name, a foreign key constraint should be defined to the indicated table
I037	Constraint list for attribute <Attribute Name> is > <Max Coded Values> and may be better expressed as a table lookup	Where a constraint list contains an excessive number of values, it usually indicates that a foreign key relationship would be more maintainable

W038	Attribute <Attribute Name> name is inconsistent with data type	Where a name uses a data type suffix, such as shipping_date, one reasonably expects that the data type for this attribute would be date.
W039	Primary key attribute (<Attribute Name>) is too large for inclusion in primary key	An attribute with excessive precision is a poor choice for a primary key, due to the difficulty in using the long value and the poor performance associated with large indexed attributes.
W040	Constraint for attribute <Attribute Name> is not uppercase only	Constraint values should be constrained to upper case values to avoid confusion and simplify UI programming.
W041	Attribute <Attribute Name> has constraint and is primary key attribute	Primary key attributes should not be constrained since this limits the possible values for the identifier.
W042	Attribute <Attribute Name> appears to part of repeating group	Repeating groups generally indicate poor modeling or excessive de-normalization. Avoid their use wherever possible.
W043	Attribute <Attribute Name> has precision > <Max Portable Length> which may restrict portability	Where the Oracle database may be exported to other databases, the maximum attribute length should be considered to avoid difficulties importing data.
E044	Attribute <Attribute Name> foreign key data type differs	The data type for the referenced foreign key isn't the same, causing potential problems with key referencing.
E045	Attribute <Attribute Name> foreign key precision differs	The precision for the referenced foreign key isn't the same, causing potential problems with key referencing.
E046	Attribute <Attribute Name> foreign key scale differs	The scale for the referenced foreign key isn't the same, causing potential problems with key referencing.
W047	Attribute <Attribute Name> foreign key attribute should be indexed	Foreign key attributes are frequently referenced in join operations and often benefit from the creation of a duplicates allowed index.
E048	Index name <Index Name> is a reserved word	No object identifiers should be reserved words since this may cause incorrect operation
W049	Index name <Index Name> is a keyword	No object identifiers should be keywords words since this may cause incorrect operation
W050	Attribute <Attribute Name> in index <Index Name> is too large	An attribute with excessive precision is a poor choice for a indexing, due to the

	for efficient indexing	difficulty in using the long value and the poor performance associated with large indexed attributes.
W051	Index <Index Name> has too many attributes	Indexes with many attributes perform poorly for insert and update operations and are often difficult to compose through a UI.
C052	Attribute <Attribute Name> is disenvowled	The mainframe practice of removing vowels from object names can lead to short, cryptic names and is unnecessary given Oracle's 30-character object name limit.
C053	Attribute <Attribute Name> - identifier may be too short	Short identifiers may be difficult to understand, unless it is well-understood acronym or conventional short form.
C053	Table <Entity Name> has no table or column comments	Comments are invaluable in providing additional information for a schema. The absence of comments is often a leading indicator of other poor modeling practices.

Tnsgen Output

SchemaScanner optionally produces output scripts that are compatible with tnsgen, TNS Software's test data generation tool. These tools ideally complement each other within development and test environments, particularly when you've got to get it right the first time.

There are a number of possible scenarios that require the generation of test data. Let's consider the most common possibilities.

Legacy Conversion

In this scenario, data resides in a legacy system such as RMS, VSAM or other non-relational repository, or exists in a primitive relational form without referential or data integrity constraints. In either case, the quality of the source data is suspect and direct data conversion will undoubtedly encounter errors. SchemaScanner in conjunction with tnsgen can help you define and populate your new database using generated data, while your legacy conversion team is free to perfect their data conversion routines.

Our test data generation method starts with a normalized, constrained and unpopulated database. We must consider the database constraints to be sufficient to correctly govern the generation of test data. A brief article regarding schema naming and constraint practices may be found at http://www.tns-soft.com/tns_schema_standards.pdf.

Strict Privacy Requirements

In this scenario, the source of the data is largely irrelevant, since the use of the data is proscribed either by privacy policy or by privacy legislation. In many cases, techniques such as using production data, which has undergone *data scrambling*, are insufficient to protect the source data; confidential information still resides within the database and may be inferred and extracted for malicious purposes. Where the data source is a legacy system, the problem is compounded; the data is not reliable and may not readily be available.

Where data is readily available, test data generation again starts with a normalized, constrained and unpopulated database. Again, we must consider the database constraints to be sufficient to correctly govern the generation of test data.

Where data does exist and is readily available test data generation method starts with a normalized, constrained and populated database. Where consistency constraints exist and are enabled, and where no constraint is violated, the current database is a valid model and satisfies finite requirements. This will permit the creation of a reasonably faithful copy of the source data, with no possibility of a breach of privacy or confidentiality.

Load Testing / Benchmarking

In this scenario, insufficient data volumes exist to load test or benchmark the target application. This may again be compounded by privacy policy / legislation or legacy conversion issues.

Our test data generation method starts with a normalized, constrained and unpopulated database. We must again consider the database constraints to be sufficient to correctly govern the generation of test data.

New Application

In this scenario, no data exists to populate the target system. Quality assurance testing and performance testing require reasonable data volumes to populate and test the target database.

Our test data generation method starts with a normalized, constrained and unpopulated database. We must again consider the database constraints to be sufficient to correctly govern the generation of test data.

Summary

These four common scenarios lead to two simple types of test data generation:

- Test Data generation where valid data may be sampled and extended attribute domain characteristics collected to create a test database that substantially mirrors the source database.
- Test Data generation where data is unavailable allowing the creation of a test database that relies upon constraint definitions to define the target database.

Example

The following is the tnsngen script generated by SchemaScanner and used to generate test data for the previous sample table. This script was derived from the initial, unpopulated schema that included attribute constraints. This file was subsequently loaded into the database using tnsloader and finally SchemaScanner was again used to generate the XML file used in previous examples.

```
; Generated By Schema-Scanner Version: 1.0 on 2003-10-21 11:47:41
;
; File Name: tnsngen.ini
; RCS_ID $Id
; Copyright (c) tns software inc., 1993-2003. All Rights Reserved
;
[options]
debug                true
random-seed          false
date-mask            "%D-%h-%Y"
datetime-mask        "%D-%h-%Y %l:%j:%k"
null-value           ""

[input-definition]
file-name            ./data/cdn_area_codes.csv
field-count          3
field-names           cdn_area_code cdn_area_city cdn_area_prov

file-name            ./data/cdn_cities.csv
field-count          2
field-names           cdn_city cdn_prov_code

file-name            ./data/passwords.csv
field-count          1
field-names           password

file-name            ./data/country.csv
field-count          2
field-names           country_code country_desc

file-name            ./data/currency.csv
field-count          2
field-names           currency_code currency_desc

file-name            ./data/first_names_female.csv
Field-count          2
field-names           female_name female_sex_ind

file-name            ./data/first_names_male.csv
field-count          2
field-names           male_name male_sex_ind

file-name            ./data/first_names.csv
field-count          2
field-names           first_name sex_ind

file-name            ./data/surname.csv
field-count          2
field-names           surname surname_initial

file-name            ./data/cdn_provinces.csv
field-count          2
field-names           prov_code prov_desc
```

```

file-name ./data/us_states.csv
field-count 2
field-names state_code state_desc

file-name ./data/stocks.csv
field-count 2
field-names ticker_symbol company_name

file-name ./data/street_names.csv
field-count 1
field-names street_name

file-name ./data/us_area_codes.csv
field-count 3
field-names us_area_code us_area_city us_area_state

file-name ./data/us_cities.csv
field-count 2
field-names us_city_name us_state

[output-definition]
file-name "donor.dat"
output-mode replace
output-format csv
record-name donor
record-count 10000

repeat-record 1

field-name donor_id
field-length 10
field-mask "%d"
field-value sequential-number 1

field-name first_name
field-length 30
field-mask "%s"
; data source is derived from name match
field-value first_name

field-name last_name
field-length 30
field-mask "%s"
; data source is derived from name match
field-value surname

field-name last_name_upper
field-length 30
field-mask "%s"
; data source is derived from name match
field-value dup-prev
conversion upper

field-name sex
field-length 1
field-mask "%s"
field-value sex_ind

field-name record_status_code
field-length 15
field-mask "%s"
; random-list is derived from schema constraint
field-value random-list "A,T,P"
; conversion based upon schema constraint
conversion upper

```

```

field-name    street_number
field-length  5
field-mask    "%d "
percent-null  30
field-value   random-integer
low-value     1
high-value    4000
skip-formatting true

field-name    Address1
field-length  60
field-mask    "%s,"
field-value   street_name
skip-prior-null true
skip-formatting true

field-name    address_2
field-length  60
field-mask    "%s"
; percent-null value is estimated
; data source is derived from name match
field-value   ""
skip-prior-null true

field-name    address_3
field-length  60
field-mask    "%s"
; percent-null value is estimated
; data source is derived from name match
field-value   ""
skip-prior-null true

field-name    city
field-length  30
field-mask    "%s"
field-value   cdn_city
skip-prior-null true

field-name    province
field-length  4
field-mask    "%s"
field-value   cdn_prov_code
conversion    upper
skip-prior-null true

field-name    postal_code
field-length  11
field-mask    "%s"
; data source is derived from name match
field-value   random-po
; conversion based upon conversion rules
conversion    upper
skip-prior-null true

field-name    home_phone
field-length  12
field-mask    "%s"
percent-null  30
; data source is derived from name match
field-value   random-phone
; conversion based upon conversion rules
conversion    upper

field-name    work_phone
field-length  12
field-mask    "%s"
percent-null  50

```

```
; data source is derived from name match
field-value      random-phone
; conversion based upon conversion rules
conversion       upper

field-name       sin
field-length     9
field-mask       "%s"
; data source is derived from name match
field-value      random-sin
percent-null     35
; conversion based upon conversion rules
conversion       upper

field-name       date_of_birth
field-length     15
field-mask       "%s"
field-value      random-date
; low/high values based upon name match
low-value        01-jan-1925
high-value       01-jan-2003

field-name       year_to_date_contributions
field-length     10
; field mask based upon schema scale
field-mask       "%.2f"
field-value      random-float
; low/high values based upon schema definition
low-value        "0"
high-value       "9999"

field-name       comments
field-length     255
field-mask       "%s"
; percent-null value is estimated
percent-null     95
field-value      random-word
```

Appendix A - SchemaScanner DTD

<!ELEMENT Attribute (AvgLength | AvgValue | Comment | Constraints | DataType | Default | DistinctValues | EnumValues | Histogram | MaxDuplicates | MaxLength | MaxValue | MinLength | MinValue | NullValues | Nullable | PctNull | Precision | Scale | Sequence | SumValue)* >

<!ATTLIST Attribute Name ID #REQUIRED >

<!ELEMENT Attributes (Attribute+) >

<!ELEMENT AvgLength (#PCDATA) >

<!ELEMENT AvgValue (#PCDATA) >

<!ELEMENT Comment (#PCDATA) >

<!ELEMENT Constraint (ConstraintAttribute) >

<!ATTLIST Constraint Name NMTOKEN #REQUIRED >

<!ATTLIST Constraint Type CDATA #REQUIRED >

<!ELEMENT ConstraintAttribute (#PCDATA) >

<!ELEMENT Constraints (#PCDATA) >

<!ELEMENT Count (#PCDATA) >

<!ELEMENT DataPoint (DataRange, Count, PercentOfTotal) >

<!ELEMENT DataRange (#PCDATA) >

<!ELEMENT DataType (#PCDATA) >

<!ELEMENT Default (#PCDATA) >

<!ELEMENT DistinctValues (#PCDATA) >

<!ELEMENT Entities (Entity) >

<!ELEMENT Entity (Owner, RowCount, PrimaryKeyConstraint, Indexes, SchemaCheck, Attributes) >

<!ATTLIST Entity Name NMTOKEN #REQUIRED >

<!ELEMENT EnumValue (#PCDATA) >

<!ATTLIST EnumValue Count NMTOKEN #REQUIRED >

<!ATTLIST EnumValue PercentOfTotal NMTOKEN #REQUIRED >

<!ELEMENT EnumValues (EnumValue+) >

<!ELEMENT Histogram (DataPoint+) >

<!ELEMENT Index (IndexAttribute) >

<!ATTLIST Index Name NMTOKEN #REQUIRED >

<!ATTLIST Index Unique NMTOKEN #REQUIRED >

<!ELEMENT IndexAttribute (#PCDATA) >

<!ELEMENT Indexes (Index) >

<!ELEMENT MaxDuplicates (#PCDATA) >
<!ELEMENT MaxLength (#PCDATA) >
<!ELEMENT MaxValue (#PCDATA) >
<!ELEMENT Message (#PCDATA) >
<!ATTLIST Message Type NMTOKEN #REQUIRED >
<!ELEMENT MinLength (#PCDATA) >
<!ELEMENT MinValue (#PCDATA) >
<!ELEMENT NullValues (#PCDATA) >
<!ELEMENT Nullable (#PCDATA) >
<!ELEMENT Options (SchemaScannerVersion, SchemaScannerRunDate, SchemaScannerOwners, SchemaScannerTables, SchemaScannerMode, SchemaScannerAnalyze, SchemaScannerStatistics) >
<!ELEMENT Owner (#PCDATA) >
<!ELEMENT PctNull (#PCDATA) >
<!ELEMENT PercentOfTotal (#PCDATA) >
<!ELEMENT Precision (#PCDATA) >
<!ELEMENT PrimaryKeyConstraint (Constraint) >
<!ELEMENT RowCount (#PCDATA) >
<!ELEMENT Scale (#PCDATA) >
<!ELEMENT SchemaCheck (Message+) >
<!ELEMENT SchemaScanner (Options, Entities) >
<!ELEMENT SchemaScannerAnalyze (#PCDATA) >
<!ELEMENT SchemaScannerMode (#PCDATA) >
<!ELEMENT SchemaScannerOwners (#PCDATA) >
<!ELEMENT SchemaScannerRunDate (#PCDATA) >
<!ELEMENT SchemaScannerStatistics (#PCDATA) >
<!ELEMENT SchemaScannerTables (#PCDATA) >
<!ELEMENT SchemaScannerVersion (#PCDATA) >
<!ELEMENT Sequence (#PCDATA) >
<!ELEMENT SumValue (#PCDATA) >

Appendix B – Oracle Keywords

ADMIN	AFTER	ALLOCATE	ANALYZE
ARCHIVE	ARCHIVELOG	AUTHORIZATION	BACKUP
BECOME	BEFORE	BEGIN	BLOCK
BODY	CACHE	CANCEL	CASCADE
CHANGE	CHARACTER	CHECKPOINT	CLOSE
COMMIT	COMPILE	CONSTRAINT	CONSTRAINTS
CONTENTS	CONTINUE	CONTROLFILE	CYCLE
DATABASE	DATAFILE	DBA	DEC
DECLARE	DISABLE	DISMOUNT	DOUBLE
DUMP	EACH	ENABLE	END
ESCAPE	EVENTS	EXCEPT	EXCEPTIONS
EXECUTE	EXPLAIN	EXTENT	EXTERNALLY
FLUSH	FORCE	FOREIGN	FREELIST
FREELISTS	FUNCTION	GROUPS	INCLUDING
INDICATOR	INTRANS	INSTANCE	INT
KEY	LAYER	LINK	LISTS
LOGFILE	MANAGE	MANUAL	MAX
MAXARCHLOGS	MAXDATAFILES	MAXINSTANCES	MAXLOGFILES
MAXLOGMEMBERS	MAXTRANS	MAXVALUE	MIN
MINEXTENTS	MINVALUE	MOUNT	NEW
NEXT	NOARCHIVELOG	NOCACHE	NOCYCLE
NOMAXVALUE	NOMINVALUE	NONE	NOORDER
NORESETLOGS	NORMAL	NOSORT	NUMERIC
OBJNO	OFF	OLD	ONLY
OPCODE	OPEN	OPTIMAL	OWN
PACKAGE	PARALLEL	PCTINCREASE	PCTUSED
PLAN	PRIMARY	PRIVATE	PROCEDURE
PROFILE	QUOTA	RBA	READ
REAL	RECOVER	REFERENCES	REFERENCING
RESETLOGS	RESTRICTED	REUSE	ROLE
ROLES	ROLLBACK	SAVEPOINT	SCHEMA
SCN	SEGMENT	SEQUENCE	SHARED
SNAPSHOT	SOME	SORT	STATEMENT_ID
STATISTICS	STOP	STORAGE	SWITCH
SYSTEM	TABLES	TABLESPACE	TEMPORARY
THREAD	TIME	TRACING	TRANSACTION
TRIGGERS	TRUNCATE	UNDER	UNLIMITED
UNTIL	USE	USING	WHEN
WORK	WRITE		

Appendix C – Oracle Reserved Words

ACCESS	ADD	ALL	ALTER
AND	ANY	ARRAYLEN	AS
ASC	AUDIT	BETWEEN	BY
CHAR	CHECK	CLUSTER	COLUMN
COMMENT	COMPRESS	CONNECT	CREATE
CURRENT	DATE	DECIMAL	DEFAULT
DELETE	DESC	DISTINCT	DROP
ELSE	EXCLUSIVE	EXISTS	FILE
FLOAT	FOR	FROM	GRANT
GROUP	HAVING	IDENTIFIED	IMMEDIATE
IN	INCREMENT	INDEX	INITIAL
INSERT	INTEGER	INTERSECT	INTO
IS	LEVEL	LIKE	LOCK
LONG	MAXEXTENTS	MINUS	MODE
MODIFY	NOAUDIT	NOCOMPRESS	NOT
NOTFOUND	NOWAIT	NULL	NUMBER
OF	OFFLINE	ON	ONLINE
OPTION	OR	ORDER	PCTFREE
PRIOR	PRIVILEGES	PUBLIC	RAW
RENAME	RESOURCE	REVOKE	ROW
ROWID	ROWLABEL	ROWNUM	ROWS
SELECT	SESSION	SET	SHARE
SIZE	SMALLINT	SQLBUF	START
SUCCESSFUL	SYNONYM	SYSDATE	TABLE
THEN	TO	TRIGGER	UID
UNION	UNIQUE	UPDATE	USER
VALIDATE	VALUES	VARCHAR	VARCHAR2
VIEW	WHENEVER	WHERE	WITH