Software Release Notes for Extract, Transform, and Load (ETL) Software and Data Warehouse Installation Package

United States Environmental Protection Agency

Office of Wetlands, Oceans and Watersheds

Contract No. 68-W-99-002
Task Order No. 014
Product Control No. SDC-0002–014-JW-6010
Deliverable No. 21-2.1, 21-3.2

February 13, 2004

SDC-0002-014-JW-6010
February 13, 2004

SOFTWARE RELEASE NOTES
FOR
EXTRACT, TRANSFORM, AND LOAD (ETL) SOFTWARE AND DATA WAREHOUSE INSTALLATION PACKAGE

CONTRACT NO. 68-W-99-002
TASK ORDER NO. 014

Prepared for:

United States Environmental Protection Agency
Office of Wetlands, Oceans and Watersheds
Assessment and Watershed Protection Division
401 M Street SW
Washington, DC 20460

Task Order Project Officer:

Robert King

Prepared by:

Systems Development Center
Science Applications International Corporation
6565 Arlington Boulevard
Falls Church, VA 22042

CONTENTS

1.0 INTRODUCTION 1
1.1 Scope 1
1.2 Identification 1

2.0 REFERENCES 1

3.0 SUMMARY OF FUNCTIONALITY 1
3.1 New Functionality 1
3.2 Improved Functionality 2

4.0 DEGREE OF FUNCTIONALITY BEING PROVIDED 2

5.0 INSTALLATION INSTRUCTIONS 2

1.0 INTRODUCTION

The STORET Central Warehouse operates as a retrieval mechanism for data submitted to the Environmental Protection Agency (EPA). The Extract, Transform, and Load (ETL) software populates the STORET Central Warehouse with data. The Data Warehouse installation package provides connectivity to the database, and executes the ETL software in the proper sequence.

1.1 Scope

These release notes provide detailed installation instruction and an overview of the ETL requirements provided in the software package. This release is an upgrade to the STORET Central Warehouse and completely replaces previous versions.

1.2 Identification

This version of the STORET Central Warehouse is identified and tracked under the configuration identifier C2.0.2, and is being provided to the Office of Wetlands, Oceans, and Watersheds (OWOW), Assessment and Watershed Protection Division (AWPD), Monitoring Branch, STORET Task Order Project Officer (TOPO).

2.0 REFERENCES

The STORET Project Plan dated September 17, 2003 was the basis for developing the STORET Central Warehouse upgrade. The requirements were initiated and documented through the Change Control Board (CCB) process.

3.0 SUMMARY OF FUNCTIONALITY

The requirements associated with this release have been integrated with prior releases of the STORET Central Warehouse. The software for creating tables, populating data (i.e., ETL), and creating and operating the Web interface is included in this deliverable. The requirements specific to this release are described below.

3.1 New Functionality

This release includes the following new functionality:

- Added Station Activity Point information to the Result fact tables.

- Included maximum and minimum Activity Start dates to the ETL software populating the STATION_CHAR table.

- Made the following columns and associated data available in the Station fact table: PRIMARY_TYPE, SECONDARY_TYPE, HORIZONTAL_DATUM, GEOPOSITIONING_METHOD, COUNTRY_CODE, STATE_POSTAL_CODE, ELEVATION_DATUM, ELEVATION_METHOD, PRIMARY_ESTUARY, and SECONDARY_ESTUARY. All of these columns will have names beginning with the prefix “FK”.

3.2 Improved Functionality

Moved Oracle connect statements to a single ETL script. This change simplifies the process of altering connect statements for specific user names, passwords, and connect strings.

4.0 DEGREE OF FUNCTIONALITY BEING PROVIDED

The software provided by this release fulfills the requirements agreed to with the TOPO. No unresolved discrepancies are known to exist upon conclusion of Unit, Integration, and System Acceptance testing.

5.0 INSTALLATION INSTRUCTIONS

This section contains the step-by-step installation instructions to create the STORET Central Warehouse, and install the associated Web Application. There are several noteworthy items regarding the installation:

- Edit this script for database usernames, passwords, and connect strings.

- Edit all ETL scripts for tablespace names.

- Scripts must be run in the order listed to ensure correct software installation.

- A rollback segment named RBSBIG must exist and be on-line.

- Oracle archiving should be turned off when running the ETL scripts.

- Log files are created for each of the Structured Query Language (SQL) scripts listed in these instructions. These log files have parallel names and directory locations as the

scripts that create them. Log files should be checked for errors after each step of the installation process to ensure the installation completes successfully.

- This release of the Central Warehouse requires Oracle 9i and JServer.

- Copy the storetw directory from the "Release C2.0.2\software" directory to your C:\ drive.

The installation package is not dependent upon a specific drive letter. If the installation package is placed on a different drive, replace "C:" with the letter of that drive in the following instructions.

1. Open a SQL*Plus session.

2. Create storetw schema with appropriate system and object privileges.
   CONNECT SYSTEM/STO2ET@STORET.SDC
   @C:\storetw\storetw\create_storetw.sql;
   CONNECT STORET/STO2ET@STORET.SDC
   @C:\storetw\storetw\storet_grants.sql;
   CONNECT STORET1/STO2ET1@STORET.SDC
   @C:\storetw\storetw\storet1_grants.sql;

3. Create the functions used by the ETL scripts.
   CONNECT STORETW/STORETW@STORET.SDC
   @C:\storetw\functions\f_createflbrmk.sql;
   @C:\storetw\functions\f_estry.sql;
   @C:\storetw\functions\f_mad.sql;
   @C:\storetw\functions\f_station_visited.sql;
   @C:\storetw\functions\f_fieldset.sql;
   @C:\storetw\functions\f_createFNAICS.sql;
   @C:\storetw\functions\f_biopt.sql;
   @C:\storetw\functions\f_prmvl.sql;
   @c:\storetw\functions\f_char_name.sql;

4. Create FA_STATION table.
   @C:\storetw\fa_statn\fa_station_table.sql;

5. Create Characteristic tables.
   @C:\storetw\char_stn\char_stn_table.sql;
   @C:\storetw\char_stn\char_table.sql;

6. Create Drainage Basin tables.
   @C:\storetw\drainagebasin\db_tables.sql;

7. Create Geographical tables (state/county).
   @C:\storetw\geopa\geo_tables.sql;

8. Create Method and Datum (MAD) code tables.
   @C:\storetw\mad\mad_tables.sql;

9. Create Organization table.
   @C:\storetw\org\org_tables.sql;

10. Create Station Types table.
    @C:\storetw\statn_types\statn_types_tables.sql;

11. Create Estuary tables.
    @C:\storetw\estuary\estry_tables.sql;

12. Create DI_DATE table.
    @C:\storetw\di_date\di_date_table.sql;

13. Create FA_REGULAR_RESULT table.
    @C:\storetw\fa_regular_result\fa_regular_result_table.sql;
    @C:\storetw\fa_regular_result\fa_regular_result_data1.sql;
    @C:\storetw\fa_regular_result\fa_regular_result_pdl_data1.sql;
    @C:\storetw\fa_regular_result\create_temp_tables.sql;
    @C:\storetw\fa_regular_result\fa_regular_result_data2.sql;

14. Create FA_BIOLOGICAL_RESULT table.
    @C:\storetw\fa_biological_result\fa_bio_result_table.sql;
    @C:\storetw\fa_biological_result\fa_bio_result_data1.sql;
    @C:\storetw\fa_biological_result\create_bio_temp_tables.sql;
    @C:\storetw\fa_biological_result\fa_bio_result_data2.sql;

15. Create FA_HABITAT_RESULT table.
    @C:\storetw\fa_habitat_result\fa_habitat_result_table.sql;
    @C:\storetw\fa_habitat_result\fa_habitat_result_data1.sql;
    @C:\storetw\fa_habitat_result\create_habitat_temp_tables.sql;
    @C:\storetw\fa_habitat_result\fa_habitat_result_data2.sql;

16. Create DI_ACT_MEDIUM table.
    @C:\storetw\di_act_medium\di_act_medium.sql;

17. Create result fact table indexes.
    @C:\storetw\fa_regular_result\fa_regular_result_index.sql;
    @C:\storetw\fa_biological_result\fa_bio_result_index.sql;
    @C:\storetw\fa_habitat_result\fa_habitat_result_index.sql;

18. Create LU_CHAR_ALIAS AND LU_CHAR_ALIAS_TYPE tables.
    @C:\storetw\lu_char_alias\lu_char_alias.sql;

19. Create DI_PROJECT table and the associated relationship tables.
    @C:\storetw\di_project\di_project_table.sql;
    @C:\storetw\di_project\project_rel_tables.sql;

20. Create the LU_STATION_ALIAS table.
    @C:\storetw\lu_station_alias\lu_station_alias.sql;

21. Create the remaining dimension tables (i.e., DI_ACTIVITY_INTENT, DI_COMMUNITY_SAMPLED, DI_SUBJECT_TAXON, DI_GROUP_TYPE, DI_BIOPART, and DI_ACTIVITY_MATRIX)
    @C:\storetw\di_activity_intent\di_activity_intent.sql;
    @C:\storetw\di_community_sampled\di_community_sampled.sql;
    @C:\storetw\di_subject_taxon\di_subject_taxon.sql;
    @C:\storetw\di_group_type\di_group_type.sql;
    @C:\storetw\di_biopart\di_biopart.sql;
    @C:\storetw\di_activity_matrix\di_activity_matrix.sql;

22. Create the Binary Large Object (BLOB) table and add BLOB IDs to the fact tables.
    @C:\storetw\fa_blob\fa_blob.sql;

23. Finish building the FA_REGULAR_RESULT table.
    @C:\storetw\fa_regular_result\fa_regular_result_const.sql;
    @C:\storetw\fa_regular_result\fa_regular_result_drop_cols.sql;

24. Finish building the FA_BIOLOGICAL_RESULT table.
    @C:\storetw\fa_biological_result\fa_bio_result_const.sql;
    @C:\storetw\fa_biological_result\fa_bio_result_drop_cols.sql;

25. Finish building the FA_HABITAT_RESULT table.
    @C:\storetw\fa_habitat_result\fa_habitat_result_const.sql;
    @C:\storetw\fa_habitat_result\fa_habitat_result_drop_cols.sql;

26. Drop unnecessary columns from FA_STATION table.
    @C:\storetw\fa_statn\fa_station_drop_cols.sql;

27. Analyze database objects and performance tune.
    @C:\storetw\analyze_objects.sql;
    @C:\storetw\performance_tuning.sql;

28. Delete temporary tables.
    @C:\storetw\drop_temp_tables.sql;

The ETL process is now complete. Perform the remaining steps for installing the web interface.

29. Create report customization related table and materialized views.
    CONNECT STORETW/STORETW@STORET.SDC
    @C:\storetw\application\APP_COLUMN_NAME.sql;
    @C:\storetw\materialized_views.sql;

30. Alter the file C:\storetw\application\DW_glob_var.sql to reflect the directory structure of the environment where the software is being installed. The global variables established in this package should have values identical to the values in the glob_var.sql file included with the existing STORET Web Application with one exception; the file name section of the variable lv_script_path should be set to “DW_storet.js”.

31. Create the global variables package as storetw.
    @C:\storetw\application\DW_glob_var.sql;

32. Compile Java procedures used by the web application procedures.
    @C:\storetw\application\procedures\DW_exec_os.java;
    @C:\storetw\application\procedures\DW_blob_handler.java;

33. Create the web application procedures as storetw.
    @C:\storetw\application\procedures\DW_huc_popup.sql;
    @C:\storetw\application\procedures\DW_counties_popup.sql;
    @C:\storetw\application\procedures\DW_top_of_page.sql;
    @C:\storetw\application\procedures\DW_bottom_of_page.sql;
    @C:\storetw\application\procedures\DW_display_calendar.sql;
    @C:\storetw\application\procedures\DW_geo_select.sql;
    @C:\storetw\application\procedures\DW_date_select.sql;
    @C:\storetw\application\procedures\DW_char_select.sql;
    @C:\storetw\application\procedures\DW_project_select.sql;
    @C:\storetw\application\procedures\DW_station_select.sql;
    @C:\storetw\application\procedures\DW_medium_select.sql;
    @C:\storetw\application\procedures\DW_community_select.sql;
    @C:\storetw\application\procedures\DW_selection_criteria.sql;

@C:\storetw\application\procedures\DW_station_count.sql;
@C:\storetw\application\procedures\DW_station_download_custom.sql;
@C:\storetw\application\procedures\DW_station_hub_custom.sql;
@C:\storetw\application\procedures\DW_proj_popup.sql;
@C:\storetw\application\procedures\DW_char_alias_popup.sql;
@C:\storetw\application\procedures\DW_station_popup.sql;
@C:\storetw\application\procedures\DW_extref_popup.sql;
@C:\storetw\application\procedures\DW_result_criteria_geo.sql;
@C:\storetw\application\procedures\DW_result_criteria_project.sql;
@C:\storetw\application\procedures\DW_result_criteria_station.sql;
@C:\storetw\application\procedures\DW_bio_result_criteria_geo.sql;
@C:\storetw\application\procedures\DW_bio_result_criteria_project.sql;
@C:\storetw\application\procedures\DW_bio_result_criteria_station.sql;
@C:\storetw\application\procedures\DW_hab_result_criteria_geo.sql;
@C:\storetw\application\procedures\DW_hab_result_criteria_project.sql;
@C:\storetw\application\procedures\DW_hab_result_criteria_station.sql;
@C:\storetw\application\procedures\DW_result_count.sql;
@C:\storetw\application\procedures\DW_result_download_custom.sql;
@C:\storetw\application\procedures\DW_result_hub_custom.sql;
@C:\storetw\application\procedures\DW_home.sql;
@C:\storetw\application\procedures\DW_blob_download.sql;
@C:\storetw\application\procedures\DW_station_download_zip.sql;
@C:\storetw\application\procedures\DW_result_download_zip.sql;

34. Copy the application's JavaScript library file into the scripts directory indicated in the global variables script (see step 30).
C:\storetw\application\scripts\DW_storet.js

35. Copy the application's Help and the associated image files into the Hyper-Text Markup Language (HTML) documents directory indicated in the global variables script (see item 30).
C:\storetw\application\doc\DW_storet_help.html
C:\storetw\application\doc\lat_long_ex.gif

36. Alter the file C:\storetw\application\procedures\DW_PERMISSIONS.sql to reflect the environment where the software is being installed. This script includes Java permissions that require a username/password with Database Administrator (DBA) privileges. The script must also be modified to include the correct directory location of the GZIP file compression utility and the location of the user reports directory (established in DW_glob_var).

37. Grant appropriate privileges and create required synonyms.
@C:\storetw\application\DW_GRANTS.sql;
CONNECT STORETWEB/STORETWEB@STORET.SDC
@C:\storetw\application\DW_SYNONYMS.sql;
CONNECT SYSTEM/STO2ET@STORET.SDC
@C:\storetw\application\DW_PERMISSIONS.sql;
DISCONNECT