

CANARY Webinar: Connecting CANARY to Data Sources

CANARY Webinar #2

David Hart, Sandia National Laboratories

August 12th, 2009

CANARY was developed through an Inter-Agency Agreement between the U. S. Environmental Protection Agency and Sandia National Laboratories. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.



Webinar Outline

- Understand the different CANARY run modes
- Describe how to connect CANARY to on-line and off-line data streams, including:
 - On-line and off-line generic databases
 - EDDIES specific configuration
- Questions and Answers

CANARY RUN MODES

On-line Run Mode Comparison

“EDDIES” mode

- Requires an “EDDIES” control mode
- Requires the use of an EDDIES database
- Parameters (SCADA signals and algorithms) must be defined both in EDDIES and in CANARY
- EDDIES handles I/O between EDS tool and the SCADA/HMI system

“RealTime” mode

- Requires “Internal” control mode
- Can use any database
- Connection to SCADA system must be set up manually
- Interpretation of CANARY results by SCADA system must be done by the SCADA/HMI contractor or DBA

CANARY Configuration File

Input Tables

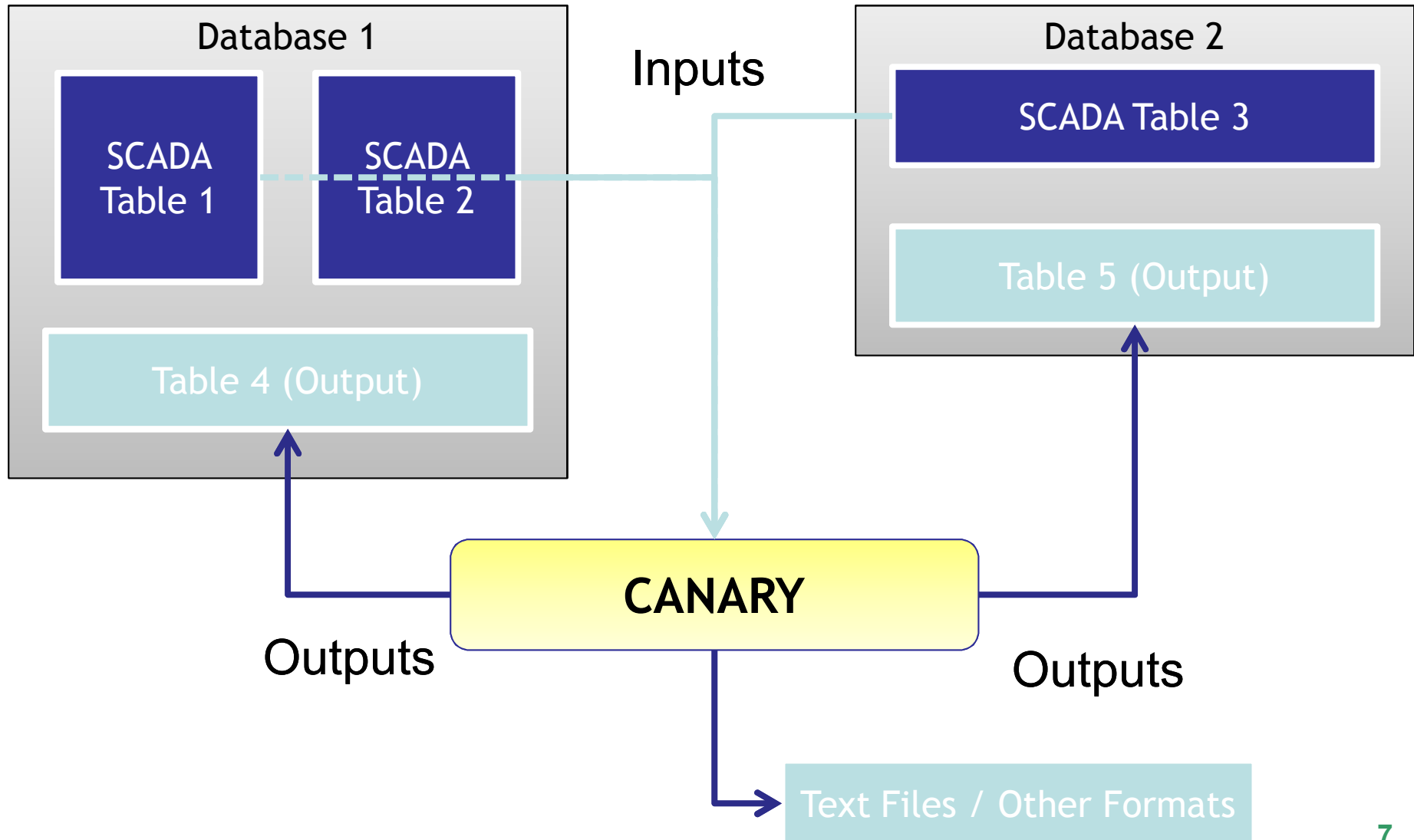
Output Tables

GENERIC DATABASE CONFIGURATION

Generic Databases

- Database connections are accomplished using Java connectors
 - All the major database vendors provide free Java connection drivers for their software
 - Installing the driver is usually as simple as unzipping a file into a directory, and including the location of the driver in CANARY's configuration file
- The database will need to have a username and password that CANARY can use
 - Adding a separate CANARY user allows the SCADA database administrator (DBA) to set specific permissions for CANARY, such as permission to only read the SCADA tables but write to a CANARY table

Generic Database/CANARY Interaction



Generic Database Configuration

- To configure the driver, use the “Timing and Control Definitions” page in the CANARY configuration editor

The screenshot shows the 'CANARY Timing and Control Definitions' dialog box. It has two main sections: 'Control Settings' and 'Timing Settings'. In the 'Control Settings' section, 'Run Mode' is set to 'Batch' and 'Control Type' is set to 'Internal'. A blue box highlights these two dropdown menus, with an arrow pointing from the text 'Make sure the run-mode is "Batch" or "RealTime" with a control type of "Internal"'. The 'Timing Settings' section includes fields for 'Data Interval' (00:06:00), 'Dynamic Start and Stop Dates?' (unchecked), 'Start Date' (2009-08-05 13:54:00), 'Stop Date' (2009-09-14 04:48:00), 'Date/Time Format' (yyyy-mm-dd HH:MM:SS), and 'Poll Interval' (00:00:10). On the right, the 'Driver ".jar" Files' section has 'Add Jar File' and 'Remove Selected' buttons. An arrow points from the text 'Add the .jar file that came with the JDBC driver download for your database' to the 'Add Jar File' button. Below this is a list box containing the path 'C:\Program Files\CANARY\jdbc14.jar'. At the bottom are 'Save and Close' and 'Cancel' buttons.

Make sure the run-mode is “**Batch**” or “**RealTime**” with a control type of “**Internal**”

Add the .jar file that came with the JDBC driver download for your database

Generic Database Configuration

- The following is a graphic of the CANARY configuration editor screen setting up CANARY to use a generic database

The image displays two overlapping windows from the CANARY configuration editor. The primary window, titled "CANARY Input and Output Definitions", is in the "Data Source Definition" tab. It contains several sections: "Standard Options" with fields for "Data Source Name" (input_db), "Data Source Type" (Database (DB)), and "Location" (jdbc:oracle:thin:@localhost:1521/xs), along with a "Set Location" button and a checked "Data Source is Enabled" checkbox; "Database Login Information" with an unchecked "Ask at connect" checkbox, "Username" (EDS_DATA_ARC), and "Password" (CANARY); and "Additional Details" with fields for "Timestep Field (CSV,DB)" (Timestep), "Connection Class" (oracle.jdbc.pool.OracleDataSource), "Input Table" (DH_PETRO), "Output Table" (empty), "String-to-Date Function" (To_Date), and "String-to-Date Format" (YYYY-MM-DD HH24:MI:SS). Action buttons on the right include "Clear/Create New", "Load Datasource", "Save Datasource", "Save and Close", and "Close". A secondary window titled "Where is the d..." is open on top of the first, showing fields for "URL Prefix" (jdbc:oracle:thin:@), "Hostname/IP" (localhost), "Port" (1521), and "Instance" (xs), with "OK" and "Cancel" buttons at the bottom.

CANARY Input and Output Definitions

Data Source Definition Clear/Create New

Standard Options

Data Source Name: input_db Set Location

Data Source Type: Database (DB) ☒ Data Source is Enabled

Location: jdbc:oracle:thin:@localhost:1521/xs

Database Login Information

☐ Ask at connect Username: EDS_DATA_ARC Password: CANARY

Additional Details

Timestep Field (CSV,DB): Timestep

Connection Class: oracle.jdbc.pool.OracleDataSource

Input Table: DH_PETRO

Output Table:

String-to-Date Function: To_Date

String-to-Date Format: YYYY-MM-DD HH24:MI:SS

Load Datasource
Save Datasource
Save and Close
Close

Where is the d...

URL Prefix: jdbc:oracle:thin:@

Hostname/IP: localhost

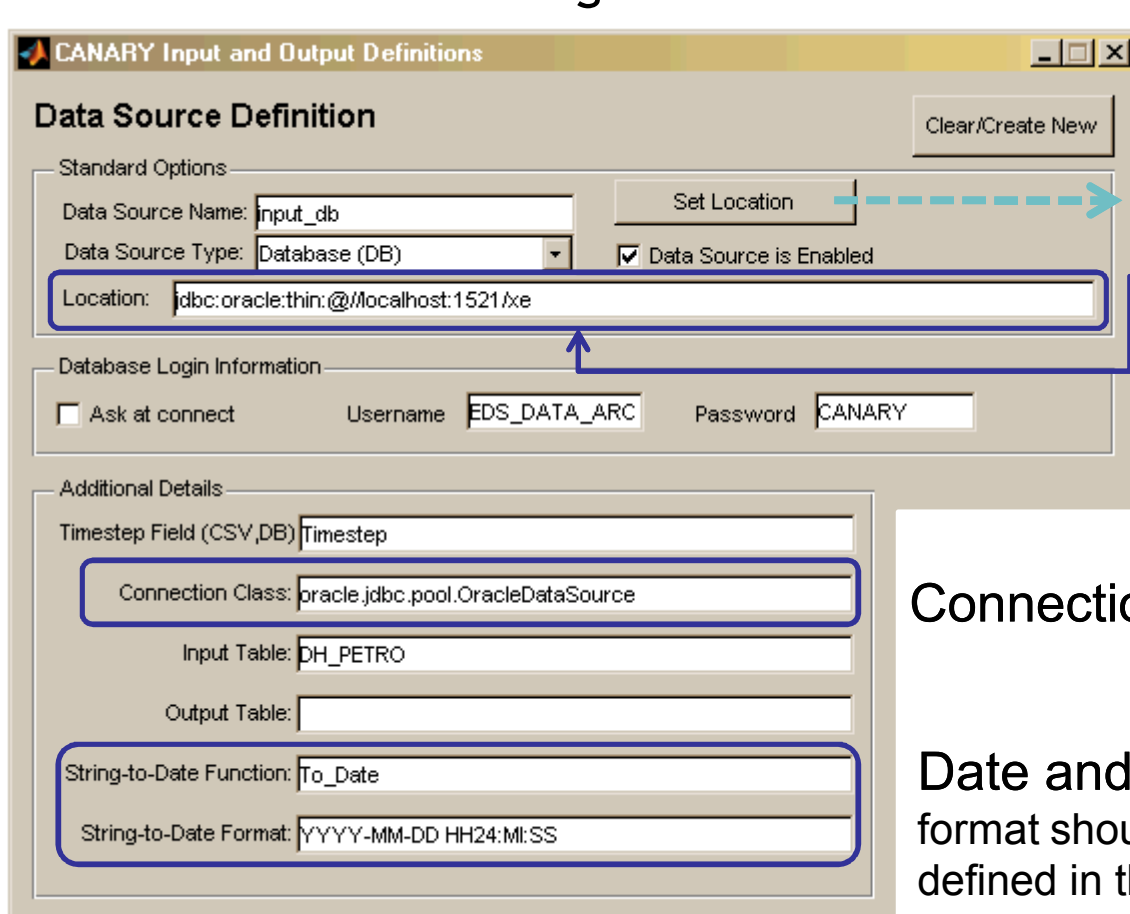
Port: 1521

Instance: xs

OK Cancel

Generic Database Configuration

- The JDBC driver downloaded from the database vendor will specify how to set the following items:



The screenshot shows the 'CANARY Input and Output Definitions' dialog box. The 'Data Source Definition' section includes fields for 'Data Source Name' (input_db), 'Data Source Type' (Database (DB)), and 'Location' (jdbc:oracle:thin:@localhost:1521/xe). A 'Set Location' button is highlighted with a dashed arrow pointing to a secondary dialog box titled 'Where is the d...'. This secondary dialog box contains fields for 'URL Prefix' (jdbc:oracle:thin:@), 'Hostname/IP' (localhost), 'Port' (1521), and 'Instance' (xe). The 'Database Login Information' section includes a checkbox for 'Ask at connect', 'Username' (EDS_DATA_ARC), and 'Password' (CANARY). The 'Additional Details' section includes 'Timestep Field (CSV,DB)' (Timestep), 'Connection Class' (oracle.jdbc.pool.OracleDataSource), 'Input Table' (DH_PETRO), 'Output Table' (empty), 'String-to-Date Function' (To_Date), and 'String-to-Date Format' (YYYY-MM-DD HH24:MI:SS).

URL

Connection Class

Date and time function format should match up with the one defined in the control section

Generic Database Configuration

- Your SCADA database administrator will specify the following:

CANARY Input and Output Definitions

Data Source Definition Clear/Create New

Standard Options

Data Source Name: Set Location

Data Source Type: ☒ Data Source is Enabled

Location:

Database Login Information

☐ Ask at connect Username Password

Additional Details

Timestep Field (CSV,DB)

Connection Class:

Input Table:

Output Table:

String-to-Date Function:

String-to-Date Format:

User ID and password

Field with time stamp info

Table names to read from or write to

Generic Database Inputs

- CANARY reads from tables that are set up in the following manner, where the SCADA tags are mapped to field names

ID	TIME_STEP	C_TURB_VAL	C_CL2_VAL	C_PH_VAL	C_COND_VAL	C_TOC_A_VAL	C_TOC_B_VAL	C_TEMP_VAL	C_PRES_OP
1	03/13/2008 00:00:00	.406	.58	8.164	358.505	.647	888	23.012	16.323
2	03/13/2008 00:02:00	.409	.58	8.202	359.26801	.623	880	23.158	16.323
3	03/13/2008 00:04:00	.407	.58	8.17	359.26801	.621	880	23.121	16.323
4	03/13/2008 00:06:00	.408	.58	8.152	359.077	.609	885	22.46	16.296
5	03/13/2008 00:08:00	.409	.58	8.144	359.26801	.611	885	23.212	16.285
6	03/13/2008 00:10:00	.408	.58	8.168	358.505	.603	880	23.178	16.285
7	03/13/2008 00:12:00	.407	.58	8.197	359.26801	.608	880	23.238	16.285
8	03/13/2008 00:14:00	.406	.58	8.192	359.26801	.589	880	23.149	16.259
9	03/13/2008 00:16:00	.409	.58	8.147	358.505	.599	880	22.549	16.248
10	03/13/2008 00:18:00	.41	.58	8.153	358.505	.6	896	23.255	16.247
11	03/13/2008 00:20:00	.409	.58	8.157	359.26801	.602	896	23.287	16.247
12	03/13/2008 00:22:00	.406	.58	8.198	359.26801	.596	883	23.198	16.244

Generic Database Outputs

- Below is a screenshot showing the structure of the CANARY generic output table. If a table of the appropriate name does not exist, CANARY will try to create one.
 - TIME_STEP and LOCATION_ID form the primary index key
 - DETECTION_INDICATOR is the event code from CANARY
 - DETECTION_PROBABILITY is the [0..1] probability of an event
 - ANALYSIS_COMMENTS contains error text or warnings
 - CONTRIBUTING_PARAMETERS contains a list of outlying signals

```
1 SELECT * FROM _canary_output c;
```

TIME_STEP	LOCATION_ID	DETECTION_INDICATOR	DETECTION_PROBABILITY	ANALYSIS_COMMENTS	CONTRIBUTING_PARAMETERS
2009-08-06 10:52:00	TANK	2	0	Starting up by filling window	NULL
2009-08-06 10:54:00	TANK	2	0	Insufficient history, unable to predict	NULL

CANARY Configuration File

Input Table

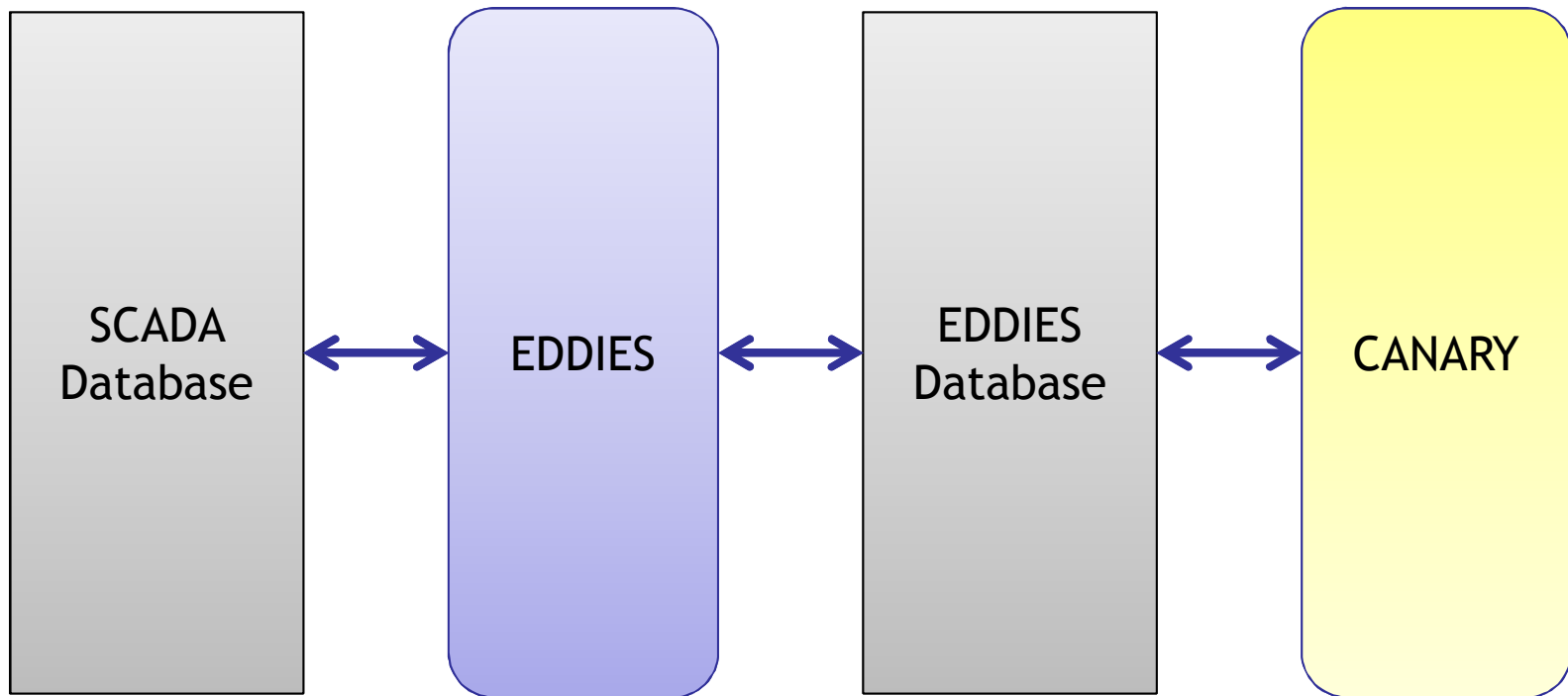
Output Table

Extending the Database

EDDIES DATABASE CONFIGURATION

EDDIES/CANARY Interaction

- The EPA's EDDIES software uses a separate database to insulate CANARY from the SCADA database.



EDDIES/CANARY Configuration

- To configure CANARY to use EDDIES, there are a few changes that have to be made in the “Input & Output” configuration page

The data source type must be “EDDIES”

The time step field is “TIME_STEP”

The tables are named “ANALYSIS_SAMPLES” and “ANALYSIS_RESULTS”

The screenshot shows the 'CANARY Input and Output Definitions' window. The 'Data Source Definition' section has 'Data Source Name' set to 'eddies' and 'Data Source Type' set to 'EDDIES (EDDIES)'. The 'Database Login Information' section has 'Ask at connect' unchecked, 'Username' set to 'CANARY', and 'Password' set to 'CANARY'. The 'Additional Details' section has 'Timestep Field (CSV,DB)' set to 'TIME_STEP', 'Input Table' set to 'analysis_samples', and 'Output Table' set to 'analysis_results'. The 'String-to-Date Function' is set to 'To_Date' and the 'String-to-Date Format' is set to 'MM/DD/YYYY HH:MI AM'. Buttons for 'Clear/Create New', 'Set Location', 'Load Datasource', 'Save Datasource', 'Save and Close', and 'Close' are visible.

CANARY Input and Output Definitions

Data Source Definition Clear/Create New

Standard Options

Data Source Name: eddies Set Location

Data Source Type: EDDIES (EDDIES) ☒ Data Source is Enabled

Location: jdbc:oracle:thin:@localhost:1521/xe

Database Login Information

☐ Ask at connect Username: CANARY Password: CANARY

Additional Details

Timestep Field (CSV,DB): TIME_STEP Load Datasource

Connection Class: oracle.jdbc.pool.OracleDataSource Save Datasource

Input Table: analysis_samples Save and Close

Output Table: analysis_results

String-to-Date Function: To_Date Close

String-to-Date Format: MM/DD/YYYY HH:MI AM

EDDIES/CANARY Configuration

- In CANARY's "Timing & Control Settings" page, the following must be correct:

Make sure that the run-mode is "EDDIES" and that the control type is also "EDDIES" and linked to the data source you created earlier

Don't forget to add the .jar file

The screenshot shows the "CANARY Timing and Control Definitions" window. The "Timing & Control Settings" tab is active. The "Control Settings" section has a blue box around it, containing "Run Mode: EDDIES", "Control Type: EDDIES", and "Linked to Datasource: eddies". The "Timing Settings" section has "Data Interval: 00:02:00", "Dynamic Start and Stop Dates?" checked, "Start Date: 08/05/2009 12:00 AM", "Stop Date: 08/08/2009 12:00 AM", "Date/Time Format: mm/dd/yyyy HH:MM PM", and "Poll Interval: 00:00:02". The "Driver 'jar' Files" section has a blue box around it, containing "Add Jar File", "Remove Selected", and a list box with "C:\Program Files\CANARY\ojdbc14.jar". At the bottom are "Save and Close" and "Cancel" buttons.

EDDIES Inputs

- EDDIES uses an input format similar to the generic output table for input variables.
 - Each row in the table represents a single value for a single SCADA tag
 - EDDIES defines that a specific table (ANALYSIS_SAMPLES) be used

TIME_STEP	PARAMETER_ID	SAMPLE_VALUE	SAMPLE_QUALITY	EVENT_STATUS	RECEIPT_TIME
31-MAR-08	DST_CWS_ _H2OxYSI_COND_V	293	Normal	0	-
31-MAR-08	DST_CWS_ _H2OxYSI_ORPx_V	693	Normal	0	-
31-MAR-08	DST_CWS_ _H2OxYSI_TEMP_V	10.2	Normal	0	-
31-MAR-08	DST_CWS_ _H2OxYSI_TURB_V	0.08	Normal	0	-
31-MAR-08	DST_CWS_ _MONxUPSxON_ALM	1	Normal	0	-
31-MAR-08	DST_CWS_ _MONxSCAN_TOCx_ALM	0	Normal	0	-
31-MAR-08	DST_CWS_ _MONxSIEV_TOCx_ALM	1	Normal	0	-
31-MAR-08	DST_CWS_ _MONxUSF_CL2x_ALM	0	Normal	0	-

EDDIES Inputs

- If CANARY needs to connect to a database that is set up like the EDDIES input tables, it is possible to configure an “EDDIES” type data source for a non-EDDIES system
 - This allows CANARY to read an EDDIES style table in a generic “RealTime” mode
 - The table name *can* be changed
 - The field names *cannot* be changed, so the format must be exactly like the EDDIES table
- **However,**
 - Your DBA may be able to set up a view that will convert an existing table into something with the EDDIES field names without having to change the SCADA database or duplicate data

EDDIES Outputs

- The EDDIES output table is roughly the same as the generic database output table. However, the EDDIES database schema defines two stored procedures to add data to the tables.
 - CANARY must use the stored procedures to output to an “EDDIES” formatted data source
 - EDDIES stores the results in the “ANALYSIS_RESULTS” table
 - EDDIES stores the “Contributing Parameters” in a separate table, called “PARAMETER_TYPE_RESULTS”
 - All the tables should be created during EDDIES setup

The CANARY FAQ is available at
<https://software.sandia.gov/trac/canary>
Answers to today's questions will be added soon

Q & A