 Talend User Components:

tJobInstance\* - Collection

http://www.cimt-ag.de

# Purpose

This collection consists of:

|  |  |
| --- | --- |
| Component | Purpose |
| tJobInstanceStart | Register a job run and provide information about the previous job run.  Setup the logging facility. |
| tJobInstanceEnd | Updates the job run data with metrics, job results and error information. |
| tJobDataRangeScanner | Collects min/max values of timestamps or numeric values within a data flow.  These min/max values for timestamps and any other values are set for the job-time-range and job-value-range. |
| tJobInstanceLiveCheck | Checks the entries of the job run registry for dead or broken job instances and cleans up the job registry. |

These components help to track the execution of jobs in a database table.

Advantages of these components:

* Provides a unique numeric id for the job to mark all data sets processed by the job
* Start-/Stop timestamps
* Return code and error messages (collects all messages)
* Host and PID of the process running this job
* Supports incremental loading
* Supports restart capabilities by providing information about the previous job run
* Key figures about moved data sets
* Snapshot of the context at the start and at the end of the job
* Detects the minimum and maximum of value for a flow
* Tracks the memory usage and detects peaks
* Detects parallel job runs (also based on the work item) and provide steering possibilities

# Talend-Integration

This component can be found in the palette under Management

This component provides several return values.

# Component tJobInstanceStart

## Basic Settings

|  |  |
| --- | --- |
| Property | Content |
| Use data source (connection pool) | If true, the component takes the database connection from a database connection pool. |
| Data source alias | The name of the database connection pool providing the database connection. |
| Database Connection | Any database connection pointing to the schema with the control tables.  The main table is JOB\_INSTANCES holding all key information. |
| Job Name | Name of the job. The default is using the build-in variable jobName |
| Task name (Job Display Name) | To distinguish job executions of the same job for different purposes, you can provide an additional declaration of the current task (mostly relevant for more generic jobs, that are called for different tasks). Best practice is to extend the job name with "-" or "#" and add a name extension to declare this way the task name. This information will be stored in the field job\_display\_name. |
| Process Instance Name | Since multiple jobs / job instances might be executed within the same process (via tRunJob) this property can be set to a more general job name, representing the whole process. |
| Job Work Item | Text describing the work item (e.g. a file name or a source object name) processed by this job instance. It can be used by later executions to determine already processed work items or to track individual increment boundaries for multiple source instances. |
| Take empty as null | If the value for work item is an empty string, it will be stored as null |
| Time range start | If the job has to process data selected by a time range, declare the start of the timerange here. |
| Time range end | If the job has to process data selected by a time range, declare the end of the timerange here. |
| Value range start | If the job has to process a partition of data selected by an id range or any other value ranges, you can store the start of the requested range here. |
| Value range end | If the job has to process a partition of data selected by an id range or any other value ranges, you can store the end of the requested range here. |
| Write Job instance ID to | To inject the job instance id into the processd data, typically a context variable will be used.  Provide the context variable to be loaded with the job insance id here. |
| Singleton Behavior | See the section Singleton Behavior later in this document. |
| Return last instance result | Fetches the information about the last run of this job. All information available as return values of the tJobInstanceStart component. |
| Last successful | The last run must have been a successful run of this job (all others will be ignored) |
| Last run must have inserted or deleted data | The last run must have data inserted or deleted. This will be detected via the couter columns. See the properties of tJobInstanceEnd. |
| Last run must have read data | The last run must have read data to be relevant as choice for the last run. This will be detected via the couter columns. See the properties of tJobInstanceEnd. |
| For the current work item | If true: only jobs with the same work item will be used to get the last instance results. This allows you to have one job for multiple purposes with the full support of the “Return last instance result” feature. |
| Collecting job instances ids running after previous run | Returns as comma separated list all instance ids of all job, which was running after the last run of this job. This helps to implement incremental jobs. It is necessary to write the job instance id into every data set proceed by the job. |
| Source job names | Filter the jobs which should part of the list above. This helps to keep the list small in case of having a lot of unrelated jobs in the system. |
| OK Result Codes | This is a String containing a comma-separated list of all return codes, which are related to a successful run. If you want using different return codes for OK please take care the tRunJob components does not die. |
| Replacement for prev… | These values can be used to simplify the job design because you can avoid the check if the job has a previous run. Please refer to the return values. |
| …-job-instance-id | Set here a default long value for the return value PREV\_JOB\_INSTANCE\_ID |
| …-job-start-date | Set here the Date typed default value for a previous job start date |
| …-time-range-end | Set here the Date typed default value for a previous time range end. |
| …-value-range-end | Set here the String typed default value for a previous value range end. |
| …-result-item | Set here the String typed default value for a previous job result item. |
| Set UTC as default time zone | This changes the default setting of the virtual machine for time zone from the local time zone to UTC. It affects the current JVM instance (=also jobs called by tRunJob that are not an independent child jobs). |
| Memory Usage Monitoring | This starts a thread, which collects every second, the used memory and detect the maximum and when it happened. Values are provided in global map after tJobInstanceEnd has finished. |
| Print job instance id to the console | If true the “Expreesion to pring” will be printed to System.out |
| Expression to print | Define here how to print the job instance id to system.out. You can add any available data as you like. It is recommended to print at least the jobname and the job instance id |

## Advanced Settings

|  |  |
| --- | --- |
| Max message length |  |
| Generate Job Instance Id | (experimental) This generates the Job Instance Id from systemtime and a large random number. The generated number should be unique even, when multiple instances are started in the same second. |
| Job instance ID is auto increment | This have to be switched on if the table use an auto increment for the job\_instance\_id column e.g. this is supposed for MySQL. |
| Sequence expression | The valid SQL Expression to get an unused job\_instance\_id from the database. In general this is a “nextval” of the DB sequence. Syntax depends on the database.  Examples:  Oracle: job\_instance\_id\_seq.nextval  PostgreSQL: nextval('dwh\_manage.job\_instance\_id\_seq')  DB2: NEXTVAL FOR dwh\_manage.job\_instance\_id\_seq |
| Reduce resume logs |  |
| Enable JMX | (Experimental JMX Connector to retrieve current values via JMX) |
| tStatCatcher Statistik | Enable trace output to talend statcatcher |

## Return values of tJobInstanceStart

|  |  |
| --- | --- |
| Return value | Content |
| ERROR\_MESSAGE | Last error message. Unfortunately this is not the error message from the actually running job. This message is build from the tRunTask component. The current TAC web service does not provide this message. |
| JOB\_INSTANCE\_ID | The job instance id used for this job run. |
| SOURCE\_JOB\_INSTANCE\_ID\_LIST | List of all job instance ids which are executed after the last run if this job. This way it is possible to implement incremental steering.  The list can easily be used in SQL e.g. :  ...where job\_instance\_id in (“ + ((String)globalMap.get(“tJobInstanceStart\_1\_SOURCE\_JOB\_INSTANCE\_ID\_LIST”) + “) …..” |
| JOB\_START\_DATE | The start date of the current job run. |
| PREV\_JOB\_EXISTS | If true means the job was running in the past at least one time. |
| PREV\_JOB\_START\_DATE | If a previous job run exists (otherwise null):  Contains the start date of the previous job |
| PREV\_JOB\_STOP\_DATE | If a previous job run exists (otherwise null):  Contains the stop date of the previous job |
| PREV\_JOB\_INSTANCE\_ID | If a previous job run exists (otherwise null):  Contains the ID of the previous job |
| PREV\_JOB\_TALEND\_PID | If a previous job run exists (otherwise null):  Contains the Talend-PID of the previous job |
| PREV\_JOB\_HOST\_PID | If a previous job run exists (otherwise null):  Contains the Host-PID (means the process ID of the operating system for this JVM) of the previous job |
| PREV\_JOB\_HOST\_NAME | If a previous job run exists (otherwise null):  Contains the name of the host where the previous job was running |
| PREV\_TIME\_RANGE\_START | If a previous job run exists (otherwise null):  Contains the time range start of the previous job |
| PREV\_TIME\_RANGE\_END | If a previous job run exists (otherwise null):  Contains the time range end of the previous job |
| PREV\_VALUE\_RANGE\_START | If a previous job run exists (otherwise null):  Contains the value range start of the previous job |
| PREV\_VALUE\_RANGE\_END | If a previous job run exists (otherwise null):  Contains the value range end of the previous job |
| PREV\_JOB\_RETURN\_CODE | If a previous job run exists (otherwise null):  Contains the return code of the previous job |
| PREV\_WORK\_ITEM | If a previous job run exists (otherwise null):  Contains the previous work item of the previous job |
| PREV\_RESULT\_ITEM | If a previous job run exists (otherwise null):  Contains the result item of the previous job |
| PREV\_COUNT\_INPUT | If a previous job run exists (otherwise null):  Contains the count inserts of the previous job |
| PREV\_COUNT\_OUTPUT | If a previous job run exists (otherwise null):  Contains the count outputs of the previous job |
| PREV\_COUNT\_UPDATED | If a previous job run exists (otherwise null):  Contains the count updates of the previous job |
| PREV\_COUNT\_DELETED | If a previous job run exists (otherwise null):  Contains the count deletes of the previous job |
| PREV\_COUNT\_REJECTS | If a previous job run exists (otherwise null):  Contains the count rejects of the previous job |

# Singleton Behavior

This component has the ability based of the information of the JOB\_INSTANCE\_STATUS table to detect already running instances of the same job (optional with the same work item).

The necessary basic settings to use this feature are:

|  |  |
| --- | --- |
| Property | Content |
| Check if another job instance of the same job is already running | If true, the component checks the JOB\_INSTANCE\_STATUS table for a already running job instance |
| Singleton for the work item | Use the work item to identify a job instance. |
| Prevent creating a job instance status entry if another job instance is already running | The option helps to avoid useless entries in the JOB\_INSTANCE\_STATUS table. |

Dedicated return values

|  |  |
| --- | --- |
| Return value | Content |
| JOB\_RUNS\_ALONE | The result of the singleton-check. This variable is not set if the check has not happened, this should help to prevent using this value if the actual check was not performed. |
| ALREADY\_RUNNING\_JOB\_START\_DATE | Start date (+time) of the already running job |
| ALREADY\_RUNNING\_JOB\_INSTANCE\_ID | The job instance id of the other job still running |
| ALREADY\_RUNNING\_JOB\_HOST\_NAME | On which server the other job instance is still running |
| ALREADY\_RUNNING\_JOB\_HOST\_PID | The native process id of the other job instance |
| ALREADY\_RUNNING\_JOB\_WORK\_ITEM | What work item the other job is still processing |

# Component tJobInstanceEnd

## Basic Settings

|  |  |
| --- | --- |
| Property | Content |
| Use separate connection | If true: the end component uses a separate database connection. This could help for jobs running very long to avoid problems with long standing connections. |
| Connection | Choose here the separate connection if you have chosen the option above. It must be a different connection component as used for the start component. |
| Job Instance Start Component | Choose here the tJobInstanceStart component. Both components depend on each other. |
| Job Result | A string representation of the result of the current job. In case the job creates a file, it is a good idea to put here the file path. |
| Time range start | If the job has processed data of a defined time interval, you can store the start of the interval here. |
| Time range end | If the job has processed data of a defined time interval, you can store the end of the interval here. This can be used as a starting point for the next iteration. |
| Value range start | If the job has process a partition of data selected by an id range or any other value ranges, you can store the start of the retrieved/requested range here |
| Value range end | See Value range start. This is the end of the retrieved/requested range. |
| Save named counters | Counters can be named; in this case the counter value will additionally be inserted in the table JOB\_INSTANCE\_COUNTERS |
| Delete previous successful job instances by work item | If checked, the component deletes all successful previous job instances with the same work item. This helps in case of the table JOB\_INSTANCE\_STATUS will be used to keep track of the current data in the DWH and repeated job runs with the same work item replaces previous data. |
| Close Connection | Closes the connection used for managing the job registration |
| Input Counters | Counters describing the result of the job can be added here. The sum of all counters will be written in the JOB\_INSTANCE\_STATUS table in COUNT\_INPUTS. By switching off the flag “Add” you can subtract a value instead of adding it. The name column provides the name to be used in the JOB\_INSTANCE\_COUNTERS table (see Save named counters option) |
| Output Counters | See Input Counters. Will be used for column COUNT\_OUTPUTS |
| Update Counters | See Input Counters. Will be used for column COUNT\_UPDATED |
| Reject Counters | See Input Counters. Will be used for column COUNT\_REJECTED |
| Delete Counters | See Input Counters. Will be used for column COUNT\_DELETED |

As Counter typically the NB\_LINE return values of the input or output components can be used. In case of the job has more the one output it is recommended to set names for particular counters to keep the distinct counter values.

## Return values of tJobInstanceEnd

|  |  |
| --- | --- |
| Return value | Content |
| ERROR\_MESSAGE | Last error message. |
| RETURN\_CODE | The retrieved return code of the current job |
| RETURN\_MESSAGE | The created return messages. This message contains all error messages from all components throwing an error. |
| MEMORY\_AVAILABLE | This is the memory (in byte) what is maximum available for the job.  Typically it is set with the JVM parameter -Xmx1024m (e.g. for 1GB RAM) |
| MEMORY\_MAX\_USED | The maximum of the used memory (in byte) what was allocated in the JVM.  Please keep in mind, if you call other jobs with tRunJob (not independently) they must be taken into account because they use the same JVM instance. |
| MEMORY\_MAX\_USED\_PERCENTAGE | The percentage between the available memory and the maximum used memory as value between 0 and 100. |

# Component tJobDataRangeScanner

This component measures the min/max values from any values (like IDs) or timestamps (like last\_modified).

These min/max values will be used to set the column values in the table job\_instance\_status (value\_range\_start, value\_range\_end, time\_range\_start, time\_range\_end).

All these values can be retrieved for the last run as return values of the tJobInstanceStart component.

Please refer to the return values of the tJobInstanceStart component.

## Basic settings

|  |  |
| --- | --- |
| Property | Content |
| Job Instance Start Component | Choose here the tJobInstanceStart component. Both components depend on each other. |
| Schema | This is necessary to have the schema column available. It is not supposed to change anything here |
| Configure Extraction | For every schema column you can define for which range it will be checked: Time range or Value range. The min and max values will be found even the component runs in iteration. |

## Return values

|  |  |
| --- | --- |
| Return value | Content |
| ERROR\_MESSAGE | Last error message in case of the range detection fails for a column. |
| TIME\_RANGE\_START | The min value for the measured time range. |
| TIME\_RANGE\_END | The max value for the measured time range. |
| VALUE\_RANGE\_START | The min value for the measured value range as Long or String |
| VALUE\_RANGE\_END | The max value for the measured value range as Long or String |
| NB\_LINE\_AGGREGATED | The number or rows for this component measured over all iterations |

# Component tJobInstanceLiveCheck

This component checks if jobs still alive. To do this, you have to build a very simple job (checkout the scenarios) and let them run on every job server you have.

## Basic settings

|  |  |
| --- | --- |
| Property | Content |
| Database Connection | Any database connection pointing to the schema with the control tables.  The main table is JOB\_INSTANCE\_STATUS holding all key information. |
| Close Connection | If true the connection will be closed at the end of the component processing |
| Schema | This component provides an input flow providing information about cleaned job instances |
| Last system start | If the last system start could be determined (currently there is not platform independent implementation to get this information automatically) all older job instance starts will be cleaned. |

## Return values

|  |  |
| --- | --- |
| Return value | Content |
| ERROR\_MESSAGE | Error message if something in the processing of the component it self went wrong |
| COUNT\_RUNNING\_PROCESSES | The number of all running processes on the current server (regardless if this is a Talend job or not) |
| COUNT\_RUNNING\_JOB\_INSTANCES | The number of as running declared job instances |
| COUNT\_BROKEN\_JOB\_INSTANCES | The number of recognized broken job instances |
| NB\_LINE | Number of rows in the data input flow |

The schema is fully commented and provides the values of the JOB\_INSTANCE\_STATUS table for the broken instances.

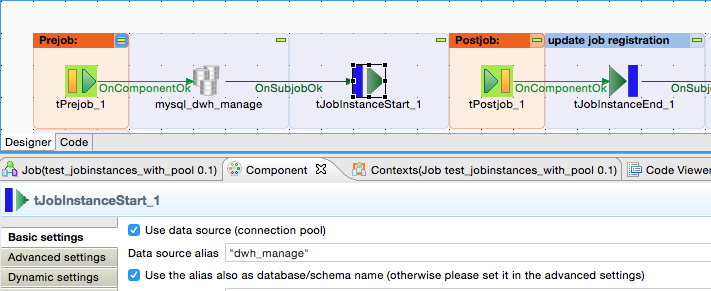
# Scenarios

## Scenario 1: Simple Job monitoring



The typical usage is to use tPrejob component to trigger the tJobInstanceStart component and the tPostjob component to trigger tJobInstanceEnd component.

## Scenario 2: Using a connection pool



The connection pool can be established anywhere unless it is before the initialization of the tJobInstanceStart component. Also, child jobs can use the same connection pool.

It is a good practice to name the pool like the addressed database. This way the configuration more convenient.

## Scenario 3: Measure the time ranges and/or value ranges



In this Scenario the flow will be scanned for the start and end values for a time range and the value range. These values could be used to ensure the job quality or to start the next run from the previous end.

Scenario for tJobInstanceLiveCheck



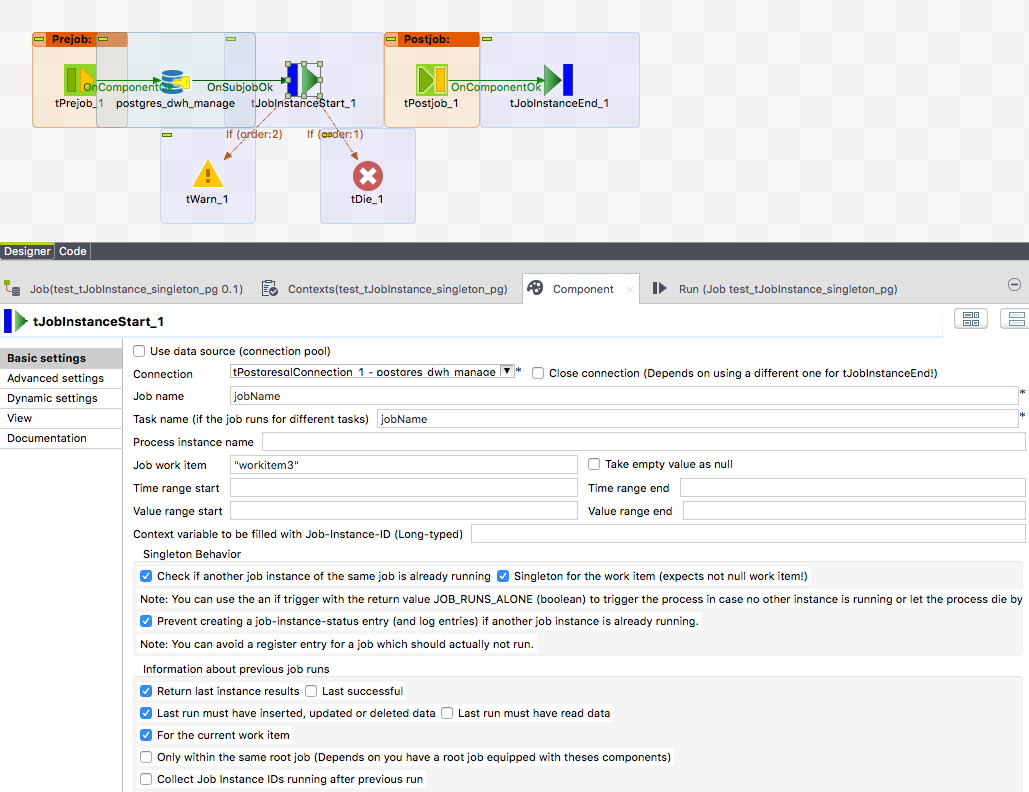
This example job shows the main purpose of the component. Such kind of job has to run frequently on every job server (servers on which the jobs run).

The component set for broken job instances the return code=999 and as return message “Process died”.

This information can be used to clean up all depending data structures.

## Scenario 4: Using the singleton check

The check depends in the information of the job\_instance\_status table. If you decide to use this check it is highly recommended to establish also a live check for job instances (use tJobInstanceLiveCheck component).



The if-trigger contains the return variable JOB\_RUNS\_ALONE to trigger the actual job or in this case to let the job die.

The tWarn component is an very simple example of a job. Start here with your actual work.

# Checklist to use this component suite

1. Use an OLTP database for the tables used by this component. Column oriented database are mostly to slow for the possible high frequency insert/updates from a large number of simultaneous running jobs! Typical bad choices: Teradata, Infobright engine of MySQL… Typical capable databases: Oracle, MySQL, DB2, PostgreSQL, H2
2. Build a separate database schema (e.g. call it dwh\_manage or dwh\_meta or what you want ;-)
3. Think about the primary key in the JOB\_INSTANCE\_STATUS table. You can use a database sequence (this is the preferred way) or you can use a self-incrementing data type (e.g. serial or identity column types). If you use a sequence, please set the SQL code to get the next value in the advanced settings of the tJobInstanceStart component. By the way, in large projects it could be helpful to set these settings per default directly in the component in the tJobInstanceStart\_java.xml file of the component.
4. If you use the component, please check if the job is long running job and, in this case, use a separate connection for the tJobInstanceEnd component or consider the usage of a connection pool to avoid problems with server-side disconnected database connections.

## Table: job\_instance\_counters

|  |  |  |
| --- | --- | --- |
| **column name** | **column type** | **description** |
| job\_instance\_id | bigint | Related job\_instance\_id |
| counter\_name | varchar(128) | Name of the counter. Set this name in the tJobInstanceEnd as name for an entry in the counter table. |
| counter\_type | Varchar(20) | Type of the counter (input, output,…) |
| counter\_value | integer | The value of the counter |

# Create table scripts for the tables

Not all databases are capable to work for this use case. Generally, all OLTP databases work fine. It could be problematic to host these tables on a column-based database or a database with distributed storage like Teradata.

Such database tends to be very slow for single inserts and updates or they lock the whole table while such operations and this could lead to significant performance problems!

The well-tested databases are MySQL (MyISAM or INNODB), Oracle, PostgreSQL, H2 and IBM DB2

In case of MySQL it is recommended using a serial data type for the column JOB\_INSTANCE\_STATUS(JOB\_INSTANCE\_ID).

In the advanced settings of the tJobInstanceStart component it is possible to declare the schema and the table names. The option Job Instance ID is auto increment allows the usage of auto increment column for JOB\_INSTANCE\_ID in the table JOB\_INSTANCE\_STATUS. In former releases some tables had slightly different names, but the meaning and structure is the same. You can adapt old names in the configuration if the tJobInstanceStart advanced settings.

## MySQL (use auto increment)

CREATE TABLE JOB\_INSTANCE\_STATUS (

JOB\_INSTANCE\_ID BIGINT UNSIGNED NOT NULL AUTO\_INCREMENT,

PROCESS\_INSTANCE\_ID BIGINT DEFAULT NULL,

PROCESS\_INSTANCE\_NAME VARCHAR(255) DEFAULT NULL,

JOB\_NAME VARCHAR(255) NOT NULL,

JOB\_PROJECT varchar(128),

JOB\_DISPLAY\_NAME VARCHAR(255) DEFAULT NULL,

JOB\_GUID VARCHAR(100) NOT NULL,

JOB\_EXT\_ID VARCHAR(255) DEFAULT NULL,

JOB\_INFO VARCHAR(255) DEFAULT NULL,

ROOT\_JOB\_GUID VARCHAR(100) DEFAULT NULL,

WORK\_ITEM VARCHAR(1024) DEFAULT NULL,

TIME\_RANGE\_START TIMESTAMP(3) NULL DEFAULT NULL,

TIME\_RANGE\_END TIMESTAMP(3) NULL DEFAULT NULL,

VALUE\_RANGE\_START VARCHAR(512) DEFAULT NULL,

VALUE\_RANGE\_END VARCHAR(512) DEFAULT NULL,

JOB\_STARTED\_AT TIMESTAMP(3) NULL DEFAULT NULL,

JOB\_ENDED\_AT TIMESTAMP(3) NULL DEFAULT NULL,

JOB\_RESULT VARCHAR(1024) DEFAULT NULL,

COUNT\_INPUT INT(11) DEFAULT NULL,

COUNT\_OUTPUT INT(11) DEFAULT NULL,

COUNT\_UPDATED INT(11) DEFAULT NULL,

COUNT\_REJECTED INT(11) DEFAULT NULL,

COUNT\_DELETED INT(11) DEFAULT NULL,

RETURN\_CODE INT(11) DEFAULT NULL,

RETURN\_MESSAGE TEXT,

HOST\_NAME VARCHAR(255) DEFAULT NULL,

HOST\_PID INT(11) DEFAULT NULL,

HOST\_USER VARCHAR(128) DEFAULT NULL,

PRIMARY KEY (JOB\_INSTANCE\_ID)

) DEFAULT CHARSET=UTF8;

CREATE INDEX JOB\_INSTANCE\_STATUS\_JOB\_GUID ON JOB\_INSTANCE\_STATUS(JOB\_GUID);

CREATE INDEX JOB\_INSTANCE\_STATUS\_JOB\_NAME ON JOB\_INSTANCE\_STATUS(JOB\_NAME);

CREATE TABLE JOB\_INSTANCE\_COUNTERS (

JOB\_INSTANCE\_ID BIGINT NOT NULL, -- reference to the job instance

COUNTER\_NAME VARCHAR(128) NOT NULL, -- name of the counter set in tJobInstanceEnd for a counter

COUNTER\_TYPE VARCHAR(20), -- type of the counter

COUNTER\_VALUE INTEGER, -- value of the counter

CONSTRAINT PK\_JOB\_INSTANCE\_COUNTERS PRIMARY KEY (JOB\_INSTANCE\_ID, COUNTER\_NAME));

## PostgreSQL (uses a sequence)

--drop table dwh\_manage.job\_instances;

create table dwh\_manage.job\_instance\_status (

job\_instance\_id bigint not null,

process\_instance\_id bigint,

process\_instance\_name varchar(255),

job\_name varchar(255) not null,

job\_project varchar(128),

job\_info varchar(512),

job\_display\_name varchar(255),

job\_guid varchar(100) not null,

job\_ext\_id varchar(255),

root\_job\_guid varchar(100),

work\_item varchar(1024),

time\_range\_start timestamp,

time\_range\_end timestamp,

value\_range\_start varchar(512),

value\_range\_end varchar(512),

job\_started\_at timestamp not null,

job\_ended\_at timestamp,

job\_result varchar(1024),

count\_input integer,

count\_output integer,

count\_updated integer,

count\_rejected integer,

count\_deleted integer,

return\_code integer,

return\_message varchar(1024),

host\_name varchar(255),

host\_pid integer,

host\_user varchar(128),

constraint job\_instances\_pkey primary key (job\_instance\_id));

create index job\_instances\_job\_guid on dwh\_manage.job\_instance\_status(job\_guid);

create index job\_instances\_job\_name on dwh\_manage.job\_instance\_status(job\_name);

--drop sequence dwh\_manage.seq\_job\_instance\_id;

create sequence dwh\_manage.seq\_job\_instance\_id start with 1;

--drop table dwh\_manage.job\_instance\_counters;

create table dwh\_manage.job\_instance\_counters (

job\_instance\_id bigint not null,

counter\_name varchar(128) not null,

counter\_type varchar(20),

counter\_value integer not null);

create index job\_instance\_counters\_idx on dwh\_manage.job\_instance\_counters(job\_instance\_id, counter\_name);

## Oracle (uses a sequence)

-- drop table JOB\_INSTANCE\_STATUS;

CREATE TABLE JOB\_INSTANCE\_STATUS (

JOB\_INSTANCE\_ID NUMBER(19) NOT NULL,

PROCESS\_INSTANCE\_ID NUMBER(19),

PROCESS\_INSTANCE\_NAME VARCHAR2(255),

JOB\_NAME VARCHAR2(255) NOT NULL,

JOB\_PROJECT VARCHAR2(128),

JOB\_INFO VARCHAR2(512),

JOB\_DISPLAY\_NAME VARCHAR2(255),

JOB\_GUID VARCHAR2(100) NOT NULL,

JOB\_EXT\_ID VARCHAR2(255),

ROOT\_JOB\_GUID VARCHAR2(100),

WORK\_ITEM VARCHAR2(1024),

TIME\_RANGE\_START TIMESTAMP,

TIME\_RANGE\_END TIMESTAMP,

VALUE\_RANGE\_START VARCHAR2(512),

VALUE\_RANGE\_END VARCHAR2(512),

JOB\_STARTED\_AT TIMESTAMP NOT NULL,

JOB\_ENDED\_AT TIMESTAMP,

JOB\_RESULT VARCHAR2(1024),

COUNT\_INPUT INTEGER,

COUNT\_OUTPUT INTEGER,

COUNT\_UPDATED INTEGER,

COUNT\_REJECTED INTEGER,

COUNT\_DELETED INTEGER,

RETURN\_CODE INTEGER,

RETURN\_MESSAGE VARCHAR2(1024),

HOST\_NAME VARCHAR2(255),

HOST\_PID INTEGER,

HOST\_USER VARCHAR(128),

CONSTRAINT JOB\_INSTANCES\_PKEY PRIMARY KEY (JOB\_INSTANCE\_ID));

CREATE INDEX JOB\_INSTANCES\_JOB\_GUID ON JOB\_INSTANCE\_STATUS(JOB\_GUID);

CREATE INDEX JOB\_INSTANCES\_JOB\_NAME ON JOB\_INSTANCE\_STATUS(JOB\_NAME);

CREATE SEQUENCE SEQ\_JOB\_INSTANCE\_ID START WITH 1;

CREATE TABLE JOB\_INSTANCE\_COUNTERS (

JOB\_INSTANCE\_ID NUMBER(19) NOT NULL,

COUNTER\_NAME VARCHAR2(128) NOT NULL,

COUNTER\_TYPE VARCHAR2(20),

COUNTER\_VALUE INTEGER NOT NULL);

CREATE INDEX JOB\_INSTANCE\_COUNTERS\_IDX ON JOB\_INSTANCE\_COUNTERS(JOB\_INSTANCE\_ID, COUNTER\_NAME);

## IBM DB2

--drop table dwh\_manage.job\_instance\_status;

create table dwh\_manage.job\_instance\_status (

job\_instance\_id bigint not null,

process\_instance\_id bigint,

process\_instance\_name varchar(255),

job\_name varchar(255) not null,

job\_project varchar(128),

job\_info varchar(512),

job\_display\_name varchar(255),

job\_guid varchar(100) not null,

job\_ext\_id varchar(255),

root\_job\_guid varchar(100),

work\_item varchar(1024),

time\_range\_start timestamp,

time\_range\_end timestamp,

value\_range\_start varchar(512),

value\_range\_end varchar(512),

job\_started\_at timestamp not null,

job\_ended\_at timestamp,

job\_result varchar(1024),

count\_input integer,

count\_output integer,

count\_updated integer,

count\_rejected integer,

count\_deleted integer,

return\_code integer,

return\_message varchar(1024),

host\_name varchar(255),

host\_pid integer,

host\_user varchar(128),

constraint job\_instances\_pkey primary key (job\_instance\_id));

create index dwh\_manage.job\_instances\_job\_guid on dwh\_manage.job\_instance\_status(job\_guid);

create index dwh\_manage.job\_instances\_job\_name on dwh\_manage.job\_instance\_status(job\_name);

--drop sequence dwh\_manage.job\_instance\_id;

create sequence dwh\_manage.seq\_job\_instance\_id start with 1;

--drop table dwh\_manage.job\_instance\_counters;

create table dwh\_manage.job\_instance\_counters (

job\_instance\_id bigint not null,

counter\_name varchar(128) not null,

counter\_type varchar(20),

counter\_value integer not null);

create index job\_instance\_counters\_idx on dwh\_manage.job\_instance\_counters(job\_instance\_id, counter\_name);

## Exasol

-- drop table dwh\_manage.job\_instance\_status;

create table dwh\_manage.job\_instance\_status (

job\_instance\_id bigint identity primary key,

process\_instance\_id bigint,

process\_instance\_name varchar(255),

job\_name varchar(255) not null,

job\_project varchar(512) UTF8,

job\_info varchar(512) UTF8,

job\_display\_name varchar(255) UTF8,

job\_guid varchar(100) UTF8 not null,

job\_ext\_id varchar(255) UTF8,

root\_job\_guid varchar(100) UTF8,

work\_item varchar(1024) UTF8,

time\_range\_start timestamp,

time\_range\_end timestamp,

value\_range\_start varchar(512) UTF8,

value\_range\_end varchar(512) UTF8,

job\_started\_at timestamp not null,

job\_ended\_at timestamp,

job\_result varchar(1024) UTF8,

count\_input integer,

count\_output integer,

count\_updated integer,

count\_rejected integer,

count\_deleted integer,

return\_code integer,

return\_message varchar(4000) UTF8,

host\_name varchar(255) UTF8,

host\_pid integer,

host\_user varchar(128) UTF8);

create table dwh\_manage.job\_instance\_counters (

job\_instance\_id bigint not null,

counter\_name varchar(128) not null,

counter\_type varchar(20),

counter\_value integer not null);