# *DWH Processes Controlling and Monitoring Application*



# Process Dispatch Center

#### Petr Stefanek, October, 2011

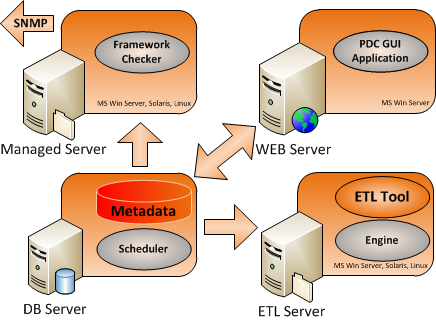
## Preface

Process Dispatch Center (PDC) is controlling framework application which is used for optimalization of ETL/ELT process and its supporting tasks. The main task of PDC is launching separate jobs in sequence based on their dependencies and priorities. While doing so it utilizes optimally systems resources and uses parallelism as much as possible. The frameworks logic takes care about optimal launch of all jobs, while the internal logic of the job itself is invisible to the framework. Framework runs all jobs in the same manner – using a child process which interprets the content of the job command line – and waits for the job exit code. In dependency on the exit code the framework considers the job finished either successfully or failed. All necessary information needed by the framework is stored as metadata in Oracle database. Metadata are used for:

* job configuration
* operational information about current and historical behavior of the jobs
* interface between all PDC parts
* source for representation of the current job status at GUI interface
* controlling of job behavior

There is no limitation of job type which can be executed by PDC application; the only limitation is that the interface used for job launching has to be installed on system where the framework Engine is running (e.g. for running an Informatica Job the Informatica Client need to be present on the system).

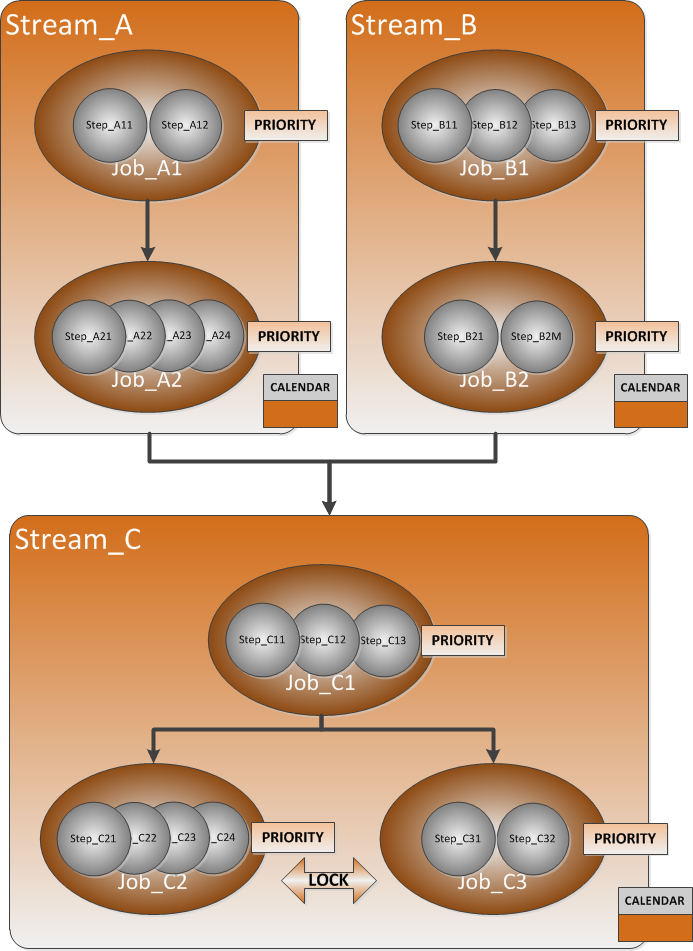
## PDC Architecture



PDC application uses for its work several parts – each taking care about specific functionality:

|  |  |
| --- | --- |
| ENGINE | executive part which launches jobs and tracks their run |
| SCHEDULER | hand-over of jobs to the Engine |
| METADATA | interdependency logic stored in Oracle procedures as well as configuration data |
| GUI | front end, GUI enables all necessary type of actions with PDC application.   * monitoring - tracking progress of processing and controlling job work. * controlling metadata creation * change management support and related tasks. |
| CHECKER | Checker detects a suspicious or erroneous situation it sends a SMTP trap to the alerting system. |

### Task Hierarchy



Tasks controlled by PDC application belongs to three hierarchical levels:

|  |  |
| --- | --- |
| STREAM | Virtual class joining jobs which together take care about some specific part of processing. The launching of jobs on the stream level is controlled using the calendar functionality. |
| JOB | the most important level is the job level. PDC application performs all controlling work on this level the, and job represents separate lowest level of a launch able task. All configuration information is directly or indirectly linked to job level. |
| STEP | breakdown of jobs. Steps are invisible from PDC application point of view. Step is subpart of a job and all steps of a job are running in a sequence. Step task is to provide specific functionality as load data into database, perform transformation and so on. In order to allow correct parallelism control the job can contain only steps of similar task type. Even if steps are for PDC application invisible, PDC can work with steps in case of job restart after failure – as resume functionality enables continuation of job processing from a failed step. |

### Parallelism

Parallelism is controlled by limitations imposed on total number of jobs running in parallel together with limitations on concurrency of jobs belonging to the same categorization (category and subcategory).

## Prerequisites

The below represent pre-requisites for use of PDC:

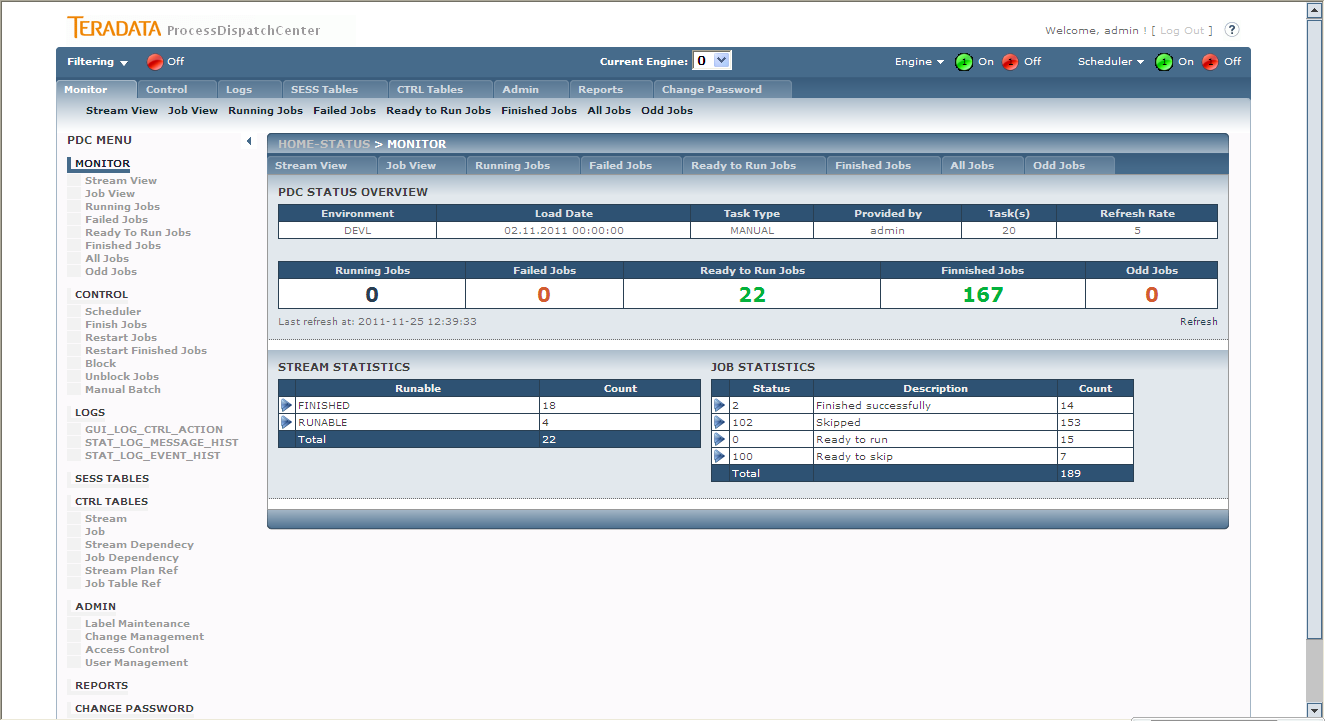
|  |  |
| --- | --- |
| **Engine, Checker** | **GUI** |
| Windows, Solaris or Linux system Perl 5.10 or higher (preferably Activestate) Oracle 11 or higher Client ETL Tool (e.g. Informatica) Client | Wintel Server with MS Windows 2xxx MS Internet Information Server .NET Framework 4 Java 1.6.0\_16 or higher Oracle 11 Client |

It is also expected that:

* Oracle schema will be dedicated for the PDC use (on any server).
* Installation of Engine is performed on system where ETL process is running to eliminate any performance bottlenecks.

## PDC GUI Application

Metadata permits PDC application administers all parts of its functionality. GUI application is WEB based application supporting standard browser functionality and using Teradata Viewpoint graphical layout. The user must have appropriate permissions for working with any part of the application. The main ask on GUI application is to enable comfortable monitoring and controlling of job processing, but GUI application is not necessary for PDC work. Jobs can be located in different instances of PDC; these instances are totally independent and are used for controlling and monitoring independent data processing. GUI shows only jobs located in selected instance of PDC but on status line are displayed statuses of all instances simultaneously. For selecting only specific part of objects a filter can be used on stream and job level. GUI consists from several parts whose meaning and functionality is described below. The monitoring screen during job processing phase is shown on the below screenshot.



### Status line

Status line is located in the top level of GUI application and shows the number of the engine which is selected for monitoring and controlling. Statuses of other Engines and Schedulers are also displayed. For faster touching of information which is looked for a filter can be used.

### Monitor

Monitor is used for displaying of current status of processing. The information can be presented from stream or job point of view. Drill down functionality enables drill for further details, so from stream view the user can drill down the information of how many jobs are located in the stream, what they are and what their status is. User can also drill down to parameters of selected job. On job level the user operates the job; it means he can abort running job, restart or finishing failed job and so on. Jobs and streams are divided into processing classes which represents objects state such as prepare for run, running, finished, failed and so on. The environment, load date and task type is shown as well as display refresh rate and maximal number of concurrently running jobs. The status overview part contains number of currently running jobs, number of failed jobs, number of jobs prepared for run and number of already finished jobs shows. All these numbers support drill down functionality, it means you can directly get a list of jobs in a category by clicking on the appropriate number.

### Control

Control enables access to all necessary settings used for controlling of PDC application. Same basic functionality for job processing is also located on the monitor part, but Control part enables driving parallelisms, temporary stop the Scheduler or doing control task on job level simultaneously for group of jobs. User can also simply stop executing jobs on selected dependency branch by blocking its parent job. Manual Batch functionality enables recalculation of selected jobs for chosen load date. It’s often used for datamarts recalculation for the day.

### Logs

Logs display controlling steps done on PDC application – it enables potential supervisor cooperation on problem solving, and also all problems captured by Framework checker application.

### SESS TABLES

SESS (session) tables contain production metadata and store actual status of objects. SESS TABLES part enables changes in the table content. Such change can represent increasing actual job priority, command line changes and so on.

### CTRL TABLES

CTRL (control) tables store configuration metadata of jobs, streams, their dependences, locks and calendars. This part of GUI application is used for creating or changing this metadata. Application automatically checks correctness of typed values. Changes are enabled only when a label is selected which is necessary for change management process package creation.

### Admin

Admin section contains necessary administrative tasks such as label maintenance, change management process, access control and so on.

### Reports

Reports page is a signpost for customer reports related for PDC processing. Customers can simply use their reporting tools for creating reports based on PDC metadata and place them on this page.