Oracle® Data Miner

Oracle Data Miner Release Notes

This document provides late-breaking information and information that is not yet part of the formal documentation.

This document contains the following topics:

New Features in Oracle Data Miner

The new features in Oracle Data Miner include:

- Oracle Data Mining Features
- Oracle Data Miner Features
- Oracle Database Features

Oracle Data Mining Features

The new Oracle Data Mining features include:

- Association Model Aggregation Metrics
- Enhancements to Algorithm Settings
- Support for Explicit Semantic Analysis Algorithm
- Enhancement to Data Mining Model Detail View
- Enhancements to Filter Column Node
- Mining Model Build Alerts
- R Build Model Node
- Support for Partitioned Models

Association Model Aggregation Metrics

Oracle Data Miner supports the enhanced Association Rules algorithm and allows the user to filter items before building the Association model.

The user can set the filters in the Association Build node editor, Association model viewer, and Model Details node editor.

> ✏ **See Also:**
>
> *Oracle Data Miner User's Guide*

Enhancements to Algorithm Settings

Oracle Data Miner has been enhanced to support enhancements in Oracle Data Mining that includes build settings for building partition models, sampling of training data, numeric data preparation that includes shift and scale transformations, and so on.

> ✏ **Note:**
>
> These settings are available if Oracle Data Miner 18.4 is connected to Oracle Database 12.2 and later.

Changes to the algorithms include:

Changes to Decision Tree Algorithm Settings

The setting Maximum Supervised Bins `CLAS_MAX_SUP_BINS` is added in the Decision Tree algorithm.

Changes to Expectation Maximization Algorithm Settings

The setting `Level of Details` replaces the current setting `Gather Cluster Statistics`.

The underlying algorithm setting used is `EMCS_CLUSTER_STATISTICS` where `All=ENABLE`, and `Hierarchy=DISABLE`. Some additional settings are added and some settings are deprecated.

Settings Added:
- Random Seed
- Model Search
- Remove Small Components

Settings Deprecated:

Approximate Computation `ODMS_APPROXIMATE_COMPUTATION`

Changes to Generalized Linear Models Algorithm Settings

The following changes are included in the Generalized Linear Model algorithm settings. The changes apply to both Classification models and Regression models.

Settings Added:

- Convergence Tolerance `GLMS_CONV_TOLERANCE`

- Number of Iterations `GLMS_NUM_ITERATIONS`

- Batch Rows `GLMS_BATCH_ROWS`

- Solver `GLMS_SOLVER`

- Sparse Solver `GLMS_SPARSE_SOLVER`

Settings Deprecated:

- Approximate Computation `ODMS_APPROXIMATE_COMPUTATION`

- Categorical Predictor Treatment `GLMS_SELECT_BLOCK`

- Sampling for Feature Identification `GLMS_FTR_IDENTIFICATION`

- Feature Acceptance `GLMS_FTR_ACCEPTANCE`

Changes to *k*-Means Algorithm Settings

The following changes are incorporated to the *k*-Means algorithm settings.

Settings Added:

- Levels of Details `KMNS_DETAILS`

- Random Seeds `KMNS_RANDOM_SEEDS`

Settings Deprecated:

- Growth Factor

Changes to Support Vector Machine Algorithm Settings

The following changes are included in the Support Vector Machine algorithm settings. The changes are applicable to both Linear and Gaussian kernel functions.

Settings Added:

- Solver `SVMS_SOLVER`

- Number of Iterations `SVMS_NUM_ITERATIONS`

- Regularizer `SVMS_REGULARIZER`

- Batch Rows`SVMS_BATCH_ROWS`

- Number of Pivots `SVMS_NUM_PIVOTS`

> **✎ Note:**
>
> Applies to Gaussian kernel function only.

Settings Deprecated:

- Active Learning
- Cache Size `SVMS_KERNEL_CACHE_SIZE`

> **✎ Note:**
>
> Applies to Gaussian kernel function only.

Changes to Singular Value Decomposition and Principal Components Analysis Algorithm Settings

The following changes are included in the Singular Value Decomposition and Principal Components Analysis algorithm.

Settings Added:

- Solver `SVDS_SOLVER`
- Tolerance `SVDS_TOLERANCE`
- Random Seed `SVDS_RANDOM_SEED`
- Over sampling `SVDS_OVER_SAMPLING`
- Power Iteration `SVDS_POWER_ITERATION`

Settings Deprecated:

- Approximate Computation `ODMS_APPROXIMATE_COMPUTATION`

Support for Explicit Semantic Analysis Algorithm

Oracle Data Miner 18.4 and later supports a new feature extraction algorithm called Explicit Semantic Analysis algorithm.

The algorithm is supported by two new nodes, that are Explicit Feature Extraction node and Feature Compare node.

Explicit Feature Extraction Node

The Explicit Feature Extraction node is built using the Explicit Semantic Analysis algorithm.

You can use the Explicit Feature Extraction node for the following:

- Document classification

- Information retrieval

- Calculations related to semantics

Feature Compare Node

The Feature Compare node enables you to perform calculations related to semantics in text data, contained in one Data Source node against another Data Source node.

The requirements of a Feature Compare node are:

- Two input data sources. The data source can be data flow of records, such as connected by a Data Source node or a single record data entered by user inside the node. In case of data entered by users, input data provider is not needed.

- One input Feature Extraction or Explicit Feature Extraction Model, where a model can be selected for calculations related to semantics.

Enhancement to Data Mining Model Detail View

The model viewers in Oracle Data Miner have been enhanced to reflect the changes in Oracle Data Mining.

Enhancements to the model viewers include the following:

- The computed settings within the model are displayed in the **Settings** tab of the model viewer.

- The new user embedded transformation dictionary view is integrated with the **Inputs** tab under **Settings.**

- The build details data are displayed in the **Summary** tab under **Summary**

- The Cluster model viewer detects models with partial details, and displays a message indicating so. This also applies to k-Means model viewer and Expectation Maximization model viewers.

Enhancements to Filter Column Node

Oracle Data mining supports unsupervised Attribute Importance ranking. The Attribute Importance ranking of a column is generated without the need for selecting a target column. The Filter Column node has been enhanced to support unsupervised Attribute Importance ranking.

Mining Model Build Alerts

Oracle Data Miner logs alerts related to model builds in the model viewers and event logs.

After a model build, Oracle Data Miner server queries Oracle Data Mining for any alerts related to the model build. The alerts are logged in:

- Model viewers: The build alerts are displayed in the **Alerts** tab.

- Event log: All build alerts are displayed along with other details such as job name, node, sub node, time, and message.

Support for Partitioned Models

Oracle Data Miner supports the building and testing of partitioned models.

The following models are enhanced to support partitioned models:

- Build Nodes
- Apply Nodes
- Test Nodes

R Build Model Node

Oracle Data Mining provides the feature to add R model implementations within the Oracle Data Mining framework. To support R model integration, Oracle Data Miner has been enhanced with a new R Build node with mining functions such as Classification, Regression, Clustering, and Feature Extraction.

Oracle Data Miner Features

The new Oracle Data Miner features include:

- Aggregation Node Support for DATE and TIMESTAMP Data Types
- Enhancement to JSON Query Node
- Enhancement to Build Nodes
- Enhancement to Text Settings
- Refresh Input Data Definition
- Support for Additional Data Types
- Support for In-Memory Column
- Support for Workflow Scheduling
- Workflow Status Polling Performance Improvement

Aggregation Node Support for DATE and TIMESTAMP Data Types

The Aggregation node has been enhanced to support DATE and TIMESTAMP data types.

For DATE and TIMESTAMP data types, the functions available are `COUNT(), COUNT (DISTINCT()), MAX(), MEDIAN(), MIN(), STATS_MODE().`

Enhancement to Build Nodes

All Build nodes are enhanced to support sampling of training data and preparation of numeric data.

The enhancement is implemented in the **Sampling** tab in all Build nodes editors. By default, the Sampling option is set to `OFF`. When set to `ON`, the user can specify the sample row size or choose the system determined settings.

> ✏️ **Note:**
>
> Data preparation is not supported in Association Build model.

The **Sampling** option is available in the following Build node editors:

*   Edit Anomaly Detection Node

*   Edit Association Build Node

*   Edit Classification Build Node

*   Edit Clustering Build Node

*   Edit Explicit Feature Extraction Build Node

*   Edit Feature Extraction Build Node

*   Edit Regression Build Node

Enhancement to JSON Query Node

The JSON Query node allows to specify filter conditions on attributes with data types such as `ARRAY, BOOLEAN, NUMBER` and `STRING`.

The user can apply filters to the data in hierarchical order using the option `All` or `Any` in the Filter Settings dialog box. The user also has the option to specify whether to apply filters to data that is used for relational data projection or aggregation definition or both by using any one of the following options:

*   JSON Unnest — Applies filter to JSON data that is used for projection to relational data format.

*   Aggregations — Applies filters to JSON data that is used for aggregation.

*   JSON Unnest and Aggregations — Applies filter to both.

Enhancement to Polling Performance

Polling performance and resource utilization functionality has been enhanced with new user interfaces.

The enhancement is supported by the following features:

*   The repository property `POLLING_IDLE_ENABLED` is added to determine whether the user interface will use automatic query or manual query for workflow status. This applies to the Workflow Jobs and Scheduled Jobs user interface. However, the Workflow Editor will continue to poll automatically when monitoring a workflow that is running.

> **Note:**
>
> When `POLLING_IDLE_ENABLED` is set to `TRUE`, then automatic query for workflow status sets in. When `POLLING_IDLE_ENABLED` is set to `FALSE`, then manual query sets in.

A new dockable window Scheduled Workflow has been added that displays the list of scheduled jobs and allows the user to manage the scheduled jobs.

- The Workflow Jobs window is enhanced with the following new features:
  - Manual refresh of workflow jobs.
  - Administrative override of automatic updates through Oracle Data Miner repository settings.
  - Access to Workflow Jobs preferences through the new Settings option.

Enhancement to Text Settings

Text settings are enhanced to support the following features:

- Text support for synonyms (thesaurus): Text Mining in Oracle Data Miner supports synonyms. By default, no thesaurus is loaded. The user must manually load the default thesaurus provided by Oracle Text or upload his own thesaurus.

- New settings added in **Text** tab:
  - Minimum number of rows (documents) required for a token
  - Max number of tokens across all rows (documents)
  - New tokens added for BIGRAM setting:
    * `BIGRAM`: Here, `NORMAL` tokens are mixed with their bigrams
    * `STEM BIGRAM`: Here, `STEM` tokens are extracted first and then stem bigrams are formed.

Refresh Input Data Definition

Use the **Refresh Input Data Definition** option if you want to update the workflow with new columns, that are either added or removed.

The **Refresh Input Data Definition** option is equivalent to `SELECT*` capability in the input source. The option allows you to quickly refresh your workflow definitions to include or exclude columns, as applicable.

> **Note:**
>
> The Refresh Input Data Definition option is available as a context menu option in Data Source nodes and SQL Query nodes.

Support for Additional Data Types

Oracle Data Miner allows the following data types for input as columns in a Data Source node, and as new computed columns within the workflow:

- `RAW`
- `ROWID`
- `UROWID`
- `URITYPE`

The `URITYPE` data type provides many sub type instances, which are also supported by Oracle Data Miner. They are:

- `HTTPURITYPE`
- `DBURITYPE`
- `XDBURITYPE`

Support for In-Memory Column

Oracle Data Miner supports In-Memory Column Store (IM Column Store) in Oracle Database 12.2 and later, which is an optional static SGA pool that stores copies of tables and partitions in a special columnar format.

Oracle Data Miner has been enhanced to support In-Memory Column in nodes in a workflow. For In-Memory Column settings, the options to set Data Compression Method and Priority Level are available in the Edit Node Performance Settings dialog box.

Support for Workflow Scheduling

Oracle Data Miner supports the feature to schedule workflows to run at a definite date and time.

A scheduled workflow is available only for viewing. The option to cancel a scheduled workflow is available. After cancelling a scheduled workflow, the workflow can be edited and rescheduled.

Workflow Status Polling Performance Improvement

The performance of workflow status polling has been enhanced.

The enhancement includes new repository views, repository properties, and user interface changes:

- The repository view `ODMR_USER_WORKFLOW_ALL_POLL` is added for workflow status polling.
- The following repository properties are added:
  - `POLLING_IDLE_RATE`: Determines the rate at which the client will poll the database when there are no workflows detected as running.

- POLLING_ACTIVE_RATE: Determines the rate at which the client will poll the database when there are workflows detected running.

- POLLING_IDLE_ENABLED: Determines whether the user interface will use automatic query or manual query for workflow status. This applies to the Workflow Jobs and Scheduled Jobs user interface. However, the Workflow Editor will continue to poll automatically when monitoring a workflow that is running.

> ✏️ **Note:**
>
> When POLLING_IDLE_ENABLED is set to TRUE, then automatic query for workflow status sets in. When POLLING_IDLE_ENABLED is set to FALSE, then manual query sets in.

- POLLING_COMPLETED_WINDOW: Determines the time required to include completed workfows in the polling query result.

- PURGE_WORKFLOW_SCHEDULER_JOBS: Purges old Oracle Scheduler objects generated by the running of Data Miner workflows.

- PURGE_WORKFLOW_EVENT_LOG: Controls how many workflow runs are preserved for each workflow in the event log. The events of the older workflow are purged to keep within the limit.

- New user interface includes the Scheduled Jobs window which can be accessed from the **Data Miner** option in both Tools menu and View menu in SQL Developer 18.4 and later.

Oracle Database Features

The new Oracle Database feature includes Support for Expanded Object Name.

Support for Expanded Object Name

The support for schema name, table name, column name, and synonym that are 128 bytes are available in the upcoming Oracle Database release. To support Oracle Database, Oracle Data Miner repository views, tables, XML schema, and PL/SQL packages are enhanced to support 128 bytes names.

Supported Platforms

For details on supported platforms, see *Oracle SQL Developer Installation Guide*.

Prerequisites for Oracle Data Miner

Before you can use Oracle Data Miner, ensure the following:

1. Install SQL Developer 18.4 or later on your system.

2. Secure access to an Oracle Database:

   - Minimum version: Oracle Database 11.2.0.4 Enterprise Edition, with the Data Mining option.

   - Preferred version: Oracle Database 12.2.0.1 Enterprise Edition.

3. Create a database user account for data mining.

4. Create a database connection within SQL Developer for the Oracle Data Miner user.

5. Install the Oracle Data Miner repository.

> **Note:**
>
> The SH sample schema is not shipped with Oracle Database 12.2. To install the sample schema, go to DB Sample Schemas.

Known Problems and Limitations

Known problems and limitations in this release includes:

- Association Model Build node cannot consume data coming directly from JSON Query node.

  Users must persist the data coming from the JSON Query node through Create Table node, and then use the persisted data as input to the Associate Model Build node.

- Classification nodes and Regression Model Build nodes are unable to consume data directly coming from JSON Query node if JSON Aggregations (with Sub Group By) are defined.

  Users must persist the data coming from the JSON Query node through Create Table node, and then use the persisted data as input to these Build nodes.

> **Note:**
>
> Build nodes can consume data directly coming from JSON Query nodes if JSON Aggregations (without Sub Group By) are not defined.

- Setting Parallel Query for a node that queries JSON data can result in a workflow runtime error. JSON queries will fail if they are run with the database Parallel Query set to ON. The following error message is displayed ORA-12805: Parallel Query server died unexpectedly.

  You can configure Parallel Query through Oracle Data Miner at the node level:

- The Node context menu has the option to set Parallel Query. Click **Parallel Query** and select the nodes to configure the parallel settings.

- The View Data viewer provides the option to set Parallel Query to `ON` when querying the selected Data Nodes.

  In both the cases, the error occurs and the same error message is displayed.

- Multi byte character data is not supported in Oracle Data Miner with Oracle Database 12.1 because of database issues. To address the multibyte issue, apply the Oracle Database 12.1.0.2 patch. It is also recommended to use the AL32UTF8 character set.

> **Note:**
>
> Request the Oracle Database 12.1.0.2 patch through Oracle Support.

- When installing Oracle Data Miner repository, error messages are generated which can be ignored. The error messages are related to database objects that are loaded during installation. After all the database objects are installed, a complete re-compilation is performed. If there are invalid objects, then an exception is raised. If the script does not raise any exceptions, then it means that the installation of Oracle Data Miner repository is successful.

> **Note:**
>
> You can ignore these error messages during Oracle Data Miner 18.4 or later installation, if no exceptions are generated.

Bug Fixes

Oracle Data Miner has the following bug fixed:

Bug: 27568325

Issue: When starting a workflow in Oracle Data Miner, the message states that the workflow has started, and the status shows as *Running.* But if the user clicks **Stop**, then the functionality to stop the workflow does not work as expected.

The issue occurs when the user migrates Oracle Data Miner from Oracle Data Miner 4.1 or before. However, this is issue does not work when migrating from Oracle Data Miner 4.2 or later.

Resolution: If you are migrating from Oracle Data Miner 4.1 or before and encounter this issue, then run the following statements by logging in to the `sys` account.

```
CREATE OR REPLACE VIEW ODMRSYS.ODMR_ALL_WORKFLOW_JOBS AS SELECT * FROM
(
  WITH WF_JOBS AS
  (
```

```
          SELECT WORKFLOW_ID, WORKFLOW_JOB_ID, CREATION_TIME FROM
ODMRSYS.ODMR$WORKFLOW_JOBS
    ),
    CHAIN_JOBS AS
    (
    SELECT JOB_NAME, STATE FROM ALL_SCHEDULER_JOBS WHERE JOB_TYPE =
'CHAIN'
    )
    SELECT
        ODMR$PROJECTS.USER_NAME,
        ODMR$PROJECTS.PROJECT_ID,
        ODMR$PROJECTS.PROJECT_NAME,
        ODMR$WORKFLOWS.WORKFLOW_NAME,
        ODMR$WORKFLOWS.CHAIN_NAME "WORKFLOW_CHAIN_NAME",
        WF_JOBS.WORKFLOW_ID,
        WF_JOBS.WORKFLOW_JOB_ID "WORKFLOW_JOB_NAME",
        WF_JOBS.CREATION_TIME "JOB_CREATION_TIME",
        CASE UPPER(NVL(CHAIN_JOBS.STATE, 'completed'))
        WHEN 'DISABLED' THEN 'FAILED'
        WHEN 'SCHEDULED' THEN 'SCHEDULED'
        WHEN 'RUNNING' THEN 'RUNNING'
        WHEN 'COMPLETED' THEN 'SUCCEEDED'
        WHEN 'STOPPED' THEN 'STOPPED'
        WHEN 'BROKEN' THEN 'FAILED'
        WHEN 'RETRY SCHEDULED' THEN 'SCHEDULED'
        WHEN 'SUCCEEDED' THEN 'SUCCEEDED'
        WHEN 'CHAIN_STALLED' THEN 'FAILED'
        ELSE CHAIN_JOBS.STATE END "WORKFLOW_STATUS"
    FROM
        WF_JOBS,
        ODMRSYS.ODMR$WORKFLOWS,
        ODMRSYS.ODMR$PROJECTS,
        CHAIN_JOBS
    WHERE
        ODMR$PROJECTS.PROJECT_ID = ODMR$WORKFLOWS.PROJECT_ID (+)
        AND ODMR$WORKFLOWS.WORKFLOW_ID = WF_JOBS.WORKFLOW_ID (+)
        AND WF_JOBS.WORKFLOW_JOB_ID = CHAIN_JOBS.JOB_NAME (+)
);


CREATE OR REPLACE VIEW ODMRSYS.ODMR_ALL_PROJECT_WORKFLOW AS SELECT *
FROM
(
    WITH WF_JOBS AS
    (
    SELECT WORKFLOW_ID, WORKFLOW_JOB_ID, CREATION_TIME
        FROM (
        SELECT t.*, row_number () OVER (partition by WORKFLOW_ID ORDER BY
CREATION_TIME DESC) rid
        FROM ODMRSYS.ODMR$WORKFLOW_JOBS t )
        WHERE rid = 1
    ),
```

```
    CHAIN_JOBS AS
    (
    SELECT JOB_NAME, STATE FROM ALL_SCHEDULER_JOBS WHERE JOB_TYPE =
'CHAIN'
    )
    SELECT
        ODMR$PROJECTS.USER_NAME,
        ODMR$PROJECTS.PROJECT_ID,
        ODMR$PROJECTS.PROJECT_NAME,
        ODMR$PROJECTS.CREATION_TIME PJ_CREATION_TIME,
        ODMR$PROJECTS.LAST_UPDATED_TIME PJ_LAST_UPDATED_TIME,
        ODMR$PROJECTS.COMMENTS PJ_COMMENTS,
        ODMR$WORKFLOWS.WORKFLOW_ID,
        ODMR$WORKFLOWS.WORKFLOW_NAME,
        ODMR$WORKFLOWS.WORKFLOW_DATA,
        ODMR$WORKFLOWS.CHAIN_NAME,
        ODMR$WORKFLOWS.CREATION_TIME WF_CREATION_TIME,
        ODMR$WORKFLOWS.LAST_UPDATED_TIME WF_LAST_UPDATED_TIME,
        ODMR$WORKFLOWS.COMMENTS WF_COMMENTS,
        CASE UPPER(NVL(CHAIN_JOBS.STATE, 'completed'))
        WHEN 'DISABLED' THEN 'FAILED'
        WHEN 'SCHEDULED' THEN 'SCHEDULED'
        WHEN 'RUNNING' THEN 'RUNNING'
        WHEN 'COMPLETED' THEN 'SUCCEEDED'
        WHEN 'STOPPED' THEN 'STOPPED'
        WHEN 'BROKEN' THEN 'FAILED'
        WHEN 'RETRY SCHEDULED' THEN 'SCHEDULED'
        WHEN 'SUCCEEDED' THEN 'SUCCEEDED'
        WHEN 'CHAIN_STALLED' THEN 'FAILED'
        ELSE CHAIN_JOBS.STATE END "STATUS"
    FROM
        WF_JOBS,
        ODMRSYS.ODMR$WORKFLOWS,
        ODMRSYS.ODMR$PROJECTS,
        CHAIN_JOBS
    WHERE
        ODMR$PROJECTS.PROJECT_ID = ODMR$WORKFLOWS.PROJECT_ID (+)
        AND ODMR$WORKFLOWS.WORKFLOW_ID = WF_JOBS.WORKFLOW_ID (+)
        AND WF_JOBS.WORKFLOW_JOB_ID = CHAIN_JOBS.JOB_NAME (+)
);
```

Documentation Accessibility

For information about Oracle's commitment to accessibility, visit the
Oracle Accessibility Program website at http://www.oracle.com/pls/topic/lookup?
ctx=acc&id=docacc.

Access to Oracle Support

Oracle customers that have purchased support have access to electronic support through My Oracle Support. For information, visit http://www.oracle.com/pls/topic/lookup?ctx=acc&id=info or visit http://www.oracle.com/pls/topic/lookup?ctx=acc&id=trs if you are hearing impaired.