Oracle® Endeca Web Acquisition Toolkit

Installation Guide

Version 3.1.1 • December 2013

ORACLE®

Copyright and disclaimer

Table of Contents

Preface

Oracle® Endeca Information Discovery Integrator is a powerful visual data integration environment that includes:

The Information Acquisition System (IAS) for gathering content from delimited files, file systems, JDBC databases, and Web sites.

Integrator ETL, an out-of-the-box ETL purpose-built for incorporating data from a wide array of sources, including Oracle BI Server.

In addition, Oracle Endeca Web Acquisition Toolkit is a Web-based graphical ETL tool, sold as an add-on module. Text Enrichment and Text Enrichment with Sentiment Analysis are also sold as add-on modules. Connectivity to data is also available through Oracle Data Integrator (ODI).

About this guide

This guide describes how to download, install, and start the Web Acquisition Toolkit.

Who should use this guide

This guide is intended for data developers who are responsible for loading source data into an Endeca data domain.

Conventions used in this guide

The following conventions are used in this document.

Typographic conventions

The following table describes the typographic conventions used in this document.

Typographic conventions

Typeface	Meaning
User Interface Elements	This formatting is used for graphical user interface elements such as pages, dialog boxes, buttons, and fields.
`Code Sample`	This formatting is used for sample code phrases within a paragraph.
`<Variable Name>`	This formatting is used for variable values, such as <install path>.
`File Path`	This formatting is used for file names and paths.

Symbol conventions

The following table describes symbol conventions used in this document.

Symbol conventions

Symbol	Description	Example	Meaning
>	The right angle bracket, or greater-than sign, indicates menu item selections in a graphic user interface.	File > New > Project	From the File menu, choose New, then from the New submenu, choose Project.

Contacting Oracle Customer Support

Oracle Customer Support provides registered users with important information regarding Oracle software, implementation questions, product and solution help, as well as overall news and updates from Oracle.

You can contact Oracle Customer Support through Oracle's Support portal, My Oracle Support at *https://support.oracle.com*.

Chapter 1
Overview of the Web Acquisition Toolkit

This section provides a brief overview of the Web Acquisition Toolkit (WAT) and its system requirements.

About the Web Acquisition Toolkit

The Web Acquisition Toolkit is a platform for application integration and process automation. It can integrate applications that weren't built to be connected and automate processes across such heterogeneous systems; cloud/SaaS applications with premise systems, legacy systems with modern Web applications, back office systems with partner Web sites.

With the visual editor Design Studio, you click through the applications and data sources you want to integrate and create an automated workflow.

In the Web Acquisition Toolkit, these workflows are known as *robots*. As you build a robot, you are free to navigate through the applications as you integrate them. You can log in to applications, extract data parts of a page, enter data into forms or search boxes, make menu selections, and scroll through multiple pages. Your robot can also access databases, files, APIs, and Web services, exporting data from one application and loading it into another; transforming data as necessary along the way.

Once built, robots are uploaded to a repository in the *Management Console*. From here, they can be scheduled for batch-execution on the RoboServer or executed on-demand via Java or C# APIs, tailored REST services that are instantaneously available once the robots have been added to the repository, or exposed as special-purpose end-user web applications called *Kapplets*.

Integration of the Web Acquisition Toolkit with Integrator ETL

The Web Acquisition Toolkit can provide source data for a Studio application. When you run a robot, WAT stores its record output in a database of your choice. (For a list of supported databases, see the "Supported Platforms" topic of the Kapow Katalyst help available at http://help.kapowsoftware.com/9.2/index.jsp.)

You incorporate the WAT records into an Integrator ETL graph by creating a Reader Database component and configuring it to read from the WAT database. After reading the WAT records into an Integrator ETL

graph, you perform any additional data transformation that is necessary, and write the graph output to Endeca Server.

Supported operating systems

WAT supports the following operating systems:

Operating System	Description
Oracle Linux	• Oracle Linux Release 6 for x86 and x64 (both 32-bit and 64-bit platforms). • Oracle Linux Release 5.6 for x86 and x64 (both 32-bit and 64-bit platforms). Both the Oracle Unbreakable Enterprise Kernel and the Red Hat Compatible Kernel are supported. **Note:** The Design Studio is supported only on version 6. **Note:** WAT provides a 32-bit version of Design Studio in the 64-bit distribution for Linux. The 32-bit version of Design Studio can be run on a 64-bit Linux system using several additional package ia32-libs.
RHEL Linux	• Red Hat Enterprise Linux Server version 6 for x86 and x64 (both 32-bit and 64-bit platforms). • Red Hat Enterprise Linux Advanced Platform version 6 for x86 and x64 (both 32-bit and 64-bit platforms). • Red Hat Enterprise Linux Server version 5.6 for x86 and x64 (both 32-bit and 64-bit platforms). • Red Hat Enterprise Linux Advanced Platform version 5.6 for x86f and x64 (both 32-bit and 64-bit platforms). **Note:** The Design Studio is supported only on version 6. **Note:** WAT provides a 32-bit version of Design Studio in the 64-bit distribution for Linux. The 32-bit version of Design Studio can be run on a 64-bit Linux system using several additional package ia32-libs.
Windows 2008	Windows Server 2008 R2 Enterprise.

Hardware requirements

The hardware requirements are described in the "System Requirements" topic of the Kapow Katalyst help available at http://help.kapowsoftware.com/9.2/index.jsp.

Database requirements

The Web Acquisition Toolkit supports a variety of databases to store output from a robot. See the table of database support in the "Supported Platforms" topic of the Kapow Katalyst help available at http://help.kapowsoftware.com/9.2/index.jsp.

Chapter 2

Installing the Web Acquisition Toolkit

This section describes how to download, install, and license the Web Acquisition Toolkit.

Downloading the WAT for 64-bit systems

You download the WAT from the Oracle Software Delivery Cloud.

There are some differences in product packaging depending on the **Platform** you select:

* On Windows, the 64-bit distribution of the WAT includes a 64-bit Roboserver and Design Studio.

* On Linux, the 64-bit distribution of the WAT includes a 64-bit Roboserver and a 32-bit version of Design Studio. The 32-bit version of Design Studio can be run on a 64-bit Linux system using several additional package ia32-libs.

To download the WAT for 64-bit systems:

1. Sign in to `https://edelivery.oracle.com`.

2. Accept the terms and restrictions and click **Continue**.

3. On the **Media Pack Search** page:

    (a) From the **Select a Product Pack** drop-down list, select **Oracle Endeca**.

    (b) From the **Platform** drop-down list, select the platform you are installing on.

    (c) Click **Go**.

4. In the media pack list, click **Oracle Endeca Web Acquisition Toolkit (<version>) Media Pack for <platform>**.

5. Download **Web Acquisition Toolkit (<version>) for <platform>** to a local directory.

6. Optionally, download the **Oracle Endeca Web Acquisition Toolkit (<version>) Documentation** if you want a local copy of the documentation.

Downloading the WAT for 32-bit Linux systems

You download the WAT from the Oracle Software Delivery Cloud.

To download the WAT for 32-bit Linux systems:

1. Sign in to `https://edelivery.oracle.com`.

2. Accept the terms and restrictions and click **Continue**.

3. On the **Media Pack Search** page:
   (a) From the **Select a Product Pack** drop-down list, select **Oracle Endeca**.
   (b) From the **Platform** drop-down list, select **Linux x86**.
   (c) Click **Go**.

4. Click **Download** for the **Oracle Endeca Web Acquisition Toolkit (<version>) Media Pack for Linux x86**.

5. Optionally, download the **Oracle Endeca Web Acquisition Toolkit (<version>) Documentation** if you want a local copy of the documentation.

Installing the WAT on Windows

The installation process consists of extracting a ZIP file to put the WAT installer and licensing file on disk, and then running the WAT installer.

Before installing, ensure that you have administrator privileges on the local machine.

To install the WAT on Windows:

1. In your local environment, locate the WAT software that you downloaded from the Oracle Software Delivery Cloud.

2. Extract the ZIP file to a local directory.

   On Windows, the ZIP file contains:

   - `EID_<version>WebAcquisitionToolkit.msi`

   - `EID_<version>WebAcquisitionToolkit_License.txt`

3. Double-click `EID_<version>WebAcquisitionToolkit.msi` to start the installer.

4. If an **Open File - Security Warning** displays, select **Run**.

5.   In the **Welcome ...** dialog, click **Next**.

For example:
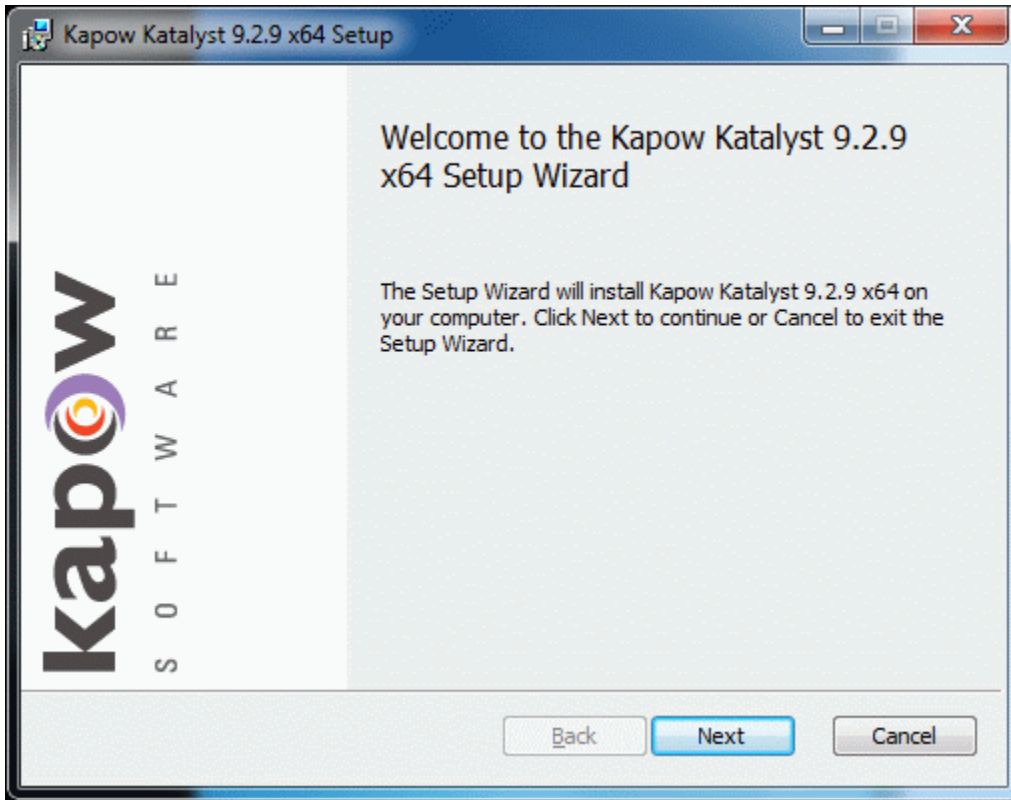
6. In the **End-User License Agreement** dialog, accept the license agreement and click **Next**.

For example:

7.   In the **Destination Folder** dialog, accept the default installation path or change it if necessary and
       click **Next**.

       For example:

8.   In the **Ready to install ...** dialog, click **Install**.

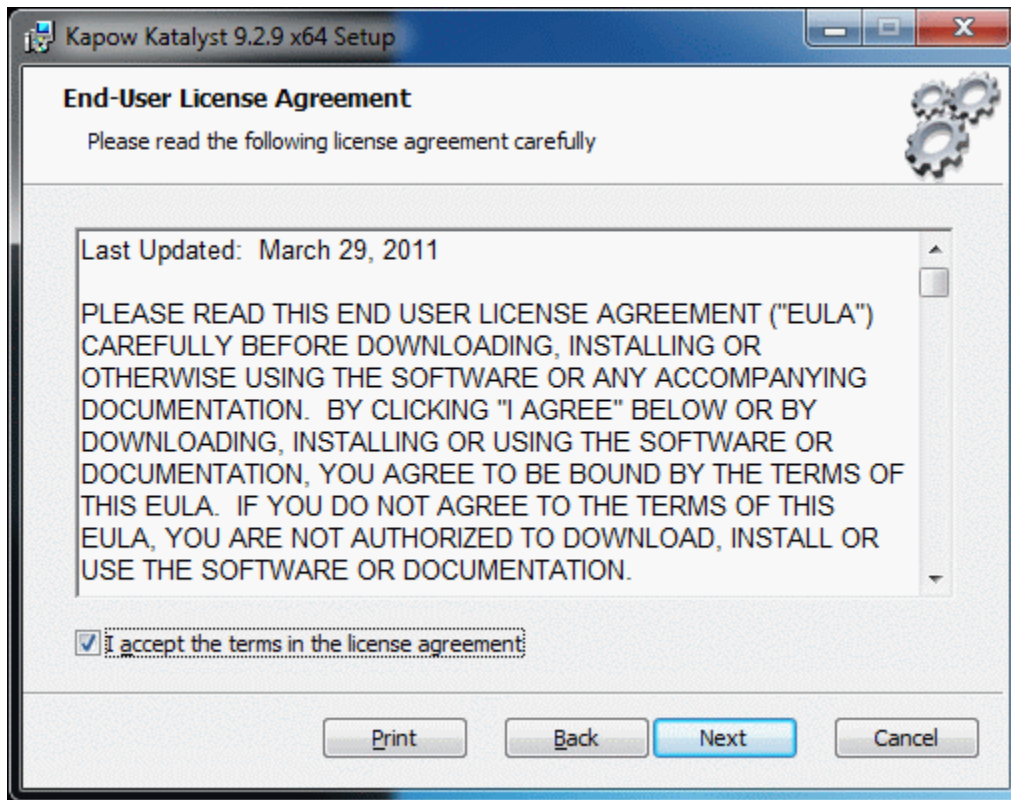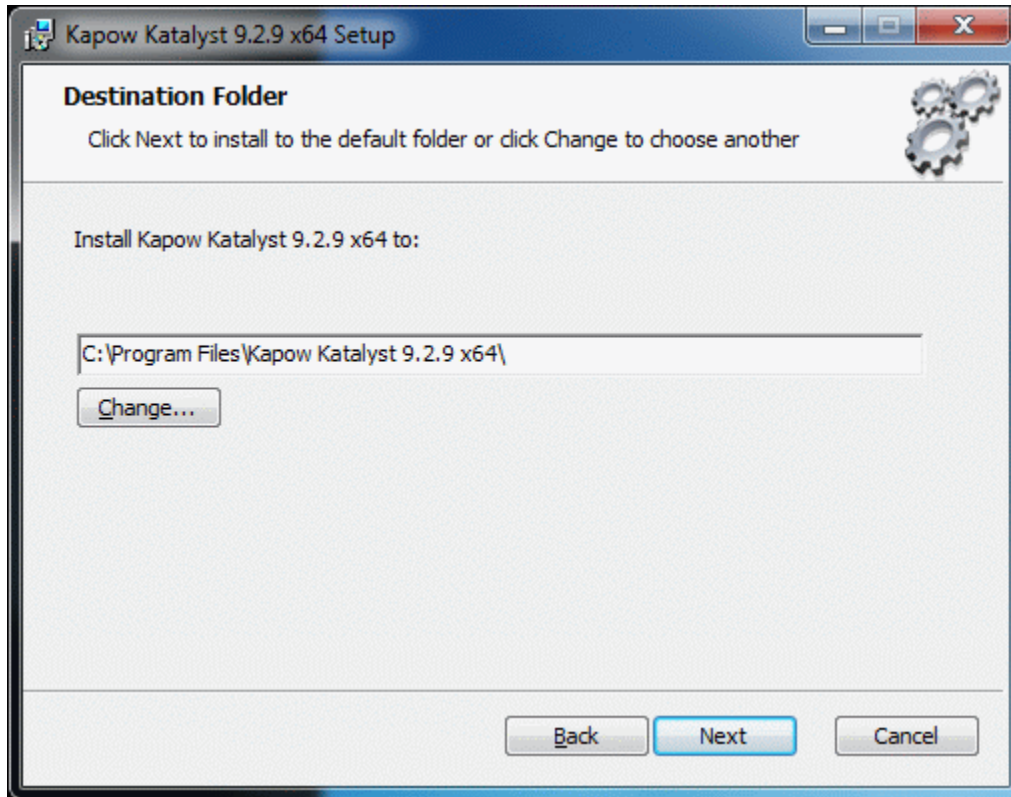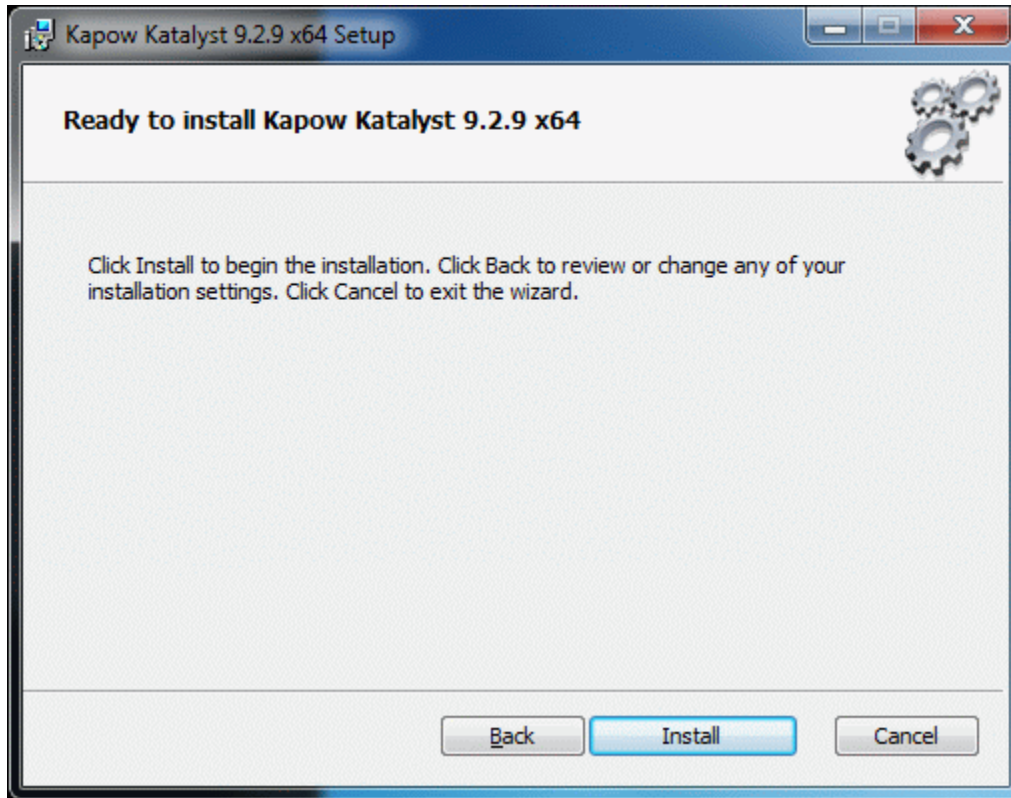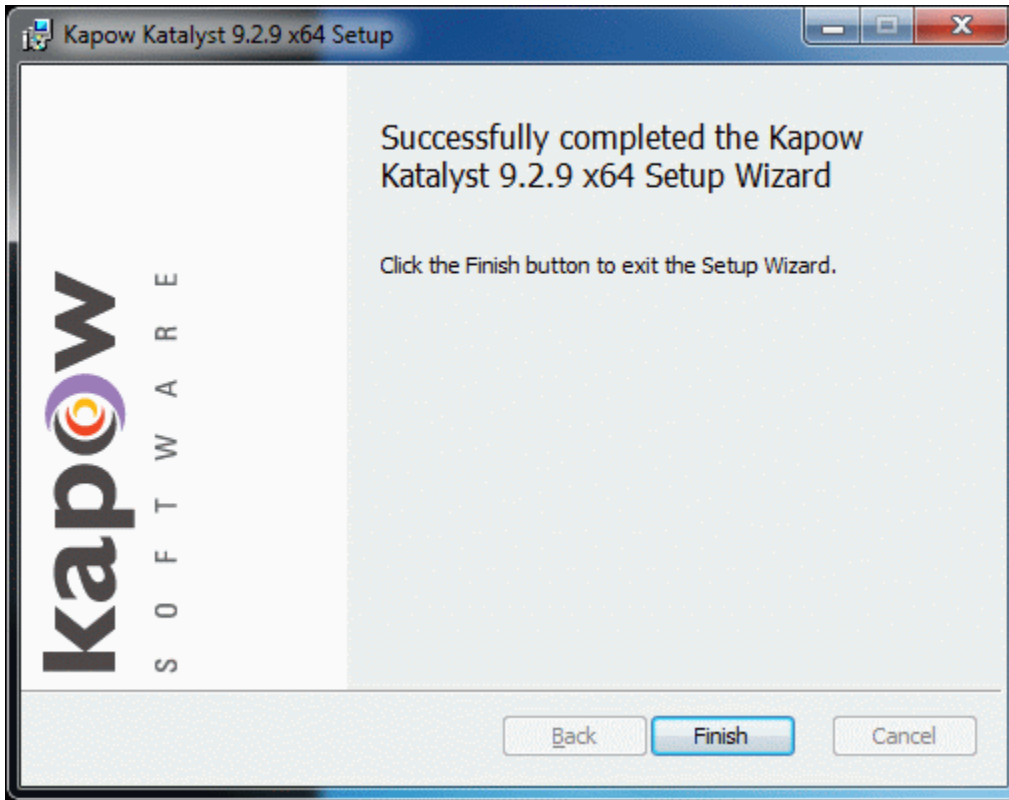For example:

9.   When the Setup Wizard completes, click **Finish**.

For example:

By default, the WAT installs into `C:\Program Files\Kapow Katalyst 9.2.9 x64.`

After installation, go on to *Licensing the WAT on page 17*.

Installing the WAT on 64-bit Linux systems

The installation process consists of extracting a ZIP file to put the WAT files on disk. Also, in order to run the 32-bit version of Design Studio on a 64-bit system, you need to install several additional libraries described in steps 4 - 5.

To install the WAT on 64-bit Linux systems:

1.   In your local environment, locate the WAT software that you downloaded from the Oracle Software Delivery Cloud.

2.   Extract the ZIP file to a local directory.

On Linux, the ZIP file contains:

- `EID_<version>WebAcquisitionToolkit_Server.tar.gz`

- `EID_<version>WebAcquisitionToolkit_License.txt`

3.   Extract the `tar.gz` file with the following command:

`$tar xzf EID_3.2_WebAcquisitionToolkit_Server.tar.gz`

4. Obtain the following libraries from your system administrator:
   - `glibc.i686`
   - `libXext.so.6`
   - `libXtst.so.6`

5. Install the libraries by running the following command:

```
yum install glibc.i686
yum install libXext.so.6
yum install libXtst.so.6
```

By default, the 64-bit distribution of WAT extracts into a directory named `../Kapow Katalyst 9.2.9 x64`.

After installation, go on to .

Installing the WAT on 32-bit Linux systems

The installation process consists of extracting a ZIP file to put the WAT files on disk.

To install the WAT on 32-bit Linux systems:

1. In your local environment, locate the WAT software that you downloaded from the Oracle Software Delivery Cloud.

2. Extract the ZIP file to a local directory.

   On Linux, the ZIP file contains:
   - `EID_<version>WebAcquisitionToolkit.tar.gz`
   - `EID_<version>WebAcquisitionToolkit_License.txt`

3. Extract the `tar.gz` file with the following command:

   `$tar xzf EID_3.2_WebAcquisitionToolkit.tar.gz`

By default, the 32-bit distribution of WAT extracts into a directory named `../Kapow Katalyst 9.2.9`.

After installation, go on to .

Files and directory structure after installation

For more information about the files and directory structure after installation, see the "Installation Folder" and "Project Folder" topics of the Kapow Katalyst help available at http://help.kapowsoftware.com/9.2/index.jsp.

Licensing the WAT

After running the installation program on Windows or Linux, you must license the Web Acquisition Toolkit by specifying a license key.

To license the WAT:

1. Start the Management Console.

    *   On Windows, go to **Start>All Programs>Kapow Katalyst 9.2.9 x64>Start Management Console 9.2.9**.
    *   On Linux, open a command prompt, change to `../Kapow Katalyst 9.2.9 x64/bin`, and run: `$./RoboServer -p 50000 -MC`.

    A command prompt window remains running in the background for the Management Console.

2. Start a Web browser and go to `http://localhost:50080/`.

3. In the browser, click **I Agree** to accept the End User License Agreement.

    For example:
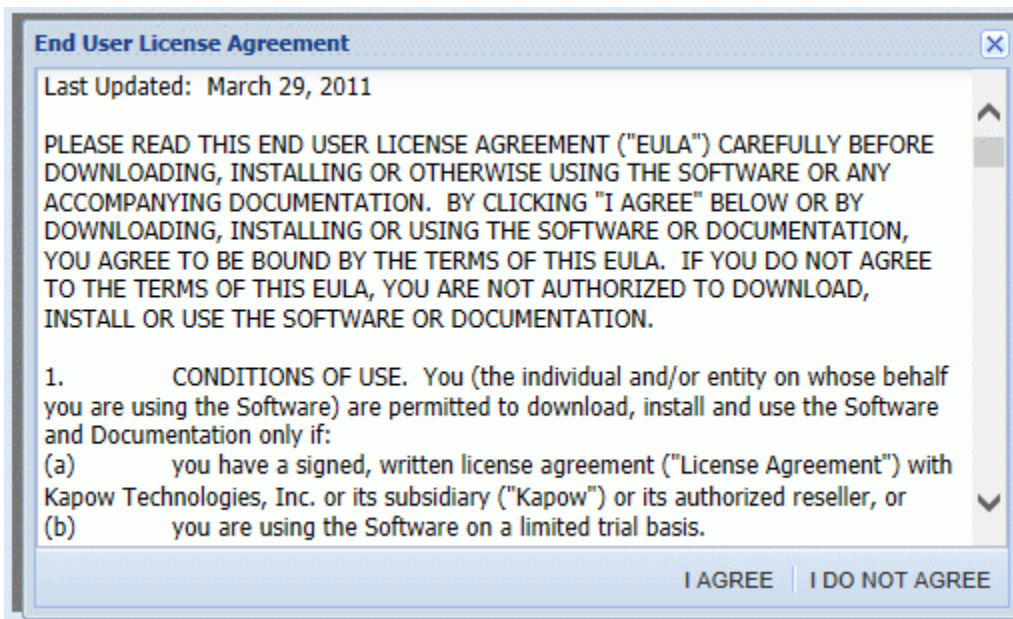
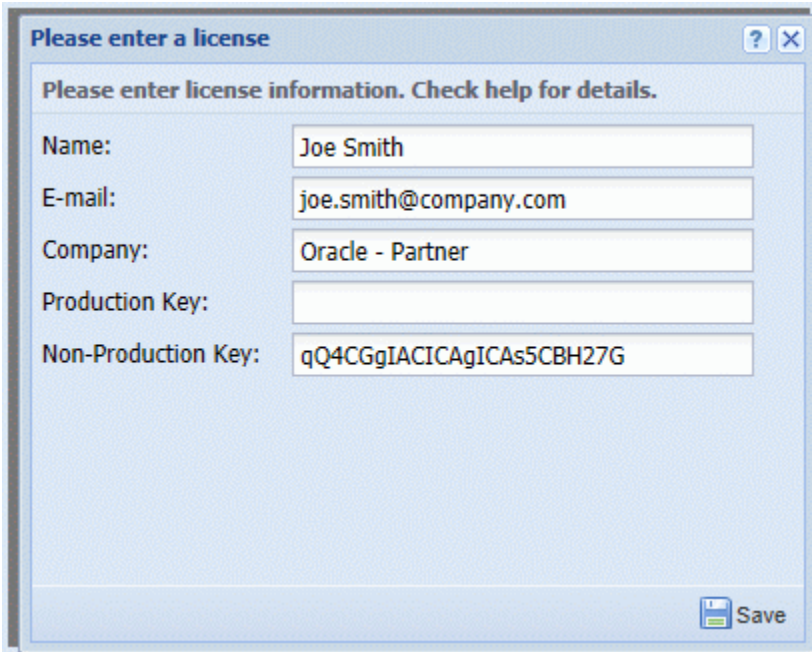   The license information screen displays.

4. Specify Name, E- mail, Company, Production Key, and Non-Production Key and then click **Save**.

    You can find the production key values in `EID_<version>WebAcquisitionToolkit_License.txt`.

The Company value must be set to `Oracle - Partner`, and the Production Key value should be blank.

For example:

5. At the Success screen, click **OK**.

For example:
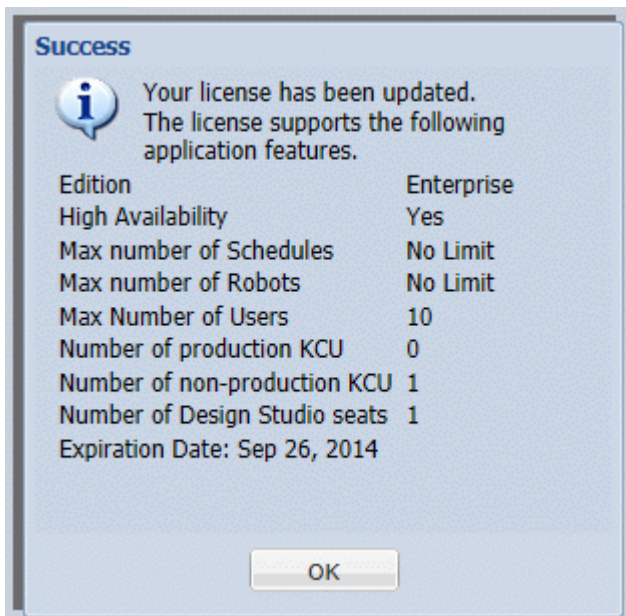
The Oracle Endeca Web Acquisition Toolkit is provided with a provisional license key. Please contact Oracle licensing (licensecodes_ww@oracle.com) to request a permanent key.

Chapter 3

Starting the Web Acquisition Toolkit

This section describes how to start components of the Web Acquisition Toolkit.

Starting Design Studio

What's next?

Starting Design Studio

Remember that on Windows, you run the 64-bit version of Design Studio, but on Linux, you run the 32-bit version.

To start Design Studio:

1. Start the Design Studio.
    -   On Windows, go to **Start>All Programs>Kapow Katalyst 9.2.9 x64>Design Studio 9.2.9**.
    -   On Linux, open a command prompt, change to `../Kapow Katalyst 9.2.9 x32/bin`, and run: `$./DesignStudio`.

    A licensing dialog displays.

2. In the licensing dialog, do the following:
    (a) Select **License Server**.
    (b) Leave **User name** and **Password** empty.
    (c) Click **OK**.

Design Studio displays with an empty project space.

What's next?

Once the WAT is installed and running, you can configure the system for your environment and then build robots to acquire data. For details, see the Kapow Katalyst help available at http://help.kapowsoftware.com/9.2/index.jsp, or see the *Oracle® Endeca Web Acquisition Toolkit Usage Guide*.