



Syncsort DMX

Amazon EC2 Edition

User Guide



© Syncsort® Incorporated, 2016

All rights reserved. This document contains proprietary and confidential material, and is only for use by licensees of DMExpress. This publication may not be reproduced in whole or in part, in any form, except with written permission from Syncsort Incorporated. Syncsort is a registered trademark and DMExpress is a trademark of Syncsort, Incorporated. All other company and product names used herein may be the trademarks of their respective owners.

The accompanying DMExpress program and the related media, documentation, and materials ("Software") are protected by copyright law and international treaties. Unauthorized reproduction or distribution of the Software, or any portion of it, may result in severe civil and criminal penalties, and will be prosecuted to the maximum extent possible under the law.

The Software is a proprietary product of Syncsort Incorporated, but incorporates certain third-party components that are each subject to separate licenses and notice requirements. Note, however, that while these separate licenses cover the respective third-party components, they do not modify or form any part of Syncsort's SLA. Refer to the "Third-party license agreements" topic in the online help for copies of respective third-party license agreements referenced herein.

Table of Contents

1	Introduction	1
2	Support and Charges.....	2
2.1	Support.....	2
2.2	Charges.....	2
3	Launching DMX.....	3
3.1	1-Click Launch	3
3.2	EC2 Console Launch	4
4	Connecting to the DMX Instance.....	6
5	Running the Use Case Accelerators	7
6	Managing your DMX Instance.....	8
6.1	Naming the DMX Instance	8
6.2	Stopping and Restarting the DMX Instance.....	8
6.3	Persistent Storage.....	8
6.3.1	Job and Task Metadata	8
6.3.2	Source and Target Data.....	8
6.4	Terminating your DMX Instance	8
Appendix A	Amazon Web Services.....	A-1

1 Introduction

DMX¹ is Syncsort's ETL software, which can run on Amazon's EC2 platform in the cloud.

DMX enables powerful ETL processing in the cloud without the need for procuring licenses and installing software on your own hardware. With DMX, you can:

- connect to any data source or target, including mainframe data
- develop and maintain your ETL jobs in an easy-to-use graphical user interface
- jump-start your ETL productivity with the included library of use case accelerators

DMX is available in the Amazon Web Services (AWS) Marketplace. For a background on AWS and the component services applicable to DMX, see Appendix A.

¹ Note that the documentation outside of this guide refers to the product as DMExpress.

2 Support and Charges

2.1 Support

Support is available for DMX via the [Syncsort Online Community](#).

Also, check out our [resources page](#), where you will find helpful videos, a link to the community, a link to this document, and the use case accelerators.

2.2 Charges

Syncsort charges are based on instance type size, and accrue only while the DMX instance is running. You can start with a small instance for development, then switch to a larger instance for running production jobs.

In addition, the following Amazon services will accrue charges from the time the instance is launched until it is terminated:

- EBS-backed root volume (defaults to deletion upon termination)
- EC2 instance while running

Any additional EBS volumes you add/attach to your instances, as well as any S3 storage you use, will incur additional charges.

3 Launching DMX

To launch a DMX instance, go to the Syncsort DMX, Amazon EC2 Edition product page in the AWS Marketplace at <https://aws.amazon.com/marketplace/pp/B00IY0KKW6>, and do the following:

1. Review the details on the Syncsort DMX product offering.
2. Click on the **Continue** button at the top of the page to get to the **Launch on EC2** page.
 - a. If you're not already signed in, you will be prompted to sign in or create an AWS account; follow the instructions on screen.
3. Select one of the following tabs for launching:
 - a. **1-Click Launch** – this method is recommended for running with a standard configuration to try the use case accelerators.
 - b. **Launch with EC2 Console** – this method allows for more instance customization when developing and running your own jobs, such as adding an EBS storage volume or setting the Availability Zone to match that of any existing EBS volume you will attach to the instance.

3.1 1-Click Launch

On the **1-Click Launch** tab of the **Launch on EC2** page, review the default settings and click on the arrow to expand each section to make any necessary changes as follows:

1. Select the desired **Version** of DMX.
2. Select the **Region** in which your AWS resources are located to reduce latency and minimize costs.
3. Select the **EC2 Instance Type** that meets your hardware requirements. Review the updated **Monthly Estimate** based on the selected instance type.
4. Select **Create new based on seller settings** to allow AWS to auto-generate a new **Security Group** that includes the required port for DMX (RDP 3389), or choose a previously defined **Security Group** that minimally includes that port.
 - a. For security purposes, it is recommended that the **Source** for the RDP port not be set to **Anywhere**; instead, select **My IP** to automatically detect the public IP address of your local machine, or **Custom IP** to enter another IP address or range. Only the specified IP address(es) will be able to access the instance via RDP.
5. Select an existing **Key Pair** if you already created one. Otherwise, create a new one as described in <http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ec2-key-pairs.html>, then come back to this screen and select the newly created one. You will need to supply the private key file when attempting to log into the instance, so be sure to have it accessible.
6. Once all settings are correct and you've read and agreed to the terms as outlined and linked to on this page, click on the **Accept Terms & Launch with 1-Click**

button at the top of the page. Note that once subscribed to a given AMI, you will not need to accept terms on subsequent launches of that AMI.

7. When first subscribing to an AMI, a pop-up will be displayed indicating that charges will begin to accrue once the instance is deployed upon subscription completion; on subsequent launches, it indicates that the instance is deploying. Instance details are shown, along with a link to the AWS Management Console, where you can start managing your DMX instance.
8. If you plan to develop and run your own jobs in an instance launched via 1-Click, you will need to [create](#) and/or [attach](#) an (existing) EBS volume to your instance to store your job and task files. Be sure that this EBS volume is not set to be deleted on instance termination, and that it is in the same Availability Zone as your instance.

3.2 EC2 Console Launch

On the **Launch with EC2 Console** tab of the **Launch on EC2** page, do the following:

1. If not already subscribed to this AMI, read the terms as outlined and linked to on the page, then click on the **Accept Terms** button at the top right.
2. Select a **Version**, then click on the **Launch with EC2 Console** button corresponding to the **Region** in which your AWS resources are located.

This will take you to the EC2 Console, where you can configure the instance on each successive screen by clicking on the **Next** button at the bottom of the page.

1. On the **Choose Instance Type** page, select the instance type that meets your hardware requirements by clicking on the categories in the left pane and then selecting the desired instance type in the table.
2. On the **Configure Instance Details** page, set the **Availability Zone** in which the instance will be launched to match the Availability Zone of any existing EBS volume that you will attach to the instance. Make any other appropriate selections to meet your requirements.
3. On the **Add Storage** page, click on **Add New Volume** to add a new EBS volume to store your job and task files. Be sure the **Delete on Termination** checkbox is unchecked if you want the EBS volume to persist after the instance is terminated. Note that this will result in additional EBS fees from Amazon. If you have an existing EBS volume that you want to attach instead, follow the instructions [here](#) after the instance is launched.
4. On the **Tag Instance** page, enter a meaningful name in the **Value** field for the **Name** tag to easily identify your ETL instance in the EC2 Console.
5. On the **Configure Security Group** page, choose **Select an existing security group** to select an existing group from the list, or choose **Create a new security group**, and enter an appropriate name and description.
 - a. For security purposes, it is recommended that the **Source** for the required RDP port (3389) not be set to **Anywhere**; instead, select **My IP** to automatically detect the public IP address of your local machine, or **Custom IP** to enter

another IP address or range. Only the specified IP address(es) will be able to access the instance via RDP.

6. Click on the **Review and Launch** button at the bottom right. You will get a security warning about the open RDP port if not restricted to a specific IP/range. Verify your selections on the **Review Instance Launch** page, and click on **Launch** to launch the instance as configured.
7. You will be prompted to select an existing [Key Pair](#) or create a new one. You will need to supply the private key file when attempting to log into the instance, so be sure to have it accessible, and check the box to acknowledge that you do.
8. Click on the **Launch Instances** button. This will launch the DMX instance and charges will begin to accrue.

4 Connecting to the DMX Instance

The Windows EC2 instance may take some time to come up and pass its status checks (displayed on the **Instances** page of the EC2 console). Once the status checks have passed, you can connect to it as follows:

1. From the EC2 Console, select the Windows EC2 instance, and click on the **Connect** button at the top of the page.
2. In the **Connect To Your Instance** screen:
 - a. Click on **Get Password** to retrieve the initial Administrator password; note that this can take up to 30 minutes from the time the Windows instance was started.
 - b. When prompted, paste in the full contents of the private key file used to launch the instance, and click on the **Decrypt Password** button to populate the **Password** field. Keep this window open or copy the password, though you can always come back and retrieve it again if the password hasn't been manually changed.
 - c. Click on the **Download Remote Desktop File** button to download the RDP file, then open it to connect to the Windows EC2 instance. You will get an Unknown publisher warning; click **Connect** to continue.
 - d. When prompted, enter the decrypted password to log in as Administrator. You will get a security certificate warning; click **Yes** to continue.

At this point, you can run the Use Case Accelerators as described in Chapter 5, or start developing and running your own DMX jobs as indicated at the end of that chapter.

When you are done running DMX jobs, you should stop or terminate your instance as described in Chapter 6 until you are ready to start or launch it again.

5 Running the Use Case Accelerators

Syncsort DMX comes with a built-in set of use case accelerators to quickly and easily demonstrate both the development and running of DMX jobs in EC2. These cover a variety of common ETL use cases.

The following use case accelerators are included in the DMX installation at C:\Program Files\DMExpress\Examples\UseCaseAccelerators. See “Use Case Accelerators” in the DMExpress Help for details on each application and how to run it.

- **BatchChangeDataCapture** – Compare two versions of the same data to identify added, deleted, or updated records
- **ChangeDataCaptureWithAttunity** – Change data capture using Attunity Stream to capture transactions
- **DataCleansing** – Standardize and merge data files with different layouts
- **DMExpressJobLogProcessing**
 - **CDCActivityReporting** – Analyze and monitor data changes in CDC job runs using Excel
 - **PerformanceBenchmarking** – Monitor performance of CDC job runs using Excel
- **MessageQueues** – Move transaction information between systems in near-real time.
- **MultiLevelAggregation** – Aggregate data across multiple dimensions
- **RecordPivoting** – Pivot to collapse multiple records into single records for a given entity
- **WebLogProcessing** – Analyze web traffic data
- **WebServiceProcessing** – Read sales data from a web service and produce a flat file of account information.

Note: To run use case accelerators that read from or write to Excel, you will need to install the [Microsoft Office ODBC driver](#). See [KB article 161](#) on how to configure ODBC connectivity on 64-bit machines.

See “Connecting to Amazon S3 from DMExpress” and “Connecting to Amazon Redshift data warehouses/clusters/databases” in the DMExpress Help for details on accessing Amazon S3 and Redshift from DMExpress.

6 Managing your DMX Instance

6.1 Naming the DMX Instance

If you launched via EC2 Console, you had an opportunity to create a Name tag in the **Tag Instance** step of the instance configuration sequence prior to launching. However, instances launched via 1-Click or the launch script will not have the Name field populated in the EC2 Console.

To make it easier to keep track of your instance in the EC2 Console, add a name by clicking on the **Name** field for a given instance and typing in a meaningful name.

6.2 Stopping and Restarting the DMX Instance

The DMX Windows instance can be stopped and restarted as needed.

On the EC2 Console, select the desired instance, click on the **Actions** dropdown at the top of the page, and select **Stop** or **Start** to stop/start the instance.

When an instance is stopped and restarted, it will be assigned a new public DNS name. For the DMX Windows instance, this will mean that you'll need to download a new RDP connection file, but the previously decrypted password will remain unchanged for the life of the instance.

6.3 Persistent Storage

6.3.1 Job and Task Metadata

When you create DMX jobs and tasks on the Windows instance, the job and task files should be saved on a persistent EBS volume attached to the DMX instance.

Be sure that the **Delete on Termination** checkbox is unchecked for the EBS volume on which you save your job/task files so they persist after the life of the instance. When you launch a new DMX instance, you can [attach](#) a previously created EBS volume containing your jobs and tasks to the new instance.

6.3.2 Source and Target Data

When you run DMX jobs, be sure that any source or target data that needs to persist after the life of the instance is stored on a persistent EBS volume (one that is not set to be deleted upon instance termination) or in S3. Reading from and writing to S3 from the Windows instance can be done with DMExpress custom tasks, as demonstrated in [KB article 46692](#).

6.4 Terminating your DMX Instance

Once all running jobs are completed, and any temporary data that you wish to keep has been copied to permanent storage, you can terminate your DMX instance as follows:

1. On the EC2 Console, select the DMX Windows instance, click on the **Actions** dropdown at the top of the page, and select **Terminate** from the bottom of the menu. You will be warned about losing any non-persistent storage:

- a. Click **Cancel** if you need to go back and save files to your persistent storage; any persistent EBS volumes will be automatically detached and can be reattached to a new instance.
 - b. Otherwise, click **Yes, Terminate** to terminate the instance. Be sure that the **Instance State** shows **terminated**; this may take some time.
2. If you want to re-launch after terminating, click on **Your Account** at the top of the AWS Marketplace page, then click on **Manage your software subscriptions**, and then click on the **Launch more software** button for the desired DMX subscription, which will take you back to the corresponding DMX Launch page.

Appendix A Amazon Web Services

[Amazon Web Services \(AWS\)](#) is a cloud computing platform comprised of a set of remote computing services. Following are the primary AWS services/components that pertain to running DMX:

- [Amazon Elastic Compute Cloud \(EC2\)](#) provides resizable computing capacity in the AWS cloud, allowing quick and easy launching of any number of virtual servers, or instances, based on your computing needs.
- An [Amazon Machine Image \(AMI\)](#) is a software configuration template consisting of an operating system, applications, libraries, data, and configuration settings. This template is used to launch an EC2 instance.
- An [EC2 instance](#) is a virtual server that comprises some combination of CPU, memory, storage, and networking capacity, based on the selected instance type.
 - Each EC2 instance is defined by the instance type, the AMI used to launch it, and the configured storage, security, and network access.
- [Amazon Elastic Block Store \(EBS\)](#) provides block level storage volumes that can be attached to EC2 instances for persistent storage.
- [Amazon Simple Storage Service \(S3\)](#) is a scalable, reliable web-based data storage service.
- The [AWS Marketplace](#) is an online store where you find, buy, and launch software that runs on the AWS cloud.
- The [AWS Management Console](#) is the web interface for managing your AWS services. Clicking on EC2 will take you to the EC2 Console.
- [Amazon EC2 Security Groups](#) define a set of rules for network access to a given EC2 instance.

About Syncsort

Syncsort provides data-intensive organizations across the big data continuum with a smarter way to collect and process the ever expanding data avalanche. With thousands of deployments across all major platforms, including mainframe, Syncsort helps customers around the world overcome the architectural limits of today's data integration and Hadoop environments, empowering their organizations to drive better business outcomes, in less time – with fewer resources and lower total cost of ownership. For more information, please visit www.syncsort.com.

Syncsort Inc.

50 Tice Boulevard, Suite 250, Woodcliff Lake, NJ 07677

201.930.8200