Informatica® B2B Data Exchange
10.2

Advanced Performance Tuning Guide

Informatica B2B Data Exchange Advanced Performance Tuning Guide
10.2
April 2017

Publication Date: 2019-10-10

Table of Contents

Preface

The *B2B Data Exchange Advanced Performance Tuning Guide* provides recommendations for improving throughput and scalability of the hardware and software that you use to run B2B Data Exchange and Data Exchange transformations in PowerCenter workflows. The guide is written for advanced B2B Data Exchange and PowerCenter users that want to tune the machine performance. It assumes that you have advanced knowledge of B2B Data Exchange, PowerCenter, and database repository management, as well as the required throughput and event processing load in your organization.

Informatica Resources

Informatica Network

Informatica Network hosts Informatica Global Customer Support, the Informatica Knowledge Base, and other product resources. To access Informatica Network, visit https://network.informatica.com.

As a member, you can:

- Access all of your Informatica resources in one place.
- Search the Knowledge Base for product resources, including documentation, FAQs, and best practices.
- View product availability information.
- Review your support cases.
- Find your local Informatica User Group Network and collaborate with your peers.

As a member, you can:

- Access all of your Informatica resources in one place.
- Search the Knowledge Base for product resources, including documentation, FAQs, and best practices.
- View product availability information.
- Find your local Informatica User Group Network and collaborate with your peers.

Informatica Knowledge Base

Use the Informatica Knowledge Base to search Informatica Network for product resources such as documentation, how-to articles, best practices, and PAMs.

To access the Knowledge Base, visit https://kb.informatica.com. If you have questions, comments, or ideas about the Knowledge Base, contact the Informatica Knowledge Base team at KB_Feedback@informatica.com.

Informatica Documentation

To get the latest documentation for your product, browse the Informatica Knowledge Base at https://kb.informatica.com/_layouts/ProductDocumentation/Page/ProductDocumentSearch.aspx.

If you have questions, comments, or ideas about this documentation, contact the Informatica Documentation team through email at infa_documentation@informatica.com.

Informatica Product Availability Matrixes

Product Availability Matrixes (PAMs) indicate the versions of operating systems, databases, and other types of data sources and targets that a product release supports. If you are an Informatica Network member, you can access PAMs at https://network.informatica.com/community/informatica-network/product-availability-matrices.

Informatica Velocity

Informatica Velocity is a collection of tips and best practices developed by Informatica Professional Services. Developed from the real-world experience of hundreds of data management projects, Informatica Velocity represents the collective knowledge of our consultants who have worked with organizations from around the world to plan, develop, deploy, and maintain successful data management solutions.

If you are an Informatica Network member, you can access Informatica Velocity resources at http://velocity.informatica.com.

If you have questions, comments, or ideas about Informatica Velocity, contact Informatica Professional Services at ips@informatica.com.

Informatica Marketplace

The Informatica Marketplace is a forum where you can find solutions that augment, extend, or enhance your Informatica implementations. By leveraging any of the hundreds of solutions from Informatica developers and partners, you can improve your productivity and speed up time to implementation on your projects. You can access Informatica Marketplace at https://marketplace.informatica.com.

Informatica Global Customer Support

You can contact a Global Support Center by telephone or through Online Support on Informatica Network.

To find your local Informatica Global Customer Support telephone number, visit the Informatica website at the following link:
http://www.informatica.com/us/services-and-training/support-services/global-support-centers.

If you are an Informatica Network member, you can use Online Support at http://network.informatica.com.

CHAPTER 1

Performance Tuning

This chapter includes the following topics:

Performance Tuning Overview

The goal of performance tuning is to optimize throughput, which is the number of events that B2B Data Exchange processes every second.

Performance tuning involves adjustments to the hardware and software configuration in your environment, which assist you to accommodate the different throughput requirements in the organization. The hardware and software adjustments also help to improve the production environment scalability, which is the ability to manage the environment resources in a way that maintains the desired performance when the data volume increases.

This chapter contains recommendations for adjustments to your hardware and software configurations. The recommendations are based on benchmark tests that measure the performance and scalability of B2B Data Exchange and PowerCenter workflows on different databases, operating systems, and software configurations.

Hardware Configuration

You can tune performance and optimize throughput by setting up and configuring your hardware to maximize throughput. Throughput is the number of events that B2B Data Exchange and PowerCenter process every second.

You can modify various hardware aspects to accommodate the required throughput. For example, you can install additional CPU cores, replace slower hard disks with faster and more reliable hard disks, or cluster multiple computers to handle the different product installations instead of a single computer.

CPU Cores

To improve the performance of the machines on which B2B Data Exchange runs, you can incorporate additional CPU cores based on the required throughput. You determine how many CPU cores to use based on the throughput requirements in your organization.

The recommendations in this section address throughput for B2B Data Exchange. For more information about the CPU core recommendations for PowerCenter and Data Transformation, see the *PowerCenter Performance Tuning Guide* and the *Data Transformation Performance Benchmark Guide*.

You can add CPU cores to a single machine, or you can use multiple machines in a cluster configuration. If you add CPU cores to a single machine, each CPU core adds full processing power to the machine overall processing power.

The following table describes the recommended minimum CPU cores that B2B Data Exchange requires for different throughput:

Events per Second	Minimum CPU Cores
100	8
200	16
300	24
400	32
Adding CPU cores to the machine may not accommodate processing of more than 400 events per second.	

Physical Storage and Memory

To improve the performance of the machines on which B2B Data Exchange and PowerCenter run, consider the following recommendations:

- Store the Data Exchange document store on a high-performance storage system, such as Network-Attached Storage (NAS), to increase the processing reliability and performance when PoweCenter processes large documents that B2B Data Exchange passes by reference. If you require higher throughput, do not use a Network File System (NFS).
- Store the database repository on a high-performance file system, such as Storage Area Network (SAN), to improve the data transfer speed and reliability.
- Store the B2B Data Exchange transaction log and the database repository on a high-performance hard drive, such as Solid-State Drive (SSD), to accommodate increased amounts of log entries.
- Run B2B Data Exchange on a machine with at least 8GB of RAM.

Cluster Environment

To improve the performance of the machines on which B2B Data Exchange and PowerCenter run, consider the following recommendations:

- Install the B2B Data Exchange server on a computer cluster instead of on a single machine.
- Store the database repository on a separate machine instead of on the same machine on which B2B Data Exchange and PowerCenter run.

Software Configuration

You can configure settings in B2B Data Exchange, PowerCenter, and the database repository software to tune performance and optimize throughput.

For example, you can pass documents from B2B Data Exchange to PowerCenter by reference, set up batch event creation requests that PowerCenter sends to B2B Data Exchange, or increase the maximum size of the redo log in the database repository.

The following sections describe the software modifications you can make to tune performance and increase throughput. The recommendations are based on performance benchmark tests in B2B Data Exchange and PowerCenter.

B2B Data Exchange Settings and Utilities

To improve the performance of B2B Data Exchange, consider the following recommendations:

- Set the log to INFO mode instead of DEBUG mode. By default, the B2B Data Exchange log generates less log entries in INFO mode. You can set the log to DEBUG mode troubleshoot issues and set it back to INFO mode to improve performance.
- In the dxruntime.sh utility for the Linux, AIX, or Solaris operating system, or the dxruntime.bat utility for the Windows operating system, increase the Java heap size for the B2B Data Exchange server. For more information about the utility, see the *B2B Data Exchange Installation and Configuration Guide*.
- In the server and console copies of the dx-configuration.properties file, you can tune several properties to improve throughput. For more information about the configuration file, see the *B2B Data Exchange Installation and Configuration Guide*.
  - Increase the number of maximum statements in the dx.jdbc.maxPoolSize property to 200.
  - To improve file send performance set the dx.mft.pool.maxSessionsPerEndpoint default setting to allow five files to be sent concurrently.
  - Tune the outbound threads control with the dx.queue.pool.outbound.threads.core property and the dx.queue.pool.outbound.threads.max property default setting to allow up to 20 messages to be processed concurrently.
  - Tune the inbound threads control with the dx.queue.pool.internal.threads.core property and the dx.queue.pool.internal.threads.max property default setting to allow up to 20 messages to be processed concurrently.
- When you define File Receive or Managed File Transfer endpoints for documents larger than 4KB, choose to pass documents by reference instead of by value.

B2B Data Exchange Transformations in PowerCenter Workflows

To improve the performance of PowerCenter workflows that use B2B Data Exchange transformations, consider the following recommendations:

- In the **Config Object** settings of each PowerCenter workflow, increase the number of partitions that PowerCenter uses when it processes files to a minimum of 8 partitions for a throughput of 100 events per second. For example, define 24 partitions for a throughput of 300 events per second.
- In the Generate_File_Path transformation, select the **Do not use a temporary folder in the generated file path** check box to copy the file directly to the Data Exchange document store instead of in a temporary location. By default, PowerCenter copies each file first to a temporary location and then to the Data Exchange document store. If you select this check box, PoweCenter maintains only one copy of each file.

- The Add_Document_To_Event transformation may significantly impact the workflow processing time. If you include the transformation in the workflow, you can minimize the performance impact in the following ways:
  - Add a document to an event only when a processing error occurs.
  - For a parent event with multiple child events, aggregate the documents to a single file and attach it to the parent event.
  - If the average row size in the document is smaller than 4KB, pass the document by value. This prevents PowerCenter from writing too many documents to the Data Exchange Document Store.
  - If the average row size in the document is larger than 4KB, pass the document by reference with the Generate_File_Path transformation.
- On an SQL server with the Windows operating system:
  - Use the Profile_Parameter transformation once in the beginning of the workflow to retrieve all the parameters instead of multiple times in the workflow to retrieve individual parameters.
  - The Set function of the Event_Attributes transformation may drastically reduce performence. If you need to use the function, you can set all of the attributes once for the parent event.
  - The Initiate_Correlation and Complete_Correlation transformations may drastically reduce performance.
- To optimize API calls to B2B Data Exchange from PowerCenter, set the PowerCenter Integration Service environment variable **DX_JCT_MAX_ROWS_IN_BATCH** to 10.

The recommendations in this section are based on Data Exchange transformation benchmark test results. For more information, see " Data Exchange Transformation Benchmark Tests" on page 13.

Database Repository

To improve the database repository performance, consider the following recommendations:

- Increase the number of concurrent sessions that the database repository can run.

  Note: Only the database administrator can perform this action. For more information, consult your database administrator.
- For Oracle servers, increase the size and the number of instances of the redo log. Each time the log file reaches the maximum size, the database creates a new log file. Increase the log file size to reduce the number of log files that the database creates during run time. Increasing the number of instances of the log file enables load balancing and increases the availability of the log files. The benchmark tests that are described in this document used three instances of the redo log with a file size of 2 GB for each instance .
- When the number of events in the event table exceeds three million, archive the events by using Data Archive. Otherwise, the B2B Data Exchange console can run slower or the database queries might fail. For more information about how to archive the events, see the *B2B Data Exchange Administrator Guide*.

B2B Data Exchange JMS Broker

To improve the performance of the B2B Data Exchange JMS broker, consider the following recommendations:

- If you split messages with one PowerCenter workflow process the messages with another PowerCenter, send the messages between the PowerCenter workflows in batches of at least 20 messages instead of sending each message separately.
- Disable persistency when you send messages between PowerCenter workflows.
- In the activemq.xml file, increase the memory limit and the storage size to match the JVM size. To do this, enter a higher value in the **memoryUsage limit** and the **storeUsage limit** parameters. For example, if the

average message size is 10 KB, and your organization requires the queue to hold up to 1,000 messages, set the memory limit to at least 10 MB.

- Tune the values for the failover URL in the JNDI connection. For example, to change the timeout, maximum attempts, and initial reconnect relay, you might use the following settings:

  ```
  jms.closeTimeout=30000&timeout=30000&maxReconnectAttempts=5&initialReconnectDelay=1000
  ```

- If you expect to create hundreds of AMQ connections either from B2B Data Exchange or from PowerCenter, set the AMQ broker to use the NIO configuration. Locate the following parameter settings:

  ```
  <transportConnector name="openwire" uri="tcp://0.0.0.0:18616"...
  ```

  Change to the following parameter settings:

  ```
  <transportConnector name="openwire1" uri="nio://0.0.0.0:18616?
  useQueueForAccept=false"/>
  ```

Configuration for Thousands of Endpoints

To improve the performance of the B2B Data Exchange endpoints when working with thousands of endpoints, configure the following settings:

- Improve the file input processing speed.
  - Use a fast file system. Ensure that the file input folders reside on computers with a fast file system.
  - Configure endpoints to multiple folders to avoid any individual folders that contain thousands of files.
- Adjust the file endpoint settings.
  - Tune the value for time interval to scan a folder for new files. Change the dx.endpoint.file.scan.interval system setting from the default value of **5** to the maximum number of seconds that the customer accepts, for example **120**.
  - Tune the value for the number of files expected each time that B2B Data Exchange scans a folder for new files. If more files are sent during peak periods than this value, the files wait for subsequent scans, which improves performance. Set the dx.endpoint.file.scan.max parameter to the expected number, for example **25**, instead of the default value of **250**.
  - Tune the value for the file backlog to a lower number. The backlog is the set of files that DX buffers during processing. Set the dx.endpoint.file.backlog.max parameter to a lower value, for example **250**, instead of the default value of **5000**. To change the expiration time for the backlog, set the endpoint.file.backlog.timeout parameter, whose default value is **10**.
- Schedule the MFT endpoints that act as inbound client to collect files at an appropriate frequency to your communication needs. A schedule that is too short creates many requests to the remote server that fetch few, small files. A schedule that is too long might create large processing peaks.

CHAPTER 2

Performance Benchmark Tests

This chapter includes the following topics:

Benchmark Tests Overview

A benchmark tests measures the performance and scalability of B2B Data Exchange and PowerCenter workflows that use Data Exchange transformations.

The test results help determine which hardware and software configuration adjustments you can make to tune the performance and optimize throughput in B2B Data Exchange and PowerCenter.

This chapter includes the following benchmark test types:

- Performance benchmark. Tests a single PowerCenter workflow that uses Data Transformation parsers and Data Exchange transformations under different CPU core combinations on single and multiple machines.
- Data Exchange transformation benchmark. Tests individual transformations with different document sizes under different conditions, such as passing by value or by reference, on different server vendors and operating systems.
- B2B Data Exchange JMS Broker benchmark. Tests the performance of the B2B Data Exchange JMS broker with different transaction volumes.

Transformation Benchmark Test Machine Specifications

The transformation benchmark tests were performed on single machines and on all machines in a cluster configuration. The test configuration contained one Oracle database machine and three identical application machines with B2B Data Exchange, PowerCenter, and Data Transformation.

The following table lists the test machine specifications for all test machines:

Category	Specification
Server Model	HP ProLiant BL460c G7 ServerOracle database
CPU Cores	2 X 8 Cores Westmere-EX 9350 1.73GHz
RAM	48GB
Local Disk	2 X 300GB 10,000RPM SAS hard disk
Shared Storage	HP P2000 disk array with 49 X 146GB 15,000RPM SAS hard disks
Operating System	Red Hat Linux 5.6 64-bit

Transformation Performance Benchmark Test

The Transformation performance benchmark test determined the optimal hardware configuration and PowerCenter partitions to use when you want B2B Data Exchange to process a specific throughput of events per second with PowerCenter and Data Transformation.

The test was performed on a single machine with an 8 CPU core processing power, and in a cluster configuration with a combined 16 CPU core processing power. B2B Data Exchange sent documents with different file sizes to PowerCenter, which processed the documents by value with the X12 Parser transformation.

For the single machine, PowerCenter was configured to use 32 partitions. For the machine cluster, each PowerCenter installation was configured to use 16 partitions, with a total of 32 partitions.

The following table describes the performance benchmark test results:

File Size	Number of Machines	Events per Second	PowerCenter Partitions
2KB	1	385	32 partitions
2KB	**2**	**500**	**2 x 16 partitions**
6.5KB	1	189	32 partitions
6.5KB	2	367	2 x 16 partitions
16KB	1	98	32 partitions
16KB	2	187	2 x 16 partitions

The table illustrates that the machines reached maximum throughput when both machines in a cluster use 32 PowerCenter partitions to process a 2K file size document with no processing errors.

PowerCenter Workflow Components

The PowerCenter workflow that the performance benchmark test used includes multiple Data Exchange transformations and Data Transformation parsers. The workflow is provided with this document.

The following table lists the order of the workflow components:

Order	Component
1	JMS Input
2	DX_Event_Details
3	DT_Streamer
4	DX_Create_Event **Note:** PowerCenter divided the file between the PowerCenter partitions and created an event for each partition.
5	DX_Add_Document_To_Event
6	DT_X12_Parser
7	DX_Add_Document_To_Event
8	DX_Set_Event_Attribute
9	DX_Event_Details
10	File Output

Data Exchange Transformation Benchmark Tests

The Data Exchange transformation benchmark tests help to determine the performance of individual transformations under different conditions, such as row size, number of documents, and server vendor.

The following sections include tests that were performed on transformations that can impact overall performance. Transformations that do not create substantial load on the workflow are not included.

For more information about recommendations for performance tuning of specific Data Exchange transformations and PowerCenter workflows, see "B2B Data Exchange Transformations in PowerCenter Workflows" on page 8.

DX_Add_Document_To_Event Transformation

The following table lists the transformation benchmark test results for an Oracle database with the Linux operating system:

Processing Type	Files	Rows	Events	Row Size	Minutes	Seconds	Throughput
By Reference to Different events	10	10000	100000	1KB	3	53	429
By Reference to Different events	1	100000	100000	1KB	2	30	667
By Value to Different Events	10	10000	100000	1KB	1	48	926
By Value to Different Events	1	100000	100000	1KB	0	37	2667
By Value to Different Events	10	10000	100000	10KB	3	5	540
By Value to Different Events	1	100000	100000	10KB	2	57	565
By Value to Different Events	10	2000	20000	100KB	3	41	90
By Value to Different Events	1	20000	20000	100KB	3	10	105
By Value to Different Events	10	300	3000	1MB	9	50	5
By Value to Different Events	1	3000	3000	1MB	7	47	6
By Value to Same Event	10	300	3000	1MB	5	43	9
By Value to Same Event	1	3000	3000	1MB	6	25	8

The following table lists the transformation benchmark test results for a Microsoft SQL Server database with the Windows operating system:

Processing Type	Files	Rows	Events	Row Size	Minutes	Seconds	Throughput
By Reference to Different events	10	10000	100000	1KB	4	55	339
By Reference to Different events	1	100000	100000	1KB	2	50	588
By Value to Different Events	10	10000	100000	1KB	0	49	2041
By Value to Different Events	1	100000	100000	1KB	0	31	3226
By Value to Different Events	10	10000	100000	10KB	1	38	1020
By Value to Different Events	1	100000	100000	10KB	1	33	1075
By Value to Different Events	10	2000	20000	100KB	2	24	138
By Value to Different Events	1	20000	20000	100KB	2	32	131
By Value to Different Events	10	300	3000	1MB	7	40	7
By Value to Different Events	1	3000	3000	1MB	2	0	25

Processing Type	Files	Rows	Events	Row Size	Minutes	Seconds	Throughput
By Value to Same Event	10	300	3000	1MB	4	8	12
By Value to Same Event	1	3000	3000	1MB	5	40	9

DX_Create_Event Transformation

The following table lists the transformation benchmark test results for an Oracle database with the Linux operating system:

Files	Rows	Events	Minutes	Seconds	Throughput
10	100000	1000000	5	20	3120
1	1000000	1000000	5	22	3105

The following table lists the transformation benchmark test results for a Microsoft SQL Server database with the Windows operating system:

Files	Rows	Events	Minutes	Seconds	Throughput
10	100000	1000000	3	47	4405
1	1000000	1000000	3	48	4386

DX_Event_Attribute Transformation

The following table lists the transformation benchmark test results for an Oracle database with the Linux operating system:

Operation	Attributes	Files	Rows	Events	Row Size	Minutes	Seconds	Throughput
Get Event Attribute from Different Events	1	10	10000	100000	10KB	0	45	2222
Get Event Attribute from Different Events	10	10	1000	100000	10KB	0	18	5556
Set Event Attribute to Same Event	1	10	10000	100000	10KB	0	50	2000
Set Event Attribute to Same Event	10	10	1000	100000	10KB	0	7	14286

Operation	Attributes	Files	Rows	Events	Row Size	Minutes	Seconds	Throughput
Set Event Attribute to Different Events	1	10	10000	100000	10KB	0	54	1852
Set Event Attribute to Different Events	10	10	1000	100000	10KB	0	28	3571

The following table lists the transformation benchmark test results for a Microsoft SQL Server database with the Windows operating system:

Operation	Attributes	Files	Rows	Events	Row Size	Minutes	Seconds	Throughput
Get Event Attribute from Different Events	1	10	10000	100000	10KB	0	11	9091
Get Event Attribute from Different Events	10	10	1000	100000	10KB	0	9	11111
Set Event Attribute to Same Event	1	10	10000	100000	10KB	0	48	2083
Set Event Attribute to Same Event	10	10	1000	100000	10KB	0	23	4348
Set Event Attribute to Different Events	1	10	10000	100000	10KB	3	19	502
Set Event Attribute to Different Events	10	10	1000	100000	10KB	2	40	625

DX_Event_Details Transformation

The following table lists the transformation benchmark test results for an Oracle database with the Linux operating system:

Operation	Files	Rows	Events	Row Size	Minutes	Seconds	Throughput
Get Event Details	10	100000	1000000	10KB	1	50	9091
Set Event Details	10	100000	1000000	10KB	5	30	3030

The following table lists the transformation benchmark test results for a Microsoft SQL Server database with the Windows operating system:

Operation	Files	Rows	Events	Row Size	Minutes	Seconds	Throughput
Get Event Details	10	100000	1000000	10KB	2	29	6711
Set Event Details	10	100000	1000000	10KB	4	7	4048

DX_Generate_Temporary_File Transformation

The following table lists the transformation benchmark test results for an Oracle database with the Linux operating system:

'Do not use a temporary folder in the generated file path' check box	Files	Rows	Events	Row Size	Minutes	Seconds	Throughput
Selected	1	10000	100000	10KB	2	7	787
Cleared	1	10000	100000	10KB	3	22	493

DX_Increment_Profile Transformation

The following table lists the transformation benchmark test results:

Database	Operating System	Files	Rows	Events	Row Size	Minutes	Seconds	Throughput
Oracle	Linux	1	100000	100000	10KB	1	10	1428
Microsoft SQL Server	Windows	1	100000	100000	10KB	1	5	1538

DX_Initiate_Correlation and DX_Complete_Correlation Transformations

The following table lists the transformation benchmark test results:

Database	Operating System	Files	Rows	Events	Row Size	Minutes	Seconds	Throughput
Oracle	Linux	1	5000	5000	10KB	0	12	416
Microsoft SQL Server	Windows	1	5000	5000	10KB	3	14	26

DX_Profile_Parameter Transformation

The following table lists the transformation benchmark test results:

Database	Operating System	Files	Rows	Events	Row Size	Minutes	Seconds	Throughput
Oracle	Linux	1	100000	100000	10KB	1	22	1219
Microsoft SQL Server	Windows	1	100000	100000	10KB	2	15	741

B2B Data Exchange JMS Broker Benchmark Test

The B2B Data Exchange JMS Broker benchmark test helps to determine the optimal hardware configuration to use when you want B2B Data Exchange to process a specific throughput of events per second with the B2B Data Exchange JMS Broker.

The following table describes the benchmark test results:

Number of Messages in the Transaction	Actual Throughput
1	~650
5	~2200
10	~3000
20	~4500
100	- Producer ~4900 - Consumer ~4200

INDEX

B

P