



Tungsten Search and Matching Server Getting Started Guide

Version: 6.10.0

Date: 2024-09-17

TUNGSTEN
AUTOMATION

© 2024 Tungsten Automation. All rights reserved.

Tungsten and Tungsten Automation are trademarks of Tungsten Automation Corporation, registered in the U.S. and/or other countries. All other trademarks are the property of their respective owners. No part of this publication may be reproduced, stored, or transmitted in any form without the prior written permission of Tungsten Automation.

Table of Contents

Preface.....	4
Related documentation.....	4
Training.....	4
Getting help for Tungsten products.....	5
Chapter 1: Concept and principles.....	6
Fuzzy searches.....	7
Security aspects.....	7
Security aspects for Administration.....	7
Security aspects for searches.....	9
Security aspects for service accounts.....	11
Security aspects for data storage.....	11
Performance considerations.....	11
Server browser service.....	12
Chapter 2: Search and Matching Server.....	13
Monitor the performance.....	13
Server performance.....	13
Chapter 3: The Load Balancer.....	16
Provide granted access for load balancing.....	19
Chapter 4: Tungsten Search and Matching Server Administration.....	20
Access the Tungsten Search and Matching Server.....	20
Set up a fuzzy database.....	21
Configure and maintain existing database.....	22
Access rights for a Microsoft SQL Server database.....	22
Configure a Microsoft SQL Server connection.....	23
Access rights for an Oracle database.....	23
Configure an Oracle database connection.....	24
Configure an ODBC database connection.....	24
Considerations for fuzzy database creation.....	25
Rename a database.....	29
Activate or deactivate a database.....	30
Delete a database.....	30
Export a database.....	30
Import a database.....	31
Configure automatic database update.....	32
Set up load balancing.....	32

Preface

This guide contains information about installing Tungsten Search and Matching Server and assumes that you have a thorough understanding of Windows standards, applications, and interfaces. Also, since Tungsten Search and Matching Server can be installed alongside Tungsten Transformation Modules and Tungsten Capture or Tungsten TotalAgility, it is necessary to fully understand those applications as well.

This guide is for solution integrators who are installing Tungsten Search and Matching Server or who need a description of the installation procedures and requirements.

If you downloaded Tungsten Search and Matching Server from the Tungsten Automation Electronic Delivery site, you can extract the files from ISO file. Refer to the instruction on the Site. Or, you may have received Tungsten Search and Matching Server as a package that includes electronic media. In this guide, we refer to both the extracted files and the electronic media as the product files.

Related documentation

In addition to this *Tungsten Search and Matching Server Installation Guide*, the following documentation is available.

***Tungsten Search and Matching Server* Getting Started Guide**

The *Tungsten Search and Matching Server Getting Started Guide* provides information on how to start using Tungsten Search and Matching Server and the Administration tool.

Help for Tungsten Search and Matching Server

The Help provides detailed information about the concept and principles of this remote fuzzy database server and how to use Tungsten Search and Matching Server Administration for setting up and maintaining fuzzy indexes.

The product document set for Tungsten Search and Matching Server is available from <https://docshield.tungstenautomation.com/Portal/Products/TSMS/6.10.0-8fsknjo9y4/TSMS.html>


Training

Tungsten Automation offers both classroom and computer-based training that help you make the most of your Tungsten Automation solution. Visit the Tungsten Automation Web site at www.tungstenautomation.com for complete details about the available training options and schedules.

Getting help for Tungsten products

The [Tungsten Knowledge Base](#) repository contains articles that are updated on a regular basis to keep you informed about Tungsten products. We encourage you to use the Knowledge Base to obtain answers to your product questions.

To access the Tungsten Knowledge Base, go to the [Tungsten website](#) and select **Support** on the home page.

 The Tungsten Knowledge Base is optimized for use with Google Chrome, Mozilla Firefox or Microsoft Edge.

The Tungsten Knowledge Base provides:

- Powerful search capabilities to help you quickly locate the information you need.
Type your search terms or phrase into the **Search** box, and then click the search icon.
- Product information, configuration details and documentation, including release news.
Scroll through the Tungsten Knowledge Base home page to locate a product family. Then click a product family name to view a list of related articles. Please note that some product families require a valid Tungsten Portal login to view related articles.
- Access to the Tungsten Customer Portal (for eligible customers).
Click the **Customer Support** link at the top of the page, and then click **Log in to the Customer Portal**.
- Access to the Tungsten Partner Portal (for eligible partners).
Click the **Partner Support** link at the top of the page, and then click **Log in to the Partner Portal**.
- Access to Tungsten support commitments, lifecycle policies, electronic fulfillment details, and self-service tools.
Scroll to the **General Support** section, click **Support Details**, and then select the appropriate tab.

Chapter 1

Concept and principles

Tungsten Search and Matching Server handles search requests from various Tungsten Automation applications on large volumes of data located at a remote site. For example, customer or supplier databases.


In contrast to other database servers, such as Microsoft SQL or Oracle server, the Tungsten Search and Matching Server allows performing non-exact (fuzzy) searches. This means that search results are returned even if the search string contains spelling mistakes or does not match the exact value that is stored in the database. This is especially important for address data where a street name can be stored in different variations (for example, "Canyon Avenue" or "Canyon Ave" or "Canyon Av."). It provides an independent fuzzy search functionality based on Microsoft SQL, Oracle, ODBC database tables and views, or for delimited text files (csv) as input that is only referenced through the Tungsten Search and Matching Server.

Performing fuzzy searches for data that is located at a remote site reduces the startup time at client site as the databases are configured only once on the server and then can be accessed from different clients. At the client site, no extra loading time or memory is needed when accessing data at a remote site. In addition, maintenance costs are reduced for databases that change regularly, as the Tungsten Search and Matching Server allows automatic update scheduling so that the client is provided with the most accurate data automatically without needing to manually update the local fuzzy database on each client.

To be able to cope with large enterprise databases Tungsten Search and Matching Server is a server application that runs on 64-bit Microsoft Windows operating systems. It is designed to use multi-threading and support multi-core environment for quick access and fast response time. If the capacity limit for a [single standalone Search and Matching Server](#) is reached, you can use a [Load Balancer](#) to distribute the computational workload across multiple computers by accessing a set of different Search and Matching Servers. For example, due to an increased number of clients that send search requests, a Load Balancer is needed.

Various Tungsten Search and Matching Server applications can send search requests for one or more fuzzy databases that are maintained on the same server. The Tungsten Automation applications can send search requests for one or more fuzzy databases that are maintained on the same server. The Tungsten Search and Matching Server handles all search requests most efficiently (for example, by prioritizing search requests from user interactive modules, such as Tungsten Transformation Modules Validation, over requests from Tungsten Transformation Modules Server for results for a Database Locator) and can perform database updates in parallel to the search requests.

To provide efficient search and automatic updates, the Tungsten Search and Matching Server consists of several Microsoft Windows services and a [configuration tool](#), Tungsten Search and Matching Server Administration, that is used to set up new and maintain existing fuzzy databases, and to configure automatic update and load balancing.

 Tungsten Search and Matching Server and the Administration tool both run on 64-bit operating systems only.

Fuzzy searches

The Tungsten Search and Matching Server provides a fuzzy index for structured data that is imported from a relational database. The fuzzy index is created on the server based on a relational database and provides a very flexible and efficient way to "search and match" for specific data during document transformation with Tungsten Transformation Modules or Tungsten TotalAgility. The fuzzy index allows non-exact searching at the character level and word level. The non-exact search at the character level is typically required in case of OCR errors, where single characters of a word are missing or wrong.

The non-exact search at the word level allows using the fuzzy index in a search dialog, where the user only needs to type in one or two words to find a specific record. This allows a user to run a fuzzy search in a large database with several million records on any column that was included in the fuzzy index.

It is also possible to search for a specific record using multiple words. The result is then the record with the best match where most of the words in the query can be matched with the record. For example, this is used in the Database Locator in a project where the complete content of a page is used as the query. In this specific case, the query contains many more words than in the record that is being searched. A typical use case for this scenario are invoices that are matched against a vendor database to find the vendor that sends the invoice or mailroom documents that are matched against a customer database to identify the customer that wrote the letter or who is referenced on the letter.

Finally, it is also possible to search for a record and require that the query is identical to the matched record. For example, this feature is used by the Database Evaluator and allows matching several fields with potential OCR errors against a set of columns in a fuzzy index.

Security aspects

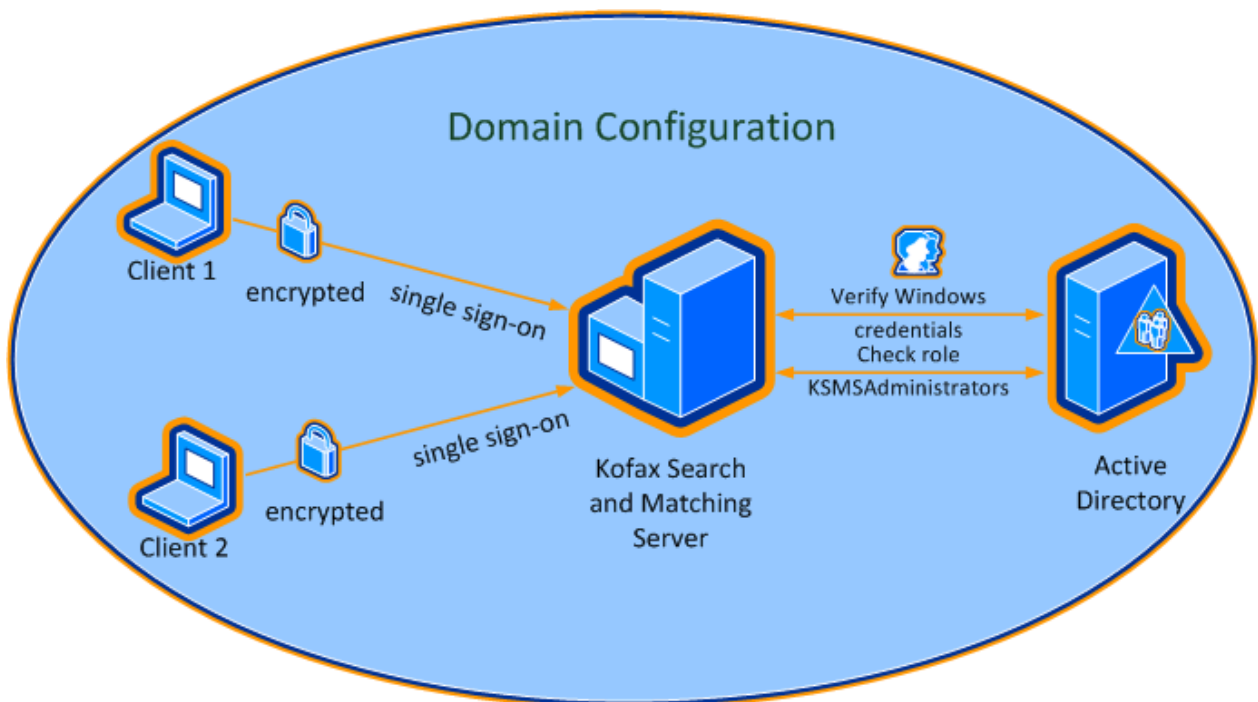
The Tungsten Search and Matching Server is a data driven application that typically is used to work with sensitive data like customer or supplier databases. In addition, the internal configuration files may contain connection information to external databases including user name and password combinations. All information is sent over the network during administration and stored on disk of the computer where the Tungsten Search and Matching Server and the Load Balancer is installed. The following topics explain how the sensitive data is protected during communication and storage.

Security aspects for Administration

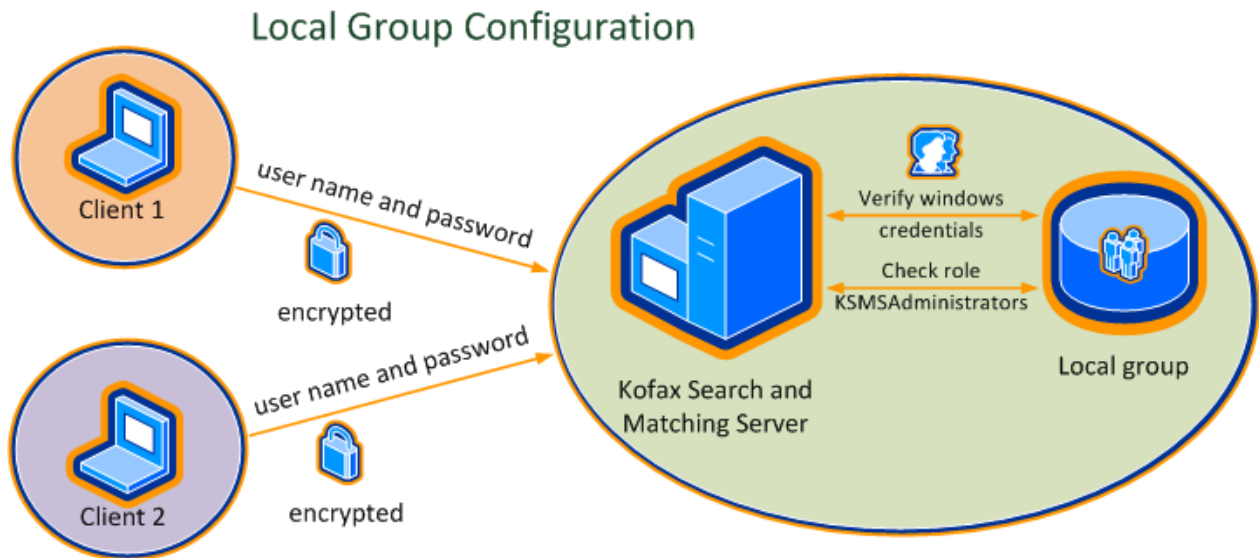
The administration of the Tungsten Search and Matching Server is restricted to Windows users that are members of the "TSMS Administrators" user group configured during installation. This user group can either be a domain or a local user group. A local user group is created during installation if it does not yet exist. Optionally, for a local group installation the installer can add the current user

to the local user group. A domain user group must always be created by the domain administrator. For both cases the system or domain administrator is responsible for adding and removing users to the TSMS Administrators user group to grant or deny access to the Tungsten Search and Matching Server configuration.

i If the TSMS Administrators user group exists as a local group and a domain group at the same time on a computer, the access restriction is only checked against the local group. Any membership in the domain user group is not relevant in that case.



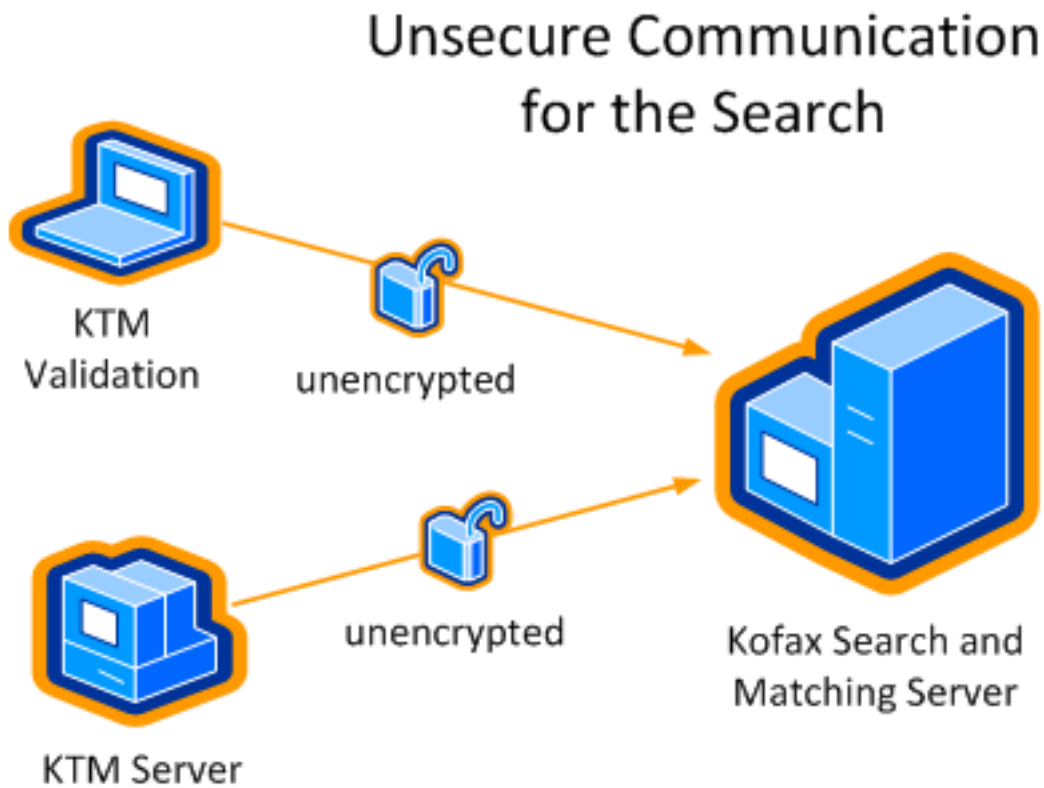
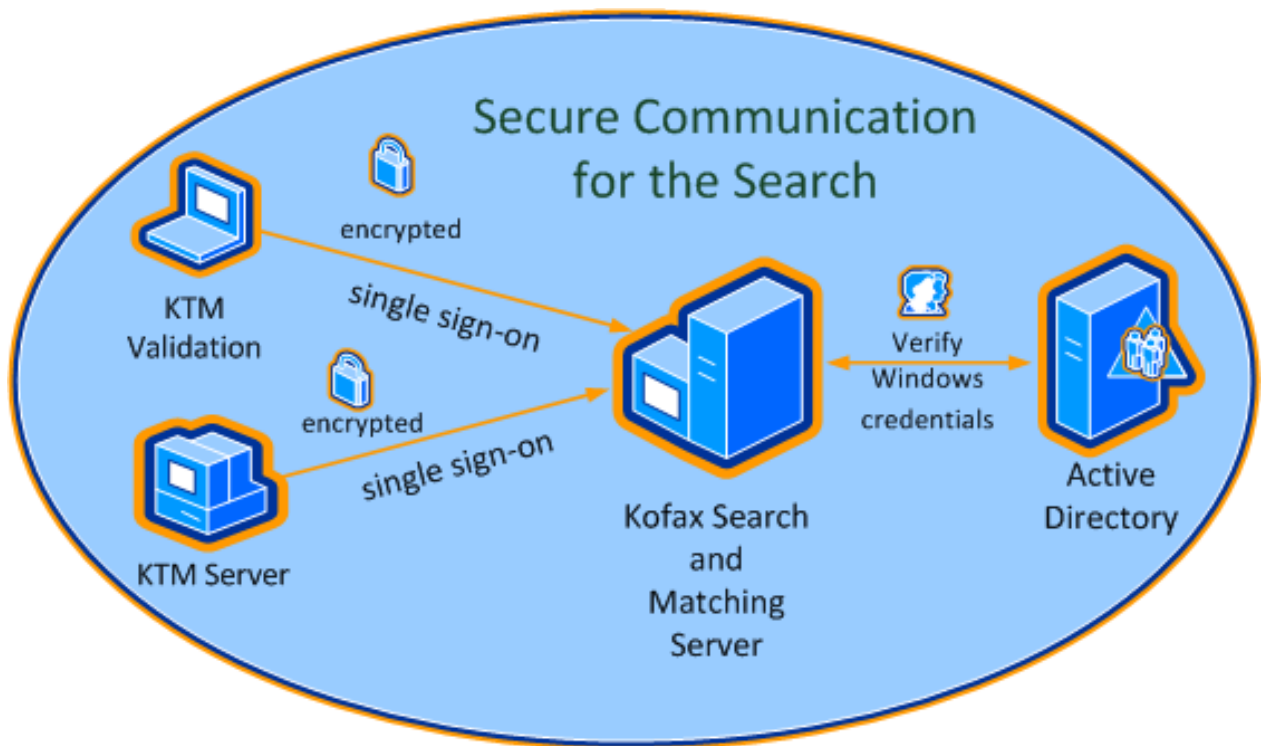
The communication between the Administration application and the Tungsten Search and Matching Server is always encrypted. The Tungsten Search and Matching Server uses Windows Communication Foundation (WCF) in order to encrypt the communication, which is based on the Transmission Control Protocol (TCP). The encryption is provided by using the Transport Layer Security (TLS) over TCP. The TLS implementation is provided by the operating system.



The Administration application can connect to the Tungsten Search and Matching Server using single sign-on if the current user is a member of the TSMS Administrators user group. In this case the user does not need to type the password, instead the Administration application sends the current user token directly to the Tungsten Search and Matching Server. If the current user is not a member of the TSMS Administrators user group it is also possible to type in a different user name and password. In this case the user name and password are sent to the Tungsten Search and Matching Server using the encryption described above.

Security aspects for searches

Search requests from a client application such as Project Builder in Tungsten Transformation Modules or Transformation Designer in Tungsten TotalAgility, or any runtime module such as Tungsten Transformation Modules Server or Validation never require an extra user name and password. Technically, the encryption is the same as for the Administration using the Transport Layer Security over TCP.



Security aspects for service accounts

All services that are installed as part of the Tungsten Search and Matching Server are by default executed with the built-in "Network Service" user account. It may be necessary to run the services with a domain user account so that the access to network shares or a Microsoft SQL Server can be accomplished using the domain user permissions. When you change the service user you have to ensure that the [read-write permissions](#) of the database related directories are adjusted accordingly. In a cluster environment the load balancer and the server nodes must run under the same service account.

Security aspects for data storage

During the installation you can select the directory for data storage. Tungsten Search and Matching Server maintains three subdirectories for storing database related files. The "Databases" directory contains the necessary files for all configured fuzzy indexes. The "Upload" directory contains database text files that were uploaded from users through the Administration application. The "Log" folder contains log files and its access is not restricted.

By default, the access to the "Databases" and the "Uploads" subdirectories is restricted to the "NETWORK SERVICE" user. It is sufficient to grant the read and write access to these directories to only the service account that runs the Tungsten Search and Matching Server. If the user account that is running the Tungsten Search and Matching Server is changed to a different user then the access permissions to these two directories must be adjusted accordingly.

Performance considerations

The performance of a Tungsten Search and Matching Server can be measured in two ways. For a single specific query, the time to answer the query can be measured. For the entire server, it is also possible to measure the throughput in queries per second, if there are enough queries. The fastest response time for a single query depends on the size of the imported database and the complexity of the query. The size of the imported database is generally based on the number of records and the number of columns. The complexity of the query is determined by the number of words in the query, but also highly depends on the number of words inside the query that matches a record. The match score calculation also becomes more complex the more words are matched in the same record. In particular, searching for words that are displayed very frequently in the database can slow down the response time, because a lot of records needs to be evaluated to calculate the score.

The matching algorithm in the Tungsten Search and Matching Server fully supports multi-core systems. An increasing number of CPUs inside a server system can reduce the minimal response time to a certain extent, especially, for fuzzy indexes from larger databases. There may be no impact for databases with less than 1 million records if no further parallel processing is possible. In any case an increasing number of CPUs increases the maximum throughput of search queries that can be processed per second.

If the maximum throughput in a given hardware environment is not adequate, it is possible to build a load balancing cluster to increase the overall throughput for search queries.

A load balancing cluster is a set of multiple computers that are linked together and share the computational workload, but are displayed as a single virtual computer. The overall throughput of a load balancing cluster scales almost linearly as computers are added to the cluster. If the maximum throughput for a single standalone Tungsten Search and Matching Server is known and the required throughput for the complete installation is also known, the number of computers for a load balancing cluster can be calculated.

You can monitor the performance of a load balancer by displaying the [performance counters](#) of each Search and Matching Server that is a member of the cluster.

Server browser service

You can use the Server Browser service to display the Search and Matching Servers and Load Balancers that are available in your network to a client. The Server Browser is a Windows service that listens on port 21200 for incoming browsing requests. The communication for server discovery relies on network broadcasts to locate available servers, which means that it functions best when servers are located on the same subnet as the clients. It is possible, though not always easy, to use the server browsing across subnets. Routers need to be configured to forward directed broadcasts on large networks. For more details see appropriate technical literature.

To find available Search and Matching Servers and Load Balancers in your network you can use the Server Discovery button in the Administration and in client applications, such as Project Builder once the Server Browser service is running. When you select a server from the provided list the computer name and its configured port number are returned so that you can use this information to establish the connection.

In case you do not want the Tungsten Search and Matching Server services that are running on your computer to be visible on the network, you can stop the Server Browser service. This has no impact on already configured connections to the Tungsten Search and Matching Server.

Chapter 2

Search and Matching Server


The Search and Matching Server is used to handle simultaneous search requests from one or more client applications and provide search results in a most efficient way. Multiple users can access the Search and Matching Server non-concurrently to [configure](#) a new and [maintain](#) existing fuzzy databases via the Tungsten Search and Matching Server Administration.

A Fuzzy Database (also called Fuzzy Index) allows fast searching and matching of text values using non-exact (fuzzy) string matching either in record sets or in specific columns. The fuzzy index is created either from a CSV file or from a single table or view of a relational database that is stored as a delimited text file in CSV format, which can either be a Microsoft SQL Server, an Oracle Server and an ODBC database.

Monitor the performance

With the installation of Tungsten Search and Matching Server new counters are added to the Windows monitoring system that allow monitoring the number and speed of various server operations.

1. From the **Start** menu, click **Control Panel**,
2. Click **Administrative Tools**, and then click **Performance** to display the Microsoft Windows performance counter.
This sequence may differ depending on your operating system.
3. Right-click the graph, and then click **Add Counters** on the shortcut menu. The **Add Counters** window is displayed.
4. Select **Tungsten Search and Matching Server** from the Performance object list to choose the performance counters you want to monitor.
5. Select a counter and click **Add** to include the selected counter to the performance monitor. If wanted, you can select multiple counters by holding the Shift key.

 You can also use Tungsten Monitor to display the performance counters. Refer to the Tungsten Monitor documentation for more information about how to use the wizard in order to view these counters.

You may change the appearance of the graph by clicking **Properties** on the shortcut menu.

Server performance

The following lists show the items that you can track.

Tungsten Search and Matching Server

The Tungsten Search and Matching Server has counters that monitor the number and speed of various service operations.

Search Request Count

This is the total number of searches performed since the Tungsten Search and Matching Server was started.

Search Requests per Hour

This is average number of search requests performed per hour. The average is calculated based on a floating window over the last minute.

Server Runtime (hours)

This is the total amount of time the Tungsten Search and Matching Server is running.

Average Search Duration (ms)

This is the average search request time in milliseconds for all search requests performed per hour. The average is calculated based on a floating window over the last minute.

Search Hit Count

This is the total number of search results for all search requests since the Tungsten Search and Matching Server was started.

Records Retrieved Count

This is the total number of database records from the available search results that are retrieved since the Tungsten Search and Matching Server was started.

Average Record Retrieval Time (ms)

This is the average database record retrieval time in milliseconds for all search requests performed. The average is calculated based on a floating window over the last minute.

Active Database Count

This is the number of currently activated databases in the Tungsten Search and Matching Server.

Manual Import Count

This is the number of automatic updates for all databases since the Tungsten Search and Matching Server was started.

Automatic Update Count

This is the number of automatic updates for all databases since the Tungsten Search and Matching Server was started.

Skipped Automatic Update Count

This is the number of skipped automatic updates that were not performed since the Tungsten Search and Matching Server was started. The update is typically skipped because the database is not modified.

Failed Import Count

This is the number of failed manual imports and automatic updates since the Tungsten Search and Matching Server was started.

Number of Database Loads on Demand

This is the number of times database are loaded on demand since the Tungsten Search and Matching Server was started. Such a database is loaded only when it is configured to be loaded on demand and accessed for the first time.

Number of Database Unloads by Timeout

This is the number of times databases are unloaded since the Tungsten Search and Matching Server was started. A database is unloaded when it is configured to be loaded on demand and idle for a specified time.

Chapter 3

The Load Balancer


You can use load balancing if the capacity limit for a single standalone Search and Matching Server is reached so that search requests can no longer be responded to in an efficient way. Load balancing makes it possible to use several computers to share computational workload and to distribute the search requests to a set of different Search and Matching Servers. The clients using the Load Balancer access it as if it was a simple Tungsten Search and Matching Server with the same interface. The Load Balancer is totally transparent to them. The Load Balancer runs on a classic Microsoft Windows service.

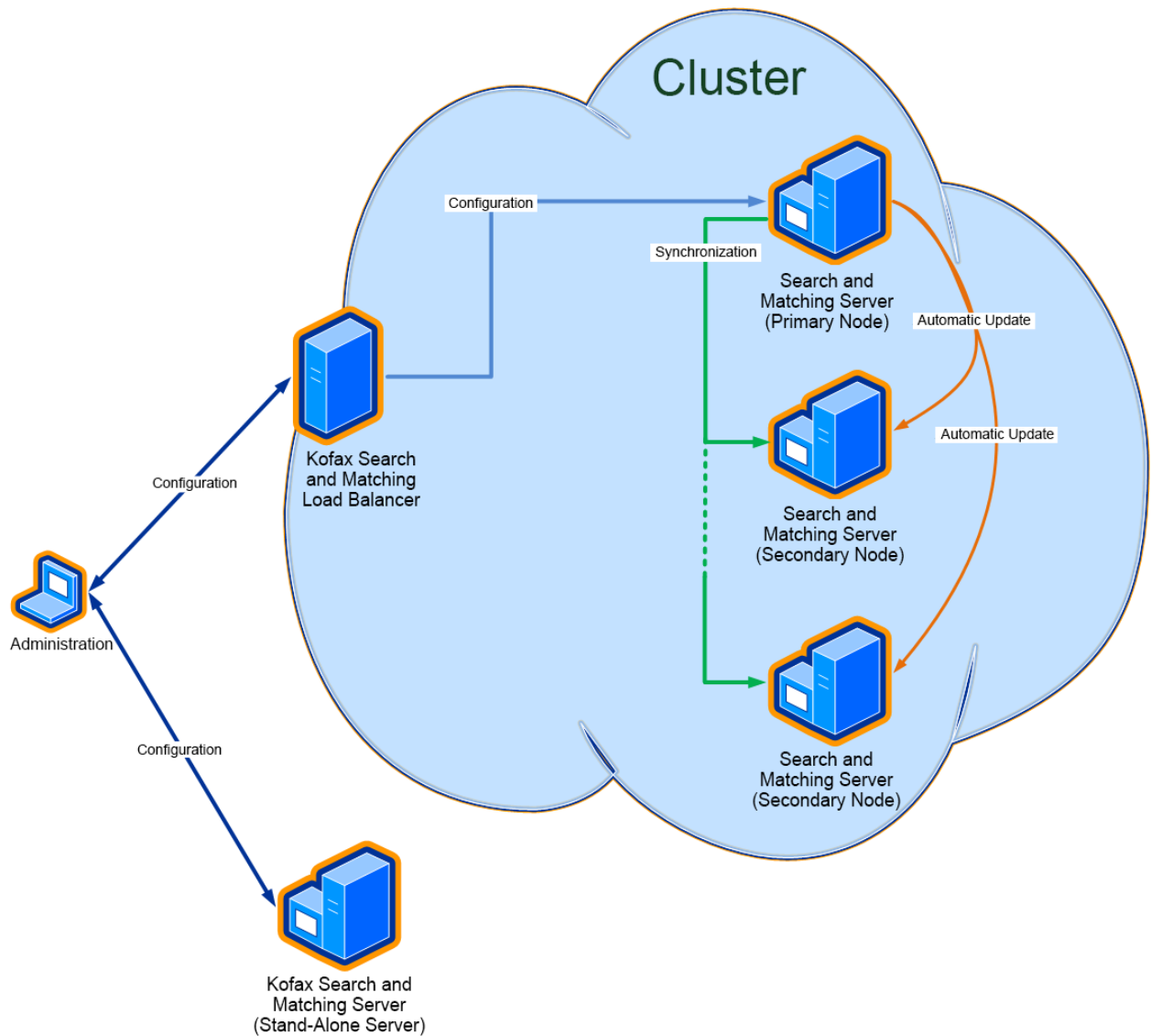
The Load Balancer handles a cluster of several Search and Matching Servers that are defined via the Tungsten Search and Matching Server Administration. In the first step you define which Search and Matching Servers [form the load balancing cluster](#) and in the second step you set up the databases..

A standalone server that is added to a cluster becomes a cluster node and any search requests are then forwarded through the Load Balancer. To minimize the configuration and maintenance effort a cluster always consists of a "primary node" and one or more "secondary" nodes. The databases are defined and configured in the Load Balancer. The selected Load Balancer will push the creation and import of the databases to the primary node. When the primary node is ready, it synchronizes the database automatically to the secondary nodes. This way, if you have a relational database, only the primary node will access to it, and it will prepare the fuzzy index for the secondary nodes. By default, the first server that is added to a cluster is the primary node. If you add more servers to the cluster they become secondary nodes. All existing databases defined on a server are deleted when a Search and Matching Server is added to the cluster as a secondary node. The databases on the secondary node are synchronized automatically from the primary node. If you connect to a secondary node via the Administration tool the menu items that are used to configure databases are not available. If needed, for example as a fallback when the Load Balancer is no longer accessible, you can convert a secondary node to become a standalone server again and configure it or access to it normally. A cluster node that is converted to a standalone server is automatically removed from the cluster.

A Search and Matching Server can be either a standalone server or a member of one single cluster.

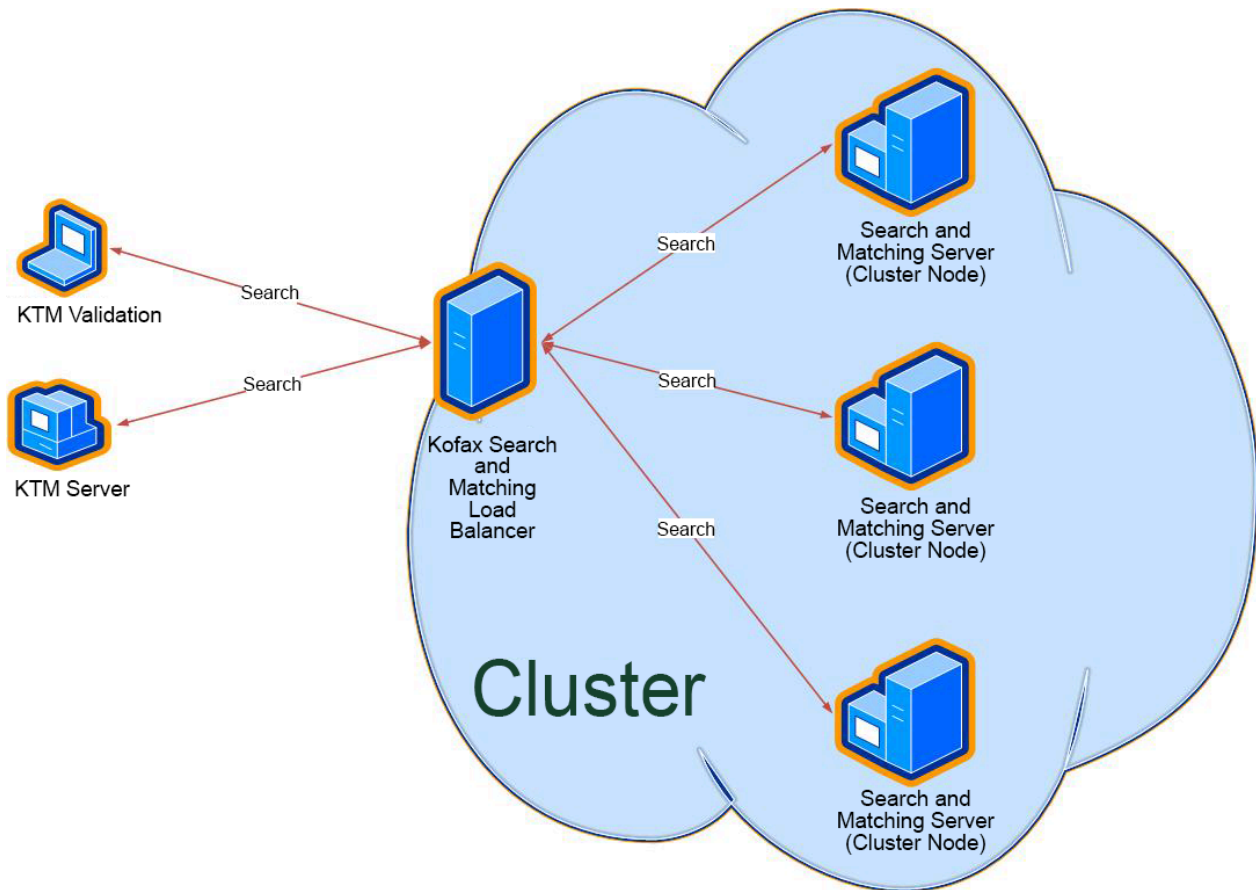
If you delete the primary node from the cluster you have to select another server (secondary node) from the list of cluster nodes to become the primary node. If for any reason no primary node is defined in the cluster, then database configuration is not possible, and the automatic update feature from the databases is skipped. If you add a server that is already member of a different cluster, this server is automatically removed from the other cluster. As a kind of fall back you can convert any cluster node to a standalone server.

 An optimal size of a load balancing cluster consists of 2 to 16 cluster nodes.



On the client side you configure a remote fuzzy database on a standalone Search and Matching Server in the same way as one from a Load Balancer by connecting to either the standalone server or the load balancer. For more details see *Project Builder Help*.

The Load Balancer provides the possibility to handle the search requests from different clients by gathering all search requests from all clients and sending the search requests one after the other to the next available Search and Matching Server in the cluster. This means that each search request is performed by a single Search and Matching Server and not distributed to multiple servers.




In parallel to handling search requests the Load Balancer synchronizes changes from the primary node to all cluster nodes within the cluster. For example, for a newly-created database or an automatic database update. The Load Balancer is fault tolerant in regards to handling search requests or performing synchronization for any cluster node that is temporarily unavailable. As soon as the node is available again it is reincorporated in the cluster for the search. If needed, the node is updated so that the databases get the latest version..

When you set up load balancing you have to consider that the Load Balancer and all cluster nodes need to share the same settings for the [security](#) to ensure a granted access. This means that the services have to run as a user that is member of the TSMS Administrators user group configured during installation and this user must have read and write permissions to the [data storage directory](#).

For maintenance and security reasons the best practice is to use the Load Balancer in a domain group installation only.

By default, the Load Balancer and Search and Matching Server services are executed for a [built-in user account](#), which is the NETWORK SERVICE user, after the installation.

 If the service is running under a user that has a changed password then you need to also update the services account.

Provide granted access for load balancing

To set the services account without the need to take care from the databases directory, you can run the installation from a command line that allows you to set the optional parameters, `SERVICEUSERNAME` and `SERVICEUSERPASSWORD`. If the parameters are defined the installed Load Balancer or Search and Matching Server services are executed for the defined user and the read and write permissions to the data storage folder are set accordingly. You can [provide a granted access](#) by changing the user accounts and read-write permission for the data storage directory manually on any computer that runs a Search and Matching Server you want to add to a load balancing cluster.

You can configure granted access and set up a load balancing cluster by following these steps:

1. Contact your system or domain administrator to add a new service user to the "TSM Administrators" user group created during installation.
2. From the control pane, select **Administrative Tools** and then select **Services** to stop the Load Balancer service (Tungsten Search and Matching Load Balancer service).
3. Change the service account for the Load Balancer service (Tungsten Search and Matching Load Balancer service) to the newly-created service user.
4. Change the read and write permissions for the data storage directory for any computer that runs the Tungsten Search and Matching Server service.
5. Restart the Load Balancer service.
6. Repeat steps 1 to 5 on each computer that runs the Tungsten Search and Matching Server services that is a member of the load balancing cluster.

Chapter 4

Tungsten Search and Matching Server Administration

After you have installed Tungsten Search and Matching Server Administration you can use the tool to create and maintain fuzzy databases for different Tungsten Search and Matching Servers or to set up a cluster of several standalone servers in order to configure load balancing. Multiple users can log on to the same Tungsten Search and Matching Server and maintain the server in parallel.

During installation of the Tungsten Search and Matching Server you define the configuration type, a domain group or a local group, in respect to the security aspects within your company and how to access the server for search requests.

Like Tungsten Search and Matching Server, the Administration tool runs on a 64-bit server only.

Start the Tungsten Search and Matching Server Administration to perform the following actions:

- [Accessing](#) the Tungsten Search and Matching Server
- Setting up a [new fuzzy database](#)
- Configuring import settings for a fuzzy database that is created from a table or view from a text file or a relational database
- [Configuring automatic update](#) and other general settings for the error management and loading options
- Displaying logging reports
- [Maintaining](#) an existing database

Access the Tungsten Search and Matching Server


You can use the Administration to configure both the standalone Search and Matching Server, but also the Load Balancer to set up a cluster of several Search and Matching Servers.


To ensure a granted access you need to know how the Search and Matching Server or Load Balancer you want to connect to is installed. If the current user is a member of the configured TSMS Administrators user group you can use single-sign on. If the current user is not a member of the TSMS Administrators user group you can type the user name and password for a user that is member of the user group.


In order to run the Tungsten Search and Matching Server in a domain you need to contact your domain administrator to create the "TSMS Administrators" group and assign users to it before installing Tungsten Search and Matching Server or one of its components.

1. From the control pane, select **Search and Matching Server** and then select **Administration**.

The **Log On to Server** window is displayed.

2. Type the name by which the local area network identifies the computer that runs the standalone Search and Matching Server or Load Balancer. You can also type the IP address of the server, or type "localhost" if the Tungsten Search and Matching Server is installed on your local computer. Note that you can search for all available servers within your local network by clicking .

 The search may take some time, depending on the size of your network and the speed of your network connection.

 If you connect to a server a second time, you can click the Down Arrow to select a server from the list.

3. If needed, change the port number that is used for communication with the standalone Search and Matching Server (the default is 22000) or Load Balancer (the default is 22001). The port number must correspond to the value that is configured during the installation.
4. The following steps are only needed if you need to connect to the server for a different user account as you are logged on to the computer.
 - a. Type the domain\user name.
 - b. Type the password.
5. Select **Connect** to set up the connection the selected server. After a successful connection a list of available databases is displayed.

Set up a fuzzy database

In order to set up a fuzzy database you add a new fuzzy database, configure the import settings and if wanted, configure the general settings.

1. To add a new fuzzy database to the list of available databases click **Add** and perform the following steps when the **Add New Database** window is displayed:

The new fuzzy database is added to the list of **Available Databases** and its database status is set to **Created**.

- a. Type the database name.
 - b. Select the database source either **Text file** if the database source is a delimited text file or **Relational database** if the database source is a Microsoft SQL Server, an Oracle database or an ODBC file.
 - c. Click **OK** to save the settings.
2. Configure the newly created fuzzy database by clicking **Import Settings** from the **Database Settings** tab. For a text file database source the **Import Settings for Database (Text File Source)** window is displayed and for a relational database source the **Import Settings for Database (Relational Database Source)** window is displayed, in order to define the options for the path or connection string, the columns and if needed, string substitutions.

On the **Database Settings** tab the details for the newly created database are updated and a preview of the first 20 records of the imported database are displayed.

3. Optionally, click **General Settings** from the **Database Settings** tab to display the **General Settings** window in order to define load options, automatic update and the error management.
4. Click **Activate** to make the newly-created database accessible.

Configure and maintain existing database

The initial view of the Tungsten Search and Matching Server enables you to maintain the fuzzy databases to do the following.


- [Rename](#) a database
- [Activate or deactivate](#) a database
- [Delete](#) a database
- [Import](#) a database
- [Export](#) a database
- Configure [Microsoft SQL Server connection](#)
- Configure [ODBC database connection](#)
- Configure [Oracle Server connection](#)

Access rights for a Microsoft SQL Server database

Several permissions in the Microsoft SQL Server database are required by Tungsten Search and Matching Server in order to use it for the fuzzy database source. These permissions are necessary in order to access tables, views, synonyms, and stored procedures.

- **Server-level permissions:**
 - The logon object must have the following permissions:
 - `CONNECT SQL`
 - `VIEW ANY DATABASE`
 - The logon object must **not** have the following permission denied:
 - `VIEW ANY DEFINITION`
- **Database-level permissions:**
 - The database user must be a member of the public role. This happens by default when the user is created, but can be removed.
 - The database role or the database user if you are not using roles, must have the following permissions:
 - `CONNECT` must be granted to the database user so that the mapped logon is able to access the database.
 - `VIEW DEFINITION` must be granted or not denied in order to access the available tables, views, synonyms, and stored procedures.
 - `SELECT` must be granted either on the database level or for each of the table, views, or synonym objects.
 - `EXECUTE` must be granted for the stored procedure and synonym that references a stored procedure.

- In the case of a synonym, the user or database role also requires the corresponding `SELECT/EXECUTE` permissions for the underlying object referenced by the synonym.
- **Database-level system metadata permissions:**
 - In order to list the databases available on the server as well as a list of tables, views, synonyms, or stored procedures that are available on the databases, you must grant some system metadata:
 - `SELECT` the following syntax views:
`sys.databases, sys.schemas, sys.objects, and sys.synonyms`
 - In order to check if a table was modified since the last fuzzy database import some system metadata must be set.
 - `SELECT` the following syntax view:
`sys.dm_db_index_usage_stats`

 By default, these system metadata permissions are granted to the public role.

Configure a Microsoft SQL Server connection

This window enables you to set up the connection string for a Microsoft SQL Server.

You can configure a Microsoft SQL Server connection that connects to a table, view or a stored procedure, or a synonym.

In order to use a synonym, the following permissions are needed:

Limits

Synonyms have the following limitations:

- Synonyms are supported for the following objects only:
 - Table
 - View
 - Stored Procedure

You can configure an Microsoft SQL Server database connection by following these steps:

1. Select the database server from the **Server Name** list.
2. Select the type of authentication and enter logon information if needed.
3. Select the database from the **Database Name** list.
4. Click **OK** to save the connection string.

Access rights for an Oracle database

Several permissions in an Oracle database are required by Tungsten Search and Matching Server in order to use it for the fuzzy database source. These permissions are necessary in order to access tables, views, synonyms, and stored procedures.

At a minimum, the Server login requires the following permissions.

- `CONNECT SQL`
- `CREATE SESSION`
- `SELECT ANY TABLE` or `READ ANY TABLE`

The following permissions are also required by the "role" or "user" for each relational object.

- `SELECT`
This is required to access views, tables, and synonyms based on a view or table.
- `EXECUTE`
This is required for the stored procedure and synonyms based on a stored procedure.
- In the case of a synonym, the user or database role also requires the corresponding `SELECT/EXECUTE` permissions for the underlying object referenced by the synonym.

Configure an Oracle database connection

The Configure Oracle Connection window is displayed to set up the connection string for a Oracle database server.

Before you can configure an Oracle database connection, the Oracle driver required. If this is not installed, you are warned and you cannot add an Oracle database. For more information about installing the Oracle driver, refer to the *Tungsten Search and Matching Server Installation Guide*.

You can configure an Oracle connection that connects to a table, view, store procedure, or a private synonym.

Limits

Synonyms have the following limitations:

- Only private synonyms are supported.
- Synonyms are supported for the following objects only:
 - Table
 - View
 - Stored Procedure

You can configure an Oracle database connection by following these steps:

1. Select the database server from the **Server Name** list.
2. Select the type of authentication and enter logon information if needed.
3. Select the database from the **Database Name** list.
4. Click **OK** to save the connection string.

Configure an ODBC database connection

The Configure ODBC Connection window is displayed to set up the connection string for a ODBC database.

1. Type the name of the system DNS.
2. Select the type of authentication and enter logon information if needed.

3. Select the database from the **Database Name** list.
4. Click **OK** to save the connection string.

Considerations for fuzzy database creation

In order to set up a fuzzy database you need to configure the import settings and the general settings. For example, the minimum user settings and rights for an Oracle database that is used as source for a fuzzy index or considerations for the selection of the columns or how to use string substitutions.

The following aspects need to be taken into account.

- Consider [minimum user right definitions](#) for an Oracle database source
- Define [columns for the import](#)
- Define [string substitutions](#)
- Set up [import and update](#)
- Use [stored procedures](#) for relational database sources

Fuzzy database CSV and text file formats

A fuzzy database needs to comply with the following features before it can be added and successfully processed.

- The database file contains one record per line only.
- Line breaks can be either CR+LF (Carriage Return/Linefeed) or LF (Linefeed).
- The database format or encoding is either ANSI, or UTF-8 with the BOM (Byte Order Marker). If a UTF-8 file is missing a BOM character, it is interpreted as an ANSI file instead.
- Values are separated by a field delimiter that is configurable in the database properties.
- Different field delimiter characters are allowed to indicate where one field ends and another begins, but double characters like `::` are not supported as field delimiters.
- Values can be enclosed in double quotes yet still contain a field delimiter character to represent a string of fields that are to be kept together. For example, `Chris;Booker;Wetzingerstr 19;79100;Freiburg.`
- The first line of the database file can contain the column headers, but this is not mandatory.
- Every record has the same number of columns. If a record does not have a value for a specific field, it still needs to be represented with an empty value between two field delimiter characters. If a record contains a different number of columns, it is ignored when the database is imported and a message is displayed to the user.

Imported columns

As databases often contain additional columns such as internal customer IDs or contact names and are not displayed on the document. You need to consider the following aspects for the creation of the fuzzy database and the columns that you select for the Column Configuration import setting for the use with Tungsten Transformation Modules or Tungsten TotalAgility.

Extraction

When you use a fuzzy database for extraction such as for a Database Locator, but not in validation, you need to clear the selection of columns whose values are not displayed on the document as

these decrease the confidence of the returned alternatives. The complete record inclusive the non selected columns is returned to the database locator and will be used for the final confidence computation.


Tungsten Transformation Modules Validation

If you use a fuzzy database in validation such as a database lookup button on a validation form, but not for extraction, you have to select all database columns you want to search for, even if their values are not displayed on the document. Only if those columns are imported can they be displayed in the Tungsten Transformation Modules Validation module within the search results window. The complete alternative is returned and includes all columns, even if the columns are not selected and mapped to the fields normally.

Extraction and validation


If you use a fuzzy database for both extraction and validation, you have to select a column even if its values are not displayed on the document so that they are imported and you can search for them in Validation. For the extraction, in order that additional invisible columns do not decrease the confidence, you have to change the settings for the Database Locator and clear the selection of such a column in the properties on the Search Masks tab.

You can change the name of the columns by selecting a column, then clicking the column a second time and typing the new column name.

 If you import a database that has customized column names a second time, but the number of columns in the source text file or relational database have changed, the previously defined column names are automatically replaced by the newly imported columns.


String substitutions

The string substitution table substitutes search text in the document and in the database. It is used to normalize the database text with the query texts. The Search Text should be unique and should not be a substring from the Replacement Text. A typical example is when street names are normalized from "Avenue", "Aven." and "Ave." to "Av." Substring replacements apply not only to whole words but also to word substrings.

 When you want to replace a substring or a word where the characters you add are on the ignore list, these characters are removed before the string substitution. The conversion to lower case also apply if "Ignore Case" setting is checked.

You can export your current list of string substitutions to be used for other databases as well. You can import string substitutions from another database from the same or another Tungsten Search and Matching Server, a Tungsten Transformation Modules, or Tungsten TotalAgility fuzzy database.

Add string substitutions

1. In the **String Substitution** area, type the search text into the **Search Text** field.
2. Type the replace text into the **Replacement Text** field.
3. Click **Add** .

The sting substitution is added to the list.

Modify string substitutions

1. Select the string substitution you want to modify. Its value populates the editable text fields above.
2. Type in your modifications and click **Modify** ➔.

Import string substitutions

1. In the **String Substitution** area, click **Import**.
2. In the Windows Explorer window, navigate to the location of the file you are importing, select the *.sdl file and click **Open**.

Export string substitutions


If you use the same dictionary in more than one project, you can create a list of string substitutions and export them to use in other projects.

1. In the String Substitution area, click **Export**.
2. In the Windows Explorer window, navigate to the save location, type in a file name for the *.sdl file and click **Save**.

Import and update

If you add or delete columns from the database source of an existing and ready to use fuzzy index that has automatic update configured, remember to change the import settings before the automatic update is performed. Otherwise the update re-imports the fuzzy database, and records that do not have the same number of configured columns are not imported; therefore you end up with a fuzzy database that contains no records. Though the server logging report for the update stores the information about the number of imported and not imported records, the import is successfully completed and the database status is still set to Ready. Any search requests for such a database do not return any results.

By default, if an automatic update of a fuzzy index fails, searches can still be performed as the system does not switch to the updated fuzzy index, but continues to use the previous version. For example, if the database source was temporarily unavailable because virus detection locked the file. You can change this behavior by selecting the Error Management setting. Then, if the automatic update fails the database status of the fuzzy database is changed to Error. A fuzzy database that has this database status cannot perform any search requests and it does not change the status until it is successfully imported again. You have to either check the import settings and import the database again, or wait for the next successful update.

 By clearing the Error Management setting you cannot revert to the previous database version.

Stored procedures for relational databases

For a Microsoft SQL Server or an Oracle database you can also use stored procedures for the database import and automatic update. The use for the database import is that it returns the database records to create the fuzzy index from and allows to perform additional steps. For example, to log the access for your sensitive data. A stored procedure for the automatic update needs to return a date and a time that indicates when the database source was last changed.

This means that you have to ensure that the date for altering the database is stored and updated for any change. For example, it is necessary to know when a bank account is changed or when a customer is removed from the database altogether. The automatic update is performed only if your data has changed since the previous update.

The following are examples of functions and stored procedure syntax:

- Stored procedure for [database import](#)
 - Oracle Server sample
 - Microsoft SQL Server sample
- Stored procedure for [automatic update](#)
 - Oracle Server sample
 - Microsoft SQL Server sample

Stored Procedures for Database Import

The following sample procedures return all records for the "SampleDatabase" for both an Oracle database and a Microsoft SQL Server database.

The following sample procedure is used to check for a "SampleDatabase" if the Vendor has changed. It accesses the table, `VendorLastChanged` for the field, `LastDate`. Note that for any change that altered the Vendor this field must be updated, otherwise the automatic update is never performed.

Oracle Server

```
CREATE OR REPLACE PROCEDURE sp_GetSampleTable (p_recordset OUT SYS_REFCURSOR) AS
BEGIN
  -- Insert statements for example to log the data access
  OPEN p_recordset FOR
    SELECT * FROM SampleDatabase;
END sp_GetSampleTable;
```

Microsoft SQL Server

```
USE [SampleDatabase]
GO
/***** Object: StoredProcedure [dbo].[ p_GetSampleTable] Script Date: 04/13/2011
04:47:06 *****/
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
CREATE PROCEDURE [dbo].[sp_GetSampleTable]
-- Add the parameters for the stored procedure here
AS
BEGIN
  -- SET NOCOUNT ON added to prevent extra result sets from
  -- interfering with SELECT statements.
  SET NOCOUNT ON;

  -- Insert statements for example to log the data access
  -- Insert statements for procedure here
  SELECT [SupplierNumber]
    , [Name_1]
    , [Name_2]
    , [Street]
```

```
,[PostCode_1]
,[City]
From SampleDatabase.dbo.Vendor;
END
```

Stored Procedures for Automatic Update

The following sample stored procedures check for a "SampleDatabase" if the Vendor has changed. It accesses the table, VendorLastChanged for the field, LastDate. Note that for any change that altered the Vendor this field must be updated, otherwise the automatic update is never performed. Examples for both Oracle Server and Microsoft SQL Server are shown.

Oracle Server

```
CREATE OR REPLACE PROCEDURE sp_GetLastChangedDateForSampleDatabase (l_result OUT DATE)
AS
BEGIN
SELECT LastDate INTO l_result FROM VendorLastChanged WHERE rownum < 2;
END sp_GetLastChangedDateForSampleDatabase;
```

Microsoft SQL Server

```
USE [SampleDatabase]
GO
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
CREATE PROCEDURE [dbo].[sp_ChangedDate]
@lastChangeDate as DateTime output
AS
BEGIN
-- SET NOCOUNT ON added to prevent extra result sets from interfering with SELECT
statements.
SET NOCOUNT ON;

-- Insert statements for procedure here
SET @lastChangeDate = (SELECT TOP 1 LastDate FROM
SampleDatabase.dbo.VendorLastChanged);
END
```

Rename a database


In order to rename a selected fuzzy database the Rename Database window is displayed. Note that you can only rename a fuzzy database that is not activated. That means an activated fuzzy database must be deactivated first. Before an activated database can be deactivated any search requests for this database need to be finished. Further search requests for the deactivated database result in an error and the batch is sent to Tungsten Capture Quality Control.

1. Select a fuzzy database from the list of **Available Databases** you want to rename.
2. If necessary, click **Deactivate** to disable the selected database for further access.
The icon in front of the selected database changes.
3. Click **Rename** and type a new name for the database.
4. Update the mapping from the database on the client side.
5. Click **OK** to close the window.

The list of **Available Databases** is updated.



Activate or deactivate a database

You have to activate a fuzzy database so that it can be accessed. To disable the access of an activated database you need to deactivate it.

 In order to change the settings of a selected database, or to rename or delete it, the database must be deactivated.

In Tungsten Transformation Modules, if a database that is used in production is deactivated, search requests from a Tungsten Transformation Modules application result in an error and the batch is sent to Tungsten Capture Quality Control.


You can select multiple databases that have the Ready status in order to activate them, or multiple databases that have either the Ready or Error status to deactivate them. If the selected databases have different activation statuses then the status of the first selected database defines the status that is assigned to all databases. For example, if the first database is activated then all selected databases are deactivated.

1. Select a fuzzy database you want to activate or deactivate from the list of **Available Databases**.
2. Click **Activate**  or **Deactivate**  for the selected fuzzy database.

An icon is displayed in front of the database in the list of **Available Databases** to indicate whether the database is activated or deactivated.

Delete a database

You can delete a selected fuzzy database. Note that you can only delete a fuzzy database that is not activated. That means an activated fuzzy database must be deactivated first.

 In Tungsten Transformation Modules, if search requests for this database are performed, this can take a while. The outstanding search requests are finished before the database is deactivated. Further search requests for the deactivated database result in an error and the batch is sent to Tungsten Capture Quality Control.

1. Select a fuzzy database from the list of **Available Databases** you want to delete.
2. If necessary, click **Deactivate** to disable the selected database.
The icon in front of the selected database changes.
3. Click **Delete** to remove the fuzzy database from the list of **Available Databases**.

Export a database

Use the Export Databases window to export one or more fuzzy databases in order to store a backup so that they can be imported to another Tungsten Search and Matching Server. When you export a database you can choose a full export that saves the fuzzy database and its configured import and general settings. Alternatively, you can save the configured settings for the fuzzy database only, which means that the fuzzy database is created from the exported settings when you import the database.

For a fuzzy database that was created from an uploaded delimited text file you can only select Full export. This is because the server does not have access to an uploaded file and therefore cannot create the fuzzy database from the exported settings during import.

1. Click **Export**.

The **Export Databases** window is displayed.

2. Select the location where the export file is saved to.
3. Select one or more fuzzy databases you want to export.
4. For all selected databases you can configure the export type. By default, **Full export** is selected. This saves the fuzzy database and its settings.

To export the configured settings but not the fuzzy database itself, select **Export settings only**.

5. Click **Export** to save the zip file and return to the main window.


Import a database

The Import Databases window is displayed to import fuzzy databases that were exported from a Tungsten Search and Matching Server. Note that the import of a database that is exported with the Export settings only setting can take longer as the Fuzzy Database is automatically created during import.

1. Click **Import**.


The **Import Databases** window is displayed.

2. Select a zip file and then one or more files you want to import.

 If the zip file was created in a previous version of Tungsten Search and Matching Server, it is possible that a password was assigned to protect the database. Passwords are no longer supported so it is necessary to use a third-party zip tool such as 7-Zip to remove the password before you import the file.

3. If wanted, you can change the following import options for the selected files:
 - a. Select **Activate databases after import** to activate the databases immediately after the import.
 - b. Select **Replace existing databases** to replace existing databases with the imported databases that have the same name.

If **Replace existing databases** is not selected, but a database with the duplicate name exists, the imported database is automatically renamed and a consecutive number is added at the end of the file name.

 If **Replace existing databases** is not selected, but a database with the duplicate name exists, the imported database is automatically renamed and a consecutive number is added at the end of the file name.

4. Click **Import** to import the all selected databases and return to the main window.

Configure automatic database update

For databases that are updated regularly you can define an hourly, daily or weekly update. On the configured update time the Tungsten Search and Matching Server checks whether the source file has changed and if so, starts the upload using the configured import settings to create an updated fuzzy database.

All search requests that are sent while the database is being updated are performed on the "old" database. As soon as the update is complete search requests are then performed on the updated database.

You can define the settings for an automatic database update on the General Settings window for databases that are imported from relational databases or delimited text files that have an URL/UNC location. This is because the Tungsten Search and Matching Server must have access rights to automatically upload and re-import the database. For fuzzy databases that are imported from a manually uploaded delimited text file that automatic update options are not available.

If a source file changes that you have manually uploaded to the server, you have to manually upload the source file again and then re-import the fuzzy database.

For a Microsoft SQL Server or an Oracle database you can select a stored procedure that returns a date and time when the database source was last changed. The update is performed only when the change date is newer than the date for last update. If no stored procedure is selected, the system checks if the database source has changed and performs the update only if it has changed or if the check for changes did not return a result.

You can configure automatic database updates by following these steps:

1. Select a database from the list of **Available Databases**.
2. From the **Database Settings** tab, click **General Settings**.
3. Optionally, you can select a stored procedure for a Microsoft SQL Server and an Oracle database that starts the update only if the database has changed.
4. Select **Use automatic update** and select the days (if weekly is selected) and the hour(s), when the upload is performed.
5. Click **OK** to save the settings.

Set up load balancing


You set up the load balancer by connecting the Administration tool to the computer that runs the Load Balancer. In a first step you create the load balancer cluster and then configure the fuzzy databases for the Search and Matching Server that is set as the primary node. Alternatively, you can set up a standalone Search and Matching Server first that you want to use as primary node for the load balancer, and then set up the load balancer cluster. You only have to add the configured server as first server to the cluster so that it is set as the primary node. Otherwise, if the cluster contains a primary node already the server is added as secondary node and all database are deleted from it when it is added to the cluster.

In order to change the databases and its settings in a cluster you have to update the configuration by connecting to the Load Balancer from the Administration. Any changes are synchronized to the cluster nodes.

i When a Load Balancer is deployed and one or more of its Tungsten Search and Matching Servers are also configured as alternative hosts, the extraction process uses the load balancer by default.

1. From the control pane, select **Search and Matching Server** and then select **Administration**. The **Log On to Server** window is displayed.

2. Type the name by which the local area network identifies the computer that runs the Load Balancer.

You can also type the IP address, or type "localhost" if the Load Balancer is installed on your local computer. Note that you can search for all available load balancer within your local network by clicking .

i The search may take some time, depending on the size of your network and the speed of your network connection.

i If you connect to a server a second time, you can click the Down Arrow to select a server from the list.

The **Server Discovery** window is displayed to select the load balancer.

3. If needed, change the port number (the default for the load balancer is 22001) that is used for communication with the Load Balancer. The port number must correspond to the value that is configured during the installation of the load balancer. If you select the load balancer from the **Server Discovery** window the port number is broadcasted automatically.
4. Click **Connect** to set up the connection the selected load balancer using single-sign on.
5. From the **Cluster** menu, select **Cluster Management** to set up the cluster of different Search and Matching Servers that are managed by connected load balancer.
The Cluster Management window is displayed to perform one of the following steps.
Repeat the steps until all wanted servers are added to the cluster and click **OK**.
6. Set up the fuzzy databases as needed that are then automatically synchronized to all nodes in the cluster.