

INFOWATCH

InfoWatch Data Discovery. Installation Guide

12/01/2026

© JSC “InfoWatch”

Tel/Fax +7 (495) 229-00-22

http://www.infowatch.com

Table of Contents

1 Who should read this guide 5

2 Set of documents 6

3 Technical support 7

4 Installation/Update Types, Compatibility with Other Products, Hardware and Software Requirements 8

4.1 Installation types and compatibility with other InfoWatch products 8

4.2 Types of update 10

4.3 Exporting data to Traffic Monitor 10

4.4 Hardware and Software Requirements 10

4.4.1 Hardware requirements 10

4.4.2 Software Requirements 11

5 Preparing the Server 13

5.1 The steps to prepare the server: 13

6 Configuring Network Access Rules 14

7 Installing Data Discovery 17

7.1 Options for starting the Data Discovery installation 18

7.2 Content Analysis Module 18

7.3 Installing Data Discovery Interactively 18

7.4 Installing Data Discovery with Parameters on the Command Line 22

7.5 Installing Content Analysis Module Interactively 26

7.6 Installing Content Analysis Module with Parameters on the Command Line 29

8 Installing the License and Plugin 34

8.1 Exporting data to Traffic Monitor 34

8.2 Importing data from Traffic Monitor 34

9 Troubleshooting 36

10 Updating Data Discovery 37

10.1 System settings and data 37

10.2 Data Discovery update options 37
10.3 Updating Data Discovery by Resetting the Installation 38
10.4 Reinstalling Data Discovery to Another Cluster 41
11 Uninstalling Data Discovery 44

This guide contains information on how to install, update, and remove InfoWatch Data Discovery (also referred to as the System or Data Discovery), software designed to search for confidential information on shared network resources, workstations, servers, and in document storages.

1 Who should read this guide

This guide is intended for deployment engineers and security officers who are going to work with the System and administrate it. This guide is intended for users familiar with the basics of the Linux operating system and PostgreSQL DBMS.

2 Set of documents

The documentation includes:

- "InfoWatch Data Discovery. Installation Guide". The document describes how to install, update, and uninstall the System.
- "InfoWatch Data Discovery. Administration Guide". The document contains information on how to configure the System.
- "InfoWatch Data Discovery. User Guide". The document describes how to work with scanning tasks.

Related documentation for InfoWatch Traffic Monitor (also referred to as Traffic Monitor) includes:

- "InfoWatch Traffic Monitor. Installation Guide". The guide describes how to install, configure, update, and uninstall InfoWatch Traffic Monitor.
- "InfoWatch Traffic Monitor. Administration Guide". The guide describes how to administrate InfoWatch Traffic Monitor (database, back end).
- "InfoWatch Traffic Monitor. User Guide". The guide describes how to work with InfoWatch Traffic Monitor (set up configuration, export/import data, prepare policies for processing objects).
- "InfoWatch Traffic Monitor. Configuration Files Guide". Provides explanations for frequently used configuration files.

3 Technical support

If you have any problems or questions related to the operation of the System, please contact our technical support service:

- If you purchased the product from InfoWatch partner, please contact the technical support service of the partner
- If you purchased the product from InfoWatch directly, please contact InfoWatch technical support at support@infowatch.com.

Technical Support Service is available to assist you Mon-Fri, 7 AM – 9 PM (Moscow time, GMT+3), except weekends and public holidays in the Russian Federation. You can also visit technical support section on our website: http://infowatch.com/support.

4 Installation/Update Types, Compatibility with Other Products, Hardware and Software Requirements

4.1 Installation types and compatibility with other InfoWatch products

InfoWatch Data Discovery is a part of InfoWatch Investigation Centre.

InfoWatch Investigation Centre is a single interface that enable operation and effective interaction between the following InfoWatch products:

- Activity Monitor 3.2
- Data Access Tracker 2.2
- Data Discovery 2.0
- Prediction 3.2
- Vision 3.4

Each product can be installed and operated separately. However, using multiple products simultaneously provides more possibilities for the Security Officer to ensure the company's information security.

The main part for installation of each Investigation Centre product is the Platform. The installer of each product includes both the Platform and the product components. Installing both parts is required when you deploy the products.
Both product components and the Platform have two installation modes:

- `central`
- `office`

You can select the installation mode for the Platform and product when running the installer. If you do not select any mode, the installer will use the default `central` value for both components.

The Investigation Centre supports **distributed installation**: the central (`central`) and office (`office`) nodes of the Platform are combined into a cluster.
You can install components of products both on the central and on the office node of Platform. This allows you to create the Platform cluster with multiple products taking into account the infrastructure, security requirements and available capacity of the company.

Important!

When using a distributed installation, take into account the following key aspects:

- For the products in a cluster to operate correctly, you must install them on the same Platform version. In case you install a product on a newer Platform version, update all the other products in the cluster.
- In one cluster, you can deploy only one central node of the Platform and a single copy of each product in the central mode.
- For operation of Data Discovery in the `office` mode, there must be Data Discovery in the `central` mode in the cluster.
- If the cluster already contains a product that was installed in the `central` mode, then another installation of this product in the `central` mode within the same cluster will fail with an error.
- Data Access Tracker and Prediction can be installed only in the `central` mode. Both products support installation both on the central and on the office node of the Platform.

Product of the same cluster have:

- access to all functionality in one management console;
- common settings;
- common role model;
- access to data regardless of which product is integrated with the source, as well as access to all processed data.
  For example, for Prediction, data that was processed in Activity Monitor has higher priority than data that is received directly from the source - Device Monitor, which is a module of the Traffic Monitor product.

Important!

To connect to the Investigation Centre web interface, you should **always** use the address or name of the server on which the **Platform is installed in the `central` mode**.

If in the distributed installation you installed a product in the `central` node on the Platform office node, you should still use the Platform central node (`central`) for work.

Products of InfoWatch Investigation Centre support **joint installation** on one server.

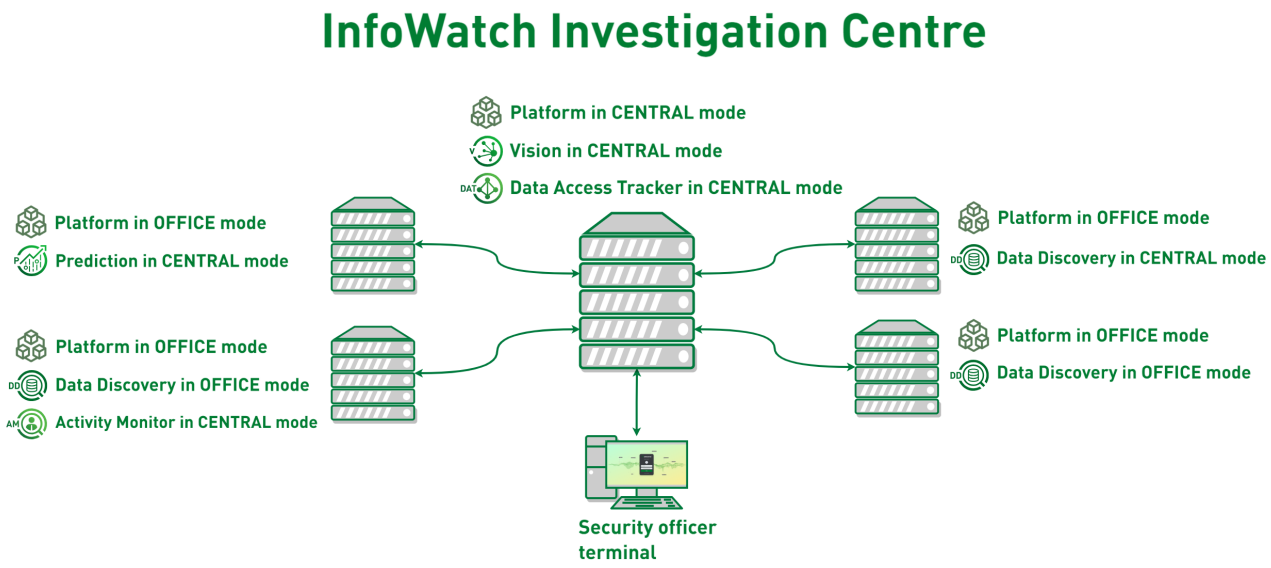
You can install **Data Discovery 2.0 together** with:

- **Activity Monitor 3.2**
- **Data Access Tracker 2.2**
- **Prediction 3.2**
- **Vision 3.4**

Note:

To calculate hardware requirements for joint installations, consider the requirements of each product. To accurately calculate the requirements for a specific project, contact InfoWatch technical support.

Below is an **example** of the Investigation Centre distributed cluster with joint installations of multiple products.



For more information about the installation process, see "Installing Data Discovery".

4.2 Types of update

Data Discovery supports only update by resetting and reinstalling over past data.

For more information about the update process, see "Updating Data Discovery".

4.3 Exporting data to Traffic Monitor

Data Discovery can export data to InfoWatch Traffic Monitor. Data Discovery can work with Traffic Monitor:

- version 6.10.x (6.10.15 and higher);
- version 7.x (7.1 and higher).

For more information about configuring export, see:

- Installing the License and Plugin
- "InfoWatch Data Discovery. Administration Guide", "Zones and servers".

4.4 Hardware and Software Requirements

4.4.1 Hardware requirements

Hardware requirements for Data Discovery depend on the network environment and expected load. Below is an example Data Discovery server configuration for processing 50 hosts within one scanning task:

Component	Recommended requirements
Disk	• Disk array: RAID 5 • Disks: 10K HDD • Disk array capacity: 400 GB **Note**: The required disk space depends on the number and size of files you are going to scan. **Warning!** In case you are using the functionality of InfoWatch Data Analysis Service or/and Content Analysis Module, Data Discovery will need additional disk space to store extracted text contents. We recommend increasing the available disk space by 40%. Please keep in mind that the required additional space depends on text data volume and company's document flow
RAM	16 GB
CPU	8 cores, 16 threads Intel Xeon or equivalent, with support for SSE 4.2 instructions
Network	1 Gbps LAN

Note:

When calculating hardware requirements for a joint installation, we recommend that for each component you should sum up requirements for Data Discovery and other products.

For example: if a product requires 24 GB of RAM, then for this product and Data Discovery to be able to work together, you will need 40 GB of RAM.

You can install the Content Analysis Module, which is used for detecting protected objects in text data, on the same server as Data Discovery or separately.
If you are going to use **the Content Analysis Module on the same server as Data Discovery**:

- To scan storages using one task and one scanner, hardware requirements for Data Discovery must be supplemented with **at least**:
  - 2 CPU cores;
  - 1 GB RAM;
  - 50 GB disk space.
- To scan file storages with a large amount of data using two or more scanning tasks simultaneously with one scanner in each task, add **at least**:
  - 4 CPU cores;
  - 2 GB RAM;
  - 50 GB disk space.

If the expected load increases, we recommend installing additional Content Analysis Modules on separate servers. The load will be distributed evenly between the servers.
If you are going to use **the Content Analysis Module on a separate server**, ensure **at least**:

- 4 CPU cores;
- 12 GB RAM;
- 100 GB disk space.

Important!

We recommend not installing Data Discovery and Traffic Monitor on the same server. Increased load from Traffic Monitor may lead to incorrect operation of Data Discovery.

4.4.2 Software Requirements

To install and operate the System, you should use the following software:

Software type	Options
Operating system	• Red Hat Enterprise Linux: • versions 7.x (starting from 7.7); • versions 8.x (starting from 8.0); • Oracle Linux: • versions 7.x (starting from 7.9); • versions 8.x (starting from 8.4). **Important**: The **Netfilter** and **NAT** modules must be loaded into the OS kernel

Software type	Options
Browser	• Google Chrome; • Yandex.Browser. **Important**: The browser must have hardware support for WebGL/WebGL2

Примечание:

Data Discovery uses DBMS PostgreSQL version 16.4. The DBMS is to be installed together with the product, you are not required to install it separately.

The Content Analysis Module supports installation on the same operating systems as Data Discovery.

5 Preparing the Server

Before installing Data Discovery, follow the steps to prepare the server.

5.1 The steps to prepare the server:

1. Data Discovery installation program is written in Python 2.7. Install the Python 2.7 interpreter. The distribution kit does not include the interpreter package.
2. Configure POD network rules or disable the firewall (for details, see "Firewalld and Kubernetes Interaction Conflicts" and "Firewall Complete Shutdown").
3. Configure network access rules for the correct operation of the product (for details, see "Configuring Network Access Rules").
4. Data Discovery uses Kubernetes. To enable correct operation of Kubernetes, install the following packages:
   - **socat**;
   - **conntrack** or **conntrack-tools** (you should install this package even if **libnetfilter_conntrack** is already installed).

   > **Note:**
   >
   > The distribution kit does not include these packages.
   >
   > Packages may not be available in the repositories of some operating systems. In this case, you download and install the required packages manually.

5. If you use an antivirus, configure it to operate together with the System (see the Knowledge Base article: "Antivirus Configuration for Platform Products").
6. If the OS uses SELinux, install the **container-selinux** package. This package is required for Data Discovery to operate correctly together with SELinux.

   > **Note:**
   >
   > The distribution kit does not include the package.
   >
   > Package may not be available in the repositories of some operating systems. In this case, download and install the required package manually.

6 Configuring Network Access Rules

For correct operation, the following connections must be allowed:

Connection	Port	Description
Access to the web interface and SSH console		
Administrator workstation → Data Discovery server	TCP 22	Data Discovery server port. Used for managing Data Discovery Server via the SSH protocol
Security Officer workstation → Investigation centre	TCP 443	Port for connection to the Investigation centre web interface (central node of Platform). Default value: **443**. You specify the value during the Platform installation. To change the port number: 1. Reset the System installation (see "Troubleshooting"). 2. Reinstall the System with the new value.
Connections to external systems		
Data Discovery server → Traffic Monitor server	TCP 9100	Traffic Monitor server port. Used for sending data from Data Discovery to Traffic Monitor via XAPI
Data Discovery server → LDAP server	TCP 3268 TCP 389	LDAP server ports. Used for integrating Data Discovery with LDAP directories
Data Discovery server → LDAPS server	TCP 3269 TCP 636	LDAPS server ports. Used for integrating Data Discovery with LDAPS directories
Data Discovery server → DNS server	UDP 53	DNS server port. Used for resolving DNS queries
Data Discovery server → NTP server	UDP 123	NTP server port. Used for time synchronization
Data Discovery server → Mail server	For the protocol: • TLS - 465	The mail server ports. The default values are specified.

Connection	Port	Description
	• STARTTLS - 587	Used to send mail notifications from Prediction and Activity Monitor. Data Discovery uses mail server only for regular report sending
Connection to scanning resources (protocols and storages are specified)		
Data Discovery server → SMB	TCP 139, 445	Port of a host with SMB connection. For scanning shared network resources and workstations running MS Windows OS
Data Discovery server → SSH	TCP 22	Port of a host with SSH connection. For scanning servers and workstations running Linux OS
Data Discovery server → SharePoint	TCP 443	Port of a host with SharePoint connection. For scanning document libraries in SharePoint Server 2016 via HTTPS
Data Discovery server → FTP	TCP 21	Port of a host with FTP connection. For scanning file resources via FTP
Data Discovery server → WebDAV (Nextcloud)	TCP 443	Port of a host with WebDAV connection. For scanning Nextcloud data storages via WebDAV
Data Discovery server → Alfresco	TCP 443	Port of a host with Alfresco system. For scanning server of the Alfresco Community Edition content management system via HTTP/HTTPS
Note: The ports specified for connection to scanning resources are default. If your network infrastructure uses other connection ports, take this into account when creating scanning tasks and adding hosts.		
System operation in a cluster		
Platform server in the `office` mode → Platform server in the `central` mode	TCP 6443 TCP 2379, 2380 TCP 10251, 10252, 10255	Platform server ports in the `central` mode. Used for communication between Platform in the `office` mode and Platform in the `central` mode

Connection	Port	Description
Platform server in the `central` mode → Platform server in the `office` mode	TCP 10251, 10255 TCP 30000–32767	Platform server ports in the `office` mode. Used for communication between the Platform in the `central` mode and the Platform in the `office` mode
Platform server in the `central` mode ← → Platform server in the `office` mode	UDP 8472 TCP 10250	Used to tunnel traffic between the cluster nodes
Platform server in the `central` mode ← → Platform server in the `office` mode	TCP 10256	The port must be available on all nodes of the cluster. Used to check the status of the `kube-proxy` component

7 Installing Data Discovery

The distribution kit is an archive whose name is **`iw_discovery_setup_2.1.X.xxx.tar.xz`**, where `xxx` is the build number. The archive contains a set of binary modules of container images required for deploying the Platform and Data Discovery components without access to the Internet (offline installation) as well as the installer.

To install Data Discovery, use command line.

You can install Data Discovery:

- on a separate server;
- in a cluster of Data Discovery servers;
- on a server together with other products;
- in a cluster with other products.

Warning!

In case you are installing Data Discovery in the cluster together with other InfoWatch products, then make sure that the Platform version is the same in all the products on all nodes. Otherwise, it is required to update the products installed on the earlier version of Platform.

To view the Platform version of the product you are installing:

1. Unpack the archive with product distributive. In our example, into the `/ddiscovery` directory. To do this go to the directory and run the following command:

   Example for version 2.0.0.15

   ```
   tar -xvf iw_discovery_setup_2.1.0.15.tar.xz
   ```

2. Run the command:

   ```
   cat /ddiscovery/platform/setup.yaml | grep version
   ```

Warning!

It is necessary that all nodes in a cluster are available for each other in the network. For that, open the required ports (see "Configuring Access Network Rules").

Before installation, also make sure that:

- DNS resolution is set up for node names
  or
- for each node, `/etc/hosts` file specifies accordance between node addresses and names.

To check if nodes are available, run the command between the cluster nodes:

```
ping <node_name>
```

7.1 Options for starting the Data Discovery installation

You can start the Data Discovery installation:

- interactively - the installer will sequentially request the installation parameters (see "Installing Data Discovery Interactively"),
- specifying installation parameters in the installer launch command (see "Installing Data Discovery with Parameters on the Command Line").

In both cases, the installer supports joint installation on the same server with other Investigation Centre products as well as distributed installation in a cluster.

7.2 Content Analysis Module

To detect protected objects in scanned text files without sending data to Traffic Monitor for analysis, Data Discovery uses a separate Content Analysis Module.

The Content Analysis Module must be installed in the same cluster as Data Discovery. You can install it on a separate server or together with Data Discovery.

You can run installation of the Conent Analysis Module:

- **Interactively** - the installer will sequentially ask for installation parameters (see "Installing Content Analysis Module Interactively");
- **Specifying installation parameters** in the installer launch command (see "Installing Content Analysis Module with Parameters on the Command Line").

In both cases, the installer supports both joint installation on the same server with other Investigation Centre products and distributed installation in a cluster.

7.3 Installing Data Discovery Interactively

> **Note:**
>
> In case you are planning a joint installation of the product and Activity Monitor on one server, install Activity Monitor first.
>
> If you are installing Data Discovery on the server where other Investigation Centre products are already installed, the installer will not ask for some of the parameters. This happens because these directories and settings were already specified during the initial installation of the Platform.

To install Data Discovery:

1. Make sure that your server meets the hardware and software requirements (see "Data Discovery Installation Types").
2. Follow the steps to prepare the server for installing Data Discovery (see "Preparing the Server").
3. Create a new directory on the server (for example, `distr_discovery`):

```
mkdir distr_discovery
```

4. Copy the **`iw_discovery_setup_2.0.X.xxx.tar.xz`** archive to the created directory. Here, we are using version `2.0.0.15`, so the archive name is `iw_discovery_setup_2.0.0.15.tar.xz`.
5. Open the directory to which you have copied the archive.
6. Unpack the archive with the distribution kit to this directory:

```
tar -xvf iw_discovery_setup_2.0.0.15.tar.xz
```

7. Run the installer using the command like:

```
./setup.py install --nodemode=<Platform_installation_mode> --
productmode=<product_installation_mode>
```

> **Note:**
>
> To install:
>
> - **the product in the central mode on the Platform central node:**
>
>   ```
>   ./setup.py install --nodemode=central --productmode=central
>   ```
>
> - **the product in the office mode on the Platform central node:**
>
>   ```
>   ./setup.py install --nodemode=central --productmode=office
>   ```
>
> - **the product in the central mode on the Platform office node:**
>
>   ```
>   ./setup.py install --nodemode=office --productmode=central
>   ```
>
> - **the product in the office mode on the Platform office node:**
>
>   ```
>   ./setup.py install --nodemode=office --productmode=office
>   ```
>
> If you do not specify the installation modes for the Platform and the product, the istaller will use the default `central` mode.
>
> If Data Discovery is the first component you are installing on the server and you specify only the `--nodemode` parameter, then the installer will use the same value for the `--productmode` parameter.

> **Important!**
>
> During the installation, the Platform products that are already installed on the server may be unavailable.

8. Read the license agreement. It contains several pages. To go to the next page, press **Enter**.

9. To accept the license agreement, type "**y**" and press **Enter**.

> **Note:**
>
> If you reject the license agreement, the installer will abort the process and exit.

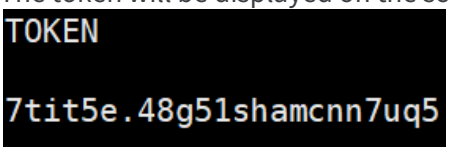
10. If you are intalling the **Platform central node** (`--nodemode=central`), enter the IP address of the network interface to communicate with the cluster and press **Enter**. The IP address should be in IPv4 format: "xxx.xxx.xxx.xxx" (by default, `0.0.0.0`).
If you specify `0.0.0.0` , all available network interfaces will be used.

> **Note:**
>
> Here and below, to use the default values, press **Enter** without entering any values.

11. If you are installing the **Platform office node** (`--nodemode=office`):
    a. Enter the address of the Platform central node in IP address or domain name format (e.g., `192.0.2.0` or `host.example.com`).
    b. Specify the token to connect to the Platform central node.
    To get a token to connect to the Platform central node:
        i. Connect to the Platform central node console via SSH.
        ii. Run the command in the terminal to get the token:

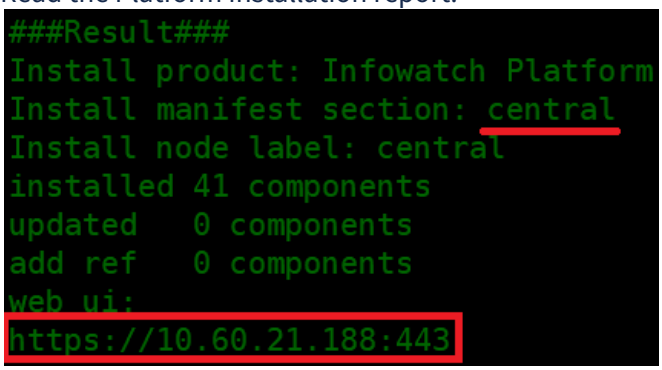
```
kubeadm token list
```

        The token will be displayed on the screen:



        Here, the token is `7tit5e.48g51shamcnn7uq5` .
        iii. Specify the node label (by default, `office`) and press **Enter**.
        The label may consist of letters of the Latin alphabet and/or numbers. The length of the label must not exceed 32 characters.
12. Allocate RAM to host Clickhouse data (by default, `80%`) and press **Enter**.
13. Specify the path to place Clickhouse data (by default, `/opt/chdata`) and press **Enter**.
14. Specify the path to place PostgreSQL data (by default, `/opt/pgdata`) and press **Enter**.
15. Specify the path to place NATS data (by default, `/opt/natsdata`) and press **Enter**.
16. Specify the path to place binary data (by default, `/opt/dsdata`) and press **Enter**.
17. If you are installing the **Platform central node**, specify the port for connecting to the web interface (by default, `443`) and press **Enter**.
18. Wait for the installation of the Platform to finish.

19. Read the Platform installation report.



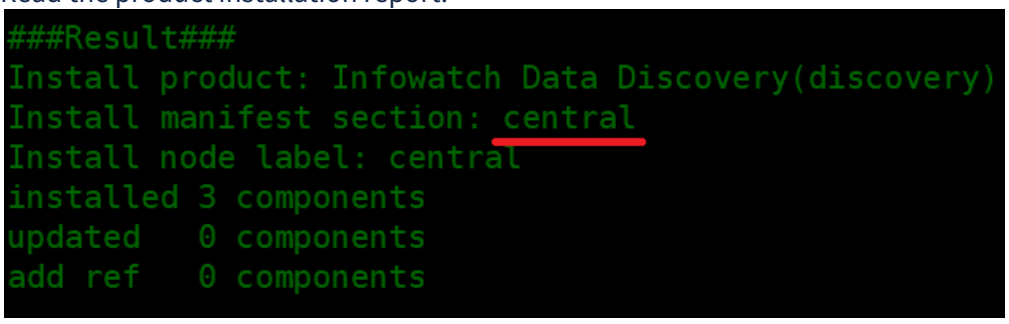
If you install the Platform, the **web ui** column contains the address and port for connecting to the web interface (here, `https://10.60.21.188:443`)

20. Specify the path for storing scanned files (by default, `/opt/disctps`) and press **Enter**. This path will be used on all servers of the Data Discovery cluster.

> **Important!**
>
> The storage directory must be empty.
>
> After installation, the storage should contain only files and directories that were added by Data Discovery. If you add other files or directories to the storage, Data Discovery may not work correctly.

> **Note:**
>
> We recommend locating the storage on a separate partition or disk. On configuring the server storage, see "*InfoWatch Data Discovery. Administration Guide*", "Zones and servers".

21. Wait for the installation of Data Discovery to finish.
22. Read the product installation report:



23. To display information about the Platform and products in the cluster, run the command:

```
./setup.py showproducts
```

24. Make sure that all services are running. For this, use the following command:

```
kubectl get pods -n infowatch
```

The command will display a list of services. Each service should have the `Running` status.

Note:

To increase the speed of scanning resources, you can add Inotify subsystem settings in the OS.

For this:

a. `Open the /etc/sysctl.conf` configuration file for editing.
b. Add the following lines at the end of the file:

```
fs.inotify.max_user_instances = 32768
fs.inotify.max_user_watches = 32768
fs.inotify.max_queued_events = 262144
```

c. Save the changes and close the configuration file.
d. To apply the new subsystem settings, run the command:

```
sysctl -p
```

25. To start using the product, connect to the Investigation Centre web interface. To do this, enter the address and port of the Platform central node in your browser.

7.4 Installing Data Discovery with Parameters on the Command Line

Note:

In case you are planning a joint installation of the product and Activity Monitor on one server, install Activity Monitor first.

The Data Discovery installer allows you to install the product with minimal user interaction. To do this, run the installation program specifying all the required parameters on the command line.

Before running the Data Discovery installer:

1. Make sure that your server meets the hardware and software requirements (see "Data Discovery Installation Types").
2. Follow the steps to prepare the Data Discovery server for installation ("Preparing the Server").
3. Create a new directory on the disk (e.g., `distr_discovery`):

```
mkdir distr_discovery
```

4. Copy the **`iw_discovery_setup_2.0.X.xxx.tar.xz`** archive to the created directory. Here, we are using version `2.0.0.15`, so the archive name is `iw_discovery_setup_2.0.0.15.tar.xz`.
5. Open the directory to which you have copied the archive.
6. Unpack the archive with the distribution kit to this directory:

```
tar -xvf iw_discovery_setup_2.0.0.15.tar.xz
```

Check out the list of keys in the table below. You can use these keys to specify the product installation parameters for the installer. By default, the values that the installer offers interactively are specified.

> **Note:**
>
> If you do not specify the installation mode for the Platform and product, the installer will use the `central` value.

Key	Default value	Available values	Description
`--nodemode`	`central`	`central, office`	The Platform installation mode
`--productmode`	`central`	`central, office`	The product installation mode
`--loc`	`ru`	`ru, en`	Product localization, the language that is used for pre-installed entities. The parameter is **optional** to run the installer
`--licagree`	`n`	`y, yes, n, no`	Accepting the license agreement
`--chmemsize`	`80%`	1-100% or a numeric value If you do not specify %, then the specified value will be treated as the volume in GB	Amount of RAM used for Clickhouse data
`--chdbpath`	`/opt/chdata`		Path for placing ClickHouse data
`--natspath`	`/opt/natsdata`		Path for placing NATS data
`--pgdbpath`	`/opt/pgdata`		Path for placing PostgreSQL data
`--dsdatapath`	`/opt/dsdata`		Path for placing binary data
`--advertise`	`0.0.0.0`		IP address of the network interface on which the `kubelet` service is running. Specify the value when

Key	Default value	Available values	Description
			installing the Platform central node
`--guiport`	`443`		Web interface connection port. Specify the value when installing the Platform central node
`--cluster`			Address of the Platform central node in the cluster. Specify the value when installing the Platform office node
`--clustertoken`			The token to connect to the Platform central node. To get the token, in the Platform central node terminal, run the command: `kubeadm token list` Specify the value when installing the Platform office node
`--nodelabel`	`office`		The node label. Specify the value when installing the Platform office node
`--tpspath`	`/opt/disctps`		Specify the path for storing scanned files. This path will be used on all servers of the Data Discovery cluster. Specify the value when installing the Platform central node

To install Data Discovery:

1. Run the Data Discovery installer with specifying parameters.

Example command to install the cental Platform and Data Discovery in the central mode:

```
./setup.py install --nodemode=central --productmode=central --licagree=yes --advertise=0.0.0.0 --chmemsize=80% --chdbpath=/opt/chdata --pgdbpath=/opt/pgdata --dsdatapath=/opt/dsdata --guiport=443 --natspath=/opt/natsdata --tpspath=/opt/disctps
```

Example command to install the branch Platform and Data Discovery in the office mode:

```
./setup.py install --nodemode=office --productmode=office --licagree=yes --nodelabel=lv-426
 --cluster=10.60.21.188 --clustertoken=bvp2e1.e9zilgwg7v2urtsr --chdbpath=/opt/chdata --
chmemsize=80% --natspath=/opt/natsdata --pgdbpath=/opt/pgdata --dsdatapath=/opt/dsdata
```

2. If you have specified all required parameters, the installer will start the installation.
   If you have not specified some of the parameters, the installer will start the interactive mode and ask you to specify the values.
3. Wait for the Data Discovery installation to complete.
4. To display information about the Platform and products in the cluster, run the command:

```
./setup.py showproducts
```

5. Make sure that all services are running. For this, run the command:

```
kubectl get pods -n infowatch
```

   The command will display the list of services. Each service should have the `Running` status.

> **Note:**
>
> To increase the speed of scanning resources, you can add Inotify subsystem settings in the OS.
>
> For this:
>
> a. `Open the /etc/sysctl.conf` configuration file for editing.
> b. Add the following lines at the end of the file:
>
> ```
> fs.inotify.max_user_instances = 32768
> fs.inotify.max_user_watches = 32768
> fs.inotify.max_queued_events = 262144
> ```
>
> c. Save the changes and close the configuration file.
> d. To apply the new subsystem settings, run the command:
>
> ```
> sysctl -p
> ```

6. To start using the product, connect to the Investigation Centre web interface. To do this, enter the central Platform address and port in your browser.

7.5 Installing Content Analysis Module Interactively

> **Note:**
>
> The Content Analysis Module must be installed in the same cluster as Data Discovery.
>
> If you are installing the Content Analysis Module on a server where other Investigation Centre products are already installed, the installer will not ask for some of the parameters, as those directories and settings were already specified during the initial installation of the Platform.

To install the Content Analysis Module:

1. Make sure that your server meets the hardware and software requirements (see "Installation/Update Types, Compatibility with Other Products, Hardware and Software Requirements").
2. Follow the steps to prepare the server for installing the Content Analysis Module (see "Preparing the Server").
3. Create a new directory on the server (for example, `distr_cas`):

   ```
   mkdir distr_cas
   ```

4. Copy the **`iw_tm_setup_7.11.0.xxx.tar.xz`** archive to the created directory. Here, we are using version `7.11.0.209`, so the archive name is `iw_tm_setup_7.11.0.209.tar.xz`.
5. Open the directory to which you have copied the archive.
6. Unpack the archive with the distribution kit to this directory:

   ```
   tar -xvf iw_tm_setup_7.11.0.209.tar.xz
   ```

7. Run the installer using the command like:

   ```
   ./setup.py install --nodemode=<Platform_installation_mode> --
   productmode=<module_installation_mode>
   ```

   > **Note:**
   >
   > To install:
   >
   > - **the module in the central mode on the Platform central node:**
   >
   >   ```
   >   ./setup.py install --nodemode=central --productmode=central
   >   ```

- **the module in the office mode on the Platform central node:**

```
./setup.py install --nodemode=central --productmode=office
```

- **the module in the central mode on the Platform office node:**

```
./setup.py install --nodemode=office --productmode=central
```

- **the module in the office mode on the Platform office node:**

```
./setup.py install --nodemode=office --productmode=office
```

If you do not specify the installation modes for the Platform and the module, the istaller will use the default `central` mode.

Important!

During the installation, the Platform products that are already installed on the server may be unavailable.

8. Read the license agreement. It contains several pages. To go to the next page, press **Enter**.
9. To accept the license agreement, type "**y**" and press **Enter**.

Note:

If you reject the license agreement, the installer will abort the process and exit.

10. If you are intalling the **Platform central node** (`--nodemode=central`), enter the IP address of the network interface to communicate with the cluster and press **Enter**. The IP address should be in IPv4 format: "xxx.xxx.xxx.xxx" (by default, `0.0.0.0`).
If you specify `0.0.0.0` , all available network interfaces will be used.

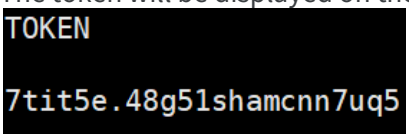
Note:

Here and below, to use the default values, press **Enter** without entering any values.

11. If you are installing the **Platform office node** (`--nodemode=office`):
    a. Enter the address of the Platform central node in IP address or domain name format (e.g., `192.0.2.0` or `host.example.com`).
    b. Specify the token to connect to the Platform central node.
    To get a token to connect to the Platform central node:
        i. Connect to the Platform central node console via SSH.
        ii. Run the command in the terminal to get the token:

```
kubeadm token list
```

The token will be displayed on the screen:



Here, the token is `7tit5e.48g51shamcnn7uq5`.

iii. Specify the node label (by default, `office`) and press **Enter**.
The label may consist of letters of the Latin alphabet and/or numbers. The length of the label must not exceed 32 characters.

12. Allocate RAM to host Clickhouse data (by default, `80%`) and press **Enter**.
13. Specify the path to place Clickhouse data (by default, `/opt/chdata`) and press **Enter**.
14. Specify the path to place PostgreSQL data (by default, `/opt/pgdata`) and press **Enter**.
15. Specify the path to place NATS data (by default, `/opt/natsdata`) and press **Enter**.
16. Specify the path to place binary data (by default, `/opt/dsdata`) and press **Enter**.
17. If you are installing the **Platform central node**, specify the port for connecting to the web interface (by default, `443`) and press **Enter**.
18. Wait for the installation of the Platform to finish.
19. Read the Platform installation report.
20. The Content Analysis Module uses the Traffic Monitor configuration. To upload the configuration, specify the address of the server with the Traffic Monitor database and press **Enter**.

> **Warning!**
>
> Interaction with **PostgreSQL DBMS only** is supported.
>
> In case you are using **Traffic Monitor 7.10 or earlier** versions, protected objects should not contain elements of Autolinguist technology. In order for content analysis module to work with configurations of these Traffic Monitor versions, remove elements of Autolinguist technology from protected objects.

> **Note:**
>
> If Traffic Monitor is installed on several servers or uses a remote database, specify the server where the DBMS is deployed.
>
> If a cluster has several servers with the Content Analysis Module, they all must have the same configuration.

21. Specify the port for connecting to the Traffic Monitor database (by default, `5433`) and press **Enter**.
22. To access the database, specify the password of the **`iwtm_linux`** user.

> **Note:**

If you did not change the user name and password during the installation of Traffic Monitor, specify the default values (see "*InfoWatch Traffic Monitor. Installation Guide*", the "*Preset server parameters*" article).

23. Wait for the installation of the Content Analysis Module to complete.
24. To display information about the Platform and products in the cluster, run the command:

```
./setup.py showproducts
```

25. Make sure that all services are running. For this, use the following command:

```
kubectl get pods -n infowatch
```

The command will display a list of services. Each service should have the `Running` status.

26. For Data Discovery to work with Content Analysis Module the protected objects are required. To add them to Investigation centre, create an integration with Traffic Monitor and run it (for details, see "*InfoWatch Data Discovery. Administration Guide*", the "*Integration with InfoWatch Traffic Monitor*" article). We recommend adding protected objects from Traffic Monitor configuration that was used during installation of the Content Analysis Module.

After the installation, **Settings** in the **Zones and servers** section of the Investigation Centre will be supplemented:

- A server with the **Content analysis** role will be added- if the Content Analysis Module is installed separately from Data Discovery;
- The **Content analysis** role will be added for the existing server - if if the Content Analysis Module is installed on the same server as Data Discovery.

Note:

You can remove the Content Analysis Module the same way as other products on the Platform:

1. Open the directory with the unpacked distribution kit.
2. Run the command:

```
./setup.py remove
```

7.6 Installing Content Analysis Module with Parameters on the Command Line

Note:

The Content Analysis Module must be installed in the same cluster as Data Discovery.

If you are installing the Content Analysis Module on a server where other Investigation Centre products are already installed, the installer will not ask for some of the parameters, as those directories and settings were already specified during the initial installation of the Platform.

The Content Analysis Module installer allows you to install the product with minimal user interaction. To do this, run the installation program specifying all the required parameters on the command line.

Before running the Content Analysis Module installer:

1. Make sure that your server meets the hardware and software requirements (see "Installation/ Update Types, Compatibility with Other Products, Hardware and Software Requirements").
2. Follow the steps to prepare the server for installing the Content Analysis Module ("Preparing the Server").
3. Create a new directory on the server (e.g., `distr_cas`):

```
mkdir distr_cas
```

4. Copy the **`iw_tm_setup_7.11.0.xxx.tar.xz`** archive to the created directory.
   Here, we are using version `7.11.0.209`, so the archive name is `iw_tm_setup_7.11.0.209.tar.xz`.
5. Open the directory to which you have copied the archive.
6. Unpack the archive with the distribution kit to this directory:

```
tar -xvf iw_tm_setup_7.11.0.209.tar.xz
```

Check out the list of keys in the table below. You can use these keys to specify the product installation parameters for the installer. By default, the values that the installer offers interactively are specified.

Note:

If you do not specify the installation mode for the Platform and product, the installer will use the `central` value.

Key	Default value	Available values	Description
`--advertise`	`0.0.0.0`		IP address of the network interface on which the `kubelet` service is running. Specify the value when installing the Platform central node
`--chmemsize`	`80%`	1-100% or a numeric value If you do not specify %, then the specified value will be treated as the volume in GB	Amount of RAM used for Clickhouse data

Key	Default value	Available values	Description
`--guiport`	`443`		Web interface connection port. Specify the value when installing the Platform central node
`--tm-db-port`	`5443`		Port for connecting to the Traffic Monitor database
`--chdbpath`	`/opt/chdata`		Path for placing ClickHouse data
`--dsdatapath`	`/opt/dsdata`		Path for placing binary data
`--natspath`	`/opt/natsdata`		Path for placing NATS data
`--pgdbpath`	`/opt/pgdata`		Path for placing PostgreSQL data
`--nodemode`	`central`	`central, office`	Platform installation mode
`--productmode`	`central`	`central, office`	Module installation mode
`--licagree`	`n`	`y, yes, n, no`	Accepting the license agreement
`--nodelabel`	`office`		The node label. Specify the value when installing the Platform office node
`--loc`	`ru`	`ru, en`	Product localization, the language that is used for pre-installed entities. The parameter is **optional** to run the installer
`--cluster`			Address of the Platform central node in the cluster. Specify the value when installing the Platform office node
`--clustertoken`			The token to connect to the Platform central node. To get the token, in the Platform central node terminal, run the command:

Key	Default value	Available values	Description
			`kubeadm token list` Specify the value when installing the Platform office node
`--tm-db-host`			Address of the server with the Traffic Monitor database. Interaction with PostgreSQL DBMS only is supported. If there are several servers with the Content Analysis Module in the cluster, they must have the same configuration. In case you are using **Traffic Monitor 7.10 or earlier** versions, protected objects should not contain elements of Autolinguist technology. In order for content analysis module to work with configurations of these Traffic Monitor versions, remove elements of Autolinguist technology from protected objects.
`--tm-db-pass-iwtm-linux`			The `iwtm_linux` user password to access the database

To install the Content Analysis Module:

1. Run the Content Analysis Module installer with specifying parameters.

   Example command to install the cental Platform and module in the central mode:

   ```
   ./setup.py install --nodemode=central --productmode=central --licagree=yes --advertise=0.0.0.0 --chmemsize=80% --chdbpath=/opt/chdata --pgdbpath=/opt/pgdata --dsdatapath=/opt/dsdata --guiport=443 --natspath=/opt/natsdata --tm-db-host=10.60.22.33 --tm-db-port=5443 --tm-db-pass-iwtm-linux=xxXX1234
   ```

   Example command to install the branch Platform and module in the office mode:

   ```
   ./setup.py install --nodemode=office --productmode=office --licagree=yes --nodelabel=lv-426 --cluster=10.60.21.188 --clustertoken=bvp2e1.e9zilgwg7v2urtsr --chdbpath=/opt/chdata --chmemsize=80% --natspath=/opt/natsdata --pgdbpath=/opt/pgdata --dsdatapath=/opt/dsdata --tm-db-host=10.60.22.33 --tm-db-port=5443 --tm-db-pass-iwtm-linux=xxXX1234
   ```

2. If you have specified all required parameters, the installer will start the installation.
   If you have not specified some of the parameters, the installer will start the interactive mode and

ask you to specify the values.

3. Wait for the Content Analysis Module installation to complete.
4. To display information about the Platform and products in the cluster, run the command:

```
./setup.py showproducts
```

5. Make sure that all services are running. For this, run the command:

```
kubectl get pods -n infowatch
```

The command will display the list of services. Each service should have the `Running` status.

6. For Data Discovery to work with Content Analysis Module the protected objects are required. To add them to Investigation centre, create an integration with Traffic Monitor and run it (for details, see "*InfoWatch Data Discovery. Administration Guide*", the "*Integration with InfoWatch Traffic Monitor*" article). We recommend adding protected objects from Traffic Monitor configuration that was used during installation of the Content Analysis Module.

After the installation, **Settings** in the **Zones and servers** section of the Investigation Centre will be supplemented:

- A server with the **Content analysis** role will be added- if the Content Analysis Module is installed separately from Data Discovery;
- The **Content analysis** role will be added for the existing server - if if the Content Analysis Module is installed on the same server as Data Discovery.

Note:

You can remove the Content Analysis Module the same way as other products on the Platform:

1. Open the directory with the unpacked distribution kit.
2. Run the command:

```
./setup.py remove
```

8 Installing the License and Plugin

8.1 Exporting data to Traffic Monitor

Data Discovery allows you to export file copies to Traffic Monitor for analysis. For this, you need **the plugin and the license for interception module** in Traffic Monitor.

To add the license:

1. Log into Traffic Monitor server management console as Security officer.
2. Go to **Administration** -> **Licenses**.
3. In the **Licenses** area, click .
4. Using a file manager, add the **tm_license_yyyy_mm_dd.license** license file you received from InfoWatch technical support service.

After the license has been successfully added, information about it is available in **Administration** -> **Licenses**.

To export data, use the preinstalled plugin in the Traffic Monitor Management Console:

- "InfoWatch Crawler" for Traffic Monitor version 7.3 and higher
- "InfoWatch Data Discovery" for Traffic Monitor whose version is lower than 7.3

The plugin is displayed only if a valid license is installed in Traffic Monitor.

Information about the plugin is displayed in the **Administration** -> **Plugins** section in the Traffic Monitor Management Console. To send data to Traffic Monitor, a token is required. The token is created automatically and available in the **Administration** -> **Plugins** -> **Tokens** section.

For more information about configuring data export, see "*InfoWatch Data Discovery. Administration Guide*", "Zones and servers".

8.2 Importing data from Traffic Monitor

To enable exporting data from Traffic Monitor, you need to configure an integration in the Investigation Centre web interface (for details, see "*InfoWatch Data Discovery. Administration Guide*", the "Integration with InfoWatch Traffic Monitor" article). For this, **the Investigation Centre plugin and license for data export** must be added in Traffic Monitor. A token for integration is automatically created after the plugin is installed. The token is available in the **Administration** > **Plugins** section of the Traffic Monitor Management Console, on the **Tokens** tab of the **InfoWatch Investigation Centre** plugin.

In **Traffic Monitor version 7.10 and later**, the unlimited license for data export to Investigation Centre and the **InfoWatch Investigation Centre** plugin are preinstalled.

In **Traffic Monitor version 7.9 and earlier**, add the license and the plugin manually as decribed below.

To add the license manually:

1. Log into Traffic Monitor server management console as Security officer.
2. Go to **Administration** > **Licenses**.
3. In the **Licenses** area, click .
4. Using a file manager, add the **tm_license_yyyy_mm_dd.license** license file you received from InfoWatch technical support service.

After the license has been successfully added, information about it is available in **Administration** > **Licenses**.

Important!

To export data to Investigation centre, the Traffic Monitor licence must contain the following information:

- company issuing a license - **IW**;
- system receiving data - **Platform**.

To add the plugin manually:

1. Log into Traffic Monitor server management console as Security officer.
2. Go to **Administration** > **Plugins**.
3. In the **Plugins** section, click .
4. In the opened window, click **Select file**.
5. Add plugin file using a file manager.
   The plugin is a zip archive `platform/tmplugin/platform.zip`. Folder with the plugin is located in the directory with the unpacked distributive.

After the plugin has been successfully added, information about it is available in **Administration** > **Plugins**.

The "Binding plugin and license" knowledge base article describes how to download a plugin and a license to InfoWatch Traffic Monitor via one archive.

9 Troubleshooting

The installation may be interrupted by the installer crash, power outage, etc. In this case, the System is installed incorrectly. If this happens, we recommend resetting the installation.

To reset the installation:

1. Stop all scanning tasks in the Data Discovery web interface.
2. Open the directory to which you unpacked the distribution archive during the installation.
3. Run the command:
   `./setup.py reset`

Resetting the installation does not delete results of the System operation. You can use them after reinstalling the System.

Warning!

Resetting the System installation automatically resets all other Platform products installed on the server.

If you reset the installation on the Platform central node, components of all products in the cluster will become unavailable. In this case, reset the installation on each office node.

After installing Data Discovery, reinstall other products in the cluster.

10 Updating Data Discovery

Data Discovery update is supported from the following versions:

- **1.8.0**;
- **1.9.0**;
- **1.9.1**;
- **1.10.0**;
- **1.10.1**.

> **Note:**
>
> To update Data Discovery from earlier version, you should first update it to version **1.8.0**, **1.9.x** or **1.10.x**.

The distribution kit is an archive whose name is `iw_discovery_setup_2.0.X.xxx.tar.xz`, where `xxx` is the build number. The archive contains a set of binary modules of container images required for deploying the Platform and Data Discovery components without access to the Internet. The archive also contains the updater.

To update Data Discovery, use command line.

10.1 System settings and data

- **Scanning tasks**:
  - Resource processing within the scanning tasks continues from where it ended before the System update.

    > **Warning!**
    >
    > Before updating, stop all the scanning tasks in the Data Discovery interface. Start the tasks manually when the update is completed on all the cluster servers.

- **Configuration files**:
  - If you update Data Discovery by resetting the current installation, some configuration files may be set to the default values. If you have edited the System configuration files, check values in the files after the update.

10.2 Data Discovery update options

All update options support joint work with other Platform products.

> **Important!**
>
> In case of joint installation of Data Discovery and Prediction, stop Prediction data calculation for the duration of Data Discovery update. See the "Required Steps to Update Products that are Installed on Platform together with Prediction" article in the knowledge base for details.

Data Discovery 2.0 supports only update by resetting the current installation and reinstalling over old data.

Note:

If Data Discovery and other products are installed in different clusters, you can update the products in the other cluster and then install the new Data Discovery version to that cluster (for details, see 'Reinstalling Data Discovery to Another Cluster").

Warning!

In case you are updating Data Discovery in the cluster together with other InfoWatch products, then make sure that the Platform version is the same in all the products on all nodes. Otherwise, it is required to update the products installed on the earlier version of Platform.

To view the Platform version of the product you are updating:

1. Unpack the archive with product distributive. In our example, into the `/ddiscovery` directory. To do this go to the directory and run the following command:

   Example for version 2.0.0.77:

   ```
   tar -xvf iw_discovery_setup_2.0.0.77.tar.xz
   ```

2. Run the command:

   ```
   cat /ddiscovery/platform/setup.yaml | grep version
   ```

10.3 Updating Data Discovery by Resetting the Installation

Important!

To update a whole distributed Investigation Centre cluster by resetting the current installation, follow the steps:

1. **Reset** installation on **office nodes** with products in `office` mode.
2. **Reset** installation on **office nodes** with products in `central` mode.
3. **Reset** installation on the **central node**.
4. **Install** the **central** node.
5. **Install office** nodes with products in `central` mode.
6. **Install office** nodes with products in `office` mode.

When updating a cluster, reset installation only once for each node.

Updating be resetting the installation consists of the following steps:

1. Resetting the installation. Resetting the installation does not delete the System data.
2. Installing new versions of the Platform and product and specifying already existing folders with data.

Note:

If you connect remotely via SSH, we recommend using the Screen utility to avoid problems in case of a lost connection with the server. When disconnecting from the utility, all the processes started are not to be interrupted and you may safely continue the System update.

To update Data Discovery:

1. **Stop all scanning tasks** in the Data Discovery web interface.
2. Follow the steps to prepare the server for the update (see "Preparing the Server").
3. Create a new directory on the disk (e.g., `distr_discovery`):

   ```
   mkdir distr_discovery
   ```

4. Copy the **`iw_discovery_setup_2.0.X.xxx.tar.xz`** archive to the created directory. Here, we are using version `2.0.0.15`, so the archive name is `iw_discovery_setup_2.0.0.15.tar.xz`.
5. Open the directory, to which you have copied the archive.
6. Unpack the archive with the distribution kit to this directory:

   ```
   tar -xvf iw_discovery_setup_2.0.0.15.tar.xz
   ```

7. In case Data Discovery and Prediction are installed jointly or you are planning to use them jointly, stop Prediction data calculation for the time of Data Discovery update. For details, see "Required Steps to Update Products that are Installed on Platform together with Prediction".
8. When you install the new version after the reset, assign the same labels to the cluster nodes as they had before the update.
   To view the list of labels in the cluster, run the command:

   ```
   kubectl get node -n infowatch -o jsonpath="{range .items[*]}
   {.metadata.annotations.flannel\.alpha\.coreos\.com/public\-ip}
   {.metadata.labels.kubernetes\.io/hostname} {.metadata.labels.iwnode}{'\n'}{end}"
   ```

9. In order to save settings for zones and servers in the new version, before resetting the installation, connect to the node where Data Discovery is installed in the **`central`** mode and stop the **Discoperator service** using the command:

   ```
   kubectl scale --replicas=0 deployment discoperator-<node_label> -n infowatch
   ```

   Example for Data Discovery in the central mode on the office node with the lv426 label:

   ```
   kubectl scale --replicas=0 deployment discoperator-lv426 -n infowatch
   ```

10. Before resetting the installation, remember or save the paths to the current data storage directories. To get the list of directories, use the command:

```
cat /var/lib/iwplatform/data/volumes.txt
```

11. **Reset** the current installation using **the installer of the new version**:

```
./setup.py reset
```

> **Important!**
>
> If there are other Investigation Centre products installed on the server, installations of these products will be reset when resetting the Platform installation. When resetting the installation on the central node, products installed on office nodes will stop working.
>
> After installing the new version of Data Discovery, reinstall the Investigation Centre products and specify directories with existing data.
>
> In case you have **already updated another product** on this server to a compatible version, **do not reset Platform installation again** when updating Data Discovery, because the Platform is already of the required version.
>
> If the product is installed **on the office node in central mode**, **access to products will be reset** for the added user roles after resetting the installation.

12. Wait for the reset installation process for the Platform and product to complete.
13. If you are also upgrading to a different server or a new OS:
    a. Copy directories with the data that were specified during installation to another partition or to a new server.

> **Note:**
>
> **In Data Discovery 1.10 and earlier**, the directories are located along the `/mnt` path by default.
> **In Data Discovery 2.0 and later**, the directories are located along the `/opt` path by default.
> Be sure to preserve rights and owners when copying the directories. For copying, use standard tools, such as the **`rsync`** utility.

    b. After resetting the current installation, copy the `/var/lib/iwplatform/` **`dbadmin.yaml`** file with the database passwords from the old server to the new one.
14. To free up some disk space, you can delete the directory with container images of previous versions `/var/lib/` **`containerd.`** For this, complete the command:

```
rm -rf /var/lib/containerd
```

15. Reboot the server using the command:

```
reboot
```

16. Install the new version of Data Discovery (see "Installing Data Discovery").

Important!

In case Activity Monitor was installed on the server before the reset, install it first, then install Prediction.

During the installation, specify existing directories to place the data. When the System asks you if you want to clear up folders before the installation, select "**n**".

If you are updating **from version 1.10 and earlier**, please note that default installation paths in version 2.0 are different. If you previously used the default paths, then during the installation of Data Discovery 2.0 or later versions, the installer will not be able to find data from the previous version and will create directories along the `/opt` path. To avoid this, move the data to the `/opt` directory or specify the old directories when installing a new version.

All nodes of the cluster with Data Discovery should contain the same path to the storage of scanned files (`/opt/disctps` by default).

17. If other products were installed on the server before the reset, install the product versions that are compatible with Data Discovery 2.0 (see "Installation/Update Types, Compatibility with Other Products, Hardware and Software Requirements").
18. If Data Discovery is installed in a distributed cluster, update the Platform and products on the other servers to versions that are compatible with Data Discovery 2.0.
19. If you need to free up disk space after updating, you can delete unused `docker` or `containerd` containers. For details, see "How to delete unused Docker and Containerd containers".
20. Clear the cache of the browser you were using before updating the product.
    The Investigation Centre web interface is available at the Platform central node address.

Important!

If Data Discovery is installed **on the office node in central mode**, **access to products will be reset** for the added user roles as a result of the previous installation reset. After updating, go to **Settings** and reconfigure access to products for these roles (see the "*Roles*" article in "*InfoWatch Data Discovery. Administration Guide*" for more information about roles).

10.4 Reinstalling Data Discovery to Another Cluster

If Data Discovery and other products are installed in different clusters, you can install the new version of Data Discovery to the cluster where another product is already installed. As a result, you will be able to work with Data Discovery and the other products in the single web interface of the Investigation Centre. Besides, the products will have common intergrations and the ability to exchange their data.

Important!

Please note that when reinstalling Data Discovery to another cluster, the following entities **will not be transferred**:

- users,
- roles,
- integrations,
- scanning task launch schedules.

As a result, the new cluster of the Investigation Centre will contain only users, roles and integrations that were there before adding Data Discovery.

After reinstallation, **reconfigure the schedules** for launching the scanning tasks.

To reinstall Data Discovery to another cluster:

1. Choose a cluster where you want to install the new version of Data Discovery. In this cluster, update the products to versions that are compatible with Data Discovery 2.0 (see "Installation/ Update Types, Compatibility with Other Products, Hardware and Software Requirements").
2. On the Data Discovery server, follow the steps to prepare the server for the update ("Preparing the Server").
3. Create a new directory on the disk (e.g., `distr_discovery`):

   ```
   mkdir distr_discovery
   ```

4. Copy the **`iw_discovery_setup_2.0.X.xxx.tar.xz`** to the created directory. Here, we are using version `2.0.0.15`, so the archive name is `iw_discovery_setup_2.0.0.15.tar.xz`.
5. Go to the directory, to which you copied the archive.
6. Unpack the archive with the distibution kit to this directory:

   ```
   tar -xvf iw_discovery_setup_2.0.0.15.tar.xz
   ```

7. Reset the current installation using the installer of the new version:

   ```
   ./setup.py reset
   ```

8. The Platform and product installation will be reset. Wait for the process to complete.
9. To free up some disk space, you can delete the directory with container images of previous versions `/var/lib/containerd`. For this, use the command:

   ```
   rm -rf /var/lib/containerd
   ```

10. Reboot the server using the command:

    ```
    reboot
    ```

11. Install the **Platform office node with the new Data Discovery version** (see "Installing Data Discovery").

    Important!

    To install Data Discovery on the Platform office node, user the following command:

```
./setup.py install --nodemode=office --productmode=central
```

When prompted by the installer, provide the address and token of the new cluster central Platform.

When the installer asks you to specify directories for storing data, specify the directories with the data of the previous version. When asked 'Should I empty folders before installation?', choose «**n**».

If you are undating **from version 1.10 and earlier**, please note that default installation paths in version 2.0 are different. If you previously used the default paths, then during the installation of Data Discovery 2.0 or later versions, the installer will not be able to find data from the previous version and will create directories along the `/opt` path. To avoid this, move the data to the `/opt` directory or specify the old directories when installing a new version.

All nodes of the cluster with Data Discovery should contain the same path to the storage of scanned files (`/opt/disctps` by default).

12. If you need to free up disk space after updating, you can delete unused Docker and Containerd containers. For details, see the article: "How to delete unused Docker and Containerd containers?".
13. Clear the cache of the browser you were using before updating the product.
The Investigation Centre web interface is available at the address of the Platform central node.

11 Uninstalling Data Discovery

To uninstall Data Discovery:

1. **Stop all tasks** in the Data Discovery web interface.
2. Open the directory on the server, to which you unpacked the distribution archive during Data Discovery installation.
3. Run the command:

```
./setup.py remove
```

4. Wait for the removal to complete. The program will remove the System components. The Data Discovery data will be preserved. If necessary, you can delete the data manually from directories that were specified during the installation.
5. Delete the product internal passwords and related data that remained on the server. For this, use the following commands:

```
kubectl delete daemonsets -l product=discovery -n infowatch
kubectl delete deployments -l product=discovery -n infowatch
kubectl delete configmaps -l product=discovery -n infowatch
kubectl delete secrets -l product=discovery -n infowatch
```

Note:

If you remove Data Discovery that was installed in the `central` mode, then the Data Discovery components that were installed in the `office` mode on other nodes are not removed automatically. You need to remove them manually.