Manual Installation of Log Analytics through Script Action

CONTENTS

AIM

This documentation aims to explain how to install log analytics manually using script action in an azure cluster(Spark, Hadoop, Storm, Kafka, HBase, Interactive Query) for getting data in log analytics workspace.

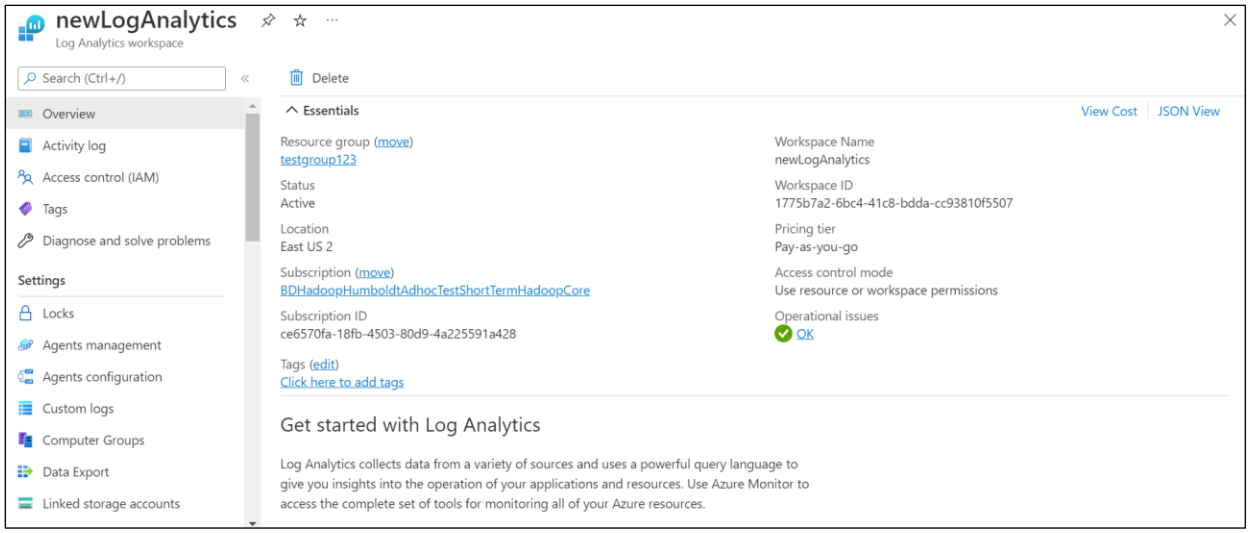
Requirements

1. Running Azure Cluster
2. Running instance of Azure Log Analytics Workspace
3. Log analytics installation script, constants and utilities in an azure storage
4. Parameters like Cluster type, workspace ID, workspace key, Resource Group name and Storage Account name.

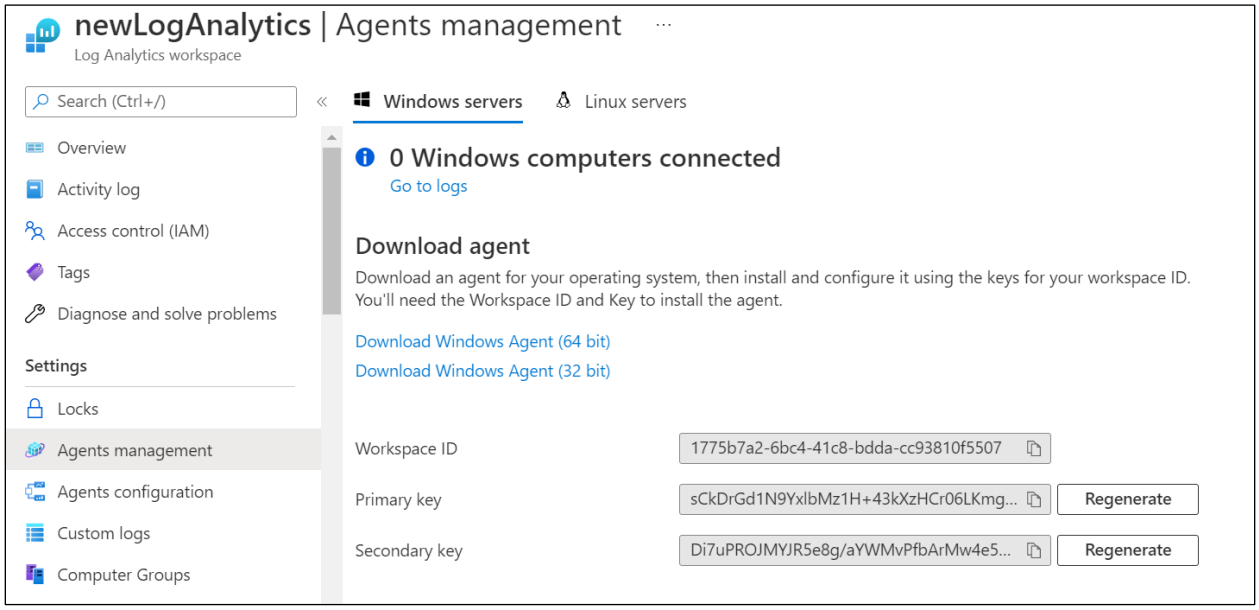
Link to the install script (log_analytics_install.sh)
https://hdiconfigactions.blob.core.windows.net/loganalyticsmonitoring/log_analytics_install.sh

Understanding Log Analytics Workspace

A Log Analytics workspace is a unique environment for log data from Azure Monitor and other Azure services such as Microsoft Sentinel and Microsoft Defender for Cloud.



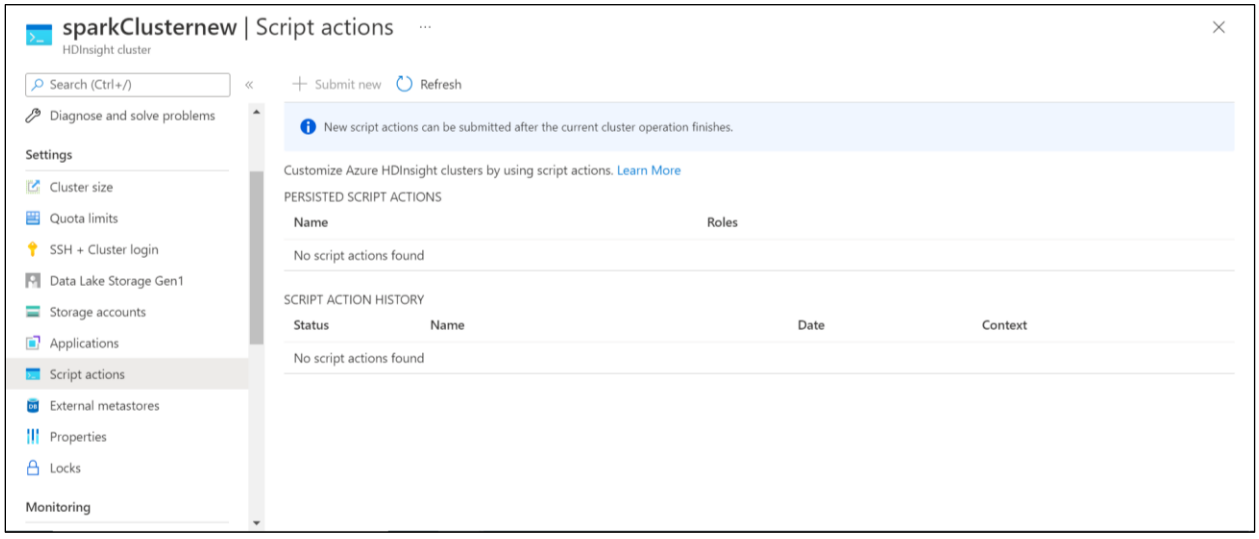
- **Resource Group** in which the workspace is present is shown in the left column.
- **Workspace ID** is present in the rightmost column which will be needed to run the script action.
- **Workspace Key** can be found in the Agents Management under the settings tab in sidebar.



Primary key is the workspace Key which is required by the installation script to connect to the workspace.

Understanding Script Action

Azure HDInsight provides a configuration method called script actions that invoke custom scripts to customize the cluster. These scripts are used to install additional components and change configuration settings. Script actions can be used during or after cluster creation.



Persisted Scripts

Persisted scripts are used to customize new worker nodes added to the cluster through scaling operations. A persisted script might also apply changes to another node type when scaling operations occur.

New Scripts can be added using the **Submit New** button.

Script Action History

It shows the scripts which have been previously run on this cluster. These history records can’t be deleted.

Understanding Parameters Used

Screenshot from log_analytics_install.sh file

```
11    CLUSTER_TYPE=$1
12    WORKSPACE_ID=$2
13
14    #turn off verbose while setting sensitive info
15    set +x
16    WORKSPACE_KEY=$3
17    RESOURCE_GROUP=${4:-N/A}
18    set -x
19    STORAGE_ACCT=${5:-hdiconfigactions}
```

This screenshot shows how each parameter is saved in a variable inside the install script of log_analytics_install.sh file.

```
 8
 9  logger -p user.info "$SCRIPTNAME - Install Geneva Monitoring Pipeline to send information to Log Analytics"
10
11  CLUSTER_TYPE=$1
12  WORKSPACE_ID=$2
13
```

The first parameter ($1) is CLUSTER_TYPE and the second parameter ($2) is WORKSPACE_ID.

```
14    #turn off verbose while setting sensitive info
15    set +x
16    WORKSPACE_KEY=$3
17    RESOURCE_GROUP=${4:-N/A}
18    set -x
19    STORAGE_ACCT=${5:-hdiconfigactions}
```

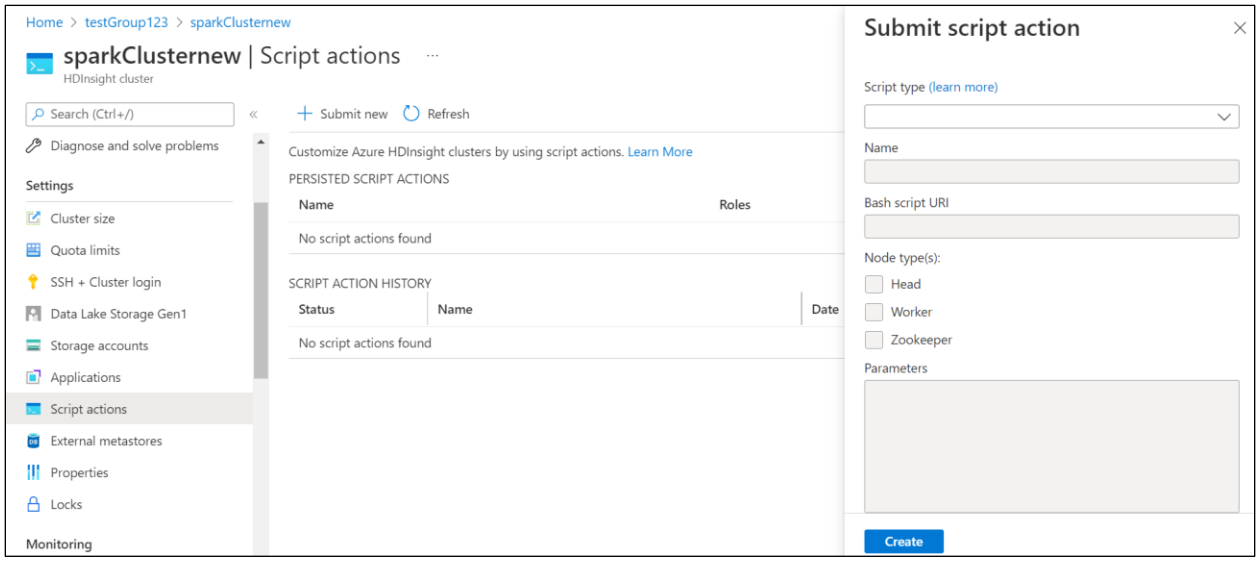
This screenshot shows that 3rd parameter ($3) is WORKSPACE_KEY, 4th parameter ($4) is RESOURCE_GROUP and 5th parameter is STORAGE_ACCT (storage account).

Note:

If no parameter is provided in 4th parameter then it automatically sets it to N/A

If no parameter is provided in 5th parameter then it automatically sets it to hdiconfigactions

Submitting New Script in Script Action



1. Select script type as custom
2. Name your script as Install Log Analytics
3. Paste the Installation Scrip URI in the Bash Script URI
4. Select all node types (Head, Worker and Zookeper)
5. Parameters need to be added in the give order with space in between.
    a) Cluster_Type
    b) Workspace_ID
    c) Workspace_KEY
    d) Resource_Group
    e) Storage_Account_Name

Note: Data Lake Storage Gen2 is not recommended to use for script actions. abfs:// is not supported for the Bash script URI. https:// URIs are possible, but those work for containers that have public access, and the firewall open for the HDInsight Resource Provider, and therefore is not recommended.

A public file-sharing service accessible through https:// paths. Examples are Azure Blob, GitHub, or OneDrive. For example URIs, see Example script action scripts.

Example Parameters

spark e981f121-adb2-4d0c-a2e9-e1e81411a2ec WJz8It+6NE0Ypq6Zu1P7pLdJLrot/+mmXSZ8AvQA3WM6EhSrgaXLYB0k/iLjjhmLtRvd2NEuidXzlbluwPtu0A== testGroup123 null

First Parameter: Type of cluster

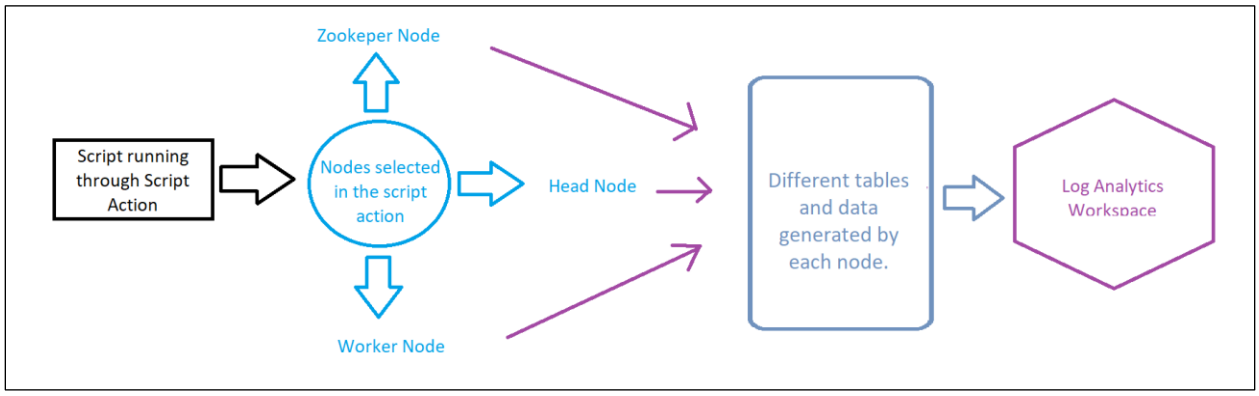
Second Parameter: Log Analytics workspace ID

Third Parameter: Log Analytics workspace key

Fourth Parameter: Resource Group name

Note: Here storage Account name is set as null , the script automatically sets the default storage account name as ‘hdiconfigactions’ in such cases.

Workflow Of Script Action



The script runs in parallel on all the specified nodes in the cluster. It runs with root privileges on the nodes.

Different tables are generated by different nodes inside a cluster and then sent to the Log Analytics Workspace connected to it, where the data populates inside the tables.

Note:

If you change the cluster user, admin, password after the cluster is created, script actions run against this cluster might fail. If you have any persisted script actions that target worker nodes, these scripts might fail when you scale the cluster.