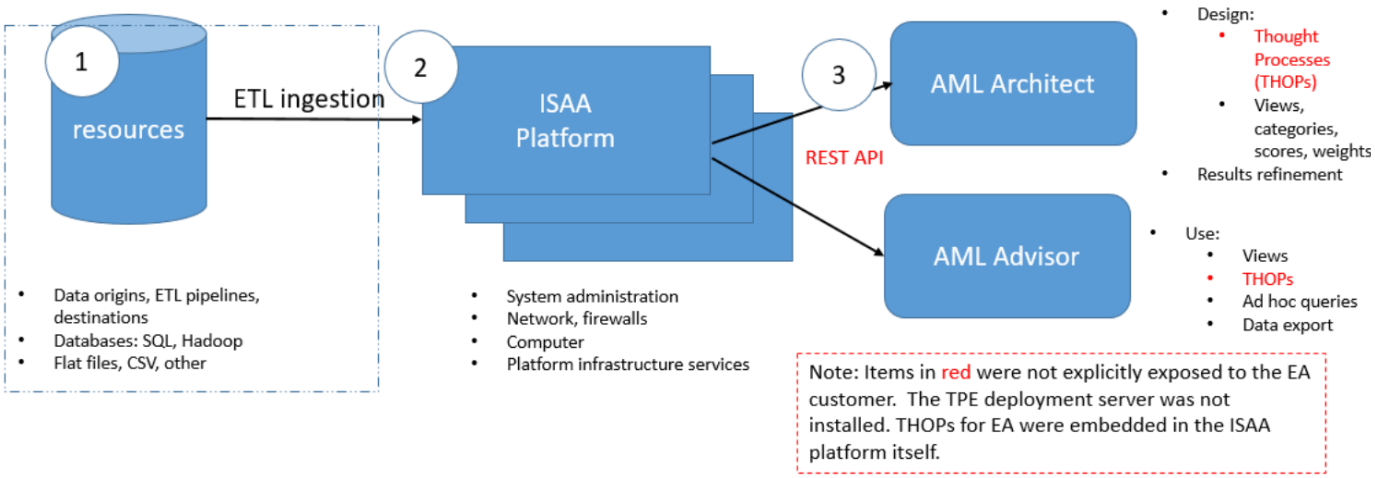
ISAA Administrator Guide, with Installation, EA edition

ISAA conceptual overview

The Intel Saffron Anti-Money-Laundering Advisor (ISAA) is a cognitive computing system for financial institutions to discover *actionable insights* in to possible crime. Based on systematic analysis of your data, you can tailor your analyses to your data and your needs to progressively refine analyses and improve insights.

Below is a simplified, at-a-glance logical view of the ISAA and its subsystems.



1	2	3
Your data is central to ISAA. A data source is called a *resource*. AML Architects design *pipelines* (which have an *origin* and a *desti nation*) to run data transformations via an *ETL* (Extract, Transfer, Load) process called *ingestion*. During ingestion, data are normalized, sent to a destination, and made available for queries via the AML Advisor for further refinement and investigation.	The ISAA platform is the central hub of the ISAA system. You configure clusters of *leader nodes* and *worker nodes*. You can also setup *zones* and *spaces* to secure containers, segregate your data, and isolate ISAA processes.	AML investigators work with ISAA's web user interfaces: the AML Architect and the AML Advisor. • With AML Architect, you design specific queries (called *views*) and *THOught Processes* (THOPs) for use via the AML Advisor. Types of views include *anomaly views* and *customer risk views*. • With the AML Advisor, users can also create ad hoc queries with LiveSearch.

Notes on Early Adopter deployment

Throughout these guides, specific details bout the ISAA deployment at a customer site are indicated with this marker:

> Note on EA deployment

In general, hese notes indicate where the EA deployment varied from the ISAA system design, where the ISAA system itself might have been immature, or where additional manual steps had to be taken to successfully deploy.

Early Adopter (EA) audiences and ISAA documentation

This collection of guides describes the Early Adopter (EA) release of ISAA, which was deployed at a customer site.

The ISAA documentation is grouped into usable collections of information by roles (or personas).

In practice at your site, these roles might be combined. For instance, in test/evaluation, these roles are often a single person.

ISAA AML Advisor Guide

The *ISAA AML Advisor Guide* is for "data explorers", persons using the AML Advisor web interface to investigate, query, and analyze results in ISAA, results based on the work of data analysts.

Recommended skills

- Curiosity
- Knowledge of AML
- Understanding of your specific goals for AML

ISAA AML Architect Guide

The *ISAA AML Architect Guide* is for the “data analyst” (sometimes called "data scientist" or "programmer") who designs the Extract, Transform, Load (ETL) programming, ingestion, categories, attributes, application of algorithms via THOught Processes (THOPs), and query design. With the AML Architect web interface, ETL tools, and JavaScript programming, the data analyst acts as “power user” in preparing data for use by data explorers via the AML Advisor.

Recommended skills

- Deep knowledge of your data and the desired goals/result of your design
- Comfort with ETL
- Familiarity or prior experience with machine learning
- Knowledge of computer programming with REST APIs and JavaScript
- Experience with StreamSets helpful

ISAA Administrator Guide, with Installation

The *ISAA Administrator Guide, with Installation* is for system administrators involved installing ISAA and third-party software, configuring servers and services, data resources, network design, maintenance and upgrades, and administration of databases, clusters, and all ISAA components.

Recommended skills

- Comfort with Linux operating system
- Familiarity with TCP/IP networks, firewalls, ports
- Familiarity with software installation using tar, gzip, RPM
- Familiarity with LDAP
- General system administration
- Familiarity with Hadoop-based systems helpful

ISAA Glossary

The ISAA Glossary, EA edition includes definitions of frequent terms in AML and ISAA.

Deployment planning and installation preparation

These are considerations for planning your deployment and preparing to install ISAA.

There are several high-level parts to installing ISAA:

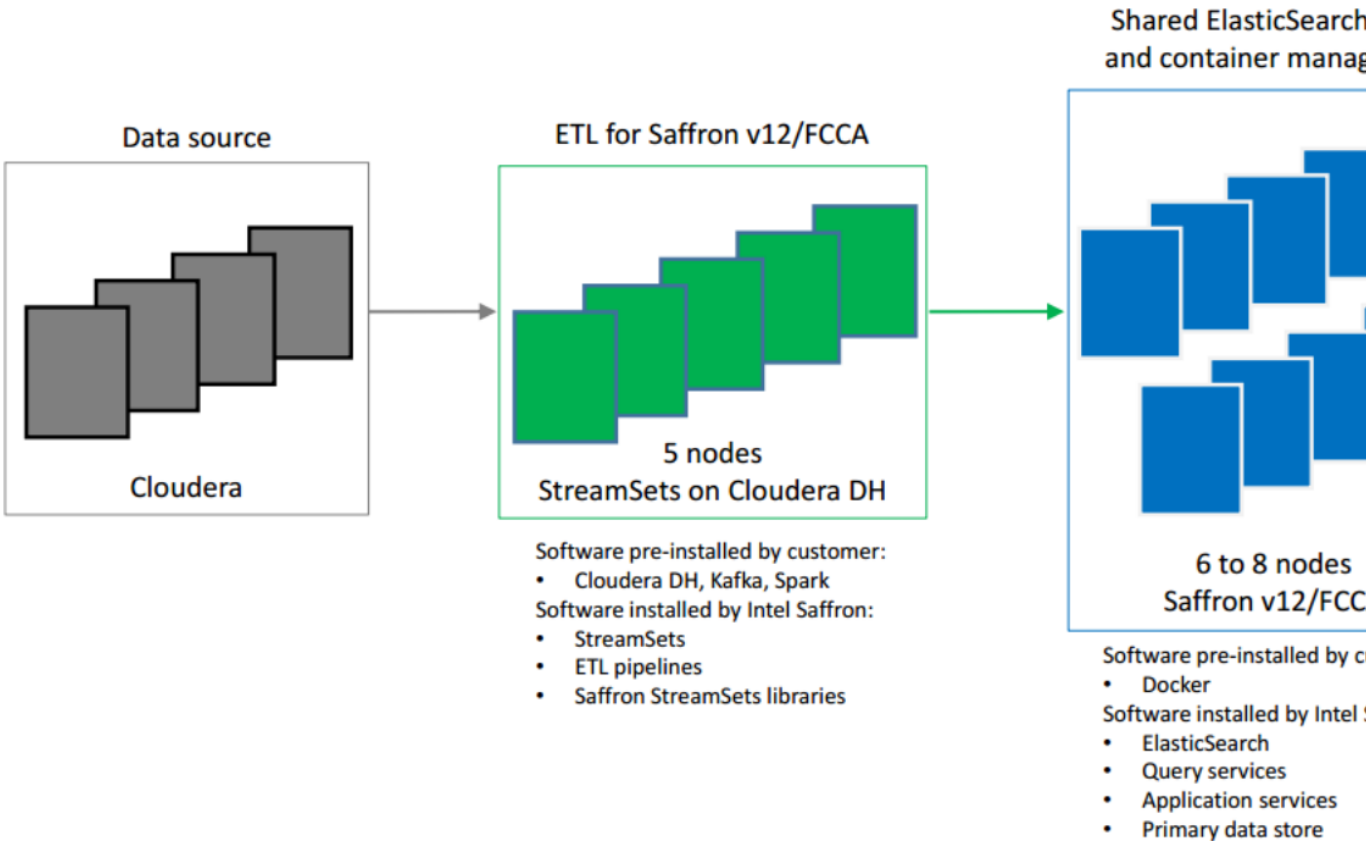
- Components or subsystems that should be pre-installed before installing the core ISAA platform:
  - Cloudera Manager and associated software, which includes Kafka and Spark
  - Docker
  - StreamSets Data Collector as a parcel in Cloudera Manager
- The core ISAA platform
- Loading Saffron programming libraries into StreamSets in Cloudera Manager
- Defining ElasticSearch resources, indexes, and aliases

Planning checklist

You can use the checklist to help plan your deployment.

Early Adopter (EA) deployment architecture

This is a logical view of the ISAA software and hardware architecture as deployed at a customer site for the Early Adopter program.



Functions of the ISAA platform

The ISAA platform serves as the central hub of the entire ISAA system and ties together the various subsystems. Some of its functions include:

- Is the Docker leader node for managing clusters of distributed worker nodes.
- Communicates with your LDAP service to allow end-user access and login.
- Maintains connections to StreamSets data resource origins and destinations for ETL.
- Hosts the Saffron Memory Base (SMB) libraries.
- Includes ElasticSearch software to drive the LiveSearch component in the AML Advisor.
- Serves various REST APIs that can be used throughout the system.
- Serves the end-user web front-ends: the AML Architect and the AML Advisor.
- Logs details about all platform services for auditing.

Third-party software description and requirements

Listed below is third-party software that is integral to ISAA.

> This software should be pre-installed before the ISAA platform.

Third-party software	Version	Vendor or Source	License required?	Bundled with ISAA?	Notes

Cloudera CDH Manager • Cloudera Kafka • Cloudera Spark	v5.8.3 • v0.10.2.0+kafka2.2.0+110 • 0.9.0-1.cdh4.6.0.p0.98	Cloudera	Yes	No	Cloudera is Hadoop-based database management software that includes the following: • Kafka is originally an open-source stream processing platform developed by the Apache Software Foundation. The software aims to provide a unified, high-throughput, low-latency platform for handling real-time data feeds. • Apache Spark is an open-source cluster-computing framework. Originally developed at the University of California, Berkeley's AMPLab, the Spark codebase was later donated to the Apache Software Foundation.
Docker	v2.6	Docker	Yes	No	From Docker, Inc., Docker is software technology for container management. Docker provides an additional layer of abstraction and automation of operating-system-level virtualization on Windows and Linux.
StreamSets Data Collector	v2.6.0.1	StreamSets	Yes	No	StreamSets is software for "big data" ingestion infrastructure and ETL. For ISAA, StreamSets should be installed as a parcel in Cloudera Manager.

Pre-install Cloudera

Cloudera manages the data sources for ISAA. It includes Cloudera Manager.

Cloudera is not bundled with ISAA. This guide assumes that you have already installed Cloudera Manager DH.

For details on Cloudera installation, see the Cloudera documentation at http://www.cloudera.com/documentation/cdh/5-1-x/CDH5-Installation-Guide/CDH5-Installation-Guide.html.

Pre-install StreamSets Data Collector in Cloudera Manager

StreamSets Data Collector is an ISAA component centered on ETL.

StreamSets is not bundled with ISAA.

ISAA recommends that you install StreamSets as a Cloudera parcel, managed by Clouder Manager. This guide assumes that you have already installed StreamSets with Cloudera Manager, as detailed by Cloudera at https://www.cloudera.com/downloads/partner/streamsets.html.

NOTE on EA deployment: Installation of StreamSets as a parcel in Cloudera Manager was in direct contradiction o fthe customer's IT security policies, which disallow installation of software as root (the superuser).

Pre-install Docker software on all nodes

Docker container management software is not bundled with ISAA.

Either the Community Edition or the Enterprise Edition of Docker is suitable for use with ISAA.

After you have planned your cluster deployment, install Docker on all nodes *before* installing ISAA. For instructions from Docker, see https://docs.docker.com/engine/installation/.

Obtain license for ISAA

If you have not already obtained a license for ISAA, contact your Intel representative for details on how to get one.

NOTE on EA deployment: No licensing was involved in the EA program.

For details on installing the license in your ISAA deloyement, see Installing ISAA license.

Operating system CentOS or Red Hat Linux

ISAA supports the following versions of operating systems. This OS must be installed on all cluster nodes:

- CentOS v7.x
- RHEL v7.x

Disable SELinux

On all nodes of the ISAA platform, disable SELinux, which is not required by ISAA:

```
setenforce 0
```

DNS for cluster nodes

Make sure that your company's Domain Name System (DNS) has entries for the nodes in your clusters to resolve hostnames of the leader and worker nodes.

You can rely on static IP addresses for the worker nodes, but the leader node to be accessed by your users should probably have a clear, human readable name for the users.

Synchronized time on all cluster nodes

Be sure the time is synchronized on all nodes in the cluster. You should rely the Network Time Protocol (NTP) daemon (ntpd).

Default ports for AML Architect, AML Advisor, and ISAA admin

The default ports are as follows:

- AML Architect and AML Advisor: 8080
- ISAA administrative interface: 8081

You can change the default ports after installation. See Reconfiguring default ports for ISAA Admin, AML Architect, and AML Advisor.

Firewall rules

Make sure there are no firewalls between leader and worker cluster nodes.

For at least the leader node, on your firewalls, whitelist the IPaddresses and ports of the servers for all *ingestion* data sources; that is, for the computers configured with the Cloudera Manager StreamSets parcel.

Fot internal user access, set appropriate firewall rules between the ISAA cluster's leader node and your intranet.

LDAP details

For user access to ISAA, you need to have an OpenLDAP service, such as Windows Active Directory (AD).

Rather than allowing the entire enterprise to access the system, you might want to create a special group or groups to control access.

Note on EA deployment: Separation of access and data by LDAP group was not part of the EA release.

Have the details of your LDAP service ready to configure in ISAA after installation, as detailed in Configure LDAP.

Docker privileges for username to install ISAA

Make sure that the username that will install the ISAA software has all Docker privileges to run Docker commands. For information on granting privileges in Docker, see the Docker documentation at https://docs.docker.com/engine/installation/linux/linux-postinstall/.

Installation path writeable by installer

The default installation directory for ISAA is `/opt/saffron`. If you want a different location, choose a directory where you want to install the software. Make a note of this path. You can specify this different path during installation.

Throughout the documentation, we refer to this path as *`installation_path`*.

Make sure that the username that will install ISAA has write rights to the *`installation_path`*.

Default administrative username and password

The default ISAA administrator credentials are:

- Username: `admin`
- Password: `administrator`

To secure your deployment, you should plan to change this default password after installation. See Changing the default administrative password .

ISAA platform installation steps

The main steps in installation of ISAA are:

- Unzip the ISAA software
- Run the installer.sh installation script

Unzip the ISAA software

ISAA is packaged as gzipped tar files (.tgz). You need to unzip only a single file. The installation script unzips all the other .tgz files.

To unzip the ISAA software:

1. Put the ISAA tarfile in a convenient location on the node.
2. Run the following command:

   ```
   tar zxvf isaa_package_name.tgz
   ```

Result: Multiple .tgz files are unzipped into subdirectories in *`isaa_package_name_directory`*.

Run the installer.sh installation script

To install the ISAA platform from the command line:

1. The Docker swarm must first be initialized. See Make sure Docker swarm initialized and worker nodes joined.
2. Be sure you have unzipped the ISAA platform package. See Unzip the ISAA software.
3. Change to the *`isaa_package_name_directory`*, which is described in Unzip the ISAA software:

   ```
   cd isaa_package_name_directory
   ```

4. Enter the following command:

   ```
   ./installer.sh
   ```

Result: The ISAA platform is installed, as indicated by the installation script's final output.

Example of running installer.sh on the command line

The following is an example of running the installation script with default settings.

The Docker swarm has already been initialized, as described in Make sure Docker swarm initialized and worker nodes joined.

```
cd isaa_package_name_directory
./installer.sh
setenforce: SELinux is disabled
Please enter the installation path [hit enter for /opt/saffron/]:
Checking if docker and docker-compose are installed and running...
Creating folders structure...
```

```
Loading SMBv12 images....
Loaded image: docker.elastic.co/elasticsearch/elasticsearch:5.5.2
Loaded image: docker.elastic.co/kibana/kibana:5.5.0
Loaded image: isaa.bigcompany.com/saffron/fcca-application-package:0.1.6
Loaded image: isaa.bigcompany.com/saffron/logstash:0.6.1
Loaded image: isaa.bigcompany.com/saffron/newidr-webservice:0.7.6
Loaded image: isaa.bigcompany.com/saffron/redis:3.2
Loaded image: isaa.bigcompany.com/saffron/ssp_apidoc:0.7.3.3-1113
Loaded image: isaa.bigcompany.com/saffron/ssp_auth:0.7.3.3-1113
Loaded image: isaa.bigcompany.com/saffron/ssp_gateway:0.7.3.3-1113
Loaded image: isaa.bigcompany.com/saffron/ssp_license:0.7.3.3-1113
Loaded image: isaa.bigcompany.com/saffron/ssp_memorystore:0.7.3.3-1113
Loaded image: isaa.bigcompany.com/saffron/ssp_report:0.7.3.3-1113
Loaded image: isaa.bigcompany.com/saffron/ssp_secret:0.7.3.3-1113
Loaded image: isaa.bigcompany.com/saffron/ssp_user:0.7.3.3-1113
Loaded image: isaa.bigcompany.com/saffron/ssp_zone:0.7.3.3-1113
REPOSITORY                                              TAG
IMAGE ID            CREATED             SIZE

isaa.bigcompany.com/saffron/fcca-application-package   0.1.6
0840b8df0234        4 weeks ago         462MB
isaa.bigcompany.com/saffron/ssp_gateway                 0.7.3.3-1113
0193566f79c2        4 weeks ago         147MB
isaa.bigcompany.com/saffron/ssp_apidoc                  0.7.3.3-1113
02f9533fc6b7        4 weeks ago         298MB
isaa.bigcompany.com/saffron/ssp_zone                    0.7.3.3-1113
4a708697bedf        4 weeks ago         287MB
isaa.bigcompany.com/saffron/ssp_report                  0.7.3.3-1113
266f4eb01260        4 weeks ago         330MB
isaa.bigcompany.com/saffron/ssp_user                    0.7.3.3-1113
156dbae0d4a0        4 weeks ago         286MB
isaa.bigcompany.com/saffron/ssp_secret                  0.7.3.3-1113
f2a73fc913f9        4 weeks ago         286MB
isaa.bigcompany.com/saffron/ssp_memorystore             0.7.3.3-1113
5bca7dfeba7a        4 weeks ago         340MB
isaa.bigcompany.com/saffron/ssp_license                 0.7.3.3-1113
5507ecd04c96        4 weeks ago         286MB
isaa.bigcompany.com/saffron/ssp_auth                    0.7.3.3-1113
f8f32bcb0abb        4 weeks ago         288MB
isaa.bigcompany.com/saffron/newidr-webservice           0.7.6
bbf9f3a97aba        4 weeks ago         245MB
isaa.bigcompany.com/saffron/fcca-application-package   0.1.6-612
2b33c9577981        4 weeks ago         462MB
isaa.bigcompany.com/saffron/logstash                    0.6.1
e467443c3047        8 weeks ago         578MB
docker.elastic.co/elasticsearch/elasticsearch           5.5.2
ca27036dd5e7        3 months ago        510MB
docker.elastic.co/kibana/kibana                         5.5.0
be0b56c8b9ee        5 months ago        630MB
isaa.bigcompany.com/saffron/redis                       3.2
7d956120fe0c        10 months ago       159MB
################################################################################
################
```

Installation is complete.

1. After installation is complete on all nodes, deploy the stack:

        cd /opt/saffron

        docker stack deploy --compose-file=docker-compose.yml saffron

        ISAA's default username and password are admin/administrator

2. To proceed to LDAP configuration, follow steps below:

        Modify configuration file: vi utils/saffronLDAP.conf

         Run: cd utils ; ./setLdapConfig.sh --file saffronLDAP.conf

How to access:

ISAA Web User Interface: http://your_ip_address:8080 To access use LDAP user/password

ISAA Platform Web Interface: http://your_ip_address:8081/

```
##########################################################################
#################
```

Installing Saffron libraries in StreamSets parcel in Cloudera Manager

ISAA assumes that you have already installed StreamSets as a parcel in Cloudera Manager, as detailed by Cloudera at https://www.cloudera.com/downloads/partner/streamsets.html.

After StreamSets installation, the Saffron program libraries need to be installed into the StreamSets parecl and configured in Cloudera Manager.

The Saffron libraries for StreamSets are delivered as a Redhat Package Manager (RPM) package separate from the ISAA platform installation package. This document assumes you already have a copy of the RPM package, which is called `saffron-streamsets-components-0.6.2.noarch.rpm`.

To install Saffron libraries in StreamSets:

1. Put the RPM package in a convenient location on the Cloudera Manager/StreamSets system.
2. Run the following command:

   ```
   rpm -i saffron-streamsets-components-0.6.2.noarch.rpm
   ```

3. In Cloudera Manager, select the StreamSets Data Collector service and click **Configuration**.
4. Configure the following settings, which are detailed in Cloudera Manager settings for StreamSets libraries:

- Java options
- Data Collector Advanced Configuration Snippet (Safety Valve) for sdc.properties
- Data Collector Advanced Configuration Snippet (Safety Valve) for sdc-security.policy

Installing ISAA license

NOTE on EA deployment: No licensing was involved in the EA program. The following assumption for GA is based on SMB v11 docs. Exactly when in the installation sequence this must be done (after unzip, before install, or after install) is not certain.

To install the license for ISAA, copy the file `license.xml` you received from Intel into your *`installation_path`*, which by default is `/opt/saffron`.

Starting, verifying, and stopping ISAA services

Make sure Docker swarm initialized and worker nodes joined

Before installing or starting ISAA, be sure the Docker swarm has been initialized on the leaders with `docker swarm init`. For more information see the Docker documentation at https://docs.docker.com/engine/reference/commandline/swarm_init/

```
docker swarm init

Swarm initialized: current node (43log8umtrregy0h65882dg5y) is now a manager.
```

Join worker nodes to swarm

In addition, be sure that all desired worker nodes have been joined to the swarm. This a two-step process:

1. On the leader node, generate a join token you will issue on all worker nodes:

   ```
   docker swarm join-token worker
   ```

   <Displays next command to issue on worker with join token, leader's IP address and port>

2. On each cluster node, issue a command displayed by the above command, which is similar to the following example.

Your token value, leader IP anddress, and port will be different from this example.

```
docker swarm join --token SWMTKN-1-2eq8226qve349-m95neiwse0ubxm 172.25.154.224:2377
```

Start by deploying ISAA stack in Docker

Starting ISAA is accomplished by deploying the ISAA stack in Docker.

For all nodes in the cluster, enter follow these steps:

1. Make sure Docker swarm initialized and worker nodes joined.
2. Change to the installation directory:

   ```
   cd installation_path
   ```

3. Deploy the ISAA stack, which is called `saffron`, as specified in the *`installation_path`*`/docker-compose.yml` file.

   ```
   docker stack deploy --compose-file=docker-compose.yml saffron
   ```

   After this command, wait some time to allow all services to start. See Verify running services.

   Result: The ISAA stack is delpoyed in Docker.

Verify running services

To verify that the services are running, enter this command:

```
docker service ls
```

Example output from docker service ls

The **MODE** column shows the word "replicated" when all services have started (even if you did not explicitly setup replication). In addition, look at the **REPLICAS** column. Depending on the number of replicas you have set up, the two numbers *x*/*y* should be equal, which indicates all services have been started. For example, for each service listed below, as shown by 1/1 in the **REPLICAS** column; in this single-node configuration there are no replicas

```
ID                  NAME                          MODE
REPLICAS
IMAGE
PORTS
0p2kdzxoyah8        saffron_report-tpe            replicated
1/1
isaa.bigco.com/saffron/newidr-webservice:0.7.6
*:3000->3000/tcp
7n5stru9rtiq        saffron_user                  replicated
1/1
isaa.bigco.com/saffron/ssp_user:0.7.3.3-1113
7wbwe9gcon0f        saffron_secret                replicated
1/1
isaa.bigco.com/saffron/ssp_secret:0.7.3.3-1113
dt2zoee94r79        saffron_memorystore           replicated
1/1
isaa.bigco.com/saffron/ssp_memorystore:0.7.3.3-1113
iad66932uuve        saffron_auth                  replicated
1/1
```

```
isaa.bigco.com/saffron/ssp_auth:0.7.3.3-1113
in3znw0uqf72        saffron_elasticsearch         replicated
1/1
docker.elastic.co/elasticsearch/elasticsearch:5.5.2
lgjc1kw7er2w        saffron_report                replicated
1/1
isaa.bigco.com/saffron/ssp_report:0.7.3.3-1113
mg8jkdti14i1        saffron_logstash              replicated
1/1                 isaa.bigco.com/saffron/logstash:0.6.1
                   *:12201->12201/udp
mnsqjes1c7hb        saffron_escore2               replicated
1/1
docker.elastic.co/elasticsearch/elasticsearch:5.5.2
mq2frl9sr11n        saffron_escore1               replicated
1/1
docker.elastic.co/elasticsearch/elasticsearch:5.5.2
niiehkbsaofz        saffron_zone                  replicated
1/1
isaa.bigco.com/saffron/ssp_zone:0.7.3.3-1113
qm6zwgk6fngs        saffron_apidoc                replicated
1/1
isaa.bigco.com/saffron/ssp_apidoc:0.7.3.3-1113
r4q994aa6vcf        saffron_application_package   replicated
1/1
isaa.bigco.com/saffron/fcca-application-package:0.1.6-612
*:8080->8080/tcp
r7rknk47494l        saffron_reportstorage         replicated
1/1                 docker.elastic.co/elasticsearch/elasticsearch:5.5.2

te67bau45odt        saffron_license               replicated
1/1
isaa.bigco.com/saffron/ssp_license:0.7.3.3-1113
uaobdhb9c4h7        saffron_gateway               replicated
1/1
isaa.bigco.com/saffron/ssp_gateway:0.7.3.3-1113
*:8081->8080/tcp
xc8kztcffbcs        saffron_escore                replicated
1/1
docker.elastic.co/elasticsearch/elasticsearch:5.5.2
xxhftp5gw8it        saffron_redis                 replicated
1/1
isaa.bigco.com/saffron/redis:3.2
z64bgtvkglxz        saffron_kibana                replicated
```

```
1/1                 docker.elastic.co/kibana/kibana:5.5.0

zone: default
```

Stop ISAA services

To stop the ISAA services:

1. On the leader node whose services you want to stop, enter the following command. `saffron` is the name of the ISAA stack.

   ```
   docker stack rm saffron
   ```

2. Repeat the above command on all leader nodes whose services you want to stop.

Administering and configuring the system

Administering and configuring the system relies on a combination of command-line tools and the ISAA Administrative UI.

Command-line:

- Starting, verifying, and stopping ISAA services
- Configure LDAP
- Reconfigure default ports for ISAA Admin, AML Architect, and AML Advisor
- Configure ElasticSearch master and data nodes

ISAA administrative UI:

- Using the ISAA administrative UI as zone admin or ISAA admin

Files for configuring the ISAA platform via the command line

Command line configuration of the ISAA platform in general consists of editing details included in your *`installation_path`* that define configurable settings.

Location and file name	Function	See Also
`installation_path``/docker-compose.yml`	Written in YAML, this file defines hosts, ports, and other details. It is loaded into Docker after configuration changes. • Docker leader and worker nodes • Ports • ElasticSearch master and data nodes	• Start by deploying ISAA stack in Docker • Reconfigure default ports for ISAA Admin, AML Architect, and AML Advisor • Configure ElasticSearch master and data nodes
`installation_path``/utils/saffronLDAP.conf`	Specifies details of the LDAP service for user access to ISAA	• Configure LDAP • saffronLDAP.conf configuration file

Configure LDAP

Configuring authentication services consists of editing your LDAP details into the *`installation_path`*`/utils/saffronLDAP.conf` file and running a script to load these details into the platfrom.

Unless your LDAP service details change, you need to run this script only once. In addition, after running the script, you can manage LDAP configuration via the ISAA administrative UI; see Edit zone LDAP details.

To configure LDAP:

1. Login to the leader node command line.
2. Change to the ISAA *`installation_path`*`/`**`utils`** directory:

```
cd installation_path/utils
```

3. With your LDAP details ready:

- Edit the `saffronLDAP.conf` file and supply the details of your LDAP service. See saffronLDAP.conf configuration file.
- Rely on the comments and example variables to add your LDAP details.
- Save the file.

1. Run the following script:

```
./setLdapConfig.sh -f saffronLDAP.conf
```

After you run this script, you can administer the LDAP details via the ISAA administrative UI. See ISAA admin--edit zone LDAP details.

Configure ElasticSearch master and data nodes

ElasticSearch is a key component of the ISAA platform that manages searchable indexes of the results of ETL. ElasticSearch is the engine underneath the AML Advisor's LiveSearch feature.

ElasticSearch nodes consist of the following

- Master nodes that execute searches
- Data nodes across which the searchable indexes are distributed for performance

ISAA's ElasticSearch services rely on Docker to distribute searchable indexes to the data nodes. Configuring ElasticSearch consists of designing its master and data nodes and defining this design in the ISAA platform's `docker-compose.yml` file via the command line. For better performance, you can also setup index aliases; see

> **Note on EA deployment:** The Early Adopter customer's Docker installation was at lower version that could not automatically deploy ElasticSearch's searchable indexes to the data nodes, which is triggered by the `docker-compose.yml` file's `deploy` keyword shown below. This was contrary to the orignal design of the system, which is discussed in Design ElasticSearch master and data nodes.

Design ElasticSearch master and data nodes

Your ElasticSearch configuration depends on your network design and functions you assign to the connected computers. Only one possible configuration, the design discussed here of the ElasticSearch configuration includes:

- The minimum required two master nodes
- One data node

The *`installation_path`*`/docker-compose.yml` file includes blocks of parameters specific to ElasticSearch. Below is the pertinent snippet of the file, which is annotated below.

```
...
escore:
   .........................
    deploy:                  #Uncomment in Docker SWARM deployment model
      mode: replicated
      replicas: 2
      endpoint_mode: dnsrr
      resources:
        limits:
          cpus: '2'
          memory: 8G

escore-data1:
  .........................
    deploy:                  #Uncomment in Docker SWARM deployment model
      mode: global
      endpoint_mode: dnsrr
...
```

Line number	Description
2	`escore`: block that defines the master nodes
4	`deploy`: keyword for distribution of ElasticSearch via Docker. This definition needs to be uncommented when deploying via the Docker swarm.
6	`replicas`: two master nodes
8	`resources` of two CPUs (line 10) and 8G RAM each (line 11)
13	`escore-data1`: block that defines a single data node
15	`deploy`: keyword for distribution of ElasticSearch via Docker. This definition needs to be uncommented when deploying via the Docker swarm.
16	`mode: global` indicates a data node.

ElasticSearch index aliasing for performance

Above a 64K threshold number of records, the performance of ElasticSearch indexes degrades. To avoid this problem, you can set up index aliases, which essentially distribute the records across several systems.

Creating aliases involves defining this threshold, creating a map between the ISAA resources space and the various indexes, and configuring the aliases with the ElasticSearch alias REST API.

A shell script for automating the creation of these indexes is detailed in manage-indices.sh script for automated creation of ElasticSearch aliases for performance.

Load the ElasticSearch configuration

To load the ElasticSearch configuration, restart the ISAA services in Docker. See Start by deploying ISAA stack in Docker

Reconfigure default ports for ISAA Admin, AML Architect, and AML Advisor

For background, see Default ports for AML Architect, AML Advisor, and ISAA admin.

To use different ports for these interfaces:

1. Login to the leader node command line.
2. Edit the *installation_path*/docker-compose.yml file and find the following lines:

```
...
ports:
    - "8081:8080"
...
```

3. Change the administrative port number 8081 and the user interfaces' port number 8080 to the desired new port numbers.
4. Stop and then restart the ISAA services. See Starting, verifying, and stopping ISAA services.

Using the ISAA administrative UI

You can do many in administrative tasks by using ISAA's administrative interface. To access the interface, in your web browser, go to the following URL:

http://*your_leadernode_hostname*:*port*/admin

Example with default port number:

http://isaa.bigco.com:8081

Difference between ISAA admin and zone admin

The ISAA system has a hierarchy of administration:

- The ISAA administrator can create new Docker *zones*, edit LDAP details, and other general functions
- A zone administrator can manage the details of specific zones and create new *spaces* in those zones.

Login as ISAA administrator

To login as the ISAA administrator:

1. With your browser, go to the URL for administration. See Using the ISAA administrative UI interface.
2. Do *not* check the **Login in as zone admin** checkbox.
3. Enter the ISAA administrator's username and password.
4. Click **Login**.

Result: You are logged in as the ISAA administrator and all defined zones are displayed.

ISAA admin--create new zone

To create a new zone:

1. Login as ISAA administrator.
2. Click **Create new zone**.
3. Enter the name of the zone.
4. Enter the ID of the zone.
   <NOTE on EA deployment: Although creating new zones via the administrative UI was part of the EA deployment, the mechanisms under the UI to use a new zone were not implemented>
5. For Auth Type, select either **NONE** or **LDAP**.
6. Click **Create** to create the zone, or **Cancel** to discard it.

Result: The new zone is created and its details are displayed.

ISAA admin--turn zone off

Turning a zone off does not remove the zone. It makes the zone unavailable for use.

1. Login as ISAA administrator.
2. From the displayed list of zones, click the name of the zone you want to disable.
3. Click **Turn Zone Off**.

Result: The zone is disabled and its status set to **OFFLINE**.

ISAA admin--edit zone LDAP details

To edit LDAP details of a zone via the ISAA administrative UI:

1. **Prerequisite:** You must have already configured your LDAP details via the command line. See Configure LDAP.
2. Login as ISAA administrator.
3. In the displayed list of zones, click the zone whose details you want to edit.
4. Click **Edit**.
5. If the **Auth Type** is not LDAP, you cannot edit the LDAP details. To change the auth type, scroll to find the **Auth Type** select list and select **LDAP**.
6. Enter the details for the LDAP service for the zone. See list of fields below. For information, compare them to the saffronLDAP.conf configuration file
7. Click **Save Changes**.

LDAP fields:

- LDAP username
- LDAP password
- LDAP URI
- Group Base Attribute
- Group Base DN
- User Base Filter
- User Base DN
- Certificate

Result: The LDAP details for this zone are saved.

Login as zone admin

Note on EA deployment: The system was configured with a single zone called `default`. Creating additional zones was not a feature of the EA deployment.

To administer the default zone:

1. With your browser, go to the URL for administration. See Using the ISAA administrative UI interface.
2. On the login screen, click the checkbox **Login in as zone admin.**
3. Enter your own username and password, not the ISAA administrator username and password.
4. For zone name, specify `default`.
5. Click **Login**.

Result: You are logged in as zone admin and the details about your zone and spaces are displayed.

Zone admin--create new space

To create a new space in the default zone:

1. Login as zone admin.
2. Click **Create New Space**.
3. Specify the name of the new space.
4. Specify the ID of the new space. <NOTE on EA deployment: Although creating new spaces via the administrative UI was part of the EA deployment, the mechanisms under the UI to use a new space were not implemented.>
5. Click **Create** to create the space or **Cancel** to discard it.

Result: With **Create**, the new space is created and listed on the page.

Zone admin--remove a space

1. Login as zone admin.
2. In the list of spaces, click the name of the space you want to remove.
3. Click **Remove Space**.
4. Click **OK** to remove the space or **Cancel** to preserve it.

Result: With **OK**, the space is removed.

Basic Docker administration commands

For reference, also see Summary of Docker commands in this guide.

Joining a worker node to the Docker swarm

To join a worker node to the cluster:

1. Login to the leader node.
2. Generate a join token with the following command. This displays the next command with token that you need to run on the worker node.

   ```
   docker swarm join-token worker
   ```

3. Take a copy of the displayed output from above.
4. Login to the worker node.
5. Enter the command that was displayed by the `docker swarm join-token worker` command shown above.

For an example, see the Docker documentation at https://docs.docker.com/engine/swarm/join-nodes/#join-as-a-worker-node.

Removing a worker node from the Docker swarm

You might need to remove a Docker worker node from the swarm for maintenance or other reason.

To remove a worker node from the swarm:

1. Login to the worker node.
2. Enter the following command:

   ```
   docker swarm leave
   ```

Other Docker management commands

Docker has many other management commands that are not documented here.

See the Docker documentation at https://docs.docker.com/engine/reference/commandline/docker/.

Programming StreamSets ETL – *ISAA AML Architect Guide*

Programming aspects of ISAA are beyond the scope of installation and administration. Details about developing StreamSets processors for ETL are in the *ISAA AML Architect Guide*.

Reference

Preinstallation preparation checklist

Deployment planning and installation preparation as a checklist.

What	Done?
Pre-install Cloudera	
Pre-install and configure StreamSets as a parcel in Cloudera Manager	
Pre-install Docker on all nodes	
Obtain license for ISAA	

CentOS or RedHat installed on all nodes. SELinux disable on all nodes.	
DNS names for nodes	
Time synchronized on all nodes	
Firewalls programmed: • No firewall among cluster nodes • Appropriate rules from cluster to intranet • Appropriate rules for ingestion resources	
LDAP details ready	
Username for installation granted Docker privileges and disk write access	
Installation path writeable by installing username	

Summary of Docker commands in this guide

Function	**Command**
Initialize Docker on leader nodes	`docker swarm init`
Join worker nodes to Docker swarm: 1. On a leader node, the `docker swarm join-token worker` command displays the command to issue on the worker nodes, including the *token_string* and the IP address and port of the leader node . 2. On a worker node, you run the `docker join` command displayed by `docker join-token worker` in #1, above.	1. On leader node: `docker swarm join-token worker` <output of command to run on worker node> 2. On worker nodes: `docker join --token` *`token_string leadernode_ip_address:port`*
Promote worker node to leader, where *`hostname`* is the *FQDN* of the worker node.	`docker promote` *`hostname`*
Start ISAA platform services	1. `cd` *`installation_path`* 2. On a single line, no space before or after dashes: `docker stack deploy --compose-file=docker-compose.yml saffron`
Status of ISAA platform services	`docker service ls`
Stop ISAA platform services	`docker stack rm saffron`

saffronLDAP.conf configuration file

For background, see Configure LDAP.

```
# Keep as-is unless default password was changed. Provide Saffron's admin
password, defaul it 'administrator'
ADMINPASSWORD=administrator
# Should be updated: Provide FQDN for LDAP Server: ldap.company.local
LDAPSERVER=ldap.company.local
# Should be updated: LDAP protocol: ldap or ldaps
LDAPPROTOCOL=ldap
# Should be updated: LDAP Port: 389 LDAPS: 636 or custom: 38972
LDAPPORT=389
# Should be updated: LDAP Bind account
LDAPBINDACCOUNT=admin
# Should be updated: LDAP Password for Bind account
LDAPBINDPASSWORD=admin
# Should be updated: LDAP search BaseDN for Bind account: i.e.
"DC=company,DC=local"
LDAPBASEDN="dc=company,dc=local"
# Should be updated: LDAP search GroupBaseDN for Saffron Group: i.e.
"cn=Saffron,ou=Groups,DC=company,DC=local"
LDAPGROUPBASEDN="cn=Saffron,ou=Groups,dc=company,dc=local"
# Should be updated: LDAP type OpenLDAP: 'open'. WindowsAD: 'win'
LDAPTYPE=open
# Keep as-is. StreamSets ETL LDAP role mapping: groupBaseDn:admin
STREAMSETSROLE=groupBaseDn:admin
```

Cloudera Manager settings for StreamSets libraries

For background, see Installing Saffron libraries in StreamSets parcel in Cloudera Manager.

- Java options:
  `-Xmx64g -Doverrun.reader.read.limit=3000000`
- Data Collector Advanced Configuration Snippet (Safety Valve) for sdc.properties:
  `parser.limit=40000000`
- Data Collector Advanced Configuration Snippet (Safety Valve) for sdc-security.policy:

> Be careful with line 1 and line 6, which include the path `/opt/saffron/`. `/opt/saffron/` is the default installation directory for the ISAA platform. If you chose to install ISAA in a different directory, change `/opt/saffron` to your *`installation_path`*. For background, see Installation path writable by installer.

```
grant codebase "file:///opt/saffron/-" {
permission java.util.PropertyPermission "*", "read,write";
permission java.lang.RuntimePermission "accessDeclaredMembers";
permission java.lang.reflect.ReflectPermission "suppressAccessChecks";
permission java.lang.RuntimePermission "getenv.*";
permission java.io.FilePermission "/opt/saffron/-", "read";
permission java.util.logging.LoggingPermission "control";
permission java.net.SocketPermission "*", "connect,resolve";
};

grant codebase "file:///opt/user-libs/-" {
permission java.util.PropertyPermission "*", "read,write";
permission java.lang.RuntimePermission "accessDeclaredMembers";
permission java.lang.reflect.ReflectPermission "suppressAccessChecks";
permission java.lang.RuntimePermission "getenv.*";
permission java.util.logging.LoggingPermission "control";
permission java.net.SocketPermission "*", "connect,resolve";
};

grant {
permission java.io.FilePermission "/tmp/-", "read,write,delete";
};

grant {
permission java.io.FilePermission "/var/lib/sdc/data/-", "read";
};
```

manage-indices.sh script for automated creation of ElasticSearch aliases for performance

For background, see ElasticSearch index aliasing for performance.

Notes on usage for your particular deployment

To use this script:

- You need to have `curl` installed and defined in your $PATH environment variable, which is included by default in the distribution of the CentOS and RedHat operating systems.
- You need write access to the ISAA spaces where you will create the aliased indexes.
- You need write access on the hosts where you will create the aliases.
- You need to modify the script to match the details of your deployment. This consists of editing a copy of the script to add your details, as deacribed below.

Line	What	Modification
3	value for `SPACE`	Instead of `default`, specify the name of the space where you will create the index aliases.
4	value for `INDICES`	For the supplied values "`customers sars transactions case`", substitue the names of the data resources that have been created for your configuration.
8, 41	`some_username` `some_password`	Supply the username and password that has write access specific to your configuration.

8	`"zone":"default"`	If you are creating aliases in a zone other than `default`, specify that zone name here.
8	`localhost:8080`	If you run this script on a system other than the ElasticSearch server, change `localhost` to the FQDN on that server name.

18	`number_of_shards`
	
	`number_of_replicas`
		Indicate the number of shards and replicas you have defined for ElasticSearch. See Design ElasticSearch master and data nodes

manage-indexes.sh script

```
#!/bin/sh

SPACE="default"
INDICES="customers sars transactions cases"

function create() {

  SLEDGE_TOKEN=$(curl -H "Accept: application/jwt" -H "Content-Type:
application/json" -d '{"username":"some_username",
"password":"some_password", "zone":"default"}'
http://localhost:8080/auth/token)
  echo $SPACE
  curl -v -X POST -H "Authorization: Bearer ${SLEDGE_TOKEN}" -H
"Content-Type: application/json" -d "{\"name\":\"${SPACE}\"}"
http://localhost:8080/spaces/

  curl -X DELETE http://localhost:9200/default:${SPACE}

  alias=""
  for index in $INDICES
  do
     echo "\n\nCreating index $index ... \n"
     curl -X PUT -H "Content-Type: application/json" -d '{"settings" : {
"index" : { "number_of_shards": "5", "number_of_replicas": "1",
"mapping.total_fields.limit": 100000 } } }' http://localhost:9200/$index

     echo "\n\nSetting up mapping for index $index ... \n"
     curl -X PUT -H "Content-Type: application/json" -d
'{"dynamic_templates" : [ { "strings" : { "match_mapping_type": "string",
"mapping" : { "fields" : { "raw" : { "type" : "text" } }, "ignore_above" :
65536, "type": "keyword" } } } ] }'
http://localhost:9200/$index/_mapping/$SPACE

    if [[ ! -z $alias ]]
    then
       alias="$alias , "
    fi
    alias="$alias {\"add\": {\"index\":\"${index}\",
\"alias\":\"default:${SPACE}\"} }"
  done

  curl -X POST -H "Content-Type: application/json" -d "{ \"actions\" : [
```

```
${alias} ] }" http://localhost:9200/_aliases

}

function remove() {
  for index in $INDICES
  do
     echo "\n\nRemoving index $index ... \n"
     curl -X DELETE http://localhost:9200/$index
  done

 echo "\n\n Removing the space from platform...\n"

  SLEDGE_TOKEN=$(curl -H "Accept: application/jwt" -H "Content-Type:
application/json" -d '{"username":"brian", "password":"doesNotMatter",
"zone":"default"}' http://localhost:8080/auth/token)
  curl -X DELETE -H "Authorization: Bearer ${SLEDGE_TOKEN}"
http://localhost:8080/spaces/$SPACE
}

case $1 in
    create)
        create
        ;;

    remove)
        remove
        ;;

    *)
        echo "usage: $0 create|remove" >&2
        ;;
esac
```

ISAA Glossary

The glossary is oriented to *AML* and specific uses of the *ISAA*.

Note: The glossary does not include definitions of many common programming/computing terms, such as HTML, JavaScript, JSON, or R.

AML

Anti-Money-Laundering

anomaly

An unusual pattern that does not conform to expected behavior, sometimes also called an *outlier*. Examples of anomalies include:

- Any sudden and substantial increase in funds
- A substantial increase in the velocity (frequency) of transactions
- A large withdrawal
- Moving money to a bank secrecy jurisdiction.
- Smaller transactions that meet certain criteria might also be flagged as suspicious.

Compare *similarity* and *novelty*.

attribute

A value and a *category* with which the value is associated. Each category can be assigned a type as part of a space definition; the type is not stored in the *resource*. The supported types are string (default) and number. Example:

Category	Value	Attribute
`ocean`	`atlantic`	`ocean.atlantic`

An attribute is sometimes called an "entity".

Compare the programming construct *name/value pair*.

Bank Secrecy Act

US law for combating money laundering and terrorist financing. Codified in Title 31 USC 5311.

BSA

See *Bank Secrecy Act*.

category

A classification of a value. The left hand side of a *name/value pair*. Sometimes a category is a *hypernym*. See also *attribute*.

CDH

Cloudera open source big data software with integrated Apache Hadoop

data drift

A common phenomenon in a machine learning or other AI systems: data changes over time, requiring re-evaluation and perhaps redesign or reprogramming.

destination

StreamSets term for where data that has been transformed via *processors* is sent. The end of a *pipeline*, the sink for ouput from ETL. See also *origin*.

dimension

An ordered relationship in a data continuum, such as time or physical space. A secondary aspect that modifies or constrains another datum. Typically described with the word "by", as in "transactions **by time**" or "outgoing transfers **by location**".

distance

The result of a calculation of the *similarity* between two or more objects. Some kinds of distance are:

- inherent, such as with time or numbers
- geographical distance-based
- feature-based
- psychological

entity

Synonym for *attribute*.

ETL

"Extract, Transform, Load." A process in computing for pulling data out of source systems, changing the data, and making it available to other systems (sometimes by placing it into a data warehouse).

FQDN

Fully qualified domain name of an Internet-connected computer

geocode

Formal notation for the longitude and latitude of a location on the surface of the Earth.

Geocode information is supplied by the GeoNames postal and city downloads available under the Creative Commons Attribution 4.0 License.

hypernym

A word with a broad meaning that more specific words fall under. A superordinate. For example, "color" is a hypernym for the following:

- red
- green
- blue

ingestion

Transferring data from one system to another, usually transforming it for use in the new system. See also *ETL*.

ISAA

Intel Saffron AML Advisor

Know your customer

A key goal of AML involving analysis of patterns of customer behavior to establish common financial characteristics about that customer, such the kinds of transactions in which the customer is likely to engage. By knowing one's customers, financial institutions can often identify unusual or suspicious behavior, termed *anomalies*, which may be an indication of money laundering.

lemmatization

Part of *NLP*, a subtask for processing text with the use of a vocabulary and morphological analysis of words. See also *stemming*.

Lemmatization, like stemming, tries to group related words, but it goes farther than stemming in that it tries to resolve ambiguity by grouping words by their word sense, or meaning, not by their specific grammatical form. The same word may represent two meanings—for example, “wake” can mean "to wake up" or a “funeral”.

name/value pair

In programming, a data structure that assigns a value to a variable. The name of the variable is similar to a classification or *category* for the value.

The left-hand-side is the name. The right-hand side is the value.

The name is often a *hypernym*, a superordinate of the value.

Arrays of name/value pairs are often combined to form a *namelist*, which is useful in *Named Entity Recognition*.

namelist

Programming construct for input or output of whole groups of variables, or input of selected items in a group of variables, usually in the form of an array. It specifies a group name to list the variables and arrays belonging to that group.

NER

Named Entity Recognition. Part of *NLP*, a subtask of information extraction that seeks to locate and classify named entities in text into pre-defined *categories* such as the names of persons, organizations, locations, expressions of times, and so on.

NLP

Natural Language Processing. Some terms in NLP include:

- *hypernym*
- *lemmatization*
- *NER*
- *stemming*

novelty

A previously unnoticed observatiion of a pattern in the data not originally included or accounted for by *processors*. Distinct from *anomaly*. The novel pattern is typically added back to the data transform processors to account for the previously unobserved pattern and thus remove the novelty. Compare *similarity* and *anomaly*.

origin

StreamSets term for where particular input data comes from, a data source. The start of a *pipeline,* which ends in a *destination*.

outlier

Synonym for *anomaly*.

path

In machine learning, a probability path is designed for humans who require a deep understanding of advanced probability for their research or applied use in statistics, biology, operations research, mathematical finance (such as *AML*), engineering, and other disciplines.

In topology, a path is a continuous mapping, with an initial point, a final point, and the space of continuous functions between them. In a topological space *X*, a path is a continuous function *f* from the unit interval *I* = [0,1] to *X*. *f* : *I* ? *X*. The initial point of the path is *f*(0) and the terminal point is *f*(1).

In graph theory, a path in a graph is a finite or infinite sequence of edges which connect a sequence of vertices which, by most definitions, are all distinct from one another.

See also *signature*.

pipeline

StreamSets term for a communications/transformation channel for incoming data. With an *origin* and a *destination*, a pipeline includes discrete *sta ges* that run *processors* to perform a particular change (or "transformation") on the incoming data.

processor

A defined programatic function that transforms incoming data, included as a *stage* in a *pipeline*. From StreamSets.

regex

Regular expression, a text pattern matching mask. See https://en.wikipedia.org/wiki/Regular_expression.

resource

A collection of *attributes* and optional structural information. The *origin* of data for a *pipeline*.

Saffron risk score

A measure of the probability of risk based on the distance from the established pattern of customer behavior, based on specific attributes. See Metrics and Scores.

SAR

See *Suspicious Activity Report*.

segment

An ordered list of *attributes* or other segments. Segments are identified by a label, which is a string.

signature

A mathematical expression that quantifies a *path*, an evolving or time-ordered sequence of events, parameterized by a continuous variable.

similarity

The state of “likeness” between two or more objects expressed by a mathematical formula. The formula is a quantification of the degree of similarity, which is called *distance*. See also *anomaly* and *novelty*.

space

In a Docker multi-tenancy deployment, a *zone* containing spaces is a segregated area for protecting and isolating processes and data for specific purposes and specific groups of users.

In analogy with a physical apartment building with many tenants, a zone is a single, locked apartment. The zone/apartment is further subdivided into individual rooms, one per person (or group of users). The rooms an analogy for Docker spaces, which protect data specific to that group of users.

stage

A discrete, identified portion of a *pipeline* where *processors* transform incoming data. From StreamSets.

stemming

In linguistic morphology and information retrieval, stemming is the process of reducing inflected (or sometimes derived) words to their word base or root form, which is generally a written word form. Example: "send" is the stem of:

- send
- sending
- sent

See also *lemmatization*.

Suspicious Activity Report

After a suspected incident of money laundering or fraud, financial institutions must file a SAR report with the Financial Crimes Enforcement Network (FinCEN) of the US government. These reports are required by the United States Bank Secrecy Act (BSA) of 1970.

THOP

THOught Process. A JavaScript program you write for computing results from a Saffron memory store, relying on algorithms you implement to produce meaningful results. These THOPs are packaged into a library you create and load into the *TPE* deployment service for use with the *AML* Advisor.

THOught Process

See THOP.

TPE

Thought Process Engine. *ISAA*'s computing service that processes *THOP*s.

zone

In a Docker multi-tenancy deployment, a zone is a segregated area for protecting and isolating processes and data for specific purposes and specific groups of users.

In analogy with a physical apartment building with many tenants, a zone is a single, locked apartment. The zone/apartment is further subdivided into individual rooms, one per person (or group of users). The rooms an analogy for Docker *spaces*, which protect data specific to that group of users.

Revision history: Draft: ISAA Administrator Guide, with Installation, EA edition

Date	Description
2018-01-30	Inspection session
2017-12, 2018-01	Working drafts for internal reviews