Table of Contents

1. System Overview ........................................................................................................................................... 3
2. Deployment Diagram ......................................................................................................................................... 4
   2.1. Managed Kubernetes Deployment Diagram .................................................................................................. 4
   2.2. Bare-Metal Kubernetes Deployment Diagram .................................................................................................. 5
3. System Requirements ......................................................................................................................................... 6
   3.1. Signals Data Factory ....................................................................................................................................... 6
   3.2. Signals Lead Discovery ................................................................................................................................... 9
1. System Overview
A Signals Lead Discovery environment embodies the following subsystems:

- A client machine for managing the cluster running under Linux Server OS
- A Kubernetes cluster with the ingress controller enabled
  a. If you don’t have a Kubernetes cluster, you should prepare three machines to be used as nodes of a Kubernetes cluster running under Linux Server OS. For more information about the requirements of the node machines, please refer to the Prerequisites section of the PerkinElmer Signals Lead Discovery v2.4 Installation Guide (Bare-Metal Kubernetes) documentation
  b. If you have already had a Kubernetes cluster, please make sure that your Kubernetes cluster is running on version 1.19 or above. For more information about the requirements of the cluster, please refer to the Prerequisites section of the PerkinElmer Signals Lead Discovery v2.4 Installation Guide (Managed Kubernetes) documentation
- A Spotfire Server running under Windows Server or Linux OS.
- A Spotfire Web Player worker node running under Windows Server OS.
- An Oracle or SQL Server database for persistence of the Spotfire Server data.

Note: For pilot environment, it is typically acceptable to consolidate the Spotfire Server, Spotfire Database, and Web Player worker node onto a single Windows Server environment.

The detailed requirements specification for the above servers is outlined below.
2. Deployment Diagram

2.1 Managed Kubernetes Deployment Diagram

The following diagram shows the deployment of Signals Lead Discovery 2.4 in a managed Kubernetes environment.

In the above diagram:

- **sld-metastore-service** microservice will be used to maintain the metadata needed by Signals Lead Discovery. The data will be stored in the backend MongoDB NoSQL database.

- **sld-nginx** microservice acts as a gateway to the internal Signals Lead Discovery APIs.

- Signals Data Factory service components provides RESTful APIs for managing projects, designing metadata, searching for or downloading compounds and assays, and scheduling backend jobs for data transformation and indexing.
• Master Job Service is a back-end service component which provides the capability of scheduling and managing data import and publish jobs. It will communicate with the Spark and ElasticSearch cluster for data processing.

Client components, such as data publishing tool and global search GUI, will communicate with Signals Lead Discovery services to achieve the data publishing, querying and downloading features.

### 2.2 Bare-Metal Kubernetes Deployment Diagram

The following diagram shows the deployment of Signals Lead Discovery 2.4 in a bare-metal Kubernetes environment.

In the above diagram:

- `slm-metastore-service` microservice will be used to maintain the metadata needed by Signals Lead Discovery. The data will be stored in the backend MongoDB NoSQL database.
- `slm-nginx` microservice acts as a gateway to the internal Signals Lead Discovery APIs.
- Signals Data Factory service components provides RESTful APIs for managing projects, designing metadata, searching for or downloading compounds and assays, and scheduling backend jobs for data transformation and indexing.
• Master Job Service is a back-end service component which provides the capability of scheduling and managing data import and publish jobs. It will communicate with the Spark and ElasticSearch cluster for data processing.
• Client components, such as data publishing tool and global search GUI, will communicate with Signals Lead Discovery services to achieve the data publishing, querying and downloading features.

3. System Requirements

3.1 Signals Data Factory

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Recommended Configuration</th>
</tr>
</thead>
</table>
| Operating System     | • Ubuntu 18.04 LTS (64-bit)  
|                      | • Centos 7.x (64-bit)    
|                      | • RHEL 7.x (64-bit)      |
| CPU                  | • For client machine that is used for running the installation and administrative tools, 2 CPUs at minimal, 4 CPUs is recommended  
|                      | • For each of the node machine that is used by the Kubernetes cluster, 8 CPUs at minimal |
| RAM                  | • For client machine that is used for running the installation and administrative tools, 16 GB RAM at minimal  
|                      | • For each of the node machine that is used by the Kubernetes cluster, 64 GB RAM at minimal |
| Disk Space           | • For the client machine that is used for running the installation and administrative tools, 50 GB storage at minimal  
|                      | • For each of the node machine that is used by the Kubernetes cluster, 1 TB storage at minimum |

**Note:** For each of the node machine that is used by a pre-existing Kubernetes cluster:

- **Storage Class prerequisites:**
  - There must be a default storage class available that can dynamically provision volumes
  - It must provide at least 3000 IOPS (Input/Output Operations per second) and 250MB/s throughput for each node
  - 1.75 TB of space must be available to this storage class in the default (3 node) configurations
  - For each additional node, an extra 400GB should be allocated to the storage class

- **Storage class recommendations:**
  - To allow for future expansion of volumes handling the stored data, it is recommended to setup the storage class setting the allowVolumeExpansion field to true as shown in the sample configuration below.
## Requirement

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Recommended Configuration</th>
</tr>
</thead>
</table>

**Note:** For each of the node machine that is going to be used by the new Kubernetes cluster:

- It is required that all cluster machine storage is using Solid State Hard Drives (SSDs). The recommendation is that these SSDs should provide at least 3000 IOPS (Input/output Operations per second) and 250MB/s throughput for the better performance on data importing, publishing and searching
- You should reserve 850GB at minimal on your storage device for storing the Signals Lead Discovery metadata and the imported and indexed assay data
- You should reserve 150GB at minimal on your storage device for storing docker images and runtime data. By default, this data is stored in /var/lib/docker, but the location can change depending on your OS distribution. If you don’t have the sufficient disk space being mounted to the root (/), you need to change the docker daemon directory to the location that is on a mounted storage which has sufficient disk space. For the detailed steps, please check the Docker official documentation: https://docs.docker.com/config/daemon/#docker-daemon-directory
- A small amount of runtime data is stored at /var/lib/rook and this is currently not configurable. Please ensure that at least 1GB of space is available in this file system location

### Storage

High performance direct-attached disk or SAN

**Note:** NAS is not supported for index storage.

### Network Connection

- For the three Kubernetes cluster node machines, they should use the same physical VLAN and IP subnet for the internal communication. Specifically:
  - Following ports should be allowed for inbound traffic for the nodes to communicate with each other: 22, 80, 443, 2376, 2379, 2380, 6443, 8472/udp, 9099, 10250, 10254, 10256, 30000-32767, 30000-32767/udp
  - Following ports should be allowed for outbound traffic for the nodes to communicate with each other: 443, 2379, 2380, 6443, 8472/udp, 9099, 10250, 10254, 10256
<table>
<thead>
<tr>
<th>Requirement</th>
<th>Recommended Configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>▪ It is recommended using machines that have 25Gbps network bandwidth, and that lower networking bandwidth may reduce performance</td>
</tr>
<tr>
<td></td>
<td>▪ For the client machine that is used for running the installation and administrative tools, it should be on a separate subnet and be able to communicate with any of the node machines in the cluster</td>
</tr>
</tbody>
</table>

**Web browser**

**Windows Clients:**

- Google Chrome latest
- Mozilla Firefox latest
- MS Edge latest

**Mac Clients:**

Apple Safari latest
### 3.2 Signals Lead Discovery

#### Software Requirements

<table>
<thead>
<tr>
<th>Category</th>
<th>Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIBCO Spotfire</td>
<td><strong>TIBCO Spotfire Client:</strong></td>
</tr>
<tr>
<td></td>
<td>• TIBCO Spotfire Analyst 10.3.x LTS</td>
</tr>
<tr>
<td></td>
<td>• TIBCO Spotfire Business Author/Consumer 10.3.x LTS</td>
</tr>
<tr>
<td></td>
<td>• TIBCO Spotfire Analyst 10.10.x LTS</td>
</tr>
<tr>
<td></td>
<td>• TIBCO Spotfire Business Author/Consumer 10.10.x LTS</td>
</tr>
<tr>
<td>TIBCO Spotfire Server</td>
<td>• TIBCO Spotfire Server 10.3.x LTS</td>
</tr>
<tr>
<td></td>
<td>• TIBCO Spotfire Server 10.10.x LTS</td>
</tr>
</tbody>
</table>

#### Hardware Requirements

As an extension of Spotfire, Signals Lead Discovery uses the same hardware requirements as Spotfire.

Please refer to the following website for the Spotfire hardware requirements:

http://support.spotfire.com/sr.asp

Please refer to the ‘PerkinElmer Lead Discovery Premium 2.1 System Requirements.pdf’ for the hardware requirements of Lead Discovery Premium modules.

#### Software Requirements

As an extension of Spotfire, Signals Lead Discovery uses the same software requirements as Spotfire.

Please refer to the following website for the Spotfire software requirements:

http://support.spotfire.com/sr.asp

**Note:** Signals Lead Discovery 2.4 does not support MS Internet Explorer.
Please refer to the ‘PerkinElmer Lead Discovery Premium 2.1 System Requirements.pdf’ for the software requirements of Lead Discovery Premium modules.