

# HPC Performance Monitor

## Application User Manual

March 7, 2017

---



# Contents

<b>1</b>	<b>Introduction .....</b>	<b>4</b>
<b>2</b>	<b>System Requirements.....</b>	<b>5</b>
2.1	Client Side .....	5
2.2	Server Side.....	5
2.3	Cluster Configuration .....	6
2.3.1	<i>Bright Cluster Managed Network Topology.....</i>	<i>6</i>
<b>3</b>	<b>HPC-PM Client .....</b>	<b>8</b>
3.1	Installation .....	8
3.1.1	<i>HPC-PM Windows Client.....</i>	<i>8</i>
3.1.2	<i>HPC-PM Linux Client.....</i>	<i>8</i>
3.1.3	<i>License .....</i>	<i>8</i>
3.2	Initial Launch of Application .....	9
3.3	Icons.....	10
3.4	Menus .....	10
3.4.1	<i>File Menu .....</i>	<i>10</i>
3.4.2	<i>Operation Menu.....</i>	<i>11</i>
3.4.3	<i>View Menu .....</i>	<i>11</i>
3.4.4	<i>Help Menu.....</i>	<i>11</i>
3.5	Global Settings .....	12
3.6	Settings .....	14
3.6.1	<i>Servers.....</i>	<i>14</i>
3.6.2	<i>Connections.....</i>	<i>17</i>
3.6.3	<i>Operations.....</i>	<i>18</i>
3.7	Results .....	20
3.7.1	<i>General .....</i>	<i>20</i>
3.7.2	<i>Line Charts.....</i>	<i>22</i>

3.7.3	<i>Bar Graphs</i> .....	23
3.7.4	<i>Core Data</i> .....	24
3.7.5	<i>Raw Data</i> .....	25
3.7.6	<i>Log</i> .....	27
<b>4</b>	<b>HPC-PM Server Agent</b> .....	<b>28</b>
4.1	Installation .....	28
4.1.1	<i>Preparation</i> .....	28
4.1.2	<i>On a Bright Cluster</i> .....	28
4.1.3	<i>On a Non-Bright Cluster</i> .....	29
4.1.4	<i>License</i> .....	30
4.2	Setup .....	31
4.3	Execution.....	31
<b>5</b>	<b>Licensing</b> .....	<b>32</b>
5.1	Client .....	32
5.2	Server.....	33
<b>6</b>	<b>Troubleshooting</b> .....	<b>34</b>
6.1	HPC-PM Client Application.....	34
6.2	HPC-PM Server Agent .....	35
6.3	Known Issues.....	35

## **1 Introduction**

HPC Performance Monitor (HPC-PM) is a client-server application that allows a user to:

- Configure a variety of testing scenarios between any number of remote servers
- Configure and execute RDMA and TCP performance operations on remote servers
- Measure multiple metrics of operations run over a fabric or network
- Monitor server CPU statistics during operation execution

HPC-PM is a valuable tool for a variety of use cases including for diagnostics and engineering purposes. On the surface it can provide a quick visualization of the performance characteristics of a network but at its core it can provide in depth comparisons between multitudes of different network and operation configurations. In addition to the powerful GUI controls on the client side, HPC-PM features cutting edge performance monitoring tools on its server side which were engineered specifically for interfacing with the HPC-PM client.

HPC-PM features optional integration with computing clusters managed by the Bright Computing, Inc. Bright Cluster Manager. This integration includes discovery of all networks configured on Bright Cluster Manager and discovery of all servers, along with their IP addresses, on those networks. Bright Computing, Inc. is in no way affiliated with HPC-PM or Software Forge, Inc. HPC-PM does not include any components of Bright Cluster Manager and merely interfaces with utilities on an existing Bright Cluster Manager installation.

## 2 System Requirements

### 2.1 Client Side

- **Operating System:**
  - Windows 7+
  - Linux: CentOS 7.2, Ubuntu 16.04
- **Software**
  - Any PDF viewer

### 2.2 Server Side

- **Operating System:** CentOS 7.2
  - Must support RDMA packages
- **Software**
  - Mellanox OFED: MLNX\_OFED\_LINUX-3.0-2.0.1:
  - Java Runtime Environment (JRE) 8+
  - lm\_sensors & cryptopp packages
- **CPU:** Xeon 4 core or better
- **Memory:** 8 GB or more
- **RDMA Devices**
  - Mellanox InfiniBand HCAs
  - Mellanox InfiniBand Switches
  - Mellanox RoCE Channel Adapters
  - Mellanox RoCE Switches

## 2.3 Cluster Configuration

### 2.3.1 Bright Cluster Managed Network Topology

Discovery is performed by querying Bright Cluster Manager for its configuration and so all discovered components are based solely on the Bright Cluster Manager configuration and no other configuration (hardware- nor software-based). The documentation of HPC-PM regarding Bright Cluster Manager assumes basic knowledge of Bright Cluster Manager. Refer to the Bright Cluster Manager documentation for more information regarding the specifics of Bright Cluster Manager.

When running HPC-PM over a cluster controlled by Bright Cluster Manager there are four main considerations:

- 1) **External network (externalnet)**: This is shown in the top right of Figure 1. The Bright Cluster Manager head node must be connected to a network that is also accessible by the device running the HPC-PM client.
- 2) **Management network (internalnet)**: This is shown in **red** in Figure 1. This network must be connected to **all** nodes involved in the performance testing scenario as well as the head node. This includes the head node and all compute nodes. During discovery only the nodes connected to the head node via the management network will be discovered. In Figure 1 this includes nodes SM-1, SM-2, SM-3, and SM-4. SM-5 will not be discovered since it is not connected to the management network.
- 3) **Test network (ibnet)**: This is shown in **blue** in Figure 1. This network must be connected to all compute nodes that will be involved in the performance testing scenario. However, not all compute nodes must be part of the same test network.
- 4) **Secondary networks**: Additional management or test networks may be present on the cluster.
  - a. An example of an additional test network is Ethernet 2 shown in **green** in Figure 1.
    - i. Ethernet 2 is not a valid management network because it is not connected to all nodes on the cluster.
    - ii. Ethernet 2 is, however, a valid test network between compute nodes SM-1 and SM-2. Compute node SM-5 is not a valid compute node because it is not connected to a valid management network containing the head node and so it cannot be used in the test network.

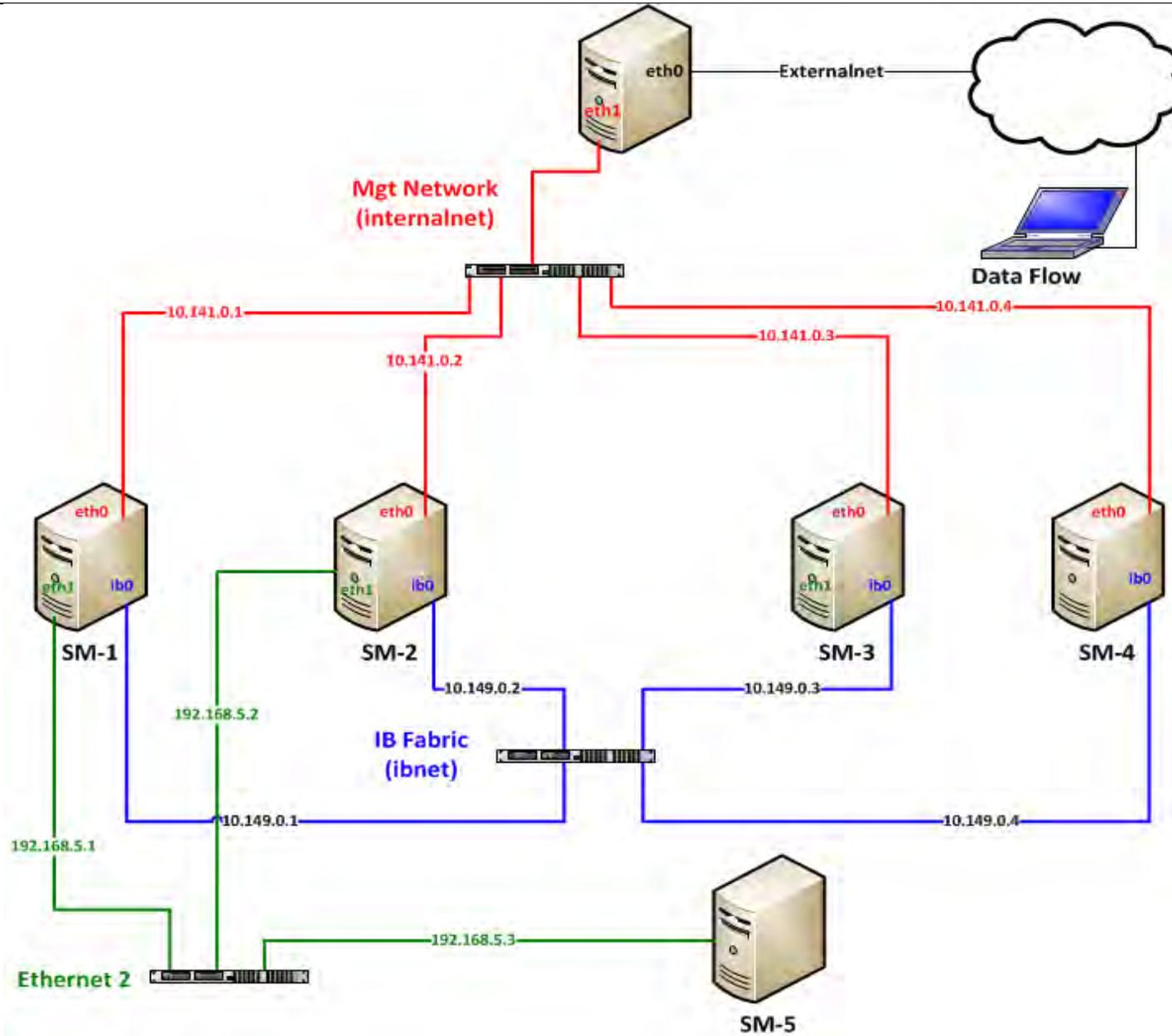


Figure 1: Bright Cluster Managed Network Topology

## **3 HPC-PM Client**

### **3.1 Installation**

#### **3.1.1 HPC-PM Windows Client**

- 1) Obtain the latest HPC-PM Windows installation package.
- 2) Double click the installation package. Follow the on-screen instructions to complete the installation procedure.

#### **3.1.2 HPC-PM Linux Client**

- 1) The command to launch HPC-PM from a Linux system: `/opt/hpc-performance-monitor/hpc-pm`
- 2) The client install location is `/opt/hpc-performance-monitor`.

##### **3.1.2.1 Manual Install**

- 1) Download the latest HPC-PM client RPM.
- 2) Install the client RPM by executing the following command:
  - a. `rpm -ivh <rpm-name>`
  - b. Where `<rpm-name>` is the name of the client RPM downloaded in step 1.

#### **3.1.3 License**

Follow the instructions in the Client Licensing section for installing client license files.

### 3.2 Initial Launch of Application

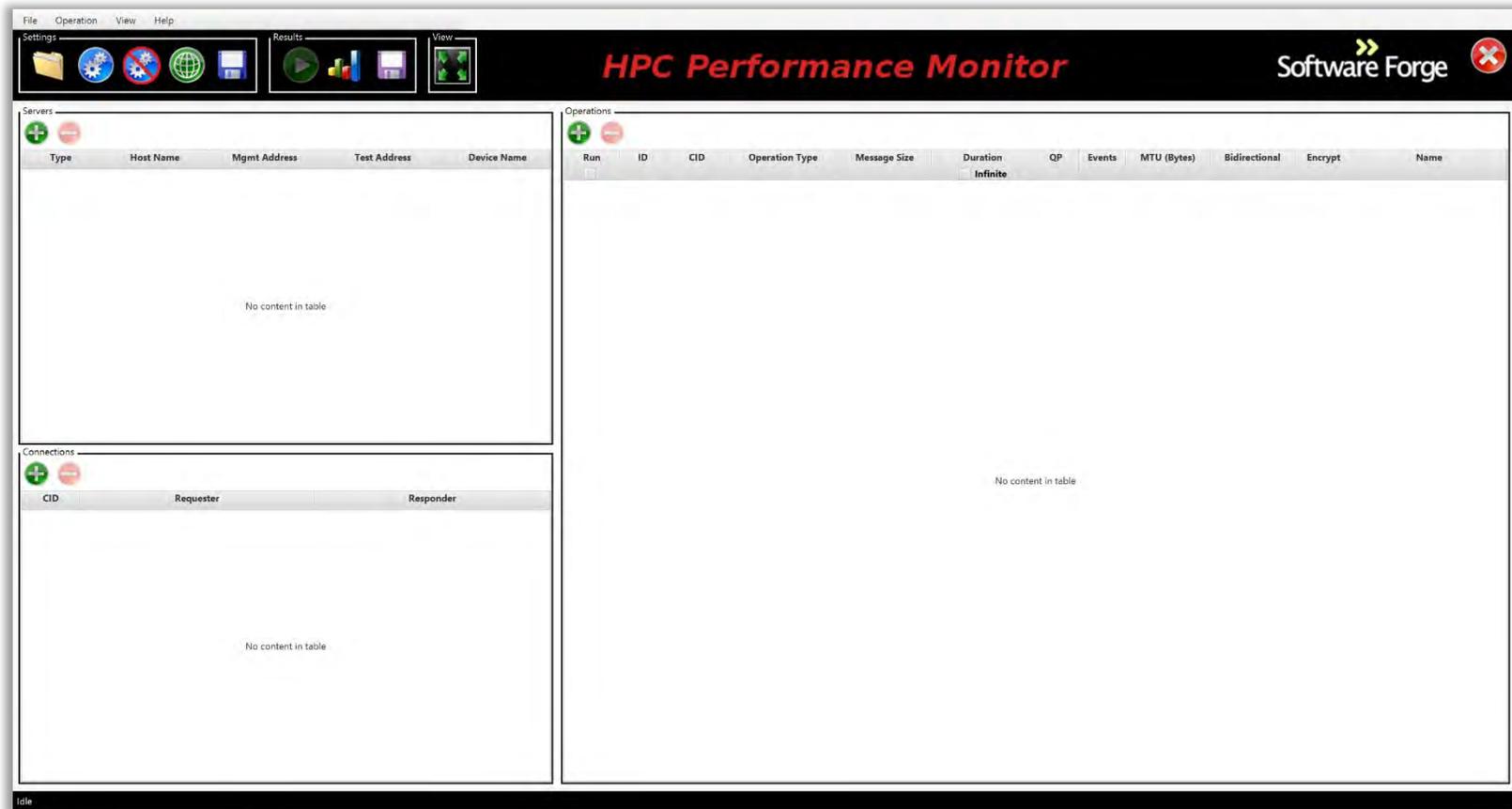


Figure 2: Blank settings screen which is shown on first execution of the application, after settings have been cleared manually or if no settings have been saved.

### 3.3 Icons



Figure 3: Application icon descriptions.

### 3.4 Menus



Figure 4: Application menu.

#### 3.4.1 File Menu

Menu Item	Description
New	Creates a new settings file. See section 3.6 Settings.
Open Settings	Open an existing settings file. The default locations are: <ol style="list-style-type: none"> <li>Windows: C:\Users\<username>\Documents\Software Forge\HPC Performance Monitor</username></li> <li>Linux: /home/&lt;username&gt;/software-forge/hpc-performance-monitor</li> </ol>
Save Settings	Save current settings file.
Save Settings As	Save current settings to a new file.
Save Results	Save the results of the last operations executed. The default locations are: <ol style="list-style-type: none"> <li>Windows: C:\Users\<username>\Documents\Software Forge\HPC Performance Monitor\results</username></li> <li>Linux: /home/&lt;username&gt;/software-forge/hpc-performance-monitor/results</li> </ol>
Exit	Closes the application.

### 3.4.2 Operation Menu

Menu Item	Description
Start	Start the operations using the current configuration.
Stop	Stop the current operations being executed.

### 3.4.3 View Menu

Menu Item	Description
Toggle Full Screen	Toggles full screen and regular windowed mode. You can also exit full screen mode using the Esc key.
Settings	Displays the settings view. Equivalent to selecting the Display Settings icon on the button bar.
Global Settings	Opens the global settings window.
Graphs	Displays the graphs view. Equivalent to selecting the Display Results icon on the button bar.
Open Logs	Opens a file browser which will default to the location of the stored log files on the system. Selecting one of the logs will open the log file in the system's default text editor if available.
Monitor Log	Opens a real-time log window to show the state of the application. Note that only logs which occur after the log window is opened will be displayed.

### 3.4.4 Help Menu

Menu Item	Description
Load License	Opens the license loading window for loading license keys. See section 5 Licensing.
Display Help	Opens this user manual in the system's default PDF reader if available.
About	Opens the global settings window. See section 3.5 Global Settings.

### 3.5 Global Settings

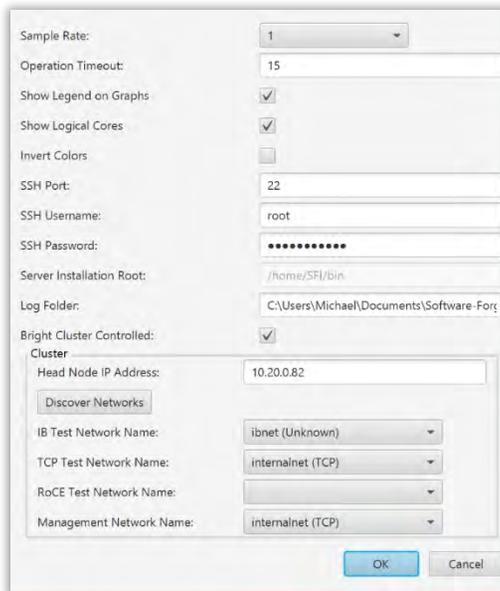


Figure 5: Global settings menu.

- 1) **Sample Rate:** The frequency in seconds that results are sampled. The default value is 1 second. Increasing the sample rate would cause updates in the GUI to appear slower but could reduce load on both the client and server.
- 2) **Operation Timeout:** The operation initiation timeout. After selecting the run operations button, if no operations have started running after the timeout period, the user is notified and the application returns to idle. The default timeout is 15 seconds with one second added per operation that is selected to run (i.e. if 5 operations are selected to run then the total timeout would be 20 seconds).
- 3) **Show Legend on Graphs:** Controls whether or not a legend is displayed for each graph in the results view.
- 4) **Show Logical Cores:** Controls whether or not data for each logical core is displayed in the core load data for hyper threaded systems. If unchecked, the data displayed for hyper threaded systems is the average of the two logical cores that exist on each physical core. This setting affects all Core Load graphs and the Raw Data pane.
- 5) **Invert Colors:** The default color scheme is white on a black background in the Results view. Selecting this setting will invert the color scheme to black on white.
- 6) **SSH Port:** The SSH port on which to connect to the servers. The default is port 22.

- 7) **SSH Username:** The username to use for SSH connections. This must be a valid SSH user that has been added to all servers on the network. The user must be granted sudo rights – see the server notes for details.
- 8) **SSH Password:** The password associated with the specified SSH username.
- 9) **Server Installation Root:** This is the path where the server application is installed on the Linux servers. This path is not modifiable and is set internally to the same path that the server installer uses.
- 10) **Log Folder:** This is the path where the application's logs will be saved to.
- 11) **Bright Cluster Controlled:** Check this box if Bright Cluster Manager controls the cluster being tested. This setting allows the cluster's compute nodes to be automatically discovered and listed in the Servers panel of the Settings view. If this box is not checked, servers must be manually entered in the Servers panel. Toggling this setting from unchecked to checked will cause all server/connection/operation settings to be cleared after a warning.
- 12) **Cluster Settings:** This is where information is entered about a specific cluster. These settings are not applicable if Bright Cluster Controlled is not checked. The IP address of the cluster's head node must be entered here in order for Network Discovery to work.
- 13) **Network Discovery:** In order for the application to be able to identify nodes on the cluster, it needs to know the internal network names on the Bright cluster. All available networks on the cluster may be discovered by pressing the Discover Networks button. This operation will query the Bright Cluster Manager for all networks available to each node on the cluster and will also identify the type of network it is (InfiniBand, RoCE, or TCP.). Once the operation completes, the network name combo boxes will be populated with the discovered networks labeled with their name and type. The application will attempt to determine which network should be used in each combo box based on the network name and type but this can be changed for each network (other than the Management Network) using the combo boxes. The Management Network will be set to the value reported by Bright Cluster Manager and cannot be changed. If the network topology of the cluster changes, then the networks will need to be re-discovered. This only needs to be done if any networks are added or removed, not necessarily if any nodes are added or removed since Node Discovery is a separate operation.

### 3.6 Settings

#### 3.6.1 Servers

##### 3.6.1.1 Manual Server Entry

Manual server entry is enabled when “Bright Cluster Controlled” is disabled in the Global Settings menu. When manual server entry is enabled, servers may be manually added or removed using the add and remove buttons at the top of the server configuration table. When a new server is added to the table, all entries will be blank and they must **all** be filled in manually before the server will be valid for use in a connection.

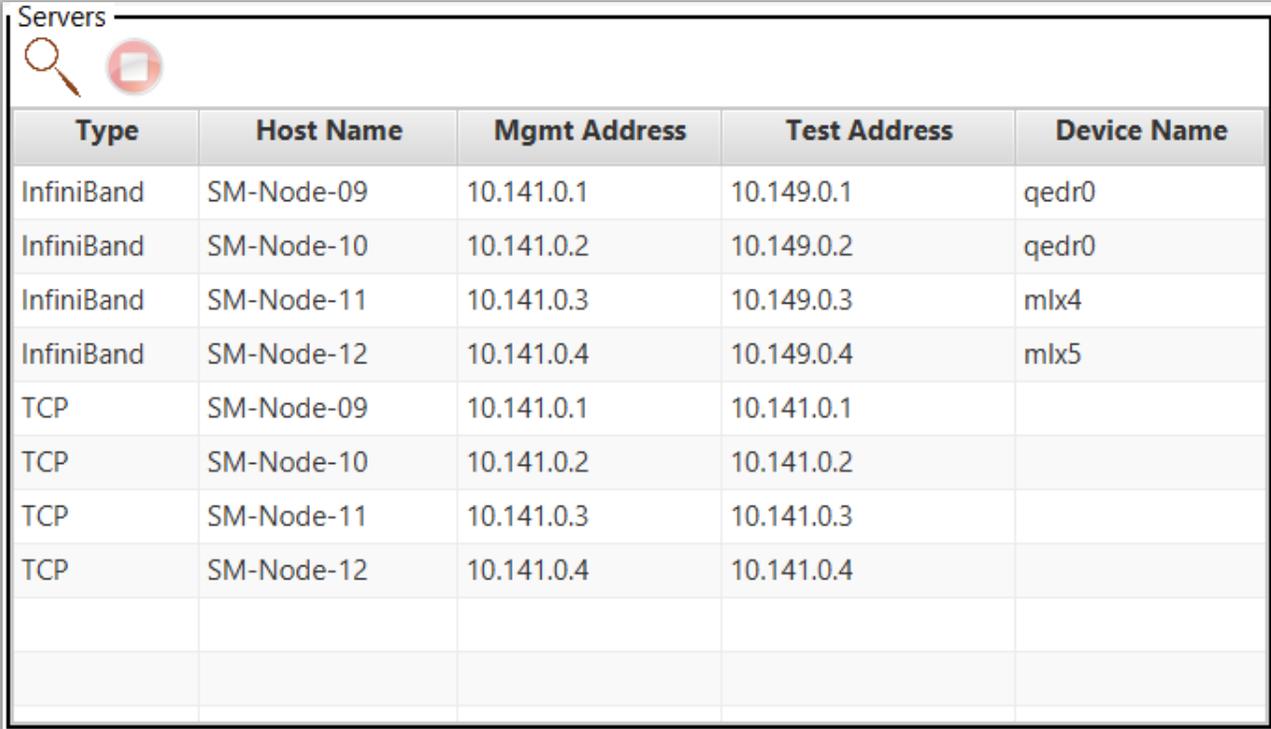
Type	Host Name	Mgmt Address	Test Address	Device Name
InfiniBand	SM-Node-09	10.141.0.1	10.149.0.1	qedr0
InfiniBand	SM-Node-10	10.141.0.2	10.149.0.2	qedr0
InfiniBand	SM-Node-11	10.141.0.3	10.149.0.3	mlx4
InfiniBand	SM-Node-12	10.141.0.4	10.149.0.4	mlx5
TCP	SM-Node-09	10.141.0.1	10.141.0.1	
TCP	SM-Node-10	10.141.0.2	10.141.0.2	
TCP	SM-Node-11	10.141.0.3	10.141.0.3	
TCP	SM-Node-12	10.141.0.4	10.141.0.4	

Figure 6: Example server configuration table with manual server entry enabled.

- 1) **Type:** The type of the network interface matching the test address. This affects filtering when setting up connections. Note that there is no difference between InfiniBand and RoCE types for operations.
- 2) **Host Name:** Add a hostname for the server you are connecting to. This can be the actual hostname or any name you wish as it is purely for identifying the server and has no effect on connecting to the server.
- 3) **Management Address:** The Ethernet IP address for the node on the server you are connecting to. This is typically either the eth0 or eth1 network interface.
- 4) **Test Address:** This is the IP address of the interface to test.
  - a. **InfiniBand:** This is the IPoIB address that has been assigned to the HCA – typically either ib0 or ib1.
  - b. **RoCE:** This is the Ethernet IP address for the RoCE CA – typically this is eth2.
  - c. **TCP:** This is the Ethernet IP address for the Ethernet NIC – typically this is eth0 or eth1.
- 5) **Device Name:** The name of the RDMA network device to use. This is the port device name that is displayed in `ibv_devinfo` output. This setting may not be necessary since a default device will be selected but grants additional control for devices with multiple ports which are loaded as separate interfaces in the OS. This setting has no effect on TCP operations.
- 6) **Important Note:** All addresses for each network a server belongs to will be saved for the currently selected server type. Changing the type of the server will toggle which test address to use by the server. If there has yet to be a test address configured for a specific server type for that server, then the test address will be blank and needs to be entered. Alternatively, an additional server can be created with the same management address but different type and test address. This provides the functionality of running operations with different network interfaces on the same server.
- 7) **Additional Notes:** Be sure to validate IP addresses before running the application.

### 3.6.1.2 Bright Cluster Controlled

When “Bright Cluster Controlled” is enabled in the Global Settings menu, servers may be automatically discovered from the Bright cluster using the Node Discovery button at the top of the table. Note that device names will not be discovered and will need to be entered manually if needed.



Type	Host Name	Mgmt Address	Test Address	Device Name
InfiniBand	SM-Node-09	10.141.0.1	10.149.0.1	qedr0
InfiniBand	SM-Node-10	10.141.0.2	10.149.0.2	qedr0
InfiniBand	SM-Node-11	10.141.0.3	10.149.0.3	mlx4
InfiniBand	SM-Node-12	10.141.0.4	10.149.0.4	mlx5
TCP	SM-Node-09	10.141.0.1	10.141.0.1	
TCP	SM-Node-10	10.141.0.2	10.141.0.2	
TCP	SM-Node-11	10.141.0.3	10.141.0.3	
TCP	SM-Node-12	10.141.0.4	10.141.0.4	

Figure 7: Example server configuration table with "Bright Cluster Controlled" enabled.

- 1) **Node Discovery:** The Discover button starts discovery of servers within a Bright cluster.
  - This operation requires that the head node IP address and any relevant Bright cluster network names have been selected in the Global Settings menu.
  - Note that duplicate host names and management addresses may be displayed in the table if the server contains multiple interfaces discovered on the Bright cluster.

- 2) **Cancel:** The cancel button is enabled while node discovery is in progress. Pressing this button will cancel the currently running node discovery. Any servers which have already been added to the server configuration table as a result of the discovery will remain after cancellation.

### 3.6.2 Connections

Connections can be created between servers listed in the server configuration table in order to specify which servers will be used and in which role for an operation.

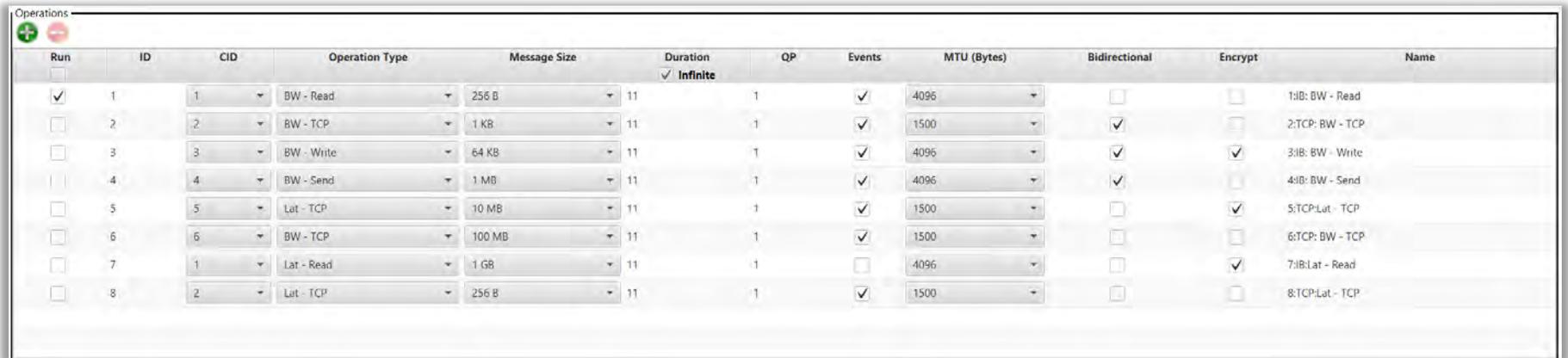
CID	Requester	Responder
1	InfiniBand : SM-Node-1 (10.20.0.51)	InfiniBand : SM-Node-2 (10.20.0.52)
2	InfiniBand : SM-Node-3 (10.20.0.53)	InfiniBand : SM-Node-4 (10.20.0.54)
3	RoCE : SM-Node-5 (10.20.0.55)	RoCE : SM-Node-6 (10.20.0.56)
4	RoCE : SM-Node-7 (10.20.0.57)	RoCE : SM-Node-8 (10.20.0.58)
5	TCP : SM-Node-9 (10.20.0.59)	TCP : SM-Node-10 (10.20.0.60)
6	TCP : SM-Node-11 (10.20.0.61)	TCP : SM-Node-12 (10.20.0.62)

Figure 8: Example connection configuration table.

- 1) **Requester:** This is the server node that will be initiating the operations (the client in the client-server model).
- 2) **Responder:** This is the server node that will be responding to the operations (the server in the client-server model).
  - a) Only valid responders will be shown when defining a connection. A valid responder uses the same server type as it's respective requester as defined in the server creation field.

### 3.6.3 Operations

Operations define the parameters for network traffic that is generated when the Run button is pressed.



Run	ID	CID	Operation Type	Message Size	Duration	QP	Events	MTU (Bytes)	Bidirectional	Encrypt	Name
<input checked="" type="checkbox"/>	1	1	BW - Read	256 B	11	1	<input checked="" type="checkbox"/>	4096	<input type="checkbox"/>	<input type="checkbox"/>	1:1B: BW - Read
<input type="checkbox"/>	2	2	BW - TCP	1 KB	11	1	<input checked="" type="checkbox"/>	1500	<input checked="" type="checkbox"/>	<input type="checkbox"/>	2:TCP: BW - TCP
<input type="checkbox"/>	3	3	BW - Write	64 KB	11	1	<input checked="" type="checkbox"/>	4096	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	3:1B: BW - Write
<input type="checkbox"/>	4	4	BW - Send	1 MB	11	1	<input checked="" type="checkbox"/>	4096	<input checked="" type="checkbox"/>	<input type="checkbox"/>	4:1B: BW - Send
<input type="checkbox"/>	5	5	Lat - TCP	10 MB	11	1	<input checked="" type="checkbox"/>	1500	<input type="checkbox"/>	<input checked="" type="checkbox"/>	5:TCP:Lat - TCP
<input type="checkbox"/>	6	6	BW - TCP	100 MB	11	1	<input checked="" type="checkbox"/>	1500	<input type="checkbox"/>	<input type="checkbox"/>	6:TCP: BW - TCP
<input type="checkbox"/>	7	1	Lat - Read	1 GB	11	1	<input type="checkbox"/>	4096	<input type="checkbox"/>	<input checked="" type="checkbox"/>	7:1B:Lat - Read
<input type="checkbox"/>	8	2	Lat - TCP	256 B	11	1	<input checked="" type="checkbox"/>	1500	<input type="checkbox"/>	<input type="checkbox"/>	8:TCP:Lat - TCP

Figure 9: Example operation configuration table.

- 1) **Run:** This specifies whether the operation is enabled to be run when the Run Operations button is pressed. Selecting the Run checkbox in the column header will select all operations.
- 2) **ID:** This is a user assigned identifier so that the operations can be sorted in any way the user wishes. Click on the ID column header to sort.
- 3) **CID:** This is the Connection ID which specifies which connection to use for an operation. Entries in the drop down menu correlate with valid connections in the connections configuration table.
  - When a new operation is created, only the CID field of the newly created row may be modified. In order to configure an operation, a CID must be selected. Once this is done, the Operation Type field will become modifiable and any invalid operation types will be filtered out based on the selected CID (for example, only TCP operation types will be selectable if the chosen CID refers to a TCP connection. After selecting the Operation Type, the remaining fields will become unlocked and will be populated with the default settings of the selected operation type.
- 4) **Operation Type:** This specifies the metric to be measured for the operation in addition to the network operation type. The supported metrics are latency and bandwidth. For InfiniBand and RoCE connections, each metric can be tested using one of the three RDMA operations – Send, Read or Write.
- 5) **Message Size:** This is the size of the messages that will be sent over the network during the test.

- 6) **Duration:** This is the time in seconds an operation will run for. Infinite allows you to run in Demo mode until the user presses the stop button.
- 7) **QP:** The number of queue pairs to use in the operation. This has no effect for TCP operations.
- 8) **Events:** Whether to use events when retrieving work completions from the completion queue. When disabled, busy polling will be used. Using events will likely result in slightly lower bandwidth but reduced CPU utilization. Busy polling will likely result in 100% CPU utilization. This has no effect on TCP operations since TCP does not have such an event system.
- 9) **MTU Size:** Maximum Transfer Unit as per the IBTA Specification. Allowable sizes are 256, 512, 1024, 2048 and 4096 but Channel Adapters are not required to support all of the sizes.
  - a. Choosing TCP when defining an operation type will defer the MTU size option to the hardware's settings (in the majority of cases this will be 1500 bytes).
  - b. For RDMA operations, the configured MTU cannot be higher than the active port MTU of the Channel Adapters. If an MTU is configured here which is higher than the active port MTU, then, when the operation is run, the MTU will be lowered automatically to the active port MTU. This will be reflected in the logs but not in the client GUI.
- 10) **Bidirectional:** Specifies that the operation will be run bidirectionally. This results in traffic being sent in both directions simultaneously across two network connections for the specified server connection and the resulting performance metric results will be added together.
- 11) **Encrypt:** Specifies that the data used in the network operation will be encrypted prior to each transmission using the AES-256 encryption algorithm. Note that the higher the message size for an operation, the longer each encryption will take which will be reflected by 0 bandwidth while the encryption algorithm is running on the server.
- 12) **Name:** This is the name of the operation which makes it easier to identify operations from the results. After defining an operation, this field contains an auto-generated string concatenating the Operation ID, Connection Type, and Operation Type fields. This field can be changed by the user if desired.

### 3.7 Results

#### 3.7.1 General

The results screen, shown in Figure 10, displays the results of the selected operations in real time. The screen is divided into 3 areas, each of which can display one of 12 different panes which are selectable via a drop down combo box, shown in Figure 11, in the upper left corner of each pane. Each area can also be resized by dragging the vertical or horizontal separator bar between them.

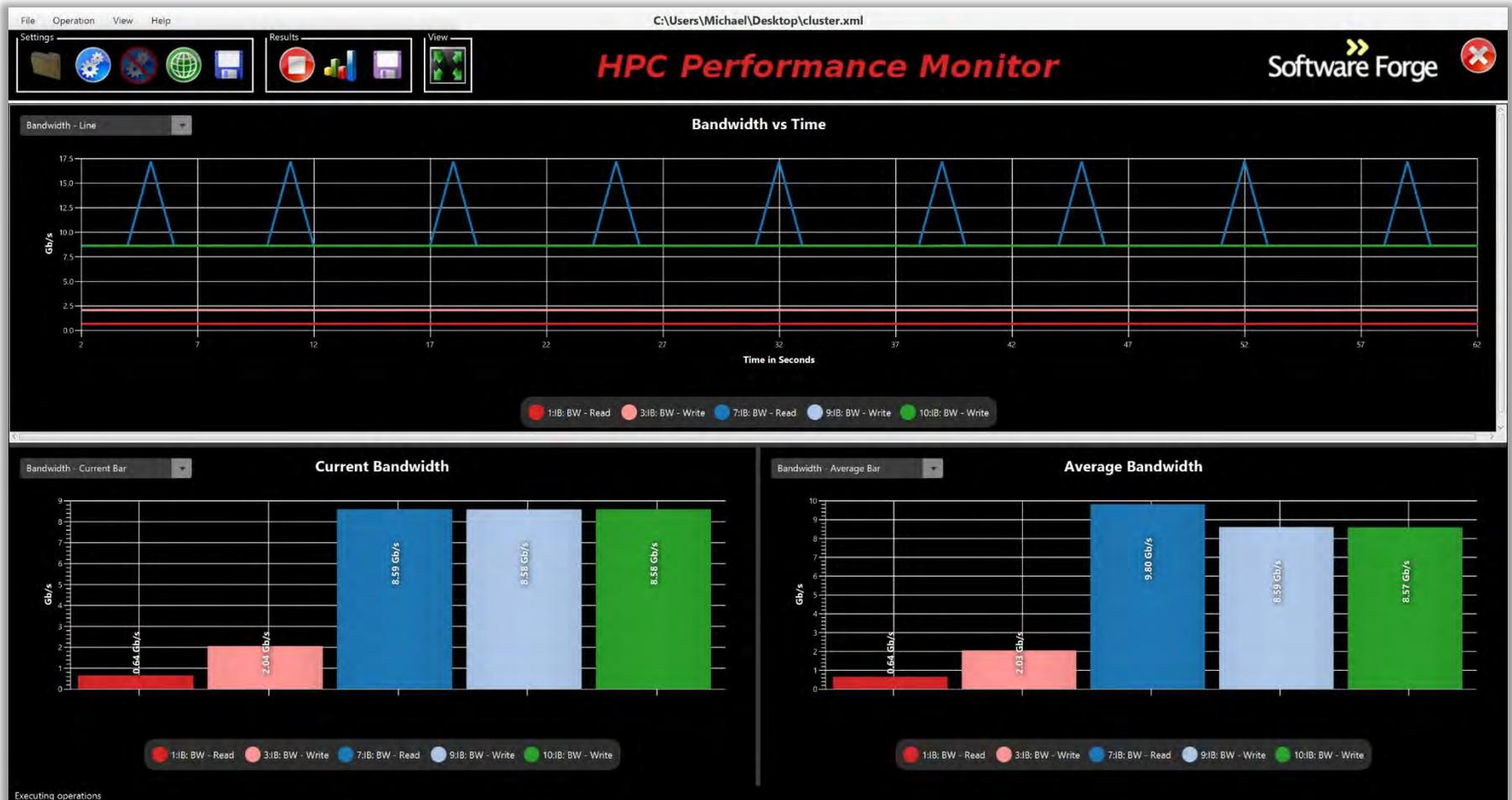


Figure 10: Example results screen.

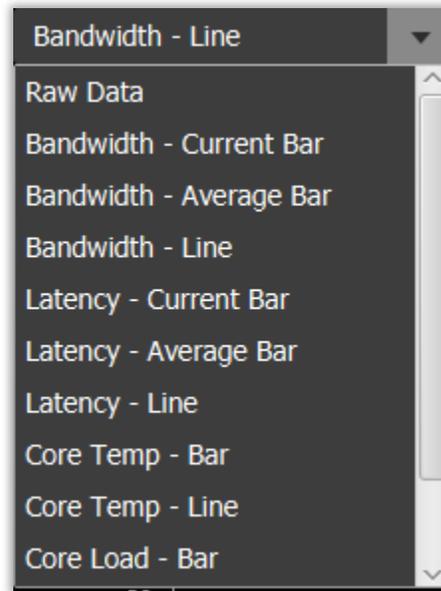


Figure 11: Results pane selection drop down menu.

Each performance metric (bandwidth and latency) has its own line chart and two bar graphs which can be displayed. One bar graph contains the real-time data (“Current” bar graph) while the other contains average data since the start of operation execution (“Average” bar graph). In addition, core temperature and core load metrics can be displayed on a line chart or bar graph for each server. The remaining pane types are the raw data pane which displays results in tabular format and the log pane to display the application logs.

### 3.7.2 Line Charts

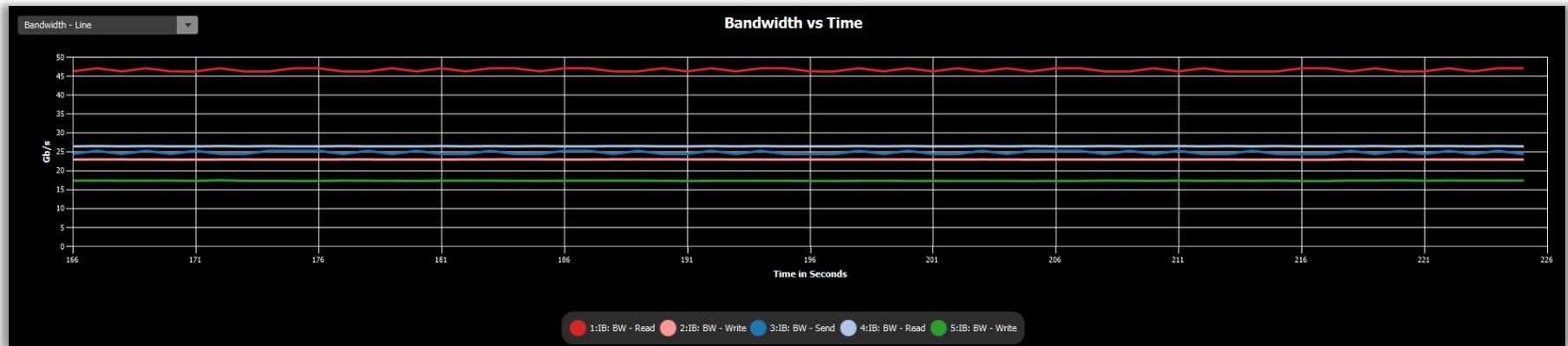


Figure 12: Example bandwidth line chart. This graph shows bandwidth in gigabits per second (Gb/s) over time. The equivalent latency line chart displays latency in milliseconds (ms) over time

- The legend is created using the name of each operation configured in the Operations section of the Settings view and can therefore be customized by the user.
- The user can zoom in on the graph to obtain more details by scrolling the middle mouse wheel.
- Moving the cursor over a data point will cause the value of the metric being measured at that data point to be displayed in a tooltip.
- Clicking on a line chart and dragging the cursor left or right will allow scrolling horizontally to allow for viewing past data. Note that if operations are currently running, each data update will move the line chart view port to the new value making scroll back effectively disabled while operations are running.
- **Note:** if two operations are producing near identical values, it may be difficult to differentiate the two colors in the line graph. However, you can always view the bar graphs and the mouse rollovers to see the details or examine the values in the raw data pane. In this case, it may be more effective to view the resulting data in a bar graph and/or the raw data pane.

### 3.7.3 Bar Graphs

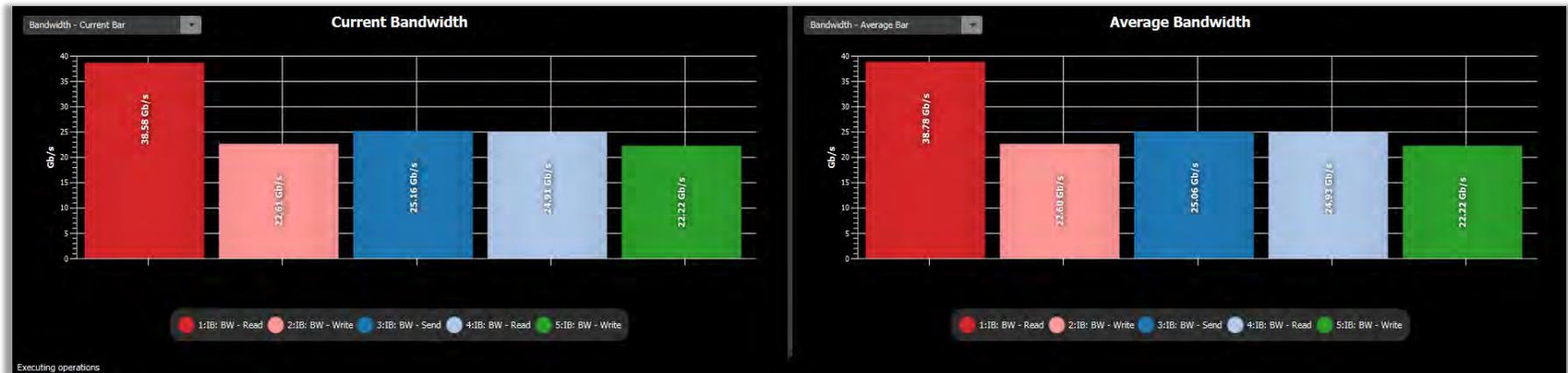


Figure 13: Example current and average bandwidth bar graphs.

- The Current bar graph shows the performance metrics in real time. The current bar graphs will not display any data if operation are not running including after a set of operations has completed.
- The Average bar graph shows the average performance metrics since the start of the last operation execution. They will be reset the next time the run button is selected.
- Any of the bars can be hovered over in order to view all of the details regarding the operation as shown in Figure 14.

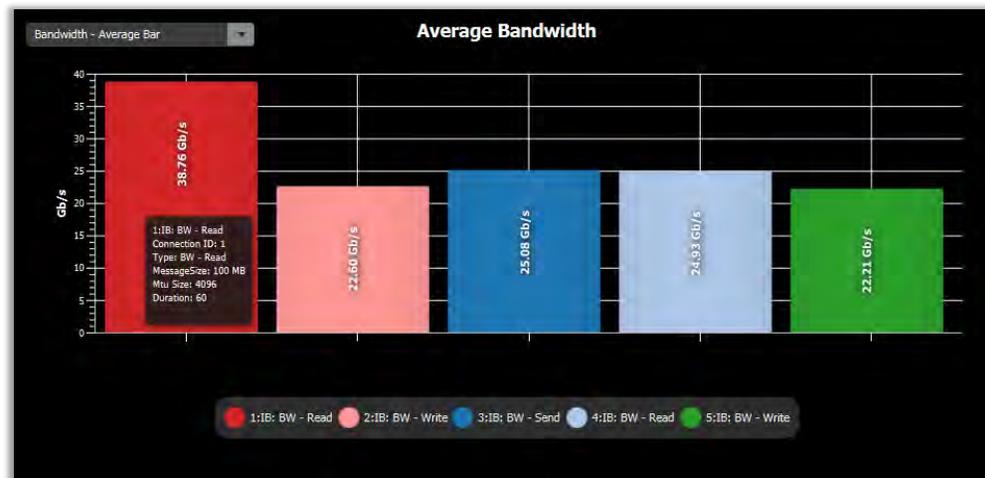


Figure 14: Example bar graph tool tip.

### 3.7.4 Core Data

Core Data consists of temperature and load metrics for processors on the remote servers. Each of these can be displayed as a line chart or bar graph. Data for each server is plotted separately and the server to be displayed can be toggled via a drop down in the upper right corner of the pane.

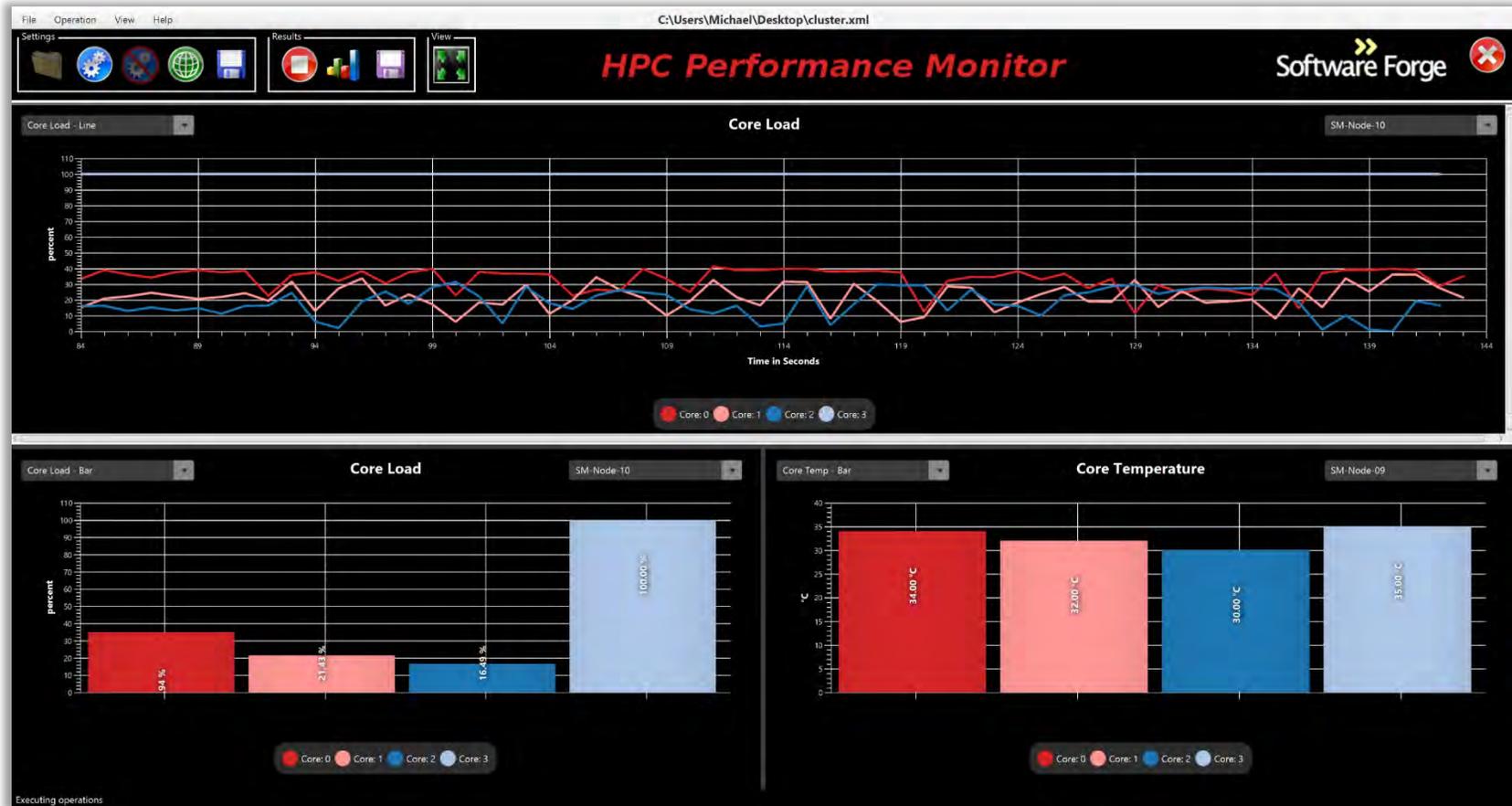
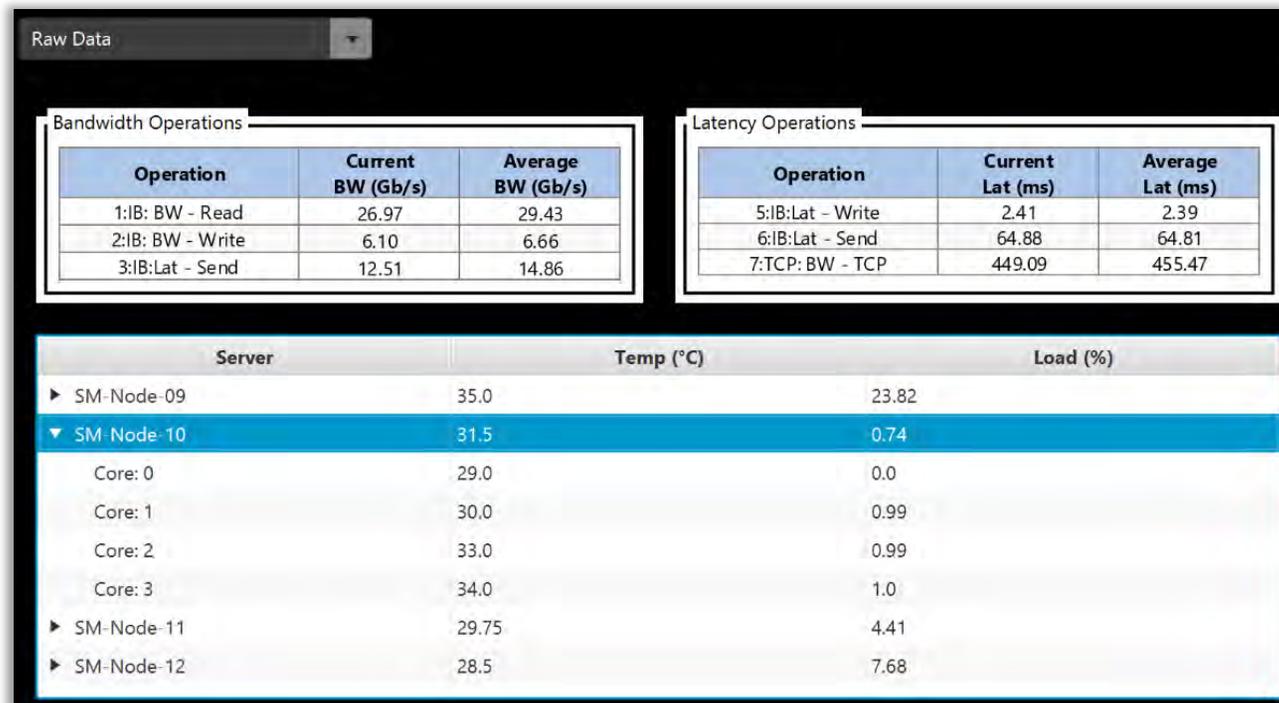


Figure 15: Example core data results.

### 3.7.5 Raw Data

The Raw Data pane shows all of the information which can be viewed in the various charts in a tabular format. Temperature and load data can be expanded to show the values from each core or logical core if applicable. When collapsed, the value displayed is the average of all the cores on a server.



The screenshot shows the 'Raw Data' pane with a dropdown menu set to 'Raw Data'. It contains two summary tables and a main data table.

**Bandwidth Operations**

Operation	Current BW (Gb/s)	Average BW (Gb/s)
1:IB: BW - Read	26.97	29.43
2:IB: BW - Write	6.10	6.66
3:IB:Lat - Send	12.51	14.86

**Latency Operations**

Operation	Current Lat (ms)	Average Lat (ms)
5:IB:Lat - Write	2.41	2.39
6:IB:Lat - Send	64.88	64.81
7:TCP: BW - TCP	449.09	455.47

**Main Data Table**

Server	Temp (°C)	Load (%)
▶ SM-Node-09	35.0	23.82
▼ SM-Node-10	31.5	0.74
Core: 0	29.0	0.0
Core: 1	30.0	0.99
Core: 2	33.0	0.99
Core: 3	34.0	1.0
▶ SM-Node-11	29.75	4.41
▶ SM-Node-12	28.5	7.68

Figure 16: Example raw data pane.

### 3.7.5.1 Network Performance Tables

- 1) The Bandwidth Operations table shows current and average results from bandwidth (BW) operation types. Operations in this table are designated by the Name field of the Operations pane in the Settings window.
- 2) The Latency Operations table shows current and average results from latency (Lat) operation types. Operations in this table are designated by the Name field of the Operations table in the Settings window.
- 3) If there are no operations of a given type (BW/Lat) then the associated table will not be shown while running operations.

### 3.7.5.2 Core Data Table

- 1) The Core Data Table is an expandable/collapsible tree table that shows temperature and load of a system with varying degrees of scope.
  - a. Temperature and load information for each *server* (averages of all processors in a server if more than 1).
  - b. Temperature and load information for each *processor* (averages of all cores in a processor if more than 1).
    - i. If a server only has a single processor then this scope will not be displayed and the cores will be displayed directly underneath the server scope.
  - c. Temperature and load information for each *core*.
  - d. Load data for each logical core.
    - i. Temperature is not a valid metric for logical cores in hyper threaded systems. When logical cores are shown, their temperature values will be displayed as N/A (not applicable).

**Please Note:** When the Show Logical Cores option in the Global Settings menu is changed, the identities of the cores are reconfigured. For example, if a hyper threaded processor has 4 cores with two threads per core, they will be displayed as cores 0 through 3 when logical cores are *not* shown. If logical cores are shown, they will be displayed as cores 0 through 7 where every other core is a logical core

### 3.7.6 Log

The log is able to be viewed in a pane in the Results window as well as through the View menu (View → Monitor Log).



Figure 17: Example log pane.

## 4 HPC-PM Server Agent

The HPC-PM server agent is meant to be used in conjunction with the HPC-PM client application to execute operations and monitor bandwidth across InfiniBand, RoCE, or TCP-based Fabrics.

### 4.1 Installation

#### 4.1.1 Preparation

- 1) Configure the Ethernet NIC on each server so that the HPC-PM client is able to communicate with the servers. This will be the Management Address.
- 2) **InfiniBand**: Be sure to configure at least one IPoIB interface on each server as this will be used for the HCA Address in the HPC-PM Client.
- 3) **RoCE**: Configure the Ethernet port on the RoCE Adapter to allow all RoCE nodes to be able to communicate with one another.

#### 4.1.2 On a Bright Cluster

##### 4.1.2.1 Scripted Install

- 1) Download the latest HPC-PM server agent RPM and install.sh script to the Bright cluster head node.
- 2) Install the server agent to the head node using the install.sh script. This example installs packages to the default image used by Bright Cluster Manager:
  - a. `./install.sh -b /cm/images/default-image server`
- 3) The install script will detect if you are missing any required dependencies and will attempt to automatically install them to the Bright image. If installation of a dependency fails, it should explain the reason of failure and provide instructions on how to finish installation.

### 4.1.2.2 **Manual Install**

- 1) Download the latest HPC-PM server agent RPM to the Bright cluster head node.
- 2) Install the following packages to the image being provisioned to cluster nodes that will run the server agent. These examples install packages to the default image used by Bright Cluster Manager:
  - a. `lm_sensors`
    - i) `yum --installroot=/cm/images/default-image install lm_sensors`
    - ii) Note: If temperature metrics are not retrieved you may need to execute the `sensors-detect` command on each node.
  - b. `cryptopp`
    - i) `yum --installroot=/cm/images/default-image install cryptopp`
    - ii) Some package repositories do not contain `cryptopp`; in which case `yum` will fail to find the package with the above command. If this is the case, an RPM will need to be obtained (from `pbone.net` for example) and installed with the following command:
      - (1) `rpm --root=/cm/images/default-image -ivh <rpm-name>`
      - (2) Where `<rpm-name>` is the name of the RPM file.
- 3) Provision compute nodes with newly modified image using `cmsh` or `cmgui`. Please see the Bright Cluster Manager Administrator Manual section 5.2.
- 4) Install the server agent by executing the following command:
  - a. `rpm -ivh <rpm-name>`
  - b. Where `<rpm-name>` is the name of the server RPM installed in step 1.
- 5) This will place the server agent files in the `/home/sfi/bin` directory which will be propagated to all compute nodes in the cluster automatically.

### 4.1.3 **On a Non-Bright Cluster**

#### 4.1.3.1 **Scripted Install**

- 1) Download the latest HPC-PM server agent RPM and `install.sh` script to **each node** that will be used for testing.
- 2) Install the server agent using the `install.sh` script:
  - a. `./install.sh server`
- 3) The install script will detect if you are missing any required dependencies and will attempt to automatically install them. If installation of a dependency fails, an explanation of the failure reason should be displayed along with instructions on how to finish installation.

### 4.1.3.2 Manual Install

- 1) Download the latest HPC-PM server agent RPM to each system that will be used for testing.
- 2) Install the following packages to the systems that will run the server agent:
  - a. lm\_sensors
    - i) `yum install lm_sensors`
    - ii) Note: If temperature metrics are not retrieved you may need to execute the `sensors-detect` command on each system.
  - b. cryptopp
    - i) `yum install cryptopp`
    - ii) Some package repositories do not contain cryptopp; in which case yum will fail to find the package with the above command. If this is the case, an RPM will need to be obtained (from pbone.net for example) and installed with the following command:
      - (1) `rpm -ivh <rpm-name>`
      - (2) Where `<rpm-name>` is the name of the RPM file.
- 3) Install the server agent by executing the following command:
  - a. `rpm -ivh <rpm-name>`
  - b. Where `<rpm-name>` is the name of the server RPM downloaded in step 1.
- 4) The server install location is **/home/sfi/bin**.

### 4.1.4 License

Follow the instructions in the Server Licensing section for installing server license files.

## 4.2 Setup

- 1) The HPC-PM client application launches the server agent remotely via an SSH connection. The Linux server must be accessible to the client machine via SSH. The username and password of the Linux account that you wish to use to access the application must be defined in the Global Settings section of the HPC-PM client application.
- 2) The SSH user must have root privileges and must be able to use the sudo command without a password for the client application to launch it successfully. This can be accomplished by adding the user to the sudoers file via visudo and adding a line similar to:
  - a. `sfi ALL=(ALL) NOPASSWD: /usr/sbin/opensm, /home/sfi/bin/hpc-pm-server`
  - b. This line adds sudo privileges for the user 'sfi' with no password for the following applications:
    - i. HPC-PM server agent which is installed in /home/sfi/bin/
    - ii. opensm in /usr/sbin/

## 4.3 Execution

- 1) Since the HPC-PM client application will launch the server agent remotely and configure all operations, there is nothing more to do on the Linux server after the HPC-PM server is installed and the sudo privileges have been set. The server agent does not need to be manually executed.

## 5 Licensing

### 5.1 Client

The HPC Performance Monitor client uses node-locked, expiring licenses. The application will not function without a license. When loading the application for the first time or any subsequent time without loading a license, you will be presented with the error dialog shown in Figure 18.

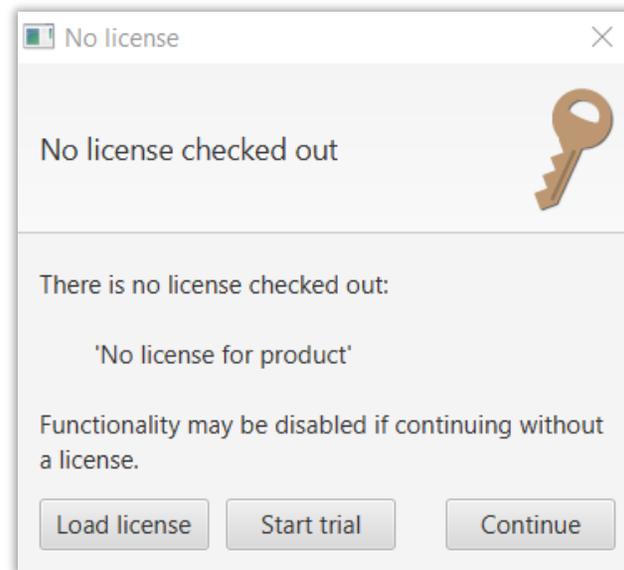


Figure 18 - License error dialog

Clicking the “Continue” button will load the application but features will be disabled until a valid license is loaded. The “Start trial” option may be selected to start a 15 day trial during which time all of the features of the application will be available but they will be disabled once again after the trial period has expired and another trial cannot be started on the same machine. The “Load license” option will bring up the window shown in Figure 19.

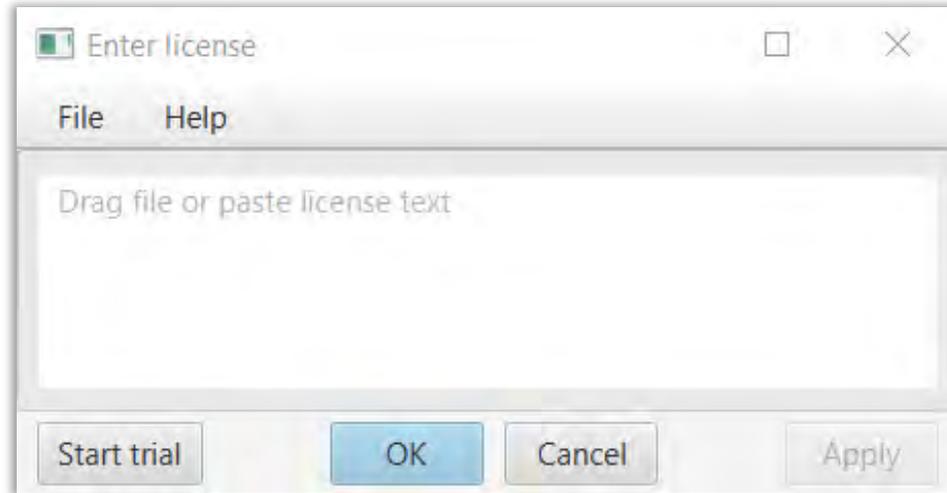


Figure 19 - License load window

In order to load a license, either copy and paste the license file text contents into the text area or drag the license file from the system file explorer to the text area, then click apply to load the license. If license loading is successful, a message will flash across the screen and full application functionality will be enabled. If license loading failed, an error dialog will be displayed. This window can also be opened via the Help→Load License menu.

## 5.2 Server

Each installation of the HPC-PM server agent requires a license file separate from the client license file. As opposed to the HPC-PM client, the HPC-PM server licenses are non-node locked meaning they can be installed on any number of servers. However, HPC-PM server licenses will expire at the same time that the client licenses expire.

In order to install an HPC-PM server license, first install the server agent according to the directions in the HPC-PM Server Agent section. Then place the license file in the **/home/sfi/bin/** directory. Without installing the server license file, the server agent applications will fail to launch and so client-side operations will not function as expected.

License details are subject to change in future releases. Contact Software Forge, Inc. at [sales@soft-forge.com](mailto:sales@soft-forge.com) for license inquiries.

## 6 Troubleshooting

### 6.1 HPC-PM Client Application

- 1) Check the log file for details regarding the operations sent to the HPC-PM Server:
  - a. View the log through the GUI via the View → Monitor Log window or in the log pane of the results view.
  - b. Alternatively view the log contents in an external text editor:
    - i. Use menu option View → Open Logs to determine the log folder.
    - ii. Log file names are time stamped so open the most recent file.
  - c. Search for the keyword “READY”. You should see “READY” for both the responding and requesting servers and you will see READY for responder first. If there is no READY for the responder, you will not see one for the requester.
- 2) Make sure that all fields are completed and **valid** in the Servers, Connections and Operations sections.
- 3) Make sure your Management Addresses are correct. On a Bright Cluster installation make sure you can “ping” each address from the head node (and review Figure 1: Bright Cluster Managed Network Topology). For non-Bright Cluster installations you must be able to “ping” each address from your client machine.
  - a. If your fabric is using a DHCP server, the management addresses on the Linux servers may have changed. Please validate these addresses when loading a previously saved settings file.
- 4) Make sure your HCA addresses are correct and that you can run the “rping” application between them.
- 5) If the operations selected in your current configuration do not run, try loading another configuration that ran previously.
- 6) If you have many operations defined and the application does not run them all, try unselecting all but one operation and then run it again. Enable other operations until you discover which one is causing the problem.
- 7) If the application or the servers do not respond, try restarting the servers and/or the client.
- 8) If you are testing a Bright Cluster, make sure you check Bright Cluster Controlled in the Global Settings and fill in all fields in the Cluster Settings section of the Global Settings window.

## 6.2 HPC-PM Server Agent

- 1) Make sure you can ping the Management Address for all the servers you entered.
- 2) Verify that the user has sudo privileges or run as root.
- 3) Make sure you have set up IPoIB addresses for the InfiniBand HCAs.
- 4) Make sure you can use the following applications between any connected servers with Infiniband or RoCE connections:
  - a. rping
  - b. ib\_write\_bw, ib\_read\_bw, ib\_send\_bw
- 5) While running the client, verify that the opensm, hpc-pm-server, and sfi\_perf executables are running on the server.
  - a. ps -ef | grep open
  - b. ps -ef | grep sfi\_perf

## 6.3 Known Issues

- 1) Text entry fields in settings are not all validated for format (i.e. ensuring numbers are numbers, IP addresses are in the valid format, etc.).
- 2) Running a configuration consisting of many operations across many connections can, on occasion, prevent the user from changing the content of the results panes or from resizing the panes in the results view. Restarting the application will resolve the issue.
- 3) Running operations on a system with a large amount of processing cores longer than the default period of time (60 seconds) can cause the application to become unstable and is not recommended for this beta release.
- 4) Currently no PDF viewer application is packaged with HPC-PM but one is required to launch the HPC-PM User Manual available through the “Help” menu.