4.0.0 Providing High Availability for Stratos

High Availability (HA) is a system design approach and associated service implementation that ensures a prearranged level of operational performance that will be met during a contractual measurement period. In a production deployment it is recommended to include HA features for all the Stratos components to make sure that the PaaS cloud is available to its highest possible level.

HA could be configured according to the Stratos deployment model:

1. Single JVM Deployment Model

In Single JVM deployment model Stratos could accommodate up to 100 cartridge instances. Here the Message Broker, Complex Event Processor (CEP) and Business Activity Monitor (BAM) can be run in the Active/Active mode and Stratos core needs to be run in Active/Passive mode. The following sections explain the configuration process in detail:

a. Message Broker (ActiveMQ)
The following are the ways that ActiveMQ supports HA:

- Shared File System Master Slave
- JDBC Master Slave
- Replicated Level DB Store

b. **Complex Event Processor (CEP)**
The following are the two methods for enabling HA in CEP:

- Distributed Cache Mode Deployment
- Persistence Mode Deployment

c. **Business Activity Monitor (BAM)**
The following are the two methods for enabling HA in BAM:

- Partially Distributed
- Fully Distributed

d. **Stratos Core (Stratos Manager, Auto-scaler and Cloud Controller)**
Stratos core products use Carbon registry for managing their states. Carbon Registry is built into each Stratos core product and could be configured with a Relational Database Management System (RDBMS) for data persistence. Therefore, HA could be enabled for Stratos core by enabling HA for the connected RDBMS and running multiple instances of each product in **Active-Pasive mode using Linux-HA**. The default RDBMS used for Carbon Registry is MySQL. Please see this article for setting up HA for MySQL.

e. **Stratos Load Balancer**
Stratos Load Balancer uses Apache Synapse for implementing session affinity. Currently, Synapse does not support distributed session management. Therefore, Stratos Load Balancer could be run in a distributed environment, but session affinity will not be supported.

2. **Distributed Deployment Model**
In a distributed deployment model Stratos could accommodate up to 1000 cartridge instances. In this approach Stratos core components are deployed as separate products and each needs to be run in Active/Passive mode. The Message Broker, Complex Event Processor (CEP) and Business Activity Monitor (BAM) can be run in Active/Active mode. The following section explain the configuration process in detail:

a. **Message Broker (ActiveMQ)**
   The following are the ways that ActiveMQ supports HA:
   - Shared File System Master Slave
   - JDBC Master Slave
   - Replicated Level DB Store

b. **Complex Event Processor (CEP)**
The following are the two methods for enabling **HA in CEP**:
- Distributed Cache Mode Deployment
- Persistence Mode Deployment
c. **Business Activity Monitor (BAM)**
The following are two methods for enabling **HA in BAM**:
- Partially Distributed
- Fully Distributed
d. **Stratos Manager**
HA could be enabled for Stratos Manager by enabling HA for the connected registry database and running multiple instances of each product in **Active-Pasive mode using Linux-HA**. Currently, Stratos Manager cannot be run in Active/Active mode.
e. **Auto-scaler**
HA could be enabled for Auto-scaler by enabling HA for the connected registry database and running multiple instances of each product in **Active-Pasive mode using Linux-HA**. Currently, Auto-scaler cannot be run in Active/Active mode.
f. **Cloud Controller**
HA could be enabled for Cloud Controller by enabling HA for the connected registry database and running multiple instances of each product in **Active-Pasive mode using Linux-HA**. Currently, Cloud Controller cannot be run in Active/Active mode.
g. **Stratos Load Balancer**
Stratos Load Balancer uses Apache Synapse for implementing session affinity. Currently, Synapse does not support a distributed session management. Therefore, Stratos Load Balancer could be run in a distributed environment, but session affinity will not be supported.

**Impact of Running Stratos in Active/Passive Mode**
Unlike other software systems, a very small downtime of the PaaS solution may not directly affect the applications running on the PaaS unless there is an immediate scaling requirement. The reason is that even if the PaaS solution is unavailable the virtual machines that run the applications will be able to run without a problem. More importantly within a very short period of time a passive instance of Stratos will be able to takeover the failed active instance. Therefore, there is very insignificant impact on running Stratos in Active/Passive mode in a production deployment.