AMELIA

IPsoft's Cognitive Agent

Amelia V3 Platform Design
(version 1.1)

Table of Contents

Document History

Author	Version	Date	Comments	Final Approval?
Randy Schneiderman	0.1	11/15/2017	Initial Document	
Randy Schneiderman	1.0	11/30/2017	Updated for V3	Yes
Randy Schneiderman	1.1	1/18/18	Added NetData and DR sections	

1. Amelia High Level Architecture Design

Amelia is the artificial intelligence platform that can understand, learn and interact as a human would to solve problems. Amelia reads natural language, understands context, applies logic, infers implications, learns through experience and even senses emotions. The diagrams below illustrate Amelia's architecture.

Amelia V3 is designed with multiple shared services, including but not limiting to Administrative Services, Conversation Engine Pods, Integration Services, and a Database Shard Architecture. The concept for this design is to separate very large JVMs, databases, daemons into smaller, faster, more easily managed fragments. The below diagrams show the service architecture and data flow overviews of Amelia V3:

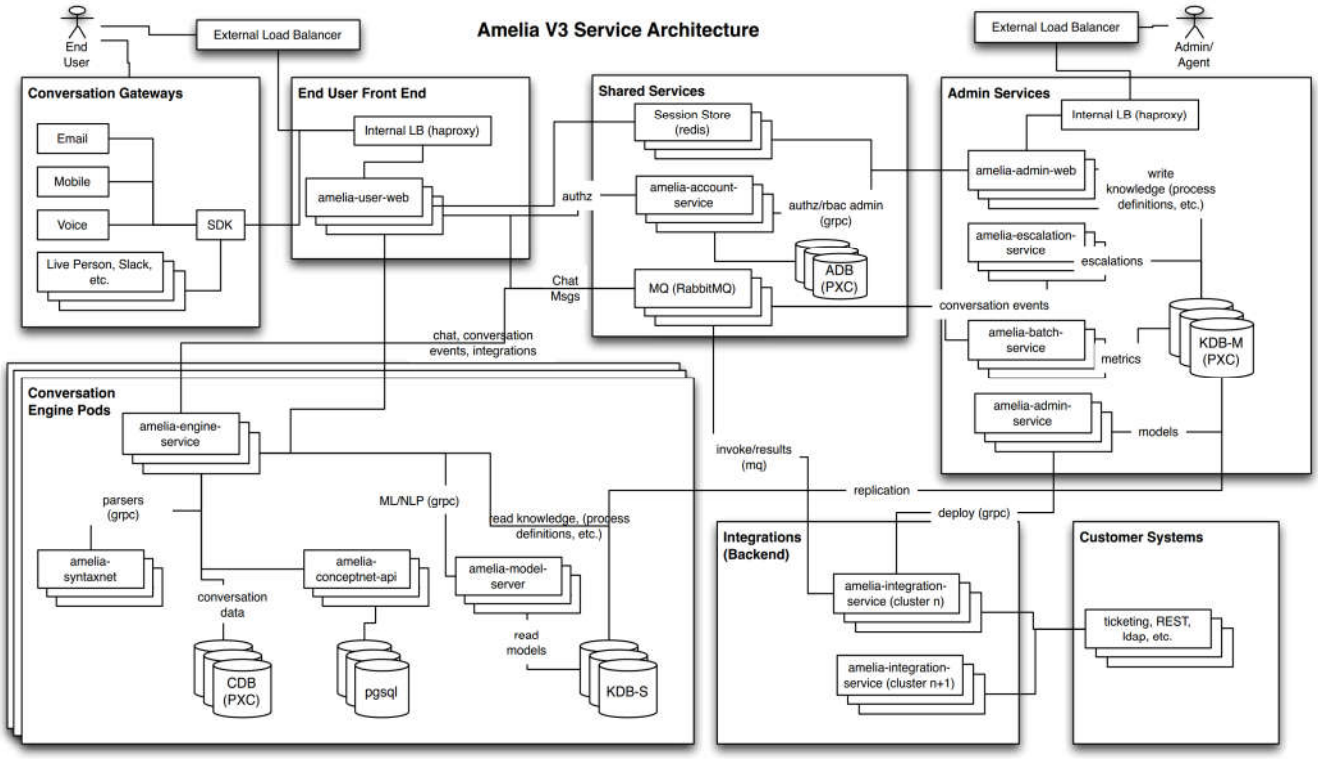


Figure 1: Service Architecture Diagram

The below sections will describe the infrastructure requirements and necessary components of Amelia V3. Some hosts will have multiple middleware components to support Amelia and her functions. Amelia supports clustering both for scaling and high availability. It is recommended that all production environments be clustered with three Application servers and three Database servers. When configured with clustering, there are no single points of failure within Amelia; any failure of a single component should at most result in a few seconds of intermittent faults. Amelia V3 is only supported on LANs and where network split-brain between

nodes is very unlikely. In the event of a network split-brain event, manual intervention may be required and data in-flight may be lost.

1.1 3-Node Cluster Architecture

IPsoft's standard deployment model, a best practice, consists of three identical servers (see Figure 1).. When clustered, there are no single points of failure within Amelia and a failure of any single component should at most incur a few seconds of intermittent errors.

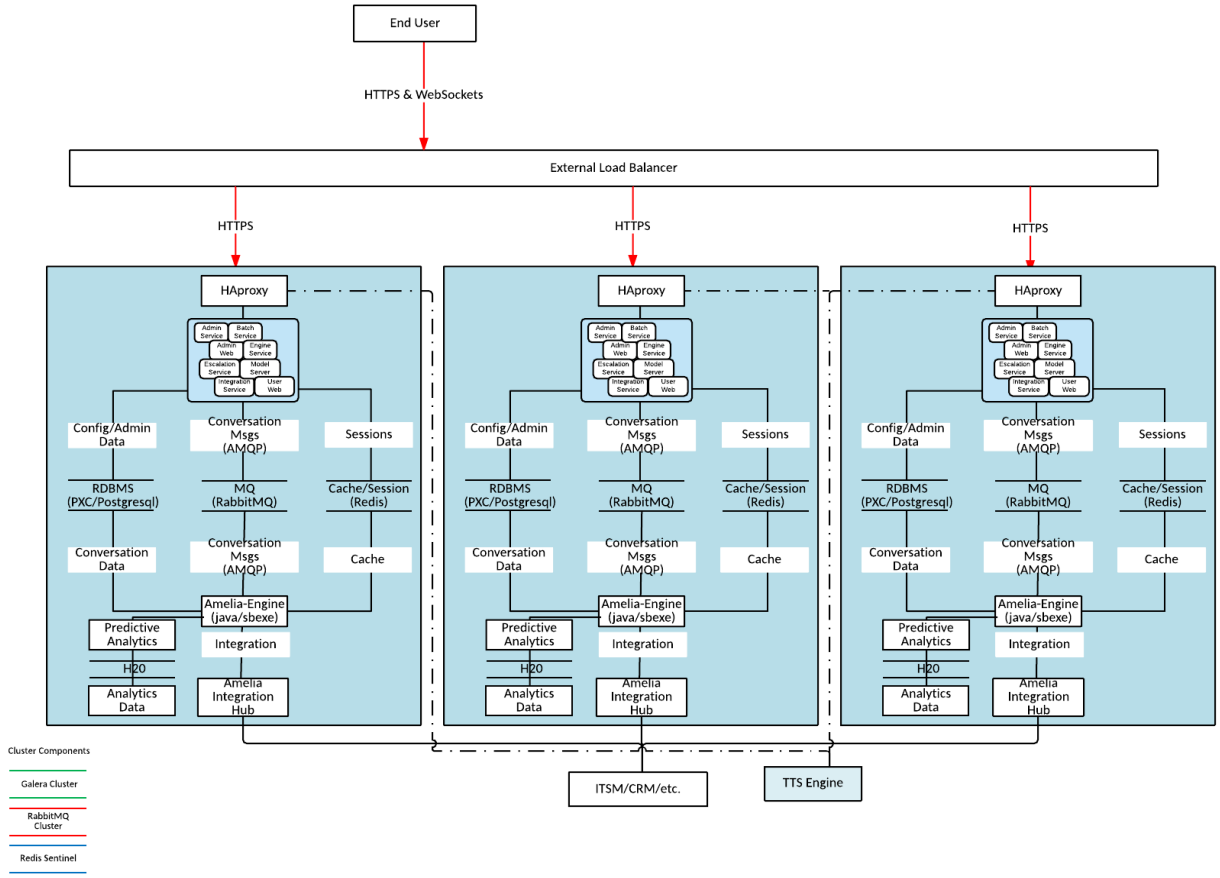


Figure 2: Initial 3-Node Cluster Environment Diagram

NOTE
Connection options — for example, to IPcenter, LDAP/SSO/SAML2, or other technology — should be discussed with IPsoft in the planning process.

1.2 6-Node Cluster Architecture

This configuration splits the application and database services into two network tiers. As Amelia V3 uses a sharding/multiple shared services architecture, the various application and database services can be further split off into additional servers to provide separation of service requirements and for large volume use cases.

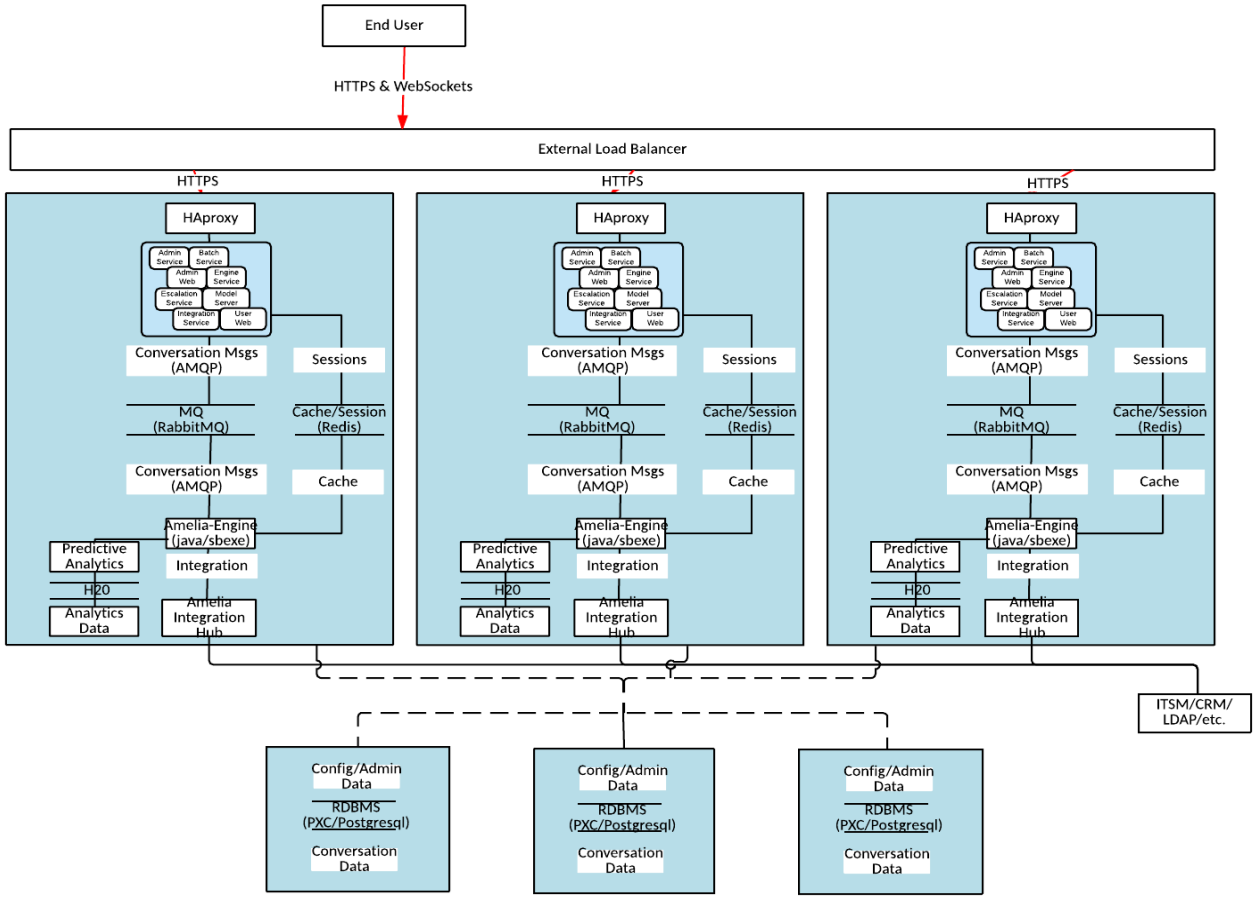


Figure 3: 6-Node Cluster Environment Diagram

NOTE
Connection options — for example, to IPcenter, LDAP/SSO/SAML2, or other technology — should be discussed with IPsoft in the planning process.

1.3 Single Node Architecture

Amelia also supports single host architectures. Single-host configurations are meant for POC/Dev environments.

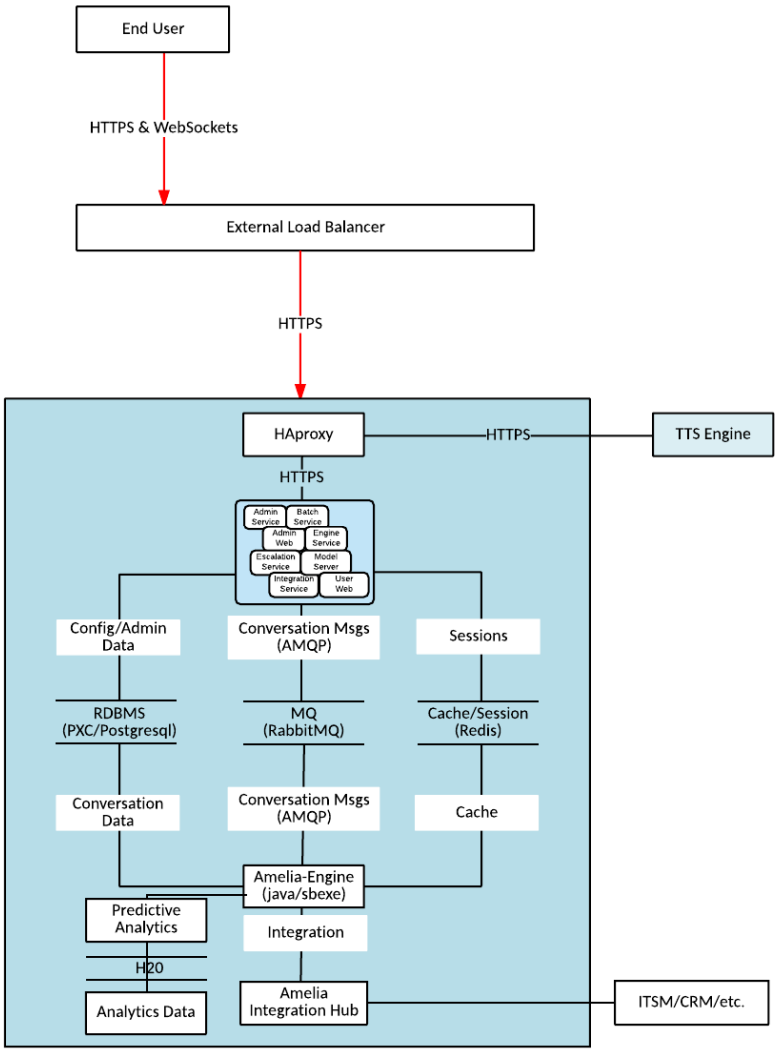


Figure 4: 1-Node Cluster Environment Diagram

NOTE
Connection options — for example, to IPcenter, LDAP/SSO/SAML2, or other technology — should be discussed with IPsoft in the planning process.

2. Deployment Models

Amelia can be delivered in several deployment models, depending on customer/partner requirements. On-Premise deployments will may require encrypted network connectivity between IPsoft and the clients/partners environments for installation of Amelia V3. Interconnectivity can be accomplished with dedicated communication circuits and/or site to site Virtual Private Network ("VPN") tunnels across the public Internet.

IPsoft has developed a self-contained tool called Amelia Deployment Center (ADC) for deployment and configuration of Amelia V3 for remote servers to install Amelia on RHEL 7 family servers, which is used to automate and organize system configuration tasks.

2.1 Amelia Hosted in IPsoft's Cloud

Amelia can be hosted at IPsoft's datacenters worldwide in IPsoft's New York Metro and Amsterdam datacenters. In this model, IPsoft assumes all responsibility for deployment and scaling of Amelia for given clients and partners. This deployment will utilize IPsoft's current hardware comprised of Dell servers, Compellent storage, VMware, and Cisco/Arista networking.

For security purposes, backend technologies and any integrations with Amelia are interconnected utilizing a secure delivery network, typically an IPsec VPN and/or MPLS.

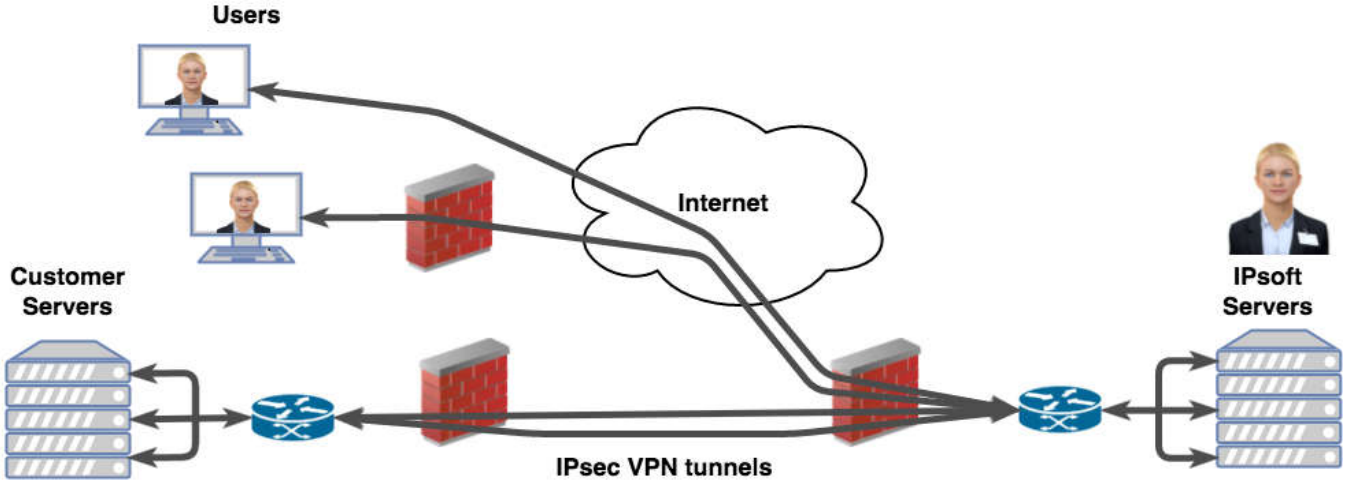


Figure 5. Amelia Hosted in IPsoft's Cloud with IPsec VPN Tunnels

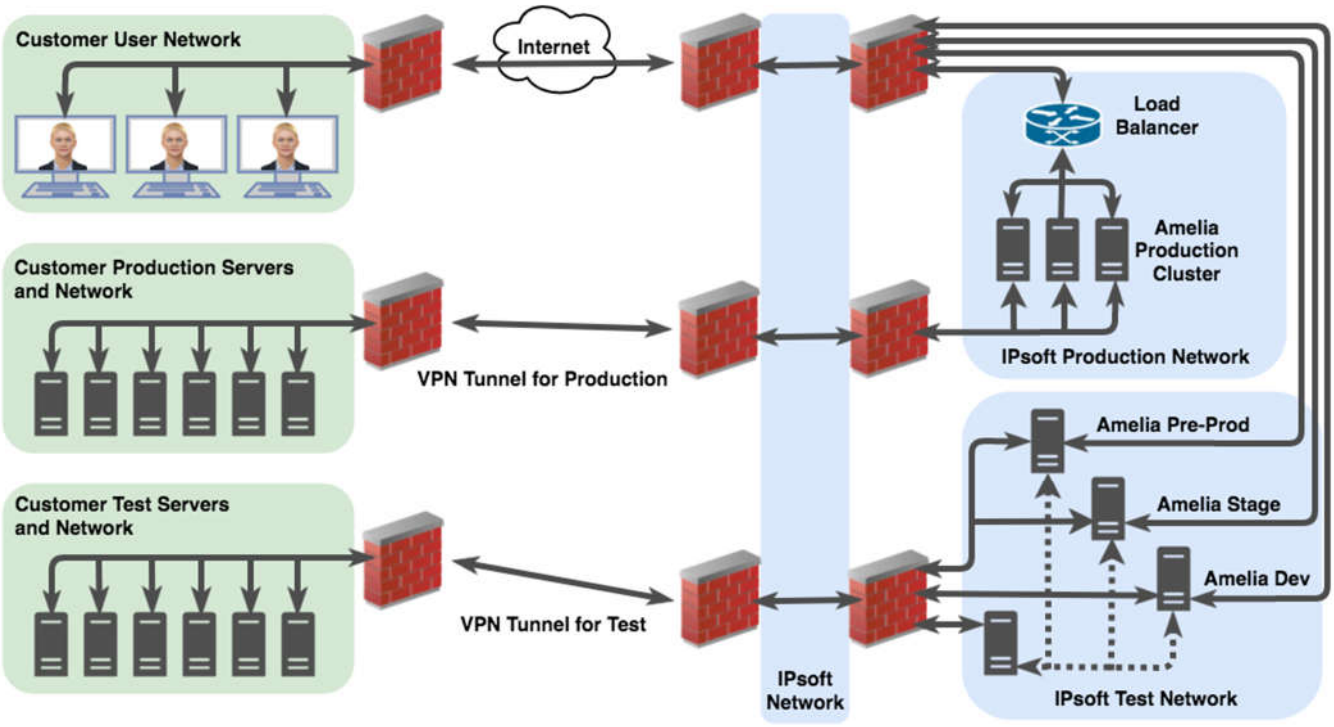


Figure 6. Amelia Hosted in IPsoft's Cloud

Figure 6 shows a conceptual view of how clients and servers are communicating with Amelia in IPsoft-hosted environment.

2.2 CUSTOMER PREMISE / CLOUD

Amelia can also be deployed within a client's/partner's facilities, without IPsoft managing the hardware and network infrastructure. This type of deployment requires the involvement of Client/Partner Architects and IPsoft's Service Design resources to develop a joint architecture for connectivity and sizing.

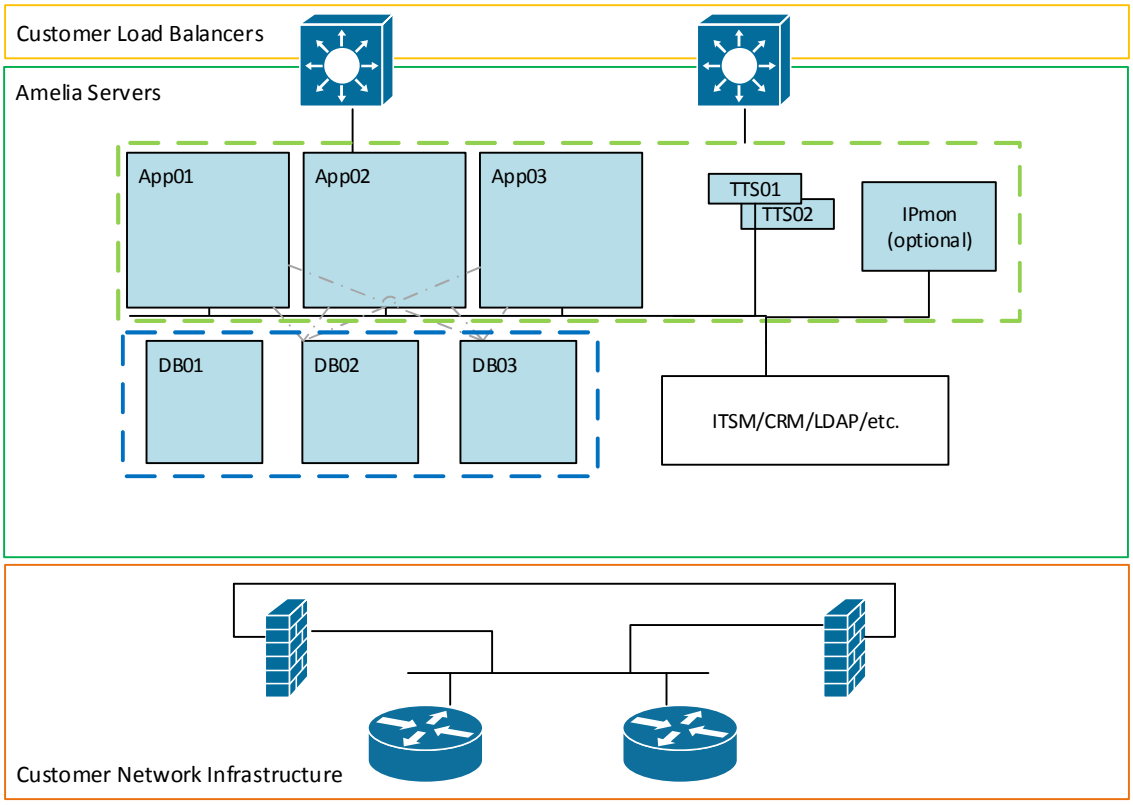


Figure 7. Amelia Hosted in Customer's Cloud

NOTE
Connection options — for example, to IPcenter, LDAP LDAP/SSO/SAML2, or other technology — should be discussed with IPsoft in the planning process.

2.3 Online Deployment

With remote connectivity enabled, IPsoft will use ADC to deploy Amelia to automate the installation of Amelia and her dependencies. Please refer to IPsoft's Amelia Deployment Center (ADC) Guide for in-depth information.

The deployment and release of Amelia to remote server(s) is managed by IPsoft's implementation of ADC. Using job templates and ADC mechanism, IPsoft can show progress and monitor the status of Amelia being deployed.

Figure 8. Online Deployment

Deployments in the public cloud (such as AWS/Azure) are considered on-premise deployments, as the roles and responsibility of the infrastructure is owned by the client/partner. IPsoft has done extensive testing within AWS for all support Operating Systems as described below in Section 3.1

At a minimum, each server should be deployed using the m4.4xlarge sizing; this does not include a backup partition if that is required. Client is responsible for any Elastic Load Balancing configuration and any OS/network security aspects.

As of this writing, Amelia is not supported using containers, for example, Docker.

2.4 Offline Deployment Infrastructure Specifications

At times, remote connectivity will not be possible between IPsoft and client/partner for deployment of Amelia on remote server(s). IPsoft refers the lack of a persistent connection as an Offline Instances. IPsoft will continue to use ADC for deployments and the initialization of all playbooks will be maintained on the server(s) locally within the client's/partner's network.

For offline deployments, IPsoft can only provide specific SLAs based on available connectivity and access.

If considering an Offline deployment of Amelia, topics of discussion should include but are not limited to:

- Backups – How are they performed? Schedule?
- Monitoring – OS, Middleware, Amelia, Web Checks
- Administration – Engineer access to Operating System
- Upgrades – OS, Amelia
- Support – Break/Fix, Root level access
- Deployment Timeline– Procurement of servers/load balancers and/or access limitations

For offline deployments with a pre-existing IPcenter instance, contact IPsoft to discuss options.

Depending on the type(s) of use cases and volume for each use case, Amelia can be scaled up by increasing the number of CPUs, RAM, and disk capacity. If Amelia is configured as a single node, It cannot be scaled-out with additional nodes.

NOTE
For deployments in a virtual environment, it is recommended to enable Memory and Disk reservations for best performance; especially in highly utilized shared infrastructure.

Amelia does not perform well in infrastructures with Memory ballooning and/or CPU scheduling issues.

It is also recommended Memory/CPU Hotplug is enabled to increase CPUs/RAM at any time to scale quickly. Oversubscribed resources are discouraged.

Amelia should be deployed utilizing at a minimum dual Intel® Xeon® Processor E5-2687W v3 (25M Cache, 3.10 GHz) and Solid State Drives (SSD) for primary storage (SAS/SATA Storage can be used for backups)

2.4.1 Clustered Setup

For Production/DR deployments, it is highly recommended to deploy a clustered setup for high availability and scaling for performance, Amelia can be deployed on physical or virtual servers. As of this writing, the following are guidelines on the infrastructure requirements for a clustered Production/DR deployment. These requirements are for each server.

Table 1. 3-Node Clustered Setup Environment Resource Requirements

	Specifications per Host			
Tier	CPU*	RAM (GB)**	Disk Capacity (GB)***	Minimum Number of Hosts
Apps	16	80	1024	3
TOTALS	48	240	3.0TB	3

*Based on underlying server's NUMA settings, multiple sockets may be suggested
**Additional RAM may be required for language pack support & gateways (1GB each pack)
***/apps is the required mount point; LVM and/or XFS file system is highly recommended. Slower disks can be utilized if slow performance is acceptable.

NOTE
The above table are requirements for each node of the cluster.

Table 2. 6-Node Clustered Production Environment Resource Requirements

	Specifications per Host			
Tier	CPU*	RAM (GB)**	Disk Capacity (GB)***	Minimum Number of Hosts
Apps	8	48	100	3
Database	8	16	2048	3
TOTALS	48	192	6.3TB	6

*Based on underlying server's NUMA settings, multiple sockets may be suggested
**Additional RAM may be required for language pack support & gateways (1GB each pack)
***/apps is the required mount point; LVM and/or XFS file system is highly recommended. Slower disks can be utilized if slow performance is acceptable.

NOTE
Clustering is only supported on LANs and where network split-brain between nodes is very unlikely. In the event of a network split-brain event, manual intervention may be required and data in-flight may be lost.

2.4.2 Single Host

For deployments of Proof-of-Concepts, Development, User Acceptance Testing (UAT), and some limited Production environments, Amelia can be deployed on a single physical or virtual server. As of this writing, the following are guidelines on the infrastructure requirements for a single host deployment

Table 3. Single Host Environment Resource Requirements

Resource	Quantity	Notes
CPUs	16	Based on underlying server's NUMA settings, multiple sockets may be suggested
RAM**	80 GB	
Disk Capacity	300 GB	/apps is the required mount; LVM and/or XFS file system is highly recommended. Slower disks can be utilized if slow performance is acceptable.

**Additional RAM may be required for language pack support & gateways (1GB each pack)

3. Components of Amelia

An Amelia instance includes database, message transport, load balancers, and other components, as well as escalation and other processes and systems to set up.

3.1 Operating System Requirements

Amelia is a Linux based Application and can be deployed onto a RHEL 7-family OS. These are:

- Scientific Linux 7.x (SL)
  - Used for IPsoft Cloud and On-Premise Online Deployments
  - IPsoft is responsible for Amelia code and OS patches
  - IPsoft's images follows the Center for Internet Security® (http://www.cisecurity.org/) benchmarks, referred to as CIS.
- CentOS 7.x
  - Licensing is provided via GPL
  - Yum repository for software dependencies is provided by client or via Internet
  - IPsoft's images follows the Center for Internet Security® (http://www.cisecurity.org/) benchmarks, referred to as CIS.
- RedHat Enterprise Linux 7.x (RHEL)
  - Licensing is provided by client at time of install
  - Yum repository for software dependencies is provided by client at time of install
  - IPsoft can provide RHEL CIS image if desired
- Oracle Linux 7.x (RHEL)
  - Licensing is provided via GNU General Public License (GPLv2). Support contracts are available from Oracle.
  - Yum repository for software dependencies is provided by client or via Internet
  - IPsoft does not provide a Oracle Linux image

IPsoft can provide an Open Virtualization Appliance (OVA) with the requisite OS, software dependencies, and Amelia herself. Leveraging IPsoft's OVA provides a quick approach to standing up an Amelia instance. On request, IPsoft also can deploy the software dependencies and Amelia on a Client's OS build. Client is responsible for any licensing and yum repository access.

3.1.1 Required RPMs

Version numbers are subject to change without notice, of course, as these come from vendor-supplied repositories.

Table 4. Required RPMs

Name	
amelia-batch-2.5.24-86.noarch	perl-PlRPC-0.2020-14.el7.noarch
amelia-engine-2.5.24-86.noarch	perl-Pod-Escapes-1.04-291.el7.noarch
amelia-escalation-2.5.24-86.noarch	perl-podlators-2.5.1-3.el7.noarch
amelia-web-2.5.24-86.noarch	perl-Pod-Perldoc-3.20-4.el7.noarch
ansible-2.2.1.0-1.el7.noarch	perl-Pod-Simple-3.28-4.el7.noarch
erlang-18.3-1.el7.centos.x86_64	perl-Pod-Usage-1.63-3.el7.noarch
haproxy-1.5.18-3.el7.x86_64	perl-Scalar-List-Utils-1.27-248.el7.x86_64
haveged-1.9.1-2.el6.x86_64	perl-Socket-2.010-4.el7.x86_64
innotop-1.11.1-1.el7.noarch	perl-Storable-2.45-3.el7.x86_64
ipsoft-amelia-qa-17.01.16-3.x86_64	perl-TermReadKey-2.30-20.el7.x86_64
jdk1.8.0_112-1.8.0_112-fcs.x86_64	perl-Text-ParseWords-3.29-4.el7.noarch
jemalloc-3.6.0-1.el7.x86_64	perl-threads-1.87-4.el7.x86_64
libaio-0.3.109-13.el7.x86_64	perl-threads-shared-1.43-6.el7.x86_64
libev-4.15-3.el7.x86_64	perl-Time-HiRes-1.9725-3.el7.x86_64
libtomcrypt-1.17-23.el7.x86_64	perl-Time-Local-1.2300-2.el7.noarch
libtommath-0.42.0-4.el7.x86_64	postgresql95-9.5.7-1PGDG.rhel7.x86_64
lsof-4.87-4.el7.x86_64	postgresql95-contrib-9.5.7-1PGDG.rhel7.x86_64
MySQL-python-1.2.5-1.el7.x86_64	postgresql95-devel-9.5.7-1PGDG.rhel7.x86_64
percona-xtrabackup-2.3.7-2.el7.x86_64	postgresql95-libs-9.5.7-1PGDG.rhel7.x86_64
Percona-XtraDB-Cluster-56-5.6.35-26.20.2.el7.x86_64	postgresql95-server-9.5.7-1PGDG.rhel7.x86_64
Percona-XtraDB-Cluster-client-56-5.6.35-26.20.2.el7.x86_64	py-bcrypt-0.4-4.el7.x86_64
Percona-XtraDB-Cluster-galera-3-3.20-2.el7.centos.x86_64	python2-crypto-2.6.1-9.el7.x86_64
Percona-XtraDB-Cluster-garbd-57-5.7.19-29.22.3.el7.x86_64	python2-ecdsa-0.13-4.el7.noarch
Percona-XtraDB-Cluster-server-56-5.6.35-26.20.2.el7.x86_64	python2-eventlet-0.18.4-1.el7.noarch
Percona-XtraDB-Cluster-shared-56-5.6.35-26.20.2.el7.x86_64	python2-paramiko-1.16.1-1.el7.noarch
perl-5.16.3-291.el7.x86_64	python2-pyasn1-0.1.9-7.el7.noarch
perl-Carp-1.26-244.el7.noarch	python-babel-1.3-1.rhel7.noarch
perl-Compress-Raw-Bzip2-2.061-3.el7.x86_64	python-greenlet-0.4.2-3.el7.x86_64
perl-Compress-Raw-Zlib-2.061-4.el7.x86_64	python-httplib2-0.7.7-3.el7.noarch
perl-constant-1.27-2.el7.noarch	python-jinja2-2.8-7.rhel7.noarch
perl-Data-Dumper-2.145-3.el7.x86_64	python-keyczar-0.71c-2.el7.noarch
perl-DBD-MySQL-4.023-5.el7.x86_64	python-markupsafe-0.23-11.rhel7.x86_64
perl-DBI-1.627-4.el7.x86_64	python-pip-7.1.0-1.el7.noarch
perl-Encode-2.51-7.el7.x86_64	python-setuptools-0.9.8-4.el7.noarch

perl-Exporter-5.68-3.el7.noarch	python-wheel-0.24.0-2.el7.noarch
perl-File-Path-2.09-2.el7.noarch	pytz-2016.6.1-2.rhel7.noarch
perl-File-Temp-0.23.01-3.el7.noarch	qpress-11-1.el7.x86_64
perl-Filter-1.49-3.el7.x86_64	rabbitmq-server-3.6.6-1.noarch
perl-Getopt-Long-2.40-2.el7.noarch	redis-2.8.19-2.el7.x86_64
perl-HTTP-Tiny-0.033-3.el7.noarch	socat-1.7.2.2-5.el7.x86_64
perl-IO-Compress-2.061-2.el7.noarch	sshpass-1.05-5.el7.x86_64
perl-libs-5.16.3-291.el7.x86_64	strace-4.8-11.el7.x86_64
perl-macros-5.16.3-291.el7.x86_64	wget-1.14-13.el7.x86_64
perl-Net-Daemon-0.48-5.el7.noarch	xinetd-2.3.15-13.el7.x86_64
perl-parent-0.225-244.el7.noarch	yum-utils-1.1.31-40.el7.noarch
perl-PathTools-3.40-5.el7.x86_64	

3.2 External Load Balancers

Amelia clusters (can also be configured for single node deployments) leverage external load balancers to handle layer 4 (transport layer) traffic. Load balancing this way will forward user/api traffic based on host and port availability. Health Checks determines if a backend server is available to process requests. The default health check is to establish a TCP connection to the server on the configured hostname/IP and port.

IPsoft recommends using the “least connections” algorithm because of the potential for longer sessions. Load balancers will forward https connections to the Amelia-Web services; any SSL certificates will be provided to IPsoft to decrypt the SSL traffic.

3.3 HAProxy

HAProxy is normally used to handle external load balancing requests for web/application loads. IPsoft uses HAproxy for service checks and load balancing internal Amelia services and dependencies. This allows for better utilization of all servers and ensures availability. IPsoft configures HAProxy by binding the loopback IP address (127.0.0.x) as the frontend listening IP/port, forwards traffic via the “Round Robin” algorithm to the other servers, with the necessary health check.

3.4 MySQL / Percona XtraDB Cluster (PXC)

The Knowledge Database (KDB) to store configuration parameters such as domain, authentication, FAQ, Grammar, transactions from SLUs/BPNs, classifiers. In addition, there is at least one slave of this database which is used at conversation time to query this information. Writes to this database are only done through the admin daemons.

The CDB (Conversation Database) is used for storing per-conversation data. There may be multiple instances of the CDB database attached to multiple engines. To support additional conversation throughput, additional instances can be added along with the associated engines. These databases are not accessible from the admin daemons.

For more information regarding PXC, here is a web link to Percona: https://www.percona.com/software/mysql-database/percona-xtradb-cluster

3.5 POSTGRESQL

Postgres is used as a read-only datasource for Conceptnet; a freely-available semantic network, designed to help computers understand the meanings of words that people use.

For more information regarding Conceptnet, here is a web link to ConceptNet: http://conceptnet.io

3.6 RABBITMQ

RabbitMQ is used for various messaging tasks and share data between the Amelia Engines and Web services as a three-node active/active/active cluster. It uses both Stomp and AMQP messaging protocols which are exposed on ports 13351 through 13354. Stomp and AMQP frontends are configured in HAProxy to distribute connections in RabbitMQ. Upon failure of a single node, the engine and web will reconnect and begin consuming and sending messages from one of the remaining nodes.

For more information regarding Rabbit MQ Clustering, here is a web link to RabbitMQ's documentation: http://www.rabbitmq.com/clustering.html

3.7 REDIS SENTINEL

Redis is a Key-Value In-Memory data structure store used for caching and configured with a single master and multiple slaves. Monitoring and automatic master promotion is handled by Redis Sentinel. Amelia connects to a Sentinel front-end in their local HAProxy to retrieve the current master Redis server information and then connect directly to that instance. Should the master Redis instance fail, a new master is elected and failover occurs automatically.

For more information regarding Redis Sentinel, here is a web link to Redis' documentation:

http://redis.io/topics/sentinel

3.8 AMELIA DAEMONS

3.8.1 Amelia-Web

Amelia-Web has a Spring Security layer to authenticate and authorize external connections to the system. Contains the REST APIs used by the UI and has read/write access to the KDB master database; no access to CDB.

3.8.2 User Web

The end user and agent conversation interface. Has read-only to a KDB slave and read/write to a CDB.

3.8.3 Engine Service

Amelia Engine Packs (AEP) are bundled units of language, process, and server-side integrations to achieve conceptual objectives.

3.8.4 Batch Service

The Batch Service provides metric calculation and other batch jobs. Has read/write to the KDB master and no access to CDB.

3.8.5 Escalation Service

Amelia can escalate during a conversation in the following manners:

- Misunderstanding: Core dialog manager or a subsystem is unable to handle the user's response.
- Explicit Escalation: Can occur only from BPN, through an "escalate" task.
- Warm Handover: A special explicit escalation, where a reason is specified.

The Escalation Service has read/write to the KDB master and no access to CDB.

3.8.6 Integration Service

The Integration Service is a process run separately from Amelia, potentially on another host or hosts, to allow Amelia to interface with external systems. Integration Flows are Apache Camel contexts created in Amelia V3 admin tools and deployed to these remote processes over GRPC. The Integration Service unpacks the bundle and deploys it in its own Spring application context separate from that of the parent context and of any other flows running on the Integration Service. Integration Service relies on Spring Integration to handle Direct RPC-Style calls over RabbitMQ, and then internally hands off to Apache Camel to execute the given request.

Integration with Amelia is complex and can vary depending on the usage. Please reach out to your IPsoft's Cognitive Lead for best practices involving integration with specific technologies

3.9 TEXT TO SPEECH (TTS)

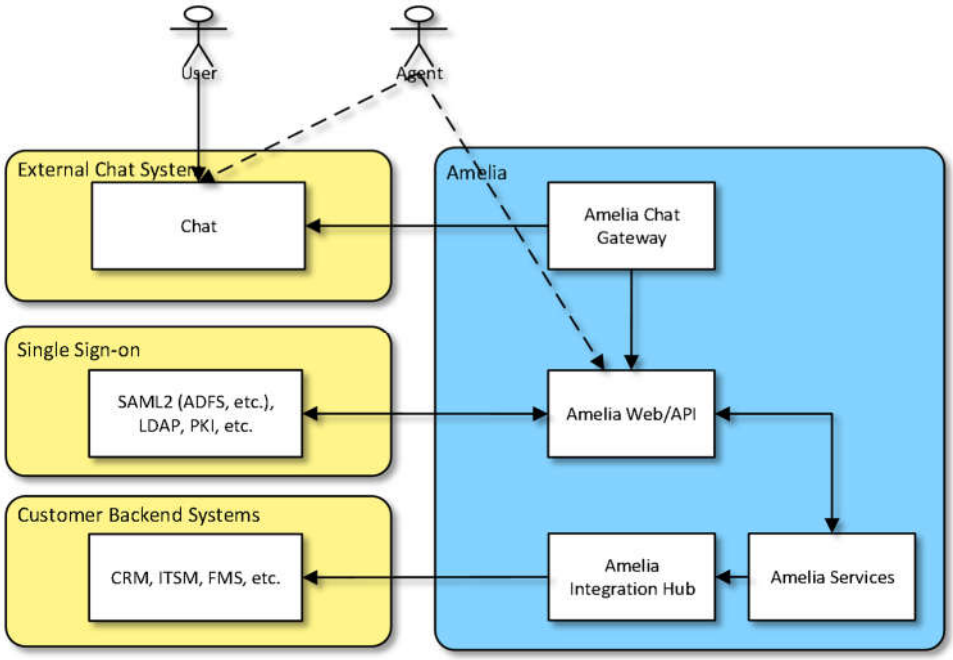
TTS may be used for speech synthesis in Amelia. Multiple TTS voice engines may be installed to facilitate synthesis of audio to satisfy specified language requirements. The TTS server receives encoding requests and also serves the front-end content.

The TTS runs on a Windows 2012 R2 server, ideally with a load balanced deployment of two or more Windows Servers in Active/Passive mode behind a provided Virtual IP (VIP) address. IPsoft recommends deploying a TTS server with a minimum of 8CPUs/16GB RAM/100GB disk capacity. Please inquire with RND on supported languages and licensing information.

TTS is not provided as part of Amelia Deployments; this is a separate, licensed installation.

4. Chat Integration

Amelia can be integrated with external chat systems like LivePerson/LiveEngage, Skype for Business, Facebook Messenger. Please reach out to your IPsoft's Cognitive Lead for best practices involving integration with specific technologies.



5. Security

5.1 Authentication Systems

Authentication systems in Amelia define where the user credentials are stored, as well as how to verify credentials. All authentication systems have the following configurable properties. Specific, or custom authentication systems, may require additional configuration or custom development.

5.1.1 Local Authentication

Passwords are stored hashed with Bcrypt in the Amelia database. Local Authentication is the default setting "out-of-the-box".

5.1.2 LDAP / Active Directory (AD)

For deployments leveraging LDAP, Amelia does not store user passwords (or hashes), but instead delegates password verification to the configured LDAP server or Windows Domain Controllers. The following additional configuration parameters are required to configure LDAP/AD properly:

5.1.3 Deny All

The Deny All authentication system is used for required accounts that should not allow interactive login. Any attempt to authenticate to this authentication system will fail and no passwords or hashes are stored. This authentication system is used for the Amelia user by default.

5.1.4 SAML 2

Amelia supports both Service Provider (SP) initiated and Identity Provider (IDP) initiated SAML2 authentication. As with all authentication methods supported by Amelia, SAML2 is integrated within the core Amelia authentication and authorization framework. In the SAML2 case, support is provided by Spring Security SAML. SAML authentication is only supported for web clients (not mobile).Default Authentication Setting

- Service Provider (SP): An application providing a service to an end-user. In this case, Amelia.
- Identity Provider (IDP): An application/service that manages user identities and provides authentication capabilities. For example, Microsoft Active Directory Federation Services (ADFS) or Ping Federate.

SERVICE PROVIDER INITIATED AUTHENTICATION

In SP initiated authentication, an end user first requests a resource within Amelia. If the resource requires authentication and the user is not yet authenticated with Amelia, the user will be sent to the IDP for authentication. Upon successful authentication, the IDP will send the user back to Amelia with a cryptographically secure assertion of their identity.

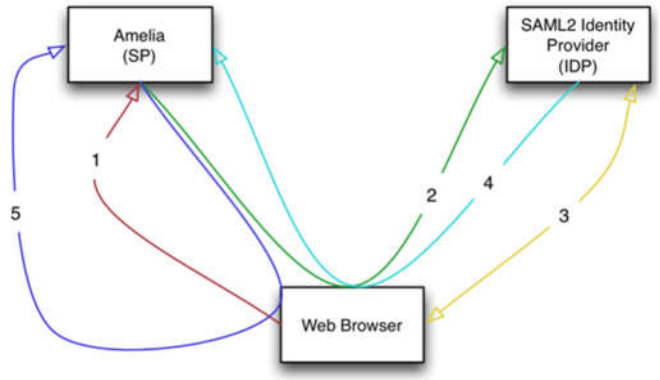


Figure 1. Amelia and Service Provider Initiated Authentication

1. An unauthenticated user attempts to access a protected resource in Amelia.
2. Amelia generates an authentication request (signed AuthnRequest) and redirects the user's browser to the IDP with this request.
3. The IDP determines the user's identity. This could involve one-time tokens, username and password, or other authentication methods. The exact mechanism is not important to the integration.
4. Once the user is authenticated, the IDP generates a signed AuthnResponse carrying the user's identity, email, and other profile information as required. The IDP then redirects the user's browser back to Amelia with this response.
5. Amelia verifies the signed response, and if valid, auto-creates (if required/configured) the user, and logs them into Amelia. Amelia then redirects the user's browser to the original protected resource they requested in step 1.

All communication takes place over TLS and is between the user's browser and the SP, and the user's browser and the IDP. No direct communication is done between the SP and IDP other than initial out of band sharing of metadata at configuration time (see below).

IDENTITY PROVIDER INITIATED AUTHENTICATION

Amelia supports IDP initiated authentication, however, SP initiated authentication should be preferred. In the IDP initiated scenario users must first authenticate to the IDP before accessing Amelia.

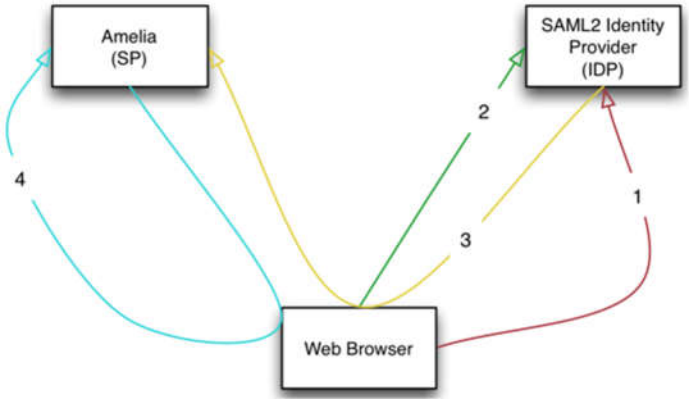


Figure 2. Amelia and Identity Provider Initiated Authentication

1. A user logs into the IDP directly. Normally by clicking a specially crafted link in some portal.

2. The user selects the service provider from a list provided by the IDP. This step will often be skipped if the user started with a special link in step 1.
3. The IDP generates and signed AuthnResponse and redirects the user's browser to Amelia.
4. Amelia verifies the signed response, and if valid, logs in the user and redirects the user's browser to the default page.

5.2 Anonymous Access

Anonymous access allows unauthenticated users to access Amelia. The unauthenticated user will be logged in as a special type of user, the Anonymous User. These users can be given access rights just like any other user in the system. The only difference between an anonymous user and any other user is that anonymous users will be auto-created when the user accesses the system and if anonymous access is enabled. Anonymous Users are given a default set of access rights based on a configured group. This group should have very minimal access rights. Most likely only End User for the anonymous domain.

Configuration of Anonymous access is a simple checkbox, which enables/disables the feature. This is also applied to the Amelia Escalation feature.

Using Anonymous Access are for specific use cases; please inquire with the Cognitive Team for specifics

5.3 SSL Certificates

Amelia requires SSL certificates to keep sensitive information sent across the network encrypted so that only the intended recipient can understand. IPsoft recommends wildcard certificates for ease of installation and future growth. If a wildcard SSL certificate is not used, four individual certificates or one “Subject Alternative Name”, or SAN certificate must be created for each instance (except for DR).

Table 5. SSL Certificate Requirements

Name	Description
amelia.client.com	HTTP/HTTPS URL for Amelia

5.4 Firewall Rules

Table 6. Firewall Rules

Destination Port	Protocol	Purpose	Source	Destination
haproxy				
80	HTTP	Redirect to HTTPS	Browser	haproxy
443	HTTPS and WebSockets	User interaction	Browser	haproxy

1936	HTTP	Statistics	Browser	haproxy
19999	HTTP	Self-Monitoring	Browser	netdata
amelia-engine-service@p001				
4431	HTTP (TLS 1.2)	Web UI	haproxy	amelia-engine-service@p001
4434	JMX	Monitoring	monitoring network	amelia-engine-service@p001
4435	GRPC	RPC	haproxy	amelia-engine-service@p001
4436	GRPC	RPC - All pods roundrobin	amelia-user-web	haproxy
44001	GRPC	RPC - Engine pod 001 roundrobin.	amelia-user-web	haproxy
amelia-user-web				
4441	HTTP (TLS 1.2)	Web UI	haproxy	amelia-user-web
4444	JMX	Monitoring	monitoring network	amelia-user-web
4449 (_only_ required for UI development)	HTTPS	UI Development	developer machine	webpack-dev-server
amelia-escalation-service				
4571	HTTP (TLS 1.2)	Web UI	haproxy	amelia-escalation-service
4574	JMX	Monitoring	monitoring network	amelia-escalation-service
amelia-batch-service				
4581	HTTP (TLS 1.2)	Web UI	haproxy	amelia-batch-service

4584	JMX	Monitoring	monitoring network	amelia-batch-service
amelia-admin-web				
4601	HTTP (TLS 1.2)	Web UI	haproxy	amelia-admin-web
4604	JMX	Monitoring	monitoring network	amelia-admin-web
4609 (_only_ required for UI development)	HTTPS	UI Development	deveveloper machine	webpack-dev-server
amelia-admin-service				
4611	HTTP (TLS 1.2)	Web UI	haproxy	amelia-admin-service
4614	JMX	Monitoring	monitoring network	amelia-admin-service
4615	GRPC	RPC	haproxy	amelia-admin-service
4616	GRPC	RPC	amelia-admin-web	haproxy
amelia-engine-service@p002				
4621	HTTP (TLS 1.2)	Web UI	haproxy	amelia-engine-service@p002
4624	JMX	Monitoring	monitoring network	amelia-engine-service@p002
4625	GRPC	RPC	haproxy	amelia-engine-service@p002
44002	GRPC	RPC	amelia-user-web	haproxy
amelia-integration-service				
4634	JMX	Monitoring	monitoring network	amelia-integration-service
4635	GRPC	RPC	haproxy	amelia-integration-service

amelia-account-service				
4641	HTTP (TLS 1.2)	Web UI	haproxy	amelia-account-service
4644	JMX	Monitoring	monitoring network	amelia-account-service
4645	GRPC	RPC	haproxy	amelia-account-service
4646	GRPC	RPC	amelia-*	haproxy
amelia-model-server				
4651	HTTP (TLS 1.2)	Web UI	haproxy	amelia-model-server
4654	JMX	Monitoring	monitoring network	amelia-model-server
4655	GRPC	RPC	haproxy	amelia-model-server
4656	GRPC	RPC	amelia-*	haproxy
amelia-rest-gateway				
4661	HTTP (TLS 1.2)	REST Endpoints	haproxy	amelia-rest-gateway
4664	JMX	Monitoring	monitoring network	amelia-rest-gateway
mysql@amelia-kdb-master	Master Knowledge DB			
13304	HTTP	Cluster Check	haproxy	xinetd
13305	TCP	SQL	haproxy	mysql@amelia-kdb-master
13306	TCP	SQL	amelia-admin-service	haproxy
13307	TCP	SST	mysql@amelia-kdb-master	mysql@amelia-kdb-master
13308	TCP	Group Communication	mysql@amelia-kdb-master	mysql@amelia-kdb-master

13309	TCP	IST	mysql@amelia-kdb-master	mysql@amelia-kdb-master
mysql@amelia-kdb-slave	Slave Knowledge DB			
13314	HTTP	Slave Check	haproxy	xinetd
13315	TCP	SQL	haproxy	amelia-kdb-slave
13316	TCP	SQL	amelia-engine-service	haproxy
mysql@amelia-cdb-p001	Pod 001 Conversation DB			
13324	HTTP	Cluster Check	haproxy	xinetd
13325	TCP	SQL	haproxy	mysql@amelia-cdb-p001
13326	TCP	SQL	amelia-engine-service@p001	haproxy
13327	TCP	SST	mysql@amelia-cdb-p001	mysql@amelia-cdb-p001
13328	TCP	Group Communication	mysql@amelia-cdb-p001	mysql@amelia-cdb-p001
13329	TCP	IST	mysql@amelia-cdb-p001	mysql@amelia-cdb-p001
mysql@amelia-cdb-p002	Pod 002 Conversation DB. Only in dev-v3-ipsoft for testing multiple cdb. Normally on 1332* ports.			
13334	HTTP	Cluster Check	haproxy	xinetd
13335	TCP	SQL	haproxy	mysql@amelia-cdb-p002
13336	TCP	SQL	amelia-engine-service@p002	haproxy
13337	TCP	SST	mysql@amelia-cdb-p002	mysql@amelia-cdb-p002
13338	TCP	Group Communication	mysql@amelia-cdb-p002	mysql@amelia-cdb-p002
13339	TCP	IST	mysql@amelia-cdb-p002	mysql@amelia-cdb-p002

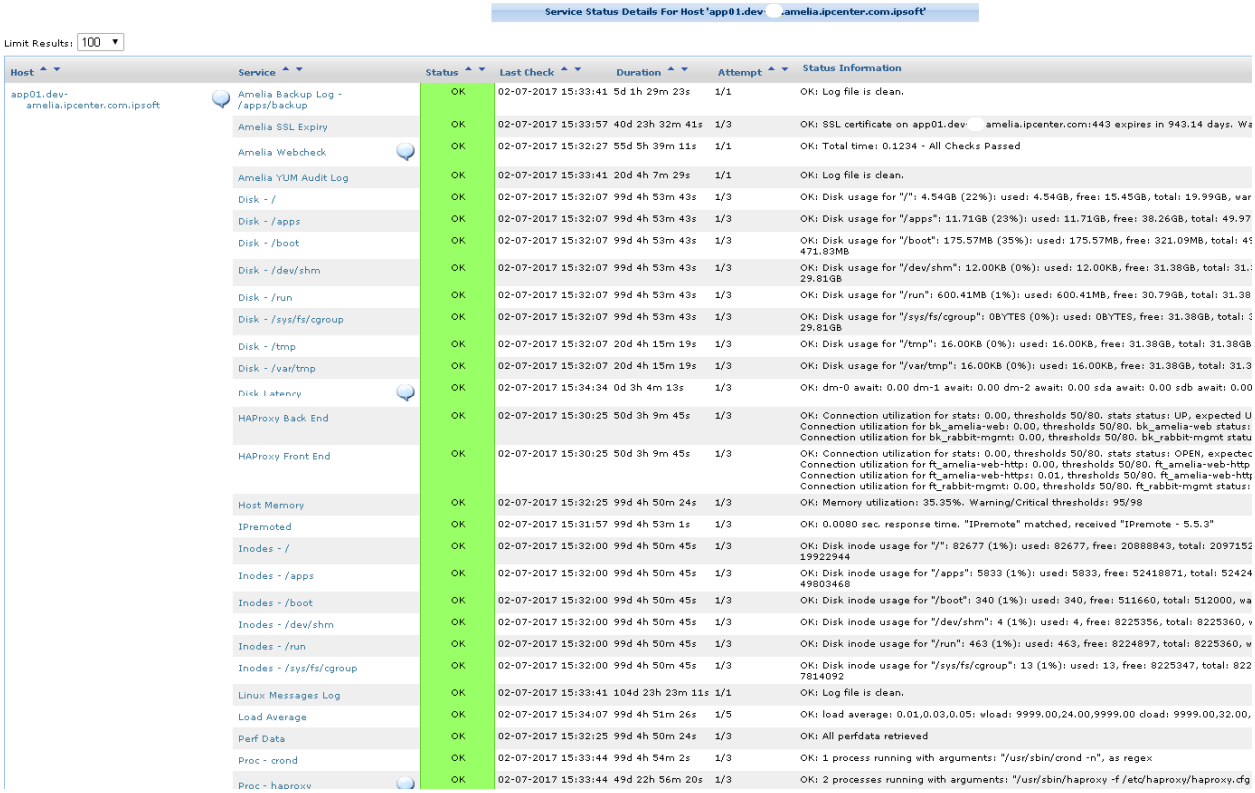
mysql@amelia-adb	Account Database			
13344	HTTP	Cluster Check	haproxy	xinetd
13345	TCP	SQL	haproxy	mysql@amelia-adb
13346	TCP	SQL	amelia-admin-service	haproxy
13347	TCP	SST	mysql@amelia-adb	mysql@amelia-adb
13348	TCP	Group Communication	mysql@amelia-adb	mysql@amelia-adb
13349	TCP	IST	mysql@amelia-adb	mysql@amelia-adb
redis				
13341	TCP	Datastore, Cache	amelia-user-web, amelia-admin-web	redis
13342	TCP	Redis Sentinel Cluster Management	redis-sentinel, haproxy	redis-sentinel
13343	TCP	Redis Sentinel VIP	amelia-user-web, amelia-admin-web	haproxy
rabbitmq				
13351	AMQP (TLS 1.2)	Messaging	haproxy	rabbitmq
13352	AMQP (TLS 1.2)	Messaging	amelia-*	haproxy
13353	STOMP (TLS 1.2)	Messaging	haproxy	rabbitmq
13354	STOMP (TLS 1.2)	Messaging	amelia-*	haproxy
13355	HTTP (TLS 1.2)	Monitoring	monitoring network	rabbitmq

13356	TCP	Clustering - Distribution	RabbitMQ	RabbitMQ
13357	TCP	Clustering - epmd	RabbitMQ/epmd	RabbitMQ/epmd
amelia-tf-serving				
13361	GRPC	question classification	haproxy	amelia-tf-serving@question_classifier
13362	GRPC	question classification	amelia-*	haproxy
13363	GRPC	named entity tagging	haproxy	amelia-tf-serving@named_entity_tagger
13364	GRPC	named entity tagging	amelia-*	haproxy
13365	GRPC	answer polarity	haproxy	amelia-tf-serving@answer_polarity
13366	GRPC	answer polarity	amelia-*	haproxy
13367	GRPC	sentiment analysis	haproxy	amelia-tf-serving@sentiment_analysis
13368	GRPC	sentiment analysis	amelia-*	haproxy
13371	GRPC	short answer extraction	haproxy	amelia-tf-serving@short_answer_extractor
13372	GRPC	short answer extraction	amelia-*	haproxy
13373	GRPC	shallow parsing	haproxy	amelia-tf-serving@shallow_parser
13374	GRPC	shallow parsing	amelia-*	haproxy
13375	GRPC	contextual lstm	haproxy	amelia-tf-serving@contextual_lstm
13376	GRPC	contextual lstm	amelia-*	haproxy
13377	GRPC	hierarchical lstm	haproxy	amelia-tf-serving@hierarchical_lstm
13378	GRPC	hierarchical lstm	amelia-*	haproxy

13379	GRPC	entailment	haproxy	amelia-tf-serving@entailment_classifier
13380	GRPC	entailment	amelia-*	haproxy
13381	GRPC	facial recognition	haproxy	amelia-tf-serving@fr_recognizer
13382	GRPC	facial recognition	amelia-*	haproxy
13383	GRPC	semnet don't know classifier	haproxy	amelia-tf-serving@dont_know_classifier
13384	GRPC	semnet don't know classifier	amelia-*	haproxy
amelia-conceptnet				
13401	TCP	database	amelia-conceptnet-api	rh-postgresql95-postgresql@amelia-conceptnet
13402	HTTP	concept net API	haproxy	amelia-conceptnet-api
13403	HTTP	concept net API	amelia-*	haproxy
amelia-tf-syntaxnet				
14000	TCP	syntaxnet parser	haproxy	amelia-tf-syntaxnet
14001	GRPC	syntaxnet parser	amelia-*	amelia-tf-syntaxnet
14009	TCP	syntaxnet parser	amelia-tf-syntaxnet	amelia-tf-syntaxnet
14090	TCP	syntaxnet parser	haproxy-test	amelia-tf-syntaxnet-test
14091	GRPC	syntaxnet parser	amelia-*-test	amelia-tf-syntaxnet-test
14099	TCP	syntaxnet parser	amelia-tf-syntaxnet-test	amelia-tf-syntaxnet-test

6. Monitoring

For IPsoft Hosted and Online deployments, all Amelia instances and supporting Operating Systems are monitored by IPcenter using IPmons. IPmons are configured to monitor each component of Amelia; each component having several individual checks to ensure thorough reporting and availability of each Amelia instance. Below is an example of the OS checks deployed:



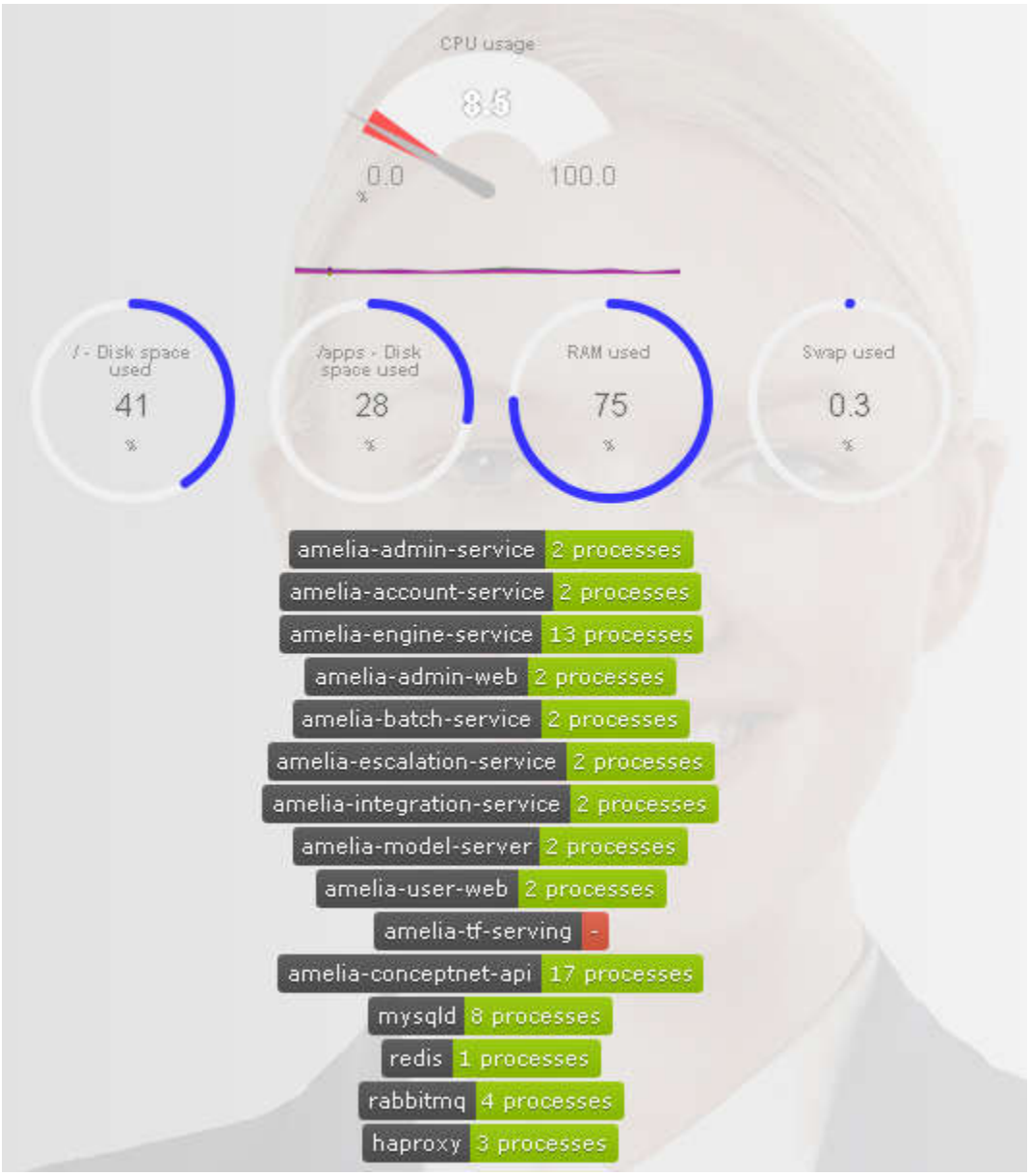
Service Status Details For Host 'app01.dev .amelia.ipcenter.com.ipsoft'

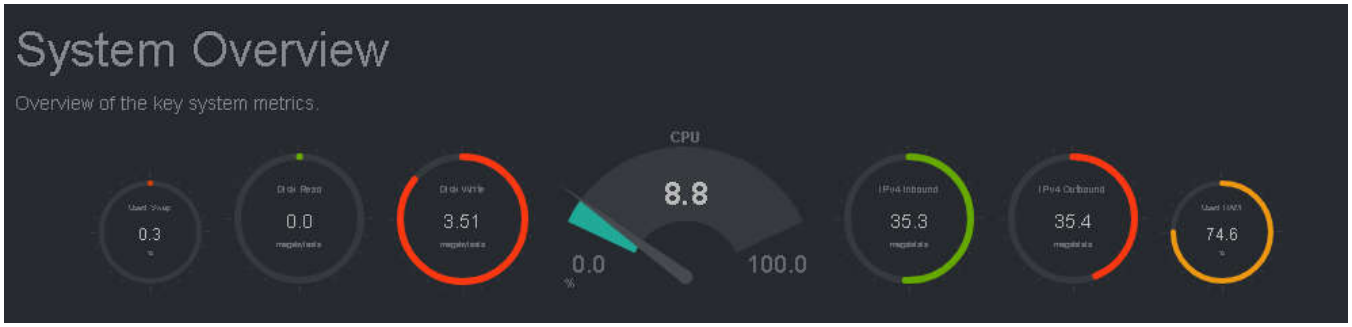
Limit Results: 100

Host	Service	Status	Last Check	Duration	Attempt	Status Information
app01.dev-amelia.ipcenter.com.ipsoft	Amelia Backup Log - /apps/backup	OK	02-07-2017 15:33:41	5d 1h 29m 23s	1/1	OK: Log file is clean.
	Amelia SSL Expiry	OK	02-07-2017 15:33:57	40d 23h 32m 41s	1/3	OK: SSL certificate on app01.dev amelia.ipcenter.com:443 expires in 943.14 days. Wa
	Amelia Webcheck	OK	02-07-2017 15:32:27	55d 5h 39m 11s	1/1	OK: Total time: 0.1234 - All Checks Passed
	Amelia YUM Audit Log	OK	02-07-2017 15:33:41	20d 4h 7m 29s	1/1	OK: Log file is clean.
	Disk - /	OK	02-07-2017 15:32:07	99d 4h 53m 43s	1/3	OK: Disk usage for "/": 4.54GB (22%): used: 4.54GB, free: 15.45GB, total: 19.99GB, war
	Disk - /apps	OK	02-07-2017 15:32:07	99d 4h 53m 43s	1/3	OK: Disk usage for "/apps": 11.71GB (23%): used: 11.71GB, free: 38.26GB, total: 49.97
	Disk - /boot	OK	02-07-2017 15:32:07	99d 4h 53m 43s	1/3	OK: Disk usage for "/boot": 175.57MB (35%): used: 175.57MB, free: 321.09MB, total: 49 471.83MB
	Disk - /dev/shm	OK	02-07-2017 15:32:07	99d 4h 53m 43s	1/3	OK: Disk usage for "/dev/shm": 12.00KB (0%): used: 12.00KB, free: 31.38GB, total: 31. 29.81GB
	Disk - /run	OK	02-07-2017 15:32:07	99d 4h 53m 43s	1/3	OK: Disk usage for "/run": 600.41MB (1%): used: 600.41MB, free: 30.79GB, total: 31.38
	Disk - /sys/fs/cgroup	OK	02-07-2017 15:32:07	99d 4h 53m 43s	1/3	OK: Disk usage for "/sys/fs/cgroup": 0BYTES (0%): used: 0BYTES, free: 31.38GB, total: 3 29.81GB
	Disk - /tmp	OK	02-07-2017 15:32:07	20d 4h 15m 19s	1/3	OK: Disk usage for "/tmp": 16.00KB (0%): used: 16.00KB, free: 31.38GB, total: 31.38GB
	Disk - /var/tmp	OK	02-07-2017 15:32:07	20d 4h 15m 19s	1/3	OK: Disk usage for "/var/tmp": 16.00KB (0%): used: 16.00KB, free: 31.38GB, total: 31.3
	Disk Latency	OK	02-07-2017 15:34:34	0d 3h 4m 13s	1/3	OK: dm-0 await: 0.00 dm-1 await: 0.00 dm-2 await: 0.00 sda await: 0.00 sdb await: 0.00
	HAProxy Back End	OK	02-07-2017 15:30:25	50d 3h 9m 45s	1/3	OK: Connection utilization for stats: 0.00, thresholds 50/80. stats status: UP, expected U Connection utilization for bk_amelia-web: 0.00, thresholds 50/80. bk_amelia-web status: Connection utilization for bk_rabbit-mgmt: 0.00, thresholds 50/80. bk_rabbit-mgmt statu
	HAProxy Front End	OK	02-07-2017 15:30:25	50d 3h 9m 45s	1/3	OK: Connection utilization for stats: 0.00, thresholds 50/80. stats status: OPEN, expectec Connection utilization for ft_amelia-web-http: 0.00, thresholds 50/80. ft_amelia-web-http Connection utilization for ft_amelia-web-https: 0.01, thresholds 50/80. ft_amelia-web-htt Connection utilization for ft_rabbit-mgmt: 0.00, thresholds 50/80. ft_rabbit-mgmt status:
	Host Memory	OK	02-07-2017 15:32:25	99d 4h 50m 24s	1/3	OK: Memory utilization: 35.35%. Warning/Critical thresholds: 95/98
	IPremoted	OK	02-07-2017 15:31:57	99d 4h 53m 1s	1/3	OK: 0.0080 sec. response time. "IPremote" matched, received "IPremote - 5.5.3"
	Inodes - /	OK	02-07-2017 15:32:00	99d 4h 50m 45s	1/3	OK: Disk inode usage for "/": 82677 (1%): used: 82677, free: 20888843, total: 2097152 19922944
	Inodes - /apps	OK	02-07-2017 15:32:00	99d 4h 50m 45s	1/3	OK: Disk inode usage for "/apps": 5833 (1%): used: 5833, free: 52418871, total: 52424 49803468
	Inodes - /boot	OK	02-07-2017 15:32:00	99d 4h 50m 45s	1/3	OK: Disk inode usage for "/boot": 340 (1%): used: 340, free: 511660, total: 512000, wa
	Inodes - /dev/shm	OK	02-07-2017 15:32:00	99d 4h 50m 45s	1/3	OK: Disk inode usage for "/dev/shm": 4 (1%): used: 4, free: 8225356, total: 8225360, v
	Inodes - /run	OK	02-07-2017 15:32:00	99d 4h 50m 45s	1/3	OK: Disk inode usage for "/run": 463 (1%): used: 463, free: 8224897, total: 8225360, w
	Inodes - /sys/fs/cgroup	OK	02-07-2017 15:32:00	99d 4h 50m 45s	1/3	OK: Disk inode usage for "/sys/fs/cgroup": 13 (1%): used: 13, free: 8225347, total: 822 7814092
	Linux Messages Log	OK	02-07-2017 15:33:41	104d 23h 23m 11s	1/1	OK: Log file is clean.
	Load Average	OK	02-07-2017 15:34:07	99d 4h 51m 26s	1/5	OK: load average: 0.01,0.03,0.05: wload: 9999.00,24.00,9999.00 cload: 9999.00,32.00,
	Perf Data	OK	02-07-2017 15:32:25	99d 4h 50m 24s	1/3	OK: All perfdata retrieved
	Proc - crond	OK	02-07-2017 15:33:44	99d 4h 54m 2s	1/3	OK: 1 process running with arguments: "/usr/sbin/crond -n", as regex
	Proc - haproxy	OK	02-07-2017 15:33:44	49d 22h 56m 20s	1/3	OK: 2 processes running with arguments: "/usr/sbin/haproxy -f /etc/haproxy/haproxy.cfg

Figure 3. Example of OS Monitoring Checks Performed

Starting with Amelia v3, IPsoft is distributing *netdata* for insights into system and application performance. It also provides a rudimentary level of monitoring and alerting. These alerts do not currently come back to IPsoft, but my be directed at a clients email if desired.


CPU usage
8.5
0.0
%
100.0
/ - Disk space used
41
%
/apps - Disk space used
28
%
RAM used
75
%
Swap used
0.3
%
amelia-admin-service 2 processes
amelia-account-service 2 processes
amelia-engine-service 13 processes
amelia-admin-web 2 processes
amelia-batch-service 2 processes
amelia-escalation-service 2 processes
amelia-integration-service 2 processes
amelia-model-server 2 processes
amelia-user-web 2 processes
amelia-tf-serving -
amelia-conceptnet-api 17 processes
mysqld 8 processes
redis 1 processes
rabbitmq 4 processes
haproxy 3 processes


System Overview
Overview of the key system metrics.
CPU
8.8
0.0
%
100.0
0.3
0.0
3.51
35.3
35.4
74.6

7. Backups

Backups for Amelia can be outfitted with the Customer/Partner's Enterprise Backup solution. This allows the customer to roll Amelia backups into their current enterprise policies. Although most deployments leverage a virtual environment, backups of entire VM can cause a degraded or unresponsive system; it is highly recommended to disable the "virtual machine memory" and "Quiesce guest file system".

File level backups would require a more unique solution. This would require standing up a new, fresh install of an Amelia host to replace an unrecoverable one, restoring the original's configuration and data from a remote backup. For a complete list of backup locations and uses, please contact the Service Technology team.

Defining Recovery Time Objective (RTO), Recovery Point Objective (RPO), and Geographic Redundancy Objective (GRO) should be considered when creating backup/recovery policy for Amelia. For IPsoft Hosted Instances, non-production instances have a 24-hour RTO and no larger than 24-hour RPO. For production instances, a 4-hr RTO and a 4-hour RPO.

8. Disaster Recovery

In its current architecture, Amelia v3 current supports an Active/Passive approach, where web traffic would be directed to the "live" datacenter and replicated to a secondary datacenter. The replication method can be achieved at either the middleware (native) or infrastructure level.

For native replication, Percona XtraDB Cluster and Postgresql would be replicating the databases across the WAN to the DR databases. In this setup all Production and DR database nodes are configured as one large logical cluster. Newer versions of Amelia, Language Packs, and Gateways would be performed separately for both the Production and DR instances; not as one upgrade to cover both instances.

Infrastructure replication can be achieved using 3rd party hardware/software vendors. SAN based replication with orchestration tools (such as Zerto/VMware SRM) is a proven DR solution, as well as software based solutions such as Veeam and Veritas. A Production instance is replicated without changing the hostnames of the VMs, however it is best to keep the IP addresses identical if possible using a stretched layer 2 network.

The length of time necessary to conduct a failover of Amelia will depend on the overall design and available automation processes. Regarding Amelia specifically, all Amelia processes can be started in parallel and can take up to 5 minutes for Amelia to be started in the secondary datacenter, assuming IP addresses are not altered.

For IPsoft Hosted instances, a 4-hr RTO and a 4-hour RPO.