AMΞLIA

IPsoft's Cognitive Agent

Amelia V3 Platform Design

(version 1.3)

Table of Contents

1. AMELIA HIGH LEVEL ARCHITECTURE DESIGN 3
1.1 3-NODE CLUSTER ARCHITECTURE 4
1.2 6-NODE CLUSTER ARCHITECTURE 5
1.3 SINGLE NODE ARCHITECTURE 6
2. DEPLOYMENT MODELS 7
2.1 AMELIA HOSTED IN IPSOFT'S CLOUD 7
2.2 CUSTOMER PREMISE / CLOUD 8
2.2.1 ONLINE DEPLOYMENT 9
2.2.2 OFFLINE DEPLOYMENT 10
3. HARDWARE/RESOURCE RECOMMENDATIONS 11
3.1 CPU REQUIREMENTS 11
3.2 STORAGE REQUIREMENTS 11
3.3 CLUSTERED SETUP 12
3.4 SINGLE HOST 13
4. COMPONENTS OF AMELIA 14
4.1 OPERATING SYSTEM REQUIREMENTS 14
4.2 EXTERNAL LOAD BALANCERS 14
4.3 HAPROXY 15
4.4 MYSQL / PERCONA XTRADB CLUSTER (PXC) 15
4.5 RABBITMQ 15
4.6 REDIS SENTINEL 15
4.7 AMELIA DAEMONS 16
4.7.1 AMELIA-ADMIN-WEB 16
4.7.2 USER WEB 16
4.7.3 ENGINE SERVICE 16
4.7.4 BATCH SERVICE 16
4.7.5 ESCALATION SERVICE 16
4.7.6 INTEGRATION SERVICE 16
4.7.7 DUCKLING SERVICE 17
4.8 TEXT TO SPEECH (TTS) 17
5. CHAT INTEGRATION 18
6. SECURITY 19
6.1 AUTHENTICATION SYSTEMS 19
6.1.1 LOCAL AUTHENTICATION 19
6.1.2 LDAP / ACTIVE DIRECTORY (AD) 19
6.1.3 DENY ALL 19
6.1.4 SAML 2 19
6.2 ANONYMOUS ACCESS 21
6.3 SSL CERTIFICATES 21

6.4 FIREWALL RULES 21
7. MONITORING 28
8. BACKUPS 30
9. DISASTER RECOVERY 31

Document History

Author	Version	Date	Comments	Final Approval?
Randy Schneiderman	0.1	11/15/2017	Initial Document	Yes
Randy Schneiderman	1.0	11/30/2017	Updated for V3	Yes
Randy Schneiderman	1.1	1/18/18	Added NetData and DR sections	Yes
Randy Schneiderman	1.2	1/29/18	Revised Sizing Requirements	Yes
Randy Scheiderman	1.3	6/21/18	Add Duckling Service and POD details, updated Figure 1	Yes

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:

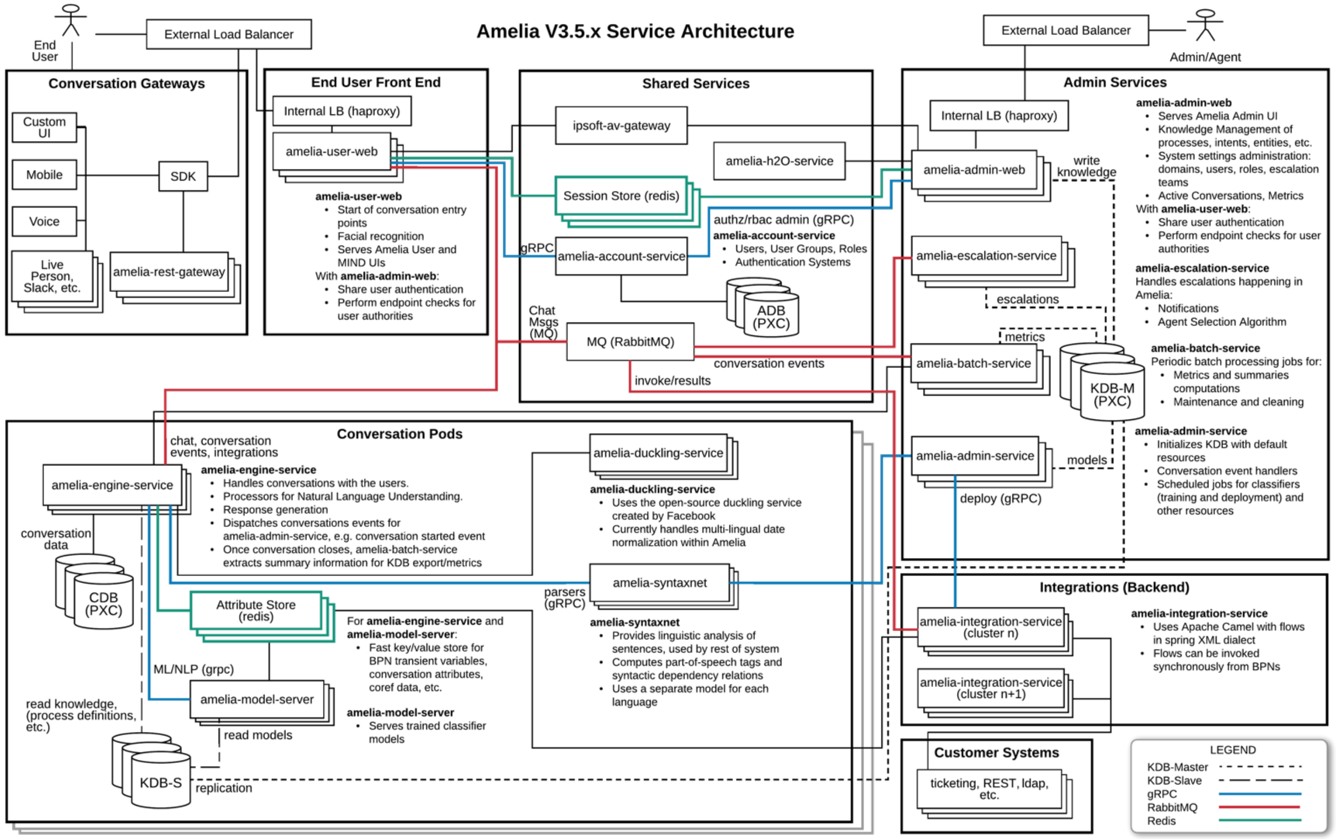

Amelia V3.5.x Service Architecture
End User
External Load Balancer
Conversation Gateways
Custom UI
Mobile
Voice
Live Person, Slack, etc.
SDK
amelia-rest-gateway
End User Front End
Internal LB (haproxy)
amelia-user-web
amelia-user-web
• Start of conversation entry points
• Facial recognition
• Serves Amelia User and MIND UIs
With amelia-admin-web:
• Share user authentication
• Perform endpoint checks for user authorities
Shared Services
ipsoft-av-gateway
amelia-h2O-service
Session Store (redis)
gRPC
amelia-account-service
authz/rbac admin (gRPC)
amelia-account-service
• Users, User Groups, Roles
• Authentication Systems
ADB (PXC)
Chat Msgs (MQ)
MQ (RabbitMQ)
conversation events
invoke/results
External Load Balancer
Admin/Agent
Admin Services
Internal LB (haproxy)
amelia-admin-web
write knowledge
amelia-escalation-service
escalations
metrics
amelia-batch-service
KDB-M (PXC)
amelia-admin-service
models
deploy (gRPC)
amelia-admin-web
• Serves Amelia Admin UI
• Knowledge Management of processes, intents, entities, etc.
• System settings administration: domains, users, roles, escalation teams
• Active Conversations, Metrics
With amelia-user-web:
• Share user authentication
• Perform endpoint checks for user authorities
amelia-escalation-service
Handles escalations happening in Amelia:
• Notifications
• Agent Selection Algorithm
amelia-batch-service
Periodic batch processing jobs for:
• Metrics and summaries computations
• Maintenance and cleaning
amelia-admin-service
• Initializes KDB with default resources
• Conversation event handlers
• Scheduled jobs for classifiers (training and deployment) and other resources
Conversation Pods
chat, conversation events, integrations
amelia-engine-service
amelia-engine-service
• Handles conversations with the users.
• Processors for Natural Language Understanding.
• Response generation
• Dispatches conversations events for amelia-admin-service, e.g. conversation started event
• Once conversation closes, amelia-batch-service extracts summary information for KDB export/metrics
conversation data
CDB (PXC)
Attribute Store (redis)
For amelia-engine-service and amelia-model-server:
• Fast key/value store for BPN transient variables, conversation attributes, coref data, etc.
ML/NLP (grpc)
amelia-model-server
amelia-model-server
• Serves trained classifier models
read knowledge, (process definitions, etc.)
read models
KDB-S
replication
amelia-duckling-service
amelia-duckling-service
• Uses the open-source duckling service created by Facebook
• Currently handles multi-lingual date normalization within Amelia
amelia-syntaxnet
parsers (gRPC)
amelia-syntaxnet
• Provides linguistic analysis of sentences, used by rest of system
• Computes part-of-speech tags and syntactic dependency relations
• Uses a separate model for each language
Integrations (Backend)
amelia-integration-service (cluster n)
amelia-integration-service (cluster n+1)
amelia-integration-service
• Uses Apache Camel with flows in spring XML dialect
• Flows can be invoked synchronously from BPNs
Customer Systems
ticketing, REST, ldap, etc.
LEGEND
KDB-Master
KDB-Slave
gRPC
RabbitMQ
Redis

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.

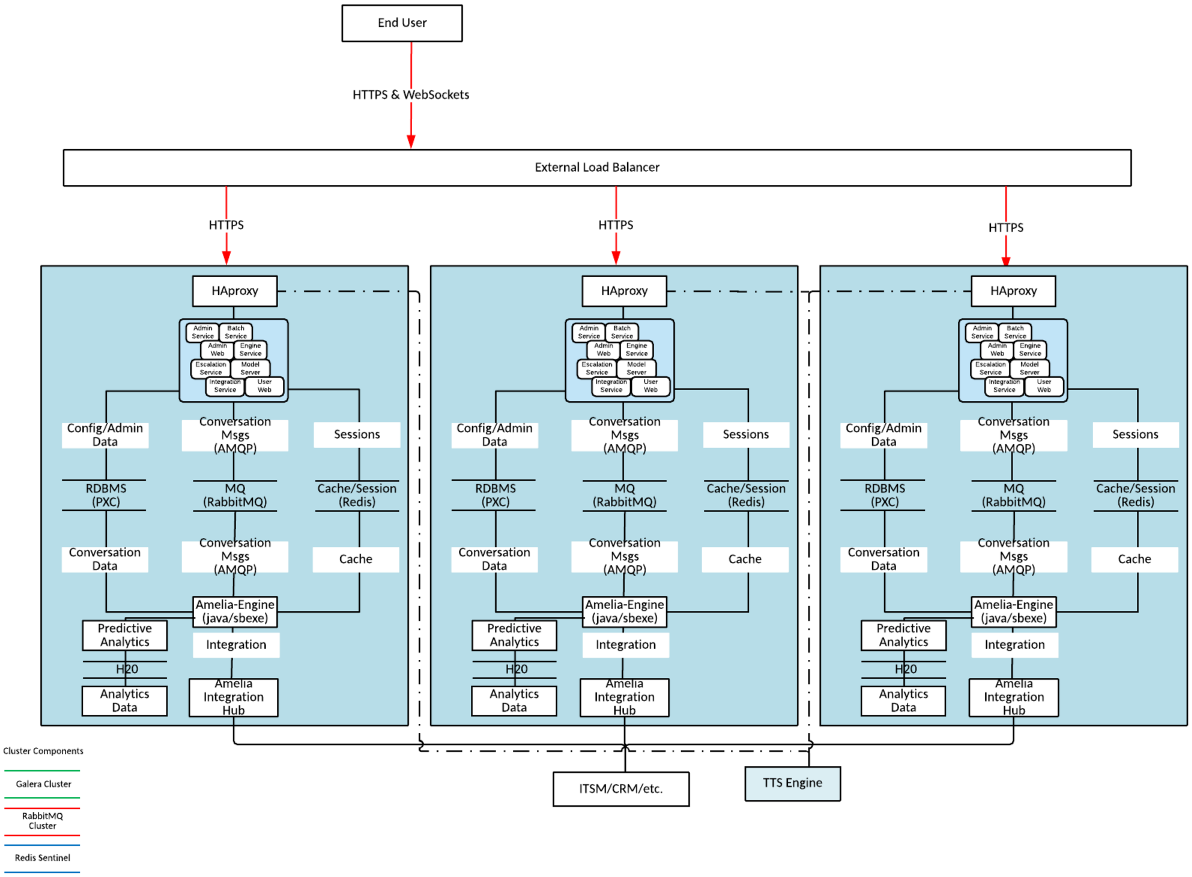


Figure 2. Initial 3-Node Cluster Environment Diagram

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.

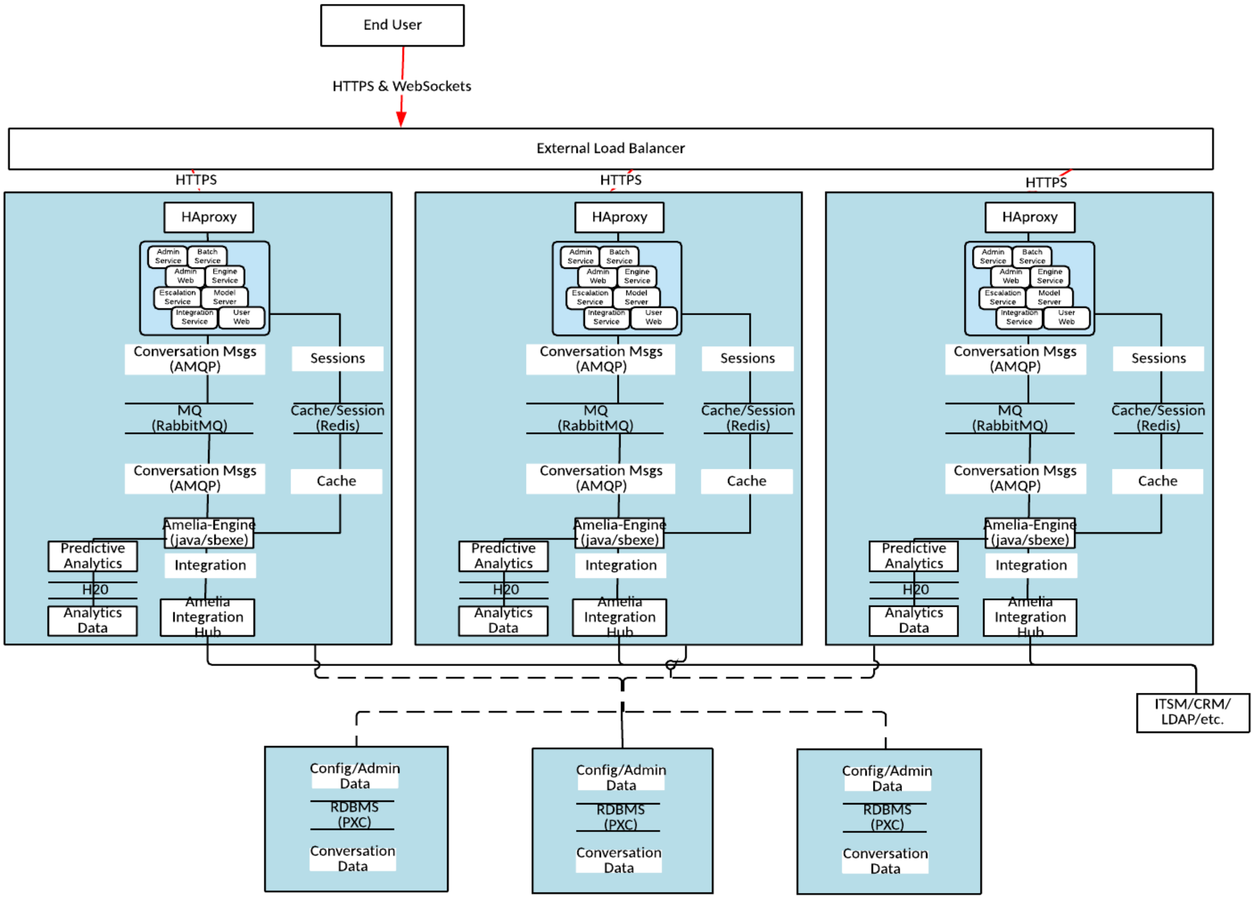


Figure 3. 6-Node Cluster Environment Diagram

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.

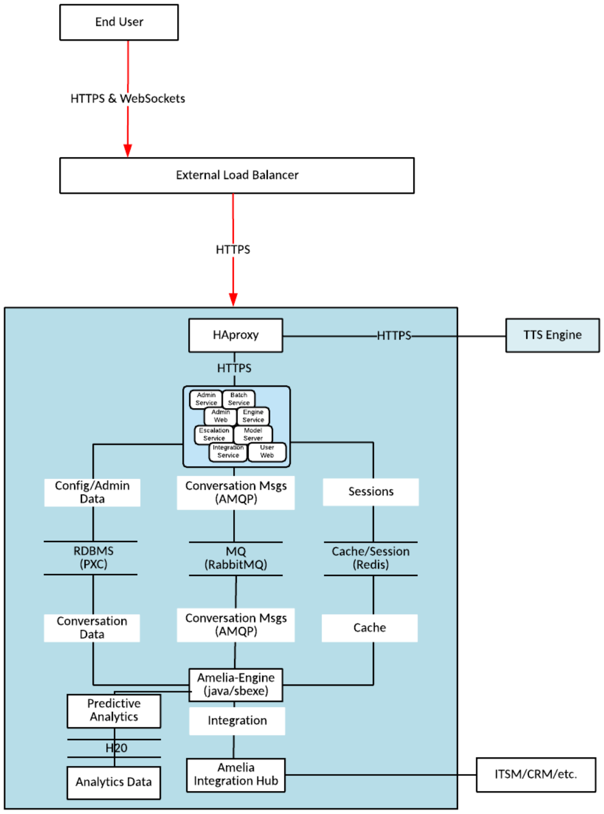


Figure 4. 1-Node Cluster Environment Diagram

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.

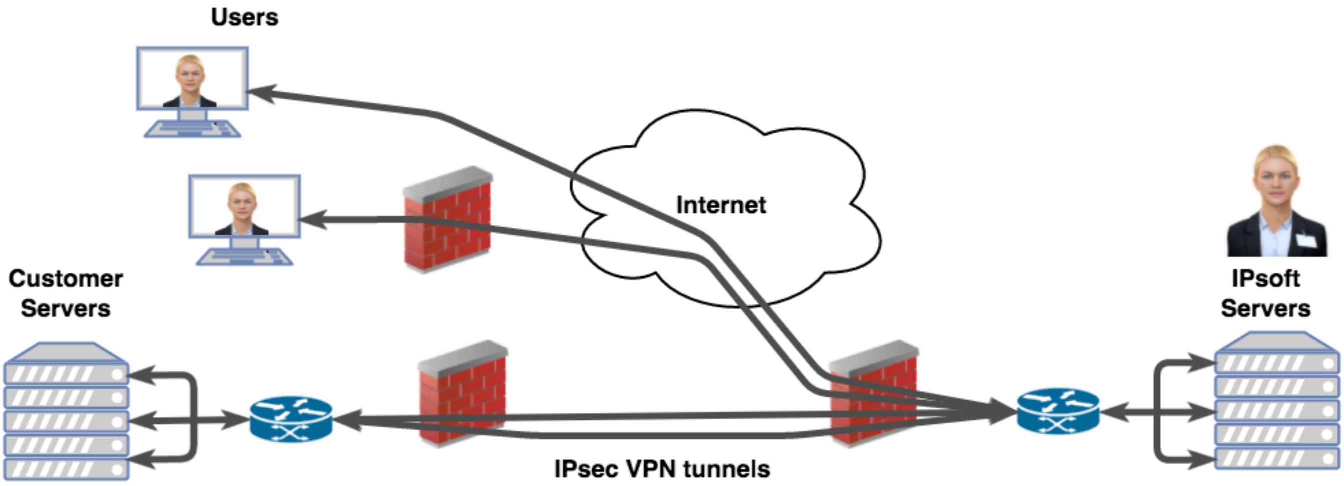


Figure 5. Amelia Hosted in IPsoft's Cloud with IPsec VPN Tunnels

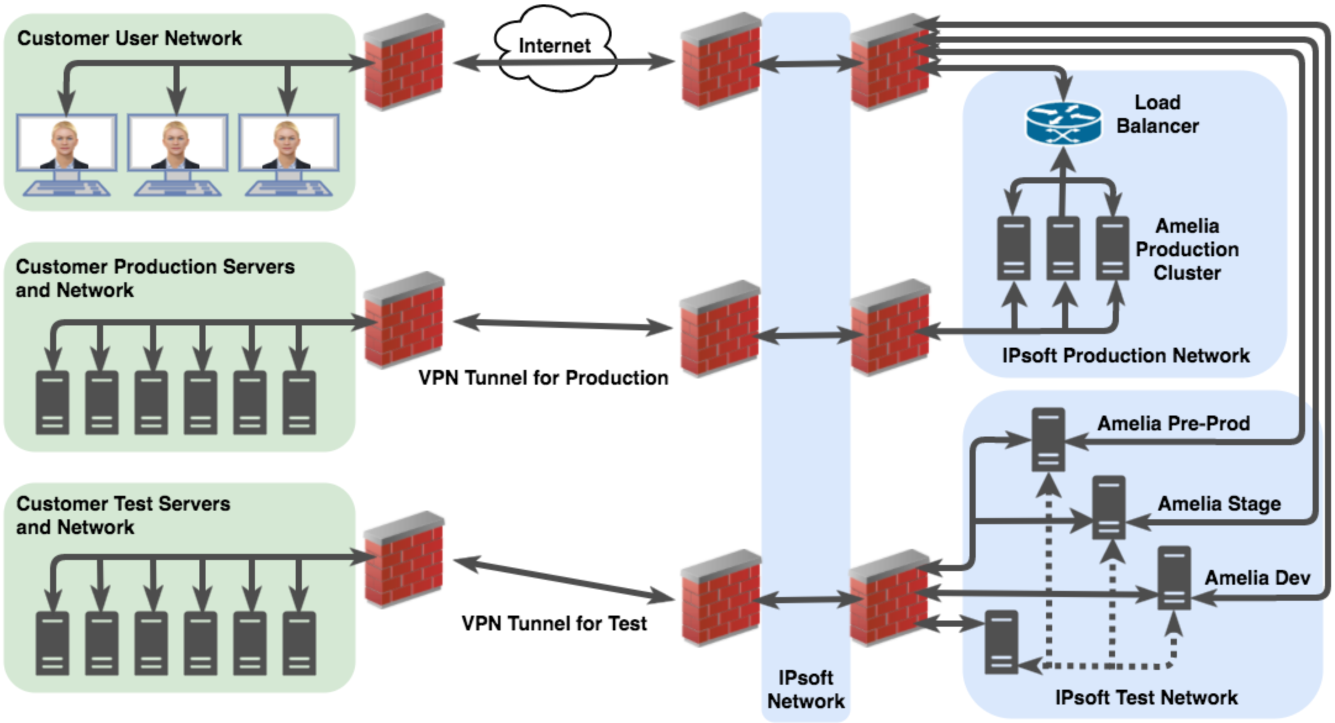

Customer User Network
Internet
Customer Production Servers and Network
VPN Tunnel for Production
Customer Test Servers and Network
VPN Tunnel for Test
IPsoft Network
Load Balancer
Amelia Production Cluster
IPsoft Production Network
Amelia Pre-Prod
Amelia Stage
Amelia Dev
IPsoft Test Network

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.

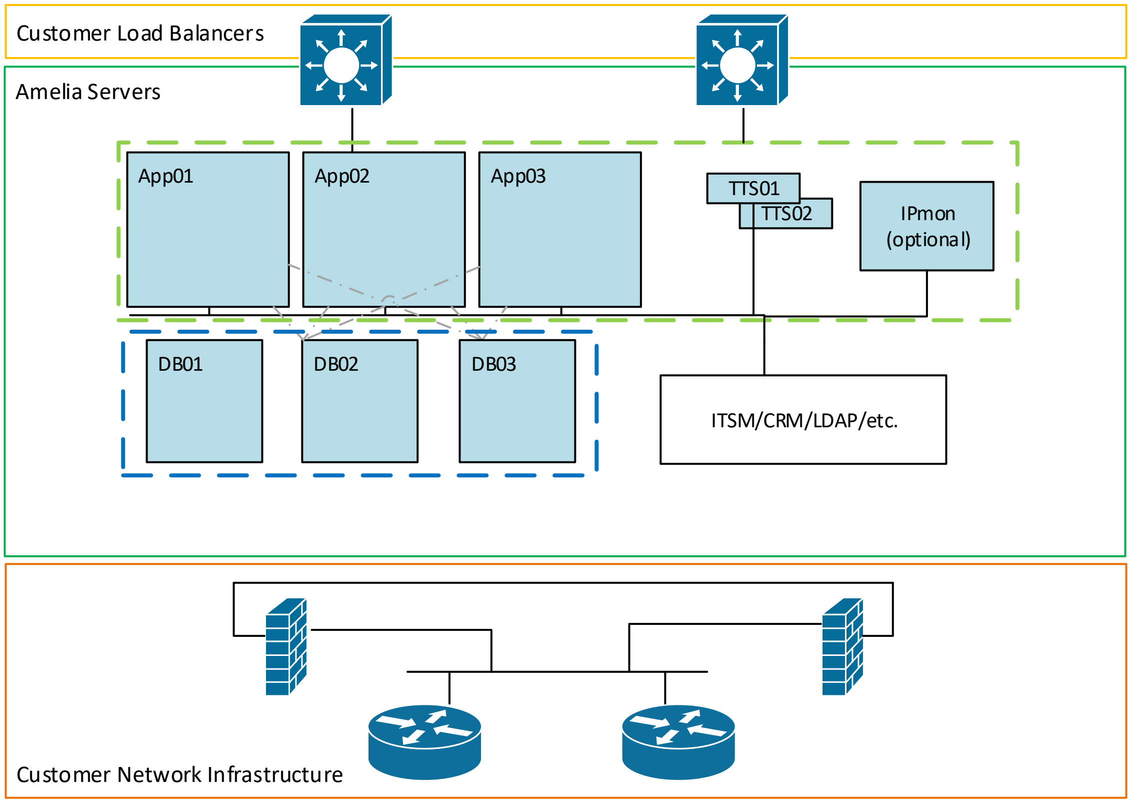


Figure 7. Amelia Hosted in Customer's Cloud

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

2.2.1 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.

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 r4.4xlarge 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.

2.2.2 Offline Deployment

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.

3. Hardware/Resource Recommendations

For deployments of Amelia on premise /cloud, there are several recommendations and requirements, depending on the type(s) of use cases and volume for each use case. Amelia can be scaled up/out by increasing the number of CPUs, RAM, and disk capacity as well the number of Conversation PODs . If Amelia is configured as a single node, It cannot be scaled-out with additional nodes.

Amelia deployments are supported in both physical and virtual environments. For virtual deployments, it is recommended to enable Memory and Disk reservations for best performance; especially in highly utilized shared infrastructure.

3.1 CPU Requirements

Amelia does not perform well in virtual infrastructures with Memory ballooning and/or CPU scheduling issues. Virtual machines (VMs) depend on available host resources (CPU, Memory), and the guest operating system consumes those resources. A problem with resource availability or scheduling inside or outside the virtual machine may cause it to become unresponsive.

Reviewing CPU performance metrics can be used to determine whether a guest operating system is actually running, whether the virtual machine monitor (VMM) is running, or whether there is scheduling contention. The metrics is also leveraging insight into the responsiveness of a virtual machine or its Guest OS:

- Run - Amount of time the virtual machine is consuming CPU resources.
- Wait - Amount of time the virtual machine is waiting for a VMkernel resource.
- Ready - Amount of time the virtual machine was ready to run, waiting in a queue to be scheduled.
- Co-Stop - Amount of time a SMP virtual machine was ready to run, but incurred delay due to co-vCPU scheduling contention.

It is recommended Memory/CPU Hotplug is enabled to increase CPUs/RAM at any time to scale quickly. Be mindful of the hardware configuration and NUMA settings, it is optimal for Amelia to access CPU/Memory resources on the same NUMA node.

Amelia should be utilizing at a minimum dual Intel® Xeon® Processor E5-2687W v3 (25M Cache, 3.10 GHz).

3.2 Storage Requirements

Amelia’s response time (latency) is important for the user’s experience , handling of high concurrent connections, and scalability. Amelia requires a low latency and high throughput storage tier. IPsoft highly recommends using Solid State Drives (SSDs) for primary storage for all Amelia Servers. SAS/SATA Storage can be used for backups if desired.

Customers and Partners can leverage existing storage platforms that are installed with SSDs. Both All-Flash-Arrays and hybrid (SSDs and Spinning Disks) are viable solutions, taking into account the initial write lands on SSD storage tier.

Amelia's databases and PODs require at minimum 10,000 IOPS, 15,000 to 20,0000 for larger deployments.

3.3 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)***	Number of Hosts
Amelia Node	16	80	512 (not including OS)	3
TOTALS	48	240	1.6TB	3

*Based on underlying server's NUMA settings, multiple sockets may be suggested
**Additional RAM/Disk Capacity for language pack support & gateways (1GB RAM / 10GB disk capacity for each)
***/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. Please see the below notes regarding proper disk capacity sizing.

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)***	Number of Hosts
Apps	12	64	300	3
Database	8	24	1024 (not including OS)	3
TOTALS	60	264	3.88TB	6

*Based on underlying server's NUMA settings, multiple sockets may be suggested
**Additional RAM/Disk Capacity for language pack support & gateways (1GB RAM / 10GB disk capacity for each)
***/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. Please see the below notes regarding proper disk capacity sizing.

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.

Databases sizes for both 3-node and 6-node clusters depends on the following:

- Data Retention Requirements/Compliance
- Local Database Backup(s)
- Use Cases
- Number of Conversation Pods

Conversation Pods are used for additional capacity for higher concurrent conversations. Conversation Pods are deployed in batches of 3 VMs to handle an additional 150 concurrent conversations. There currently is no limit on the number of Pods that can be deployed with Amelia. For deployments with high peaks, it is recommended to configure Amelia with the necessary sizing for those peaks and scale down if desired.

Table 3. Converstation POD Setup Environment Resource Requirements

	Specifications per Host			
Tier	CPU*	RAM (GB)**	Disk Capacity (GB)***	Number of Hosts
POD Node	8	32	100 (not including OS)	3
TOTALS	24	96	300 GB	3

3.4 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 4. 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/Disk Capacity for language packs & gateways (1GB RAM/10GB Disk Capacity each pack)

4. 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.

4.1 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.

4.2 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.

4.3 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.

4.4 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

4.5 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

4.6 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

4.7 AMELIA DAEMONS

4.7.1 Amelia-Admin-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.

4.7.2 User Web

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

4.7.3 Engine Service

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

4.7.4 Batch Service

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

4.7.5 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.

4.7.6 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.

4.7.7 Duckling Service

Based on Facebook's open source Duckling project, this service is implemented it within Amelia to handle multi-lingual date normalization. Duckling service uses probabilistic context free grammar based on rules consisting of patterns (regular expressions for character level and predicates for concept level matching) and productions/derivations. The modules that parse temporal expressions in English, Spanish, French, Italian and Chinese.

4.8 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.

5. 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.

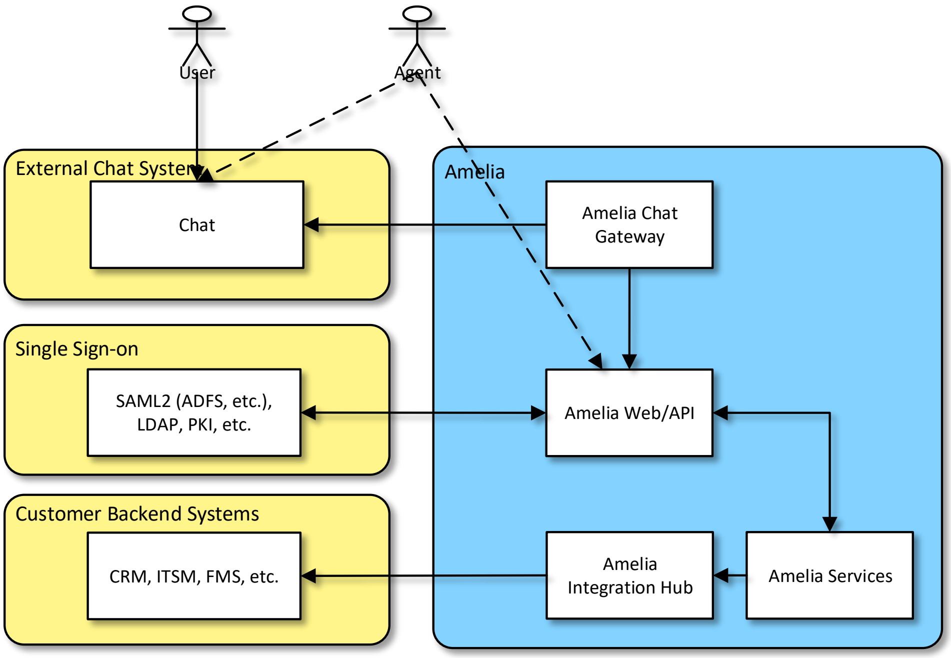


Figure 8. Chat Integration

6. Security

6.1 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.

6.1.1 Local Authentication

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

6.1.2 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.

6.1.3 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.

6.1.4 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).

- 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.

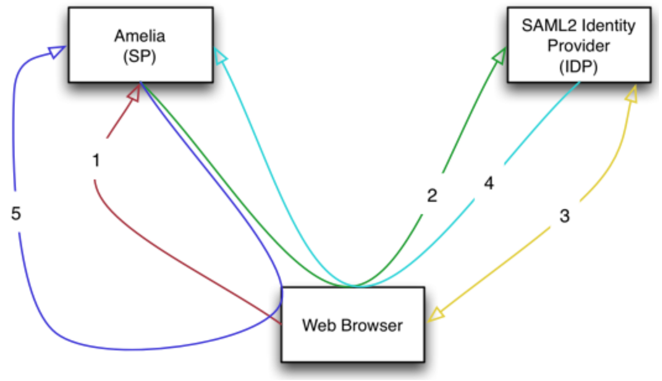


Figure 9. Amelia and Service Provider Initiated Authentication

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.

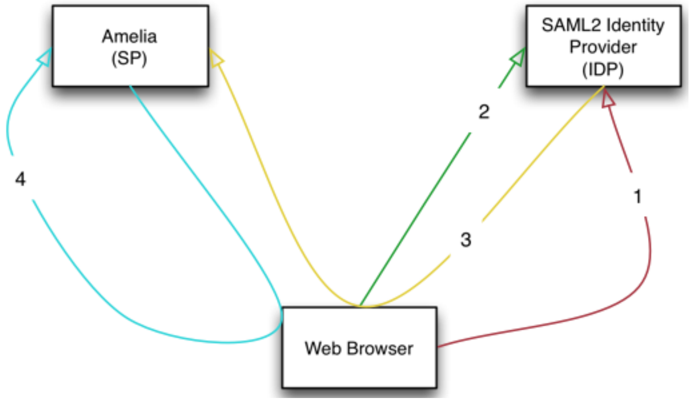


Figure 10. Amelia and Identity Provider Initiated Authentication

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.

6.2 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

6.3 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

6.4 Firewall Rules

Table 6. Firewall Rules

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

Purpose	Source	Protocol	Destination	Destination Port	Required
Statistics	Brower	HTTP	haproxy	1936	Yes
Antivirus	ipsoft-av-gateway	TCP	clamav	3310	Yes
Web UI	haproxy	HTTP (TLS 1.2)	amelia-engine-service@p001	4431	Yes
Monitoring	monitoring network	JMX	amelia-engine-service@p001	4434	Yes
RPC	haproxy	GRPC (TLS 1.2)	amelia-engine-service@p001	4435	Yes
RPC - All pods roundrobin	amelia-user-web	GRPC (TLS 1.2)	haproxy	4436	Yes
RPC - Engine pod 001 roundrobin.	amelia-user-web	GRPC (TLS 1.2)	haproxy	44001	Yes
Web UI	haproxy	HTTP (TLS 1.2)	amelia-user-web	4441	Yes
Monitoring	monitoring network	JMX	amelia-user-web	4444	Yes
UI Development	developer machine	HTTPS	webpack-dev-server	4449	For UI development only
Monitoring	monitoring network	JMX	amelia-escalation-service	4574	Yes
Monitoring	monitoring network	JMX	amelia-batch-service	4584	Yes
Web UI	haproxy	HTTP (TLS 1.2)	amelia-admin-web	4601	Yes
Monitoring	monitoring network	JMX	amelia-admin-web	4604	Yes
UI Development	deveveloper machine	HTTPS	webpack-dev-server	4609	For UI development only

Purpose	Source	Protocol	Destination	Destination Port	Required
Web UI	haproxy	HTTP (TLS 1.2)	amelia-admin-service	4611	Yes
Monitoring	monitoring network	JMX	amelia-admin-service	4614	Yes
Web UI	haproxy	HTTP (TLS 1.2)	amelia-engine-service@p002	4621	Yes
Monitoring	monitoring network	JMX	amelia-engine-service@p002	4624	Yes
RPC	haproxy	GRPC (TLS 1.2)	amelia-engine-service@p002	4625	Yes
RPC	amelia-user-web	GRPC (TLS 1.2)	haproxy	44002	Yes
Monitoring	monitoring network	JMX	amelia-integration-service	4634	Yes
RPC	haproxy	GRPC	amelia-integration-service	4635	Yes
Web UI	haproxy	HTTP (TLS 1.2)	amelia-account-service	4641	Yes
Monitoring	monitoring network	JMX	amelia-account-service	4644	Yes
RPC	haproxy	GRPC (TLS 1.2)	amelia-account-service	4645	Yes
RPC	amelia-*	GRPC (TLS 1.2)	haproxy	4646	Yes
Web UI	haproxy	HTTP (TLS 1.2)	amelia-model-server	4651	Yes
Monitoring	monitoring network	JMX	amelia-model-server	4654	Yes
RPC	haproxy	GRPC (TLS 1.2)	amelia-model-server	4655	Yes

Purpose	Source	Protocol	Destination	Destination Port	Required
RPC	amelia-*	GRPC (TLS 1.2)	haproxy	4656	Yes
REST Endpoints	haproxy	HTTP (TLS 1.2)	amelia-rest-gateway	4661	Yes
Monitoring	monitoring network	JMX	amelia-rest-gateway	4664	Yes
Monitoring	monitoring network	JMX	amelia-client-gateway	4674	Yes
Monitoring	monitoring network	JMX	amelia-client2-gateway	4684	Yes
Monitoring	monitoring network	JMX	amelia-client3-gateway	4694	Yes
Cluster Check	haproxy	HTTP	xinetd	13304	Yes
SQL	haproxy	TCP	mysql@amelia-kdb-master	13305	Yes
SQL	amelia-admin-service	TCP	haproxy	13306	Yes
SST	mysql@amelia-kdb-master	TCP	mysql@amelia-kdb-master	13307	Yes
Group Communication	mysql@amelia-kdb-master	TCP	mysql@amelia-kdb-master	13308	Yes
IST	mysql@amelia-kdb-master	TCP	mysql@amelia-kdb-master	13309	Yes
Slave Check	haproxy	HTTP	xinetd	13314	Yes
SQL	haproxy	TCP	amelia-kdb-slave	13315	Yes
SQL	amelia-engine-service	TCP	haproxy	13316	Yes
Cluster Check	haproxy	HTTP	xinetd	13324	Yes
SQL	haproxy	TCP	mysql@amelia-cdb-p001	13325	Yes
SQL	amelia-engine-service@p001	TCP	haproxy	13326	Yes

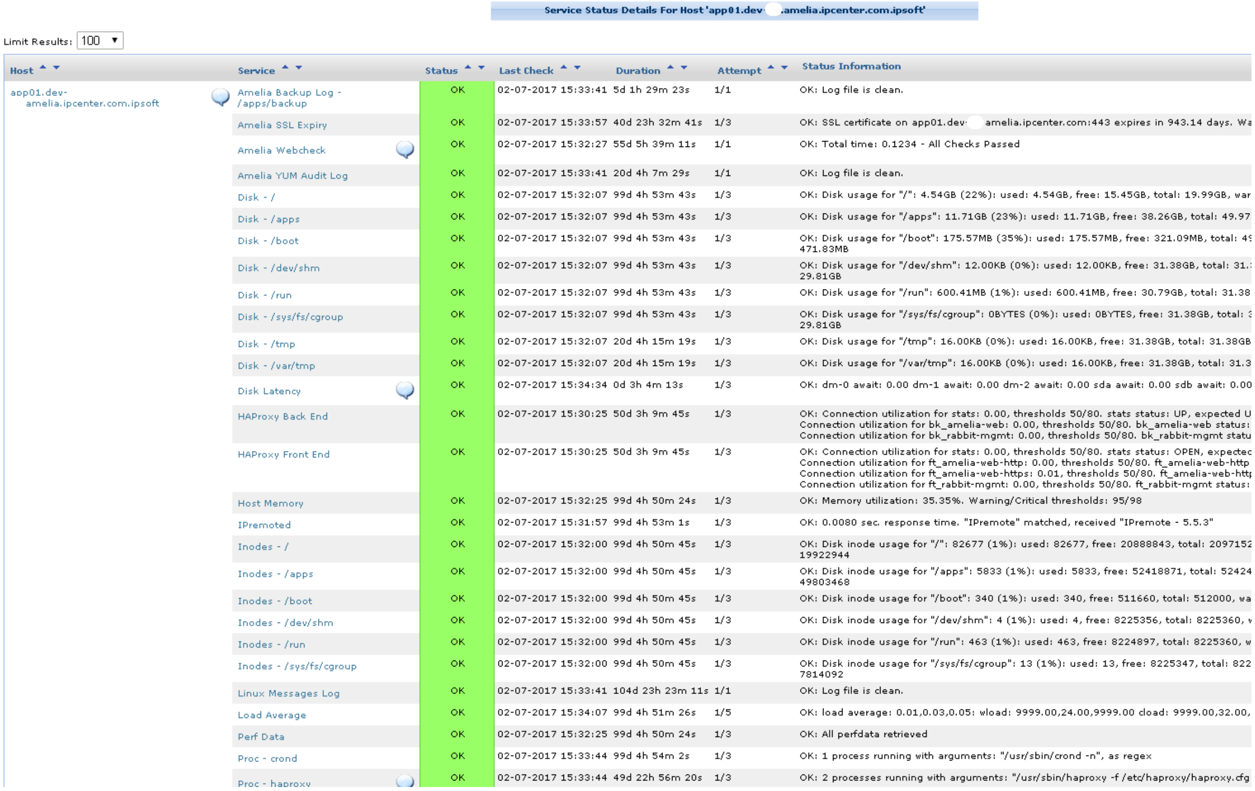
Purpose	Source	Protocol	Destination	Destination Port	Required
SST	mysql@amelia-cdb-p001	TCP	mysql@amelia-cdb-p001	13327	Yes
Group Communication	mysql@amelia-cdb-p001	TCP	mysql@amelia-cdb-p001	13328	Yes
IST	mysql@amelia-cdb-p001	TCP	mysql@amelia-cdb-p001	13329	Yes
Cluster Check	haproxy	HTTP	xinetd	13334	Yes
SQL	haproxy	TCP	mysql@amelia-cdb-p002	13335	Yes
SQL	amelia-engine-service@p002	TCP	haproxy	13336	Yes
SST	mysql@amelia-cdb-p002	TCP	mysql@amelia-cdb-p002	13337	Yes
Group Communication	mysql@amelia-cdb-p002	TCP	mysql@amelia-cdb-p002	13338	Yes
IST	mysql@amelia-cdb-p002	TCP	mysql@amelia-cdb-p002	13339	Yes
Cluster Check	haproxy	HTTP	xinetd	13344	Yes
SQL	haproxy	TCP	mysql@amelia-adb	13345	Yes
SQL	amelia-admin-service	TCP	haproxy	13346	Yes
SST	mysql@amelia-adb	TCP	mysql@amelia-adb	13347	Yes
Group Communication	mysql@amelia-adb	TCP	mysql@amelia-adb	13348	Yes
IST	mysql@amelia-adb	TCP	mysql@amelia-adb	13349	Yes
Datastore, Cache	amelia-user-web, amelia-admin-web, amelia-escalation	TCP	redis	13341	Yes

Purpose	Source	Protocol	Destination	Destination Port	Required
Redis Sentinel Cluster Management	redis-sentinel, haproxy	TCP	redis-sentinel	13342	Yes
Redis Sentinel VIP (watches all other redis clusters)	amelia-user-web, amelia-admin-web, amelia-escalation, amelia-engine, amelia-model-server	TCP	haproxy	13343	Yes
Messaging	haproxy	AMQP (TLS 1.2)	rabbitmq	13351	Yes
Messaging	amelia-*	AMQP (TLS 1.2)	haproxy	13352	Yes
Messaging	haproxy	STOMP (TLS 1.2)	rabbitmq	13353	Yes
Messaging	amelia-*	STOMP (TLS 1.2)	haproxy	13354	Yes
Monitoring	monitoring network	HTTP (TLS 1.2)	rabbitmq	13355	Yes
Clustering - Distribution	RabbitMQ	TCP	RabbitMQ	13356	Yes
Clustering - epmd	RabbitMQ/epmd	TCP	RabbitMQ/epmd	13357	Yes
Datastore, Cache	amelia-engine, amelia-model-server	TCP	redis	13361	Yes
Datastore, Cache	amelia-engine, amelia-model-server	TCP	redis	13371	Yes
Web UI	haproxy	HTTP (TLS 1.2)	amelia-model-server	13381	Yes
Monitoring	monitoring network	JMX	amelia-model-server	13384	Yes
RPC	haproxy	GRPC (TLS 1.2)	amelia-model-server	13385	Yes
RPC	amelia-*	GRPC (TLS 1.2)	haproxy	13386	Yes

Purpose	Source	Protocol	Destination	Destination Port	Required
Date-time parser	amelia-engine-service	HTTPS	haproxy	14010	Yes
Date-time parser	haproxy	HTTPS	amelia-duckling-service	14011	Yes
AV scan REST APi	amelia-*-web	HTTP (TLS 1.2)	haproxy	14020	Yes
AV scan REST API	haproxy	HTTP (TLS 1.2)	ipsoft-av-gateway	14021	Yes
Monitoring	monitoring network	JMX	ipsoft-av-gateway	14024	Yes
syntaxnet parser	haproxy	TCP	amelia-tf-syntaxnet	14000	Yes
syntaxnet parser	amelia-*	GRPC	amelia-tf-syntaxnet	14001	Yes
syntaxnet parser	amelia-tf-syntaxnet	TCP	amelia-tf-syntaxnet	14009	Yes
syntaxnet parser	haproxy-test	TCP	amelia-tf-syntaxnet-test	14090	Yes
syntaxnet parser	amelia-*-test	GRPC	amelia-tf-syntaxnet-test	14091	Yes
syntaxnet parser	amelia-tf-syntaxnet-test	TCP	amelia-tf-syntaxnet-test	14099	Yes
Flow UI	haproxy	HTTP	amelia-h2o	54321	Yes
H2O Internal Communication	amelia-h2o	TCP	amelia-h2o	54322	Yes

7. 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:



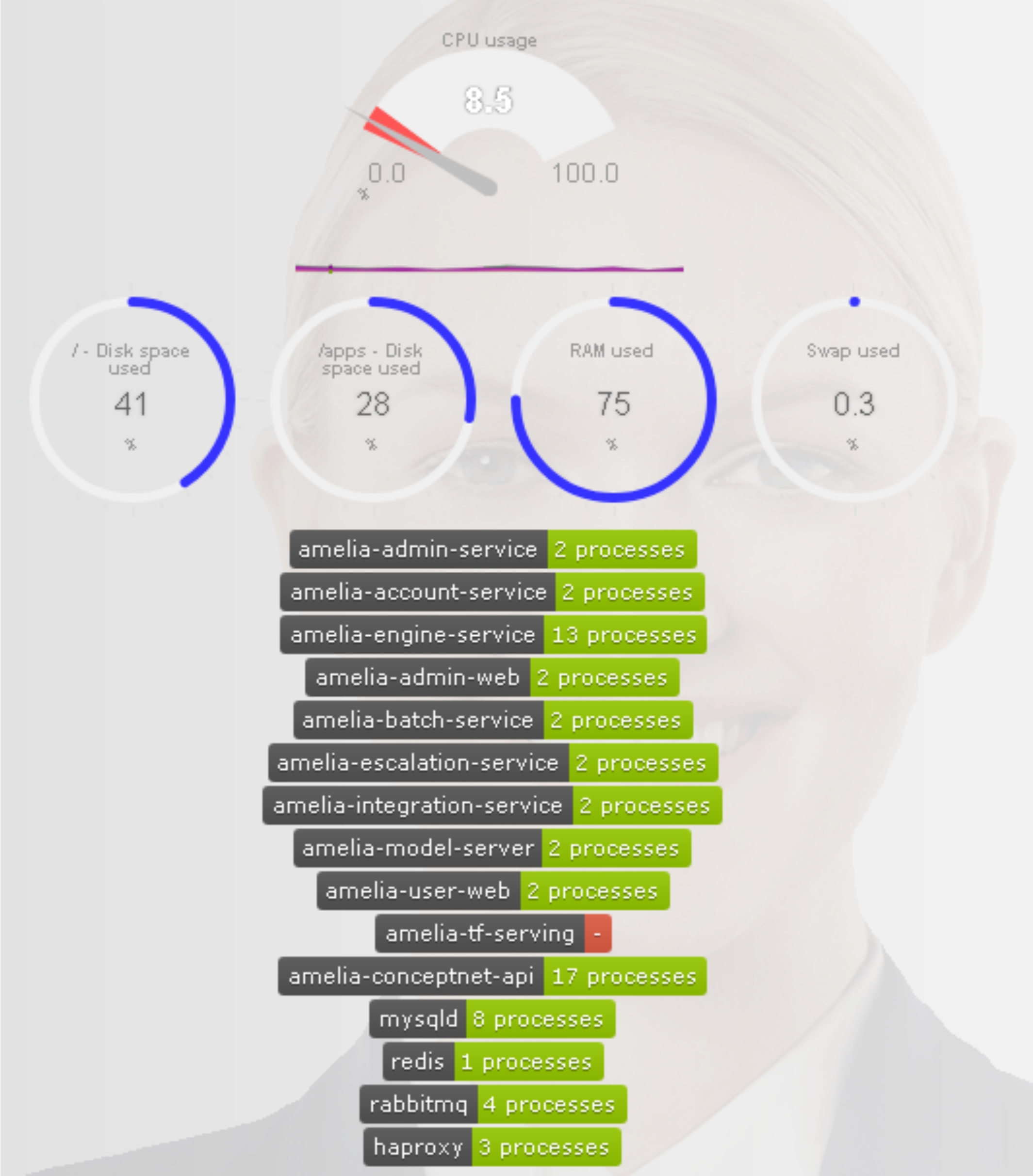
Service Status Details For Host 'app01.dev-[redacted].amelia.ipcenter.com.ipsoft'

Limit Results: 100 ▼

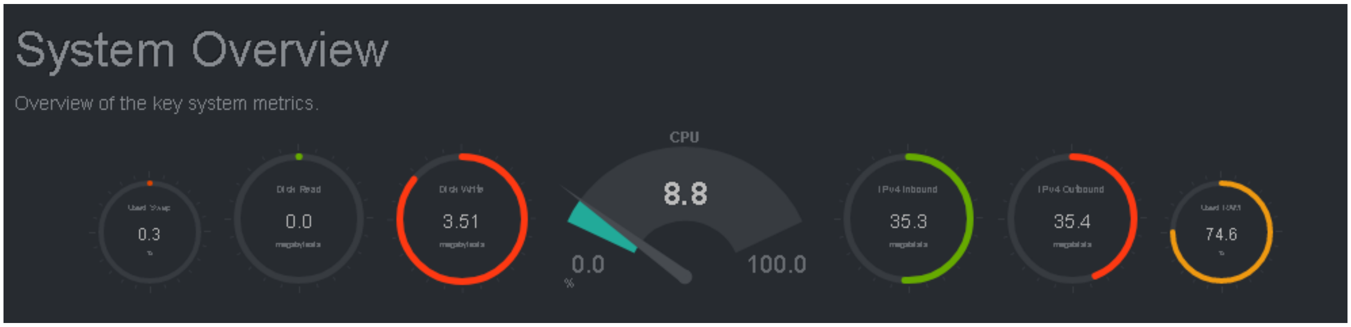
Host	Service	Status	Last Check	Duration	Attempt	Status Information
app01.dev-amelia.ipcenter.com.ipsoft	Amelia Backup Log - /apps/backup	OK	02-07-2017 15:33:41	5d 1h 29m 23s	1/1	OK: Log file is clean.
	Amelia SSL Expiry	OK	02-07-2017 15:33:57	40d 23h 32m 41s	1/3	OK: SSL certificate on app01.dev-[redacted]amelia.ipcenter.com:443 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-http 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, w
	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 11. 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.



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.

8. 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.

9. 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 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.