Virtual Design Master

# CHALLENGE 1

## THE RE-BEGINNING

ADAM POST
ADAMJPOST@GMAIL.COM
@SEMI_TECHNICAL

*Table of Contents*

# Overview and Approach

## Executive Summary

Having recovered from recent events and organized efforts to colonize Earth once more, HumanityLink Corporation is poised to move forward with an application and infrastructure re-architecture initiative.

It has been emphasized that organizational focus must remain on Terraforming efforts. Constructing, maintaining and scaling of physical facilities and infrastructure required to host HumanityLink is not feasible. Personnel have been reassigned within the organization to support the physical excavation effort, leaving limited resources dedicated to the IT environment.

Because of this constraint, and considering the availability and scalability requirements of the HumanityLink application, a Tier-1 cloud provider with a global presence has been selected. In turn, the architecture of the HumanityLink application will be adjusted to take advantage of this deployment model.

In addition to making best use of available skills and manpower, these choices prepare the application for the expected, unpredictable growth that will accompany such an ambitious initiative.

As part of the application re-architecture, HumanityLink functions will be distributed across two separate but integrated applications called HumanityLink Distributed Modeling (HLDM) and HumanityLink Fleet Management (HLFM). Appropriate infrastructure will be provided to support each.

HLDM comprises a distributed compute environment consisting of Master and Worker nodes running a proprietary processing and work-instruction engine. This environment processes future state models generated by the Design team, compares this model against current topographical maps and generates a delta model containing terraforming instructions. Excavation supervisors at the work site retrieve and distribute these instructions, with Excavation workers completing the required work according to plan.

HLFM receives uploaded health information from the Excavation fleet at the end of each shift, compares this data versus a healthy baseline, and determines repairs needed for each Excavator. These instructions are retrieved by Repair supervisors, who distribute the work to Repair workers for execution. This arrangement results in a healthy fleet, with repairs taking place every off-shift.

Within this design, components are de-coupled as much as possible, eliminating direct dependency between them. Work input and output is mediated using a highly-available and scalable hosted object storage service. Because work instructions are generated and executed in batches, operations can be completed despite unreliable connections that may be present at design or excavation sites.

To increase performance, scalability and cost-efficiency of the applications, auto-scaling is employed wherever possible within compute environments. When utilization metrics rise to a defined threshold, additional instances are deployed to handle observed load. Once this load has passed, node count will automatically return to a defined minimum, increasing overall efficiency.

Availability is enhanced by distributing compute and storage components across at least three sites, and making use of Disaster Recovery environments to protect against complete primary site failure.

In summary, it is believed that this architecture meets and exceeds all business and technical requirements and provides a solid foundation for HumanityLink that will last well into the future.

## Selection Approach

The opportunity to re-architect a demanding application like HumanityLink comes with many challenges, including selection of a proper destination platform. Although stated requirements make clear that a cloud platform is suitable based on potential scalability and high-availability advantages, selection of a specific provider can be a challenge due to near-parity in applicable, available features.

However, during the solution evaluation process it was determined that the contending solution, Microsoft Azure, was rendered inoperable by an outbreak within its facilities.

It is believed the outbreak originated at a Microsoft company gathering where the Bill and Melinda Gates foundation was prepared to demonstrate effectiveness of the serum developed as a result of its humanitarian research initiative.

As with many live demonstrations, the outcome was unexpected and the engineering team was infected when a live sample of the virus was mishandled. Available serum on-hand was not sufficient to resolve the outbreak, and full-scale production had not yet ramped up. When the engineering teams returned to their facilities, perhaps driven only by subconscious patterns embedded by routine, the infrastructure was destroyed.

Attempts to opportunistically market this “feature” to Azure cloud customers as “Meddling Zombie” have not been successful. Internally, favorable comparisons have been made to a similar open-source service named after a disorder-causing primate, but this enthusiasm is not shared by the cloud community at large. Additionally, despite biological code associated with the virus being readily available, it does not technically meet the requirement for “open source”.

Because of this catastrophic event and subsequent fallout, Microsoft Azure is considered unusable for this initiative. Amazon Web Services has thus been selected as the platform to provide infrastructure services to HumanityLink. This choice is reinforced by the superior position of AWS within the source of all truth and wisdom, the Gartner Magic Quadrant.

*New in 2017 – Azure Meddling Zombie for hands-off infrastructure resilience testing.*

## Intended Audience

This document is intended for project stakeholders within HumanityLink corporation, including business executives, technical support personnel, developers and application owners.

# Scope and Requirements

## Project Scope

The scope of this project is limited to design and implementation of an infrastructure architecture that meets the requirements of the Humanity Link Distributed Modeling and Humanity Link Fleet Management applications.

Following implementation, Design engineers will begin using an adjusted process to upload work to a new destination, which will feed the new workflow.

The team responsible for maintaining HumanityLink Scout will migrate the initial Geo data set to the new destination, as well as adjust the destination receiving ongoing updates.

Operations personnel will perform cutover of Excavation and Repair supervisor bots, both of which will retrieve work plan packages from new destinations on receipt of new object availability notifications.

Any additional work is out of scope and will require further discovery to plan and implement.

## Project Parameters

*Requirements*

| | |
|---|---|
| **R01** | Provide a globally-scalable architecture to support deployment of HumanityLink 2.0 |
| **R02** | Deliver scheduling, operations and maintenance services to the robotic fleet |
| **R03** | All systems must handle sustained load in support of 24/7 business operation |
| **R04** | Systems must adapt to unknown workloads and prevent business interruptions |
| **R05** | Systems must de-couple and take advantage of hosted services, wherever possible |
| **R06** | Data generated by HumanityLink processes must be retained indefinitely |
| **R07** | Robotics fleet must not continuously be reliant on infrastructure to perform duties |
| **R08** | Analytics functions for fleet management must be accommodated |

*Constraints*

| | |
|---|---|
| **C01** | Architecture must minimize operational burden and efficiently utilize staff |
| **C02** | Microsoft Azure has been rendered inoperable by an internal outbreak |
| **C03** | All instances must use CentOS 7 as the application has been developed against it |

*Assumptions*

| | |
|---|---|
| **A01** | The business is prepared to make whatever financial investment required for infrastructure |
| **A02** | Re-colonization of the US is highest priority, after which operations will scale globally |
| **A03** | Redundant internet connectivity is available at all excavation sites |
| **A04** | An existing Excavation and Repair fleet exists, although expansion is ongoing. |
| **A05** | Existing Excavation and Repair fleets have been prepared for cutover to the new architecture |
| **A06** | Teams responsible for HumanityLink Scout and Designer are prepared for cutover |
| **A07** | Robots will be operated in shifts to ensure continuous performance of work |

*Risks*

| | |
|---|---|
| **r01** | Environment relies exclusively on a single vendor for cloud infrastructure services |
| **r02** | Solution depends on internet connectivity between excavation and hosting sites |
| **r03** | Failure of notifications between components will result in work stoppage |
| **r04** | Production load of the HumanityLink application is unknown |

## Technical Requirements

*Availability*

| | |
|---|---|
| **RA01** | Availability of HumanityLink must be maintained in the event of component failure |
| **RA02** | Production operations must be distributed across three independent locations |
| **RA03** | Provide resiliency at every infrastructure and service layer possible |
| **RA04** | Mechanisms must be implemented to assist with transparent component failover |

*Manageability*

| | |
|---|---|
| **RM01** | Secure remote access for administration purposes must be provided |
| **RM02** | Ongoing maintenance and patching of operating systems must be accommodated |
| **RM03** | Health and utilization of the environment must be actively tracked and configured to alert support personnel |

*Performance*

| | |
|---|---|
| **RP01** | System must scale both vertically (up) and horizontally (out) as demands change |
| **RP02** | Infrastructure components must be sized to meet a moderate initial demand, as actual demand cannot be accurately projected |
| **RP03** | HLDM components require high-throughput, low-latency connectivity |
| **RP04** | Full capabilities of available hardware must be leveraged, wherever possible |
| **RP05** | The Distributed Modeling environment must accommodate a wide range of IO demands |
| **RP06** | Data resources must be immediately accessible without retrieval delay |
| **RP07** | Input and output files generated must be size optimized to reduce network load |

*Recoverability*

Description of general recoverability requirements.

| | |
|---|---|
| **RR01** | Instance backups must be taken at least once per day and retained for seven days |
| **RR02** | An RTO of one hour is required for both HLDM and HLFM environments |
| **RR03** | Future-state models, Geo data and delta models must be retained indefinitely |
| **RR04** | A fully-functional DR environment must be available for all system environments |
| **RR05** | Controls must exist to ensure activation of DR for HLDM is not performed accidentally |
| **RR06** | Bi-directional transition between primary and DR environments must be supported without excess administrative effort |
| **RR07** | Dataset consistency within HLDM and HLFM must be guaranteed following recovery |
| **RR08** | Disaster recovery environments must provide at least 50% capacity of the primary |
| **RR09** | HLFM application backups must be captured hourly and retained for thirty days |
| **RR10** | Disaster recovery for HLFM must be automatically activated during full site failure |

*Security*

| | |
|---|---|
| **RS01** | Users and components shall be provisioned lowest privileges needed to perform duties |
| **RS02** | Traffic between components must be controlled via policy and restricted, where required |
| **RS03** | Restrictions must be in place preventing unauthorized access to generated data |
| **RS04** | Administration tasks must be completed from locations with hardware IPSEC to VPC |
| **RS05** | HumanityLink Distributed Modeling must not be directly accessible from the public internet |
| **RS06** | Data transmitted to and from the HLDM and HLFM environments must be encrypted |
| **RS07** | All publicly exposed instances must possess appropriate firewall protection |

# Application Architecture

## Summary

HumanityLink Suite 2.0 contains a number of sub-systems, each of which is essential to the Terraforming initiative and overall global success of the re-colonization strategy. **(R01)**

This infrastructure re-design effort focuses on development of the new Distributed Modeling and Fleet Management environments **(R02)**. These will provide core services of the HumanityLink Suite and enable execution of Terraforming objectives to be completed at required locations.

Preparation and deployment of HumanityLink Scout and Designer has been completed **(A06)**, and the existing Excavation and Repair fleets have been prepped for cutover to the new architecture **(A04, A05)**. With this in mind, project focus will be placed solely on construction of HLDM and HLFM.

Primary components and workflow can be summarized as follows:

*Assess*

- HumanityLink Scout
    - Gathers topographic data from satellite sources, associates with GPS data
    - Uploads data updates to a dedicated S3 bucket for reference by Distributed Modeling
    - Data also serves as a starting point for engineers in HumanityLink Designer

*Design*

- HumanityLink Designer
    - CAD application operated at a secure site by a team of engineers
    - Generates a desired future-state model of land and features
    - Uploads completed work file for processing by Distributed Modeling
- **HumanityLink Distributed Modeling (Architecture Focus)**
    - Batch-based distributed compute platform for the Excavation industry
    - References topology maps gathered by HumanityLink Scout
    - Processes future-state model generated by HumanityLink Designer
    - Creates delta model and instruction set to be processed by HumanityLink Excavators

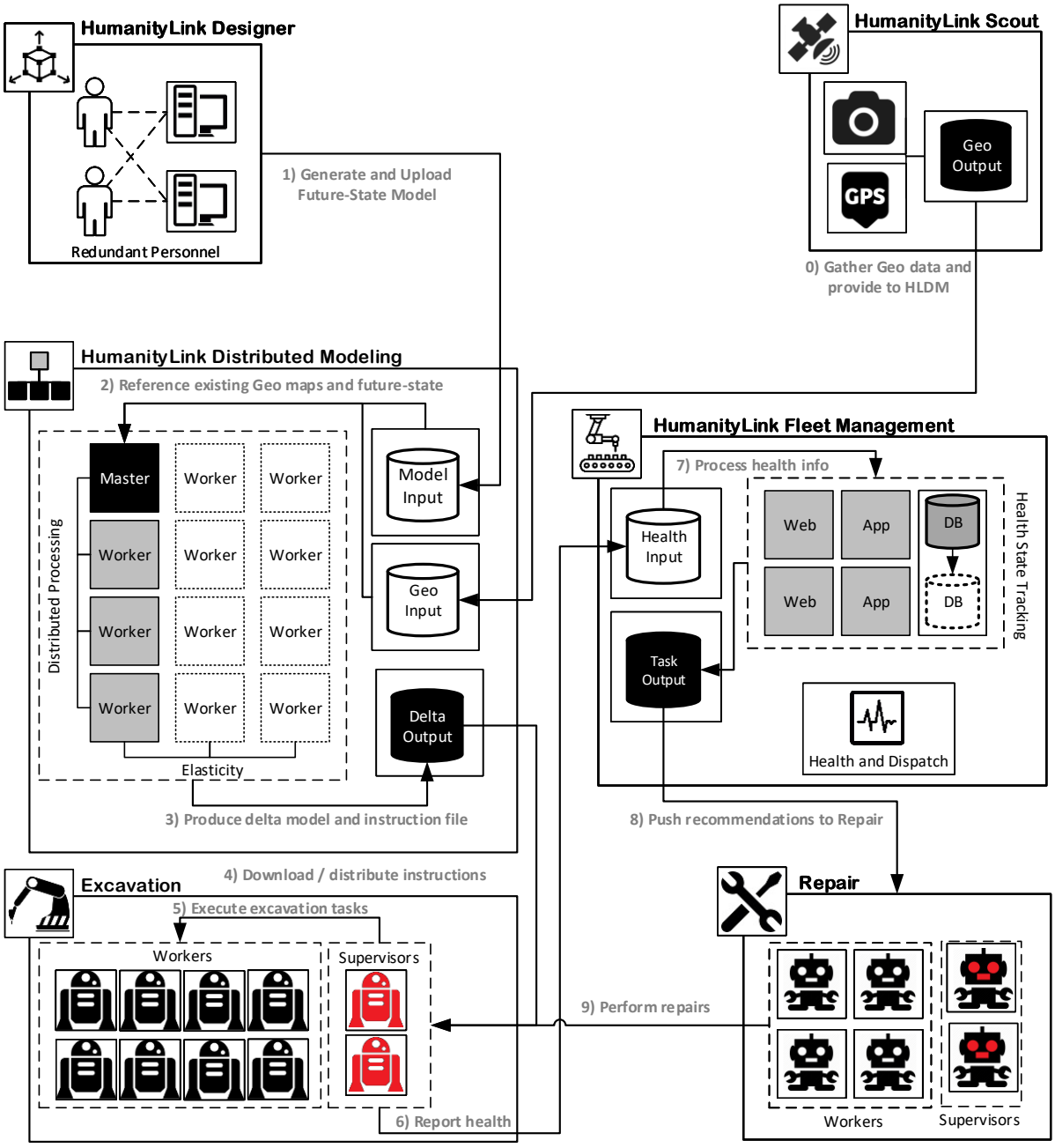
*Operate*

- HumanityLink Excavator
    - Runs on all robotic terraforming workers and supervisors
    - Retrieves delta model from HumanityLink Distributed Modeling
    - Coordinates and distributes work execution between available robots in the fleet
    - Performs Terraforming / excavation to reach desired future-state
    - Reports health status to supervisors on an ongoing basis
    - Supervisors upload fleet health status to HLFM at end-of-shift

*Maintain*

- **HumanityLink Fleet Management (Architecture Focus)**
    - Receives health data from robotic fleet and compares vs. optimal baseline spec
    - Determines repair operations and issues instructions to repair fleet
    - Provides dispatch, administration and analytics interface for fleet management

## Conceptual Diagram




*HumanityLink Suite including Distributed Modeling and Fleet Management environments.*

# Infrastructure Architecture

## HumanityLink Distributed Modeling

*Summary*

HumanityLink Distributed Modeling is a proprietary distributed computing application specifically engineered for Terraforming uses. When preparing to alter surface topography of an area, a number of key tasks are required, including:

- Conduct analysis of current topography
- Decide upon the modifications to be made
- Create a specific plan for accomplishing the objective

Distributed Modeling efficiently takes care of these tasks so that Excavation can proceed without delay **(R02)**. The distributed, scalable nature of the application also allows multiple tasks to be completed in parallel to feed multiple sites and fleets. **(R01, R03)**

Modeling begins with creation of a future-state topography design and transmission of this design to the HLDM environment. Once received, HLDM will conduct an analysis of current-state data for the referenced site, determining whether the goal can be accomplished. Passing this check allows the system to proceed to processing.

Upon initiation of a Distributed Modeling job, both current and future state data are loaded into HLDM and divided between the worker nodes equally. This processing results in a ready-to-execute plan, referred to as a delta design. Because fleet capabilities are known and built into the application, specific steps are able to be provided directly for execution.

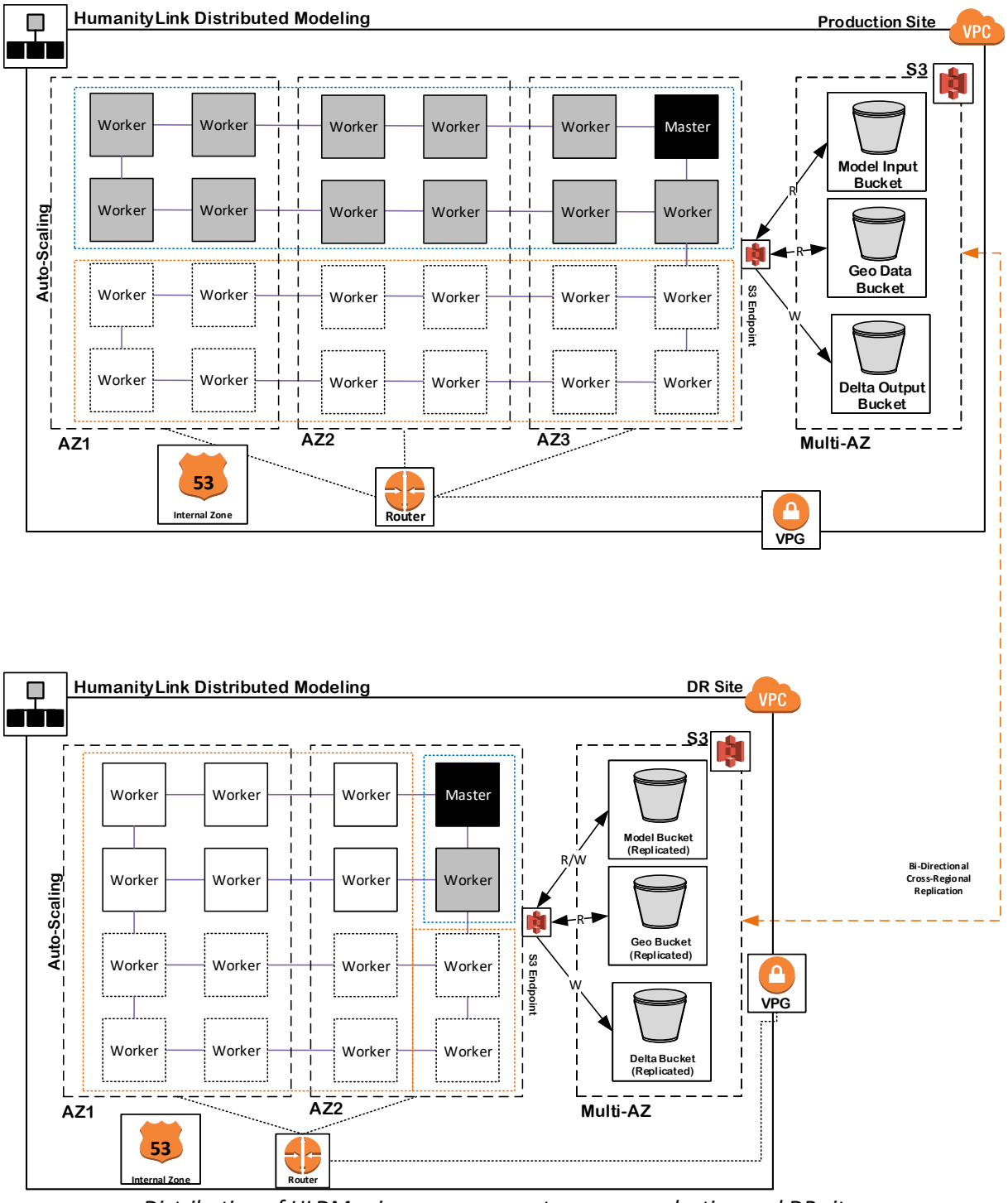
Once ready for release, the delta design is deposited in an accessible location and Excavation supervisors are notified to retrieve it **(R05)**. This data is staged to the rest of the Excavation fleet in preparation for execution.

It is important to note that dispatch functions are provided by HumanityLink Fleet Manager, not Distributed Modeling, so once the delta model is distributed the work flow for HLDM is complete.

*Solution Capacity*

| Area | Capacity (Min/Max) |
|---|---|
| vCPU | 96 / 192 |
| RAM | 732GB / 1464GB |
| Instance Storage Capacity | 6TB / 12TB |
| Steady-State IOPS | 18,000/36,000 |
| Network Throughput | 10Gbit Node-Node |
| Object Storage | Unlimited |

*Logical Diagram*




*Distribution of HLDM primary components across production and DR sites*

*Compute*

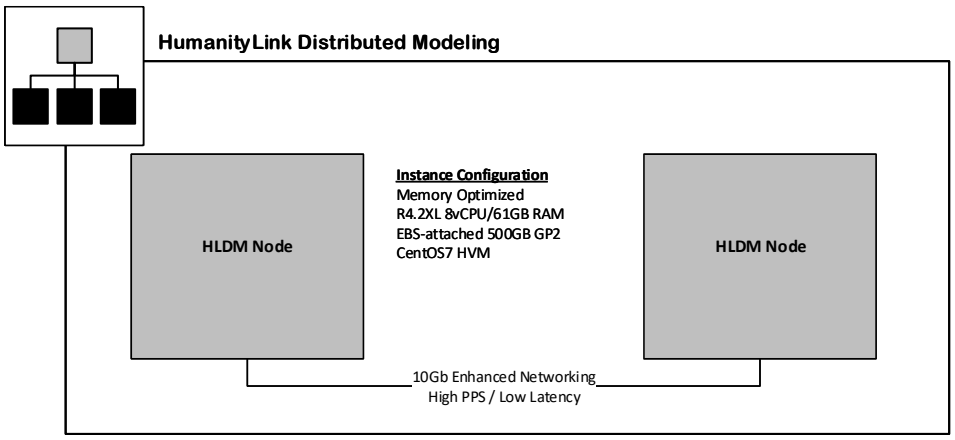
Memory-optimized instances from the R4 class have been selected to support compute-intensive operations required by HLDM. Network performance between nodes is also of high-importance, and Enhanced Networking has been enabled for all instances. All nodes will be deployed from a customized CentOS7 HVM AMI with supporting drivers present. Overall, this configuration provides balanced performance with room to scale up CPU, RAM and network throughput as new requirements emerge. Hardware GPU resources can be made available if CPU proves insufficient.

*Design Decisions*

| Area | Decision | Meets Requirement | Justification |
|---|---|---|---|
| Node Size | R4.2XL | RP02 | This selection provides a balanced mix of resources and reserves larger instances for future scale-up requirements. |
| Inter-Node Communication | Enhanced Networking | RP03 | Use of ENI (Enhanced Network Interfaces) improves packet-processing performance and reduces latency. |
| Virtualization Type | HVM | RP04 | HVM AMI's provide the highest level of performance by leveraging hardware virtualization extensions. |
| Operating System | CentOS7 Linux | C03 | HLDM was developed against this operating system and although the application will run on other distributions, management desires consistency. |
| Placement Groups | None | RA01 | A single-AZ design would not meet application availability requirements. |

*Instance Configuration*

| Quantity | Role | Type | CPU | RAM | OS | Disk |
|---|---|---|---|---|---|---|
| 12-24 | HLDM Master / Worker | R4.2XL | 8 vCPU 2GHZ Intel Broadwell | 61GB DDR4 | CentOS 7 HVM AMI | 500GB EBS GP2 |




*Virtual hardware configuration of HLDM instances and networking optimizations.*

### *Storage*

Storage workload on individual EC2 instances is expected to be mixed, with large external data loads consisting primarily of sequential write IO. Generation of delta models and work instruction is expected to be more random and IO-intensive in nature.

Each instance has been equipped with 500GB of GP2 to start, and this amount can either be scaled up (up to 16TB) by performing an online increase, or scaled out by adding disks to the instance.

SSD-based GP2 has been selected for this configuration due to its well-rounded performance characteristics. With an allocation of 500GB, 1500 steady-state IOPS are expected per-instance, with the ability to burst up to 3000 IOPS. If performance becomes constrained, volumes can be converted to IOPS-guaranteed SSD volumes (IO1) and scaled appropriately. The R4.2XL instances selected are also EBS-optimized, which guarantees dedicated bandwidth between instances and their attached disks.

In this configuration, data to be processed is distributed equally between online nodes, and it is not expected that a dataset will exceed the aggregate capacity of all nodes. Management performed by HLDM will queue work if storage consumption reaches high levels, preventing system instability.

Providing scalable object storage to the environment are several S3 buckets, each configured for specific datasets and component interactions. Because high-performance, low-latency access is a requirement, S3 Standard has been selected for all uses. Files deposited into S3 will make use of multi-part uploads to eliminate excessive restarts when or if connection quality degrades.

This configuration is expected to provide a high amount of initial performance, as well as the ability to scale in multiple dimensions to meet future requirements.

*Design Decisions*

| Area | Decision | Meets Requirement | Justification |
|---|---|---|---|
| EBS Volume Type | GP2 | R04<br>RP05<br>R04 | Meets mixed workload requirements of HLDM nodes. |
| EBS Volume Quantity | 2<br>Independent | RR01 | Independent volumes software RAID supports the requirement to leverage EBS snapshots. |
| EBS Volume Size | 60GB OS<br>440GB Data | R03<br>RP02 | This allocation provides 1500 IOPS/node and sufficient capacity to meet expected requirements. |
| EBS Volume Scaling | Up / Out | RP01 | Performance and capacity can both be optimized by increasing disk size or attaching additional disks. |
| S3 Storage Type | Standard | RP06 | Supports requirement for high-performance low-latency object storage performance. |
| S3 Archival | 30D S3 IA Transition | R06<br>RP06 | Migrating unused data to S3 IA will maintain immediate access to resources and allow indefinite archival without excess cost. |

### *Manageability*

#### *Administration and Maintenance*

Administrative access is provided to the technical team via IPSEC connection to a Virtual Private Gateway attached to each VPC. No other inbound access is provided. Because of this, administrative functions must be performed from a facility that maintains a hardware VPN connection to the VPC.

Ongoing maintenance of system images will be completed by the Operations team, and a golden, updated AMI will be maintained for HLDM nodes. System updates will be facilitated primarily via Configuration Management, but updates will continue to the AMI in order to minimize post-deployment configuration time.

#### *Monitoring*

AWS CloudWatch will be leveraged to monitor cloud infrastructure resources, and alarms will be sent to a distribution list monitored by the Operations ticketing system. Default utilization thresholds and fault monitors will be used until application workload is fully understood, after which thresholds may be adjusted. Ongoing adjustment will result in trustworthy alerts and reduction of administrative effort.

#### *Configuration Management*

Upon launch of a new node manually or via Auto-Scaling, a pre-configured Cookbook is obtained from AWS OpsWorks for Chef Automate. Upkeep of Cookbooks and Chef servers is handled by the HL Operations team.

#### *Notifications and Scheduling*

Workflow components residing downstream, including HLDM Master nodes, Repair and Excavation supervisors will be configured to receive notifications via SNS when objects are uploaded to their respective buckets.

This will allow coordination of work execution without maintenance of a steady-state connection or direct communication between components. Transmissions will be initiated only as-needed.

Should work files become available before optimal shift start time, logic built into the Excavation and Repair components will stage the plans and delay execution until triggered by HLFM.

#### *Design Decisions*

| Area | Decision | Meets Requirement | Justification |
|---|---|---|---|
| Administrative Access | IPSEC Only | RM01<br>RS04 | Security requirements dictate no access to the public internet should be present. |
| Ongoing Maintenance | OpsWorks + AMI Patching | RM02 | Use of these tools will ensure the environment remains up to date and without configuration drift. |
| Monitoring Platform | Cloudwatch | RM03 | CloudWatch is an AWS-native service and can provide underlying infrastructure metrics others cannot. |
| Event Notifications | Simple Notification Service | R05<br>r03 | Redundant SNS messages will be used to notify subscribers of new object availability in S3. |

*Availability*

To improve application availability, compute nodes have been distributed across multiple availability zones, which represent separate physical sites inter-connected by redundant high-speed links. Should a worker node or AZ fail, remaining workers process the load until a replacement can be provisioned.

If a master modeling node fails, an election takes place mediated by the HumanityLink Distributed Modeling application. In this way, health and performance of the modeling environment is maintained even during component failure.

Underlying block storage disks provided by Amazon EBS are replicated across three availability zones, as well, preventing impact from hardware failure. Replication of object data in S3 takes place in two phases. First, data is replicated across the region, resulting in three copies of each object. Secondly, cross-region replication is enabled for all buckets, enabling recovery of the operations within the DR environment should disaster occur.

On the networking front, Virtual Private Gateways provide multiple, highly-available connection destinations for administrative access. In conjunction with the recommendation to have multiple WAN connections present at design and terraforming sites, access to the environment should be continuously available. DNS services will be hosted by Route 53, ensuring these core functions are highly-available and globally-scalable.

*Design Decisions*

| Area | Decision | Meets Requirement | Justification |
|---|---|---|---|
| Cloud Platform | Amazon Web Services | R01<br>C01<br>C02 | Provides global presence, robust availability and minimizes operational overhead. Use of multiple independent regions minimizes single provider risk. |
| Compute Availability | Instance Distribution across AZ's | RA01<br>RA02<br>RA03 | This configuration results in continued operation of the application in event of node or AZ failure. |
| Instance Storage Availability | EBS Replication | RA01<br>RA02<br>RA03 | Use of EBS volumes ensures three copies of data exists within the region, offering continued availability during HW failure. |
| Object Storage Availability | S3 Standard Replication | R01<br>R04<br>RA01<br>RA02<br>RA03 | S3 standard replicates each object three times within a region, providing protection against hardware failure. Cross region replication is also used, protecting against entire region failure. |
| VPN Availability | VPG Native HA | RA03 | Virtual Private Gateways provide multiple connection addresses, ensuring connectivity to the VPC can be re-established if a VPG component fails. |
| Site Access Availability | Redundant WAN connections | r02<br>RA01<br>RA03 | Use of multiple WAN connections protects against WAN provider failure. |
| DNS Availability | Amazon Route 53 | R01<br>C01 | R53 will be used for all internal/external DNS due to its resilient characteristics |

*Scalability*

HLDM has been designed to support many modeling operations in parallel, which will allow a single environment to provide delta models and work instruction for many excavation sites. As demand fluctuates, instances will be added to and removed from the compute farm in support of scalability objectives. Available compute capacity is unconstrained and as much can be consumed as is needed.

S3 object storage is massively scalable in performance and capacity, delivering well in excess of expected demand. As HumanityLink scales globally, CloudFront can be added to support performant retrieval of work instruction packages. This will allow content to be distributed to edge locations nearest the excavation site, minimizing latency and increasing download performance. However, operations will start in the US. (A02)

Despite scaling measures described here, workload may eventually justify creation of separate Distributed Modeling environments per region where excavations are taking place. This would allow more operations to proceed in parallel without contention between them. Until that point, scaling up and out of the proposed environment is recommended.

*Design Decisions*

| Area | Decision | Meets Requirement | Justification |
|---|---|---|---|
| EC2 Scale-Out | EC2 Auto-Scaling | R03<br>R04<br>RP01<br>r03 | Combined with configuration management tools, use of AutoScaling provides ongoing environment right-sizing in support of application demand. |
| EC2 Scale-Up | Manual Adjustment | R03<br>R04<br>RP01<br>RP02 | If scaling up is required, instances can be adjusted offline. Instances have been sized to ensure resource increase options remain. |
| EBS Scale-Out | Additional Disks | RP01 | Scaling out of EBS can be accomplished by adding up to 40 volumes to a Linux-based instance. |
| EBS Scale-Up | Size Increase or IO1 Conversion | R04<br>r03<br>RP01 | With a ratio of 3 IO/GB, performance of GP2 can be increased by increasing volume size. Conversion to IO1 will also accomplish this goal with up to 50 IO/GB |
| S3 Performance Scaling | Native load distribution | R01<br>R03<br>RP01 | S3 Standard is designed to increase in performance as quantity of parallel operations increases. |

*Auto-Scaling Rules*

1) **HLDM_Node_AutoScale – Min 12, Max 24, Desired 12**
    - Launch Config - HLDM AMI, R4.2XL, HLDM Role and Security Groups, Auto-IP
    - Scale Up Plan - CPU Utilization >90% for 15min, add 3 instances
    - Scale Down Plan – CPU Utilization <40% for 60min, subtract 3 instances
2) **HLDM_DR_Node_AutoScale – Min 2, Max 16, Desired 2**
    - Scale Up Plan - CPU Utilization >90% for 5min, add 6 instances
    - Scale Down Plan – CPU Utilization <40% for 60min, subtract 3 instances

### *Recoverability*

Because the configuration of master and worker nodes is easily reproducible and all data of value is stored in S3, scheduled daily EBS snapshots will suffice for instance recovery purposes. These automated snapshot actions will be performed by CloudRanger, an external automation solution. Snapshots will only be taken of OS disks, and these disks have been sized to allow recovery in under an hour (60GB).

Snapshots will persist for a week, and upon creating the eighth snapshot the oldest will be deleted. Should an instance fail, recovery from snapshot or re-deployment using a preconfigured AMI will be possible. OS state will be crash-consistent, so re-processing will take place if application-level integrity checks fail.

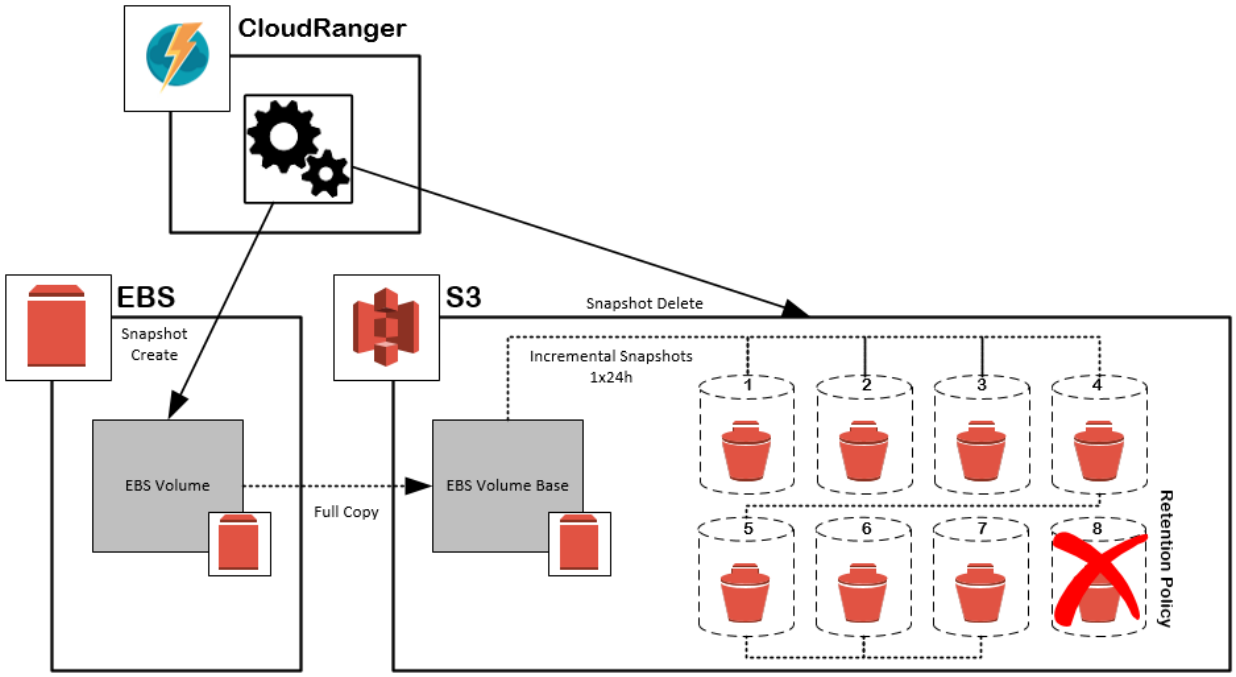
Data stored in S3 will be retained indefinitely until storage patterns of the application are fully understood. At that time, a new retention policy can be created that performs archival to Glacier or deletion, as appropriate. In support of data preservation, delete rights have been restricted and versioning has been enabled.

As part of the initial deployment, a lifecycle rule will migrate data unused in over 30 days to S3-Infrequent Access, which will reduce cost but maintain immediate, low-latency access to the data.

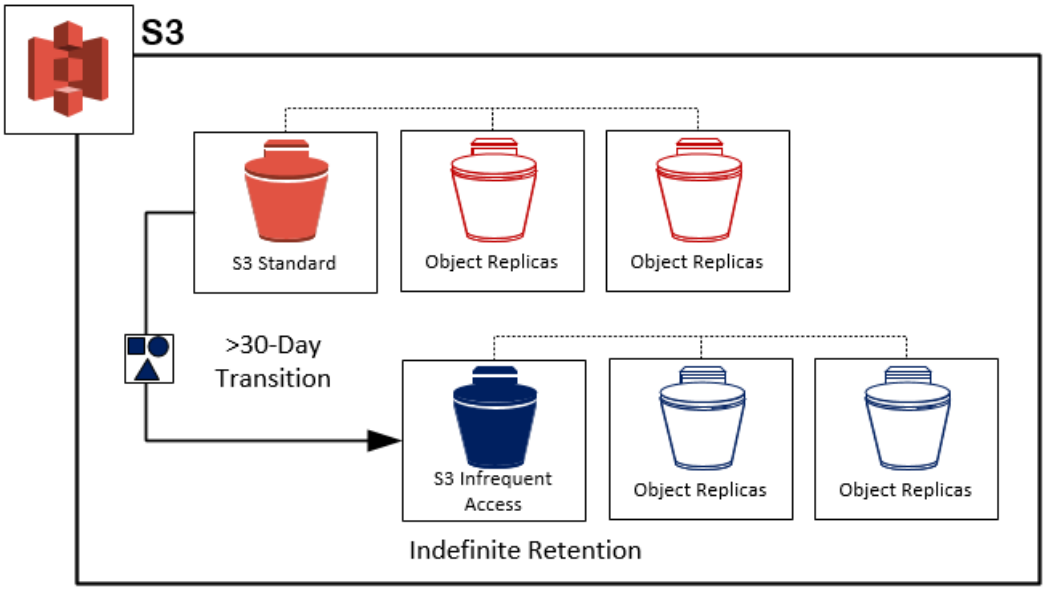
*Design Decisions*

| Area | Decision | Meets Requirement | Justification |
|---|---|---|---|
| EC2 Backup Method | EBS Snapshots | RR01<br>RR02 | Instance OS disks can be recovered using EBS snapshots within the defined RTO of 1 hours. (60GB) |
| EC2 Backup Product | CloudRanger | RR01 | Facilitates automated EBS snapshots without requiring deployment of additional components. |
| EC2 Backup Frequency | Once daily | RR01 | One snapshot per day meets defined RPO of 24 hours. |
| EC2 Backup Retention | One week | RR01 | Retaining backups for one week meets the defined retention requirement. |
| Data Consistency Model | Crash Consistent + Integrity Check | RR07 | Snapshots are crash consistent. Upon resuming processing, app-level integrity checks will take place and data will be re-processed, as required. |
| S3 Retention Policy | Indefinite | R06<br>RR03 | An initial indefinite retention policy will allow storage patterns to be observed over time. Policy can then be adjusted, as required. |
| S3 Lifecycle Policy | 30D Transition to IA | R06<br>RR03<br>RP06 | Transitioning unused data to IA supports indefinite archival goals and preserves immediate access without adding excess cost. |

*Logical Diagram*




*Automated snapshot creation and deletion via API call facilitated by 3rd party service.*




*Transition of objects between classes of storage according to lifecycle rules.*

### *Disaster Recovery*

In support of the requirement to make distributed modeling services continuously available, a dedicated disaster recovery environment has been provided that uses the same architecture as the primary. Size of the initial deployment has been scaled down with the expectation that this site will only be leveraged, as needed.

S3 data is replicated to the disaster recovery environment via cross-region replication policies defined on each of the three production buckets. DR buckets are also configured to replicate data back to the primaries to support data consistency following failback.

HLDM nodes are configured to reference local buckets when loading and processing jobs. Future-state designs will be uploaded using the alternate location. Downstream components will be notified of work file availability using SNS notifications as normal, except data will originate from the DR environment.

Initiation of DR is a simple but manual process involving starting services on the HLDM master node via SSH connection. This is intended to prevent accidental activation of the Disaster Recovery environment. To improve responsiveness, a process will be defined and key staff will be educated on its use.

*Design Decisions*

| Area | Decision | Meets Requirement | Justification |
|---|---|---|---|
| DR Approach | Fully-functional Standby environment | RR04 | All components required to operate HLDM will be online and available for failover at all times. |
| DR Activation | Manual service activation | RR02<br>RR05 | Allows fast activation of DR, when needed, and ensures cutover is intentional. |
| Data Availability Mechanism | Cross-Region Replication | RR06<br>RR07 | All content within primary buckets will be replicated across regions to matching destination buckets. |
| Data Consistency Mechanism | Bi-Directional Replication Policy | RR05 | Content between primary and DR buckets will be synchronized to support seamless failover and failback. |
| DR Capacity | >50% | RR08 | HLDM DR environment is configured to exceed 50% primary capacity via scaling |

*Solution Capacity*

| Area | Capacity (Min/Max) |
|---|---|
| vCPU | 16 / 128 |
| RAM | 122GB / 976GB |
| Instance Storage Capacity | 1TB / 8TB |
| Steady-State IOPS | 3,000/24,000 |
| Network Throughput | 10Gbit Node-Node |
| Object Storage | Unlimited |

*Security*

Using security groups, administrative access to all instances is limited to SSH via the attached Virtual Private Gateway. Inter-node communication is unrestricted, and access to the S3 endpoint is allowed via HTTPS. Otherwise, an internet gateway is not present and instances have no direct access to the outside world. Access lists between networks are not employed, as requirements do not justify their use.

All communication between HLDM and outside system components is mediated through S3. Rather than maintaining a steady-state connection, those components are notified of S3 object creation activity via SNS. This intermediate layer significantly reduces the attack surface of the distributed computing environment. Communication to OpsWorks, SNS and R53 will utilize internal SSL endpoints, as well.

To facilitate access to appropriate S3 resources, all HLDM instances are launched into an IAM role with pre-configured permissions. Reads and writes originating from outside the VPC will make use of IAM users and access keys. No permissions exist outside those needed by system components to process data.

To protect against leakage of data resulting from an attack, HumanityLink components encrypt data client-side before upload to S3. During distributed processing, HLDM manages encryption of data at the application level. Because of these measures, AWS encryption features are not required.

*Design Decisions*

| Area | Decision | Meets Requirement | Justification |
|---|---|---|---|
| Administration Access Control | IAM Users | RS01 | Select administrators have global administrative rights. All other users are allocated only required permissions. |
| Instance Access Control | Security Groups | RS02 | Security groups control traffic between instances, endpoints and gateways and restrict access, where required. |
| S3 VPC Access Control | IAM Roles + Bucket Policies | RS01 RS03 | All instances have been launched into an IAM role, and this role has been given only the permissions required to conduct normal operations. |
| S3 Outside-VPC Access | IAM Users + Bucket Policies | RS01 RS03 | Components residing outside of the VPC environment are configured to use access keys provisioned to specific IAM User accounts. Only required permissions are present. |
| S3 Encryption | SSL + Client Side Encryption | RS06 | Data is encrypted by HLDM, Designer and Scout before upload to S3, protecting against unauthorized access |
| EBS Encryption | None | RS06 | HLDM manages data encryption during distributed processing. No EBS encryption is required. |
| VPC Internet Access | None | RS04 RS05 | No IGW is attached to either the primary or secondary site VPC, preventing access to the nodes via public internet. |

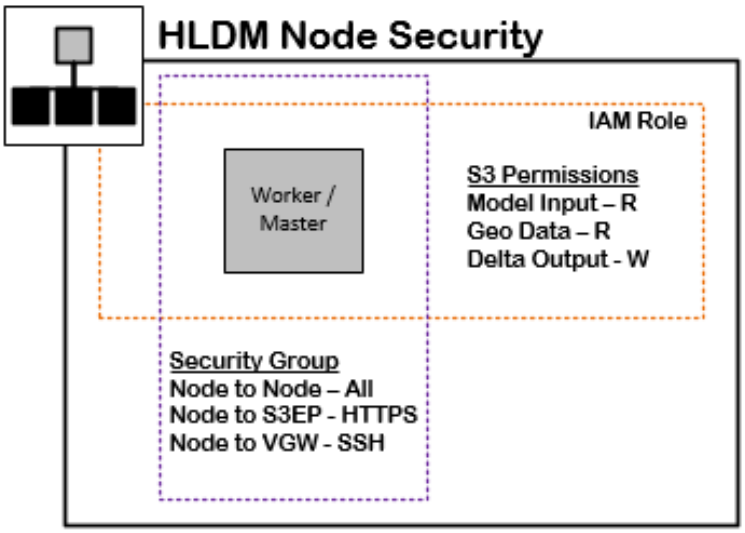
*Roles and Users*

1) **Role - HLDM Node**
    - Model Input Bucket – Read
    - Geo Data Bucket – Read
    - Delta Output Bucket – Write
2) **User – Designer**
    - Group – S3ModelUploaders
    - Write - Create object in Model Input bucket
3) **User – Excavation Supervisor**
    - Group – S3DeltaDownloaders
    - Read - Retrieve object from Delta Output bucket
4) **User – Scout**
    - Group – S3GeoUploaders
    - Write – Create object in Geo Data bucket

*Security Groups*

1) **Node to Node**
    - In / Out - Permit all traffic
2) **Node to S3 Endpoint**
    - In - HTTPS allow / Out – All

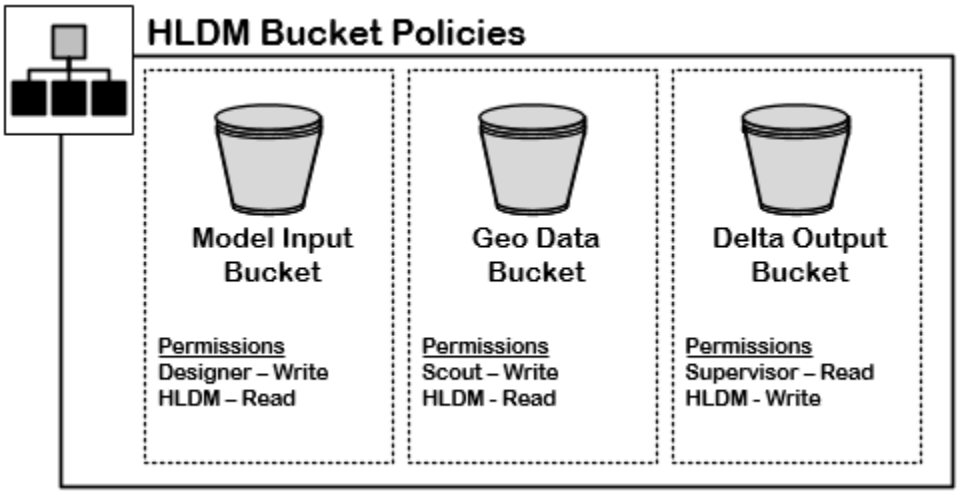
3) **Node to VGW**
    - In - SSH Allow / Out - All




*Security Groups restricting access between HLDM nodes and network resources*

*S3 Permissions*

1) **Model Input Bucket**
    - Write access for CAD users depositing models using HumanityLink Designer
    - Read access for Master and Worker nodes retrieving the future-state model
    - No delete access granted
2) **Geo Data Bucket**
    - Write access for topographical data generated by HumanityLink Scout
    - Read access for Master and Worker nodes referencing current-state data
    - No delete access granted
3) **Delta Output Bucket**
    - Write access for Master and Worker nodes depositing delta model and work plan
    - Read access for Excavation Supervisors retrieving and distributing work plan
    - No delete access granted




*Configuration of S3 Bucket Policies used by Distributed Modeling*

## *Physical Configuration*

### *Production Networks*

| Private Network | Network | Usable IP's |
|---|---|---|
| us-east-1a | 172.31.0.0/20 | 4091 |
| us-east-1b | 172.31.16.0/20 | 4091 |
| us-east-1c | 172.31.32.0/20 | 4091 |

### *Production Instances*

| Name | IP | Type | CPU | Memory | Disk 1 | Disk 2 | OS | Apps |
|---|---|---|---|---|---|---|---|---|
| HLDM01 | 172.31.0.5 | R4.2XL | 8 | 61 | 60 | 440 | CentOS7 | HLDM v1.0 |
| HLDM02 | 172.31.16.5 | R4.2XL | 8 | 61 | 60 | 440 | CentOS7 | HLDM v1.0 |
| HLDM03 | 172.31.32.5 | R4.2XL | 8 | 61 | 60 | 440 | CentOS7 | HLDM v1.0 |
| HLDM04 | 172.31.0.6 | R4.2XL | 8 | 61 | 60 | 440 | CentOS7 | HLDM v1.0 |
| HLDM05 | 172.31.16.6 | R4.2XL | 8 | 61 | 60 | 440 | CentOS7 | HLDM v1.0 |
| HLDM06 | 172.31.32.6 | R4.2XL | 8 | 61 | 60 | 440 | CentOS7 | HLDM v1.0 |
| HLDM07 | 172.31.0.7 | R4.2XL | 8 | 61 | 60 | 440 | CentOS7 | HLDM v1.0 |
| HLDM08 | 172.31.16.7 | R4.2XL | 8 | 61 | 60 | 440 | CentOS7 | HLDM v1.0 |
| HLDM09 | 172.31.32.7 | R4.2XL | 8 | 61 | 60 | 440 | CentOS7 | HLDM v1.0 |
| HLDM10 | 172.31.0.8 | R4.2XL | 8 | 61 | 60 | 440 | CentOS7 | HLDM v1.0 |
| HLDM11 | 172.31.16.8 | R4.2XL | 8 | 61 | 60 | 440 | CentOS7 | HLDM v1.0 |
| HLDM12 | 172.31.32.8 | R4.2XL | 8 | 61 | 60 | 440 | CentOS7 | HLDM v1.0 |

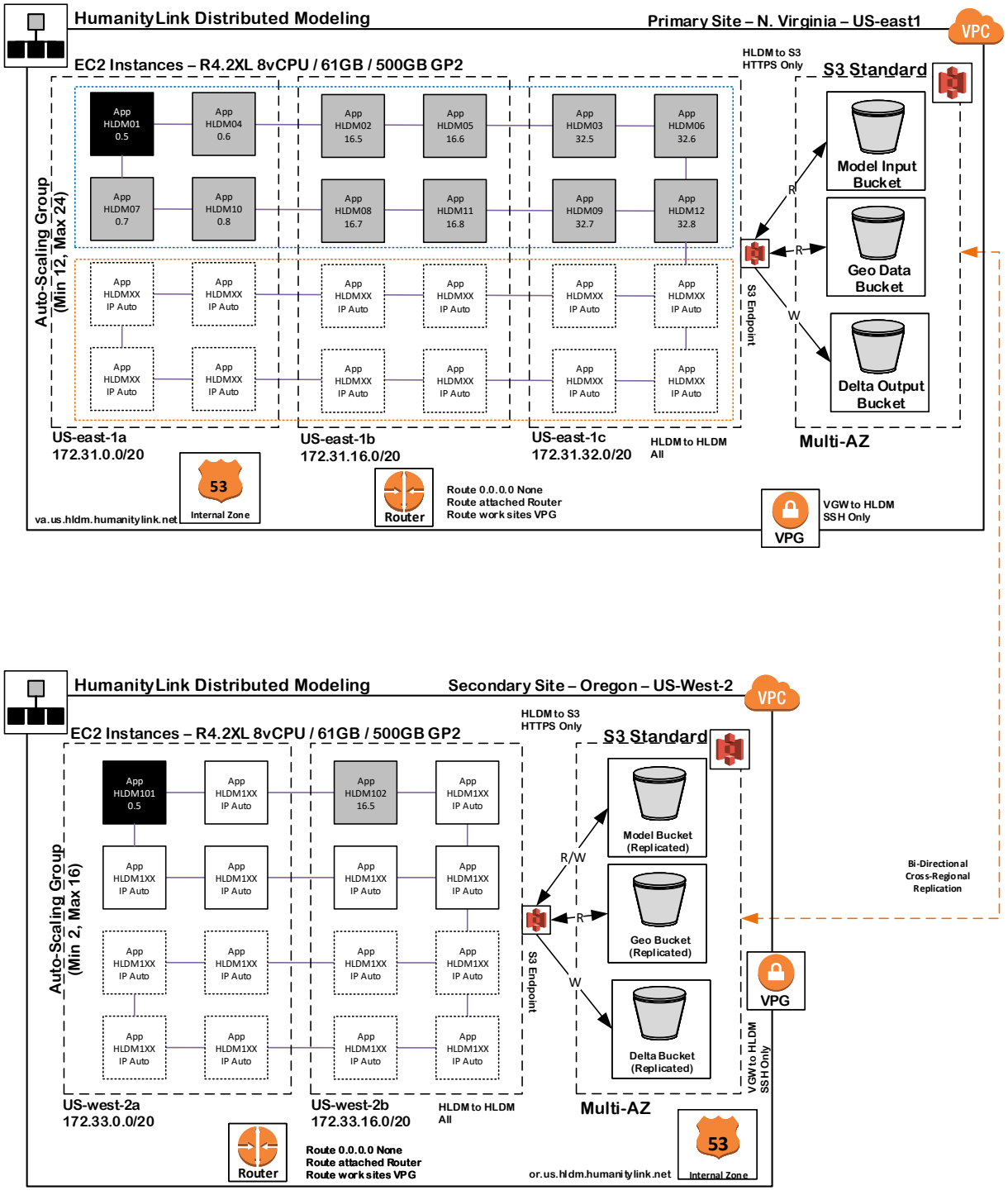
### *Disaster Recovery Networks*

| Private Network | Network | Usable IP's |
|---|---|---|
| us-west-2a | 172.33.0.0/20 | 4091 |
| us-west-2b | 172.33.16.0/20 | 4091 |

### *Disaster Recovery Instances*

| Name | IP | Type | CPU | Memory | Disk 1 | Disk 2 | OS | Apps |
|---|---|---|---|---|---|---|---|---|
| HLDM101 | 172.33.0.5 | R4.2XL | 8 | 61 | 60 | 440 | CentOS7 | HLDM v1.0 |
| HLDM102 | 172.33.16.5 | R4.2XL | 8 | 61 | 60 | 440 | CentOS7 | HLDM v1.0 |

*Physical Diagram*




*Physical configuration of Humanity Link Distributed Modeling primary and DR sites*

## HumanityLink Fleet Management

*Summary*

HumanityLink Fleet Management is a multi-purpose application with a broad range of capabilities essential to operation of robotic Excavation and Repair fleets. **(R02)**

This application receives aggregated fleet health information from Excavation supervisors at the conclusion of each work shift **(A07)**. This data is fed into the HLFM environment and processed by multiple nodes performing operations in parallel. Once health data is processed, results are stored in a database for analytics and trend identification purposes.

The primary output of health data processing by HLFM is a maintenance plan tailored specifically to the fleet and the current status of its members **(R02, A04)**. This work plan is made possible as a result of intelligence built into the application regarding Excavation and Repair bot design and capabilities.

Once the maintenance plan is available, Repair supervisors are made aware of its availability and location for retrieval via notification **(R05).**

Control of excavation and repair scheduling is provided to human site supervisors via the HLFM web portal. In addition to scheduling tasks, human supervisors can perform administrative overrides, remove workers from service, and analyze performance of the fleet **(R02).**

Should an instruction need to be issued, a task bundle can be generated using the portal. This generates a notification to the applicable Repair or Excavation supervisor, who then retrieves the task bundle and ensures it is executed by the target worker.

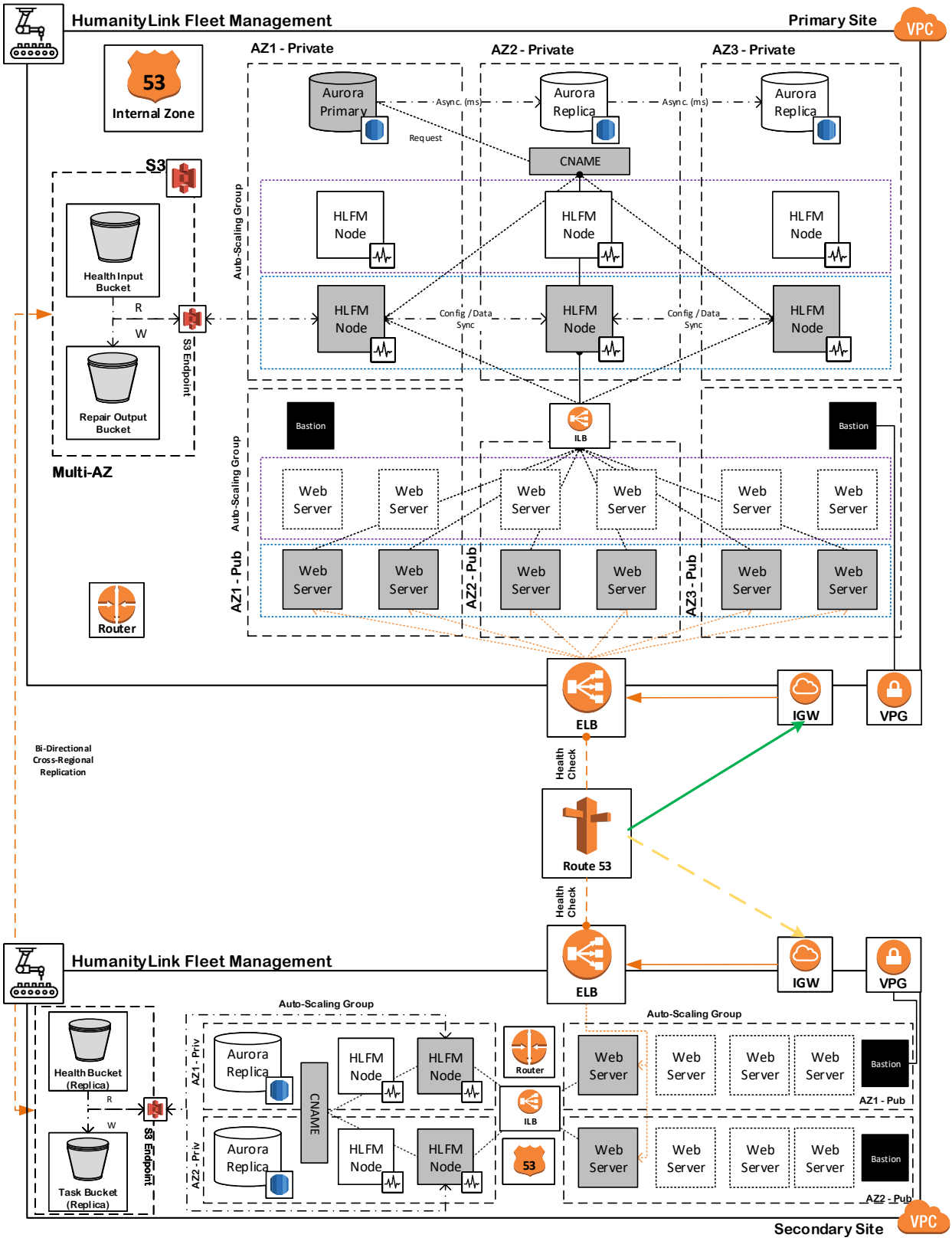
Aside from these interactions, the Excavation and Repair fleets are not reliant on the infrastructure in performing their tasks. These efforts can proceed autonomously once work plans for the shift are distributed and initiated. **(R05, R07)**

*Services Provided*

The HumanityLink Fleet Management application provides the following services:

- Monitoring of Excavation fleet performance over time
- Analysis of trends in health and efficiency
- Schedule adjustment and confirmation
- Activation and de-activation of fleet members
- Administrative overrides for individual workers
- Generation of repair plans according to health analysis
- Initiation of Excavation and Repair procedures

*Logical Diagram*




*Distribution of HLFM services across primary and disaster recovery sites*

### *Compute*

Relatively small general-purpose instances have been selected for the web tier as workload will be distributed among many servers in an auto scaling group. Compute-optimized instances have been selected for running the HLFM application, as this is expected to be a CPU-intensive workload. Database functions will be provided by a large memory-optimized, SSD-backed RDS instance. This will allow caching of frequent requests in RAM and fast reads from disk, as needed. Amazon Aurora will run on this database instance, offering several performance and scalability improvements compared to MySQL and other relational database platforms.

*Design Decisions*

| Area | Decision | Meets Requirement | Justification |
|---|---|---|---|
| Web Server Instance | M4.L | R03<br>R04<br>RP02 | General-purpose instances with a smaller amount of CPU/RAM, in conjunction with AutoScale, will meet projected web-tier requirements. |
| App Server Instance | C4.2XL | R03<br>R04<br>RP02 | HLFM performs best with higher clock speeds, so a fitting compute-optimized instance has been selected. |
| Database Type | Relational / SQL | R08 | Use of a relational DB has been selected to support analytics requirements. |
| Database Instance Type | RDS Aurora | R03<br>R04<br>R08<br>RP02 | Aurora meets performance, scalability, replication and analytics requirements. |
| Database Instance Size | DB.R3.2XL | RP02 | Emphasis has been placed on memory and disk performance to meet expected workload. |
| Bastion Instance | T2.M | RP02 | Performance requirements for administration hosts are low, so a burstable instance has been chosen. |
| Virtualization Type | HVM | RP04 | HVM AMI's provide the highest level of performance by leveraging hardware virtualization extensions. |
| Operating System | CentOS7 Linux | C03 | HLDM was developed against CentOS and although the application will run on other distributions, management desires consistency. |

*Instance Configuration*

| Quantity | Role | Type | CPU | RAM | OS | Disk |
|---|---|---|---|---|---|---|
| 6 | Web | M4 L | 2 | 8 | CentOS7 | GP2 60GB |
| 3 | App | C4.2XL | 8 | 15 | CentOS7 | GP2 100GB |
| 3 | RDS | DB.R3.2XL | 8 | 61 | Managed | GP2 1TB |
| 2 | Bastion | T2.M | 2 | 24 | CentOS7 | GP2 60GB |

*Storage*

Storage workload within the web tier is expected to be low but variable, consisting primarily of reads when serving web content.

The application tier has been designed to accommodate a mixed workload consisting of sequential S3 loads and deposits, as well as temp file usage during the work instruction creation process.

Database usage is expected to be high when health data is processed by the app tier and written to the database. Usage will be more moderate as this data is periodically referenced by the health portal. The burst capability of GP2 is being relied upon to handle this variance, providing up to 6,000 IOPS at a 1TB allocation.

If this configuration proves insufficient the storage type will be changed and an appropriate amount of dedicated IOPS will be provisioned for the database.

Database storage has been sized to accommodate long-term retention of event and health data for the Excavation and Repair fleets, which will improve identification of health trends.

HLFM uses the same S3-based approach as HLDM in communicating with HumanityLink components residing outside the cloud environment. Several S3 Standard buckets have been provisioned to store data input from the Excavation fleet and instruction output for the Repair fleet.

Files deposited into S3 will make use of multi-part uploads to eliminate excessive restarts when or if connection quality degrades. This data will be pre-compressed to reduce overall transfer time, as well.

*Design Decisions*

| Area | Decision | Meets Requirement | Justification |
|---|---|---|---|
| EBS Volume Type | GP2 SSD | R03<br>R04<br>RP01 | Meets mixed workload requirements for all HLFM tiers and provides IO burstability for periods of increased utilization. |
| EBS Volume Size (Instances) | 60GB Web<br>100GB App | RP02 | Web servers have been sized to allow OS install and a small amount of static web content. App servers have been sized to accommodate the management application. |
| EBS Volume Size (Database) | 1TB | RP02 | Historical event and health data will be retained for an extended period of time. This sizing decision will support that goal. |
| EBS Volume Scaling | Up / Out | RP01 | Performance and capacity can be optimized by increasing disk size or attaching additional disks. Disk type can also be modified to guarantee IOPS. |
| S3 Storage Type | Standard | RP06 | Supports requirement for high-performance low-latency object storage performance. |
| Data Reduction | Pre-Compression | RP07 | Data will be pre-compressed before upload to S3 to optimize transfer speed. |

### *Manageability*

#### *Administration and Maintenance*

Administrative access is provided to the technical team via IPSEC connection only from physical locations with this configuration present. Staff can perform functions from a pair of Bastion hosts accessed over this tunnel.

Ongoing maintenance of system images will be completed by the Operations team, and a golden, updated AMI will be maintained for the Web and Application tiers. System updates will be facilitated primarily via Configuration Management, but updates will continue to the AMI in order to minimize post-deployment configuration time.

#### *Monitoring*

AWS CloudWatch will be used to provide insight into the health of the Fleet Management application and associated resources. Default alarms and thresholds will be used until load is fully understood, after which tuning will take place to improve alert reliability.

#### *Configuration Management*

As with HLDM, HLFM will utilize Amazon OpsWorks for configuration management functions. Cookbooks will be maintained for each server role by the operations team, detailing required OS configuration.

#### *Notifications and Scheduling*

Amazon SNS notifications will be used to inform HLFM that health data has been uploaded by Excavation supervisors. SNS will also be used to make the Repair fleet aware of the presence of a new repair instruction package.

#### *Design Decisions*

| Area | Decision | Meets Requirement | Justification |
|---|---|---|---|
| Administrative Access | SSH to Bastion Host over IPSEC | RM01<br>RS01 | IPSEC VPN eliminates exposure of administrative ports to the public internet. Use of Bastion hosts in this configuration simplifies security policy management. |
| Ongoing Maintenance | OpsWorks + AMI Patching | RM02 | Use of these tools will ensure the environment remains up to date and without configuration drift. |
| Monitoring Platform | Cloudwatch | RM03 | CloudWatch is an AWS-native service and can provide resource metrics others cannot. |
| Event Notifications | Simple Notification Service | R04 | SNS will be used to notify subscribers of new object availability in S3. |

### *Availability*

In support of this project's primary availability requirement, compute and storage resources are distributed among three sites in the production environment. Multiple replicas of the database instance also exist, protecting the availability of that component. In the event of server instance failure, load balancers will route requests to surviving instances. Should the primary database instance fail, a secondary replica will be activated, and the referenced DNS name will be cut over to the new instance. VPN and DNS are also provided by highly-resilient services, resulting in a very fault-tolerant architecture.

*Design Decisions*

| Area | Decision | Meets Requirement | Justification |
|---|---|---|---|
| Cloud Platform | Amazon Web Services | R01<br>C01<br>C02<br>r01 | Provides global presence, robust availability and minimizes operational overhead. Use of multiple independent regions minimizes single provider risk. |
| Server Availability | Instance Distribution across AZ's | RA01<br>RA02<br>RA03 | This configuration distributes Web and App instances across multiple sites, protecting against service failure when an instance fails. |
| Database Availability | 4 RDS Replicas Asynchronous | RA01<br>RA02<br>RA03 | Two replicas of the RDS instance will reside in both the Primary and DR environments, meeting availability requirements. |
| Web Service Availability | External Elastic Load Balancer | RA01<br>RA02<br>RA03 | Use of an external load balancer provides continuous availability of the web tier and health dashboard in the event of web instance failure. |
| HLFM Service Availability | Internal Elastic Load Balancer | R03<br>RA02<br>RA04 | An internal load balancer distributes requests across multiple HLFM servers and also provides high-availability should an HLFM instance fail. |
| Instance Storage Availability | EBS Replication | R02<br>RA01 | Use of EBS volumes ensures three copies of data exists within the region, offering continued availability during HW failure. |
| Object Storage Availability | S3 Standard Replication | R02<br>RA01 | S3 standard replicates each object three times within a region, providing protection against hardware failure. Cross region replication is also used, protecting against entire region failure. |
| VPN Availability | VPG Native HA | RA01<br>RA03 | VPG's possess multiple connection addresses, ensuring connectivity to the VPC can be re-established if a VPG component fails. |
| DNS Service Availability | Route 53 Hosted DNS | RA03<br>RA04 | R53 will be used for all internal/external DNS due to its resilient characteristics |
| SNS Availability | Email + SMS | RA03<br>r03 | Notifications must be sent via multiple paths to ensure delivery |

*Scalability*

To cope with unknown and unpredictable demand, this design makes use of auto-scaling groups in both the web and HLFM application tiers. As resource utilization reaches a defined threshold and remains in excess of this value for a period of time, additional instances will be created automatically.

These instances will then retrieve appropriate configuration for their role and become active in servicing requests. Internal and external Elastic Load Balancers will be employed to handle this transition gracefully. With proper monitoring and tuning of thresholds over time, this process should not require additional management. Scaling of S3 object storage will take place automatically as the number of operations increases.

Scaling of the database tier will consist primarily of monitoring workload and adjusting instance type or allocated resources, as required. With a write-biased application, as HLFM is expected to be, addition of read replicas will not significantly improve scalability of this tier. However, these will be present as a function of availability improvement and can be utilized for scalability purposes, as well.

*Design Decisions*

| Area | Decision | Meets Requirement | Justification |
|---|---|---|---|
| Web Tier Scale-Out | EC2 Auto-Scaling | R04<br>RP01<br>r03 | Use of AutoScaling in the web tier will accommodate fluctuations in usage of the Web portal. |
| App Tier Scale-Out | EC2 Auto-Scaling | R04<br>RP01<br>r03 | Processing load created by analysis of multiple fleet's health data and generation of work plans is unknown. Auto-scaling will allow HLFM to adapt as these needs change. |
| Database Tier Scale-Up | Manual Process | RP01<br>RP02<br>r03 | Increasing Compute and Storage resources is the recommended method of scaling the HLFM database, as workload is expected to be write-biased. |
| Database Tier Scale-Out | RDS Replicas | RP01 | A number of replicas can be created of the primary RDS instance. During normal operation, these can only service reads, which will not improve performance of HLFM. |
| EBS Scale-Out | Additional Disks | RP02 | Scaling out of EBS can be accomplished by adding up to 40 volumes to a Linux-based instance. |
| EBS Scale-Up | Size Increase or IO1 Conversion | RP01<br>RP02<br>r03 | With a ratio of 3 IO/GB, performance of GP2 can be increased by increasing volume size. Conversion to IO1 will also accomplish this goal with up to 50 IO/GB |
| S3 Performance Scaling | Native load distribution | R01<br>R02<br>R03<br>RP01 | S3 Standard is designed to increase in performance as quantity of parallel operations increases. |

*Auto-Scaling Rules*

1) **Web_Node_AutoScale**
    - Parameters - GroupMinSize 6
    - GroupMaxSize 12
    - GroupDesiredCapacity 6
    - Launch Config
        i. HLFM Web AMI
        ii. Instance M4.Large
        iii. Auto-IP
    - Security Groups
        i. ELBtoWeb
        ii. WebToILB
    - Scale Up Plan
        i. CPU Utilization >90% for 5min, add 3 instances
    - Scale Down Plan
        i. CPU Utilization <40% for 60min subtract 3 instances

3) **HLFM_Node_AutoScale**
    - Parameters - GroupMinSize 3
    - GroupMaxSize 6
    - GroupDesiredCapacity 3
    - Launch Config
        i. HLFM AMI
        ii. Instance C4.2XL
        iii. Auto-IP
    - Security Groups
        i. ILBtoHLFM
        ii. HLFMtoS3
        iii. HLFMtoAurora
    - Scale Up Plan
        i. CPU Utilization >90% for 15min, add 1 instance
    - Scale Down Plan
        i. CPU Utilization <40% for 60min subtract 1 instances

### *Recoverability*

In order to protect EC2 instances in the web and application tier, once-daily EBS snapshots will take place mediated by an external automation product, CloudRanger. These events will be triggered automatically according to a defined schedule, and snapshot data will be transmitted and stored in S3.

This will allow replacement volumes to be generated for impacted web or HLFM servers and speed recovery of those elements. Both the web and HLFM servers have been sized to allow full recovery of an impacted volume within the defined RTO of one hour.

Because EC2 snapshots are only taking place once per-day, additional measures are needed within the Application tier to meet the defined application data RPO of one hour. To meet this requirement, HLFM nodes will back up configuration files to existing S3 buckets on an hourly basis. Upon recovery of an instance, HLFM can invoke recovery of its configuration files, meeting the recovery objective.

At the database level, automatic full-instance daily backups are performed, providing a baseline of protection. Layered on top of this is the use of transaction log backups, which allows for point-in-time recovery of a database instance to just minutes before the failure.

In total, these measures result in a solution that meets the defined RPO and RTO of 1 hour for the HLFM application.

*Design Decisions*

| Area | Decision | Meets Requirement | Justification |
|---|---|---|---|
| EC2 Backup Method | EBS Snapshot | RR01 | Web and App disks can be recovered using EBS snapshots within the defined RTO of 1 hour. (60GB / 100GB) |
| EC2 Backup Product | CloudRanger | RR01 | Facilitates automated EBS snapshots without requiring deployment of additional components. |
| EC2 Backup Frequency | Once daily | RR01 | This schedule in combination with hourly configuration file backups meets the RPO |
| EC2 Backup Retention | 7 days | RR01 | S3 and database backups contain data essential to the application. As a result, less EBS snapshot retention is required. |
| S3 Backup Retention | 30 days | RR09 | Requirements dictate that application data backups be retained for a month. This configuration meets the requirement. |
| RDS Backup Method | Instance Snapshot + Transaction Logging | RR02<br>RR09 | Utilizing RDS native backup capabilities, the defined RPO/RTO can be met. |
| RDS Backup Frequency | Daily (Snapshot)<br>Continuous (Log) | RR02<br>RR09 | This configuration will result in the achievement of restoration objectives. |
| RDS Backup Retention | 30 days | RR09 | This decision aligns with application data retention requirements. |

### *Disaster Recovery*

In support of defined recoverability objectives, a scaled-down but fully-functional replica of the production HLFM environment has been created in a separate AWS region distant from the primary. This environment has been designed to scale to at least 50% of primary capacity and support all functions of the application.

Because this environment must become live within one hour of disaster declaration, servers and services are kept continuously online. Initially prepared to service requests are two web servers and two HLFM application servers, with both tiers configured to scale quickly as workload increases. Web server configuration updates are pushed via configuration management, and application server configuration can quickly be synchronized against replicated data in S3.

At the database level, two replicas of the primary Aurora RDS instance are present, allowing for redundancy to be preserved in the case one of them is made primary in the DR site.

Data stored in Health input and Repair output S3 buckets is configured to replicate cross region in bi-directional fashion. This allows all current data to be accessible in DR when operations resume, as well as ensures dataset consistency when operations fall back to the primary site.

Failover to the disaster recovery environment is mediated by Route 53-hosted DNS services and SNS notifications. Route 53 health checks integrated with Elastic Load Balancers at both locations will detect a regional failure and direct users of the HLFM web portal to the healthy destination.

SNS subscription to object creation events in DR buckets will make Repair supervisors aware of new work instruction packages following failover. Excavation supervisors will be informed of the new upload location via SNS message, as well.

*Design Decisions*

| Area | Decision | Meets Requirement | Justification |
|---|---|---|---|
| DR Approach | Fully-functional Standby environment | RR04 | All components required to operate HLFM will be online and available for failover at all times. |
| DR Activation | DNS Failover + SNS Notifications | RR10 | Allows for automatic failover to DR within the defined RTO. |
| DR Data Availability | S3 Cross-region Replication + RDS Replication | RR04 RR06 | S3 and database content will continuously be replicated to the DR environment. |
| DR Data Consistency | S3 Bi-Directional Replication RDS Replication | RR06 | Bi-directional S3 replication will support failover and failback. Replication will also take place between the DR RDS primary and a standby instance in the impacted region, once available. |

*Security*

Security within the HLFM environment is approached using several AWS features, including security groups, IAM roles and users, and bucket permissions.

Security groups have been defined to allow only required communications between the various layers of the application stack. External ELB's can communicate with the web tier, the web tier is restricted to accessing the internal ELB, and the HLFM tier accepts communication only from the internal ELB. RDS instances communicate only with the HLFM tier. In this manner, only appropriate network communication is allowed. S3 access is only permitted between HLFM nodes and the S3 endpoint.

S3 buckets are protected using the same restricted-access approach. HLFM nodes are launched into an IAM role, and this IAM role is granted only needed permissions to the Health input and Repair output buckets. Similarly, IAM users are created for components residing outside the VPC, and read/write access is granted only where required.

Administrative access via SSH is allowed only over IPSEC connection to an attached Virtual Private Gateway. Admins can use this method to connect to a Bastion host and conduct their duties from that point. In this way, the bastion host is protected from internet attacks, and individual instances are only exposed to SSH from the bastion.

Data encryption extensions built into the HumanityLink Suite ensure encryption of data before upload to S3 and before writing to the database. As a result, AWS services are not needed for this task.

Using this layered approach, the HLFM environment is protected from internal attacks, external attacks and data theft, which ultimately supports the security and availability goals of the business.

*Design Decisions*

| Area | Decision | Meets Requirement | Justification |
|---|---|---|---|
| Network Access Control | Security Groups | RS02 | Security groups control traffic between instances, endpoints and gateways and restrict access, where required. |
| S3 VPC Access Control | IAM Roles + Bucket Policies | RS01 RS03 | All instances have been launched into an IAM role, and this role has been given only the permissions required to conduct normal operations. |
| S3 Outside-VPC Access | IAM Users + Bucket Policies | RS01 RS03 | Components residing outside of the VPC environment are configured to use access keys provisioned to specific IAM User accounts. Only required permissions are present. |
| Data Encryption | Application-level Encryption | RS06 | Intelligence built into the HumanityLink Suite encrypts data before depositing to S3 or writing to the RDS database. |
| Web Firewall | NAXSI | RS07 | This NGINX module will supplement security groups and assist with web server protection against exploitation. |

*Roles and Users*

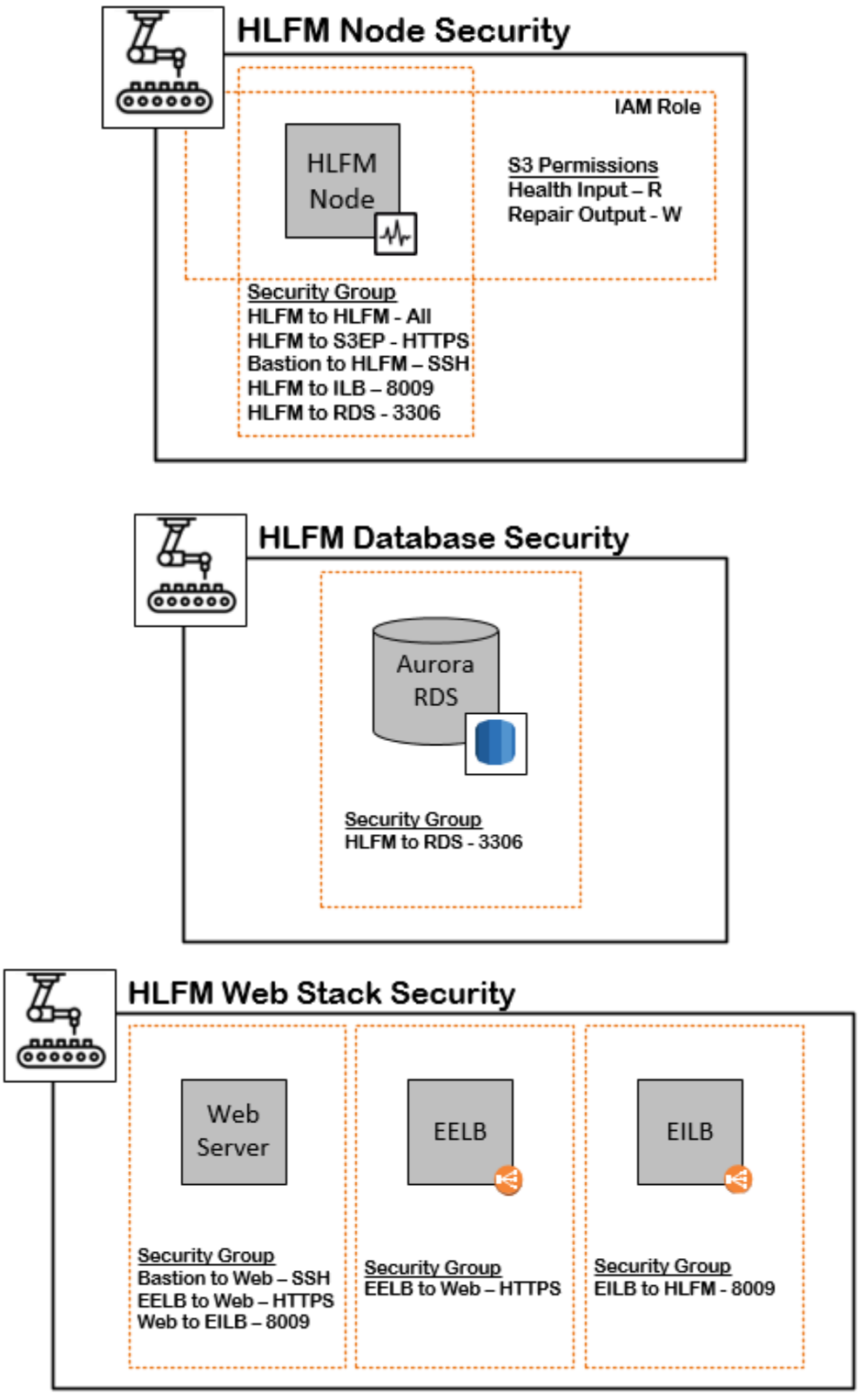
1) **Role – HLFM Node**
    - Health Input Bucket – Read
    - Repair Output Bucket – Write

2) **User – Excavation Supervisor**
    - Group – S3HealthUploaders
    - Write - Create object in Health Input bucket

3) **User – Repair Supervisor**
    - Group – S3RepairDownloaders
    - Read - Retrieve object from Repair Output bucket

*Security Groups*

1) **VGW to Bastion**
    a. SSH allow
2) **Bastion to All**
    a. SSH allow
3) **HLFM to HLFM**
    **a.** Allow all
4) **HLFM to S3EP**
    **a.** HTTPS allow
5) **ELB to Web**
    a. HTTPS allow
6) **Web to ILB**
    a. TCP 8009 Allow
7) **ILB to HLFM**
    a. TCP 8009 allow
8) **HLFM to RDS**
    a. TCP 3306 allow

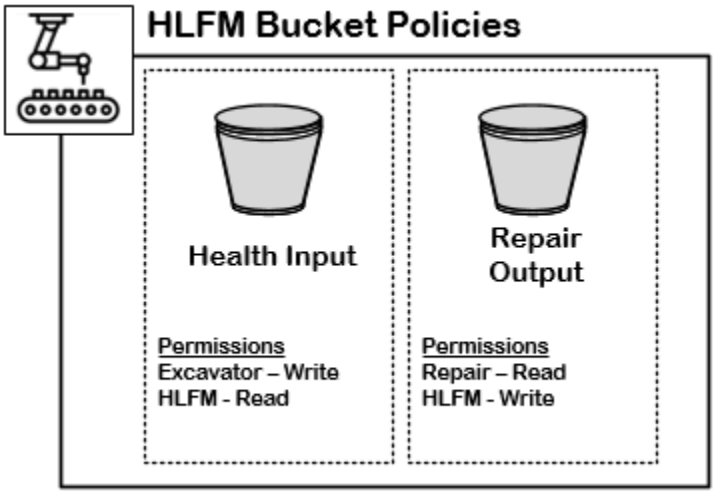



*Configuration of Security Groups associated with Fleet Management components.*

*S3 Permissions*

4) **Health Input Bucket**
    - Write access for Excavation Supervisors uploading fleet health data
    - Read access for HumanityLink Fleet Management nodes preparing for analysis
    - No delete access granted

5) **Repair Output Bucket**
    - Write access for HumanityLink Fleet Management work file uploads
    - Read access for Repair Supervisors retrieving work packages
    - No delete access granted




*Configuration of S3 Bucket Policies used by Fleet Management*

## *Physical Configuration*

### *Production Networks*

| Public Network | Network | Usable IP's |
|---|---|---|
| pub us-east-1a | 172.32.0.0/20 | 4091 |
| pub us-east-1b | 172.32.16.0/20 | 4091 |
| pub us-east-1c | 172.32.32.0/20 | 4091 |

| Private Network | Network | Usable IP's |
|---|---|---|
| priv us-east-1a | 172.32.48.0/20 | 4091 |
| priv us-east-1b | 172.32.64.0/20 | 4091 |
| priv us-east-1c | 172.32.80.0/20 | 4091 |

### *Production Instances*

| Name | IP | Type | CPU | Memory | Disk 1 | OS | Apps |
|---|---|---|---|---|---|---|---|
| BAS01 | 172.32.0.252 | T2.M | 2 | 24 | 60 | CentOS7 | Lynx, SSH |
| BAS02 | 172.32.32.252 | T2.M | 2 | 24 | 60 | CentOS7 | Lynx, SSH |
| WEB01 | 172.32.0.5 | M4.L | 2 | 8 | 60 | CentOS7 | NGINX 1.13 |
| WEB02 | 172.32.16.5 | M4.L | 2 | 8 | 60 | CentOS7 | NGINX 1.13 |
| WEB03 | 172.32.32.5 | M4.L | 2 | 8 | 60 | CentOS7 | NGINX 1.13 |
| WEB04 | 172.32.0.6 | M4.L | 2 | 8 | 60 | CentOS7 | NGINX 1.13 |
| WEB05 | 172.32.16.6 | M4.L | 2 | 8 | 60 | CentOS7 | NGINX 1.13 |
| WEB06 | 172.32.32.6 | M4.L | 2 | 8 | 60 | CentOS7 | NGINX 1.13 |
| HLFM01 | 172.32.48.5 | C4.2XL | 8 | 15 | 100 | CentOS7 | HLFM 1.0 |
| HLFM02 | 172.32.64.5 | C4.2XL | 8 | 15 | 100 | CentOS7 | HLFM 1.0 |
| HLFM03 | 172.32.80.5 | C4.2XL | 8 | 15 | 100 | CentOS7 | HLFM 1.0 |
| AuroraPri | Auto | DB.R3.2XL | 8 | 61 | 1000 | Managed | Aurora RDS |
| AuroraRepl1 | Auto | DB.R3.2XL | 8 | 61 | 1000 | Managed | Aurora RDS |
| AuroraRepl2 | Auto | DB.R3.2XL | 8 | 61 | 1000 | Managed | Aurora RDS |

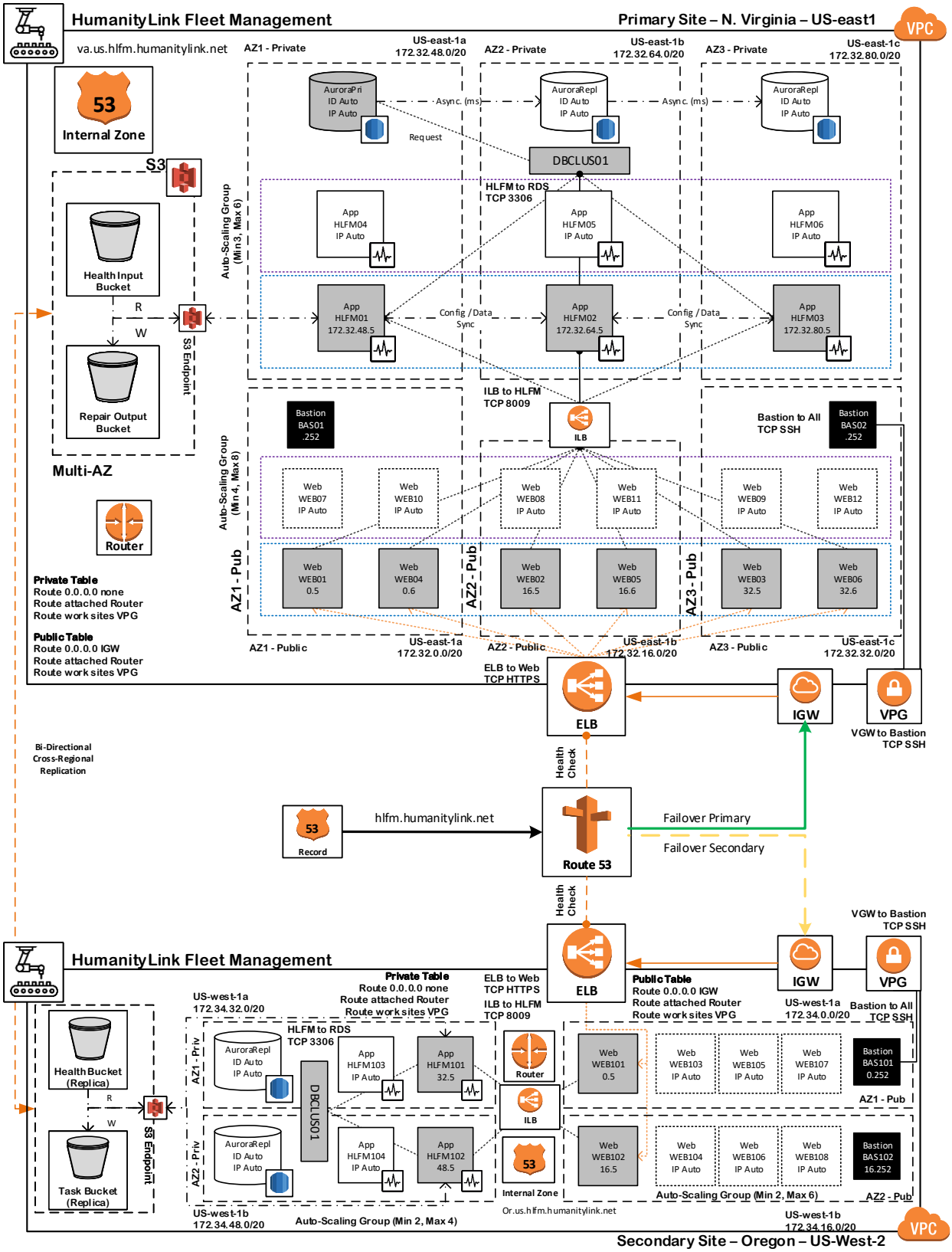
*Disaster Recovery Networks*

| Name | Network | Usable IP's |
|---|---|---|
| pub us-west-2a | 172.34.0.0/20 | 4091 |
| pub us-west-2b | 172.34.16.0/20 | 4091 |

| Private Network | Network | Usable IP's |
|---|---|---|
| priv us-west-2a | 172.34.32.0/20 | 4091 |
| priv us-west-2b | 172.34.48.0/20 | 4091 |

*Disaster Recovery Instances*

| Name | IP | Type | CPU | Memory | Disk 1 | OS | Apps |
|---|---|---|---|---|---|---|---|
| BAS101 | 172.34.0.252 | T2.M | 2 | 24 | 60 | CentOS7 | Lynx, SSH |
| BAS102 | 172.34.16.252 | T2.M | 2 | 24 | 60 | CentOS7 | Lynx, SSH |
| WEB101 | 172.34.0.5 | M4.L | 2 | 8 | 60 | CentOS7 | NGINX 1.13 |
| WEB102 | 172.34.16.5 | M4.L | 2 | 8 | 60 | CentOS7 | NGINX 1.13 |
| HLFM101 | 172.34.32.5 | C4.2XL | 8 | 15 | 100 | CentOS7 | HLFM 1.0 |
| HLFM102 | 172.34.48.5 | C4.2XL | 8 | 15 | 100 | CentOS7 | HLFM 1.0 |
| AuroraRepl3 | Auto | DB.R3.2XL | 8 | 61 | 1000 | Managed | Aurora RDS |
| AuroraRepl4 | Auto | DB.R3.2XL | 8 | 61 | 1000 | Managed | Aurora RDS |




*Physical configuration of Humanity Link Fleet Management primary and DR sites*

# Appendix

The documentation below was referenced in the creation of this solution design:

- AWS VPC Documentation
- AWS EC2 Documentation
- AWS EBS Documentation
- AWS ELB Documentation
- AWS S3 Documentation
- AWS RDS Documentation
- AWS Route 53 Documentation