Resilient Infrastructure for Media Operations

Geoff Stedman, SDVI CMO
Version 2, Revised March 2024

Introduction

Media companies require secure and robust infrastructure that is resilient even when failures or disasters hit. The requirement for robust, redundant infrastructure grew from broadcast operations that wanted better than 5 nines of availability (99.999% uptime). While their media supply chain infrastructure is perhaps not quite as sensitive as on-air broadcast infrastructure, the demands for content across multiple delivery platforms has elevated the need for resilient supply chains as well.

A common technique used by media companies to achieve their availability goals is to maintain redundancy…redundant infrastructure deployed across redundant locations. Many built a primary facility with main and backup systems, along with a second facility geographically separated with yet another backup system. This level of redundancy did work to maintain availability when something failed or a disaster struck, but it is fundamentally expensive, hard to manage, and inhibits the agility these companies need to respond to market changes.

The cloud, with its natural redundancy and geo-diversity, gives media companies another option to deploy reliable and highly-available media supply chain infrastructure. In fact, cloud infrastructure can be as, or more, reliable and available than what media companies previously built themselves – without the cost and hassle of building and maintaining data centers, and without the waste. Cloud services can now provide the same level of business continuity at a fraction of the cost of previous continuity strategies.

In this paper, we will review how the SDVI Rally media supply chain management platform takes advantage of cloud resiliency features to ensure that media supply chains continue to function in the event of a disruption.

The cloud-native SDVI Rally media supply chain platform enables customers to migrate their media supply chains to the cloud with built-in business continuity, ensuring your supply chains will continue to run, even during an emergency. Rally can maintain duplicate redundant instances of all core services across availability zones in order to avoid any interruption of service during a failure. This well-designed resilience strategy means there will be very little impact to your business should an emergency occur.

Single vs. Multi Availability Zones

Before we get into the Rally architecture, it will help to first understand how the AWS cloud is set up to support resiliency. The AWS cloud is first organized into Regions. For example, the United States is divided into seven geographic regions, such as Northern Virginia, Northern California, Ohio, and Oregon to name a few. A Region is made up of a number of isolated and physically separate Availability Zones (AZ's). Each AZ has one or more discrete data centers with redundant power, networking, and connectivity, which gives customers the ability to operate production applications that are more highly available, fault tolerant, and scalable than would be possible from a single data center. All AZs in an AWS Region are interconnected with high-bandwidth, low-latency networking,

over fully redundant, dedicated metro fiber, and all traffic between AZs is encrypted. Because network performance is sufficient to accomplish synchronous replication between AZs, it makes partitioning applications for high availability easy, which better protects those applications from disruptions caused by power failures, lightning strikes, tornadoes, earthquakes, and more.

When we first configure Rally for each customer, we provide a self-contained Rally instance with all the services that support it. We call these a Rally Silo. Customers have at least two Silos at their disposal; a production environment, and a staging/non-production environment. These will be configured either as a single or multiple Availability Zone installations. Typically, most large media companies will deploy their production system in a multi-Availability Zone configuration.

Depending on your organization's needs, you may want more silos for DEV or QA groups. Non-production systems like these generally do not have the same SLA demands as the production or staging environments, so they can be configured for a single AZ.

Figure 1) AWS Global Infrastructure Map

Practical Considerations
In a single AZ deployment, if there is a disruption to that AZ, jobs running in Rally may be interrupted for as long as the AZ remains impacted. Once the AZ is back online, Rally will restart any workflow which had not completed. There could be a small delay as work needs to be restarted once the cloud provider resolved any issue and as always, customers only incur charges for completed jobs.

In a multi-AZ deployment, instances of the Rally services are running and actively sharing load across multiple Availability Zones. If there is a disruption to one AZ, Rally will immediately re-route the workorder to another AZ. Customers would see no impact because Rally actively switches work over to another available AZ. There are no cross-region charges because both AZs are within the same Region.
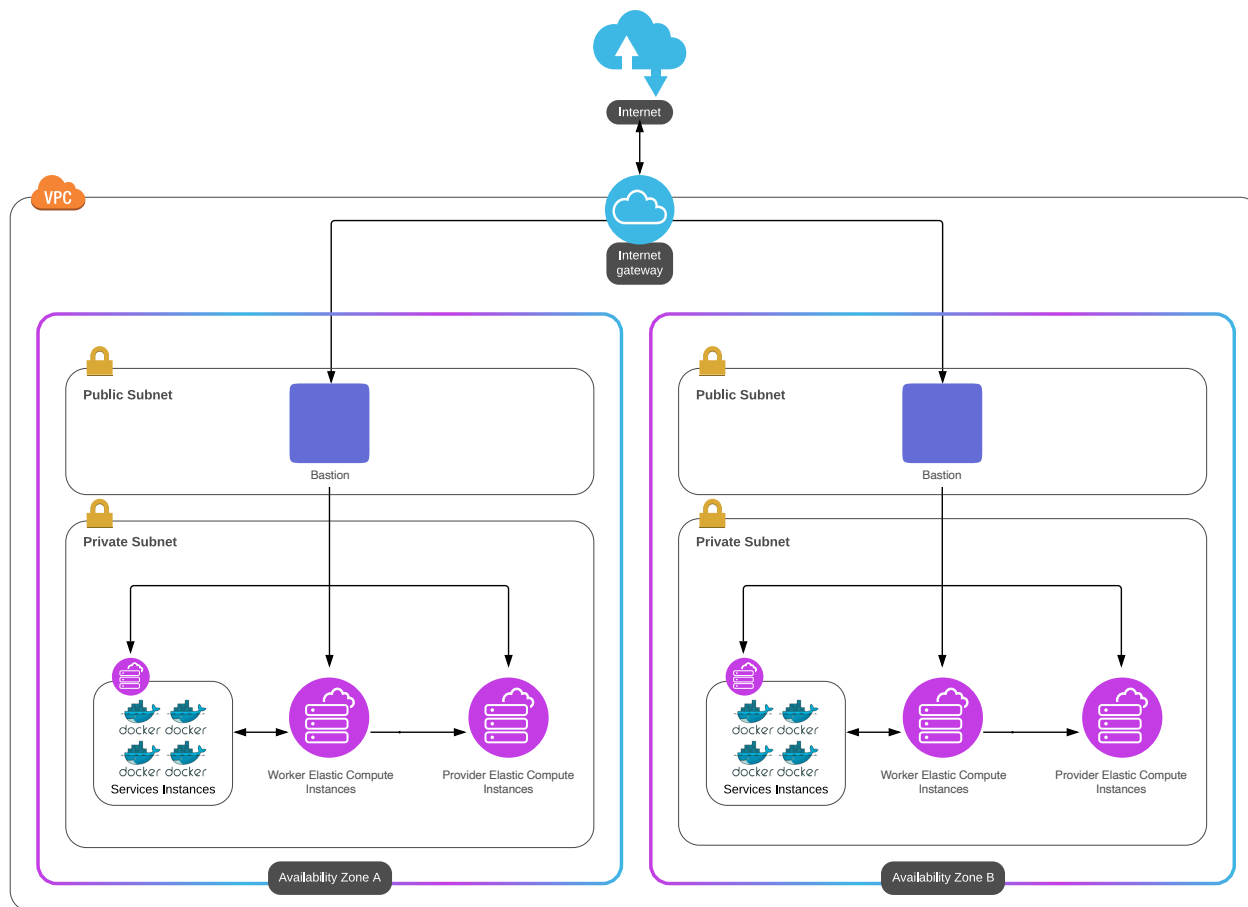
Figure 2) SDVI Rally AWS Silo Architecture

Multi Region

In addition to the multi-AZ approach discussed above, there may be reasons to utilize multiple AWS Regions for Rally deployments. Customers have the ability to locate Rally-managed storage locations and provider pools across different AWS Regions (or different Cloud providers) with the goal of providing both resilience and the ability to process jobs in the Region where the content is located. For global media companies receiving and distributing content in multiple regions of the world, this approach keeps supply chain processing in the same Region where the content resides (minimizing inter-Region data transfers). In the unlikely event that a Region goes down, all supply chain processing could be reassigned to the Region that remains operational.

It is also possible to provision multiple discrete Rally silos in different Regions, each providing active capabilities, but separate from each other. Users would need to log into each system separately, and the two systems would not be centrally managed, although they could move content and data between them (albeit incurring inter-Region data egress costs).

Finally, using the new feature of Rally Multi-Region Sync you can have two (or more) silos in two (or more) different regions that sync bidirectionally. Users can search and view information for assets in all silos, even if the remote silos have outages. This is accomplished through data replication, where the search indices are continuously copied between all silos. Capabilities gained by going this direction include but are not limited to, searching for remote assets by name and metadata; viewing remote asset name, status indicators, and metadata; getting a hyperlink that goes directly to the asset page on the remote silo; viewing a list of remote silos and their sync statuses for administration purposes. All of this is achieved through relatively simple configurations between the different silos. It also means silos in each region can be actively used with the content in that region, and syncronize their data with each other.

Reliability of Rally

The Rally architecture has been through multiple AWS reviews and certifications, including the AWS Foundational Technical Review and Well-Architected Review processes, and is designed so that there is no single point of failure. In addition to the Availability Zone configurations to enhance reliability and availability discussed above, each Rally silo is backed up nightly, ensuring that all supply chains and provider presets can be restored in the event of a critical outage.

Each of the three different disaster recovery options includes trade-offs between availability and cost, depending on which level of service is provisioned for a Rally deployment.

Scenario one: Single Availability Zone – provides the lowest cost, and a more restrictive SLA. If the AZ were to go offline, all jobs would be automatically rescheduled after the zone came back online. Because of the nature of cloud services, were this to happen, once catch-up starts on any jobs which were not completed, they will be executed on as many EC2 instances as needed. This is the most affordable option.

Scenario two: Multi-Availability Zone – provides distributed storage and providers across zones. This approach yields better resiliency and scalability, and is often used for large volume processing where the SLA has contractual delivery requirements. Multi-AZ provides a shared, load-balanced approach for a Rally system across two or more Availability Zones. If one AZ goes offline, Rally will automatically allocate jobs to the AZ(s) that remain online.

Scenario three: Multiple Region Availability – provides resources running on separate Rally silos across Regions. Discreet silos can be provisioned in different Regions, each providing duplicate, active capabilities. The different silos would synchronize bidirectionally with each other, and would be available to take on workloads when notified if an entire region had failed. This scenario would require a full duplication of the scheduled active workload. This option is the most expensive due to potential inter-region data transfers required to keep content in both regions.

Conclusion

Rally is architected to provide 24/7 availability with automated backups and options for redundancy. There's no need to operate another physical backup environment which requires service contracts, perhaps third-party management, and real estate costs. Rally offers near-infinite cloud scalability for media supply chain operations, which also extends to business continuity using a cloud-native architecture.

You can now make choices about what is appropriate based on SLA requirements, budgets and how critical your media content processing factory is to your business. By gaining the technical and business agility that comes with a solid business continuity strategy, there will never be a danger of having a catastrophe take down supply chain operations.