

cloudticity

Designing Your AWS Disaster Recovery System for Epic EHR

Table of Contents

I. DISASTER RECOVERY CONCEPTS 3

The EHR Must Go On! 3

The Many Flavors of Disaster 4

- Component Failures 5
- Human Errors 6
- Malicious Activity 6

Disaster Recovery Objectives 9

- Recovery Time Objective (RTO) 9
- Recovery Point Objective (RPO) 10

DR Options on AWS 12

- Backup and Restore: Slow, Cheap, and Reliable 12
- Pilot Light: Fast Recovery 13
- Warm Standby: Ready to Go 13
- Multi-Site Active-Active: Real-time Recovery 14

AWS Regions and Availability Zones 15

Shared Responsibility Model for Resilience 16

AWS Services for DR 17

Epic Data Architecture 17

Figure 11. Epic Data Architecture 18

Shared Responsibility Model for Resilience 18

II. DESIGNING YOUR EPIC DR SYSTEM 19

Step 1. Engage Your Third-Party Resources 19

Step 2. Establish Your RPO 20

Step 3. Establish Your RTO 22

Step 4. Map Out Your DR Architecture 23

Step 5. Choose a Recovery Architecture 24

Figure 14. Epic Pilot Light DR Implementation on AWS 24

Don't Try This at Home 25

About Sapphire Health 26

About Cloudticity 26

I. DISASTER RECOVERY CONCEPTS

This section provides an overview of disaster recovery on Amazon Web Services (AWS) and presents the key concepts that you must understand to navigate the challenges of planning and implementing an AWS-based disaster recovery (DR) system. You can access additional information on these topics using the links embedded in the copy.

The EHR Must Go On!

Since their introduction in the 1960s, electronic health records (EHRs) have transformed hospital operations. The leading vendor of EHRs is Epic, which maintains a 31 percent share of the hospital EHR market and accounts for 42 percent of hospital beds. For the hospitals and clinics that use Epic, reliable access to the Epic EHR is an essential element in care delivery (see Figure 1).

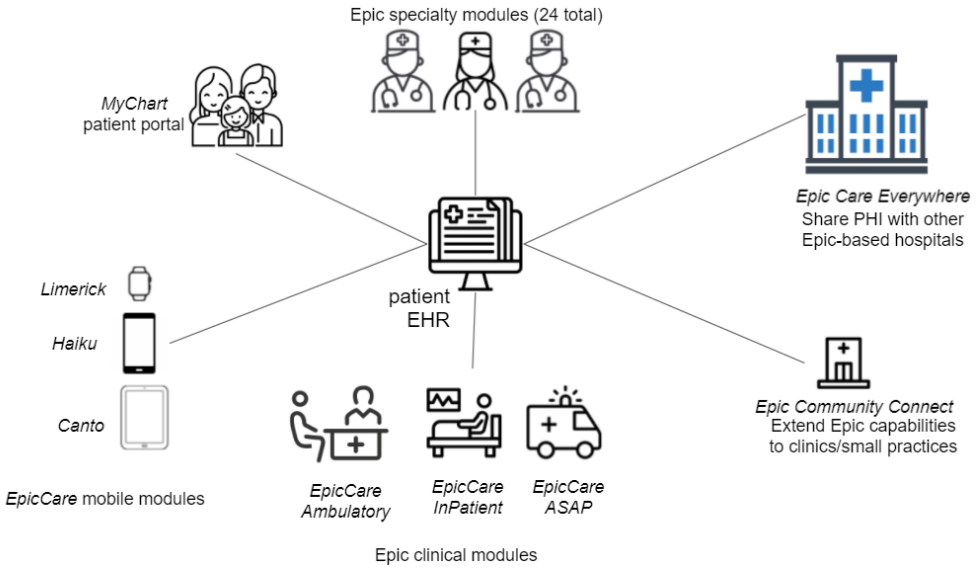


Figure 1. Typical Epic EHR Deployment

This reliance on Epic puts a great deal of pressure on hospital IT executives to prevent unplanned downtime – the Epic software must be up and running all the time. While "all the time" is certainly aspirational, the mandate for hospital CIOs is to do everything they can to make the Epic environment as resilient as possible to ensure operational continuity.

There are two aspects to resilience: high availability and disaster recovery (see Figure 2). For clarity, this discussion focuses primarily on Epic disaster recovery. However, the design of a DR system is closely tied to the requirements for high availability, so in reality these two topics are tightly interwoven.

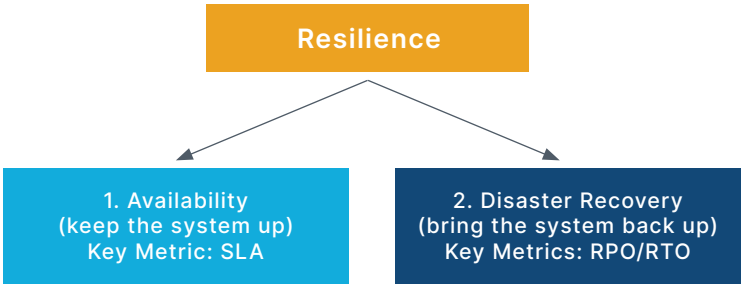


Figure 2: The Two Components Of Resilience

To set the context for a discussion of disaster recovery, let's start by examining some different kinds of disasters.

The Many Flavors of Disaster

The prospect of a significant interruption to hospital operations may seem remote, but in fact it's closer at hand than you think. In a recent survey, three-quarters of data center operations managers said they have experienced an IT service outage in the past three years. Three in ten said they had experienced an outage that had "significant impact." Severe outages are less common (one in six reported having had one in the past three years) but can have catastrophic results on hospital operations, staff productivity, and even clinical outcomes. Here are some of the kinds of disasters that hospitals must consider.

Weather Events

The most dramatic incidents that can affect a physical data center are weather events such as hurricanes, earthquakes, typhoons, and tornados. The damage mechanism can be structural failure to the data center facility due to the kinetic impact of wind and wave, or flooding caused by gradual water accumulation hours or days after the precipitating event. Events that breach the actual physical walls and physically damage equipment are extremely rare – in most cases, the failure is due to disruption of essential utilities such as power, water, cooling, or connectivity.

While the risks of catastrophic natural disasters may capture the imagination, they are in fact the easiest to mitigate. Simply deploying a disaster recovery system in a remote geographic location usually is sufficient to ensure the physical survivability of equipment, software, and digital information in the face of weather-related disasters.

Component Failures

Data centers are full of hardware and software components, and components inevitably and eventually fail. Redundant designs can help mitigate the risks of a component failure, but duplicating every piece of equipment is prohibitively expensive, so single points of failure exist and can cause outages. The impact of a component failure can be localized-for example, just affecting pharmacy or medical imaging, or they can take out power for the entire data center, effectively crippling all hospital operations that depend on the IT infrastructure.

Take the recent case of Guy's and St. Thomas Foundation Trust, one of the UK's largest healthcare organizations. An extreme heat wave caused a failure in an air conditioning system which in turn caused some critical servers to go offline. As a result, clinical staff lost access to the organization's EHR system, dramatically impacting their ability to deliver care. The ongoing outage meant doctors and nurses could not see patients' medical notes remotely and had to write down the results of consultations by hand. Access to the results of diagnostic tests, including CT and MRI scans, was also down, leaving clinicians having to phone for results. Ironically, the trust was migrating their legacy EHR system to Epic, with a go-live date less than a year away.

Human Errors

While weather events and component failures remain significant causes of downtime, networking and configuration issues are emerging as two of the more common causes of service degradation. A recent study revealed that 42% of respondents had experienced an outage due to human error in the last three years. Among those, 57% cited data center staff execution and 44% incorrect staff processes and procedures as root causes.

Where humans are involved, there are bound to be mistakes – and those mistakes can have catastrophic effects. It seems unthinkable to contemplate, but in October 2021, a single erroneous command by a Facebook IT engineer took down the entire Meta family of social media applications – Facebook, Instagram, Messenger, and WhatsApp. Those applications combined reach one in three people on earth, meaning that almost three billion people – that's a billion with a "B" – lost access to their accounts for more than six hours. The identity of the person was never revealed, but it's safe to speculate that their following performance review was less than stellar.

Malicious Activity

The previous three kinds of disaster can be characterized as accidental, but there's nothing accidental about the fourth kind – malicious activity. This category is somewhat broad: industrial espionage, hacktivism, or state-sponsored cyberattacks, but ransomware is today's most prevalent form of malicious activity. Once the province of the solo hacker, ransomware has become big business. Witness the emergence of ransomware as a service (RaaS), a dark-web business in which RaaS operators sell a wide array of easy-to-use tools and services – even customer support portals for victims! – that makes it easy for virtually anyone to launch a devastating ransomware attack.

The anatomy of a ransomware attack is familiar to every IT professional (see figure 3). Attackers gain access to the network by defeating security measures, often through simple social engineering exploits such as phishing emails or vulnerabilities in commonly used utilities such as Remote Desktop Protocol (RDP). A small packet of malware unpacks and executes, establishing a command and control link to the outside attacker.

Once on the network, the attacker can encrypt critical files, making them inaccessible to the user without the encryption key. A ransomware screen (see Figure 4) instructs the victim how to pay the ransom using cryptocurrency which usually – but not always – results in the delivery of the decryption key and restoration of access to the user's information.

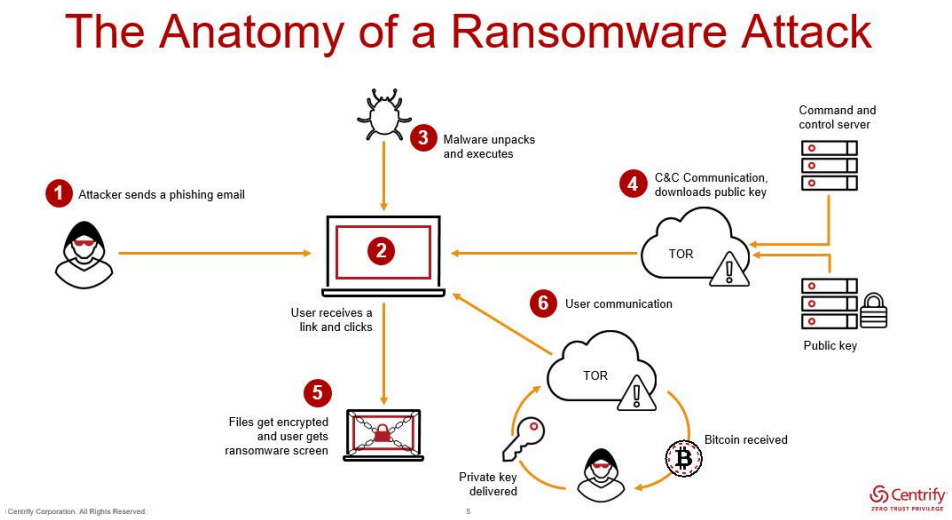


Figure 3. Anatomy Of A Ransomware Attack



Figure 4. Typical Ransomware Screen

A fast-growing variation of ransomware attacks involves multi-extortion techniques in which the attackers not only encrypt the files of an organization, but also name and shame the victims or threaten to launch distributed denial of service (DDoS) attacks to encourage the victims to pay more quickly. In 2021, the names and proof of compromise for 2,566 victims were publicly posted on ransomware leak sites, marking an 85% increase compared to 2020.

Ransomware attacks create a dilemma for healthcare organizations. The financial impact of a successful ransomware attack can be significant, but more damaging is the sudden inability for clinicians to access patient information. In addition, the threat of release of personal health information (PHI) is chilling because it would subject the institution to public humiliation, lawsuits from patients, and regulatory action due to HIPAA violations. As you will see later in this white paper, the considerations for recovering from a ransomware attack in a timely way are somewhat different from other kinds of disaster and require special attention in the designing of a DR system.



Disaster Recovery Objectives

DR system design revolves around the answers to two key questions:

- How much data can you afford to lose or recreate?
- How quickly do you need to recover?

The answers help establish two crucial design parameters: the recovery point objective (data loss) and the recovery time objective (downtime). These two parameters dictate the design of the DR system, so it's essential to get them right.

Recovery Time Objective (RTO)

The Recovery Time Objective is the maximum time that the business can tolerate the unavailability of a system, network, or service. Beyond this point, significant damage to the business will be incurred. RTO represents a business-aligned operational target that comprises part of the minimum goal of a DR plan (see figure 5 and table 1).

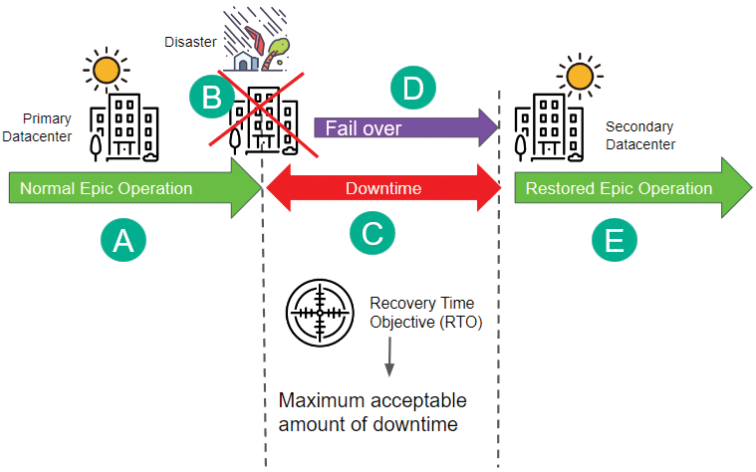


Figure 5. Recovery Time Objective

Step	State	Description
A	Normal operation	During normal operation, clinicians, business analysts and others have full access to the Epic modules they need to deliver care effectively.
B	Disaster event	An unexpected external event disables the primary data center.
C	Service interruption (downtime)	With the primary data center off-line, all stakeholders are without access to Epic. Many are forced to resort to manual note-taking and other means of communication, while some are completely unable to do their jobs. The RTO specifies the maximum amount of time that service interruption can continue without significant damage to the organization.
D	Failover	At the moment when the service interruption is detected, the remediation steps in the DR system are initiated. Most DR systems include some automated steps, but Epic recommends manual activation of the DR activities for the IRIS/Chronicle database that stores patient records.
E	Restore operation	When the failover operation finishes, Epic is back online. Now clinicians and others can be fully productive again. However, some amount of information has been lost, see recovery point objective below.

Recovery Point Objective (RPO)

The Recovery Point Objective measures the maximum data loss from an outage that the business can tolerate (see figure 6). RPO forms the second half of the minimum goal of the DR plan. Where RTO specifies how long forward from the moment of outage the system can be unavailable, RPO specifies how far back from the moment of outage that loss of data can extend.

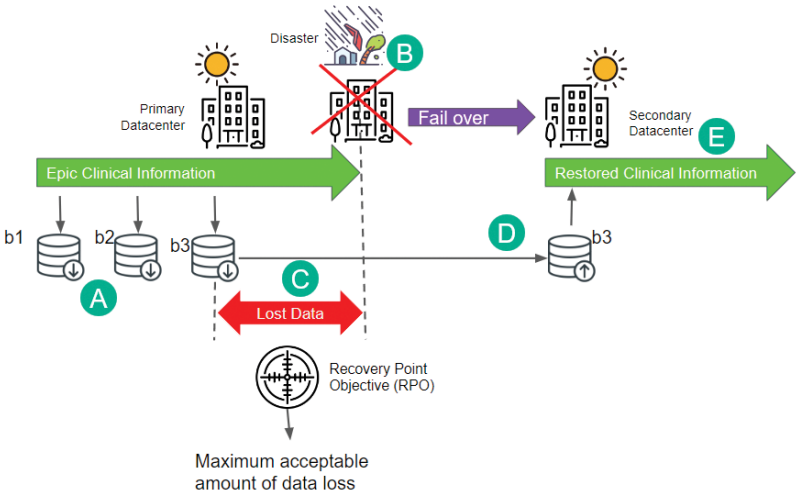


Figure 6. Recovery Point Objective

Step	State	Description
A	Normal operation	Information in the Epic databases is regularly backed up with mirroring or other backup technology. Backups are replicated to the secondary data center as they are created.
B	Disaster event	An unexpected external event disables the primary data center.
C	Lost data	Any changes to the Epic databases between the last backup and the disaster event are lost forever. The RPO is based on the maximum amount of data that your organization can afford to lose due to a disaster event.
D	Data restore	In some DR strategies, the RPO is defined by the last backup. Since backups typically occur every 4 to 24 hours, this approach can result in the loss of up to a full day's information – completely unacceptable in the healthcare world. Epic's standard implementation for the OLTP DR maintains an asynchronous replication from the primary node to the DR node. Because of this architectural strategy, the RPO is equivalent to the replication latency, usually on the order of a few minutes.
E	Restore operation	When the failover operation finishes, Epic is back online. Now clinicians and others can be fully productive again.

DR Options on AWS

Broadly speaking, there are four ways to implement DR on AWS: backup and restore, pilot light, warm standby, and multi-site (active-active) (see figure 7). Each option is described below in general terms – the next section will describe the Epic-specific aspects of each method.

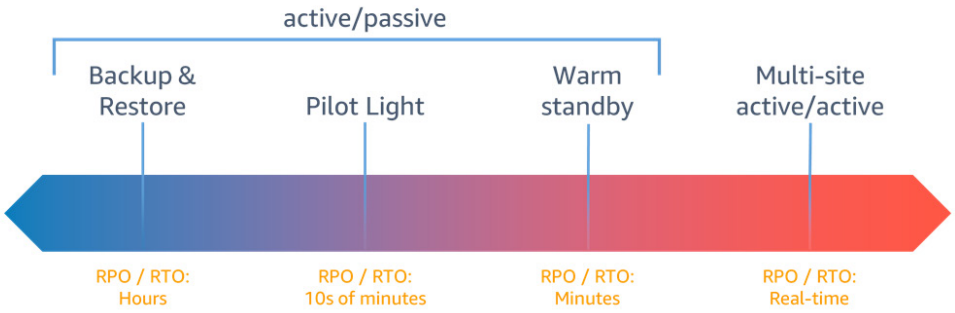


Figure 7. Four DR Options For AWS

Backup and Restore: Slow, Cheap, and Reliable

Backup and restore (B&R) has historically been a relatively slow, albeit highly reliable, way to restore information following a DR event. Traditional backup uses backup software to write to lower-cost (and therefore slower) SATA storage and archival tape that can be stored offsite for high survivability regardless of the event. This method is not only slow, but somewhat error prone-for example, a misaligned write head on a tape drive can result in an unreadable tape, which you may not discover until you actually need the information.

The basic B&R process on AWS is to take frequent snapshots of data and store them in Amazon S3 Glacier instances for later retrieval. S3 Glacier is intended for long-term archival storage and cuts storage costs significantly compared to S3 Standard. The backup data can be transferred over a dedicated network line using AWS Direct Connect. Extremely large data transfers can make use of AWS Snowball Edge, which is primarily intended as a one-time migration tool.

For many Healthcare Information Technology (HIT) groups, backup and restore with offsite copies can ensure data availability at the cost of high RTO (typically 24 hours) and RPO (>24 hours, depending on when a failure occurs). Offsite copies of backups are a common requirement of cybersecurity insurance carriers.

Pilot Light: Fast Recovery

The most common method of active-passive DR is the pilot light plan. This method takes its name from furnaces, where a small light burning continuously enables the furnace to ramp up the heat quickly without requiring manual relighting. In the IT analog, the critical core pieces of the system – for example, Elastic Load Balancing and Amazon EC2 Auto Scaling – are active, while functional elements such as EC2 instances are idle. In case of a failover, rapidly provisioning EC2 instances as needed using Amazon Machine Images (AMIs) can result in a relatively fast RTO in tens of minutes. Continuous data replication using Elastic Block Store (EBS) snapshots can reduce RPO to minutes or seconds.

The pilot light approach can be more costly than B&R because some core AWS resources are constantly running. Pilot light is an example of an active-passive strategy, as is warm standby, our next topic.

Warm Standby: Ready to Go

Also an active-passive strategy, this DR plan differs from pilot light designs in that some minimum number of EC2 instances run at all times. When failover occurs, the EC2 instances scale up automatically via EC2 Auto Scaling and traffic is rerouted via Elastic Load Balancing. Data replication uses the same method as pilot light scenarios. Warm standby can lead to faster recovery than pilot light scenarios due to ready-to-go resources. Warm standby can cut the RTO to minutes but incurs significant ongoing expenses compared to pilot light designs. Evaluating this cost-performance tradeoff is one of your organization's critical decisions to make in planning your Epic DR system.

Multi-Site Active-Active: Real-time Recovery

This method is the gold standard for DR. In a multi-site active-active DR system, a one-to-one copy of the primary infrastructure continually runs (active) in a remote location. If the primary infrastructure fails, the DR system takes over immediately, reducing RTO and RPO to near zero. It doesn't require a detailed financial analysis to realize that this method is quite expensive, possibly doubling the ongoing cost of running the infrastructure. Some implementations scale down the DR system slightly to mitigate the ongoing costs, but this method runs the risk of compromising the ability of the secondary infrastructure.

It's important to note that organizations commonly mix the available options based on the criticality of the workload. An example would be that tier 1 (critical) systems will use active-active, tier 2 (line of business) systems may use a mix of warm standby or pilot light, and tier 3 (non-production) systems will rely solely on backup and recovery.



AWS Regions and Availability Zones

Our discussion of DR on AWS so far has been abstract and detached from considerations of physical location. However, even when you run Epic in the cloud, a real-world data center still hosts your data – and the physical location has implications for effective DR design, to say nothing of compliance and other requirements.

The AWS global infrastructure spans 27 geographic regions, each containing multiple Availability Zones (AZs) and Local Zones (LZs). AWS Regions are geographically distinct areas containing Availability Zones. Each Availability Zone has 2+ physical data center locations. Local Zones act as an Availability Zone in a single data center closer to users and data for ultra-low latency applications. In disaster recovery plans, multiple AZs give network planners a powerful mechanism to increase redundancy and thus improve reliability (see figure 8).

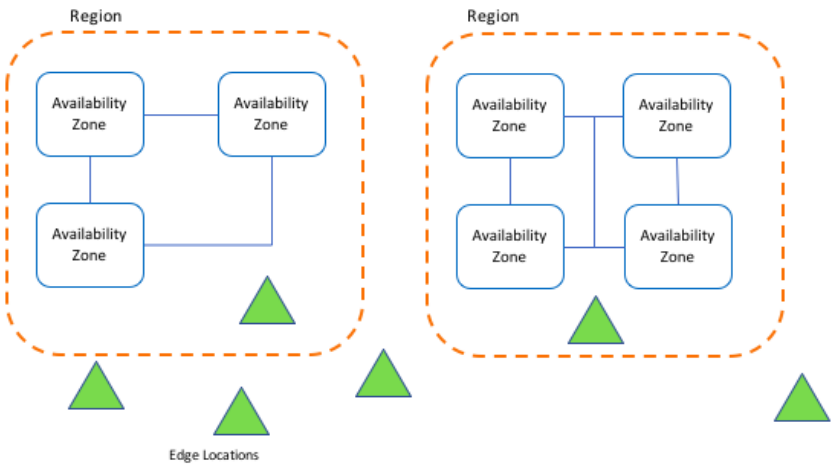


Figure 8. AWS Regions and Availability Zones

Shared Responsibility Model for Resilience

When you migrate from running Epic in your on-premises data center to cloud-based deployment, you enter into a partnership with your cloud provider where you share responsibility for security, compliance, and resiliency that once were entirely under your control. Understanding how the shared responsibility model works and operationally affects your organization can be critical to ensure a secure, resilient infrastructure.

The division of responsibilities of the two parties lies between the infrastructure and the workloads: AWS ensures the resiliency of the cloud - the hardware, software, networking and facilities - whereas the customer is responsible for the resiliency of the workloads that run upon that cloud. Service Level Agreements (SLAs) ensure Amazon's commitment to the operational efficacy of its customers.

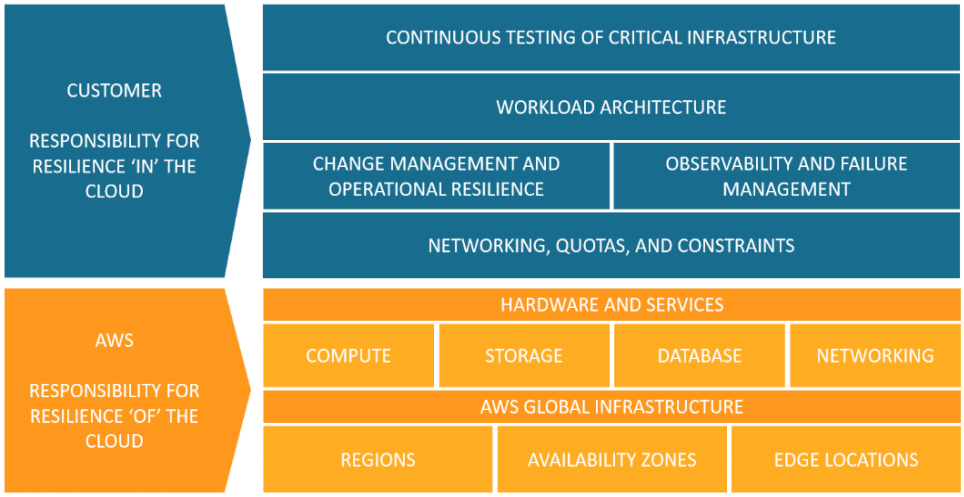


Figure 9. The AWS Shared Responsibility Model For Resilience

AWS Services for DR

AWS offers many services that can streamline and automate critical parts of the recovery process. **AWS Backup** provides a centralized view of backup activities including tools to help system administrators configure, schedule and monitor backups and automate restore processes, as well as cross-region copy capability. **AWS CloudFormation** is an infrastructure-as-code (IaC) capability that allows you to quickly deploy EC2 instances in response to an event. When data restore operations complete, **Amazon EventBridge** receives notifications. In turn, it notifies operators using **AWS System Manager** and invokes the reconnection logic using **AWS Lambda** (see figure 10).

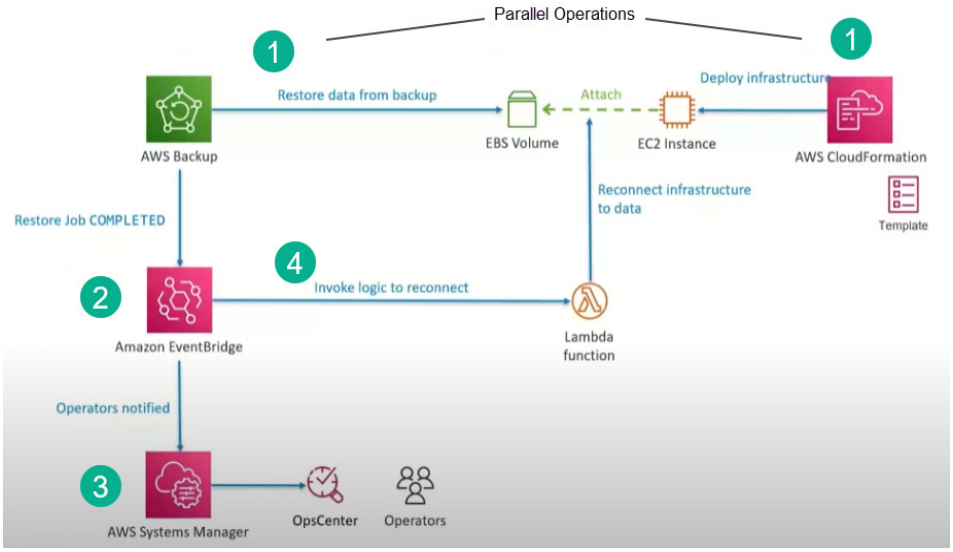


Figure 10. Typical Restore Operation Using AWS Services

Epic Data Architecture

The performance requirements and unique workflows of healthcare delivery can overwhelm traditional relational databases. To meet this need, Epic built the Chronicles online transactional database (OLTP) on InterSystems IRIS, a high-throughput key-value database engine.

Epic and InterSystems have a long history of collaboration going back to the 1970s. Learn more about InterSystems IRIS here.

An Epic deployment's typical database architecture includes IRIS, Microsoft SQL databases, and WebBLOB information in block format (see figure 11).

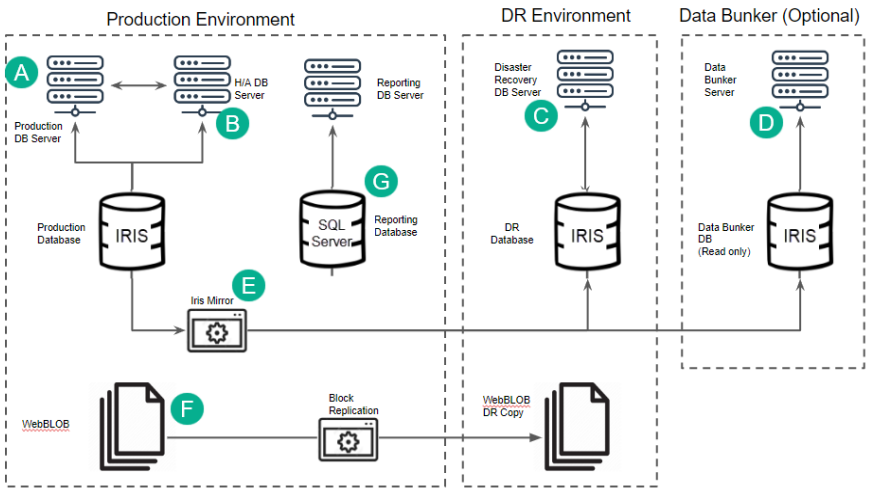


Figure 11. Epic Data Architecture



II. DESIGNING YOUR EPIC DR SYSTEM

This section addresses many vital concepts and best practices for designing an AWS-based DR system for Epic. However, the discussions here are, by necessity, general because the needs of individual institutions vary widely. It is strongly recommended that you engage with an experienced and proven managed services provider (MSP) or other expert third party to help you develop detailed requirements for your Epic DR, translate those requirements into a detailed design, and implement the deployment on AWS.

Step 1. Engage Your Third-Party Resources

InterSystems IRIS database technology.
The IRIS technology is closely coupled to the Epic Systems software architecture based on decades of collaborative development between Epic and InterSystems. Embedded technologies such as IRIS mirroring, FHIR interoperability, and database sharding make IRIS ideal for Epic, but also place demands on system architects using these technologies in Epic cloud deployments.

Epic DR best practices.
Epic Systems provides substantial guidance in data replication and DR objectives. Still, it requires a significant understanding of the Epic software and data architecture to make the most of this information.

Additional Software.
Customers commonly have software outside the Epic environment that provides additional functionality for areas such as document management, ICD-10 coding, and HL7 integration servers. Organizations must account for these systems in any DR plan.

AWS services for DR.
AWS provides a wide range of services such as EC2 Fleet, AWS CloudFormation, AWS EventBridge, and AWS Backup that can significantly simplify DR system design. However, choosing the appropriate services and integrating them for a critical application such as disaster recovery is best left to a managed services provider or other third parties with extensive experience in AWS services.

AWS pricing strategies.
AWS offers a pay-as-you-go approach that allows you to pay only for the individual services you need for as long as you use them. Furthermore, AWS pricing is completely transparent, requiring no long-term contracts or complex licensing. However, optimizing your AWS spend can be a non-trivial exercise that dramatically affects your ultimate total cost of ownership for the Epic DR system. Expert advice in this area often pays for itself many times over.

Step 2. Establish Your RPO

Calculating the optimum RPO and RTO for your organization is often as much an art as a science. Different software systems often have different RPO and RTO values depending on the importance of the system to hospital operations. For example, a hospital may consider the electronic health record (EHR) system its most critical application and assign it the shortest RPO and RTO. Email may also be a vital application necessary to communicate in the event of a disaster, so that it may have a short RTO but a longer RPO.

While this white paper focuses primarily on disaster recovery, it's important to remember that DR and high availability (HA) go hand in hand to ensure resiliency. Furthermore, security is also a critical part of the equation; ransomware and other cyber threats can cause major interruptions to hospital operations.

As defined before, RPO represents the amount of information an organization can withstand losing in an event that takes down the Epic system. A typical process for determining RPO involves gathering information from several functional areas within the organization (see table 1).

Factor	Function
Clinical impact of outages	Chief Medical Officer, Chief Nursing Officer
Regulatory mandates	Chief Compliance Officer
Risk analysis	Chief Risk Officer
Best practices	Epic application team
Budgetary limitations	Chief Financial Officer
Technology requirements	Chief Technology Officer, Chief Information Officer

Table 1. Factors Influencing Determination of RPO

One common strategy is to establish different RPOs based on the criticality of the information content by database (see table 2).

Information Type	Criticality	Typical RPO
EHR database	High	Seconds
Reporting database	Medium	One day
WebBLOB	Low	Hours

Table 2. Typical RPO Categorization

Step 3. Establish Your RTO

Each of the primary cloud service providers – Amazon Web Services (AWS), Microsoft Azure, and Google Cloud – have their own business models and service offerings, which can be tricky to navigate. An MSP who knows your cloud service provider inside and out can add significant value by helping you choose the right pricing tiers to minimize your spending for computing, storage, and networking services. In addition, your MSP can help you configure and deploy tools offered by your cloud provider in areas such as security, analytics, compliance, configuration management, and AI/machine learning.



Timeline Component	Description
Detection	The determination that a significant event has occurred that could degrade or prevent access to clinical and business information. Must be automatic, can never be human-determined.
Notification	Messages sent to all first-line decision-makers letting them know that the event has occurred and provide whatever decision-making information is available.
Escalation	If first-line decision makers perceive a threat to continued operation of the Epic software, senior management should be briefed and brought into the process.
Declaration	Formal decision to failover.
Failover	The actual process of switching users to the DR site and away from the now-compromised primary environment.

Figure 12. RTO Sequence

Step 4. Map Out Your DR Architecture

This section describes a DR system architecture with three major components: primary, secondary, and data bunker sites (see figure 13).

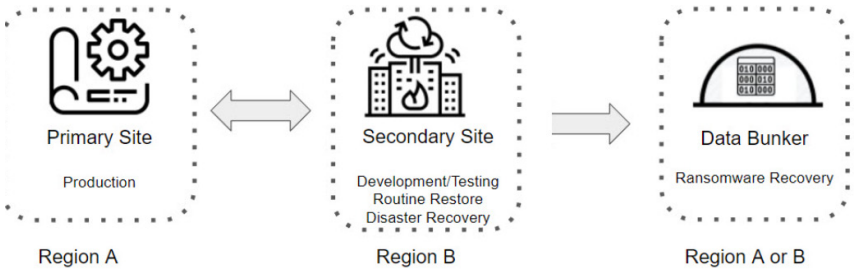
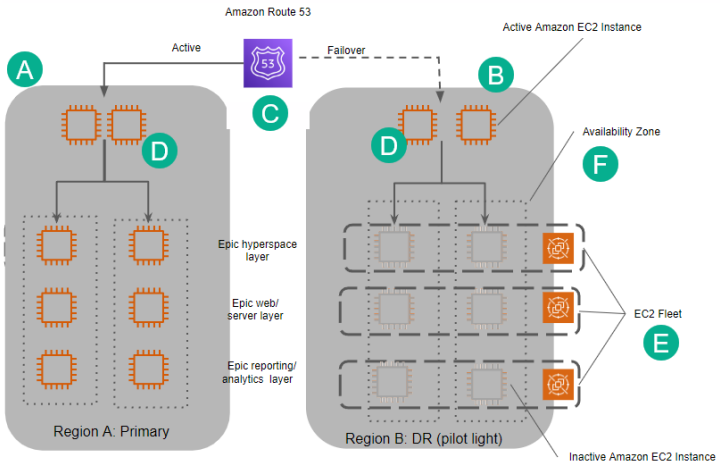


Figure 13. Overall Design Of Epic DR System

The data bunker tier is becoming more and more essential as the threat of ransomware continues to increase. Briefly, think of the bunker as a data lake that stores large amounts of data in a usable state on relatively inexpensive (and slower) media and has no egress. The bunker does not run production resources under normal circumstances but serves as the ultimate recovery point in a major disaster. Replication is one-way and continuous.

Step 5. Choose a Recovery Architecture

The pilot light architecture discussed earlier is the most common choice for Epic DR systems based on AWS. The pilot light can be as little as 50% the cost of multi-site active-active solutions and still achieve the RPO and RTO results that modern hospitals seek (figure 14).



Item	Component	Description
A	Active Region (Primary)	All Epic software runs here on EC2 instances in normal operation.
B	Passive Region (DR)	The application server architecture is in place but there are no running EC2 instances. In the event of failover, EC2 instances can be provisioned in seconds or minutes using automated processes.
C	Route 53	Amazon's DNS service is used to reroute requests to the DR system in the event of a failover.
D	Load balancing	An HA pair of EC2 instances performs the load balancing function.
E	EC2 Fleets	An EC2 Fleet contains the configuration information to launch a fleet—or group—of instances. In a single API call, a fleet can launch multiple instance types across multiple Availability Zones.
F	Availability Zones	Multiple AZs implement high availability to prevent downtime due to hardware or software failure.

Figure 14. Epic Pilot Light DR Implementation on AWS

Don’t Try This at Home

Of necessity, this paper can only be descriptive, not prescriptive, because every implementation of Epic DR on AWS must be carefully crafted to the organization’s requirements, resources, and strategic goals. One of our goals in creating this document was not only to provide useful background information about using AWS for Epic disaster recovery but also to show the level of complexity of designing a disaster recovery system that meets the stringent requirements that hospitals have for their Epic deployments.

We cannot emphasize enough the importance of contracting with highly experienced third parties when you take on the challenge of implementing an Epic DR system on AWS. The stakes are high and there is great value in the experience that partners can bring, not the least of which is the peace of mind you will get when working with an MSP or other knowledgeable third-party vendor. In addition, bringing in the outside talent you need minimizes risk and makes it far more likely that you can achieve the goals of the project within the budget and time constraints under which you are working.

To bring this challenging skill set to the marketplace, Cloudticity has recently announced a strategic alignment with Sapphire Health. Customers can now work with Sapphire Health and Cloudticity to run Epic EHR, Epic business continuity and disaster recovery (BR/DC), and Epic archived images on AWS Cloud in a safe, secure, and HIPAA-compliant environment. Read the press release here.

ABOUT SAPPHIRE HEALTH

Sapphire Health supports healthcare organizations with a team of seasoned healthcare IT professionals to modernize and secure their Epic infrastructure. Sapphire Health can assist with all aspects of planning, design, and execution for Epic platform migrations, infrastructure automation and optimization, IT managed services, and business continuity and resiliency. For more information, visit https://sapphirehealth.org.

ABOUT CLOUDTICITY

Cloudticity is a digital enablement partner for the healthcare industry generating measurable business and clinical outcomes by unlocking the full potential of the cloud. Through groundbreaking automation and deep cloud expertise, Cloudticity solutions empower healthcare organizations to create and scale the next generation of healthcare solutions. For more information, visit http://www.cloudticity.com.

READY TO TAKE THE NEXT STEPS?

When you're ready, we're ready. Schedule a free consultation with a healthcare cloud expert today.


cloudticity
Meet with an expert and map out your Epic cloud journey today!
SCHEDULE FREE CONSULTATION