AWS Kerangka Well-Architected

Pemulihan Bencana Beban Kerja di AWS: Pemulihan di Cloud

Pemulihan Bencana Beban Kerja di AWS: Pemulihan di Cloud: AWS Kerangka Well-Architected

Table of Contents

Pemulihan Bencana Beban Kerja di AWS: Pemulihan di Cloud

Tanggal publikasi: 12 Februari 2021 (Riwayat dokumen)

Pemulihan bencana adalah proses mempersiapkan dan memulihkan diri dari bencana. Suatu peristiwa yang mencegah beban kerja atau sistem memenuhi tujuan bisnisnya di lokasi utama yang digunakan dianggap sebagai bencana. Paper ini menguraikan praktik terbaik untuk merencanakan dan menguji pemulihan bencana untuk setiap beban kerja yang digunakan AWS, dan menawarkan pendekatan berbeda untuk mengurangi risiko dan memenuhi Recovery Time Objective (RTO) dan Recovery Point Objective (RPO) untuk beban kerja tersebut.

Whitepaper ini mencakup bagaimana menerapkan pemulihan bencana untuk beban kerja pada. AWS Lihat Pemulihan Bencana Aplikasi Lokal AWS untuk informasi tentang penggunaan AWS sebagai situs pemulihan bencana untuk beban kerja lokal.

Pengantar

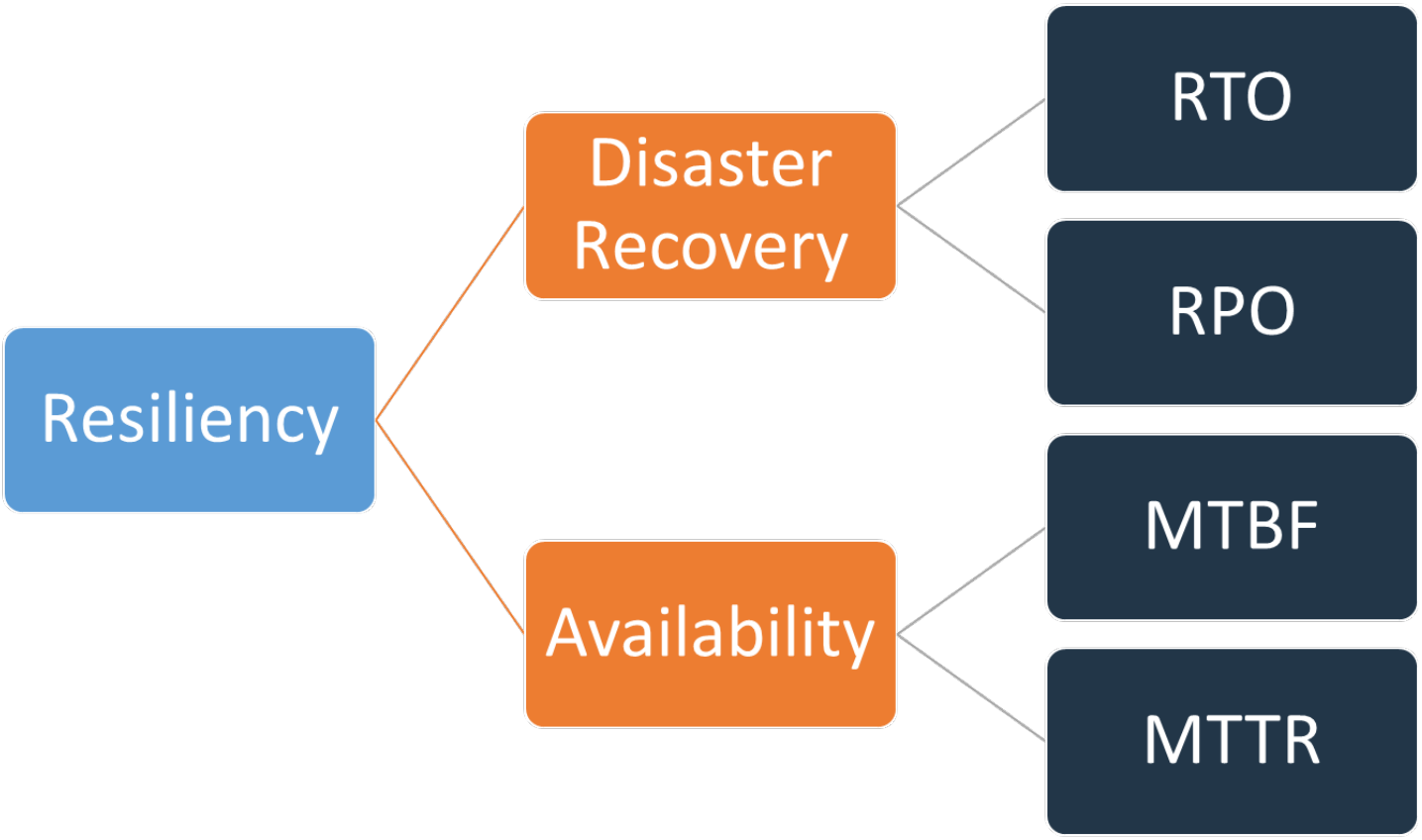
Beban kerja Anda harus menjalankan fungsi yang dimaksudkan dengan benar dan konsisten. Untuk mencapai ini, Anda harus arsitek untuk ketahanan. Ketahanan adalah kemampuan beban kerja untuk pulih dari gangguan infrastruktur, layanan, atau aplikasi, memperoleh sumber daya komputasi secara dinamis untuk memenuhi permintaan, dan mengurangi gangguan, seperti kesalahan konfigurasi atau masalah jaringan sementara.

Pemulihan bencana (DR) adalah bagian penting dari strategi ketahanan Anda dan menyangkut bagaimana beban kerja Anda merespons ketika bencana melanda (bencana adalah peristiwa yang menyebabkan dampak negatif yang serius pada bisnis Anda). Respons ini harus didasarkan pada tujuan bisnis organisasi Anda yang menentukan strategi beban kerja Anda untuk menghindari hilangnya data, yang dikenal sebagai Recovery Point Objective (RPO), dan mengurangi downtime di mana beban kerja Anda tidak tersedia untuk digunakan, yang dikenal sebagai Recovery Time Objective (RTO). Oleh karena itu, Anda harus menerapkan ketahanan dalam desain beban kerja Anda di cloud untuk memenuhi tujuan pemulihan Anda (RPO dan RTO) untuk peristiwa bencana satu kali tertentu. Pendekatan ini membantu organisasi Anda untuk menjaga kelangsungan bisnis sebagai bagian dari Business Continuity Planning (BCP).

Paper ini berfokus pada bagaimana merencanakan, merancang, dan mengimplementasikan arsitektur AWS yang memenuhi tujuan pemulihan bencana untuk bisnis Anda. Informasi yang dibagikan di sini ditujukan bagi mereka yang memiliki peran teknologi, seperti chief technology officer (CTOs), arsitek, pengembang, anggota tim operasi, dan mereka yang bertugas menilai dan mengurangi risiko.

Pemulihan dan ketersediaan bencana

Pemulihan bencana dapat dibandingkan dengan ketersediaan, yang merupakan komponen penting lainnya dari strategi ketahanan Anda. Sementara pemulihan bencana mengukur tujuan untuk peristiwa satu kali, tujuan ketersediaan mengukur nilai rata-rata selama periode waktu tertentu.



Gambar 1 - Tujuan Ketahanan

Ketersediaan dihitung menggunakan Mean Time Between Failures (MTBF) dan Mean Time to Recover (MTTR):

$$Availability = \frac{Available\ for\ Use\ Time}{Total\ Time} = \frac{MTBF}{MTBF + MTTR}$$

Pendekatan ini sering disebut sebagai “sembilan”, di mana target ketersediaan 99,9% disebut sebagai “tiga sembilan”.

Untuk beban kerja Anda, mungkin lebih mudah untuk menghitung permintaan yang berhasil dan gagal daripada menggunakan pendekatan berbasis waktu. Dalam hal ini, perhitungan berikut dapat digunakan:

$$Availability = \frac{Successful\ Responses}{Valid\ Requests}$$

Pemulihan bencana berfokus pada peristiwa bencana, sedangkan ketersediaan berfokus pada gangguan yang lebih umum pada skala yang lebih kecil seperti kegagalan komponen, masalah jaringan, bug perangkat lunak, dan lonjakan beban. Tujuan dari pemulihan bencana adalah kesinambungan bisnis, sedangkan ketersediaan menyangkut memaksimalkan waktu bahwa beban kerja tersedia untuk menjalankan fungsi bisnis yang dimaksudkan. Keduanya harus menjadi bagian dari strategi ketahanan Anda.

Apakah Anda sudah Well-Architected?

AWS Well-Architected Framework membantu Anda memahami pro dan kontra dari keputusan yang Anda buat saat membangun sistem di cloud. Enam pilar dari Kerangka Kerja ini memungkinkan Anda mempelajari praktik terbaik arsitektural untuk merancang dan mengoperasikan sistem yang andal, aman, efisien, hemat biaya, dan berkelanjutan. Menggunakan AWS Well-Architected Tool, tersedia tanpa biaya di AWS Management Console, Anda dapat meninjau beban kerja Anda terhadap praktik terbaik ini dengan menjawab serangkaian pertanyaan untuk setiap pilar.

Konsep yang tercakup dalam whitepaper ini memperluas praktik terbaik yang terkandung dalam whitepaper Reliability Pillar, khususnya pertanyaan REL 13, “Bagaimana Anda merencanakan pemulihan bencana (DR)?”. Setelah menerapkan praktik di whitepaper ini, pastikan untuk meninjau (atau meninjau ulang) beban kerja Anda menggunakan AWS Well-Architected Tool.

Model Tanggung Jawab Bersama untuk Ketangguhan

Ketahanan adalah tanggung jawab bersama antara AWS dan Anda, pelanggan. Penting bagi Anda untuk memahami bagaimana pemulihan dan ketersediaan bencana, sebagai bagian dari ketahanan, beroperasi di bawah model bersama ini.

Tanggung jawab AWS “Ketahanan Cloud”

AWS bertanggung jawab atas ketahanan infrastruktur yang menjalankan semua layanan yang ditawarkan di AWS Cloud. Infrastruktur ini terdiri dari perangkat keras, perangkat lunak, jaringan, dan fasilitas yang menjalankan layanan AWS Cloud. AWS menggunakan upaya yang wajar secara komersial untuk membuat layanan AWS Cloud ini tersedia, memastikan ketersediaan layanan memenuhi atau melampaui Perjanjian Tingkat Layanan AWS (SLAs).

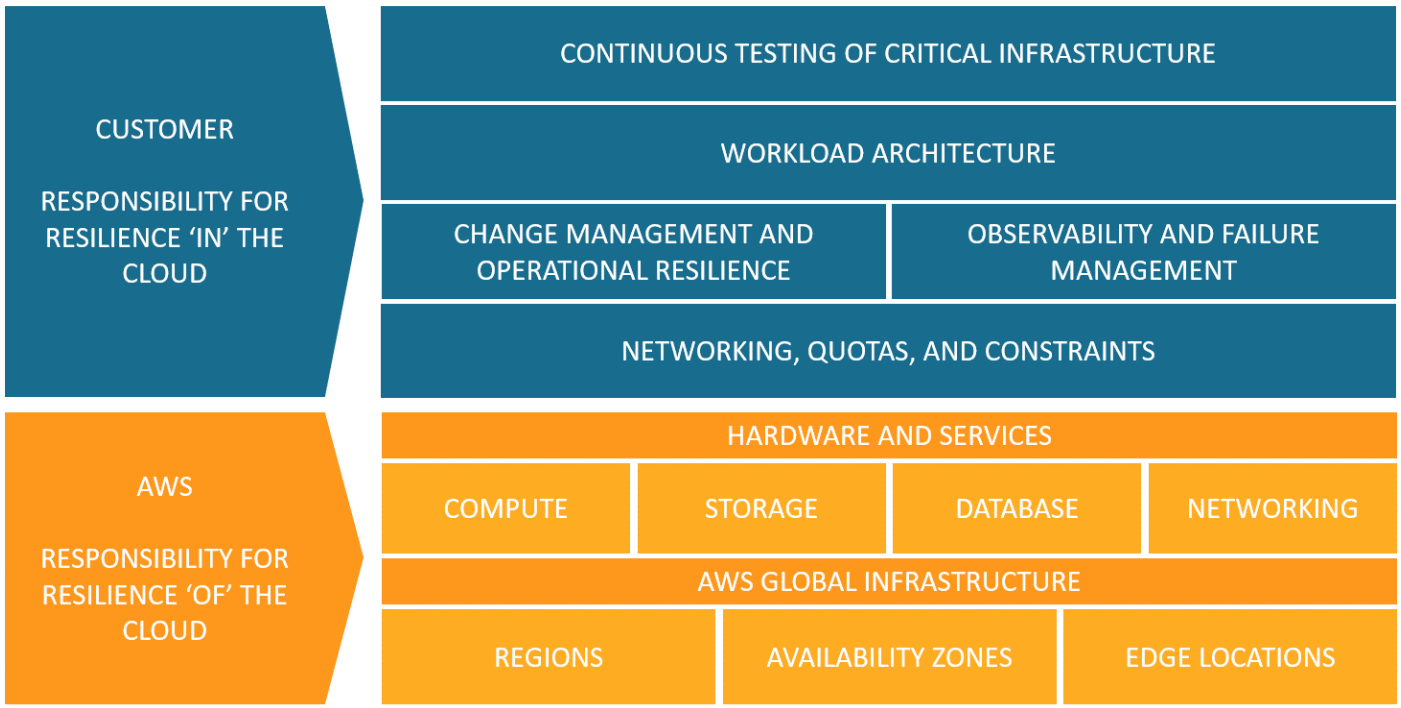
AWS Global Cloud Infrastructure dirancang untuk memungkinkan pelanggan membangun arsitektur beban kerja yang sangat tangguh. Setiap Wilayah AWS sepenuhnya terisolasi dan terdiri dari beberapa Availability Zone, yang merupakan partisi infrastruktur yang terisolasi secara fisik. Zona Ketersediaan mengisolasi kesalahan yang dapat memengaruhi ketangguhan beban kerja, yang akan mencegahnya untuk memengaruhi zona-zona lain di Wilayah. Tetapi pada saat yang sama, semua zona di Wilayah AWS saling berhubungan dengan jaringan bandwidth tinggi, latensi rendah, melalui serat metro khusus yang sepenuhnya redundan yang menyediakan jaringan throughput tinggi dan latensi rendah antar zona. Semua lalu lintas antara zona dienkripsi. Performa jaringan cukup untuk mendapatkan replikasi sinkron antara zona. Ketika sebuah aplikasi dipartisi AZs, perusahaan lebih terisolasi dan terlindungi dari masalah seperti pemadaman listrik, sambaran petir, tornado, angin topan, dan banyak lagi.

Tanggung jawab pelanggan “Ketahanan di Cloud”

Tanggung jawab Anda akan ditentukan oleh layanan AWS Cloud yang Anda pilih. Hal ini akan menentukan jumlah konfigurasi kerja yang harus Anda lakukan sebagai bagian dari tanggung jawab ketangguhan Anda. Misalnya, layanan seperti Amazon Elastic Compute Cloud (Amazon EC2) mengharuskan pelanggan untuk melakukan semua konfigurasi ketahanan dan tugas manajemen yang diperlukan. Pelanggan yang menerapkan EC2 instans Amazon bertanggung jawab untuk menerapkan EC2 instans di beberapa lokasi (seperti AWS Availability Zones), menerapkan penyembuhan mandiri menggunakan layanan seperti Amazon Auto Scaling, serta menggunakan EC2 praktik terbaik arsitektur beban kerja yang tangguh untuk aplikasi yang diinstal pada instans.

Untuk layanan terkelola, seperti Amazon S3 dan Amazon DynamoDB, AWS mengoperasikan lapisan infrastruktur, sistem operasi, dan platform, dan pelanggan mengakses titik akhir untuk menyimpan dan mengambil data. Anda bertanggung jawab untuk mengelola ketangguhan data Anda, termasuk strategi pencadangan, penentuan versi, dan replikasi.

Menerapkan beban kerja Anda di beberapa Availability Zone di Wilayah AWS adalah bagian dari strategi ketersediaan tinggi yang dirancang untuk melindungi beban kerja dengan mengisolasi masalah ke satu Availability Zone, dan menggunakan redundansi Availability Zone lainnya untuk terus melayani permintaan. Arsitektur Multi-AZ juga merupakan bagian dari strategi DR yang didesain untuk membuat beban kerja menjadi lebih terisolasi dan terlindungi dari masalah-masalah seperti pemadaman listrik, sambaran petir, angin topan, gempa bumi, dan lain-lain. Strategi DR juga dapat menggunakan beberapa Wilayah AWS. Misalnya dalam konfigurasi aktif/pasif, layanan untuk beban kerja akan gagal dari wilayah aktifnya ke wilayah DR-nya jika Wilayah aktif tidak dapat lagi melayani permintaan.



Gambar 2 - Ketahanan adalah tanggung jawab bersama antara AWS dan pelanggan

Apa itu bencana?

Saat merencanakan pemulihan bencana, evaluasi rencana Anda untuk tiga kategori utama bencana ini:

- Bencana alam, seperti gempa bumi atau banjir
- Kegagalan teknis, seperti kegagalan daya atau konektivitas jaringan
- Tindakan manusia, seperti kesalahan konfigurasi yang tidak disengaja atau akses atau modifikasi unauthorized/outside pihak

Masing-masing potensi bencana ini juga akan memiliki dampak geografis yang dapat bersifat lokal, regional, seluruh negara, benua, atau global. Baik sifat bencana maupun dampak geografis penting ketika mempertimbangkan strategi pemulihan bencana Anda. Misalnya, Anda dapat mengurangi masalah banjir lokal yang menyebabkan pemadaman pusat data dengan menggunakan strategi Multi-AZ, karena tidak akan memengaruhi lebih dari satu Availability Zone. Namun, serangan terhadap data produksi akan mengharuskan Anda untuk menjalankan strategi pemulihan bencana yang gagal untuk mencadangkan data di Wilayah AWS lain.

Ketersediaan tinggi bukan pemulihan bencana

Ketersediaan dan pemulihan bencana bergantung pada beberapa praktik terbaik yang sama, seperti pemantauan kegagalan, penyebaran ke beberapa lokasi, dan failover otomatis. Namun, Ketersediaan berfokus pada komponen beban kerja, sedangkan pemulihan bencana berfokus pada salinan diskrit dari seluruh beban kerja. Pemulihan bencana memiliki tujuan yang berbeda dari Ketersediaan, mengukur waktu hingga pemulihan setelah peristiwa skala besar yang memenuhi syarat sebagai bencana. Anda harus terlebih dahulu memastikan beban kerja Anda memenuhi tujuan ketersediaan Anda, karena arsitektur yang sangat tersedia akan memungkinkan Anda untuk memenuhi kebutuhan pelanggan jika terjadi ketersediaan yang memengaruhi peristiwa. Strategi pemulihan bencana Anda memerlukan pendekatan yang berbeda dari yang tersedia, dengan fokus pada penerapan sistem diskrit ke beberapa lokasi, sehingga Anda dapat gagal mengatasi seluruh beban kerja jika perlu.

Anda harus mempertimbangkan ketersediaan beban kerja Anda dalam perencanaan pemulihan bencana, karena akan mempengaruhi pendekatan yang Anda ambil. Beban kerja yang berjalan pada satu EC2 instans Amazon dalam satu Availability Zone tidak memiliki ketersediaan tinggi. Jika masalah banjir lokal memengaruhi Availability Zone tersebut, skenario ini memerlukan failover ke AZ lain untuk memenuhi tujuan DR. Bandingkan skenario ini dengan beban kerja yang sangat tersedia yang digunakan di multi-situs aktif/aktif, di mana beban kerja diterapkan di beberapa Wilayah aktif dan semua Wilayah melayani lalu lintas produksi. Dalam hal ini, bahkan dalam kejadian yang tidak mungkin terjadi bencana besar membuat Wilayah tidak dapat digunakan, strategi DR dicapai dengan merutekan semua lalu lintas ke Wilayah yang tersisa.

Cara Anda mendekati data juga berbeda antara ketersediaan dan pemulihan bencana. Pertimbangkan solusi penyimpanan yang terus mereplikasi ke situs lain untuk mencapai ketersediaan tinggi (seperti multi-situs, beban active/active kerja). Jika file atau file dihapus atau rusak pada perangkat penyimpanan utama, perubahan destruktif tersebut dapat direplikasi ke perangkat penyimpanan sekunder. Dalam skenario ini, meskipun ketersediaan tinggi, kemampuan untuk gagal jika terjadi penghapusan data atau korupsi akan terganggu. Sebagai gantinya, point-in-time cadangan juga diperlukan sebagai bagian dari strategi DR.

Rencana Kelangsungan Bisnis (BCP)

Rencana pemulihan bencana Anda harus menjadi bagian dari rencana kelangsungan bisnis organisasi Anda (BCP), itu tidak boleh menjadi dokumen mandiri. Tidak ada gunanya mempertahankan target pemulihan bencana yang agresif untuk memulihkan beban kerja jika tujuan bisnis beban kerja itu tidak dapat dicapai karena dampak bencana pada elemen bisnis Anda selain beban kerja Anda. Misalnya gempa bumi dapat mencegah Anda mengangkut produk yang dibeli di aplikasi eCommerce Anda - bahkan jika DR yang efektif menjaga beban kerja Anda tetap berfungsi, BCP Anda perlu mengakomodasi kebutuhan transportasi. Strategi DR Anda harus didasarkan pada persyaratan bisnis, prioritas, dan konteks.

Analisis dampak bisnis dan penilaian risiko

Analisis dampak bisnis harus mengukur dampak bisnis dari gangguan terhadap beban kerja Anda. Ini harus mengidentifikasi dampak pada pelanggan internal dan eksternal karena tidak dapat menggunakan beban kerja Anda dan efek yang ada pada bisnis Anda. Analisis harus membantu menentukan seberapa cepat beban kerja perlu tersedia dan berapa banyak kehilangan data yang dapat ditoleransi. Namun, penting untuk dicatat bahwa tujuan pemulihan tidak boleh dibuat secara terpisah; probabilitas gangguan dan biaya pemulihan adalah faktor kunci yang membantu menginformasikan nilai bisnis dalam menyediakan pemulihan bencana untuk beban kerja.

Dampak bisnis mungkin bergantung pada waktu. Anda mungkin ingin mempertimbangkan untuk mempertimbangkan hal ini ke dalam perencanaan pemulihan bencana Anda. Misalnya, gangguan pada sistem penggajian Anda cenderung memiliki dampak yang sangat tinggi terhadap bisnis sebelum semua orang dibayar, tetapi mungkin berdampak rendah setelah semua orang dibayar.

Penilaian risiko dari jenis bencana dan dampak geografis bersama dengan ikhtisar implementasi teknis beban kerja Anda akan menentukan kemungkinan gangguan yang terjadi untuk setiap jenis bencana.

Untuk beban kerja yang sangat penting, Anda dapat mempertimbangkan untuk menerapkan infrastruktur di beberapa Wilayah dengan replikasi data dan pencadangan berkelanjutan untuk meminimalkan dampak bisnis. Untuk beban kerja yang kurang kritis, strategi yang valid mungkin tidak memiliki pemulihan bencana sama sekali. Dan untuk beberapa skenario bencana, juga valid untuk tidak memiliki strategi pemulihan bencana sebagai keputusan berdasarkan probabilitas rendah bencana terjadi. Ingatlah bahwa Availability Zone dalam Wilayah AWS sudah dirancang dengan

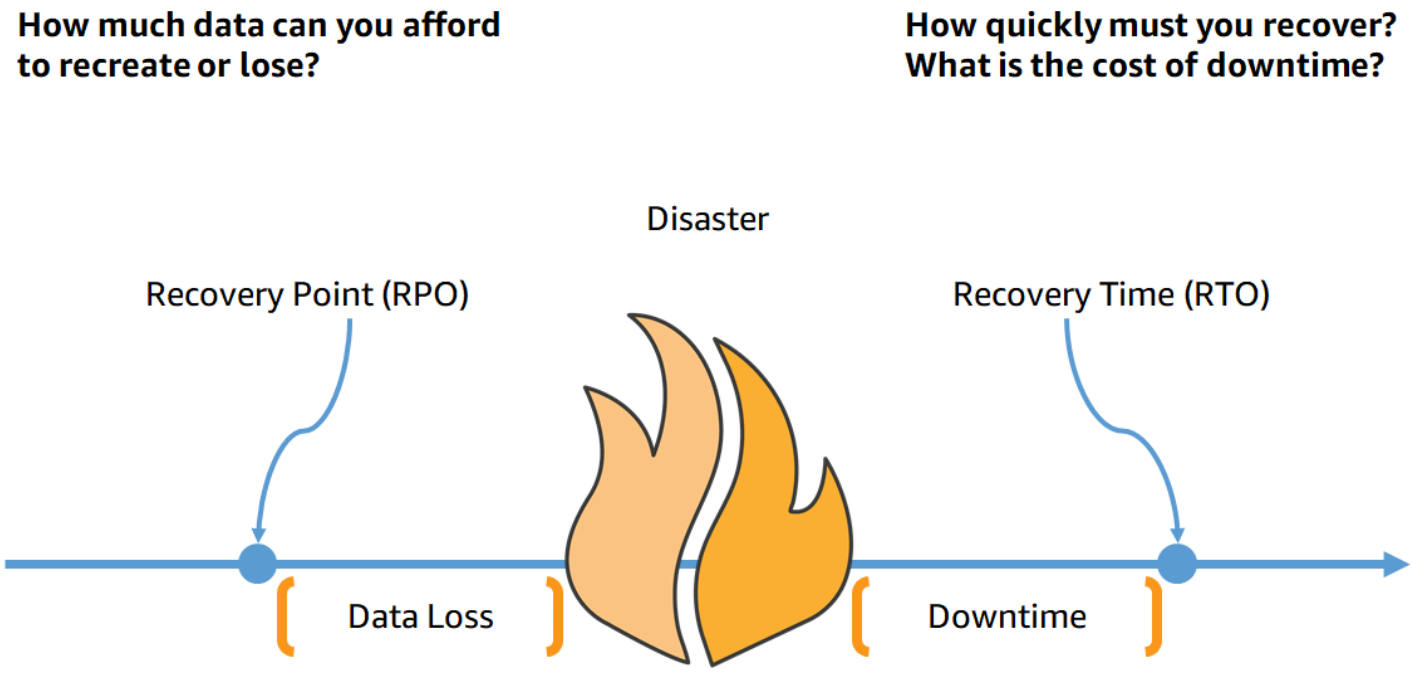
jarak yang berarti di antara mereka, dan perencanaan lokasi yang cermat, sehingga sebagian besar bencana umum hanya berdampak pada satu zona dan bukan zona lainnya. Oleh karena itu, arsitektur Multi-AZ dalam Wilayah AWS mungkin sudah memenuhi sebagian besar kebutuhan mitigasi risiko Anda.

Biaya opsi pemulihan bencana harus dievaluasi untuk memastikan bahwa strategi pemulihan bencana memberikan tingkat nilai bisnis yang benar dengan mempertimbangkan dampak dan risiko bisnis.

Dengan semua informasi ini, Anda dapat mendokumentasikan ancaman, risiko, dampak, dan biaya skenario bencana yang berbeda dan opsi pemulihan terkait. Informasi ini harus digunakan untuk menentukan tujuan pemulihan Anda untuk setiap beban kerja Anda.

Tujuan pemulihan (RTO dan RPO)

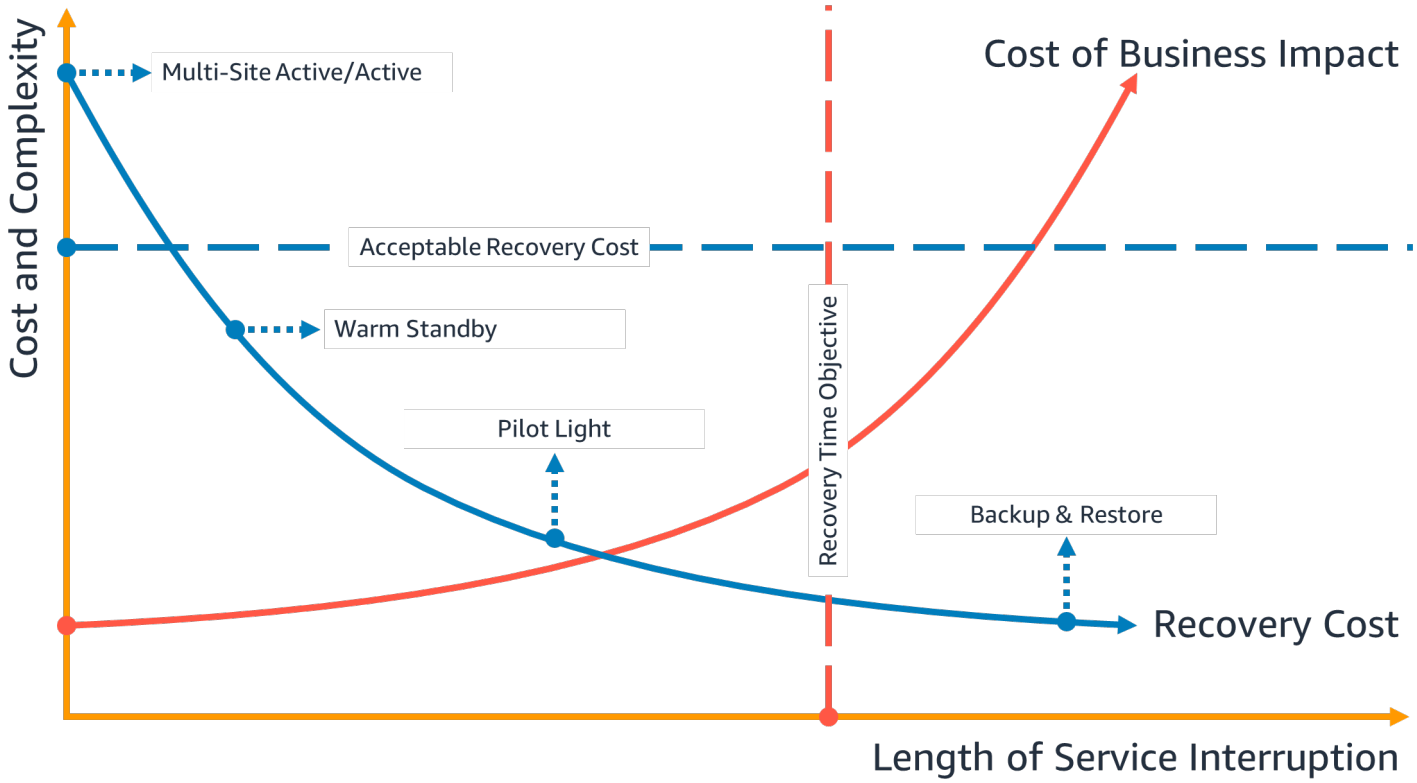
Saat membuat strategi Disaster Recovery (DR), organisasi paling sering merencanakan Recovery Time Objective (RTO) dan Recovery Point Objective (RPO).



Gambar 3 - Tujuan pemulihan

Recovery Time Objective (RTO) adalah penundaan maksimum yang dapat diterima antara gangguan layanan dan pemulihan layanan. Tujuan ini menentukan apa yang dianggap sebagai jendela waktu yang dapat diterima ketika layanan tidak tersedia dan ditentukan oleh organisasi.

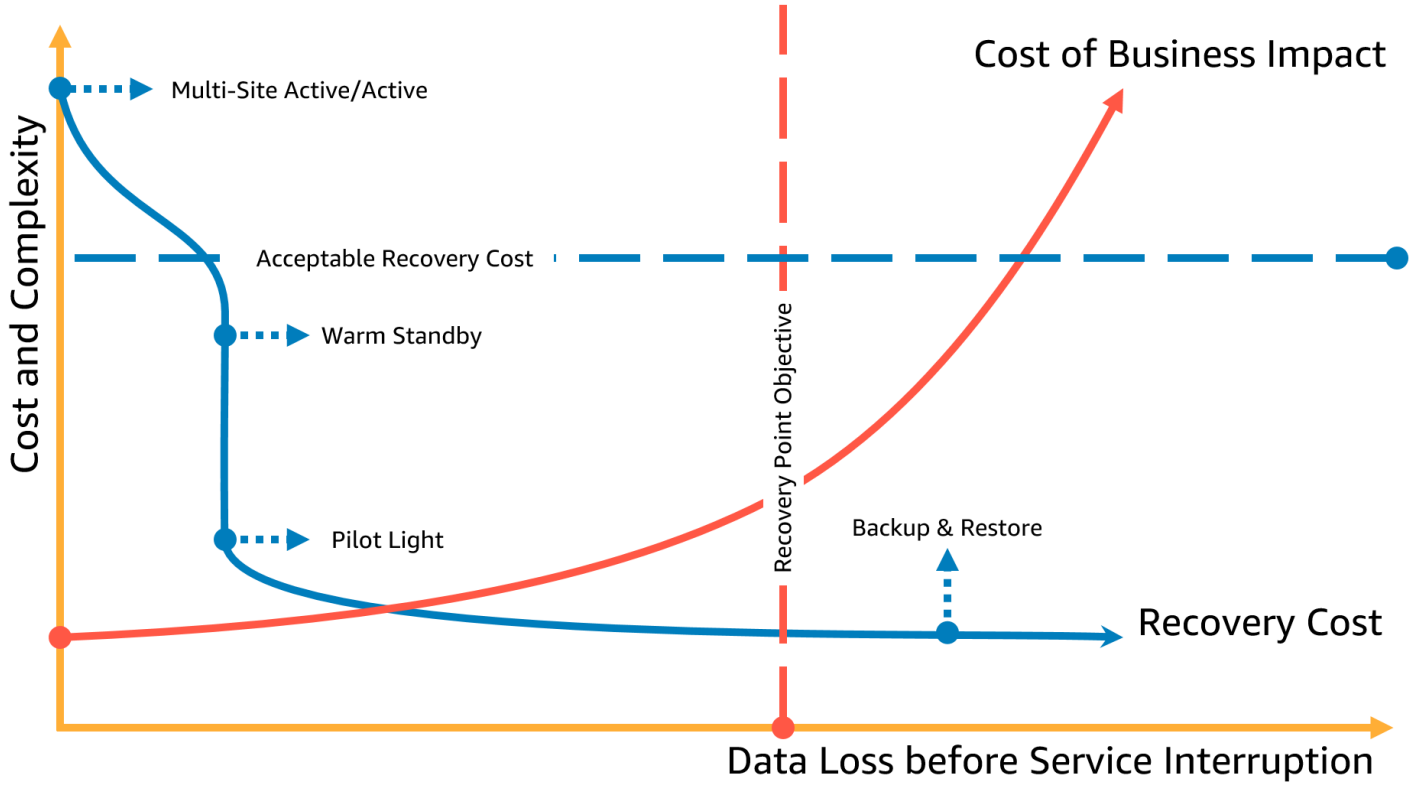
Ada empat strategi DR yang dibahas dalam paper ini: backup dan restore, pilot light, warm standby, dan multi-site active/active (lihat Disaster Recovery Options in the Cloud). Dalam diagram berikut, bisnis telah menentukan RTO maksimum yang diizinkan serta batas dari apa yang dapat mereka belanjakan untuk strategi restorasi layanan mereka. Mengingat tujuan bisnis, strategi DR Pilot Light atau Warm Standby akan memenuhi RTO dan kriteria biaya.



Gambar 4 - Tujuan waktu pemulihan

Recovery Point Objective (RPO) adalah jumlah waktu maksimum yang dapat diterima sejak titik pemulihan data terakhir. Tujuan ini menentukan apa yang dianggap sebagai hilangnya data yang dapat diterima antara titik pemulihan terakhir dan gangguan layanan dan ditentukan oleh organisasi.

Dalam diagram berikut, bisnis telah menentukan RPO maksimum yang diizinkan serta batas dari apa yang dapat mereka belanjakan untuk strategi pemulihan data mereka. Dari empat strategi DR, baik strategi Pilot Light atau Warm Standby DR memenuhi kriteria RPO dan biaya.



Gambar 5 - Tujuan titik pemulihan

> **Note**
>
> Jika biaya strategi pemulihan lebih tinggi daripada biaya kegagalan atau kerugian, opsi pemulihan tidak boleh diberlakukan kecuali ada pendorong sekunder seperti persyaratan peraturan. Pertimbangkan strategi pemulihan dengan berbagai biaya saat membuat penilaian ini.

Pemulihan bencana di cloud tidak sama dengan biasanya

Strategi pemulihan bencana berkembang dengan inovasi teknis. Rencana pemulihan bencana di lokasi mungkin melibatkan pengangkutan kaset secara fisik atau mereplikasi data ke situs lain. Organisasi Anda perlu mengevaluasi kembali dampak bisnis, risiko, dan biaya dari strategi pemulihan bencana sebelumnya untuk memenuhi tujuan DR di AWS. Pemulihan bencana di AWS Cloud mencakup keunggulan berikut dibandingkan lingkungan tradisional:

- Pulihkan dengan cepat dari bencana dengan kompleksitas yang berkurang
- Pengujian sederhana dan berulang memungkinkan Anda untuk menguji lebih mudah dan lebih sering
- Overhead manajemen yang lebih rendah mengurangi beban operasional
- Peluang untuk mengotomatisasi mengurangi kemungkinan kesalahan dan meningkatkan waktu pemulihan

AWS memungkinkan Anda untuk memperdagangkan biaya modal tetap dari pusat data cadangan fisik untuk biaya operasi variabel dari lingkungan yang berukuran benar di cloud, yang dapat secara signifikan mengurangi biaya.

Bagi banyak organisasi, pemulihan bencana lokal didasarkan pada risiko gangguan terhadap beban kerja atau beban kerja di pusat data dan pemulihan data yang dicadangkan atau direplikasi ke pusat data sekunder. Saat organisasi menerapkan beban kerja di AWS, mereka dapat menerapkan beban kerja yang dirancang dengan baik dan mengandalkan desain AWS Global Cloud Infrastructure untuk membantu mengurangi efek gangguan tersebut. Lihat whitepaper AWS Well-Architected Framework - Reliability Pillar untuk informasi selengkapnya tentang praktik terbaik arsitektur untuk merancang dan mengoperasikan beban kerja yang andal, aman, efisien, dan hemat biaya di cloud. Gunakan AWS Well-Architected Tooluntuk meninjau beban kerja Anda secara berkala untuk memastikan bahwa mereka mengikuti praktik terbaik dan panduan Kerangka Kerja Well-Architected. Alat ini tersedia tanpa biaya di AWS Management Console.

Jika beban kerja Anda menggunakan AWS, Anda tidak perlu khawatir tentang konektivitas pusat data (kecuali kemampuan Anda untuk mengaksesnya), daya, AC, pemadam kebakaran, dan perangkat keras. Semua ini dikelola untuk Anda dan Anda memiliki akses ke beberapa Zona Ketersediaan yang terisolasi kesalahan (masing-masing terdiri dari satu atau lebih pusat data diskrit).

Wilayah AWS Tunggal

Untuk peristiwa bencana berdasarkan gangguan atau hilangnya satu pusat data fisik, menerapkan beban kerja yang sangat tersedia di beberapa Availability Zone dalam satu Wilayah AWS membantu mengurangi bencana alam dan teknis. Pencadangan data yang berkelanjutan dalam Wilayah tunggal ini dapat mengurangi risiko terhadap ancaman manusia, seperti kesalahan atau aktivitas tidak sah yang dapat mengakibatkan hilangnya data. Setiap Wilayah AWS terdiri dari beberapa Availability Zone, masing-masing diisolasi dari kesalahan di zona lainnya. Setiap Availability Zone pada gilirannya terdiri dari satu atau lebih pusat data fisik diskrit. Untuk mengisolasi masalah yang berdampak dengan lebih baik dan mencapai ketersediaan tinggi, Anda dapat mempartisi beban kerja di beberapa zona di Wilayah yang sama. Availability Zones dirancang untuk redundansi fisik dan memberikan ketahanan, memungkinkan kinerja tanpa gangguan, bahkan jika terjadi pemadaman listrik, downtime Internet, banjir, dan bencana alam lainnya. Lihat AWS Global Cloud Infrastructure untuk mengetahui cara AWS melakukan hal ini.

Dengan menerapkan di beberapa Availability Zone dalam satu Wilayah AWS, beban kerja Anda lebih terlindungi dari kegagalan satu (atau bahkan beberapa) pusat data. Untuk jaminan tambahan dengan penerapan Wilayah Tunggal, Anda dapat mencadangkan data dan konfigurasi (termasuk definisi infrastruktur) ke Wilayah lain. Strategi ini mengurangi ruang lingkup rencana pemulihan bencana Anda untuk hanya menyertakan pencadangan dan pemulihan data. Memanfaatkan ketahanan multi-wilayah dengan membuat cadangan ke Wilayah AWS lain sederhana dan murah dibandingkan dengan opsi Multi-wilayah lainnya yang dijelaskan di bagian berikut. Misalnya, mencadangkan ke Amazon Simple Storage Service (Amazon S3) Simple Storage Service (Amazon S3) memberi Anda akses ke pengambilan data secara langsung. Namun jika strategi DR Anda untuk sebagian data Anda memiliki persyaratan yang lebih santai untuk waktu pengambilan (dari menit hingga jam), maka menggunakan Amazon S3 Glacier atau Amazon S3 Glacier Deep Archive akan secara signifikan mengurangi biaya strategi pencadangan dan pemulihan Anda.

Beberapa beban kerja mungkin memiliki persyaratan residensi data peraturan. Jika ini berlaku untuk beban kerja Anda di lokasi yang saat ini hanya memiliki satu Wilayah AWS, selain merancang beban kerja Multi-AZ untuk ketersediaan tinggi seperti yang dibahas di atas, Anda juga dapat menggunakan wilayah tersebut sebagai lokasi terpisah, yang dapat membantu untuk menangani persyaratan residensi data yang berlaku untuk beban kerja Anda di Wilayah tersebut. AZs Strategi DR yang dijelaskan di bagian berikut menggunakan beberapa Wilayah AWS, tetapi juga dapat diimplementasikan menggunakan Availability Zone, bukan Wilayah.

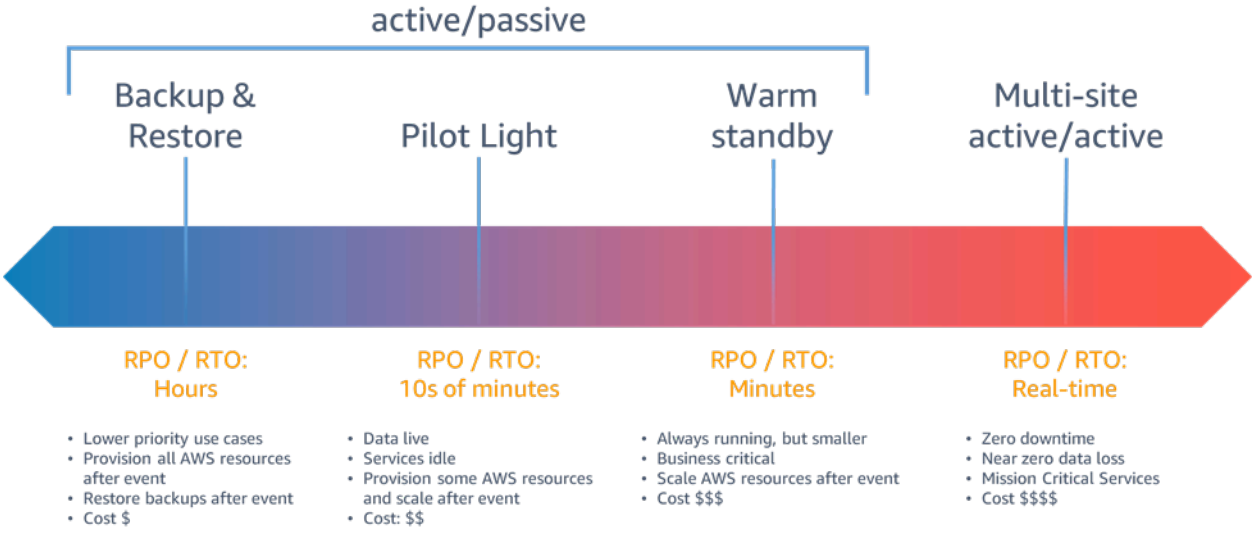
Beberapa Wilayah AWS

Untuk peristiwa bencana yang mencakup risiko kehilangan beberapa pusat data dalam jarak yang cukup jauh satu sama lain, Anda harus mempertimbangkan opsi pemulihan bencana untuk mengurangi bencana alam dan teknis yang memengaruhi seluruh Wilayah dalam AWS. Semua opsi yang dijelaskan dalam bagian berikut dapat diimplementasikan sebagai arsitektur Multi-wilayah untuk melindungi dari bencana tersebut.

Opsi pemulihan bencana di cloud

Strategi pemulihan bencana yang tersedia untuk Anda dalam AWS dapat dikategorikan secara luas menjadi empat pendekatan, mulai dari biaya rendah dan kompleksitas rendah dalam membuat cadangan hingga strategi yang lebih kompleks menggunakan beberapa Wilayah aktif. Active/passive strategi menggunakan situs aktif (seperti Wilayah AWS) untuk menampung beban kerja dan melayani lalu lintas. Situs pasif (seperti Wilayah AWS yang berbeda) digunakan untuk pemulihan. Situs pasif tidak aktif melayani lalu lintas sampai peristiwa failover dipicu.

Sangat penting untuk secara teratur menilai dan menguji strategi pemulihan bencana Anda sehingga Anda memiliki keyakinan dalam menerapkannya, jika diperlukan. Gunakan AWS Resilience Hub untuk terus memvalidasi dan melacak ketahanan AWS beban kerja Anda, termasuk apakah Anda mungkin memenuhi target RTO dan RPO Anda.



Gambar 6 - Strategi pemulihan bencana

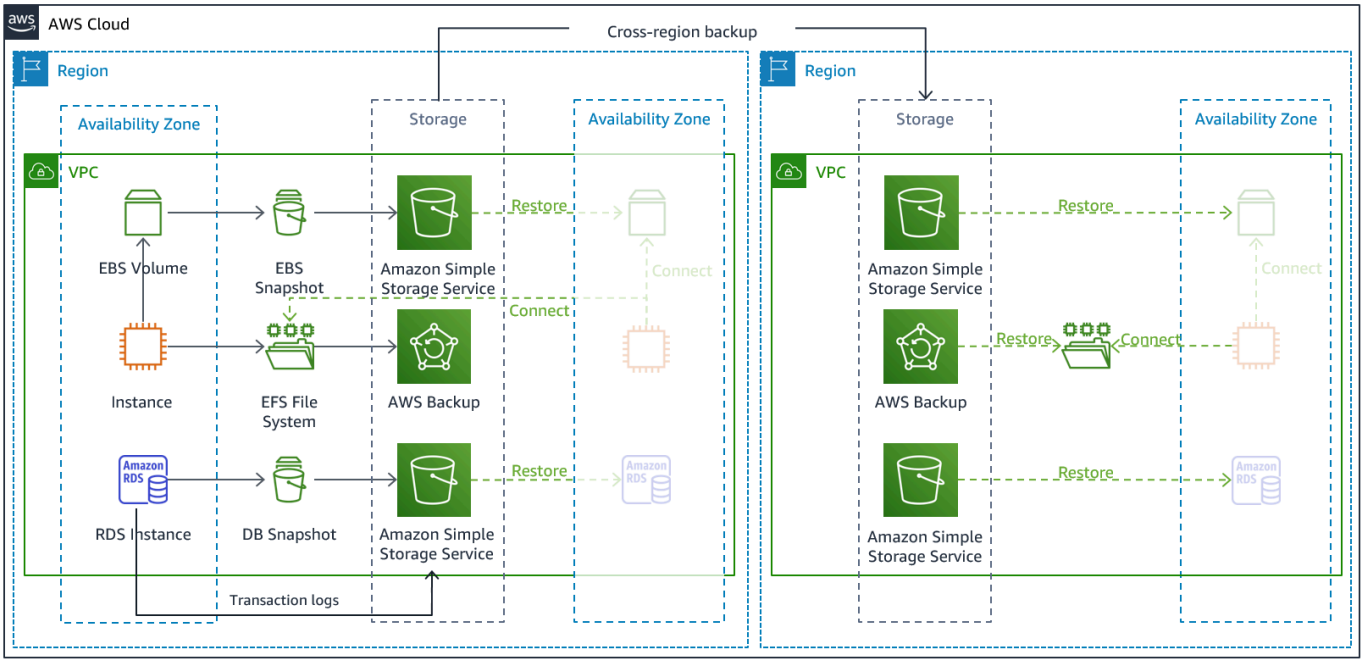
Untuk peristiwa bencana berdasarkan gangguan atau hilangnya satu pusat data fisik untuk beban kerja yang dirancang dengan baik dan sangat tersedia, Anda mungkin hanya memerlukan pendekatan cadangan dan pemulihan untuk pemulihan bencana. Jika definisi Anda tentang bencana melampaui gangguan atau hilangnya pusat data fisik ke Wilayah atau jika Anda tunduk pada persyaratan peraturan yang memerlukannya, maka Anda harus mempertimbangkan Pilot Light, Warm Standby, atau Multi-Site Active/Active.

Saat memilih strategi Anda, dan sumber daya AWS untuk mengimplementasikannya, ingatlah bahwa dalam AWS, kami biasanya membagi layanan ke dalam bidang data dan bidang kontrol. Bidang data

bertanggung jawab untuk menghadirkan layanan waktu nyata sedangkan bidang kontrol digunakan untuk mengonfigurasi lingkungan. Untuk ketahanan maksimum, Anda harus menggunakan hanya operasi pesawat data sebagai bagian dari operasi failover Anda. Ini karena pesawat data biasanya memiliki tujuan desain ketersediaan yang lebih tinggi daripada bidang kontrol.

Pencadangan dan pemulihan

Backup and restore adalah pendekatan yang cocok untuk mengurangi kehilangan data atau korupsi. Pendekatan ini juga dapat digunakan untuk mengurangi bencana regional dengan mereplikasi data ke Wilayah AWS lainnya, atau untuk mengurangi kurangnya redundansi untuk beban kerja yang diterapkan ke satu Availability Zone. Selain data, Anda harus menerapkan ulang infrastruktur, konfigurasi, dan kode aplikasi di Wilayah pemulihan. Agar infrastruktur dapat digunakan kembali dengan cepat tanpa kesalahan, Anda harus selalu menerapkan menggunakan infrastruktur sebagai kode (IAc) menggunakan layanan seperti atau. AWS CloudFormationAWS Cloud Development Kit (AWS CDK) Tanpa IAc, mungkin rumit untuk memulihkan beban kerja di Wilayah pemulihan, yang akan menyebabkan peningkatan waktu pemulihan dan mungkin melebihi RTO Anda. Selain data pengguna, pastikan juga mencadangkan kode dan konfigurasi, termasuk Amazon Machine Images (AMIs) yang Anda gunakan untuk membuat EC2 instance Amazon. Anda dapat menggunakan AWS CodePipelineuntuk mengotomatiskan redeployment kode aplikasi dan konfigurasi.



Gambar 7 - Backup dan Restore Arsitektur

Layanan AWS

Data beban kerja Anda akan memerlukan strategi cadangan yang berjalan secara berkala atau berkelanjutan. Seberapa sering Anda menjalankan cadangan Anda akan menentukan titik pemulihan yang dapat dicapai (yang harus selaras untuk memenuhi RPO Anda). Cadangan juga harus menawarkan cara untuk mengembalikannya ke titik waktu di mana ia diambil. Backup dengan point-in-time pemulihan tersedia melalui layanan dan sumber daya berikut:

- Cuplikan Amazon Elastic Block Store (Amazon EBS)
- Cadangan Amazon DynamoDB
- Cuplikan Amazon RDS
- Cuplikan Amazon Aurora DB
- Cadangan Amazon EFS (saat menggunakan AWS Backup)
- Cuplikan Amazon Redshift
- Cuplikan Amazon Neptunus
- Amazon DocumentDB
- Amazon FSx untuk Windows File Server, Amazon FSx untuk Lustre, Amazon untuk NetApp ONTAP, dan Amazon FSx untuk OpenZFS FSx

Untuk Amazon Simple Storage Service (Amazon S3), Anda dapat menggunakan Amazon S3 Cross-Region Replication (CRR) untuk menyalin objek secara asinkron ke bucket S3 di wilayah DR secara terus menerus, sambil memberikan versi untuk objek yang disimpan sehingga Anda dapat memilih titik restorasi. Replikasi data yang berkelanjutan memiliki keuntungan menjadi waktu terpendek (mendekati nol) untuk mencadangkan data Anda, tetapi mungkin tidak melindungi terhadap peristiwa bencana seperti korupsi data atau serangan berbahaya (seperti penghapusan data yang tidak sah) serta pencadangan. point-in-time Replikasi berkelanjutan tercakup dalam bagian AWS Services for Pilot Light.

AWS Backupmenyediakan lokasi terpusat untuk mengonfigurasi, menjadwalkan, dan memantau kemampuan pencadangan AWS untuk layanan dan sumber daya berikut:

- Volume Amazon Elastic Block Store (Amazon EBS)
- EC2Contoh Amazon
- Basis data Amazon Relational Database Service (Amazon RDS) (termasuk database Amazon Aurora)

- Tabel Amazon DynamoDB
- Sistem file Amazon Elastic File System (Amazon EFS)
- Volume AWS Storage Gateway
- Amazon FSx untuk Windows File Server, Amazon FSx untuk Lustre, Amazon untuk NetApp ONTAP, dan Amazon FSx untuk OpenZFS FSx

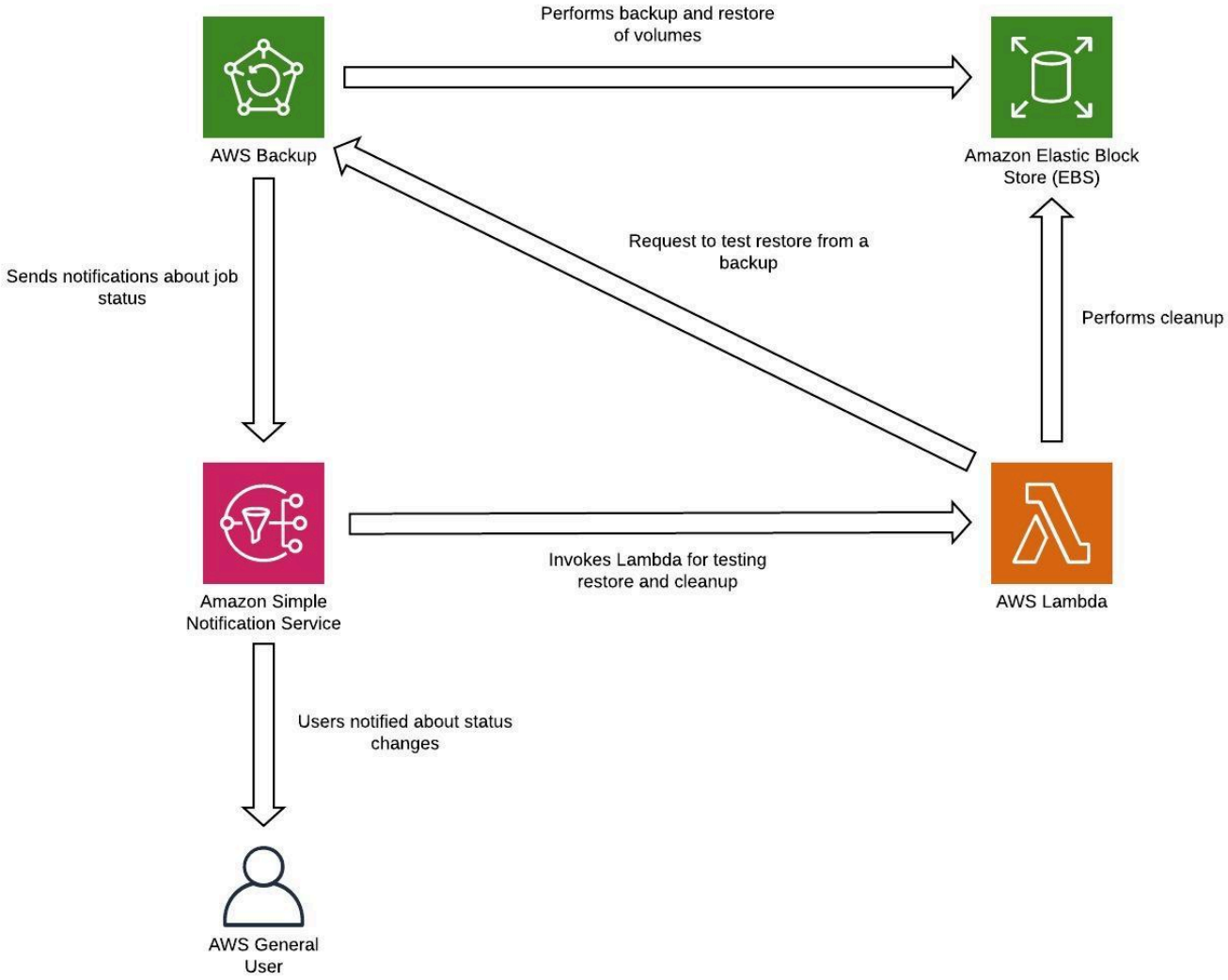
AWS Backup mendukung penyalinan cadangan di seluruh Wilayah, seperti ke Wilayah pemulihan bencana.

Sebagai strategi pemulihan bencana tambahan untuk data Amazon S3 Anda, aktifkan versi objek S3. Pembuatan versi objek melindungi data Anda di S3 dari konsekuensi tindakan penghapusan atau modifikasi dengan mempertahankan versi asli sebelum tindakan. Pembuatan versi objek dapat menjadi mitigasi yang berguna untuk bencana jenis kesalahan manusia. Jika Anda menggunakan replikasi S3 untuk mencadangkan data ke wilayah DR Anda, maka, secara default, saat objek dihapus di bucket sumber, Amazon S3 hanya menambahkan penanda hapus di bucket sumber. Pendekatan ini melindungi data di Wilayah DR dari penghapusan berbahaya di Wilayah sumber.

Selain data, Anda juga harus mencadangkan konfigurasi dan infrastruktur yang diperlukan untuk memindahkan beban kerja Anda dan memenuhi Tujuan Waktu Pemulihan (RTO) Anda. AWS CloudFormationmenyediakan Infrastructure as Code (IAc), dan memungkinkan Anda menentukan semua sumber daya AWS dalam beban kerja Anda sehingga Anda dapat menerapkan dan menerapkan ulang dengan andal ke beberapa akun AWS dan Wilayah AWS. Anda dapat mencadangkan EC2 instans Amazon yang digunakan oleh beban kerja Anda sebagai Amazon Machine Images ()AMIs. AMI dibuat dari snapshot volume root instans Anda dan volume EBS lainnya yang dilampirkan ke instans Anda. Anda dapat menggunakan AMI ini untuk meluncurkan versi EC2 instans yang dipulihkan. AMI dapat disalin di dalam atau di seluruh Wilayah. Atau, Anda dapat menggunakannya AWS Backupuntuk menyalin cadangan di seluruh akun dan ke Wilayah AWS lainnya. Kemampuan pencadangan lintas akun membantu melindungi dari peristiwa bencana yang mencakup ancaman orang dalam atau kompromi akun. AWS Backup juga menambahkan kemampuan tambahan untuk EC2 pencadangan — selain volume EBS individual instans, AWS Backup juga menyimpan dan melacak metadata berikut: jenis instans, cloud pribadi virtual (VPC) yang dikonfigurasi, grup keamanan, peran IAM, konfigurasi pemantauan, dan tag. Namun, metadata tambahan ini hanya digunakan saat memulihkan EC2 cadangan ke Wilayah AWS yang sama.

Setiap data yang disimpan di Wilayah pemulihan bencana sebagai cadangan harus dipulihkan pada saat failover. AWS Backup menawarkan kemampuan pemulihan, tetapi saat ini tidak mengaktifkan pemulihan terjadwal atau otomatis. Anda dapat menerapkan pemulihan otomatis ke wilayah DR

menggunakan AWS SDK APIs untuk AWS Backup dipanggil. Anda dapat mengatur ini sebagai pekerjaan berulang secara teratur atau memicu pemulihan setiap kali cadangan selesai. Gambar berikut menunjukkan contoh pemulihan otomatis menggunakan Amazon Simple Notification Service (Amazon AWS LambdaSNS) dan. Menerapkan pemulihan data berkala terjadwal adalah ide yang baik karena pemulihan data dari cadangan adalah operasi bidang kontrol. Jika operasi ini tidak tersedia selama bencana, Anda masih akan memiliki penyimpanan data yang dapat dioperasikan yang dibuat dari cadangan baru-baru ini.

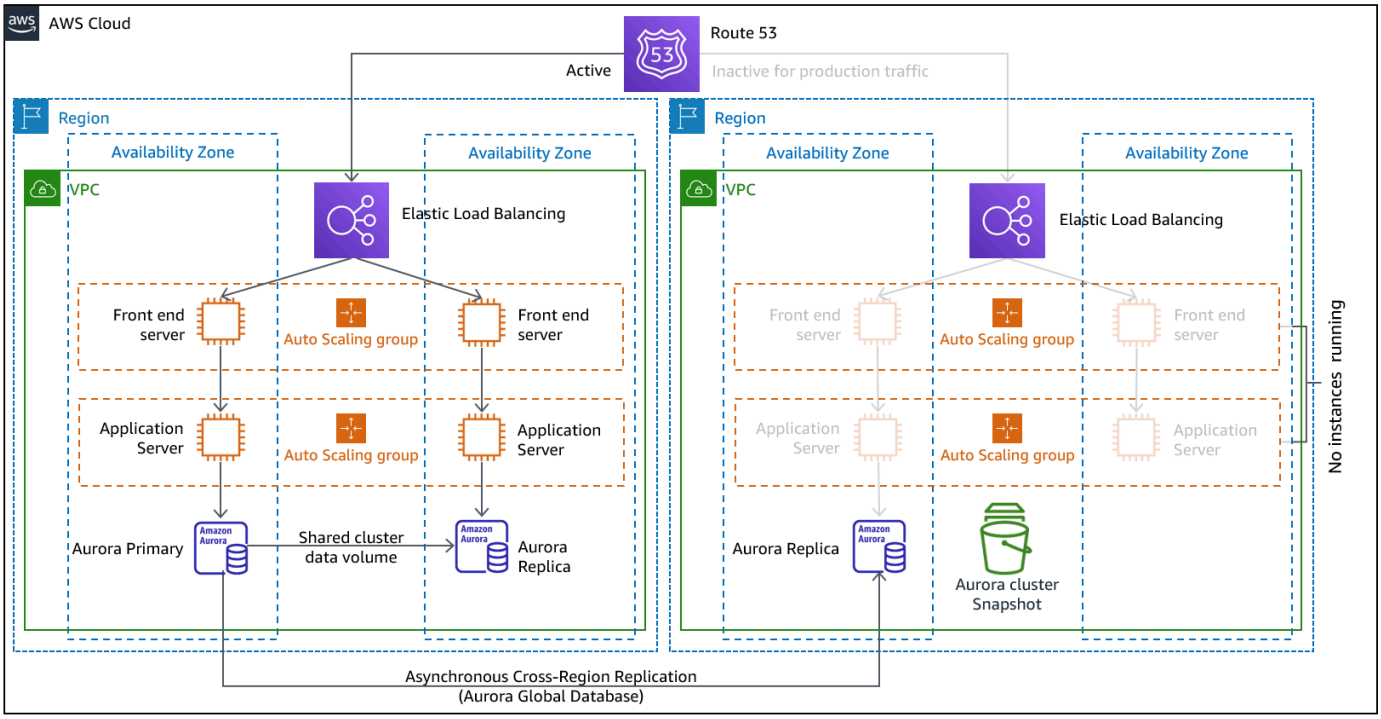


Gambar 8 - Memulihkan dan menguji cadangan

> **Note**
>
> Strategi pencadangan Anda harus mencakup pengujian cadangan Anda. Lihat bagian Menguji Pemulihan Bencana untuk informasi selengkapnya. Lihat AWS Well-Architected Lab: Menguji Backup dan Restore Data untuk demonstrasi implementasi langsung.

Pilot light

Dengan pendekatan pilot light, Anda mereplikasi data Anda dari satu Wilayah ke Wilayah lain dan menyediakan salinan infrastruktur beban kerja inti Anda. Sumber daya yang diperlukan untuk mendukung replikasi dan pencadangan data, misalnya basis data dan penyimpanan objek, selalu aktif. Elemen lain, seperti server aplikasi, dimuat dengan kode aplikasi dan konfigurasi, tetapi "dimatikan" dan hanya digunakan selama pengujian atau ketika failover pemulihan bencana dipanggil. Di cloud, Anda memiliki fleksibilitas untuk mengurangi sumber daya saat Anda tidak membutuhkannya, dan menyediakannya saat Anda melakukannya. Praktik terbaik untuk "dimatikan" adalah tidak menyebarkan sumber daya, dan kemudian membuat konfigurasi dan kemampuan untuk menerapkannya ("aktifkan") bila diperlukan. Berbeda dengan pendekatan pencadangan dan pemulihan, infrastruktur inti Anda selalu tersedia dan Anda selalu memiliki opsi untuk menyediakan lingkungan produksi skala penuh dengan mengaktifkan dan meningkatkan skala server aplikasi Anda.



Gambar 9 - Arsitektur cahaya pilot

Pendekatan pilot light meminimalkan biaya pemulihan bencana yang sedang berlangsung dengan meminimalkan sumber daya aktif, dan menyederhanakan pemulihan pada saat bencana karena persyaratan infrastruktur inti semuanya ada. Opsi pemulihan ini mengharuskan Anda untuk mengubah pendekatan penerapan Anda. Anda perlu membuat perubahan infrastruktur inti ke setiap Wilayah dan menerapkan perubahan beban kerja (konfigurasi, kode) secara bersamaan ke setiap Wilayah. Langkah ini dapat disederhanakan dengan mengotomatiskan penerapan Anda dan menggunakan infrastruktur sebagai kode (IAc) untuk menyebarkan infrastruktur di beberapa akun dan Wilayah (penyebaran infrastruktur penuh ke Wilayah utama dan penyebaran infrastruktur yang diperkecilkan/dimatikan ke wilayah DR). Disarankan Anda menggunakan akun yang berbeda per Wilayah untuk menyediakan isolasi sumber daya dan keamanan tingkat tertinggi (dalam hal kredensi yang dikompromikan juga merupakan bagian dari rencana pemulihan bencana Anda).

Dengan pendekatan ini, Anda juga harus mengurangi bencana data. Replikasi data berkelanjutan melindungi Anda dari beberapa jenis bencana, tetapi mungkin tidak melindungi Anda dari korupsi atau penghancuran data kecuali strategi Anda juga mencakup versi data yang disimpan atau opsi untuk pemulihan. point-in-time Anda dapat mencadangkan data yang direplikasi di Wilayah bencana untuk membuat point-in-time cadangan di Wilayah yang sama.

Layanan AWS

Selain menggunakan layanan AWS yang tercakup dalam bagian Backup dan Restore untuk membuat point-in-time cadangan, pertimbangkan juga layanan berikut untuk strategi pilot light Anda.

Untuk pilot light, replikasi data berkelanjutan ke database langsung dan penyimpanan data di wilayah DR adalah pendekatan terbaik untuk RPO rendah (bila digunakan selain point-in-time cadangan yang dibahas sebelumnya). AWS menyediakan replikasi data asinkron yang berkelanjutan, lintas wilayah, dan asinkron untuk data menggunakan layanan dan sumber daya berikut:

- Replikasi Amazon Simple Storage Service (Amazon S3)
- Amazon RDS baca replika
- Basis data global Amazon Aurora
- Tabel global Amazon DynamoDB
- Cluster global Amazon DocumentDB
- Datastore Global untuk Amazon ElastiCache (Redis OSS)

Dengan replikasi berkelanjutan, versi data Anda segera tersedia di Wilayah DR Anda. Waktu replikasi aktual dapat dipantau menggunakan fitur layanan seperti S3 Replication Time Control (S3 RTC) untuk objek S3 dan fitur manajemen database global Amazon Aurora.

Ketika gagal menjalankan read/write beban kerja Anda dari Wilayah pemulihan bencana, Anda harus mempromosikan replika baca RDS untuk menjadi contoh utama. Untuk instans DB selain Aurora, prosesnya membutuhkan beberapa menit untuk diselesaikan dan reboot adalah bagian dari proses. Untuk Replikasi Lintas Wilayah (CRR) dan failover dengan RDS, menggunakan database global Amazon Aurora memberikan beberapa keuntungan. Database global menggunakan infrastruktur khusus yang membuat database Anda sepenuhnya tersedia untuk melayani aplikasi Anda, dan dapat mereplikasi ke Wilayah sekunder dengan latensi tipikal kurang dari satu detik (dan dalam Wilayah AWS kurang dari 100 milidetik). Dengan database global Amazon Aurora, jika Wilayah utama Anda mengalami penurunan kinerja atau pemadaman, Anda dapat mempromosikan salah satu wilayah sekunder untuk mengambil tanggung jawab baca/tulis dalam waktu kurang dari satu menit bahkan jika terjadi pemadaman regional total. Anda juga dapat mengonfigurasi Aurora untuk memantau jeda waktu RPO dari semua cluster sekunder untuk memastikan bahwa setidaknya satu cluster sekunder tetap berada dalam jendela RPO target Anda.

Versi infrastruktur beban kerja inti Anda yang diperkecil dengan sumber daya yang lebih sedikit atau lebih kecil harus diterapkan di Wilayah DR Anda. Dengan menggunakan AWS CloudFormation,

Anda dapat menentukan infrastruktur dan menerapkannya secara konsisten di seluruh akun AWS dan di seluruh Wilayah AWS. AWS CloudFormation menggunakan parameter semu yang telah ditentukan sebelumnya untuk mengidentifikasi akun AWS dan Wilayah AWS tempat akun tersebut digunakan. Oleh karena itu, Anda dapat menerapkan logika kondisi di CloudFormation template Anda untuk menerapkan hanya versi infrastruktur yang diperkecil di Wilayah DR. EC2 Misalnya penerapan, Amazon Machine Image (AMI) menyediakan informasi seperti konfigurasi perangkat keras dan perangkat lunak yang diinstal. Anda dapat mengimplementasikan pipeline Image Builder yang membuat kebutuhan AMIs Anda dan menyalinnya ke Wilayah utama dan cadangan Anda. Ini membantu memastikan bahwa emas ini AMIs memiliki semua yang Anda butuhkan untuk menyebarkan kembali atau meningkatkan beban kerja Anda di wilayah baru, jika terjadi peristiwa bencana. EC2 Instans Amazon diterapkan dalam konfigurasi yang diperkecil (lebih sedikit instance daripada di Wilayah utama Anda). Untuk meningkatkan skala infrastruktur guna mendukung lalu lintas produksi, lihat Amazon Auto EC2 Scaling di bagian Warm Standby.

Untuk active/passive konfigurasi seperti lampu pilot, semua lalu lintas awalnya pergi ke Wilayah utama dan beralih ke Wilayah pemulihan bencana jika Wilayah utama tidak lagi tersedia. Operasi failover ini dapat dimulai secara otomatis dan manual. Failover yang dimulai secara otomatis berdasarkan pemeriksaan kesehatan atau alarm harus digunakan dengan hati-hati. Bahkan menggunakan praktik terbaik yang dibahas di sini, waktu pemulihan dan titik pemulihan akan lebih besar dari nol, menimbulkan beberapa kehilangan ketersediaan dan data. Jika Anda gagal ketika Anda tidak perlu (alarm palsu), maka Anda mengalami kerugian tersebut. Oleh karena itu, Failover yang dimulai secara manual sering digunakan. Dalam kasus ini, Anda masih harus mengotomatiskan langkah failover, sehingga inisiasi manual akan seperti menekan tombol.

Ada beberapa opsi manajemen lalu lintas yang perlu dipertimbangkan saat menggunakan AWS layanan.

Salah satu opsinya adalah menggunakan Amazon Route 53. Menggunakan Amazon Route 53, Anda dapat mengaitkan beberapa titik akhir IP di satu atau beberapa Wilayah AWS dengan nama domain Route 53. Kemudian, Anda dapat merutekan lalu lintas ke titik akhir yang sesuai di bawah nama domain tersebut. Pada failover Anda perlu mengalihkan lalu lintas ke titik akhir pemulihan, dan menjauh dari titik akhir utama. Pemeriksaan kesehatan Amazon Route 53 memantau titik akhir ini. Dengan menggunakan pemeriksaan kesehatan ini, Anda dapat mengonfigurasi failover DNS yang dimulai secara otomatis untuk memastikan lalu lintas dikirim hanya ke titik akhir yang sehat, yang merupakan operasi yang sangat andal yang dilakukan pada bidang data. Untuk menerapkan ini menggunakan failover yang dimulai secara manual, Anda dapat menggunakan Amazon Application Recovery Controller (ARC). Dengan ARC, Anda dapat membuat pemeriksaan kesehatan Route 53 yang tidak benar-benar memeriksa kesehatan, tetapi bertindak sebagai sakelar on/off yang

memiliki kendali penuh. Menggunakan AWS CLI atau AWS SDK, Anda dapat membuat skrip failover menggunakan API bidang data yang sangat tersedia ini. Skrip Anda mengaktifkan sakelar ini (pemeriksaan kesehatan Route 53) memberi tahu Route 53 untuk mengirim lalu lintas ke Wilayah pemulihan alih-alih Wilayah utama. Pilihan lain untuk failover yang dimulai secara manual yang telah digunakan beberapa orang adalah dengan menggunakan kebijakan perutean tertimbang dan mengubah bobot Wilayah primer dan pemulihan sehingga semua lalu lintas masuk ke Wilayah pemulihan. Namun, ketahuilah bahwa ini adalah operasi bidang kontrol dan oleh karena itu tidak sekuat pendekatan bidang data menggunakan Amazon Application Recovery Controller (ARC).

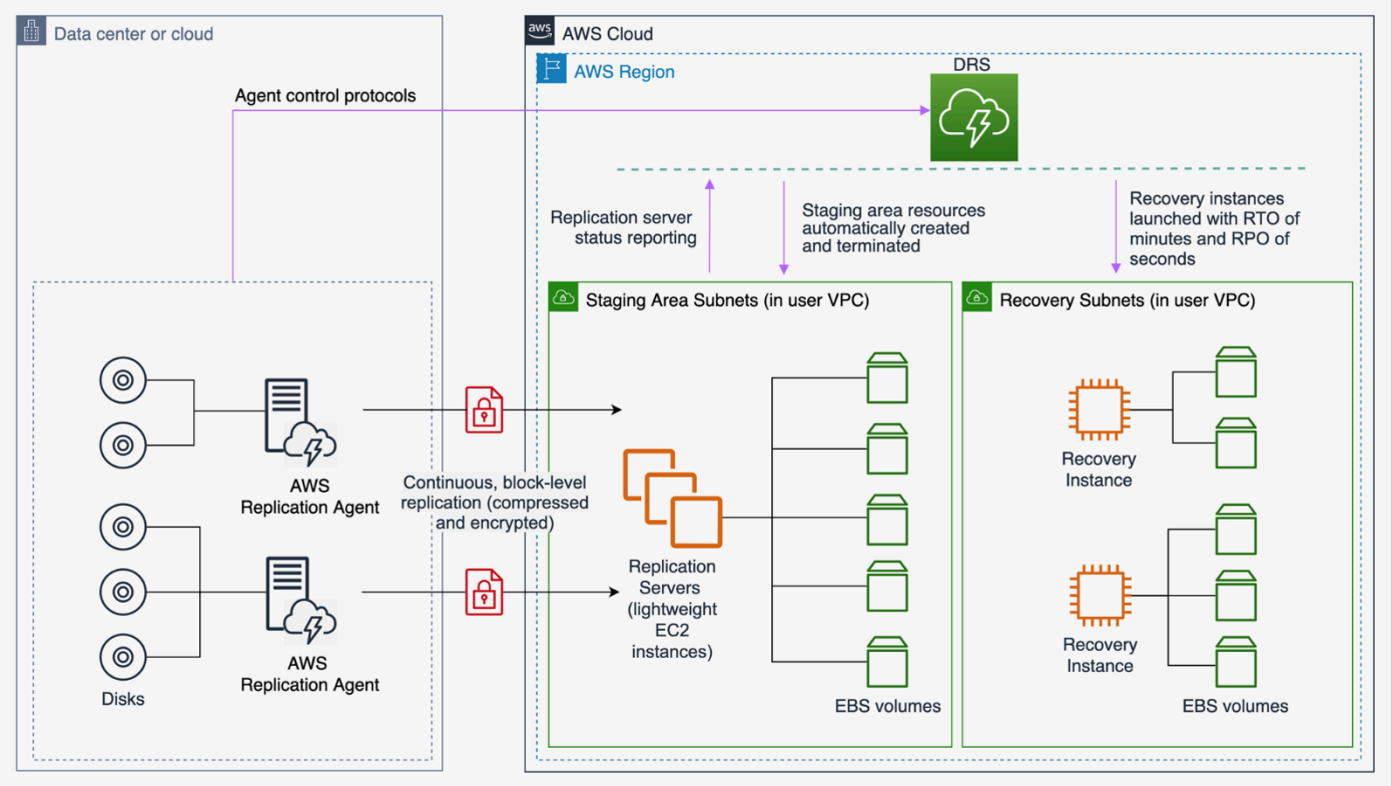
Pilihan lain adalah menggunakan AWS Global Accelerator. Dengan menggunakan AnyCast IP, Anda dapat mengaitkan beberapa titik akhir di satu atau beberapa Wilayah AWS dengan alamat atau alamat IP publik statis yang sama. AWS Global Accelerator kemudian mengarahkan lalu lintas ke titik akhir yang sesuai yang terkait dengan alamat itu. Pemeriksaan kesehatan Global Accelerator memantau titik akhir. Dengan menggunakan pemeriksaan kesehatan ini, AWS Global Accelerator memeriksa kesehatan aplikasi Anda dan mengarahkan lalu lintas pengguna secara otomatis ke titik akhir aplikasi yang sehat. Untuk failover yang dimulai secara manual, Anda dapat menyesuaikan titik akhir mana yang menerima lalu lintas menggunakan tombol lalu lintas, tetapi perhatikan bahwa ini adalah operasi pesawat kontrol. Global Accelerator menawarkan latensi yang lebih rendah ke titik akhir aplikasi karena menggunakan jaringan AWS edge yang luas untuk menempatkan lalu lintas di tulang punggung jaringan AWS sesegera mungkin. Global Accelerator juga menghindari masalah caching yang dapat terjadi dengan sistem DNS (seperti Route 53).

Amazon CloudFront menawarkan failover asal, di mana jika permintaan yang diberikan ke titik akhir utama gagal, CloudFront merutekan permintaan ke titik akhir sekunder. Tidak seperti operasi failover yang dijelaskan sebelumnya, semua permintaan berikutnya masih masuk ke titik akhir utama, dan failover dilakukan per setiap permintaan.

AWS Pemulihan Bencana Elastis

AWS Elastic Disaster Recovery (DRS) terus mereplikasi aplikasi yang dihosting server dan database yang dihosting server dari sumber mana pun menjadi AWS menggunakan replikasi tingkat blok dari server yang mendasarinya. Elastic Disaster Recovery memungkinkan Anda menggunakan Wilayah AWS Cloud sebagai target pemulihan bencana untuk beban kerja yang dihosting di tempat atau di penyedia cloud lain, dan lingkungannya. Ini juga dapat digunakan untuk pemulihan bencana beban kerja yang AWS dihosting jika hanya terdiri dari aplikasi dan database yang dihosting EC2 (yaitu, bukan RDS). Elastic Disaster Recovery menggunakan strategi Pilot Light, memelihara salinan data dan sumber daya “dimatikan” di Amazon Virtual Private Cloud (Amazon VPC) yang digunakan sebagai area pementasan. Saat peristiwa failover dipicu, sumber daya bertahap digunakan untuk

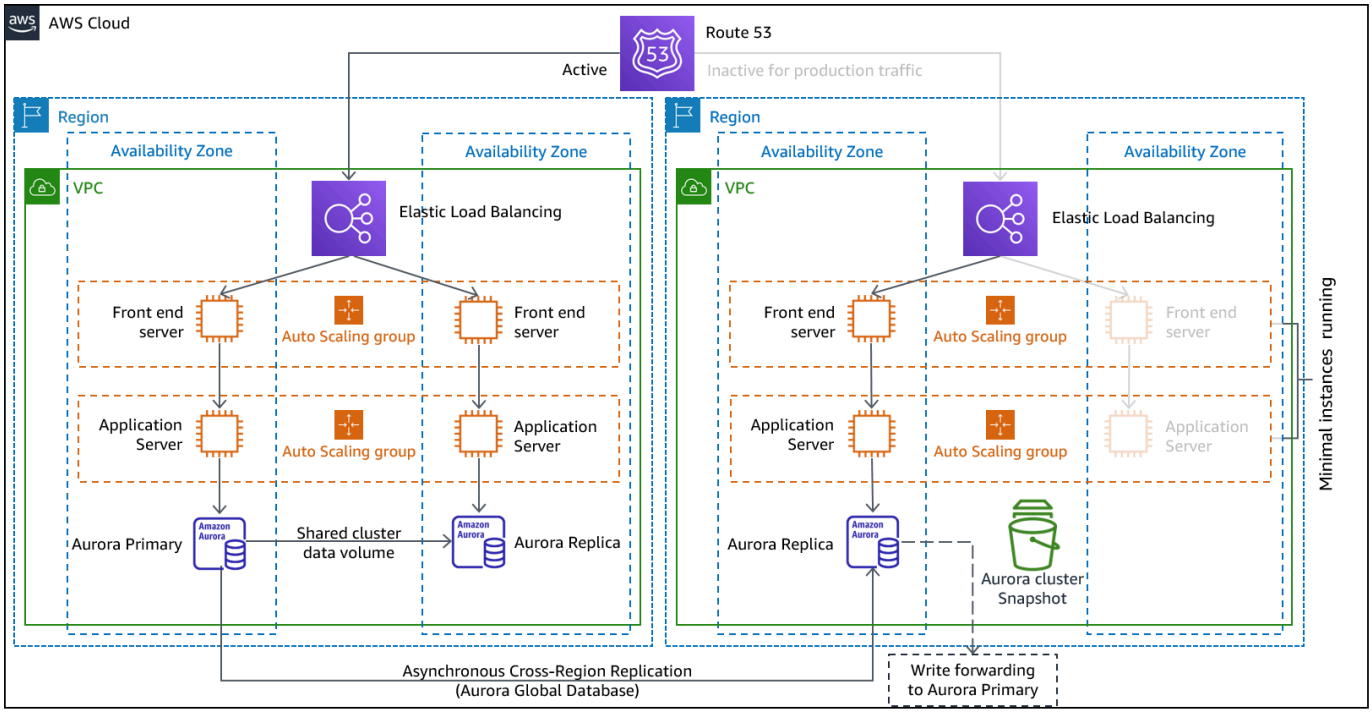
secara otomatis membuat penerapan kapasitas penuh di VPC Amazon target yang digunakan sebagai lokasi pemulihan.



Gambar 10 - Arsitektur Pemulihan Bencana AWS Elastis

Warm standby

Pendekatan warm standby melibatkan memastikan ada salinan lingkungan produksi yang skalanya diturunkan tetapi berfungsi sepenuhnya di Wilayah lainnya. Pendekatan ini memperpanjang konsep pilot light dan mempercepat waktu pemulihan karena beban kerja selalu aktif di Wilayah lainnya. Pendekatan ini juga memungkinkan Anda untuk lebih mudah melakukan pengujian atau menerapkan pengujian berkelanjutan untuk meningkatkan kepercayaan pada kemampuan Anda untuk pulih dari bencana.



Gambar 11 - Arsitektur siaga hangat

Catatan: Perbedaan antara lampu pilot dan siaga hangat terkadang sulit dipahami. Keduanya mencakup lingkungan di Wilayah DR Anda dengan salinan aset Wilayah utama Anda. Perbedaannya adalah bahwa lampu pilot tidak dapat memproses permintaan tanpa tindakan tambahan yang diambil terlebih dahulu, sedangkan siaga hangat dapat menangani lalu lintas (pada tingkat kapasitas yang dikurangi) dengan segera. Pendekatan pilot light mengharuskan Anda untuk “menghidupkan” server, mungkin menerapkan infrastruktur tambahan (non-inti), dan meningkatkan skala, sedangkan siaga hangat hanya mengharuskan Anda untuk meningkatkan (semuanya sudah digunakan dan berjalan). Gunakan kebutuhan RTO dan RPO Anda untuk membantu Anda memilih di antara pendekatan ini.

Layanan AWS

Semua layanan AWS yang tercakup dalam pencadangan dan pemulihan dan lampu pilot juga digunakan dalam keadaan siaga hangat untuk pencadangan data, replikasi data, perutean active/ passive lalu lintas, dan penyebaran infrastruktur termasuk instance. EC2

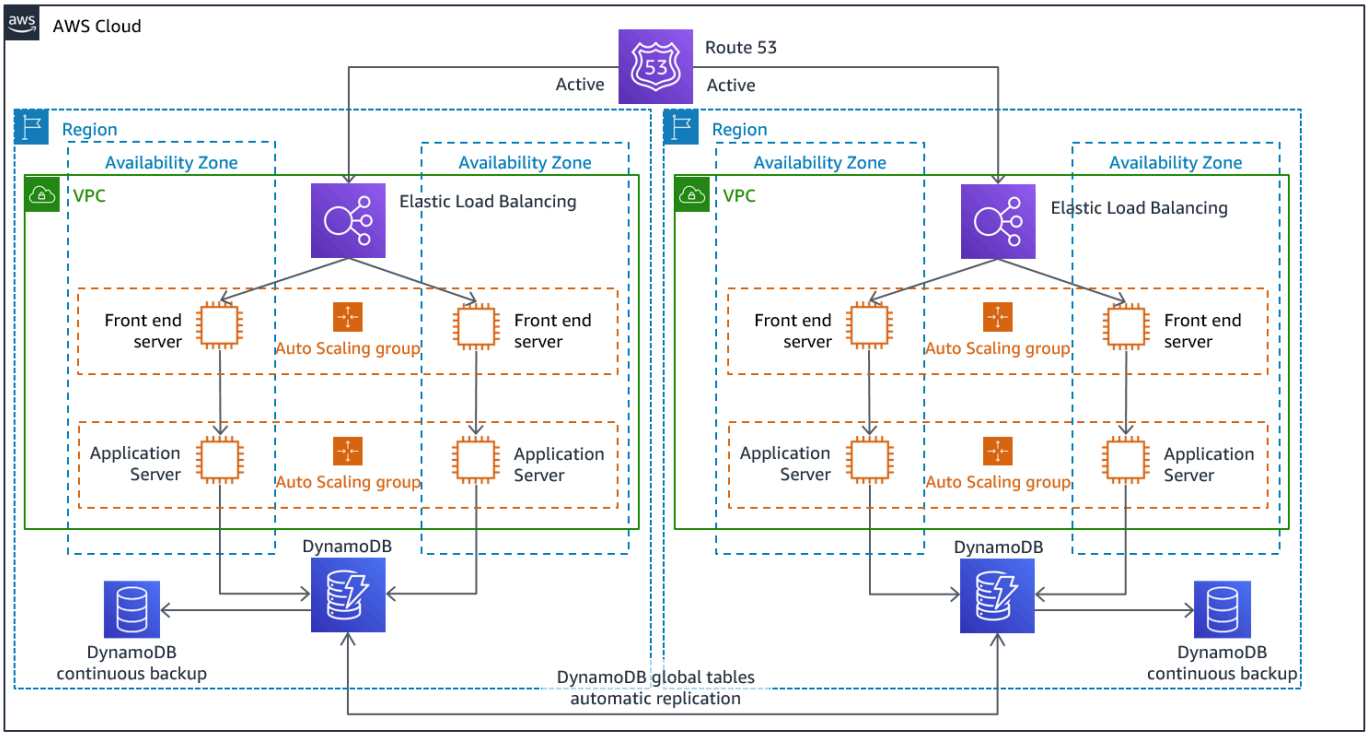
EC2 Auto Scaling Amazon digunakan untuk menskalakan sumber daya termasuk EC2 instans Amazon, tugas Amazon ECS, throughput Amazon DynamoDB, dan replika Amazon Aurora dalam Wilayah AWS. Amazon EC2 Auto Scaling menskalakan penyebaran EC2 instans di seluruh Availability Zone dalam Wilayah AWS, memberikan ketahanan dalam Wilayah tersebut. Gunakan

Auto Scaling untuk meningkatkan skala Wilayah DR Anda ke kemampuan produksi penuh, sebagai bagian dari strategi lampu pilot atau siaga hangat. Misalnya, untuk EC2, tingkatkan pengaturan kapasitas yang diinginkan pada grup Auto Scaling. Anda dapat menyesuaikan pengaturan ini secara manual melalui AWS Management Console, secara otomatis melalui AWS SDK, atau dengan menerapkan ulang AWS CloudFormation template Anda menggunakan nilai kapasitas baru yang diinginkan. Anda dapat menggunakan AWS CloudFormation parameter untuk membuat redeploying CloudFormation template lebih mudah. Pastikan bahwa kuota layanan di Wilayah DR Anda ditetapkan cukup tinggi sehingga tidak membatasi Anda dari peningkatan kapasitas produksi.

Karena Auto Scaling adalah aktivitas bidang kontrol, mengambil ketergantungan padanya akan menurunkan ketahanan strategi pemulihan Anda secara keseluruhan. Ini adalah trade-off. Anda dapat memilih untuk menyediakan kapasitas yang cukup sehingga Wilayah pemulihan dapat menangani beban produksi penuh seperti yang digunakan. Konfigurasi stabil statis ini disebut hot standby (lihat bagian selanjutnya). Atau Anda dapat memilih untuk menyediakan lebih sedikit sumber daya yang biayanya lebih murah, tetapi bergantung pada Auto Scaling. Beberapa implementasi DR akan menggunakan sumber daya yang cukup untuk menangani lalu lintas awal, memastikan RTO rendah, dan kemudian mengandalkan Auto Scaling untuk meningkatkan lalu lintas berikutnya.

Multi-situs aktif/aktif

Anda dapat menjalankan beban kerja Anda secara bersamaan di beberapa Wilayah sebagai bagian dari strategi aktif/aktif aktif atau siaga aktif multi-situs aktif/pasif. Multi-situs active/active melayani lalu lintas dari semua wilayah yang digunakan, sedangkan hot standby melayani lalu lintas hanya dari satu wilayah, dan Wilayah lainnya hanya digunakan untuk pemulihan bencana. Dengan active/active pendekatan multi-situs, pengguna dapat mengakses beban kerja Anda di salah satu Wilayah di mana ia digunakan. Pendekatan ini adalah pendekatan yang paling kompleks dan mahal untuk pemulihan bencana, tetapi dapat mengurangi waktu pemulihan Anda mendekati nol untuk sebagian besar bencana dengan pilihan dan implementasi teknologi yang benar (namun korupsi data mungkin perlu bergantung pada cadangan, yang biasanya menghasilkan titik pemulihan bukan nol). Hot standby menggunakan active/passive konfigurasi di mana pengguna hanya diarahkan ke satu wilayah dan wilayah DR tidak mengambil lalu lintas. Sebagian besar pelanggan menemukan bahwa jika mereka akan berdiri di lingkungan penuh di Wilayah kedua, masuk akal untuk menggunakannya aktif/aktif. Atau, jika Anda tidak ingin menggunakan kedua Wilayah untuk menangani lalu lintas pengguna, maka Warm Standby menawarkan pendekatan yang lebih ekonomis dan operasional kurang kompleks.



Gambar 12 - active/active Arsitektur multi-situs (ubah satu jalur Aktif menjadi Tidak Aktif untuk siaga panas)

Dengan pendekatan multi-situs active/active, because the workload is running in more than one Region, there is no such thing as failover in this scenario. Disaster recovery testing in this case would focus on how the workload reacts to loss of a Region: Is traffic routed away from the failed Region? Can the other Region(s) handle all the traffic? Testing for a data disaster is also required. Backup and recovery are still required and should be tested regularly. It should also be noted that recovery times for a data disaster involving data corruption, deletion, or obfuscation will always be greater than zero and the recovery point will always be at some point before the disaster was discovered. If the additional complexity and cost of a multi-site active/active (atau siaga panas) diperlukan untuk mempertahankan waktu pemulihan mendekati nol, maka upaya tambahan harus dilakukan untuk menjaga keamanan dan untuk mencegah kesalahan manusia untuk mengurangi bencana manusia.

Layanan AWS

Semua layanan AWS yang tercakup dalam pencadangan dan pemulihan, lampu pilot, dan siaga hangat juga digunakan di sini untuk pencadangan point-in-time data, replikasi data, perutean active/active lalu lintas, serta penyebaran dan penskalaan infrastruktur termasuk instance. EC2

Untuk active/passive skenario yang dibahas sebelumnya (Pilot Light dan Warm Standby), Amazon Route 53 dan AWS Global Accelerator dapat digunakan untuk rute lalu lintas jaringan ke wilayah aktif. Untuk active/active strategi di sini, kedua layanan ini juga memungkinkan definisi kebijakan yang menentukan pengguna mana yang pergi ke titik akhir regional aktif mana. Dengan AWS Global Accelerator Anda mengatur panggilan lalu lintas untuk mengontrol persentase lalu lintas yang diarahkan ke setiap titik akhir aplikasi. Amazon Route 53 mendukung pendekatan persentase ini, dan juga beberapa kebijakan lain yang tersedia termasuk kebijakan berbasis geoproximity dan latensi. Global Accelerator secara otomatis memanfaatkan jaringan ekstensif server AWS edge, untuk mengarahkan lalu lintas ke backbone jaringan AWS sesegera mungkin, sehingga latensi permintaan lebih rendah.

Replikasi data asinkron dengan strategi ini memungkinkan RPO mendekati nol. Layanan AWS seperti database global Amazon Aurora menggunakan infrastruktur khusus yang membuat database Anda sepenuhnya tersedia untuk melayani aplikasi Anda, dan dapat mereplikasi hingga lima Wilayah sekunder dengan latensi tipikal kurang dari satu detik. Dengan active/passive strategies, writes occur only to the primary Region. The difference with active/active merancang bagaimana konsistensi data dengan penulisan ke setiap Wilayah aktif ditangani. Merupakan hal yang umum untuk merancang bacaan pengguna untuk dilayani dari Wilayah terdekat dengan mereka, yang dikenal sebagai baca lokal. Dengan menulis, Anda memiliki beberapa opsi:

- Rute strategi global tulis semuanya menulis ke satu Wilayah. Dalam kasus kegagalan Wilayah itu, Wilayah lain akan dipromosikan untuk menerima tulisan. Basis data global Aurora sangat cocok untuk menulis global, karena mendukung sinkronisasi dengan replika baca di seluruh Wilayah, dan Anda dapat mempromosikan salah satu Wilayah sekunder untuk mengambil read/write tanggung jawab dalam waktu kurang dari satu menit. Aurora juga mendukung penerusan tulis, yang memungkinkan cluster sekunder dalam database global Aurora meneruskan pernyataan SQL yang melakukan operasi tulis ke cluster primer.
- Rute strategi lokal tulis menulis ke Wilayah terdekat (seperti membaca). Tabel global Amazon DynamoDB memungkinkan strategi semacam itu, memungkinkan baca dan tulis dari setiap wilayah tabel global Anda digunakan. Tabel global Amazon DynamoDB menggunakan penulis terakhir memenangkan rekonsiliasi antara pembaruan bersamaan.
- Strategi partisi tulis menetapkan penulisan ke Wilayah tertentu berdasarkan kunci partisi (seperti ID pengguna) untuk menghindari konflik penulisan. Replikasi Amazon S3 yang dikonfigurasi dua arah dapat digunakan untuk kasus ini, dan saat ini mendukung replikasi antara dua Wilayah. Saat menerapkan pendekatan ini, pastikan untuk mengaktifkan sinkronisasi modifikasi replika pada bucket A dan B untuk mereplikasi perubahan metadata replika seperti daftar kontrol akses objek (ACLs), tag objek, atau kunci objek pada objek yang direplikasi. Anda juga dapat mengonfigurasi

apakah akan mereplikasi penanda hapus antar bucket di Wilayah aktif atau tidak. Selain replikasi, strategi Anda juga harus menyertakan point-in-time cadangan untuk melindungi terhadap kerusakan data atau peristiwa penghancuran.

AWS CloudFormation adalah alat yang ampuh untuk menegakkan infrastruktur yang diterapkan secara konsisten di antara akun AWS di beberapa Wilayah AWS. AWS CloudFormation StackSetsmemperluas fungsi ini dengan memungkinkan Anda membuat, memperbarui, atau menghapus CloudFormation tumpukan di beberapa akun dan Wilayah dengan satu operasi. Meskipun AWS CloudFormation menggunakan YAMM atau JSON untuk mendefinisikan Infrastruktur sebagai Kode, AWS Cloud Development Kit (AWS CDK)memungkinkan Anda untuk mendefinisikan Infrastruktur sebagai Kode menggunakan bahasa pemrograman yang sudah dikenal. Kode Anda dikonversi CloudFormation yang kemudian digunakan untuk menyebarkan sumber daya di AWS.

Deteksi

Penting untuk mengetahui sesegera mungkin bahwa beban kerja Anda tidak memberikan hasil bisnis yang seharusnya mereka berikan. Dengan cara ini, Anda dapat dengan cepat menyatakan bencana dan pulih dari suatu insiden. Untuk tujuan pemulihan yang agresif, waktu respons ini ditambah dengan informasi yang tepat sangat penting dalam memenuhi tujuan pemulihan. Jika tujuan waktu pemulihan Anda adalah satu jam, maka Anda perlu mendeteksi insiden tersebut, memberi tahu personel yang sesuai, melibatkan proses eskalasi Anda, mengevaluasi informasi (jika ada) pada waktu yang diharapkan untuk pemulihan (tanpa melaksanakan rencana DR), menyatakan bencana dan pulih dalam waktu satu jam.

> **Note**
>
> Jika pemangku kepentingan memutuskan untuk tidak memanggil DR meskipun RTO akan berisiko, maka evaluasi kembali rencana dan tujuan DR. Keputusan untuk tidak menggunakan rencana DR mungkin karena rencana tersebut tidak memadai atau ada kurangnya kepercayaan dalam pelaksanaannya.

Sangat penting untuk memperhitungkan deteksi insiden, pemberitahuan, eskalasi, penemuan, dan deklarasi ke dalam perencanaan dan tujuan Anda untuk memberikan tujuan yang realistis dan dapat dicapai yang memberikan nilai bisnis.

AWS menerbitkan sebagian besar up-to-the-minute informasi kami tentang ketersediaan layanan di Dashboard Service Health. Periksa kapan saja untuk mendapatkan informasi status terkini, atau berlangganan umpan RSS untuk diberitahu tentang gangguan pada setiap layanan individu. Jika Anda mengalami masalah operasional real-time dengan salah satu layanan kami yang tidak ditampilkan di Dashboard Service Health, Anda dapat membuat Permintaan Dukungan.

AWS Health DashboardMemberikan informasi tentang AWS Health peristiwa yang dapat memengaruhi akun Anda. Informasi ini disajikan dalam dua cara: di dasbor yang menampilkan peristiwa terbaru dan mendatang yang diatur berdasarkan kategori, dan dalam log peristiwa lengkap yang menampilkan semua peristiwa dari 90 hari terakhir.

Untuk persyaratan RTO yang paling ketat, Anda dapat menerapkan failover otomatis berdasarkan pemeriksaan kesehatan. Rancang pemeriksaan kesehatan yang mewakili pengalaman pengguna dan berdasarkan Indikator Kinerja Utama. Pemeriksaan kesehatan mendalam melatih fungsionalitas

utama dari beban kerja Anda dan melampaui pemeriksaan detak jantung yang dangkal. Gunakan pemeriksaan kesehatan mendalam berdasarkan beberapa sinyal. Berhati-hatilah dengan pendekatan ini sehingga Anda tidak memicu alarm palsu karena gagal ketika tidak perlu dengan sendirinya menimbulkan risiko ketersediaan.

Menguji pemulihan bencana

Uji implementasi pemulihan bencana untuk memvalidasi implementasi dan secara teratur menguji failover ke Wilayah DR beban kerja Anda untuk memastikan bahwa RTO dan RPO terpenuhi.

Pola yang harus dihindari adalah mengembangkan jalur pemulihan yang jarang dieksekusi. Misalnya, Anda mungkin memiliki penyimpanan data sekunder yang digunakan untuk kueri hanya-baca. Saat Anda menulis ke penyimpanan data dan penyimpanan primer gagal, Anda mungkin ingin melakukan failover ke penyimpanan data sekunder. Jika Anda tidak sering menguji failover ini, Anda mungkin akan mendapati bahwa asumsi Anda tentang kemampuan penyimpanan data sekunder ternyata salah. Kapasitas sekunder, yang mungkin sudah cukup ketika Anda terakhir menguji, mungkin tidak lagi dapat mentolerir beban dalam skenario ini, atau kuota layanan di Wilayah sekunder mungkin tidak cukup.

Pengalaman kami menunjukkan bahwa satu-satunya pemulihan kesalahan yang dapat diterapkan adalah jalur yang sering Anda uji. Inilah alasan mengapa memiliki sejumlah kecil jalur pemulihan adalah yang terbaik.

Anda dapat membuat pola pemulihan dan mengujinya secara rutin. Jika Anda memiliki jalur pemulihan yang kompleks atau kritis, Anda masih perlu menjalankan kegagalan produksi secara teratur untuk memvalidasi bahwa jalur pemulihan berfungsi.

Kelola penyimpangan konfigurasi di Wilayah DR. Pastikan infrastruktur, data, dan konfigurasi Anda sesuai kebutuhan di Wilayah DR. Misalnya, periksa itu AMIs dan kuota layanan. up-to-date

Anda dapat menggunakannya AWS Configuntuk terus memantau dan merekam konfigurasi sumber daya AWS Anda. AWS Config dapat mendeteksi drift dan memicu Otomasi AWS Systems Manager untuk memperbaiki drift dan menaikkan alarm. AWS CloudFormationjuga dapat mendeteksi penyimpangan di tumpukan yang telah Anda terapkan.

Kesimpulan

Pelanggan bertanggung jawab atas ketersediaan aplikasi mereka di cloud. Penting untuk mendefinisikan apa itu bencana dan memiliki rencana pemulihan bencana yang mencerminkan definisi ini dan dampaknya terhadap hasil bisnis. Buat Recovery Time Objective (RTO) dan Recovery Point Objective (RPO) berdasarkan analisis dampak dan penilaian risiko dan kemudian pilih arsitektur yang sesuai untuk mengurangi bencana. Pastikan bahwa deteksi bencana adalah mungkin dan tepat waktu - sangat penting untuk mengetahui kapan tujuan berada dalam risiko. Pastikan Anda memiliki rencana dan validasi rencana dengan pengujian. Rencana pemulihan bencana yang belum divalidasi berisiko tidak dilaksanakan karena kurangnya kepercayaan diri atau kegagalan untuk memenuhi tujuan pemulihan bencana.

Kontributor

Para kontributor untuk dokumen ini antara lain:

- Alex Livingstone, Praktik Operasi Cloud Utama, AWS Enterprise Support
- Seth Eliot, Arsitek Solusi Keandalan Utama, Amazon Web Services

Sumber bacaan lebih lanjut

Untuk mendapatkan informasi tambahan, buka:

- AWS Pusat Arsitektur
- Pilar Keandalan, AWS Well-Architected Framework
- Daftar Periksa Rencana Pemulihan Bencana
- Menerapkan Pemeriksaan Kesehatan
- Arsitektur Disaster Recovery (DR) di AWS, Bagian I: Strategi Pemulihan di Cloud
- Arsitektur Disaster Recovery (DR) di AWS, Bagian II: Backup dan Restore dengan Pemulihan Cepat
- Arsitektur Pemulihan Bencana (DR) di AWS, Bagian III: Pilot Light dan Warm Standby
- Arsitektur Pemulihan Bencana (DR) di AWS, Bagian IV: Multi-situs Aktif/Aktif
- Membuat Mekanisme Pemulihan Bencana Menggunakan Amazon Route 53
- Meminimalkan Ketergantungan dalam Rencana Pemulihan Bencana
- Tangan di Laboratorium Pemulihan AWS Bencana yang Dirancang dengan Baik
- AWS Implementasi Solusi: Arsitektur Aplikasi Multi-Region
- AWS RE: Invent 2018: Pola Arsitektur untuk Aplikasi Aktif-Aktif Multi-Wilayah (09-R2) ARC2

Riwayat dokumen

Untuk mengetahui jika ada perubahan pada laporan resmi ini, Anda dapat berlangganan umpan RSS.

Perubahan	Deskripsi	Tanggal
Pembaruan kecil	Perbaikan bug dan berbagai perubahan kecil.	1 April 2022
Laporan resmi diperbarui	Pembaruan editorial kecil.	Maret 21, 2022
Laporan resmi diperbarui	Menambahkan informasi tentang pesawat data dan bidang kontrol. Menambahk an rincian lebih lanjut tentang cara menerapkan active/pa ssive failover. Mengganti Pemulihan CloudEndure Bencana dengan Pemulihan Bencana AWS Elastis.	Februari 17, 2022
Pembaruan kecil	AWS Well-Architected Tool informasi ditambahkan.	Februari 11, 2022
Publikasi awal	Whitepaper pertama kali diterbitkan.	12 Februari 2021

Pemberitahuan

Pelanggan bertanggung jawab untuk membuat penilaian independen mereka sendiri atas informasi dalam dokumen ini. Dokumen ini: (a) hanya untuk tujuan informasi, (b) mewakili penawaran dan praktik produk AWS saat ini, yang dapat berubah tanpa pemberitahuan, dan (c) tidak membuat komitmen atau jaminan apa pun dari AWS dan afiliasinya, pemasok, atau pemberi lisensinya. Produk atau layanan AWS disediakan “sebagaimana adanya” tanpa jaminan, pernyataan, atau ketentuan dalam bentuk apa pun, baik tersurat maupun tersirat. Tanggung jawab dan kewajiban AWS kepada pelanggannya dikendalikan oleh perjanjian AWS, dan dokumen ini bukan bagian dari, juga tidak mengubah, perjanjian apa pun antara AWS dan pelanggannya.

AWS Glosarium

Untuk AWS terminologi terbaru, lihat AWS glosarium di Referensi.Glosarium AWS

Terjemahan disediakan oleh mesin penerjemah. Jika konten terjemahan yang diberikan bertentangan dengan versi bahasa Inggris aslinya, utamakan versi bahasa Inggris.