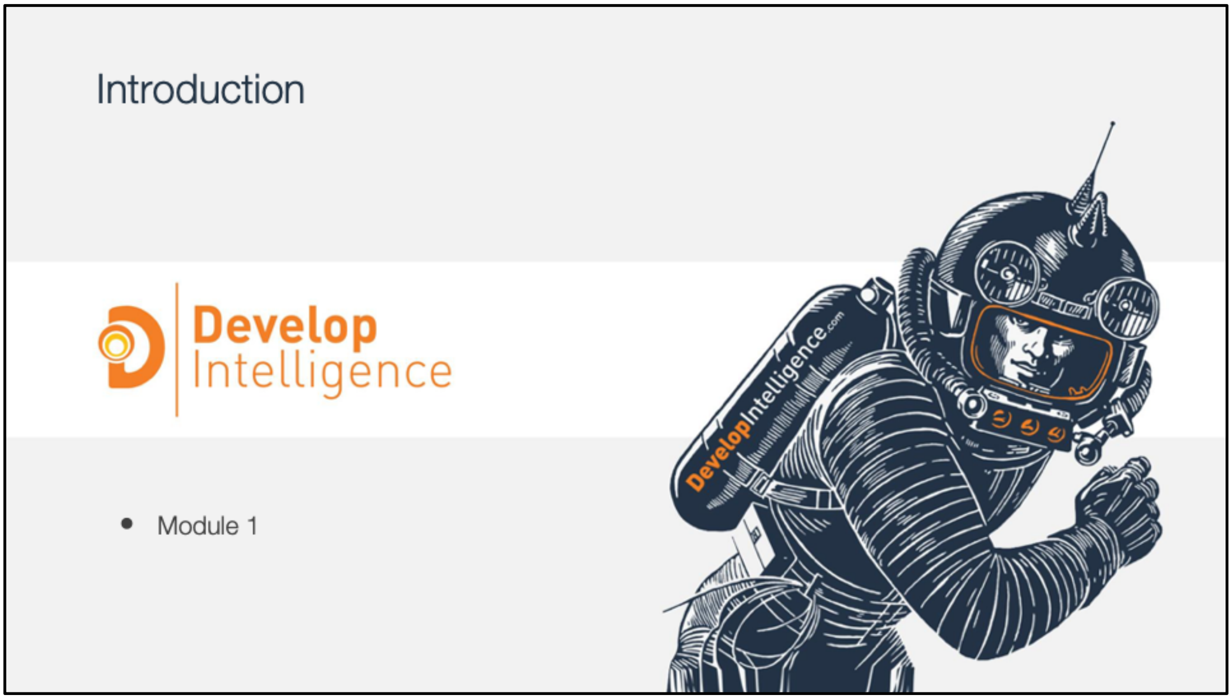
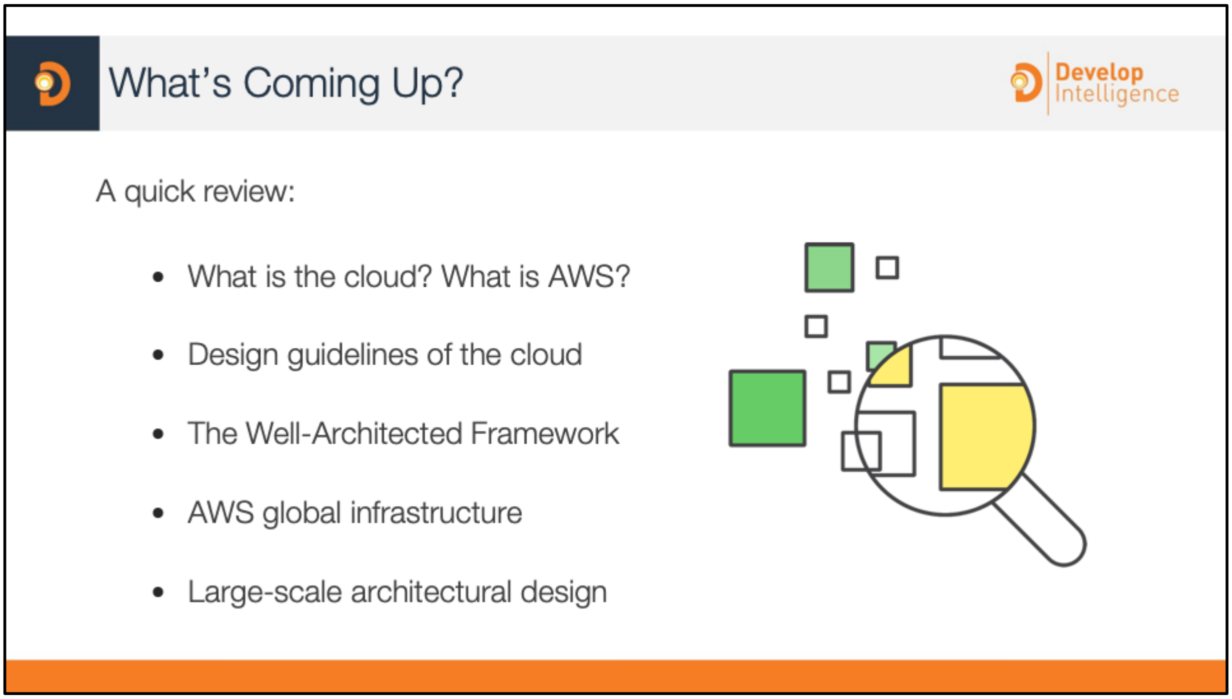


Introduction

Develop
Intelligence

- Module 1

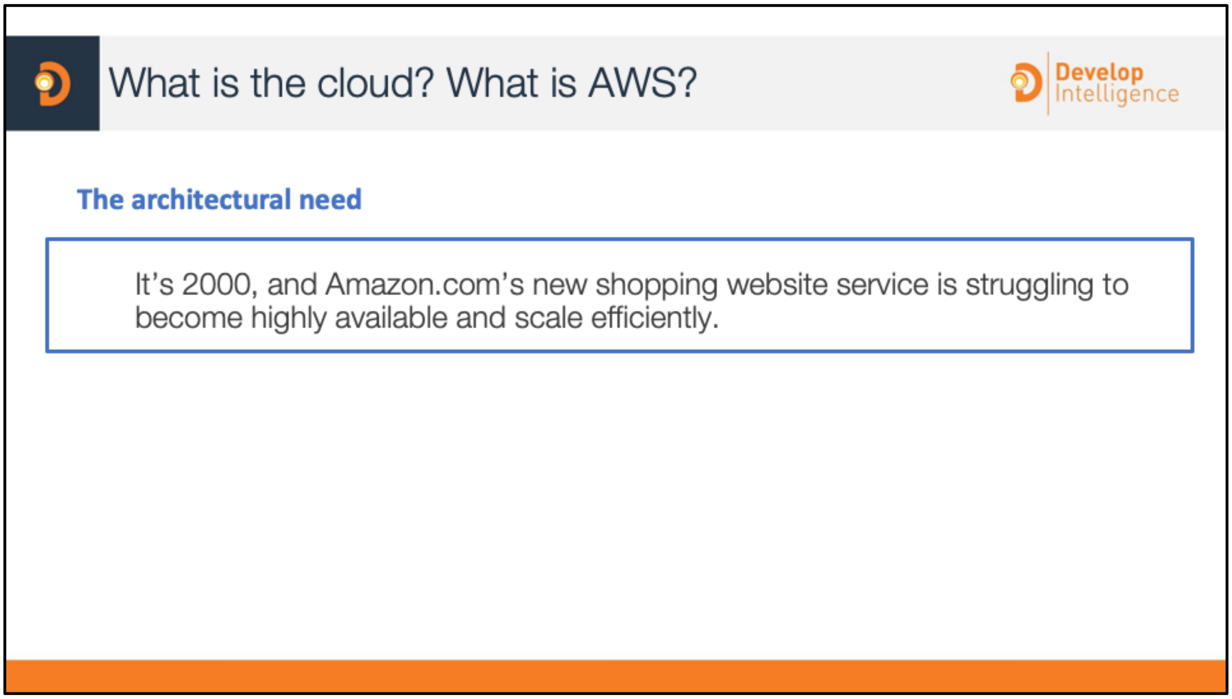


What's Coming Up?

Develop Intelligence

A quick review:

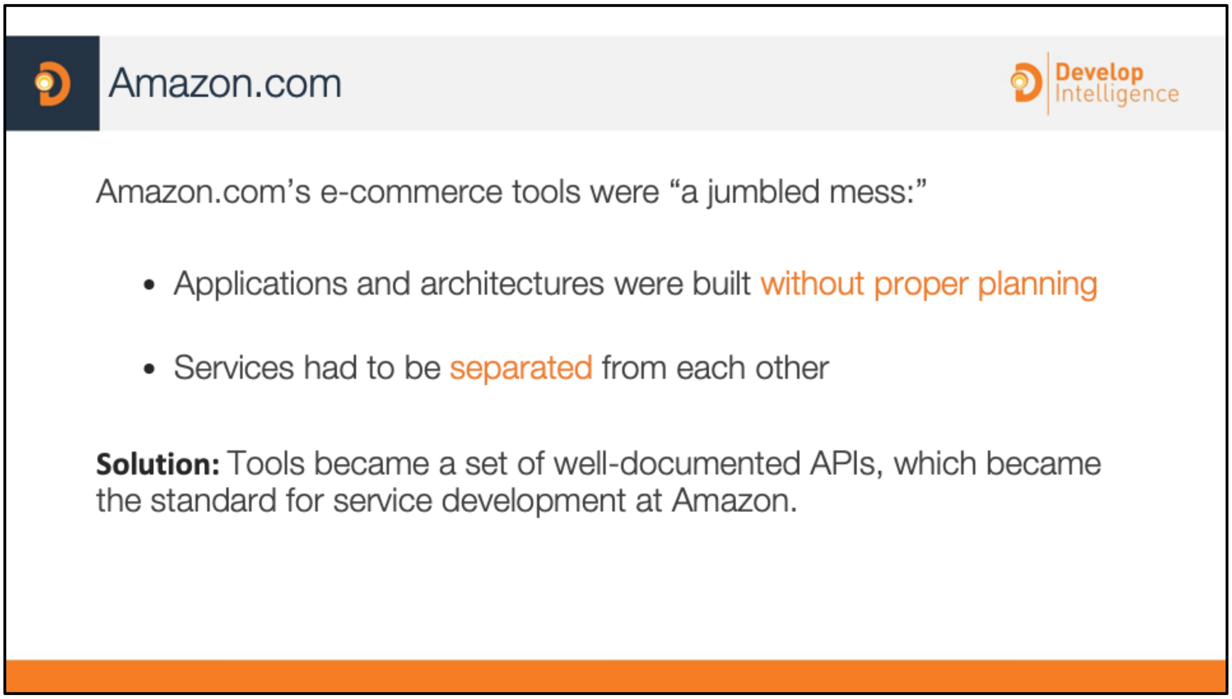
- What is the cloud? What is AWS?
- Design guidelines of the cloud
- The Well-Architected Framework
- AWS global infrastructure
- Large-scale architectural design



What is the cloud? What is AWS?

The architectural need

It’s 2000, and Amazon.com’s new shopping website service is struggling to become highly available and scale efficiently.



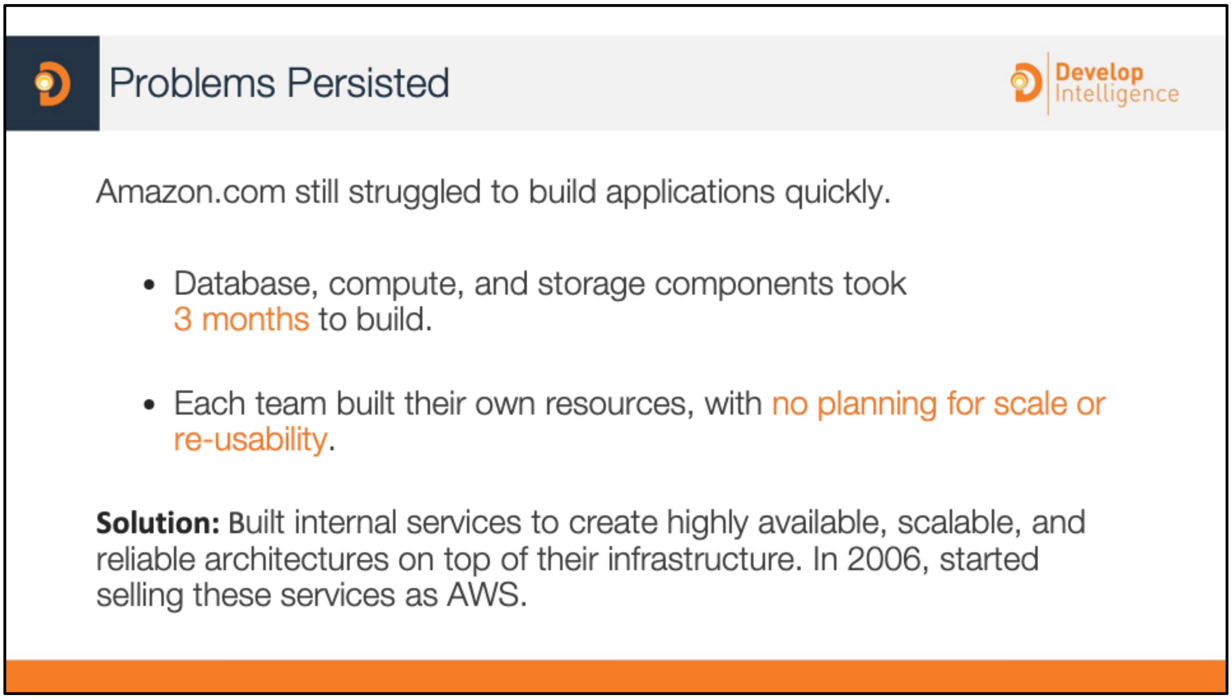
Amazon.com

Amazon.com's e-commerce tools were "a jumbled mess:"

- Applications and architectures were built without proper planning
- Services had to be separated from each other

Solution: Tools became a set of well-documented APIs, which became the standard for service development at Amazon.

https://techcrunch.com/2016/07/02/andy-jassys-brief-history-of-the-genesis-of-aws/

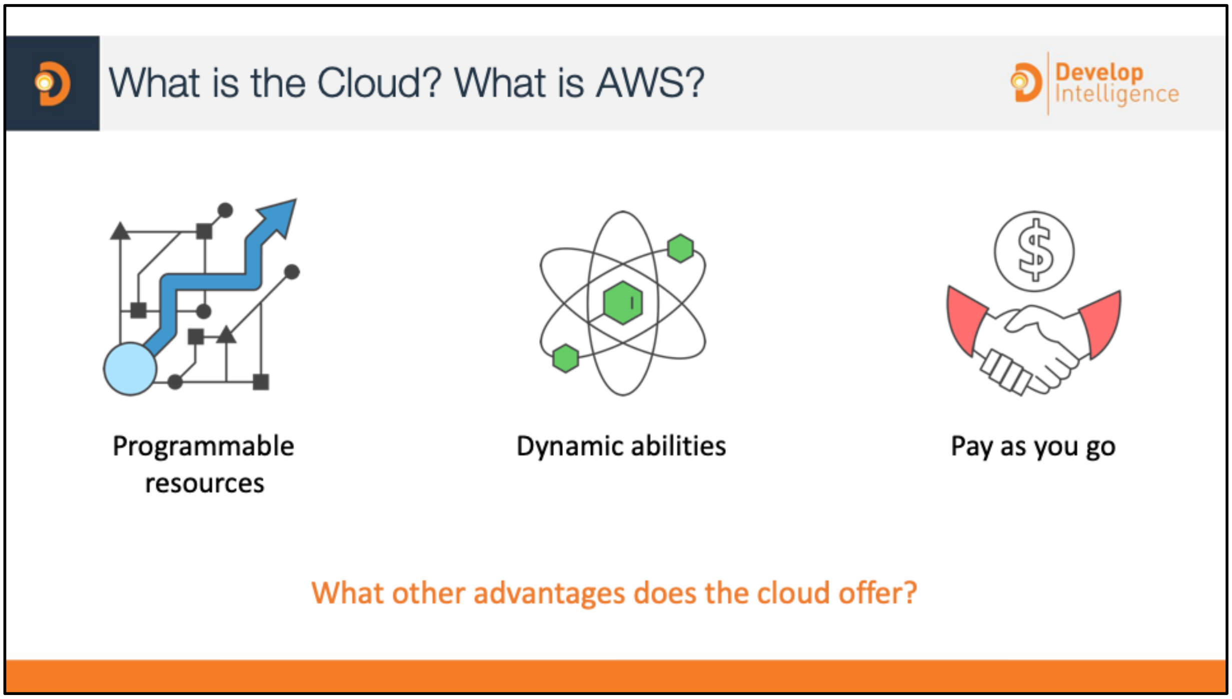


Problems Persisted

Amazon.com still struggled to build applications quickly.

- Database, compute, and storage components took 3 months to build.
- Each team built their own resources, with no planning for scale or re-usability.

Solution: Built internal services to create highly available, scalable, and reliable architectures on top of their infrastructure. In 2006, started selling these services as AWS.



What is the Cloud? What is AWS?

Programmable resources

Dynamic abilities

Pay as you go

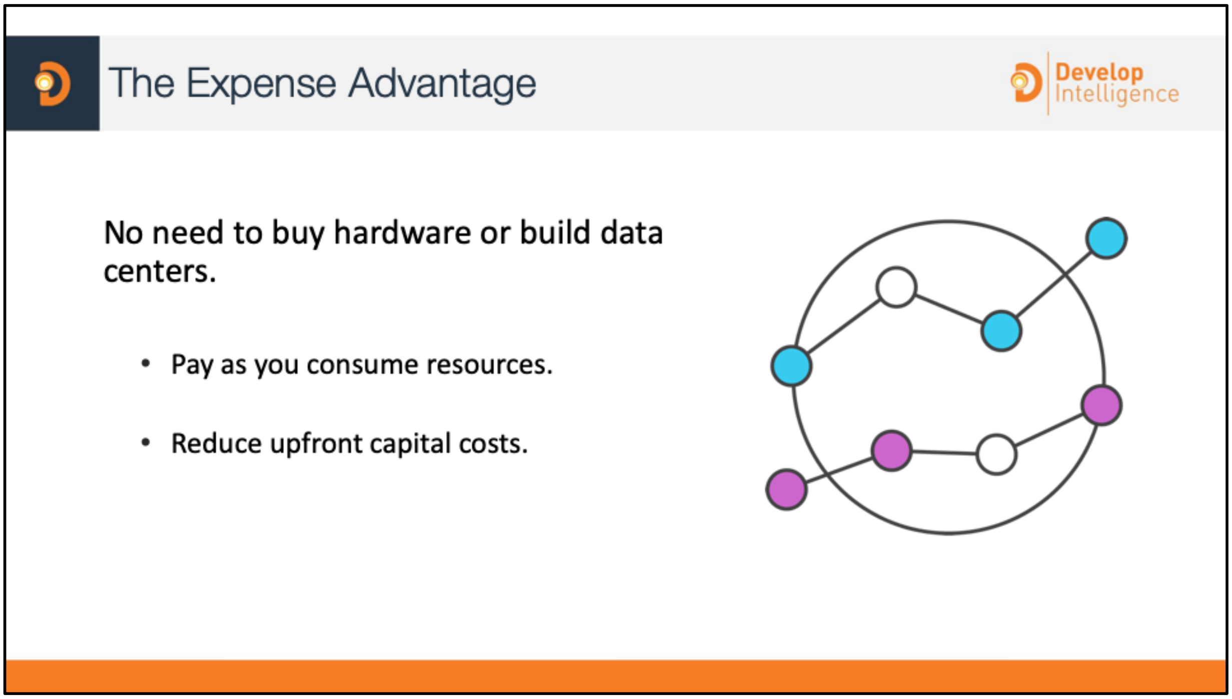
What other advantages does the cloud offer?

The cloud gives enormous advantage to those who can leverage its unique powers. The use of IT assets as programmatic resources allows you quickly set up and tear down infrastructure in a way that isn't possible with a traditional approach.

Access to these resources allows you to move forward in a very dynamic fashion. You can increase your database throughput or compute power with just a few clicks of the mouse. This provides an agility and flexibility that can really make the difference to your business.

Additionally, one of the biggest benefits of cloud computing is the ability to pay as you go. Letting you test and leverage the system without being fully committed. You can stop using these services at any time and change tactics to fit your needs.

Let's talk about the six advantages of cloud computing with AWS. For more information, see https://aws.amazon.com/what-is-cloud-computing



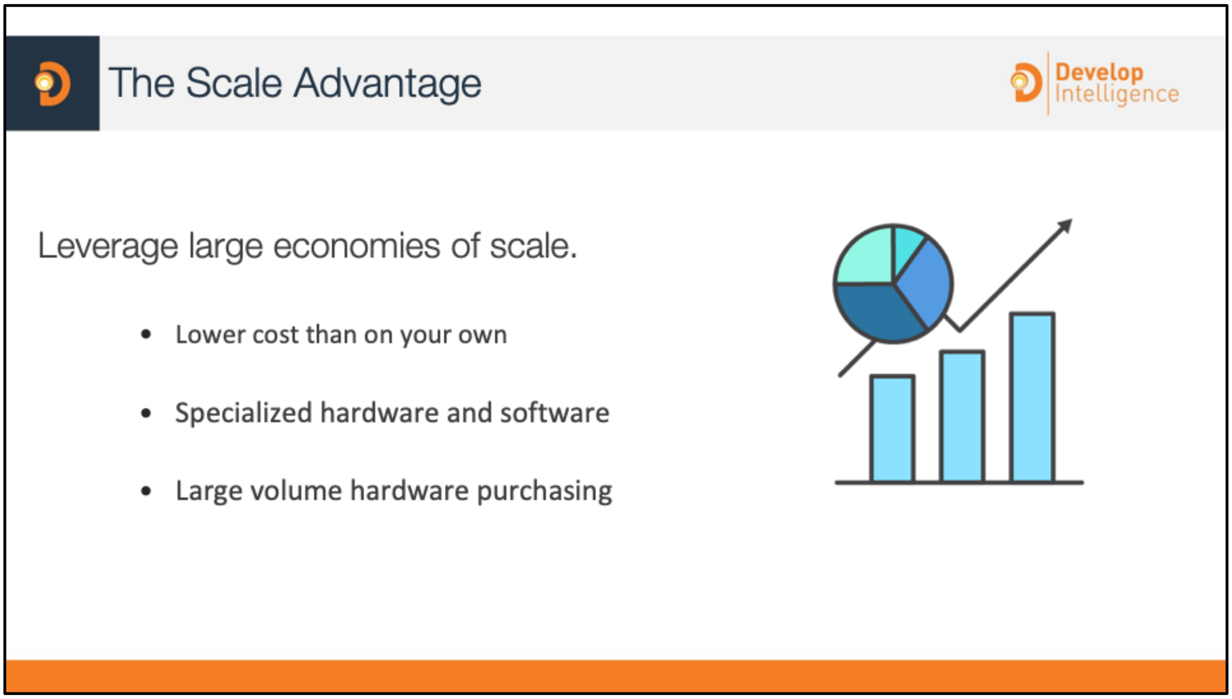
The Expense Advantage

No need to buy hardware or build data centers.

- Pay as you consume resources.
- Reduce upfront capital costs.

Instead of having to invest heavily in data centers and servers before you know how you're going to use them, you can only pay when you consume computing resources, and only pay for how much you consume.

This advantage is particularly good for startups or upfront budget-constrained projects. Being on the forward edge of technology can be risky. Building an on-premises infrastructure yourself might be cost-prohibitive, and it could stall testing, experimentation, and innovation. With the expense advantage, you can quickly get up and running while only paying for what you use.



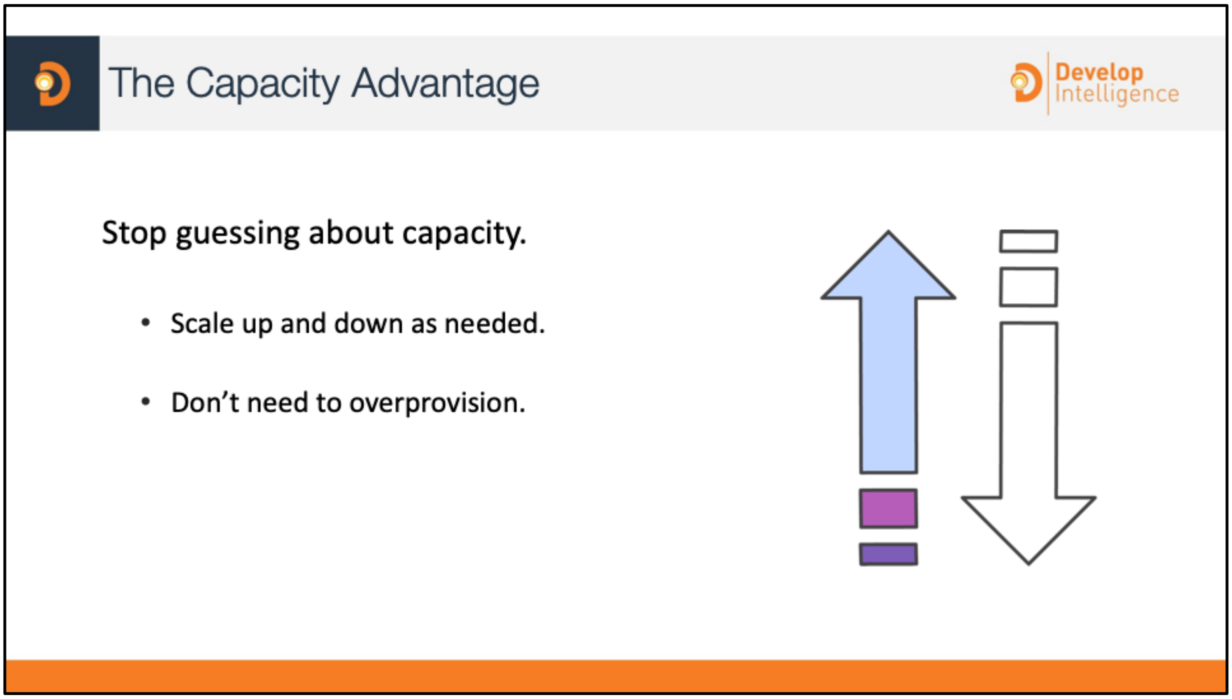
The Scale Advantage

Leverage large economies of scale.

- Lower cost than on your own
- Specialized hardware and software
- Large volume hardware purchasing

By using cloud computing, you can achieve a lower variable cost than you can get on your own. Because usage from hundreds of thousands of customers are aggregated in the cloud, providers like AWS can achieve higher economies of scale, which translates into lower pay-as-you-go prices.

AWS has developed specific hardware and software that is optimized for large-scale clouds. This, combined with the volume of these purchases, allows AWS to support lower costs and high efficiencies than those in most on-premises data centers. These savings are in turn passed on to the customer in order to lower prices and enhance the customer experience.



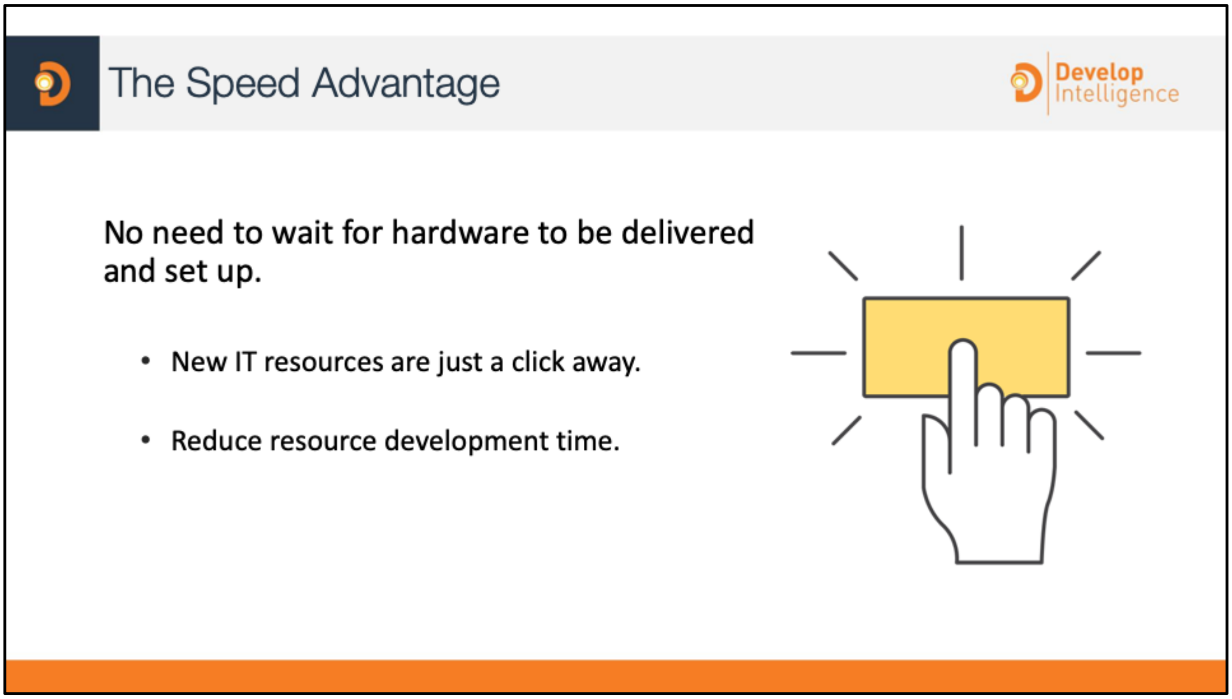
The Capacity Advantage

Stop guessing about capacity.

- Scale up and down as needed.
- Don’t need to overprovision.

Eliminate guessing on your infrastructure capacity needs. When using AWS, you only pay when you consume computing resources, and only pay for how much you consume. You can access as much or as little as you need, and scale up and down, scale in and out as required with only a few minutes’ notice.

It’s really hard to guess capacity if you don’t yet know what the customer response is—for example, if you’re launching a new product or service. Scaling your infrastructure as demand shifts and spikes offers a huge advantage over the mostly static on-premises solutions.



The Speed Advantage

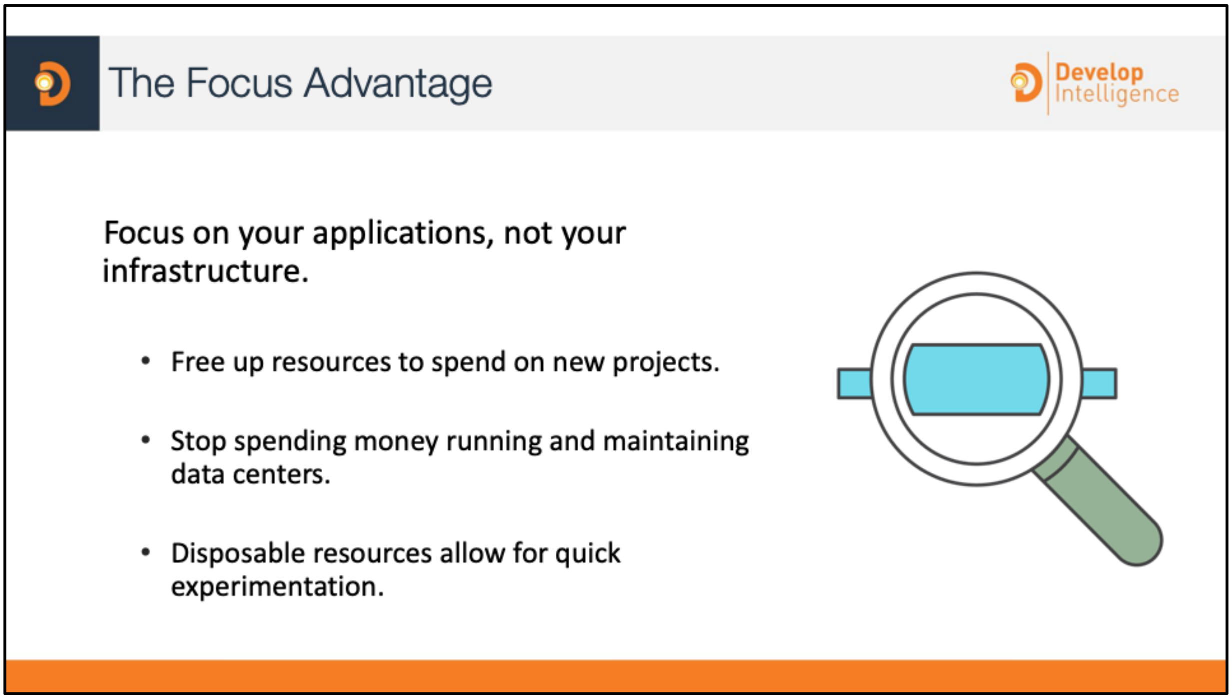
Develop Intelligence

No need to wait for hardware to be delivered and set up.

- New IT resources are just a click away.
- Reduce resource development time.

In a cloud computing environment, new IT resources are only a click away. You can make resources available to your developers in minutes, instead of weeks. This results in a dramatic increase in agility for the organization, since the cost and time it takes to experiment and develop is significantly lower.

It can take 6 to 20 weeks to provision a server in an on-premises environment. This timeframe can really stifle innovation. With AWS you can provision completely on your own, hundreds or even thousands of services in minutes. This allows you to experiment and create quickly.



The Focus Advantage

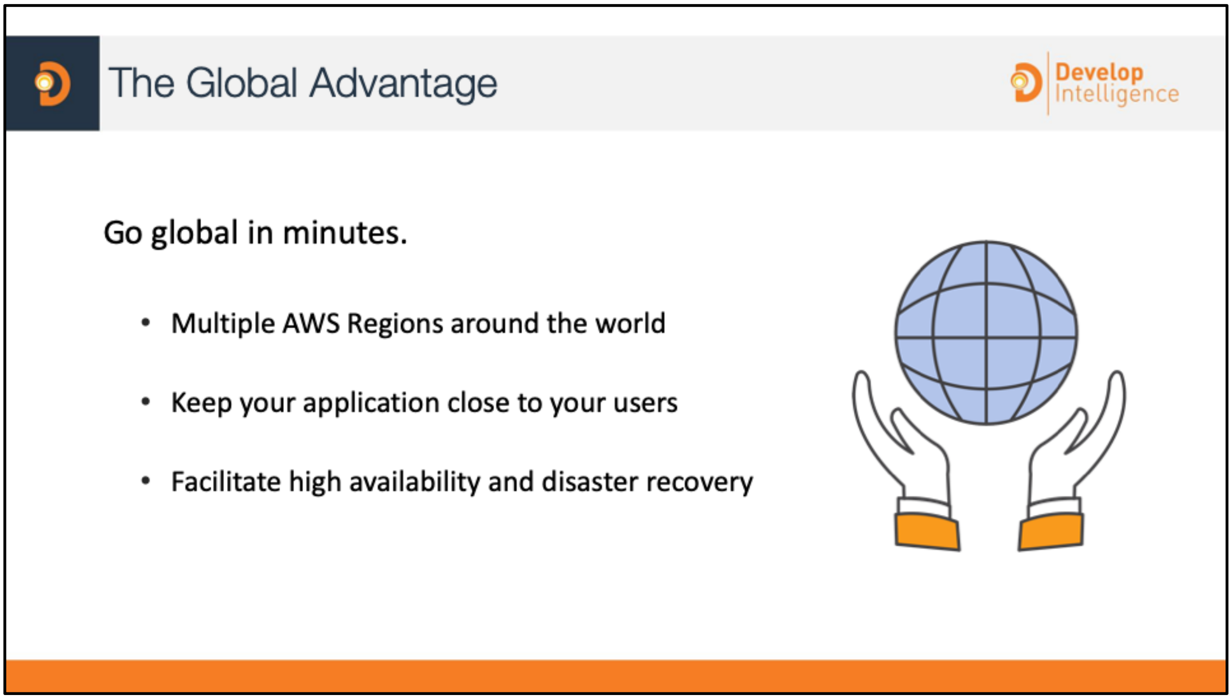
Develop Intelligence

Focus on your applications, not your infrastructure.

- Free up resources to spend on new projects.
- Stop spending money running and maintaining data centers.
- Disposable resources allow for quick experimentation.

Focus on projects that differentiate your business, not the infrastructure. Cloud computing lets you focus on your own customers, rather than on the heavy lifting of racking, stacking and powering servers.

The cloud has already done much of the heavy lifting for you. For most companies, their scarcest resource are their software development engineers. Engineering teams have a long list of priorities and tasks that need to be accomplished. It’s a significant advantage to focus that resource on projects that move your mission forward, instead of working on underlying infrastructure.

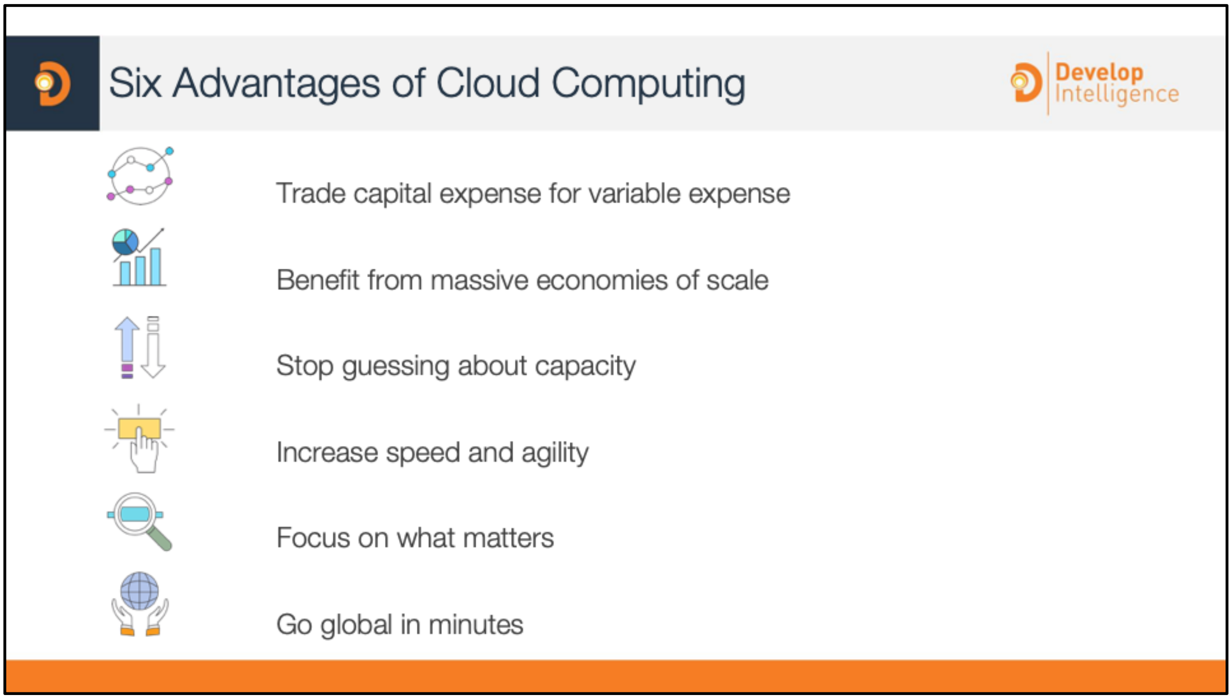


The Global Advantage

Go global in minutes.

- Multiple AWS Regions around the world
- Keep your application close to your users
- Facilitate high availability and disaster recovery

You can deploy your application in multiple regions around the world with just a few clicks—providing lower latency and better experience for your customers simply and at minimal cost.

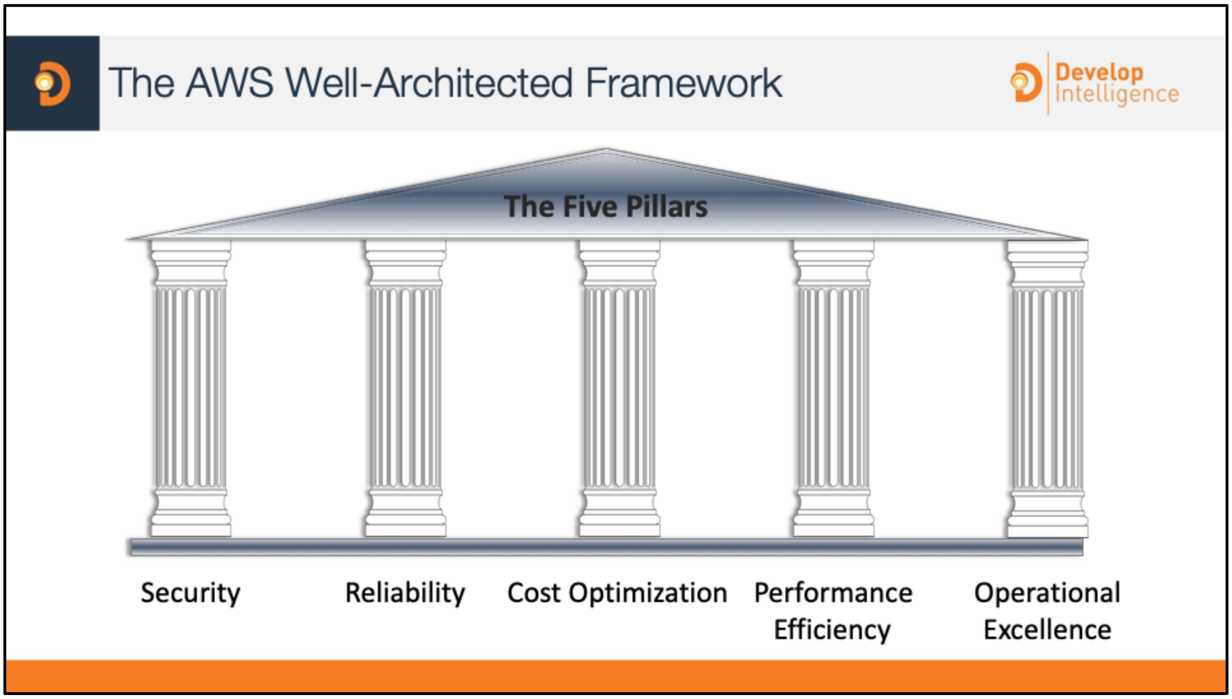


Six Advantages of Cloud Computing

Develop Intelligence

- Trade capital expense for variable expense
- Benefit from massive economies of scale
- Stop guessing about capacity
- Increase speed and agility
- Focus on what matters
- Go global in minutes

For more information about the six primary benefits of cloud computing with AWS, see
https://www.youtube.com/watch?v=yMJ75k9X5_8



The AWS Well-Architected Framework

Develop Intelligence

The Five Pillars

Security

Reliability

Cost Optimization

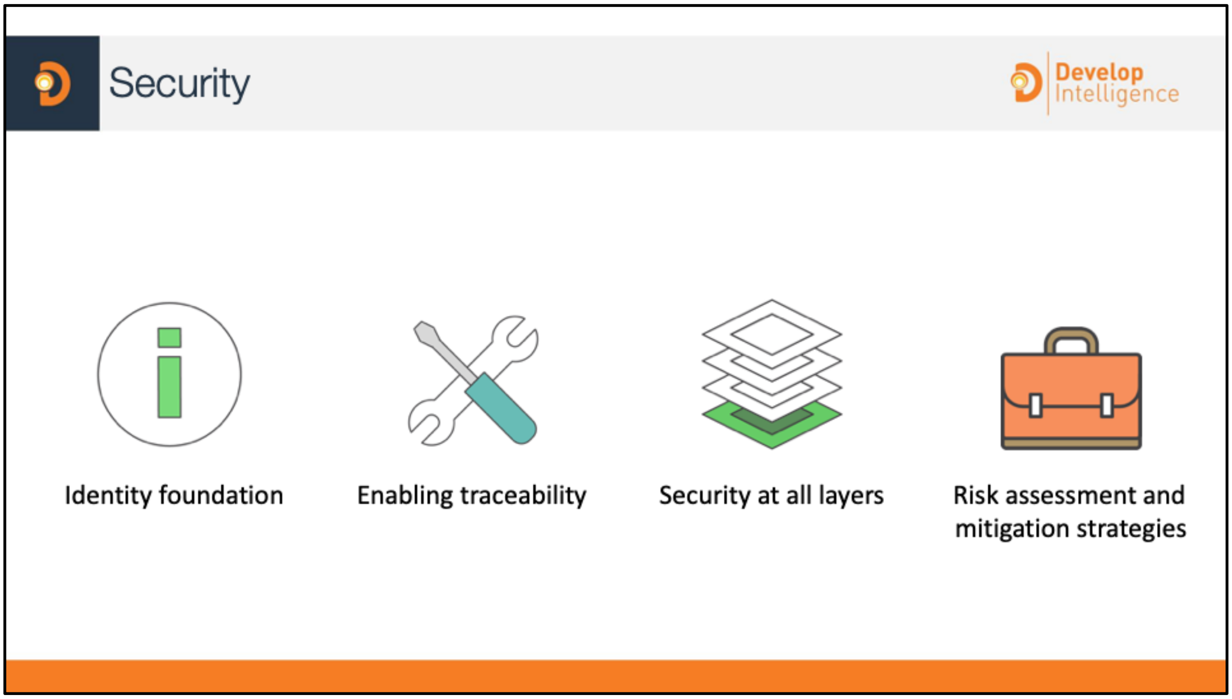
Performance Efficiency

Operational Excellence

First we'll talk about some of the goals of the design principles in the WAF.

If you would like to have some assistance with well architected design:

The AWS Well-Architected Tool is a self-service tool that provides you with on-demand access to current AWS best practices. It is designed to help architects and their managers review AWS workloads at any time, without the need for an AWS Solutions Architect. This service is based on the AWS Well-Architected Framework, which was developed to help cloud architects build secure, high-performing, resilient, and efficient application infrastructure. You can review the state of your workloads and compare them to the latest AWS architectural best practices.



Security

Identity foundation

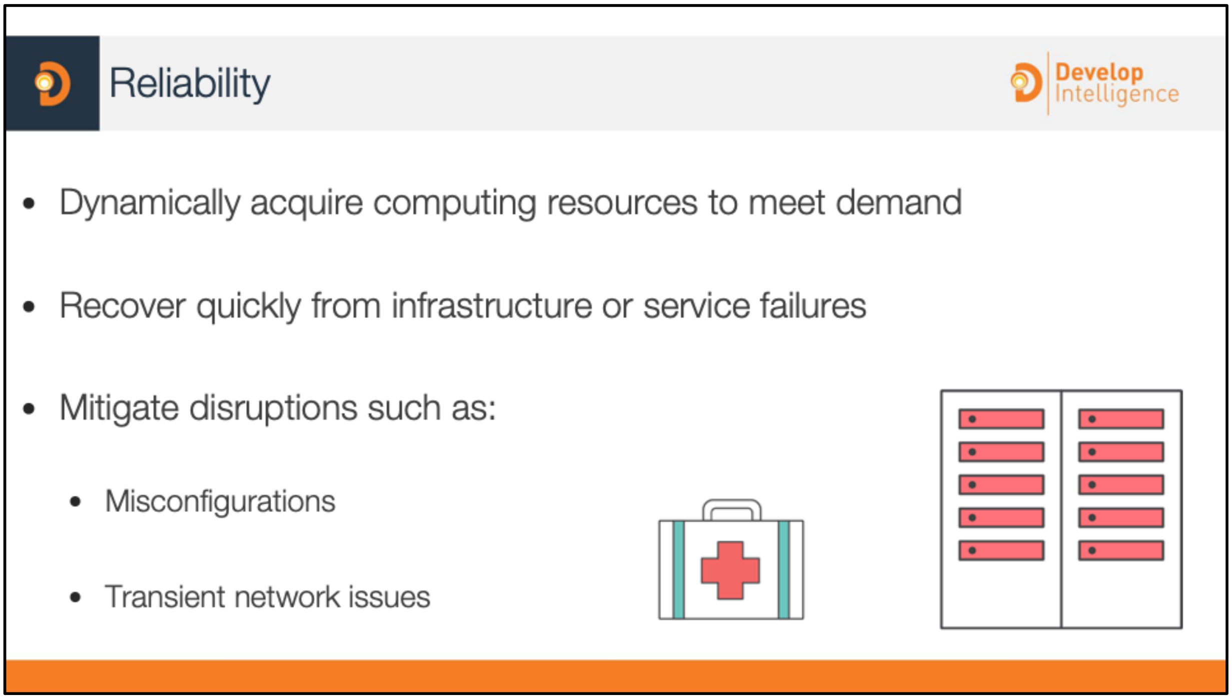
Enabling traceability

Security at all layers

Risk assessment and mitigation strategies

Security deals with protecting information and mitigating possible damage. Your architecture will present a much stronger security presence by implementing some basic security measures, like implementing a strong identity foundation, enabling traceability, applying security at all layers, automating security best practices, protecting data in transit and at rest

For more information, see https://d1.awsstatic.com/whitepapers/architecture/AWS-Security-Pillar.pdf

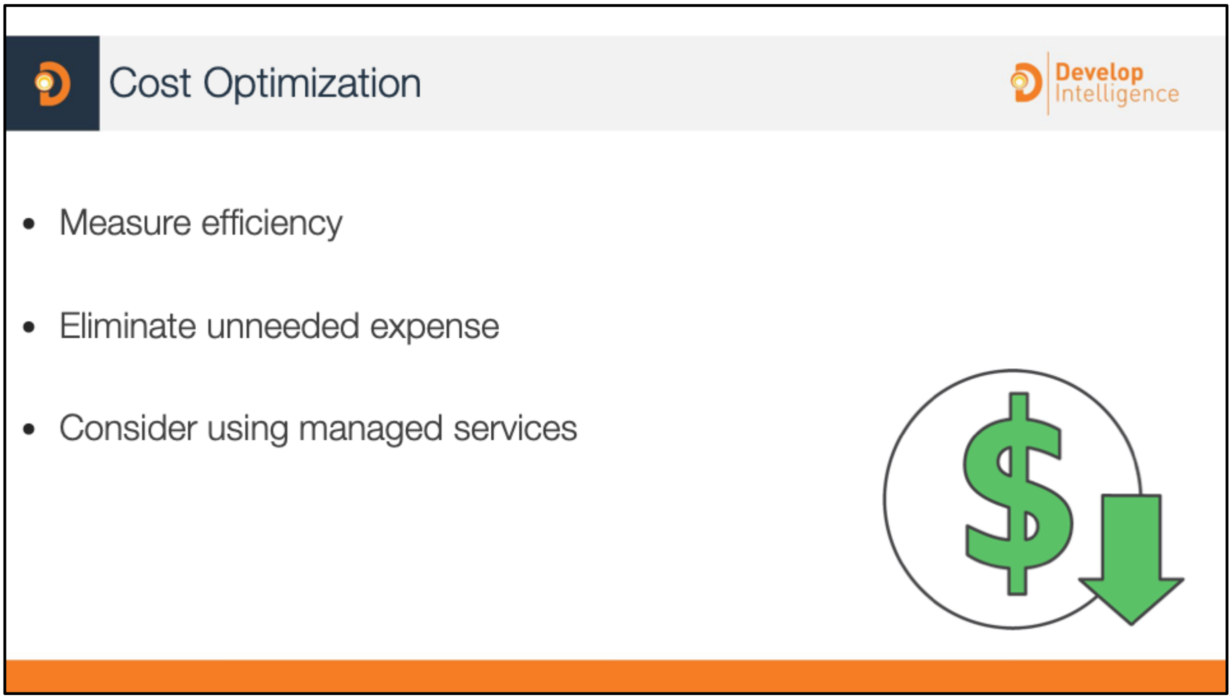

Develop Intelligence

Reliability

- Dynamically acquire computing resources to meet demand
- Recover quickly from infrastructure or service failures
- Mitigate disruptions such as:
  - Misconfigurations
  - Transient network issues

Ensuring reliability can be difficult in a traditional environment. Issues arise from single points of failure, lack of automation, and lack of elasticity. When applying the ideas of the reliability pillar, you will be able to prevent many of these issues. Properly designing your architecture in respect to high availability, fault tolerance, and overall redundancy will be helpful for you and your customers.

For more information, see https://d1.awsstatic.com/whitepapers/architecture/AWS-Reliability-Pillar.pdf



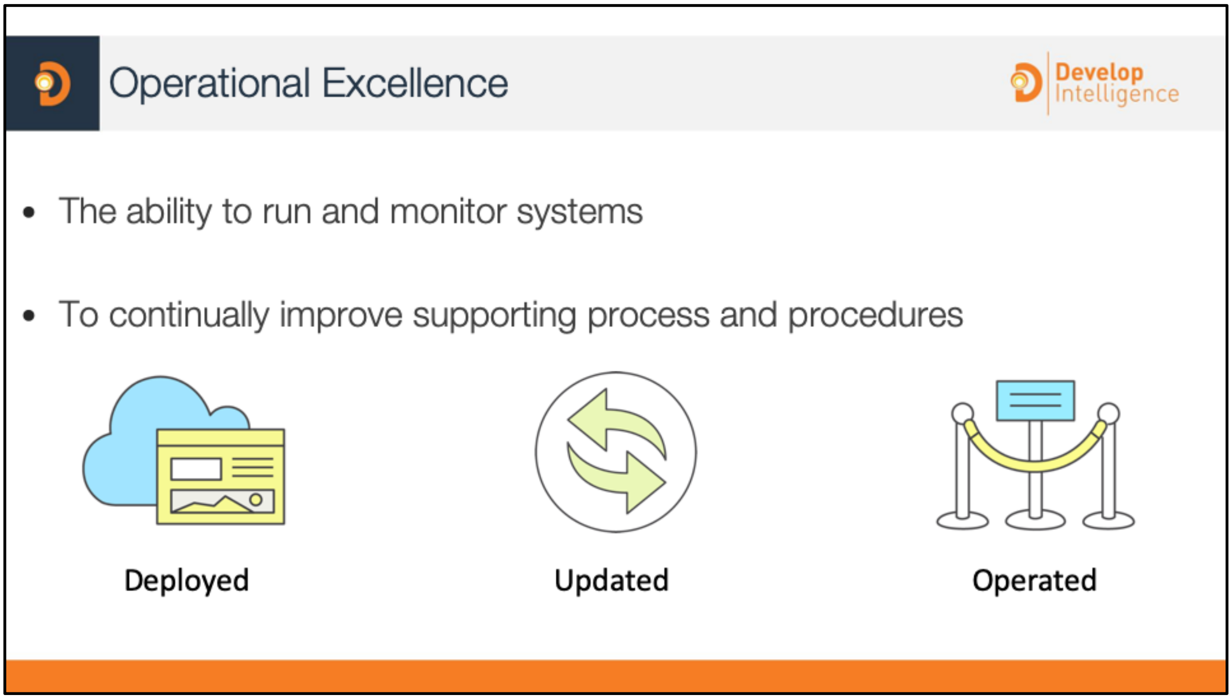
Cost Optimization

Develop
Intelligence

- Measure efficiency
- Eliminate unneeded expense
- Consider using managed services

Cost optimization is an ongoing requirement of any good architectural design. The process is iterative and should be refined and improved throughout your production lifetime. Understanding how efficient your current architecture is in relation to your goals will ultimately help with removing unneeded expense. Consider using managed services as they operate at cloud scale and can offer a lower cost per transaction or service.

For more information, see https://d1.awsstatic.com/whitepapers/architecture/AWS-Cost-Optimization-Pillar.pdf



Operational Excellence

- The ability to run and monitor systems
- To continually improve supporting process and procedures

Deployed

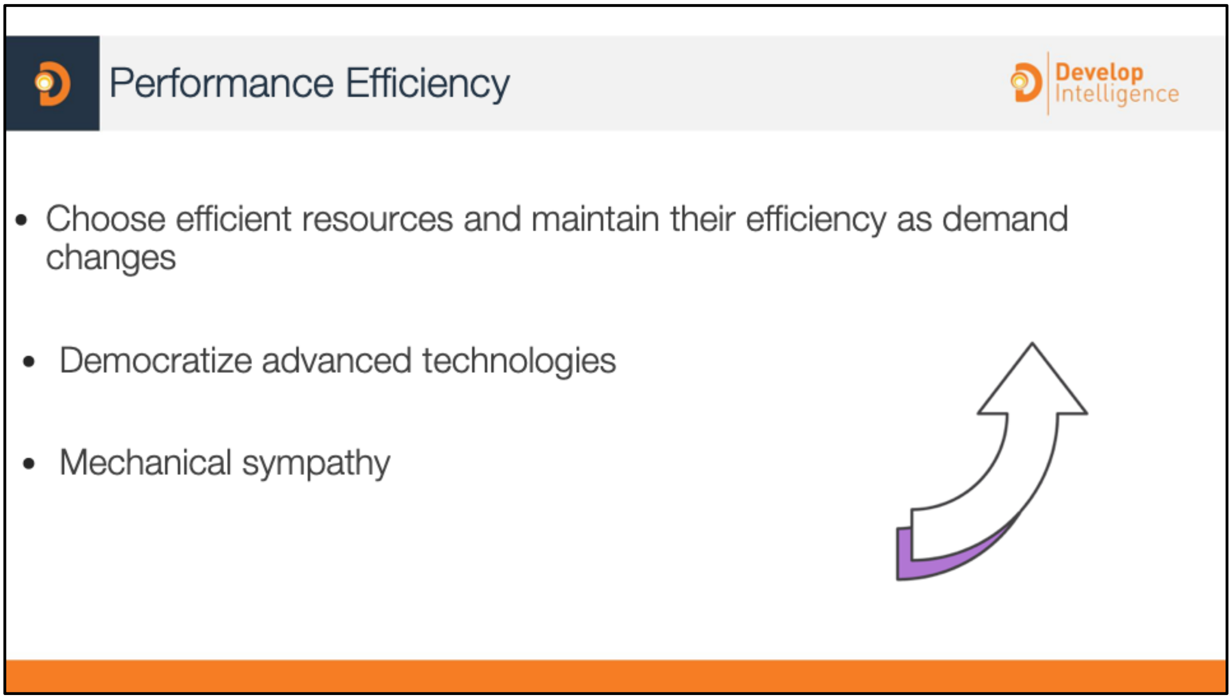
Updated

Operated

When creating a design or architecture, you must be aware of how it will be deployed, updated, and operated. It is imperative that you work towards defect reductions and safe fixes and enable observation with logging instrumentation.

In AWS, you can view your entire workload (applications, infrastructure, policy, governance, and operations) as code. It can all be defined in and updated using code. This means you can apply the same engineering discipline that you use for application code to every element of your stack.

For more information, see https://d1.awsstatic.com/whitepapers/architecture/AWS-Operational-Excellence-Pillar.pdf



Performance Efficiency

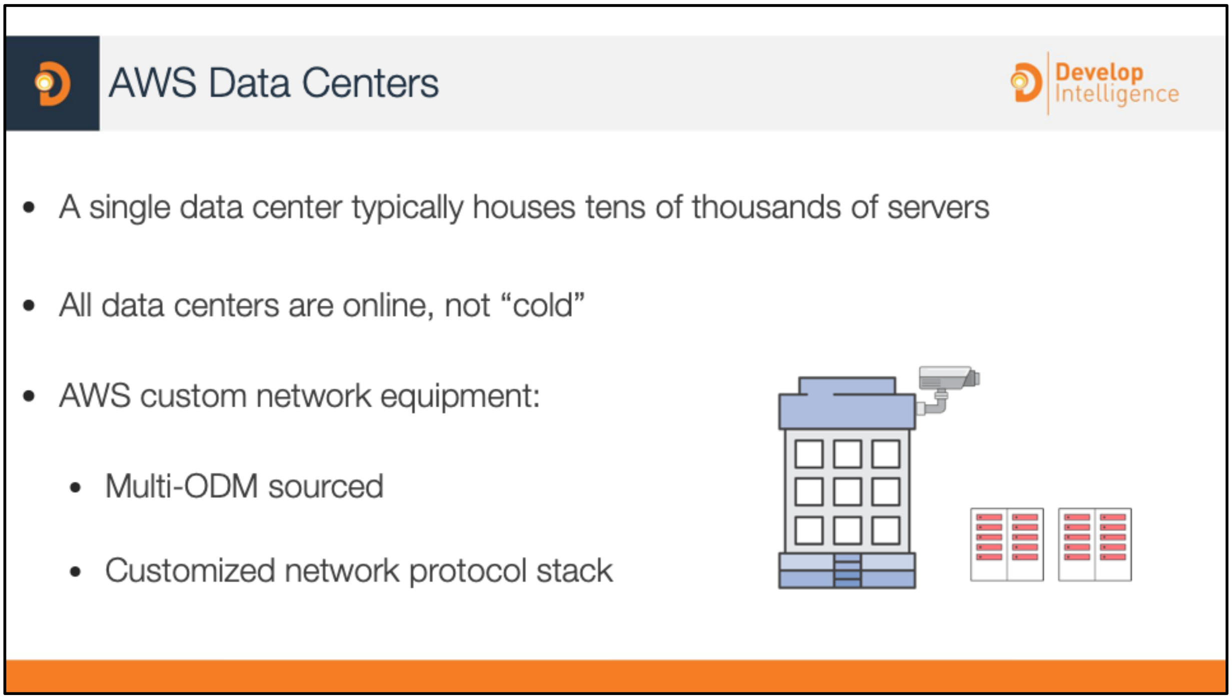
- Choose efficient resources and maintain their efficiency as demand changes
- Democratize advanced technologies
- Mechanical sympathy

When considering performance, you want to maximize your performance by using computation resources efficiently and maintain that efficiency as the demand changes.

It is also important to democratize advanced technologies. In situations where technology is difficult to implement yourself, consider using a vendor. In implementing the technology for you, the vendor takes on the complexity and knowledge, freeing your team to focus on more value-added work.

Mechanical sympathy: Use the technology approach that aligns best to what you are trying to achieve. For example, consider data access patterns when you select database or storage approaches.

For more information, see https://d1.awsstatic.com/whitepapers/architecture/AWS-Performance-Efficiency-Pillar.pdf



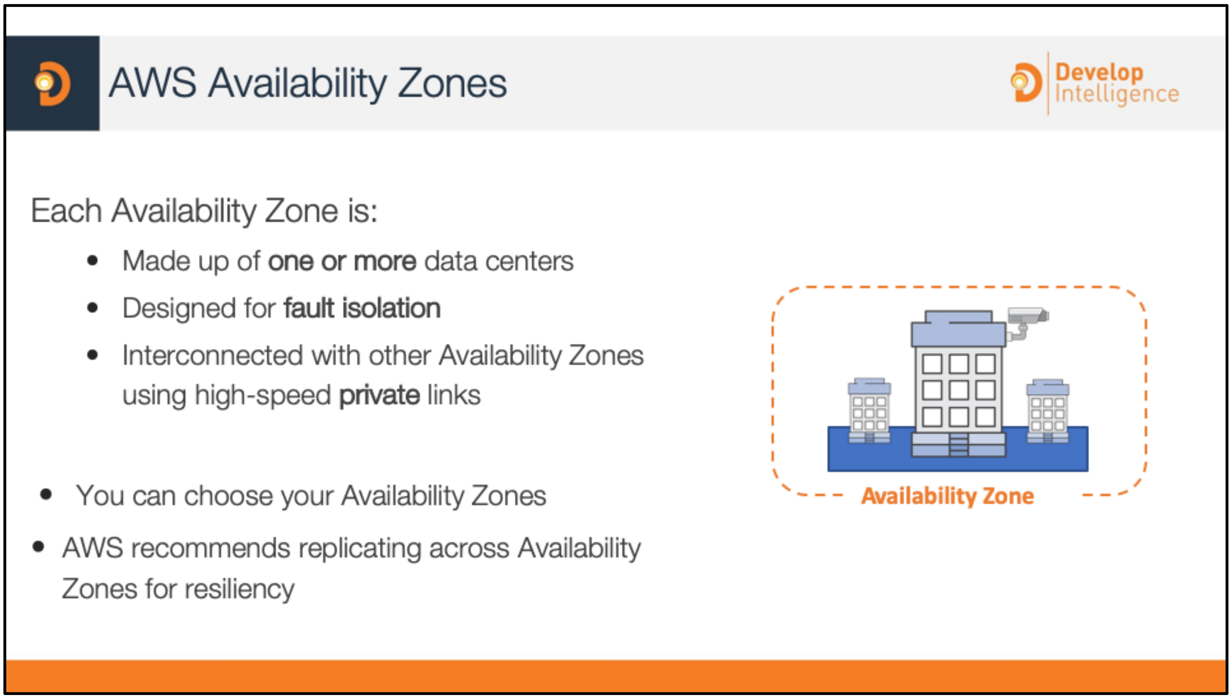
AWS Data Centers

- A single data center typically houses tens of thousands of servers
- All data centers are online, not “cold”
- AWS custom network equipment:
  - Multi-ODM sourced
  - Customized network protocol stack

AWS data centers are built in clusters in various global regions. Larger data centers are undesirable; all data centers are online and serving customers. No data center is “cold;” in case of failure, automated processes move customer data traffic away from the affected area. Core applications are deployed in an N+1 configuration, so that in the event of a data center failure, there is sufficient capacity to enable traffic to be load-balanced to the remaining sites.

Original design manufacturers, or ODMs, designs and manufactures products based on specifications from a second company. The second company then rebrands the products for sale.

For more information, see https://aws.amazon.com/compliance/data-center/



AWS Availability Zones

Develop Intelligence

Each Availability Zone is:

- Made up of **one or more** data centers
- Designed for **fault isolation**
- Interconnected with other Availability Zones using high-speed **private** links

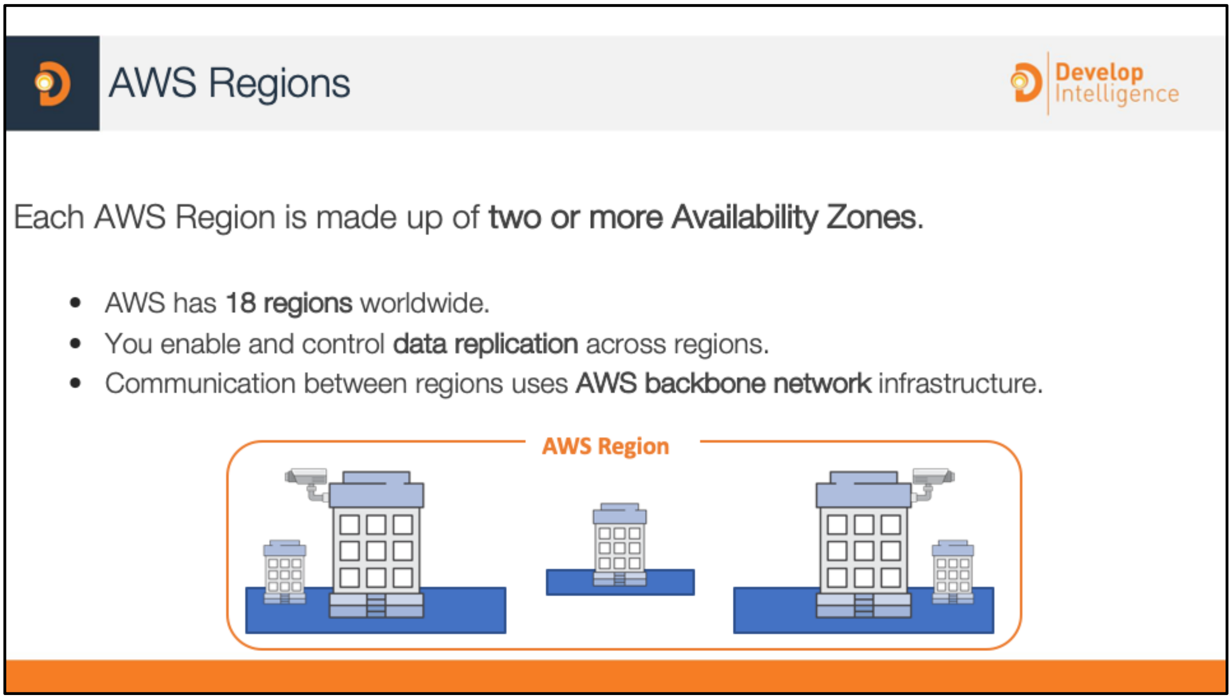
- You can choose your Availability Zones
- AWS recommends replicating across Availability Zones for resiliency

Availability Zone

AWS data centers are organized into *Availability Zones*. Each Availability Zone comprises one or more data centers, with some Availability Zones having as many as six data centers. However, no data center can be part of two Availability Zones.

Each Availability Zone is designed as an independent failure zone. This means that Availability Zones are physically separated within a typical metropolitan region and are located in lower-risk flood plains (specific flood-zone categorization varies by region). In addition to having discrete uninterruptable power supply and onsite backup generation facilities, they are each fed via different grids from independent utilities to further reduce single points of failure. Availability Zones are all redundantly connected to multiple tier-1 transit providers.

You are responsible for selecting the Availability Zones where your systems will reside. Systems can span multiple Availability Zones. You should design your systems to survive temporary or prolonged failure of an Availability Zone if a disaster occurs. Distributing applications across multiple Availability Zones allows them to remain resilient in most failure situations, including natural disasters or system failures.



AWS Regions

Each AWS Region is made up of **two or more Availability Zones**.

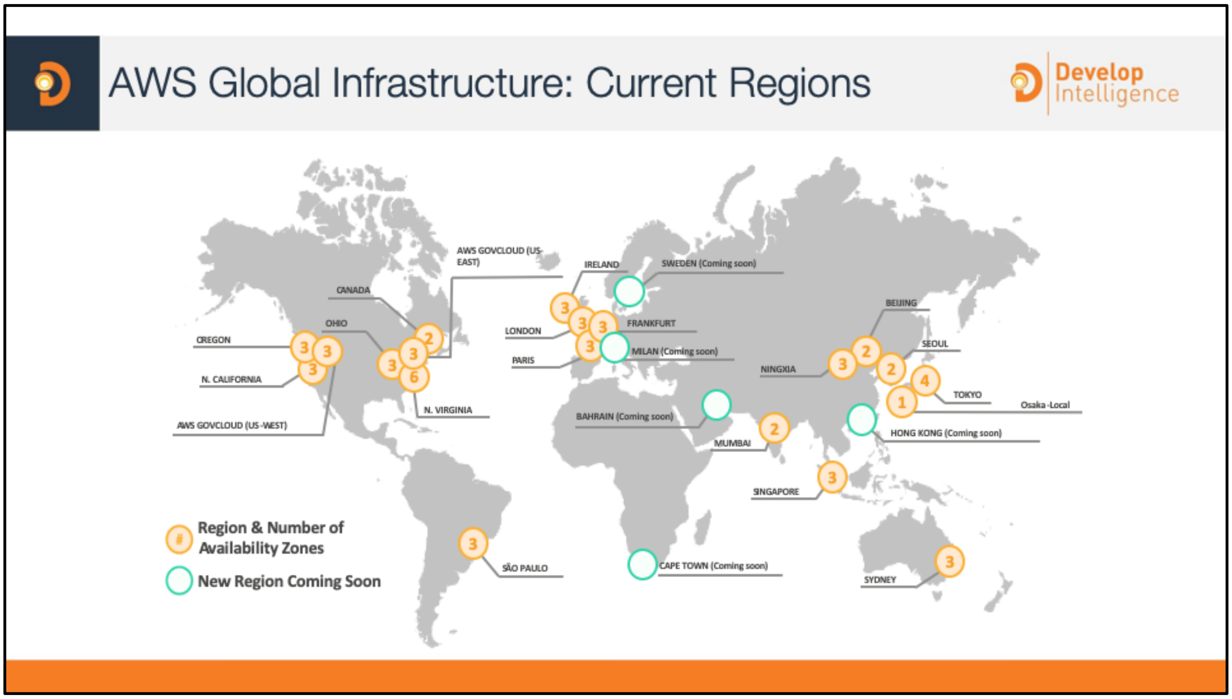
- AWS has **18 regions** worldwide.
- You enable and control **data replication** across regions.
- Communication between regions uses **AWS backbone network** infrastructure.

Availability Zones are further grouped into *AWS Regions*. Each region contains two or more Availability Zones.

When you distribute applications across multiple Availability Zones, be aware of location-dependent privacy and compliance requirements, such as the EU Data Privacy Directive. When you store data in a specific region, it is not replicated outside that region. AWS never moves your data out of the region you put it in. It is your responsibility to replicate data across regions, if your business needs require that. AWS provides information about the country, and—where applicable—the state where each region resides; you are responsible for selecting the region to store data in based on your compliance and network latency requirements.

AWS Regions are connected to multiple Internet Service Providers (ISPs) as well as to a private global network backbone, which provides lower cost and more consistent cross-region network latency when compared with the public Internet.

For more information, see https://aws.amazon.com/about-aws/global-infrastructure/#reglink-pr



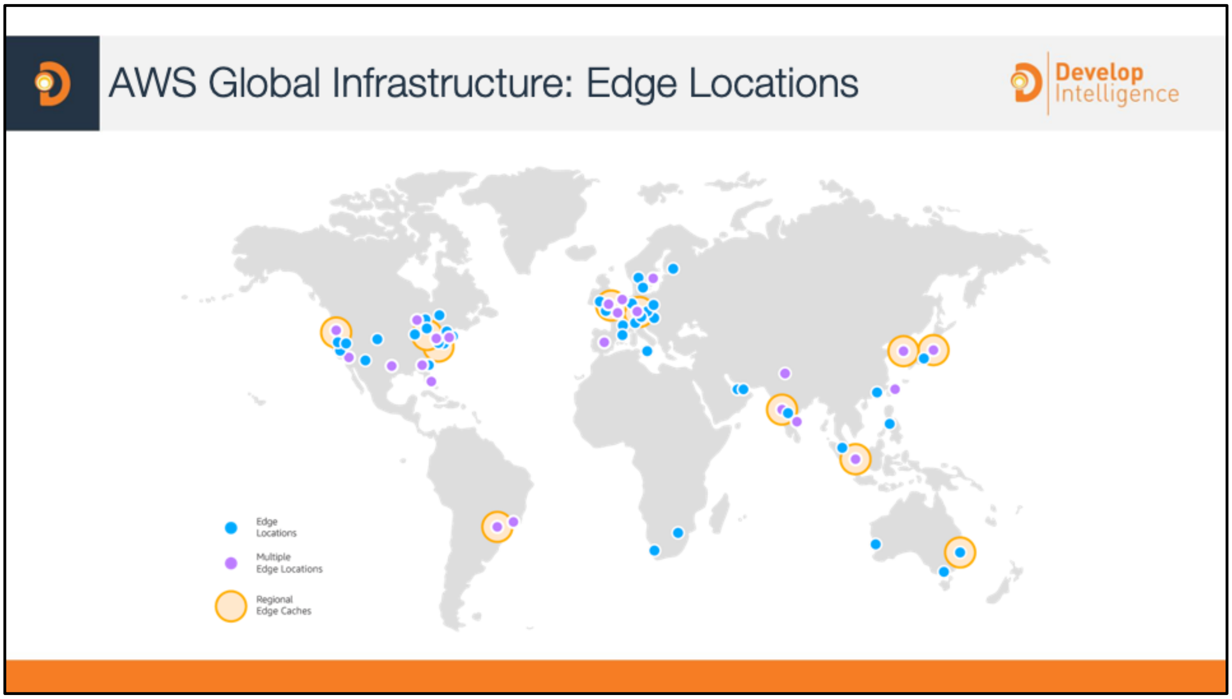
AWS Global Infrastructure: Current Regions

AWS is steadily expanding its global infrastructure to help customers achieve lower latency and higher throughput and to ensure that your data resides only in the region you specify. As you and all customers grow your businesses, AWS will continue to provide infrastructure that meets your global requirements.

The isolated GovCloud (US) Region is designed to allow US government agencies and customers to move sensitive workloads into the cloud by addressing their specific regulatory and compliance requirements.

AWS products and services are available by region so you may not see all regions available for a given service.

You can run applications and workloads from a region to reduce latency to end users while avoiding the up-front expenses, long-term commitments, and scaling challenges associated with maintaining and operating a global infrastructure.



AWS Global Infrastructure: Edge Locations

Develop Intelligence

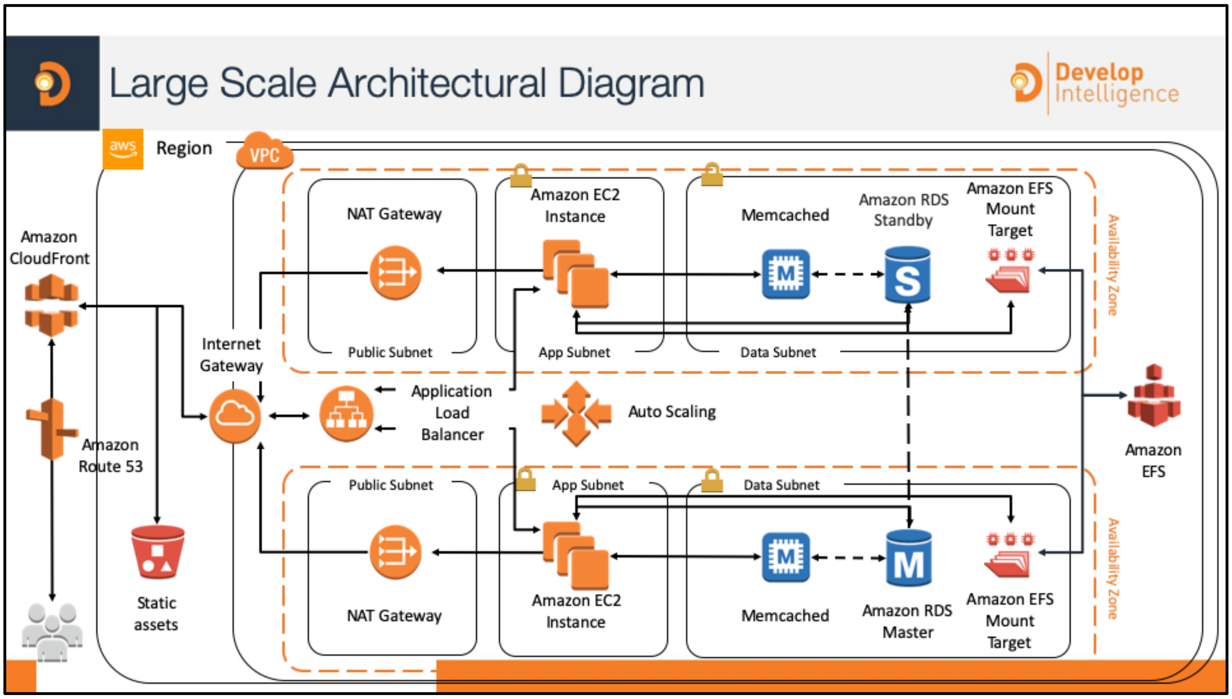
To deliver content to end users with lower latency, Amazon CloudFront uses a global network of 150 Points of Presence (139 Edge Locations and 11 Regional Edge Caches) in 65 cities across 29 countries.

Edge locations are located in: North America, Europe, Asia, Australia, and South America, and support AWS services like Amazon Route 53 and Amazon CloudFront.

Regional Edge Caches

Regional edge caches, used by default with Amazon CloudFront, are utilized when you have content that is not accessed frequently enough to remain in an edge location. Regional edge caches absorb this content and provide an alternative to that content having to be fetched from the origin server.

For more information, see https://aws.amazon.com/cloudfront/features/



Large Scale Architectural Diagram

By the end of class, you will be able to understand all of the components of this architectural diagram. You will also be able to construct your own architectural solutions that are just as large and robust.