Service Name

AWS Transit Gateway

Service Overview

AWS Transit Gateway is a fully managed network transit hub that enables customers to interconnect multiple Virtual Private Clouds (VPCs) and on-premises networks through a central gateway using a hub-and-spoke architecture. It simplifies network management by eliminating the need for complex VPC peering relationships or multiple VPN connections, allowing thousands of VPCs to connect to a single gateway. Transit Gateway acts as a regional virtual router that scales elastically based on network traffic, supporting both IPv4 and IPv6 protocols while providing centralized routing configuration and traffic inspection capabilities across multi-account and multi-region environments.

Key Use Cases

- Centralized connectivity: Connect thousands of VPCs and on-premises networks through a single hub, eliminating complex peering relationships and reducing operational overhead
- Multi-account network architecture: Enable secure communication between VPCs across multiple AWS accounts while maintaining centralized network management and routing policies
- Hybrid cloud connectivity: Integrate on-premises networks with AWS infrastructure using VPN connections, Direct Connect, and SD-WAN solutions through Transit Gateway Connect attachments
- Network segmentation: Implement isolated routing domains and traffic inspection using multiple route tables and centralized security appliances in appliance mode
- Multi-region global networks: Connect networks across AWS regions using Transit Gateway peering attachments while keeping traffic on the AWS backbone

Features

- **Hub-and-spoke architecture**: Central gateway connecting thousands of VPCs and on-premises networks with simplified routing
- **Multiple attachment types**: Support for VPC, VPN, Direct Connect, Transit Gateway Connect, peering, and network function attachments
- **Route table management**: Default and custom route tables with dynamic and static routing, route propagation, and traffic segmentation
- **Multicast support**: Native multicast routing with IGMP snooping for broadcasting applications like financial data feeds
- **Cross-region peering**: Inter-region connectivity keeping traffic on AWS backbone with automatic encryption at physical layer

- **Encryption control**: Enforce encryption-in-transit for all traffic on attached VPCs with configurable encryption policies
- **Network monitoring**: Flow logs, CloudWatch metrics, and Network Manager integration for comprehensive network visibility
- **Resource sharing**: AWS RAM integration for sharing transit gateways across multiple accounts with granular permissions
- **High bandwidth performance**: Up to 100 Gbps per VPC attachment per AZ and 20 Gbps per Connect attachment with ECMP support
- **Flexible cost allocation**: Customizable cost allocation policies for data processing charges across attachments and accounts

Value Proposition

AWS Transit Gateway provides significant advantages over traditional networking approaches by dramatically reducing operational complexity through its managed hub-and-spoke model, eliminating the need for hundreds of VPC peering connections. It offers enterprise-grade scalability supporting thousands of VPC connections per gateway while maintaining high performance with up to 100 Gbps bandwidth per attachment:

- The service provides enhanced security through built-in encryption controls and centralized traffic inspection capabilities
- Cost efficiency is achieved through simplified architecture reducing the number of required connections and flexible cost allocation options
- Unlike self-managed alternatives, Transit Gateway is fully managed by AWS, reducing operational overhead while providing native integration with AWS services like Direct Connect, VPN, and Network Manager for comprehensive network visibility and control

Service Integration

Transit Gateway natively integrates with Amazon VPC for attaching virtual private clouds and managing routing. AWS Direct Connect integration enables high-bandwidth hybrid connectivity through Direct Connect gateways:

- Site-to-Site VPN connections provide secure tunnels to on-premises networks with BGP dynamic routing support
- AWS Resource Access Manager (RAM) enables cross-account sharing of transit gateways
- Transit Gateway Connect attachments integrate with third-party SD-WAN and network appliances using GRE tunnels and BGP
- CloudWatch provides monitoring metrics and Network Manager offers centralized network management and visualization

- Integration with Route 53 Resolver enables DNS resolution across connected networks, while AWS Network Firewall can be deployed for centralized security inspection in appliance mode configurations

Onboarding and Offboarding

Customers begin by creating a Transit Gateway in their desired AWS region through the AWS Management Console, CLI, or API, specifying basic configuration like ASN and enabling optional features like DNS support and multicast. Next, they create VPC attachments by selecting subnets in each VPC they want to connect, with Transit Gateway automatically creating elastic network interfaces:

- Route tables are configured to control traffic flow between attachments, with options for route propagation from VPN and Direct Connect connections
- For hybrid connectivity, customers establish VPN connections or Direct Connect gateway attachments
- Testing involves verifying connectivity between resources in different VPCs using standard network tools like ping or application-specific tests
- AWS provides comprehensive tutorials and best practice guides for common scenarios including centralized routing, isolated networks, and multi-region deployments

Getting Started Guide: https://docs.aws.amazon.com/vpc/latest/tgw/tgw-getting-started.html

Data Removal

Offboarding requires systematic detachment of all resources before deletion. First, delete all route table associations and route propagations, then detach VPCs, VPN connections, Direct Connect gateways, and peering connections. Each attachment must be fully detached and reach ‘deleted’ state before proceeding. Finally, delete the Transit Gateway itself, which automatically removes the default route table and any custom route tables. All routing configurations, flow logs, and associated metadata are permanently removed. For shared transit gateways, the owner must revoke resource shares through AWS RAM before deletion.

Backup/Restore and Disaster Recovery

Transit Gateway configurations are not directly backed up by AWS Backup service as it’s infrastructure-as-code. Disaster recovery relies on Infrastructure as Code (IaC) templates using CloudFormation, CDK, or Terraform to recreate configurations. Route table configurations, attachments, and policies should be documented and version-controlled. For multi-region disaster recovery, Transit Gateway peering enables cross-region connectivity with automatic failover through routing updates. Network configurations can

be replicated across regions using automated deployment pipelines. Flow logs and monitoring data can be archived to S3 for historical analysis and compliance requirements.

Backup Capabilities

Transit Gateway does not support traditional backup operations as it's a managed networking service with infrastructure configurations. The service maintains high availability through AWS's redundant infrastructure across multiple Availability Zones:

- Configuration backup is achieved through Infrastructure as Code practices, exporting route tables and attachment configurations to version-controlled templates
- Transit Gateway is not compatible with AWS Backup service since it doesn't store customer data but rather provides network routing services
- Recovery relies on recreating the configuration using saved IaC templates rather than restoring from backup snapshots

Disaster Recovery

Transit Gateway supports disaster recovery through multi-region peering connections that maintain encrypted connectivity across regions on the AWS backbone. It integrates with AWS Elastic Disaster Recovery by providing network connectivity for replicated resources across regions:

- Cross-region peering enables automatic traffic rerouting during regional failures
- The service supports disaster recovery architectures by connecting recovery VPCs to production networks through peered transit gateways
- Route propagation and dynamic routing enable automatic failover scenarios
- However, Transit Gateway itself is not directly managed by AWS Elastic Disaster Recovery service but serves as the network foundation for DR architectures

Recovery Objectives

Transit Gateway supports aggressive RTO and RPO objectives through its multi-Availability Zone architecture providing sub-second failover within regions. Cross-region peering enables RPO of near-zero for network connectivity with RTO measured in minutes for DNS-based failover. Dynamic routing protocols like BGP provide automatic path selection during failures. Network configurations can be restored rapidly through IaC deployment, typically within 5-15 minutes. For critical applications, multi-region active-active architectures using peered transit gateways can achieve RPO/RTO of seconds, while disaster recovery scenarios with pre-configured cross-region peering can meet most enterprise requirements for sub-hour recovery objectives.

Service Constraints

Service Quotas: https://docs.aws.amazon.com/vpc/latest/tgw/transit-gateway-quotas.html

FAQ: https://aws.amazon.com/transit-gateway/faqs/

Technical Requirements

Service Documentation: https://docs.aws.amazon.com/vpc/latest/tgw/

User Guide: https://docs.aws.amazon.com/vpc/latest/tgw/what-is-transit-gateway.html

API Reference: https://docs.aws.amazon.com/AWSEC2/latest/APIReference/OperationList-query-ec2.html

Pricing Overview

Please see the AWS UK G Cloud 15 Pricing Document accompanying this service in the Digital Marketplace.

Public Pricing: https://aws.amazon.com/transit-gateway/pricing/

Cost Optimization

Transit Gateway offers several unique cost optimization features including Flexible Cost Allocation (FCA) that enables custom allocation of data processing charges to source, destination, or central accounts based on defined policies. This addresses common scenarios where central IT teams bear costs for partner connections or security teams pay for inspection services:

- Cost allocation tags provide granular tracking of charges across multiple teams and projects
- VPN Concentrator attachments reduce costs for multiple low-bandwidth remote sites by consolidating up to 100 connections under a single attachment
- Strategic use of Transit Gateway Connect attachments eliminates data processing charges compared to standard VPC attachments
- Multi-region peering keeps inter-region traffic on AWS backbone, potentially reducing data transfer costs compared to internet routing

Savings Considerations

Primary cost savings come from architectural simplification, replacing hundreds of VPC peering connections with a single Transit Gateway, reducing both hourly attachment costs and operational overhead. Centralized NAT gateways and internet gateways reduce duplicate infrastructure across VPCs:

- Consolidated Direct Connect and VPN connections serve multiple VPCs through single attachments rather than per-VPC connections
- Data processing charges can be optimized through careful route table design to minimize unnecessary traffic traversal
- For multi-account environments, shared Transit Gateway through AWS RAM eliminates duplicate gateways per account
- Long-term savings include reduced networking operational complexity, fewer required networking specialists, and simplified troubleshooting through centralized routing and monitoring
- Strategic placement of Transit Gateway in hub accounts optimizes cross-account data charges through careful cost allocation policy design