

Service Name

AWS Clean Rooms

Service Overview

AWS Clean Rooms is a secure collaboration workspace that enables organizations to analyze and collaborate on their collective datasets with partners without revealing underlying data. The service allows users to create clean rooms in minutes, invite partners, link data resources, and configure privacy-enhancing controls. Organizations can run SQL queries, PySpark jobs, and machine learning models on combined datasets while maintaining data privacy through built-in access controls, analysis rules, cryptographic computing, audit logs, and differential privacy. Data remains in its original location without requiring movement to another cloud provider.

Key Use Cases

- **Advertising and Marketing:** Improve advertiser reach and publisher monetization through secure collaboration on campaign data and audience insights
- **Machine Learning and Analytics:** Train custom ML models and create lookalike segments using collective datasets for fraud detection, campaign optimization, and medical research
- **Entity Resolution:** Match and link customer records from various sources using identity resolution partners like LiveRamp and Neustar to generate common IDs
- **Cross-industry Data Collaboration:** Enable secure data sharing for research and development, investment decisions, and multi-party analytics across healthcare, finance, and other regulated industries

Features

- **Multi-party Collaboration:** Create collaborations with multiple organizations and configure member roles for secure data sharing
- **SQL and PySpark Analytics:** Run SQL queries and PySpark jobs on collective datasets with Spark SQL support and configurable compute sizes
- **Privacy-Enhancing Controls:** Implement fine-grained analysis rules, differential privacy, and cryptographic computing to protect sensitive data
- **Cross-Cloud Support:** Collaborate with datasets from multiple clouds including AWS, Snowflake, and Amazon Athena
- **Machine Learning Capabilities:** AWS Clean Rooms ML for custom modeling, lookalike modeling, and synthetic dataset generation
- **Entity Resolution:** AWS Entity Resolution integration for matching customer records across different data sources

- **Cross-Region Collaboration:** Support for collaboration with data sources stored in different AWS and Snowflake regions
- **Analysis Templates:** Pre-approved SQL and Python queries that can be reused across collaborations

Value Proposition

AWS Clean Rooms provides unique value through its comprehensive privacy-preserving data collaboration platform that eliminates the need to move or share raw data.

Organizations benefit from IDC MarketScape recognition as a leader in worldwide data clean room technology, extensive privacy controls including differential privacy and cryptographic computing, seamless integration with existing AWS services and third-party platforms like Snowflake, and the ability to collaborate with thousands of companies already using AWS. The service offers flexibility through multiple analytics engines (SQL, PySpark), supports both custom and lookalike ML modeling, and provides cost-effective collaboration where compute costs can be shared between partners.

Service Integration

AWS Clean Rooms integrates deeply with core AWS services to provide comprehensive data collaboration capabilities. Key integrations include Amazon S3 for data storage, AWS Glue Data Catalog for data preparation and cataloging, Amazon Athena for serverless querying, AWS CloudFormation for infrastructure as code deployment, AWS CloudTrail for audit logging, AWS Entity Resolution for identity matching, AWS Secrets Manager for secure credential storage, Amazon EventBridge for event-driven workflows, and AWS KMS for encryption key management. The service also works with Amazon SageMaker for ML model deployment, Amazon QuickSight for visualization, and AWS Backup for data protection, creating a complete data collaboration ecosystem within the AWS cloud.

Onboarding and Offboarding

Customers get started with AWS Clean Rooms by first signing up for an AWS account and setting up necessary service roles for AWS Clean Rooms and AWS Clean Rooms ML using the AWS Management Console, CLI, or CloudFormation. The onboarding process involves creating a collaboration or joining an existing one through membership creation, configuring data sources by linking AWS Glue tables, Amazon S3 data, or Snowflake sources, setting up analysis rules and privacy controls, and optionally creating analysis templates for pre-approved queries. The service provides guided workflows through AWS Clean Rooms Solutions to simplify data preparation and collaboration setup, with comprehensive documentation and expert support available.

Getting Started Guide: <https://docs.aws.amazon.com/clean-rooms/latest/userguide/setting-up.html>

Data Removal

During offboarding, collaboration creators and members can delete their collaborations, memberships, and associated resources through the AWS Clean Rooms console or API. Data removal involves deleting configured tables, analysis templates, and any query results stored in Amazon S3. Organizations retain control over their original data sources, which remain in their own AWS accounts and are never moved to AWS Clean Rooms infrastructure, ensuring complete data sovereignty during the offboarding process.

Backup/Restore and Disaster Recovery

AWS Clean Rooms leverages underlying AWS service backup capabilities since data remains in original locations like Amazon S3 and AWS Glue. Organizations can use AWS Backup for centralized backup management, Amazon S3 versioning and replication for data protection, and cross-region data sources for geographic redundancy. The service metadata and configurations can be recreated using AWS CloudFormation templates for disaster recovery scenarios.

Backup Capabilities

AWS Clean Rooms does not provide native backup capabilities as it operates on data stored in external sources. Backup responsibilities lie with the underlying data storage services like Amazon S3, which offers versioning, replication, and integration with AWS Backup. Organizations should implement backup strategies for their source data systems and use AWS CloudFormation templates to backup collaboration configurations and analysis templates.

Disaster Recovery

AWS Clean Rooms supports disaster recovery through its distributed architecture across multiple AWS regions and integration with resilient AWS services. The service supports cross-region collaboration, allowing data sources in different regions to participate in the same collaboration. Recovery strategies can leverage AWS CloudFormation for infrastructure recreation, Amazon S3 cross-region replication for data availability, and the service's stateless query execution model for rapid recovery.

Recovery Objectives

AWS Clean Rooms helps customers meet RPO and RTO objectives through its integration with highly available AWS services and cross-region support. The service's architecture allows for quick collaboration recreation using CloudFormation templates, while underlying data services like Amazon S3 provide sub-second RPO through cross-region replication. RTO objectives can be met through automated failover to alternate regions and the service's ability to quickly recreate collaboration environments without data movement requirements.

Service Constraints

Service Quotas: <https://docs.aws.amazon.com/clean-rooms/latest/userguide/quotas.html>

FAQ: <https://aws.amazon.com/clean-rooms/faqs/>

Technical Requirements

Service Documentation: <https://docs.aws.amazon.com/clean-rooms/latest/userguide/what-is.html>

User Guide: <https://docs.aws.amazon.com/clean-rooms/latest/userguide/>

API Reference: <https://docs.aws.amazon.com/clean-rooms/latest/apireference/Welcome.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/clean-rooms/pricing/>

Cost Optimization

AWS Clean Rooms offers unique cost optimization through its pay-per-query model based on Clean Rooms Processing Units (CRPUs), allowing organizations to pay only for actual compute usage. Cost optimization features include configurable compute sizes for SQL queries, ability to cache tables for complex queries to reduce costs, customizable Spark properties for performance tuning, and shared cost model where collaboration creators can designate which member pays for compute costs. The service also supports advanced configurations to optimize SQL performance and reduce processing time, directly impacting costs.

Savings Considerations

Main cost savings come from eliminating data movement and duplication costs since data remains in original locations, shared compute cost model where partners can split collaboration expenses, efficient query execution through optimized Spark SQL engine and configurable compute resources, and elimination of traditional ETL processes for data collaboration. Organizations save on infrastructure costs by leveraging existing AWS data storage and avoiding dedicated data clean room infrastructure. The pay-per-use model ensures costs scale with actual usage rather than requiring upfront capacity planning or reserved resources.