

Service Name

Amazon Kendra

Service Overview

Amazon Kendra is a managed information retrieval and intelligent search service that uses natural language processing and advanced deep learning models for semantic and contextual search. It allows users to connect multiple data repositories to an index and ingest and crawl documents to create a unified search experience. The service uses machine learning algorithms to understand the context of queries and content, providing accurate and relevant results including specific answers to questions, not just search results. Kendra supports factoid, descriptive, and keyword-based natural language queries across enterprise content scattered across multiple locations and repositories.

Key Use Cases

- **Enterprise search:** Searching across internal documents, websites and applications when content is scattered across multiple locations and repositories
- **Knowledge management:** Building AI-powered assistants for financial planning, analysis, and data discovery to support analysts with business insights
- **Customer support:** Creating intelligent Q&A systems and chatbots that can provide specific answers from knowledge bases and documentation
- **Retrieval Augmented Generation (RAG):** Integrating with Amazon Bedrock Knowledge Bases and Amazon Q Business to create RAG-powered digital assistants

Features

- **GenAI Enterprise Edition:** Delivers highest accuracy for RAG use cases with advanced semantic models and information retrieval technologies
- **Intelligent Search:** Machine learning algorithms understand context of queries and content to provide accurate and relevant results
- **Data Source Connectors:** Native connectors for 30+ data sources including S3, SharePoint, Confluence, Salesforce, and databases
- **Experience Builder:** No-code interface to create and customize search experiences with filtering, FAQs, and sorting capabilities
- **Custom Document Enrichment:** Extract metadata and entities from documents using Amazon Comprehend for improved search accuracy
- **User Access Control:** Metadata-based user permissions filtering and integration with AWS IAM Identity Center
- **Query Suggestions:** Autocompletion suggestions and spell checking to help users refine queries

- **Search Analytics Dashboard:** Monitor search usage and performance metrics to optimize search experiences

Value Proposition

Amazon Kendra provides enterprise-grade intelligent search capabilities that eliminate the complexity of building and maintaining search infrastructure. Unlike traditional keyword-based search engines, Kendra uses advanced machine learning to understand natural language queries and provide direct answers, not just document links:

- The service offers pre-built connectors to 30+ enterprise data sources, reducing integration effort
- With features like GenAI Enterprise Edition optimized for RAG applications, seamless AWS service integration, and managed infrastructure that scales automatically, organizations can deploy production-ready intelligent search solutions without the overhead of managing search algorithms, relevance tuning, or infrastructure scaling

Service Integration

Amazon Kendra integrates deeply with core AWS services including Amazon S3 for document storage and indexing, Amazon Bedrock for generative AI capabilities and Knowledge Bases, Amazon Q Business for enterprise AI assistants, AWS IAM Identity Center for user access control, Amazon Comprehend for document enrichment and entity extraction, AWS Secrets Manager for secure credential storage, Amazon VPC for network isolation, AWS Lambda for custom processing, and Amazon CloudWatch for monitoring and logging. The service also works with AWS data services like Amazon RDS, DynamoDB, and Aurora for database content indexing, providing a comprehensive ecosystem for enterprise search and AI applications.

Onboarding and Offboarding

Customers get started by creating an AWS account and setting up an Amazon Kendra index through the AWS console, CLI, or SDKs. The process involves selecting an index type (GenAI Enterprise, Basic Enterprise, or Developer Edition), configuring data sources using pre-built connectors, and setting up IAM roles for access permissions:

- Amazon Kendra provides a free trial for GenAI Enterprise and Developer Edition indexes for the first 30 days
- Getting started guides are available for various data sources including Amazon S3, MySQL databases, and AWS IAM Identity Center, with step-by-step instructions for console, CLI, Python SDK, and Java SDK implementations

Getting Started Guide: <https://docs.aws.amazon.com/kendra/latest/dg/getting-started.html>

Data Removal

Amazon Kendra provides data removal capabilities through index deletion and data source management. Customers can delete individual documents, remove entire data sources, or delete complete indexes to remove all associated data. When deleting an index, all documents, metadata, and associated resources are permanently removed. Data source connectors can be configured to track deletions during sync operations, automatically removing documents from the index when they are deleted from source systems. For complete offboarding, customers delete all indexes, data sources, and associated IAM roles.

Backup/Restore and Disaster Recovery

Amazon Kendra does not provide explicit backup and disaster recovery capabilities as a managed service feature. The service relies on AWS infrastructure resilience and availability zones for high availability. Data durability is maintained through AWS's underlying storage systems. For disaster recovery, customers must implement their own backup strategies by maintaining copies of source documents in durable storage like Amazon S3, and re-indexing data in alternate regions if needed. Index configurations and metadata can be recreated using Infrastructure as Code approaches.

Backup Capabilities

Amazon Kendra does not have built-in backup capabilities and does not integrate with AWS Backup service. The managed service maintains data durability through AWS infrastructure, but customers are responsible for backing up their source content and index configurations. Organizations typically implement backup strategies at the source data level using services like AWS Backup for supported data sources, and maintain Infrastructure as Code templates to recreate index configurations and data source connections if needed.

Disaster Recovery

Amazon Kendra does not integrate with AWS Elastic Disaster Recovery service. Disaster recovery must be implemented through manual processes including maintaining source data backups, using multi-region index deployment strategies, and recreating indexes in alternate regions during outages. The service provides high availability within regions through AWS infrastructure, but cross-region disaster recovery requires customer-implemented solutions involving data replication and index recreation in secondary regions.

Recovery Objectives

Amazon Kendra supports recovery objectives primarily through high availability within AWS regions using multiple availability zones. For achieving specific RPO and RTO targets, customers must implement their own strategies including maintaining real-time data

source replication, automated index recreation processes, and traffic routing between regions. The service's managed nature provides operational resilience, but meeting strict recovery objectives requires combining Kendra with other AWS services like Route 53 for DNS failover and automated deployment pipelines.

Service Constraints

Service Quotas: <https://docs.aws.amazon.com/kendra/latest/dg/quotas.html>

FAQ: <https://aws.amazon.com/kendra/faqs/>

Technical Requirements

Service Documentation: <https://docs.aws.amazon.com/kendra/latest/dg/what-is-kendra.html>

User Guide: <https://docs.aws.amazon.com/kendra/latest/dg/getting-started.html>

API Reference: <https://docs.aws.amazon.com/kendra/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/kendra/pricing/>

Cost Optimization

Amazon Kendra offers unique cost optimization through multiple index editions with different pricing tiers. GenAI Enterprise Edition provides lower starting costs at \$0.32/hour with smaller capacity units optimized for RAG applications:

- Organizations can optimize costs by selecting appropriate index types based on use cases, implementing efficient data source sync schedules to minimize connector charges, right-sizing storage and query capacity units based on actual usage patterns, and using the 30-day free tier for evaluation and testing
- Cost optimization also includes optimizing document sizes and extraction processes to stay within capacity limits

Savings Considerations

Main cost savings come from selecting the right index edition for specific needs, with GenAI Enterprise Edition offering 75% lower hourly costs compared to Basic Enterprise Edition. Organizations save through efficient connector management with flat monthly rates for GenAI Edition versus per-document scanning charges in Basic Edition:

- Additional savings include optimizing query patterns to reduce capacity unit requirements, implementing effective document lifecycle management to control storage costs, and using regional availability strategically
- The service's managed nature eliminates infrastructure and operational costs associated with self-hosted search solutions, providing overall cost efficiency for enterprise search implementations