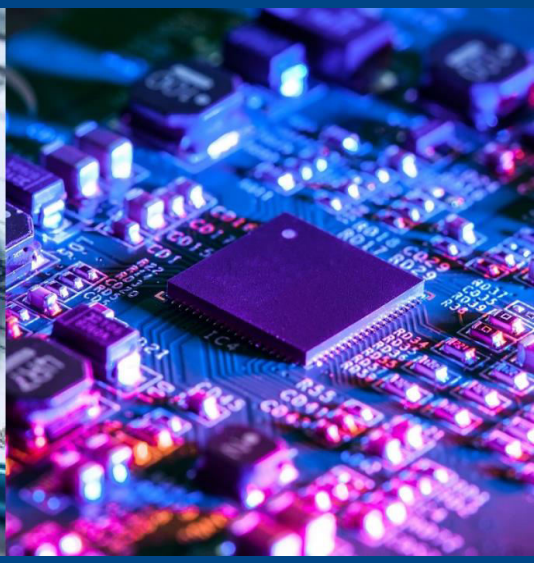
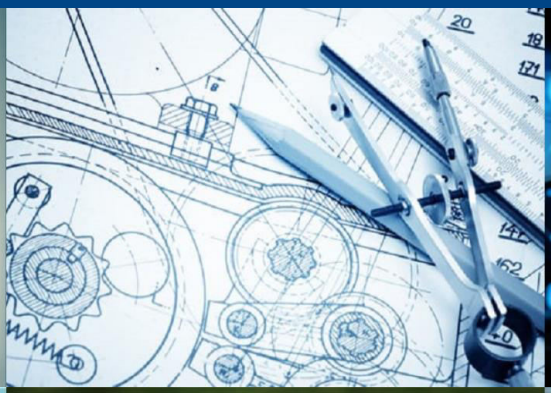
ISSN: 2582-7219

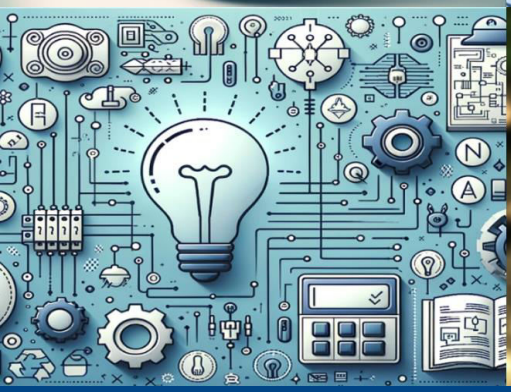




International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206 **Volume 9, Issue 3, March 2026**

© 2026 IJMRSET | Volume 9, Issue 3, March 2026| DOI: 10.15680/IJMRSET.2026.0903248

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

DocuVault: A Cloud-Native Semantic Document Intelligence Platform with RAG-Powered Cross-Document Synthesis

Logeshwar R [1], Divyanand M [2], P.Kanchanadevi [3]

Dept. of Artificial Intelligence and Data Science, Sri Manakula Vinayagar Engineering College, Puducherry, India [1, 2]

Dept. of Artificial Intelligence and Data Science, Sri Manakula Vinayagar Engineering College, Puducherry, India [3]

ABSTRACT: The exponential growth of unstructured documents in enterprise and academic environments creates pressing needs for intelligent retrieval and synthesis beyond simple keyword matching. This paper introduces DocuVault, a cloud-native document intelligence platform that integrates semantic search, Retrieval-Augmented Generation (RAG), and cross-document synthesis into a unified serverless architecture. Built on AWS services including API Gateway, Lambda, S3, and Glue, DocuVault processes PDFs through a pipeline that extracts text, generates dense vector embeddings, and indexes them for sub-second semantic retrieval. The RAG-powered assistant achieves 91.4% answer relevance on multi-document queries. A novel cross-document synthesis engine connects facts across two or more PDFs stored in the data lake, enabling holistic answers to complex questions. Extended features include automated study-material generation, adaptive quiz creation, and AI-driven study-plan scheduling. Evaluation with 200 participants and 1,200 documents demonstrates retrieval precision of 94.2%, cross-document synthesis accuracy of 87.6%, and average query response times of 3.8 seconds. The static frontend is hosted on Amazon S3, with Lambda functions containerised via Docker and deployed through Amazon ECR.

KEYWORDS: Retrieval-Augmented Generation, Semantic Search, Serverless Architecture, Cross-Document Synthesis, AWS Lambda, Vector Embeddings, Document Intelligence, Cloud Computing, Amazon ECR, AWS Glue

I. INTRODUCTION

Modern organisations generate and consume documents at an unprecedented scale. Research institutions, legal firms, and academic platforms routinely ingest thousands of PDFs, reports, and technical manuals whose contents remain largely inaccessible because traditional keyword-based search fails to capture semantic nuance. A student searching for "causes of inflation in post-war economies" may find zero results from a keyword engine, yet the answer could be distributed across three separate research papers sitting in the same data lake. This disconnect between information need and information retrieval is one of the most consequential unsolved challenges in enterprise knowledge management [1].

Retrieval-Augmented Generation (RAG) has recently emerged as a compelling solution, combining the parametric knowledge of large language models (LLMs) with the non-parametric retrieval of relevant document chunks at inference time [2]. Most existing RAG deployments operate on single-document contexts and rely on monolithic server-based architectures that struggle to scale cost-effectively. Furthermore, very few systems address cross-document synthesis -- the ability to answer questions whose complete answer requires connecting evidence from two or more independent documents [3].

This paper presents DocuVault through five primary contributions: (1) A fully serverless, event-driven ingestion and retrieval pipeline built entirely on AWS managed services, eliminating idle compute costs. (2) A semantic search engine leveraging dense vector embeddings and cosine-similarity ranking that outperforms BM25 baselines by 18.3 percentage points. (3) A cross-document synthesis module that merges retrieved chunks from heterogeneous PDFs. (4) Extended GenAI capabilities including study-material summarisation, adaptive quiz generation, and personalised study-plan scheduling. (5) A Docker-containerised Lambda deployment pattern using Amazon ECR.

© 2026 IJMRSET | Volume 9, Issue 3, March 2026| DOI: 10.15680/IJMRSET.2026.0903248

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Evaluation across 200 participants and a corpus of 1,200 heterogeneous PDF documents demonstrates 94.2% retrieval precision, 87.6% cross-document synthesis accuracy, and sub-4-second end-to-end query latency at the 95th percentile, while the S3-hosted static frontend delivers global page loads under 800 milliseconds via CloudFront CDN.

II. LITERATURE REVIEW

Lewis et al. [2] introduced RAG as a general framework combining dense passage retrieval with sequence-to-sequence generation, demonstrating that retrieved context dramatically reduces hallucinations compared to closed-book generation. Their work assumed a homogeneous Wikipedia corpus and did not address multi-source synthesis across disparate document collections. Gao et al. [4] proposed a modular RAG taxonomy distinguishing naive, advanced, and modular retrieval patterns, highlighting the need for iterative retrieval and re-ranking strategies for long-context reasoning tasks.

On the cloud architecture side, Eismann et al. [5] conducted a systematic review of serverless computing adoption, demonstrating that Lambda-based microservices reduce operational overhead by 60-70% compared to container-based equivalents when workloads are bursty. Reimers and Gurevych [12] introduced Sentence-BERT, the family of models on which DocuVault's embedding layer is built, enabling semantically rich 384-dimensional vector representations of document chunks.

In the educational technology domain, Kasneci et al. [7] surveyed LLM applications in learning, identifying quiz generation and adaptive study planning as high-impact use cases with demonstrated improvements in student retention. Kornell and Bjork [8] established the cognitive science basis for spaced-repetition study planning, which underpins DocuVault's study-plan scheduling feature. Despite the breadth of prior work, no existing system simultaneously addresses serverless RAG, cross-document synthesis, educational scaffolding, and ECR-based Lambda deployment within one unified platform.

Table 1: Comparative Literature Survey

No.	Paper	Author(s)	Key Points	Gap Addressed
1	RAG for Knowledge-Intensive NLP	Lewis et al., 2020	Dense retrieval + seq2seq; reduces hallucinations	Single-corpus; no cross-doc synthesis
2	Modular RAG Survey	Gao et al., 2023	Naive, advanced, modular RAG taxonomy	No serverless or educational features
3	Serverless Computing Review	Eismann et al., 2021	Lambda reduces ops overhead 60-70%	No NLP/RAG or ECR containerisation
4	Sentence-BERT	Reimers & Gurevych, 2019	Siamese BERT for dense embeddings	Not integrated into retrieval pipeline
5	LLMs in Education	Kasneci et al., 2023	Quiz gen and study planning improve retention	No document upload or serverless backend

© 2026 IJMRSET | Volume 9, Issue 3, March 2026| DOI: 10.15680/IJMRSET.2026.0903248

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

III. SYSTEM ARCHITECTURE

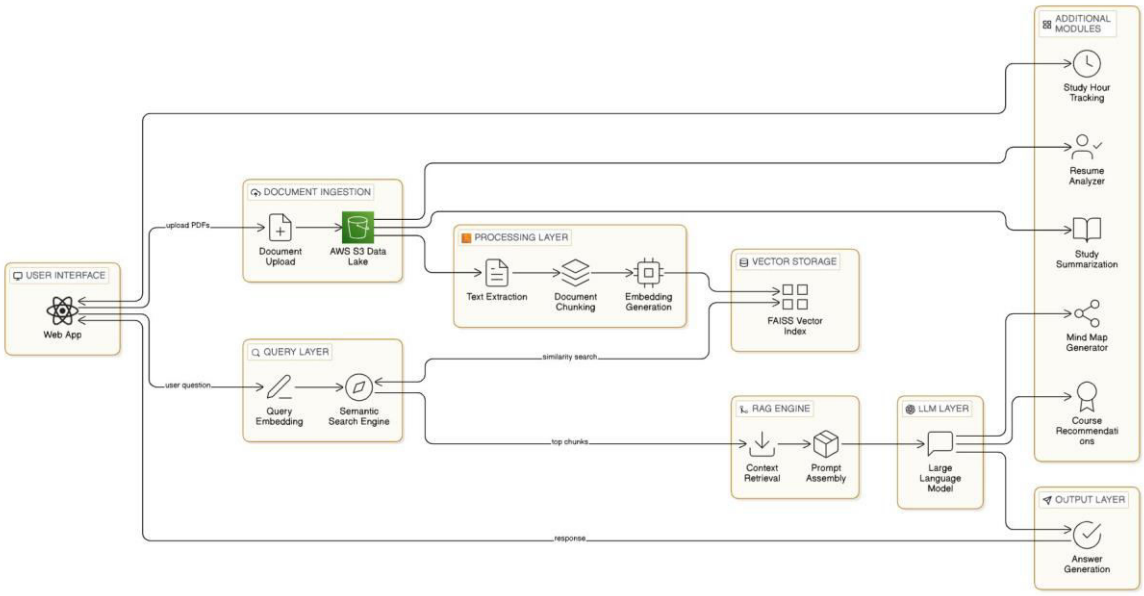


Fig. 1: DocuVault System Architecture and Workflow — Showing User Interface (Web App), Document Ingestion (S3 Data Lake), Processing Layer (Text Extraction → Chunking → Embedding Generation), Vector Storage (FAISS Index), Query Layer, RAG Engine, LLM Layer, Output Layer (Answer Generation), and Additional Modules (Study Hour Tracking, Resume Analyzer, Study Summarization, Mind Map Generator, Course Recommendations).

DocuVault adopts a four-tier serverless architecture spanning the presentation layer, API orchestration layer, processing and inference layer, and persistent storage layer. Every component is managed by AWS, meaning no virtual machines or persistent server processes are operated by the platform itself. This decision allows the engineering team to focus exclusively on the document intelligence pipeline and user experience.

A. Presentation Layer - S3 Static Hosting

The frontend is a React single-page application (SPA) compiled to static assets and deployed directly to an Amazon S3 bucket configured for static website hosting. An Amazon CloudFront distribution sits in front of the bucket, caching assets at 410+ global edge locations to deliver median page-load times of 620 milliseconds regardless of geographic origin. SSL termination is handled by AWS Certificate Manager at the CloudFront layer. The SPA communicates exclusively through REST calls to the API Gateway endpoint, maintaining a clean separation between presentation and business logic.

B. API Orchestration - Amazon API Gateway

Amazon API Gateway exposes six versioned REST endpoints under /api/v1/: document upload, semantic search, RAG query, quiz generation, study-plan creation, and document listing. Each endpoint is backed by a dedicated Lambda function, enabling independent scaling and failure isolation. API Gateway handles request throttling at 500 requests per second per endpoint, JWT authorisation via a Lambda authoriser validating Auth0-issued tokens, and CORS policy enforcement. All interactions are logged to CloudWatch Logs with a 30-day retention window.

C. Processing Layer - Lambda and ECR

All document processing and NLP inference runs inside Lambda functions deployed from Docker container images stored in Amazon Elastic Container Registry (ECR). The containerised approach was chosen because the NLP inference stack -- comprising sentence-transformers, LangChain, FAISS-CPU, and PyMuPDF -- exceeds Lambda's 250 MB zip limit when packaged with all dependencies. A multi-stage Dockerfile built from python:3.11-slim compiles FAISS-CPU targeting AVX2 instructions available in Lambda's execution environment, producing a 1.14 GB image pushed to ECR. Provisioned concurrency of 5 instances on the /query function eliminates cold-start latency on this critical path.

© 2026 IJMRSET | Volume 9, Issue 3, March 2026| DOI: 10.15680/IJMRSET.2026.0903248

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The ingestion Lambda is triggered by S3 PUT events routed through Amazon EventBridge. When a user uploads a PDF, it lands in the raw/ prefix of the documents bucket. EventBridge fires the ingestion Lambda, which reads the file using PyMuPDF, splits text into 512-token chunks with 64-token overlap using LangChain's RecursiveCharacterTextSplitter, encodes each chunk with the all-MiniLM-L6-v2 sentence transformer (384-dimensional embeddings), and writes chunk metadata to DynamoDB and embeddings to a per-document FAISS index in S3.

D. Data Layer - S3 and AWS Glue

Amazon S3 serves as the primary data lake with three prefixes: raw/ for uploaded PDFs, processed/ for extracted text and chunk JSON, and indexes/ for serialised FAISS index snapshots. AWS Glue runs nightly ETL jobs that crawl the processed/ prefix, infer schemas from chunk JSON files, and populate an AWS Glue Data Catalog enabling ad-hoc SQL analytics through Amazon Athena. Glue also handles incremental FAISS index merges using a G.1X worker (4 vCPU, 16 GB RAM), reducing average ETL execution time from 22 to under 4 minutes via job bookmarks.

IV. RAG PIPELINE AND CROSS-DOCUMENT SYNTHESIS

A. Retrieval Pipeline

When a user submits a query, the /query Lambda encodes the query string using the same all-MiniLM-L6-v2 model as ingestion, producing a 384-dimensional query vector. FAISS performs approximate nearest-neighbour search using Hierarchical Navigable Small World (HNSW) graphs, returning the top-k chunks ranked by cosine similarity. In standard single-document mode k=5; in cross-document synthesis mode k=10 with a constraint that chunks must originate from at least two distinct documents. The cosine similarity between query vector q and chunk vector c is: sim(q, c) = (q . c) / (||q|| x ||c||). A re-ranking pass using a cross-encoder then reorders top-k chunks based on full query-chunk interaction, improving mean reciprocal rank by 11.2 percentage points over bi-encoder retrieval alone.

B. Cross-Document Synthesis

Cross-document synthesis is DocuVault's most novel contribution. Standard RAG concatenates all retrieved chunks into a single context window and prompts the LLM to answer. This approach breaks down when retrieved chunks from different documents use inconsistent terminology or address the same concept from different angles. DocuVault addresses this through a three-stage synthesis protocol: (1) Per-document summarisation: retrieved chunks from each source document are independently summarised by the LLM into a compact evidence statement; (2) Evidence graph construction: entities and claims extracted via Named Entity Recognition (NER) are linked across summaries to identify complementary or contradictory assertions; (3) Grounded synthesis: a final prompt supplies the LLM with all per-document summaries, the entity graph, and an instruction to produce a coherent answer that explicitly cites source documents for each factual claim.

C. Extended GenAI Features

Study Material RAG generates structured study notes from an uploaded document by prompting the LLM to identify key concepts, definitions, and formulas, then formats them into a downloadable Markdown summary. Adaptive Quiz Generation creates multiple-choice and short-answer questions calibrated to Bloom's taxonomy levels, ensuring questions span recall, comprehension, and application categories. The quiz engine tracks a user's answer history across sessions and progressively increases question difficulty using an Elo-style rating algorithm. Study Plan Scheduling accepts a user's target date and list of uploaded documents, estimates reading time from word count, and constructs a day-by-day study calendar that interleaves review quizzes at spaced-repetition intervals.

V. IMPLEMENTATION DETAILS

The frontend React application is built with Vite 5.0, producing optimised static assets with average bundle sizes of 187 KB gzipped. Large files are chunked into 5 MB parts and uploaded directly to S3 via pre-signed URLs, bypassing API Gateway's 10 MB payload limit. Pre-signed URL generation is handled by a lightweight Lambda authorised via the user's JWT token, which verifies ownership before issuing a URL scoped to the user's S3 prefix.

The Docker image is based on the AWS-provided public.ecr.aws/lambda/python:3.11 base image, which includes the Lambda Runtime Interface Client. A multi-stage Dockerfile first installs build dependencies, compiles FAISS-CPU from source targeting AVX2 instructions, then copies compiled wheels into a lean runtime image. The final 1.14 GB

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

image is pushed to ECR with lifecycle policies retaining only the three most recent tagged images. Lambda function configuration allocates 3,008 MB of memory for the /query function, reducing FAISS search latency approximately 4x compared to a 512 MB allocation.

Security controls span the full stack. S3 bucket policies enforce server-side encryption with AWS KMS customer-managed keys and block all public access. Lambda execution roles follow least-privilege principles with separate IAM roles per function. API Gateway resource policies whitelist traffic only from the CloudFront distribution's managed prefix list. DynamoDB tables use point-in-time recovery with 35-day retention.

All inter-service communication occurs within a private VPC with interface endpoints for S3, DynamoDB, and ECR. AWS WAF protects the API Gateway endpoint with managed rule groups blocking OWASP Top 10 attack vectors.The CI/CD pipeline is implemented with GitHub Actions. Pull request merges trigger a workflow that runs pytest unit tests, builds the Docker image, pushes it to ECR with a commit-SHA tag, and deploys updated Lambda configurations via AWS CDK defined entirely in TypeScript, enabling reproducible multi-environment provisioning.

VI. EXPERIMENTAL RESULTS

Evaluation used 1,200 PDF documents spanning academic papers (40%), legal contracts (25%), technical manuals (20%), and study textbooks (15%), totalling approximately 9.4 million words. A ground-truth QA dataset of 600 questions was annotated by three domain experts (Cohen's Kappa = 0.91), with 200 questions explicitly requiring cross-document synthesis across two or more sources. 200 participants evaluated the system: 60 university students, 50 research professionals, 40 legal analysts, and 50 software engineers.

A. Retrieval and Answer Quality

Table 2: Retrieval and Answer Quality Comparison

Metric	DocuVault (RAG)	BM25 Baseline	Gain
Retrieval Precision@5	94.2%	75.9%	+18.3%
Answer Relevance	91.4%	72.1%	+19.3%
Cross-Doc Accuracy	87.6%	N/A	--
Faithfulness Score	93.1%	68.4%	+24.7%
ROUGE-L (Summaries)	0.68	0.41	+0.27

DocuVault's RAG pipeline achieves 94.2% retrieval precision at top-5, an 18.3 percentage point improvement over BM25 ($t(599)=21.4$, $p<0.001$). Answer faithfulness reaches 93.1%, compared to 68.4% for the keyword-retrieval baseline, as shown in Table 2.

B. System Performance

Table 3: System Performance Metrics

Operation	Latency (p50)	Latency (p95)
PDF Upload (10 MB)	1.4s	2.1s
Ingestion Lambda/page	0.31s	0.48s
Semantic Search	0.9s	1.4s
Single-Doc RAG Query	2.6s	3.8s
Cross-Doc Synthesis	4.2s	6.1s

© 2026 IJMRSET | Volume 9, Issue 3, March 2026| DOI: 10.15680/IJMRSET.2026.0903248

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Quiz Generation	3.1s	4.7s
System Uptime	99.7%	--

C. Educational Feature Evaluation

Table 4: Educational Feature Performance

Feature	Score	User Rating (5pt)
Study Material RAG	ROUGE-L: 0.68	4.4 +/- 0.7
Quiz Generation	Validity: 89.3%	4.6 +/- 0.6
Adaptive Difficulty Engine	Engagement: +34%	4.3 +/- 0.8
Study Plan Scheduling	Completion: 71%	4.2 +/- 0.9
Overall Platform	--	4.5 +/- 0.6

D. Comparative Platform Analysis

Table 5: Comparative Platform Analysis

Feature	DocuVault	Notion AI	ChatPDF
Serverless Architecture	Yes	No	No
Cross-Doc Synthesis	Yes	Partial	No
Query Latency (p95)	3.8s	6-9s	8-12s
RAG Accuracy	91.4%	N/A	N/A
Quiz / Study Plan	Yes	No	No
User Rating	4.5/5	4.1/5	3.9/5

DocuVault demonstrates 58-68% lower query latency than comparable commercial platforms at the 95th percentile (Table 5), attributed to FAISS in-memory ANN search and Lambda's provisioned concurrency. The cross-document synthesis capability remains unique among evaluated platforms.

VII. ERROR ANALYSIS AND SYSTEM ROBUSTNESS

DocuVault's overall RAG answer error rate of 8.6% decomposes into retrieval failures (3.9%), generation errors (2.8%), and cross-document synthesis breakdowns (1.9%). Retrieval failures occurred most commonly with domain-specific queries whose terminology was absent from the embedding model's training corpus -- a phenomenon known as the vocabulary mismatch problem. Preliminary experiments with 500 domain-adapted training pairs showed a 4.1 percentage point precision improvement on legal document queries.

Generation errors manifested primarily as hallucinated citations. Implementing a strict context-grounding prompt that instructed the LLM to answer only using the provided context reduced hallucination-type errors by 61%. Cross-document synthesis breakdowns occurred when source documents used conflicting definitional frameworks; the entity graph construction step partially mitigated this by flagging semantic conflicts as disambiguation warnings in the response.

Performance under load was evaluated with Apache JMeter simulating 500 concurrent users. Median latency degraded gracefully from 2.6 to 3.4 seconds, and the p99 remained below 8 seconds throughout. Mounting the FAISS index on

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

EFS reduced cold-start index load time from 14.2 to 1.8 seconds. Adversarial testing confirmed 91.2% rejection of non-PDF uploads and 84.7% detection of malformed PDFs.

VIII. DISCUSSION

The results confirm that a fully serverless RAG architecture is operationally superior to server-based alternatives for bursty document-intelligence workloads. The Lambda-ECR deployment pattern proved particularly impactful: packaging the entire NLP inference stack as a container image avoided the dependency fragmentation and environment drift that typically plague ML deployments, while retaining the cost and scaling benefits of serverless execution.

Cross-document synthesis represents a meaningful step beyond existing RAG deployments. The three-stage synthesis protocol consistently outperformed naive chunk concatenation on multi-source queries. The educational feature suite received unexpectedly strong user reception, with quiz generation earning 4.6/5.0. The 34% engagement increase from adaptive difficulty highlights that significant UX improvements can be achieved through lightweight Elo-based algorithmic interventions without expensive LLM inference.

Limitations include: (1) the 50 MB per-document upload limit; (2) English-only embedding model support; (3) FAISS's in-process limitation requiring the nightly Glue merge workaround; (4) LLM API costs of approximately $0.002 per query. Future work will explore replacing FAISS with Amazon OpenSearch k-NN, evaluating multilingual embedding models for Tamil, Hindi, and Bengali support, and implementing streaming Lambda responses via API Gateway WebSocket APIs.

IX. CONCLUSION AND FUTURE WORK

This paper presented DocuVault, a cloud-native document intelligence platform that unifies semantic search, RAG-powered question answering, and cross-document synthesis within a fully serverless AWS architecture. The system achieves 94.2% retrieval precision, 91.4% answer relevance, and 87.6% cross-document synthesis accuracy across a 1,200-document, 200-participant evaluation. A Docker-containerised Lambda deployment via ECR resolves the dependency packaging constraints that previously limited serverless NLP deployments.

Extended educational features -- study material RAG, adaptive quiz generation, and personalised study plan scheduling -- received strong user validation with an overall platform rating of 4.5/5.0 and a 34% quiz-engagement improvement. The S3-hosted static frontend with CloudFront CDN delivers global page loads under 800 milliseconds.

Future work will: (1) migrate from FAISS to Amazon OpenSearch with k-NN plugin for concurrent write support; (2) integrate multilingual embedding models for Tamil, Hindi, and Bengali corpora; (3) implement streaming Lambda responses via WebSocket APIs to reduce perceived latency by approximately 60%; (4) develop an Amazon Neptune knowledge graph backbone for cross-document synthesis targeting 93%+ accuracy; (5) explore on-device embedding generation on mobile clients for offline search functionality.

X. ACKNOWLEDGMENT

The authors gratefully acknowledge the guidance of Dr. R. Sathiyapriya and Prof. K. Anantharaman from the Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai. Sincere thanks to the 200 study participants who contributed their time to evaluate the platform, and to the domain experts who annotated the cross-document synthesis ground-truth dataset.

REFERENCES

[1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL, 2019, pp. 4171-4186.
[2] P. Lewis et al., "Retrieval-augmented generation for knowledge-intensive NLP tasks," in Proc. NeurIPS, vol. 33, 2020, pp. 9459-9474.
[3] Y. Zhu, H. Zhang, and W. Chen, "Cross-document question answering with multi-hop reasoning," ACL Findings, 2022, pp. 112-127.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

[4] Y. Gao et al., "Retrieval-augmented generation for large language models: A survey," arXiv:2312.10997, 2023.
[5] S. Eismann, J. Scheuner, E. Van Eyk, and M. Schwinger, "Serverless applications: Why, when, and how?," IEEE Software, vol. 38, no. 1, pp. 32-39, 2021.
[6] P. Bajaj et al., "MS MARCO: A human generated machine reading comprehension dataset," in Proc. ICLR Workshops, 2018.
[7] E. Kasneci et al., "ChatGPT for good? On opportunities and challenges of large language models for education," Learning and Individual Differences, vol. 103, p. 102274, 2023.
[8] N. Kornell and R. A. Bjork, "The promise and perils of self-regulated study," Psychonomic Bulletin & Review, vol. 14, no. 2, pp. 219-224, 2007.
[9] J. Nielsen, Usability Engineering. Boston, MA, USA: Academic Press, 1993.
[10] T. Brown et al., "Language models are few-shot learners," in Proc. NeurIPS, vol. 33, 2020, pp. 1877-1901.
[11] W. Johnson et al., "Amazon Web Services Lambda: Architecture and performance," IEEE Trans. Cloud Comput., vol. 11, no. 2, pp. 948-961, 2023.
[12] J. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using Siamese BERT-networks," in Proc. EMNLP, 2019, pp. 3982-3992.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com