Tech Stack Rationale for Document Migration

|  |  |  |
| --- | --- | --- |
| Component | Rationale | Why Chosen |
| Spring Boot Batch | A robust framework for batch processing, suitable for migrating large datasets in manageable chunks. | - Scalability: Handles large datasets efficiently. - Flexibility: Customizable workflows. - Integration: Easy integration with Oracle DB and other Spring components. |
| API-C (On-Prem) | Acts as an API management layer for routing requests between on-premises systems and AWS services, ensuring secure communication. | - Security: Ensures secure communication via Direct Connect and API Gateway. - Routing: Centralized management of API requests. - Legacy Integration: Minimizes learning curve by leveraging existing infrastructure. |
| AWS Direct Connect | Provides dedicated, high-bandwidth, low-latency connectivity between on-premises infrastructure and AWS, ensuring fast and stable data transfer. | - High Bandwidth: Handles large volumes of data (6TB). - Reliability: Stable and low-latency communication between on-prem and AWS. |
| AWS API Gateway | Manages the HTTP requests from on-premises systems (via API-C) to AWS services (Lambda), providing security, scalability, and monitoring features. | - Scalability: Automatically handles variable traffic. - Security: Integrates with IAM for secure access. - Cost-effective: Managed service reduces overhead. |
| AWS Lambda | Serverless compute service for processing document metadata, uploading BLOBs to S3, and interacting with DynamoDB for metadata storage. | - Serverless: No infrastructure management required. - Scalability: Handles multiple concurrent executions efficiently. - Cost-effective: Pay only for compute time used. |
| AWS S3 | Provides durable, scalable, and cost-effective storage for document BLOBs. | - Durability: 99.999999999% durability. - Scalability: Can store large amounts of data. - Cost-effective: Reduces storage costs for large data. |
| AWS DynamoDB | NoSQL database for storing metadata (e.g., GUID, S3 URL) associated with documents, offering low-latency reads and writes. | - Scalable: Handles large metadata with low latency. - Fully Managed: No need to manage infrastructure. - On-Demand Mode: Scales automatically without manual intervention. |
| Oracle DB (On-Prem) | The source database for the original document metadata and BLOBs that need to be migrated, used for tracking and updating records with the new GUID. | - Existing Infrastructure: Already houses relevant data. - Data Integrity: Ensures accurate migration and GUID stamping in the Oracle database. |