

PROJECT OVERVIEW

A leading publishing-tech company sought to modernize its Data and Analytics capabilities to support its print-on-demand and integrated supply chain model. With over 10 million book titles and a daily shipment volume of more than 40,000 units, the company needed a robust, reliable, scalable, and secure platform to consolidate isolated data sources and deliver actionable insights. The project involved designing and implementing a Data Lake and Analytics platform using AWS native services to optimize performance.



ASBResources
Extraordinary People.
Extraordinary Results.

CHALLENGES

- Siloed systems such as ERP, Finance, Warehouse, E-commerce hindered data integration

Fragmented Data sources



- Delayed decisions due to lack of live analytics

No Real-time Insights



- Time consuming and error-prone processes

Manual reporting



- Disconnected analytics limited strategic growth

No unified Analytics



- Legacy systems struggled with growth and compliance

Scalability and Security concern



WHY AWS

AWS is for Data!

MODERN DATA FOUNDATION

Tailored user experiences

Visualize interactive dashboards

Consume services with simplified authoring interface

Automate app deployment with GenAI

Experience

Agile data transformation

Transform your data with batch, real-time, and Zero-ETL

Query data at scale

Train ML models on your data

Act

Unified data governance



PROTECT



CURATE



UNDERSTAND

Govern

Optimized cloud datastores

Remove data silos ingest and store all your data in one place

Scale at performance with 'Unlimited' capacity in cloud at reduced costs

Reduce development overhead with automated and repeatable data pipelines

Store



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

KEY FEATURES

- **Comprehensive Data Lake Architecture**
 - Centralized, scalable data repository built on AWS S3 with structured zones (raw, cleaned, curated) for efficient data management.
 - Serverless design ensures cost-effectiveness, elasticity, and minimal operational overhead.
- **Real-Time Data Processing and Analytics**
 - Integrated ingestion pipelines using AWS AppFlow and Lambda for near-instant updates from diverse sources. Real-time dashboards powered by Amazon QuickSight
- **Predictive and Prescriptive Insights**
 - Machine learning models built with Amazon SageMaker to predict trends, optimize inventory, and recommend actions.
- **Streamlined User Management and Governance**
 - Centralized governance using AWS IAM for user roles, permissions, and data access policies. AWS KMS for encryption and CloudTrail for activity tracking
- **Future-Ready Design for Growth**
 - Modular architecture with provisions for external-facing analytics, ensuring long-term adaptability.
 - AI-driven insights pave the way for new product offerings and revenue streams.

OUR SOLUTION

We architected a comprehensive data lake and analytics platform using AWS native services.

The platform was designed to ingest, process, store, and visualize data from various sources, providing a single source of truth for the client.

Ingestion Layer: Data was ingested from multiple sources using AWS SFTP, AWS Database Migration Service, Lambda Functions, and Amazon AppFlow.

Storage Layer: Amazon S3 was used to store data in raw, cleaned, and curated zones, ensuring scalability and cost-efficiency.

Processing Layer: AWS Glue and AWS Step Functions were used to orchestrate data processing pipelines, ensuring seamless data flow.

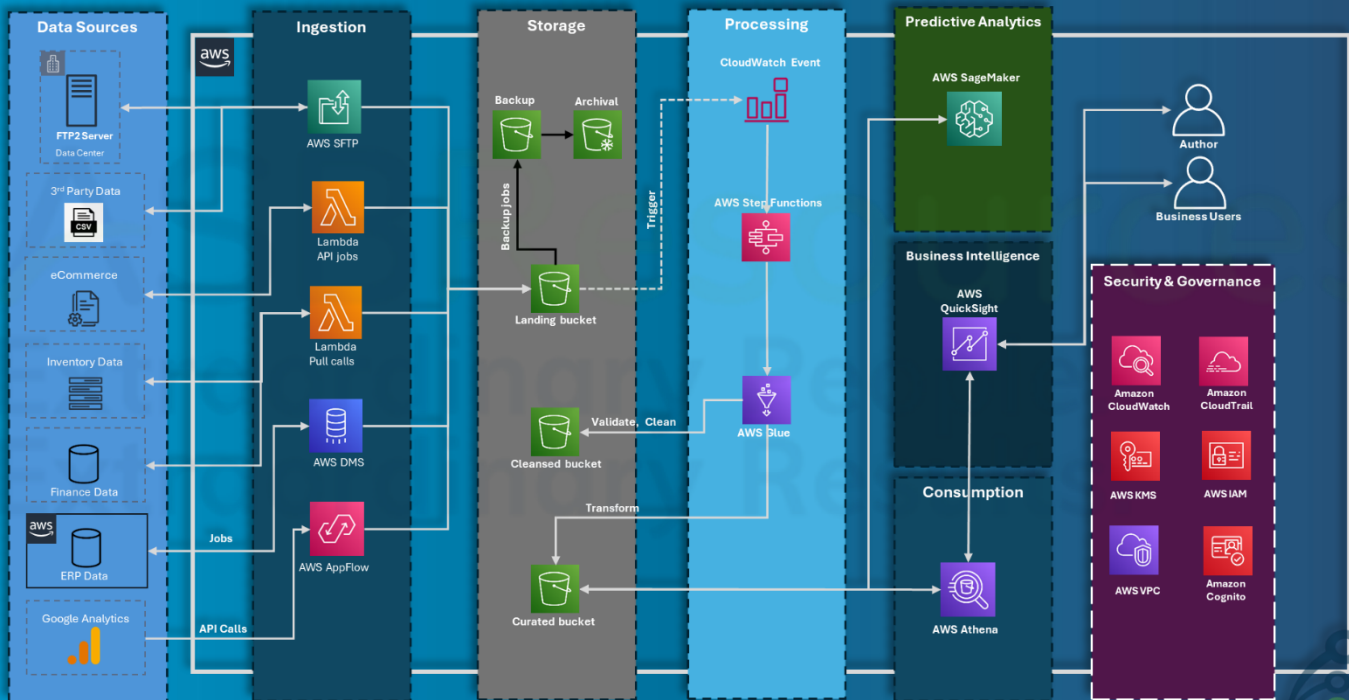
Consumption Layer: Amazon Athena was employed to run complex SQL queries against data stored in S3, providing quick insights.

Business Intelligence Layer: Amazon QuickSight was integrated to create interactive dashboards, enabling users to visualize data and gain actionable insights.

Predictive Analytics: Amazon SageMaker was used to build and deploy machine learning models for predictive and prescriptive analytics.

TECH STACK

Cloud Native Data Lake & Analytics Platform



ASB Resources Confidential

- **AWS Services:** AWS SFTP, AWS Database Migration Service, Lambda Functions, Amazon AppFlow, Amazon S3, AWS Glue, AWS Step Functions, Amazon Athena, Amazon QuickSight, Amazon SageMaker, AWS IAM, AWS KMS, Amazon CloudWatch, AWS CloudTrail, Cognito
- **Data Formats:** CSV, JSON, XML, Avro, Parquet, ORC.
- **Programming Languages:** Python, Scala.

KEY OUTCOMES

- **Data Unification:** Successfully unified data from over 12 disparate sources, creating a single, cohesive data lake.
- **Scalability:** The platform can handle up to 50TB of data per month, with the ability to scale effortlessly as data volumes grow.
- **Cost Efficiency:** Reduced data storage and processing costs by 30% compared to on-premises.
- **Real-Time Insights:** Enabled real-time analytics, reducing the time to generate insights from weeks to hours.
- **User Adoption:** Achieved a 95% adoption rate among internal users, with over 200 users accessing the platform daily.
- **Predictive Analytics:** Implemented predictive models that improved forecast accuracy by 25%, leading to better inventory management and reduced costs.
- **Security and Compliance:** Ensured data security and compliance with industry standards, protecting sensitive information.

SCHEDULE A CALL WITH ONE OF OUR EXPERTS TODAY!

HEAD OFFICE

4365 Route 1 S,
Suite 102
Princeton NJ 08540

US

330 Camp Road
Charlotte, NC
28206

CANADA

116 Albert Street,
Suite 200 & 300,
Ottawa, Ontario,
K1P 5G3, Canada

INDIA

Hubtown Solaris,
Suite 1113,
N S Phadke Marg,
Sai Wadi, Andheri East
Mumbai 400069

