

# Cloud Data Warehouse Solution for Order Management

## ORGANIZATION

Our client is a leading skincare company with over 25 years of innovation, serving millions of customers annually with cutting-edge skincare solutions.

## CHALLENGE

Our client sought to replace their existing third-party managed data warehouse solution with a modern, best-in-class data management platform built on Amazon Redshift. The new solution needed to support data ingestion, querying, extraction, reporting, and analysis while seamlessly integrating with the client's existing Tableau visualization and reporting tools. The system was also required to scale with utilization, ensuring consistent performance, cost-effectiveness, and high-quality data.

XTIVIA delivered a robust AWS cloud-based data warehouse solution, implemented by experienced data professionals, to meet the client's scalability, consistency, and efficiency requirements.

## Scope of Work:

- **Discovery and Analysis:** Conducted a detailed assessment of the existing workload and ETL processes.
- **Infrastructure Setup:** Provisioned a scalable and secure data warehouse infrastructure in a new AWS account.
- **ETL Pipeline Development:** Designed and implemented ETL pipelines and ingestion jobs for one data source using AWS DMS.
- **Technical Guidance:** Provided expertise on application architecture, scalable data storage, and efficient access patterns.
- **Application Integration Support:** Assisted the client in developing application integrations using the new data storage solution.
- **Knowledge Transfer:** Conducted knowledge-sharing sessions and hand-offs to ensure the client's teams could manage and maintain the solution effectively.

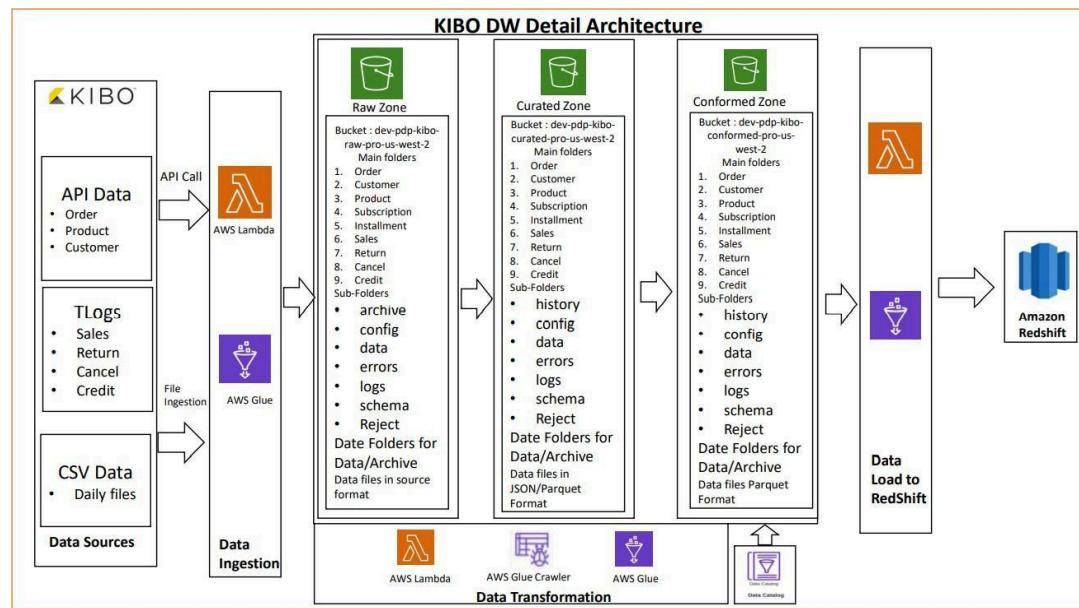
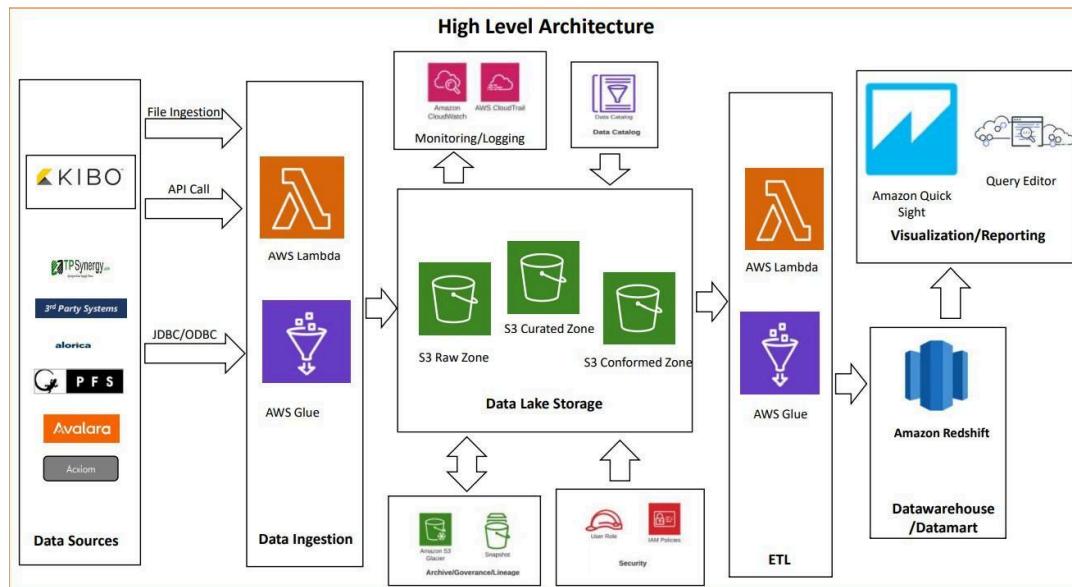
This modernized data warehouse architecture enabled the client with a scalable, high-performance platform for streamlined data management and analysis.

## TECHNICAL SOLUTION

XTIVIA used AWS data lake, including Amazon S3 and native serverless Lambda, Data Cataloging, Athena tools, and data lake engine to support data analytics vision and deliver the data warehouse and analytics. The task was completed in 12 weeks, which included implementation, a top-down approach to defining an implementation plan, and a roadmap that consisted of the following tasks:

- AWS S3 & Redshift infrastructure setup on AWS and access through secured VPN
- Architecture and Design
- Order management data consolidation and curation using Lambda and Data Catalog
- Multiple data sources/file formats consolidated and converted to compressed Parquet format as per AWS best practices

- Data transformed from source to raw to curated and pushed to conform layers as per business and technical requirements
- Sort key creation using Redshift best practices for faster data query need
- A robust architecture to support future needs and easy configuration of new data sources on-boarding



## PROJECT ACTIVITIES

### Discovery Phase:

- Reviewed the client's existing ETL (Extract, Transform, and Load) processes, application logic, and scalability challenges.
- Assessed data lake architecture and objectives:
  - Defined data strategy, pipelines, and ingestion approach.
  - Identified stages and access patterns.
  - Analyzed data preparation and transformation requirements.
  - Established business use cases.
  - Reviewed data access, lineage, and governance requirements.

### Implementation of Data Infrastructure:

- **AWS Landing Zone:** Configured a multi-account structure with centralized auditing, logging, and SSO-based authentication.
- **S3 Buckets for Data Lake Zones:**
  - Defined bucket names and partition strategies.
  - Configured IAM roles and bucket policies for system and user access through Lake Formation.
- **Data Anonymization:**
  - Identified sensitive data fields for anonymization.
  - Implemented secure anonymization techniques.
- **Redshift Cluster Setup:**
  - Designed cluster architecture, including instance type, storage requirements, workload management, and concurrency scaling.
  - Configured JDBC connectivity and user administration.

### Data Pipeline Development:

- Designed and implemented pipelines for data ingestion:
  - Selected optimal methods for ingesting historical and current data from daily transactional systems.
  - Built ETL processes to transition data from raw to curated formats, conforming to Redshift.

### Data Conversion and Transformation:

- Conducted data cleansing, transformation, and preparation for Redshift integration.
- Ensured seamless ETL processes to meet analytical and operational needs.

This structured approach ensured the client achieved a scalable, secure, and high-performance data infrastructure optimized for their business requirements.

## BUSINESS RESULT

XTIVIA identified innovative approaches to enhance the productivity and collaboration of the client's information systems. Our team implemented a scalable, extensible data warehouse solution using AWS S3 and Redshift, aligned with the client's data and analytics vision to meet current and future demands. Key features of the solution included:

- **Modern Data Management:** Delivered a best-in-class solution for efficient and effective data management.
- **Cost-Effective Performance:** Ensured consistent performance with a cost-efficient architecture.
- **Scalable Cloud Architecture:** Designed a flexible, cloud-based solution to accommodate future feature enhancements and performance improvements.
- **BI Tool Flexibility:** Enabled seamless integration with any BI tool, including Tableau and Power BI.
- **Serverless Data Transformation:** Utilized AWS Lambda for serverless transformation and load operations to minimize cloud computing costs.
- **Data Lake Foundation:** Established a robust data lake foundation for raw data analysis.
- **Real-Time Data Processing:** Designed an event-driven architecture for real-time data processing, replacing traditional batch processing.
- **Complex SQL Support:** Provided support for implementing advanced SQL capabilities.

This solution empowered the client with a future-ready, high-performing, and cost-efficient data warehouse that supports their evolving business needs.

## KEYWORDS

AWS, S3, Cloud Data Warehouse, event-driven architecture, Data Lake Foundation

## SOFTWARE

Amazon S3, Lambda, Python, Amazon RedShift, Flat Files

## HARDWARE/PLATFORM

Amazon S3 and RedShift, Azure

## ABOUT XTIVIA

At XTIVIA, we've been providing IT solutions and consulting services for over 30 years with a wide range of services, including technology assessments, IT service and asset management, software development, data analytics, cloud migration, DevSecOps, ERP, and enterprise content management. Dedicated to each discipline, ensuring that our clients receive the best possible service. Through strategic acquisitions, we've acquired talented people who are experts in their industries, passionate about what they do, and committed to providing exceptional service to our clients. Whether you need to improve your IT infrastructure or implement new software solutions, XTIVIA is here to help you achieve your business goals. XTIVIA has offices in Colorado, New York, New Jersey, Texas, Virginia, and India. [www.xtivia.com](http://www.xtivia.com)