

# Order Management Data: Cloud Data Warehouse Solution

## ORGANIZATION

Our client is a skin-care products company that serves millions of customers annually. They have been a leader in skincare innovation for over twenty-five years.

## CHALLENGE

Our client needed to replace their existing data warehouse solution managed by third-party service providers with a new solution built on Amazon Redshift to provide a modern, best-in-class data management solution. Requirements include support for data loads, querying, extraction, and reporting and analysis. In addition, visualization and reporting capabilities should tie into the client's existing Tableau and should scale in line with utilization to ensure consistent performance and cost-effectiveness with quality data.

XTIVIA provided experienced data professionals to implement a modern AWS cloud data warehouse solution to ensure scalability, consistency, and cost-effectiveness.

### Scope of Work:

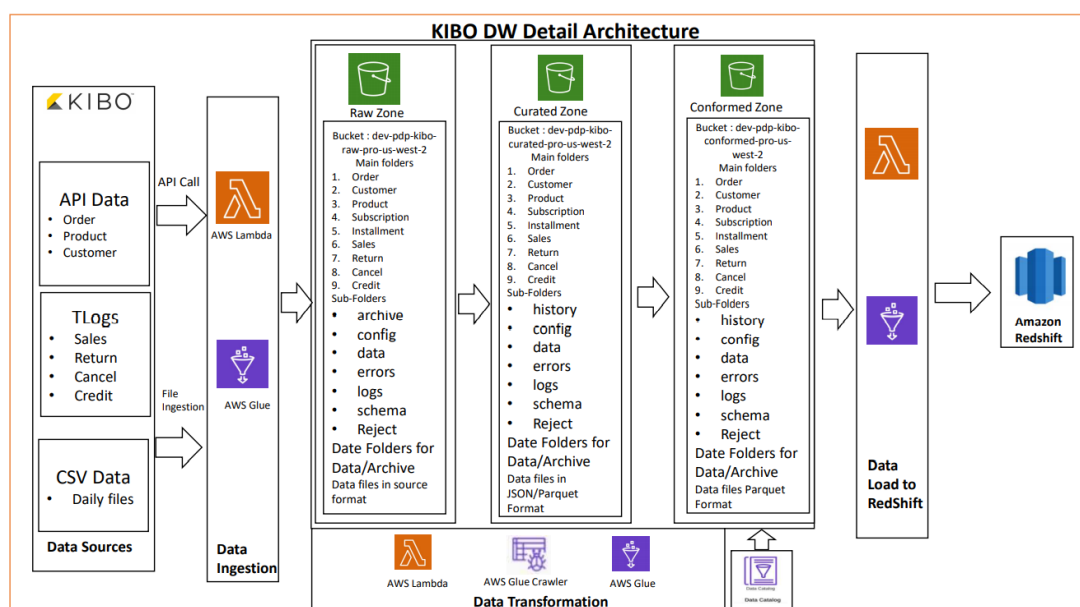
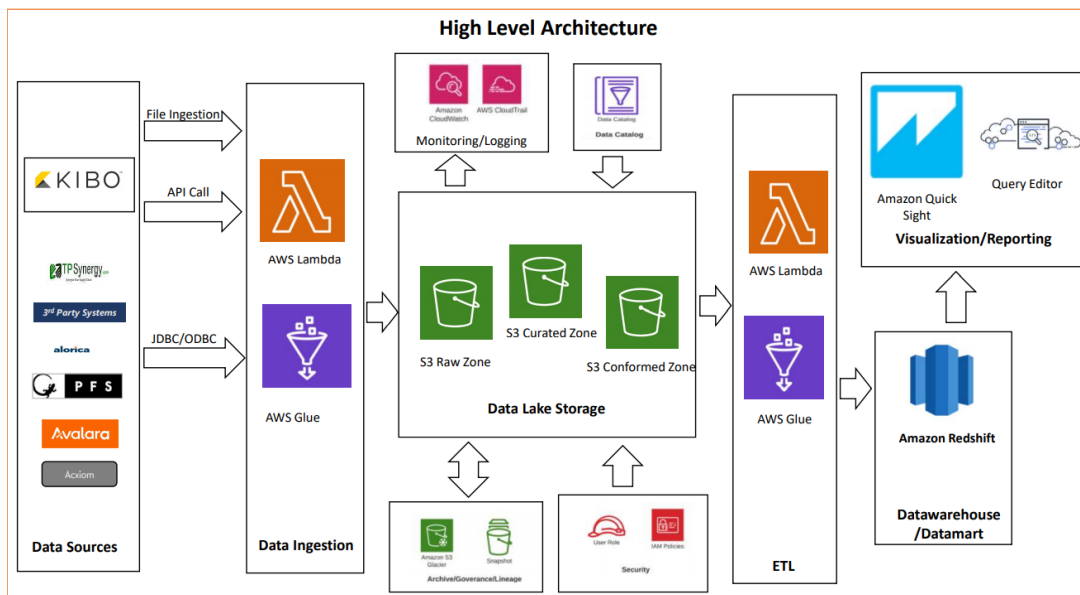
- Perform discovery of existing workload and data extract, transform, and load procedures.
- Provision of data warehouse infrastructure in a new AWS account
- Develop ETL pipeline and ingestion jobs for one data source utilizing DMS
- Provide customers with the technical guidance of application architecture, scalable data storage, and access patterns.
- Provide support and guidance to customers while developing application integrations using new data storage.
- Perform knowledge transfer and hand-off cost-effectiveness

## 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 customer's existing data extract, transform, and load procedures, application logic, and scalability concerns
- Reviewed data lake architecture and goals:
  - Defined data strategy
  - Defined data pipelines
  - Defined ingestion strategy
  - Defined stages and access patterns
  - Understood data preparation and transformation requirements
  - Defined business use case
  - Reviewed data access requirements
  - Reviewed data lineage requirements
  - Reviewed data governance requirements
- Implementation of Data Infrastructure
  - AWS landing zone with multi-account structure, centralized auditing, logging, and authentication (SSO)
  - S3 buckets for data lake zones
    - Bucket names
    - Partition strategy
  - IAM roles and bucket policies for system access
    - Lake Formation -IAM roles and bucket policies for user access
  - Data Anonymization
    - Identified data fields for anonymization
    - Implemented anonymization method
  - Redshift -- Redshift cluster design
    - Instance type
    - Storage requirements
    - Workload management
    - Concurrency scaling
    - Setup JDBC
    - User Admin
    - JDBC Connectivity to Redshift
  - Created data pipeline for Data ingestion
    - Identified best method for ingesting historical data
    - Identified best method for ingesting current data from daily transactional systems
  - Implemented ETL process from raw to curated to conform to Redshift
    - Data Conversion
    - Data Cleansing
    - Data Transformation
    - ETL to Redshift



## BUSINESS RESULT

XTIVIA found new ways to make information systems more productive and collaborative. Our team implemented a scalable and extensible data warehouse solution with AWS S3 and Redshift data warehouse that aligns to the client's data and analytics vision and meets future demands, including:

- Modern, best-in-class data management solution
- Consistent performance and cost-effectiveness
- Cloud-based scalable solution architecture keeping future data need for feature enhancements and performance improvement
- Flexibility to work with any BI tool – seamless with Tableau/Power BI
- All the transformation and load using serverless Lambda to minimize cloud compute cost
- Data lake foundation to analyze raw data
- Event-based real-time architecture and design for data processing instead of batch mode
- Support for implementation of complex SQL

### KEYWORDS

AWS, S3, Cloud Data Warehouse

### SOFTWARE

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

### HARDWARE/PLATFORM

Amazon S3 and RedShift, Azure

## ABOUT XTIVIA

XTIVIA is a proven technology integration and innovation company known for delivering leading-edge technology solutions to our clients' specific requirements, regardless of project complexity. We bring next-level business operations to reality through Application Development, Business Intelligence, Data Warehousing, Database Support & Management, Enterprise Information Management, Digital Experience Solutions, and Enterprise Resource Planning. Our success stems from a demonstrated ability to deliver deep expertise via professional services, empowering clients to leverage their chosen technology successfully, competitively, and profitably. XTIVIA has offices in Colorado, New York, New Jersey, Missouri, Texas, Virginia, and India. [www.xtivia.com](http://www.xtivia.com)