DATA TRANSFORMATION, ARCHIVAL AND ANALYTICS SOLUTION FOR WASTE MANAGEMENT SERVICES

Abstract

Data is the key asset of any organization and takes years to accumulate. Organizations analyze vast amounts of data using spreadsheets, applications, data warehouses, and business analytics solutions. As the demand for more innovation and faster responses grows, managing data is becoming an increasingly immense challenge.

Waste management companies use data to make critical decisions to uncover opportunities and challenges. Waste management has been mandated across the world by government agencies. It is one of the key components of sustainability, as our consumption increases leading to greater waste generation. Local and international communities work ceaselessly to overcome the challenges related to waste disposal.

This paper looks at how a waste management company achieved seamless data transformation, archival, and analytics by partnering with Infosys. It highlights the challenges faced in data migration and how these were overcome to migrate to the Oracle Cloud. It also demonstrates how the Infosys industry solution for data archival helps non-hazardous waste management companies manage their historical data and build analytics to navigate the next.

Infosys®

Navigate your next

Table of Contents

Overview

The process of waste management starts with waste reduction, segregation of waste at source, the process to collect waste from designated locations, and disposal of waste at the destination. It can broadly be classified into hazardous and non-hazardous waste management. The primary source of revenue comes from the collection operation which is driven mainly by residential waste. Once the waste is collected, there are opportunities available for cross-selling, recovery of key materials, and harnessing it for renewable energy.

Data is captured during each phase of the waste management process and transformed into a data model. The Oracle Cloud data model allows data to be stored and made available in specific subject areas which enables companies to build analytics and drive critical decisions.

Considerations for a Data Transformation Project

Waste management companies often grow through acquisitions and inherit disparate systems and applications. This results in the data being non-harmonized and heterogeneous, leading to increasingly complex data models. This causes challenges with scalability and makes operations more complex. The primary objective of data transformation and migration to Oracle Cloud is to reduce the total cost of ownership (TCO) and integrate disparate sources of data. This enables business units within the waste management organization to collaborate better.

The availability of appropriate and relevant data is necessary for ensuring smooth transformation to Oracle Cloud. When data is transformed by mapping business processes to Oracle Cloud products, critical end-to-end processes must be executed without any error.

Infosys recently partnered with a leading waste management company to enable data transformation, archival, and analytics for them. Before starting a data transformation project, it is important to consider certain key questions, and develop a strategy based on the answers. Some of these questions are:

1. How do we understand the scope and quantify the volume of data for the transformation?

2. What are the critical data entities in the master and transactional data that need harmonization?

3. What are the business validation rules for archived data?

4. What kind of testing needs to be performed on the archived data during implementation?
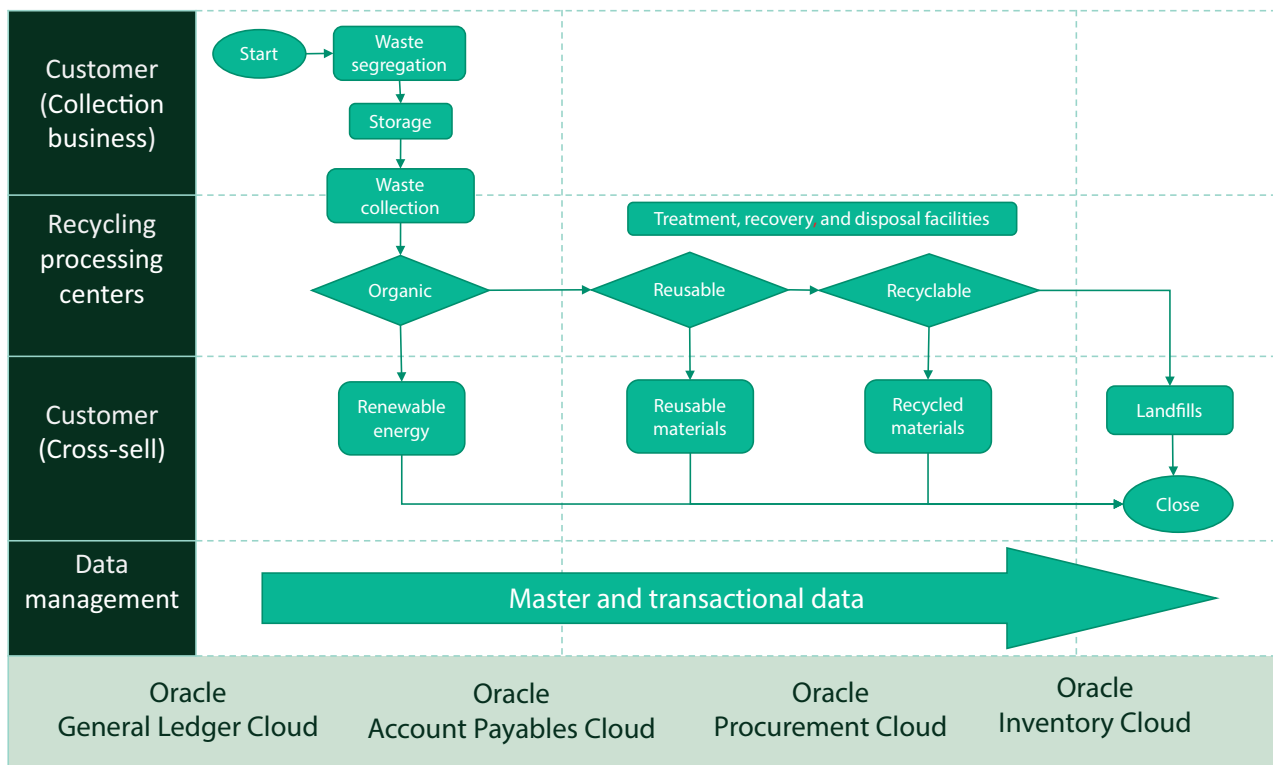
5. How can the business access the archived data?

Figure 1 – Data management in Oracle cloud during the waste disposal process

Data archival for regulatory requirements

In this instance, in addition to migrating master and transaction data to Oracle Cloud, the waste management company archived the entire data in a central database for regulatory, legal, and compliance needs. The central database serves as the source for historical data reporting.
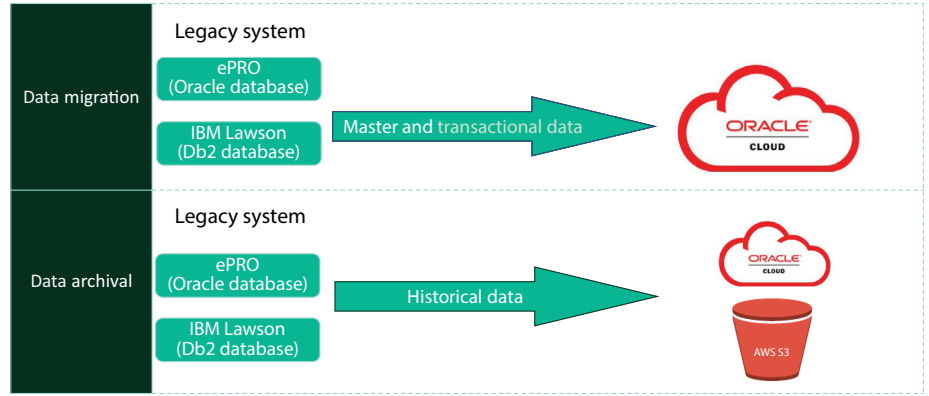

Figure 2 – Overall approach for data migration and archival on Oracle Cloud or AWS simple storage service

Infosys Data Archival and Analytics Solution

To implement data transformation for the waste management enterprise, Infosys leveraged three key solutions built on Oracle Cloud:

- Data transformation and migration solution

- Data archival solution

- Analytics and reporting solution

Data transformation and migration solution

The Infosys industry solution, Infosys Data Services Suite (iDSS), is a highly scalable and easily portable solution built with a high degree of automation to transform, validate, and migrate data from disparate on-premises systems to Oracle Cloud. The standard data conversion tasks are part of the solution framework to drive discussions with business users and streamline processes through appropriate engagements.
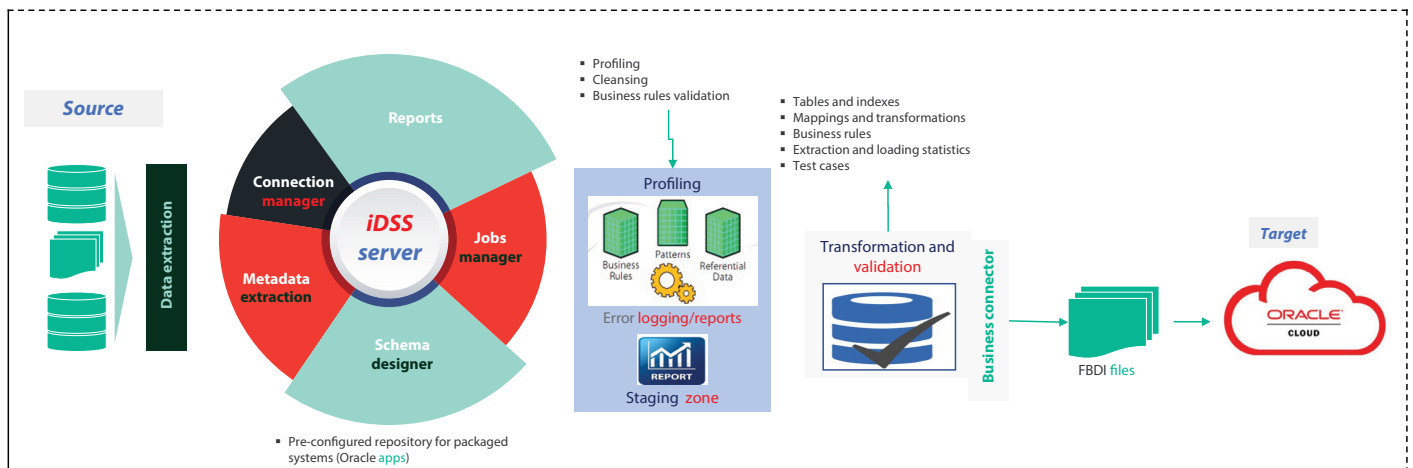

Figure 3 – System architecture diagram for Oracle Cloud transformation

Data archival solution

The Infosys industry solution is built for data archival considering:

- Business objectives of clients

- Technical and business challenges involved

- Rationale for maintaining data for several years in a scalable and easily accessible data warehouse

Data migration services (DMS) are used to migrate data to and from AWS easily and securely allowing the source data source or database to remain fully operational during the migration. AWS DMS reads the data from the source database and creates a series of parquet files that will be loaded into the S3 archival area.

The automated solution leverages learning from several engagements and facilitates the migration of data to Oracle Cloud from any ecosystem system or application. The Infosys Automated Data

Migration solution built on an ingestion framework can be easily configured and adopted based on requirements.

Key features

The data archival solution provides several key features that were used in the waste management data transformation solution:

- **Data transformation/compression**

  Leveraged Python to compress data extracts from csv format to parquet format. Parquet has tremendous compression capabilities and is also easily ingestible by several big data services and Snowflake

- **Audit logging and high-level validation**

  AWS DMS and Python were used to create audit logs to track database table and row count during the transformation and errors were captured during processing

- **Detailed cell-by-cell data validation**

  Source data and archived data were compared using the in-built data validation algorithm

- **Partitioning**

  External tables were partitioned to provide faster query response and reduce costs as only a small part of the data is scanned per query instead of scanning an entire data set.

  Partitioning divides data into parts keeping related data together based on a partition key (column values). This helps organize data in a hierarchical folder structure based on the distinct values of one or more columns. AWS Simple Storage Service (S3) data is commonly partitioned by time/date so that S3 sub-folders are created based on year/month/date values found in a timestamp or date field.
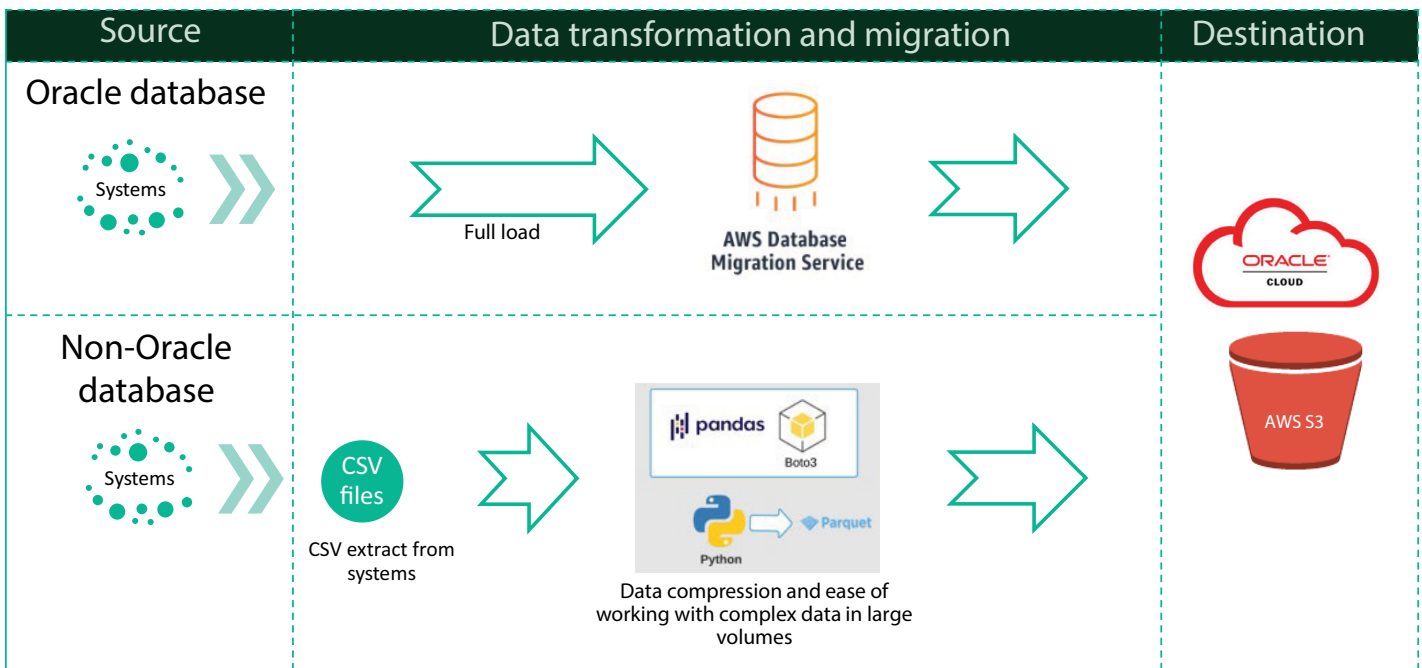

Figure 4 – Data archival system architecture diagram for Oracle Cloud or AWS S3

Analytics and reporting solution

The data warehouse is used to store historical data used for reporting and data analysis. The required data is extracted from legacy data sources and stored in an enterprise cloud data warehouse (EDW) developed using the Snowflake cloud database. The EDW stores a subset of archival data, containing up to 7 years of historical transactional data from legacy applications. All master data from

these applications is stored in the data warehouse to prevent orphaning of transactional data.

A cloud data warehouse platform such as Snowflake uses an architecture that has distributed cluster nodes along with disk storage, its own CPU, and memory. These platforms deliver performance and scale by running full table scans in parallel.

Oracle Analytics Cloud (OAC) has the capability to explore data from a variety of data sources, including Snowflake cloud databases. The OAC data model is used to present data from underlying data sources in an intuitive manner that reflects business structures and uses business terminology to support analysis and reporting needs.

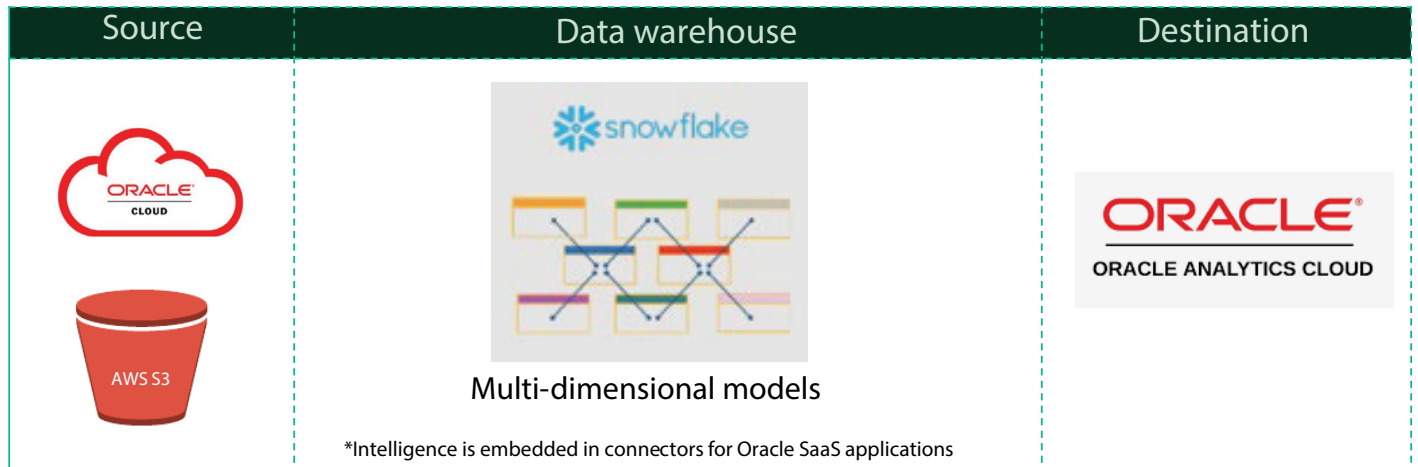| Source	Data warehouse	Destination
ORACLE CLOUD / AWS S3	snowflake Multi-dimensional models *Intelligence is embedded in connectors for Oracle SaaS applications	ORACLE ANALYTICS CLOUD

Figure 5 – Data connectors – Oracle and non-Oracle – for any source

The reporting data model for the waste management project was developed using the OAC Developer Client Tool and uploaded to Oracle Analytics Cloud. Users can filter data from reports based on historical data using prompts and selection criteria provided in the interface.

Business Benefits

A data transformation project is challenging, but it also provides several business benefits that establish the return on investment for a project of this scale. For the waste management project, the following benefits were achieved:

- Scalability – Highly scalable and easily portable solution built to migrate data efficiently to ensure predictability with reconciliation processes for large volumes of data

- Cost effectiveness – Reduces the total cost of ownership (TCO), increases productivity, and reduces the time to market

- Data analysis — Ability to analyze data in the cloud, on-premises, or on a hybrid platform. Ability to run analytics in the cloud with the data archival solution

- Enhanced security – Object-level security to restrict access to various OAC objects including dashboards, specific subject areas, and reports

Conclusion

Like any other business, waste management deals with huge volumes of data. Analysis of real-time and historical data helps drive decisions for more efficient waste management operations. Implementing an Oracle Cloud based data transformation project provided the waste management business with several benefits.

The industry solution on data archival with analytical reporting capabilities has enabled business users to better manage their historical data. It further allows them to create stunning visuals to explain their results and share them with colleagues. In addition, the transformation solution provides a strategic platform to capitalize on the collective intelligence of people, processes, and data to navigate the next.

About the Authors

Chandrashekar Poonachi
Industry Principal, Infosys

Chandra has 20 years of IT experience in consulting, project management, and pre-sales in the services, distribution, manufacturing, hi-tech, and healthcare industries. As a technology transformation leader with Oracle cloud and Oracle E-Business Suite, he has helped clients in their digital journey with Infosys across multiple locations in India, Japan, UK, and the US. Chandra is a solution architect with the Infosys Oracle practice and has contributed to the industry solution with Oracle ERP Cloud.

Abhiram Bharadwaj
Digital Solution Specialist, Infosys

Abhiram has 14 years of experience in business intelligence and data analytics projects across various domains. At Infosys, he is a part of the Oracle BI practice and is one of the primary anchors for proprietary tool development and automation within the Oracle BI Center of Excellence. Before Infosys, Abhiram was with Oracle India and worked with Oracle sales consultants and product development teams to create demo-ready solutions to take to the market along with Oracle Analytics Cloud products.

Amit Kumar Upadhyay
Senior Project Manager, Infosys

Amit has 17 years of experience in Oracle on-premises and cloud products covering transformation, upgrade, rollout, and support engagements in project lead and consultant roles. At Infosys, Amit is a part of the Oracle practice. His responsibilities include managing scope, schedule, and budget apart from quality assurance, staffing, stakeholder management, and developing metrics.

Sanjay Kanchan
Group Project Manager, Infosys

Sanjay has 25 years of professional experience in sales and marketing, Oracle consulting, program/project/delivery management, and pre-sales in manufacturing and services industries. Sanjay's experience in the Oracle ecosystem covers implementation, migration, upgrade, rollout, and support engagements in the capacities of consultant, lead, and service delivery/program lead for Oracle Cloud/EBS. He has managed and led large and complex transformation programs as delivery lead, managing multiple Oracle programs in the services segment. As a delivery lead, Sanjay has helped customers in their digital transformation journey while maintaining high service levels in terms of cost, time, and quality for key engagements.

References

1. https://www.ateam-oracle.com/post/reference-architecture-fusion-saas-data-replication-into-adw-using-odi-marketplace-and-bicc

For more information, contact askus@infosys.com

Infosys®
Navigate your next

Infosys.com | NYSE: INFY

Stay Connected