Migrating BI from On-Prem to the Cloud: A Case Study

Axis Group

Table of Contents

AXIS

What is MIPS? What are ACOs?

Merit-Based Incentive Payment System (MIPS) is a government program that encourages collaboration between health networks. The program provides scores to doctors and healthcare networks that are based on the level of preventative care that is provided to their patients. The scores are then used to determine Medicare payment adjustments.

Accountable Care Organizations (ACOs) are groups of doctors and hospitals that come together to provide low-cost care as they share financial responsibilities. ACOs were created under the Affordable Care Act to optimize the quality of care received by a Medicare patient but at a lower cost to the patient.

Background

Hospitals and other healthcare organizations are responsible for the care of millions of patients per year. As hospitals and the healthcare industry as a whole embrace digital transformations, many are choosing to store patients' data in data warehouses based in the cloud instead of a traditional on-premises option.

There are many challenges associated with using an on premises datastore. On-premises options are more expensive, slower, and typically harder to manage, although both options offer various security capabilities. Data security is critical across all industries, but especially in healthcare. Healthcare organizations can face serious repercussions if they do not share data securely. Additionally, cloud options provide more storage with less overhead, allowing for organizations to obtain insights from their data much quicker.

Healthcare systems use Electronic Medical Records (EMR) which allows for patients to have their own digital chart capturing their labs, medications, procedures, and diagnosis across time. ACOs, like our client, must combine data from EMRs to manage entire populations. As a result, a platform that would combine security, specifically for healthcare data and ease of data use, which in turn would make it easy to share across members of ACOs, was critical for the organization.

AXIS

Background *Continued*

Specifically, our client was looking for a cloud-based platform that would allow for data sharing securely that would also save money by processing data faster with less overhead. With Axis Group help, they completed an evaluation of cloud data warehousing options, and they decided upon Snowflake. Snowflake is a cloud-based data warehouse that offers a consumption-based pricing model, which allows for a reduction in costs. Additionally, Snowflake makes it easy for data sharing and exchange and offers security for health data. These were critical components that led to Snowflake being chosen by the organization as they share healthcare data across healthcare organizations. They needed the score attributes for 2.5 million patients to be generated in three hours or less on a once-a-week basis. They also wanted this warehouse to serve as the main source for patient quality and provider performance data moving forward.

The Customer

A large ACO was using an on-premises server in Oracle for their data warehouse. This warehouse contained Merit-Based Incentive Payment System (MIPS) scores for Providers, Clinics and Networks for two ACOs.

Additionally, the data warehouse contained data from over 80 Electronic Medical Record (EMR) systems. The organization was facing slow load times for their data as well as expensive daily maintenance and licensing costs, paying over ***three hundred dollars per day*** in licensing alone. They wanted to lower their costs, while maintaining the integrity of their systems; they were looking for a change and contacted Axis Group.

In the case study that follows, we will walk through how Axis Group planned and implemented this migration project, moving from an on-premises data warehouse to Snowflake (in the cloud).

Approach

Axis Group's approach to the project can be broken into the following initiatives:

- Data Migration and Validation
- Automating the conversation from Oracle to SnowSQL
- Refactoring Transformation Logic
- Performance Testing and Data Validation

Data Migration

The data migration was an iterative process that revolved around ingesting pipe-delimited flat files that originated from a backup of the Oracle database. There were nineteen Oracle PL scripts that needed to be migrated to Snowflake SQL.

Each step of the data migration process is further explained below.

Pre-Loading

The first step in the migration process was to recreate the source tables' structures in Snowflake. The data definition language (DDL) that was generated by Oracle needed to be cleaned-up and translated to Snowflake-specific SQL. To do this, a SQL developer took the DDL from a table and then ran the code through a Python script to fix syntax and then write the code to a text file. The text file needed additional edits to ensure the correct data types and syntax. This also allowed the DDL to remain in a centralized location. This can be seen below.

Staging

The Oracle tables were exported as pipe-delimited CSV files and uploaded to the Unix file server. Additionally, the files had to be uploaded and staged within Snowflake. This needed to be done with the SnowSQL Client (CLI) tool as well as for Python with Snowflake's ODBC driver.

Loading

After the files were staged and the tables were created, data could be loaded into the tables. A Python script was used to automatically generate code, log files, and cleanup code. Using a script helped ensure a consistent approach to all aspects of code management.

Validation

The validation process included validating the row count for each table and then validating the distinct row counts for each column of each table. It is important to note the sheer size of some of the tables. One table had over 1.1 billion rows and two of them had 500 million rows. During the validation process, we identified which data types were causing discrepancies.

During the validation process, Axis Group discovered two data types causing discrepancies: 1) dates and timestamps and 2) floats.

The first issue was identified by the data types in the dates and timestamps. In Oracle, when dates were formatted without times, they were still stored with their time data. However, they were only exported with date data, which meant that the data in Snowflake only contained date information resulting in fewer distinct counts in Snowflake. Floats were the other data type causing issues, as Oracle and Snowflake round differently. However, through the validation process, it was determined the rounding differences did not impact the final results.

Scorecard SQL Script Refactoring

There were nineteen Oracle PL scripts in total that needed to be migrated to Snowflake SQL. These scripts needed to be refactored to address the dissimilarities that exist between the two platforms. There is a standard for SQL, but it does not cover some of the nuances of the more elaborate functions.

The refactoring process can be broken down into the following sequence of operations that Axis Group followed:

1. We obtained copies of all involved scripts.
2. We performed an inventory of SQL functions and code elements found in all scripts.
3. We conducted research on Snowflake documentation as well as experimented with live Snowflake sessions to find and then validate functions.
4. We converted the existing scripts. This was done by translating Oracle functions to their Snowflake equivalents when possible. This was an extremely iterative process, as not every function had a Snowflake equivalent, so reworking and additional validation occurred to complete this step. The table below provides more information on this process.
5. We ran all scripts in the Snowflake environment. This step was repeated until there were no bugs, and each script was executed without error.
6. We verified that the data generated for the scores was the same as what was generated against the original code and original data. High-level summary scores were consistent, but the lower-level values were randomly assigned. However, because these values are not in the dashboard that serves as the primary source of the scoring data, the total score values matched, and the validation was accurate.

The following table depicts the differences between the functions in Oracle and Snowflake.

Oracle Function	Evaluation	Snowflake Function	Differences Notes
/	Different	;	An optional way to terminate a statement in Oracle. Needs to be ; in SF
&v_table – Dynamic SQL Tablename	Different	TABLE, IDENTIFIER	When used in a from clause, use TABLE($mytable); When used in a CREATE TABLE or INSERT statement, use IDENTIFIER($mytable). Oracle supports variable substitution natively, so it's impossible to have a variable with the name of a table and call it, i.e. update &mytable. SF has limited support for dynamic SQL. To replicate this case in particular, SF supports the function IDENTIFIER(), which can take a variable with a string literal, i.e. UPDATE IDENTIFIER($mytable)
COLUMN flu_start new_value p7_flu_start	Different	COLUMN	column…new_value is used to define a substitution variable for values selected in the specified column. Used to control output to console In this case, p7_flu_start becomes values in column flu_start when flu_start is selected.
define v_ttable = 'scores.g'&attrib_date._	Different	||	SET v_ttable = 'scores.g' || $attrib_date._ Also, variable is proceeded with $ not &. Oracle uses DEFINE to initialize variables. SF uses SET.
EXP	Different	POW	POW(x,y)
INTERVAL '65' YEAR	Different	INTERVAL '65 YEAR'	Change location of quotations. interval '65 year' vs '65' year
LAST	Different	LAST_VALUE	Use of LAST_VALUE syntax is recommended by Oracle, and equiv. function to this is available in SF
MONTHS_BETWEEN	Different	DATEDIFF	DateDiff(month, <date1> <date2>)
ROWNUM<= 10	Different	FETCH 10	ROWNUM is Oracle is used in Where clause. Fetch is used after where clause. To make dynamic, set a variable to 10 to limit return to first 10 rows. Set variable to NULL will return all rows. i.e. SELECT FROM X FETCH $y ROWS. FETCH $$$ will also return all results
SYSDATE	Different	CURRENT_DATE	

Validation Process

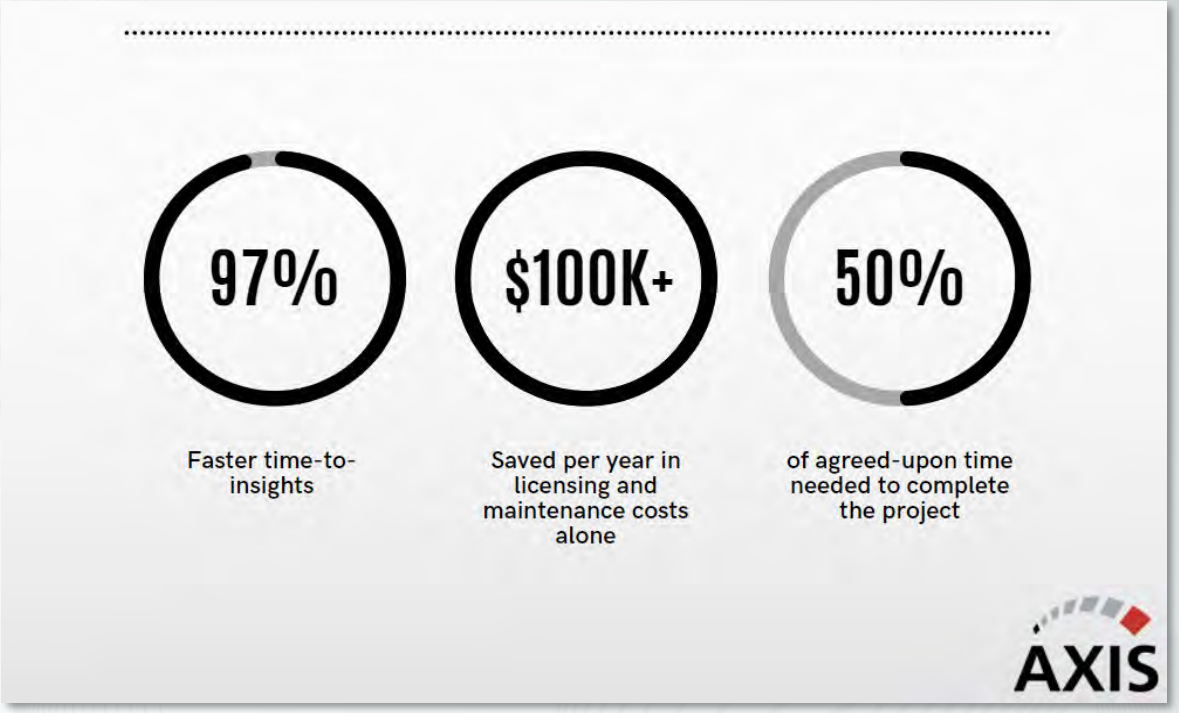
Score data from a specific timeframe was used for the validation. A copy of the original code was used to generate scores, which were then compared to the scores generated for the migrated code for the same timeframe. This was an iterative process as initially only about 80% of the values matched up. However, after looking at the migrated code, we resolved several bugs and brought the number closer to 90%.

Axis Group migrated the data and related code, then resolved the bugs, and ensured the values both pre- and post-migration were aligned.

During this phase, it was discovered that there was an issue with the original logic at the sub-grouping level. The code randomly assigned counts to one sub-grouping in cases where there were multiple sub-groupings present. However, for the higher-level score the sum of the sub-grouping values was consistent. By validating at the score name level, all but a few of the measures matched exactly, and the few that did not match were off by one, which was traced back to the original data and not represented in the Snowflake copy of the data.

AXIS

The Axis Group Impact



The healthcare organization was extremely satisfied with the results of the data warehouse migration initiative. Not only was the project completed ***two months*** earlier than originally planned, which led to it costing less than the agreed-upon budget, but they also saw tremendous improvement in time-to-insights. Their original load time for creating the scoring information had been about three hours, and after the migration, the code ran and scores were available in ***under five minutes***, in the smallest Snowflake server that could be configured.

Additionally, their daily cost before the migration was more than three hundred dollars per day. After the work was completed by Axis Group, they were paying about fifteen dollars per day, resulting in savings of over one hundred thousand dollars per year just in licensing and maintenance costs.

About Axis Group

With 25 years of experience, Axis Group delivers data and analytics consulting services and solutions to leading enterprises. Axis Group meets companies where they are on their digital transformation journeys and helps them achieve their data and analytics goals. Focusing on each company's unique culture and digital maturity, Axis Group delivers solutions from data visualization to data science. Axis Group ensures data literacy and analytics adoption to enable self-sufficiency resulting in smarter teams and better business outcomes. Axis Group combines business acumen, leadership, and industry-specific experience with technical expertise to tackle the toughest data problems. Axis Group is the Enablement Company™.

Axis Group

Southeastern Office
1100 Abernathy Road
Suite 800
Atlanta, GA 30328
Call us: (678) 367-0330

Northeastern Office
300 Connell Drive
Suite 3000
Berkeley Heights, NJ 07922
Call us: (908) 988-0200

marketing@axisgroup.com

Learn more here:
https://www.axisgroup.com/offerings/business-intelligence