Solve Tomorrow’s Business Challenges With a Modern Data Warehouse

Presented by: Dan D’Orazio
Date: September 25, 2019
Introduction

Dan D’Orazio - daniel.dorazio@matillion.com
Solution Architect at Matillion

LinkedIn: https://www.linkedin.com/in/dan-d-08b2028/

- 15+ years as a Data Engineer in the Application and Data space
- Familiar with a variety of toolsets; Informatica, Spark, Hive, SSIS
- Worked across verticals; Financial, Software, Telecommunications...
- Self-described tinkerer and data-geek
Social Media Giveaway

Share a quote or photo from this presentation for the chance to win an Ember mug.

Submission Requirements:
- Post on LinkedIn, Facebook or Twitter and include the hashtag #MatillionFutureofData or tag @Matillion.
- Winner selected at random and announced via social media at the beginning of the Booth Crawl tonight.
Agenda

• The Data Warehouse - a (brief) history
• Today’s data landscape
• What is a “Modern Data Warehouse”?
• Why modernize?
• A case for the cloud
• Do’s and Don’ts
• Be Bold
• Q & A
The Data Warehouse
A (brief) history

- 2013: Amazon Redshift first release
- 2000's: Big Data, Hadoop, REST API's, etc...
- 1995: Dotcom begins
- 1990's: Ralph Kimball and Bill Inmon develop DW methodologies
- 1980's: New Hardware / Software offerings specifically for “Information Centers”
- 1970's: “Data marts” are explored conceptually
- 1960's: First concept of “dimension” / “facts”
Today’s Data Landscape

Data is ... “Big”

Volume

1. 175 ZB by 2025
2. 49% in the public cloud
3. Enterprises are key contributors; Social Media, APIs, etc.

Variety

1. Relational Databases
2. Delimited Files
3. Web Services
4. NoSQL
5. JSON/JSOND
6. Logs
7. Media

Velocity

1. Data Acquisition
2. Conformed Models
3. Process Frequency
4. Overall Speed-to-Insight
What is a “Modern Data Warehouse”?

Traditional look and feel
- Columns and Rows
- SQL Dialect
- Robust connectivity
- Centralized repository - one-stop shopping

With a twist
- Near infinite scalability
- Native support for semi-structured data and modern file formats
- Frequent improvements
- Right-sized cost control
- Instant procurement and provisioning
Why modernize?

- Reporting and Operational Dashboards still benefit from cleaned, curated, and homogenous data.
- Centralization of raw data supports quick acquisition and efficient insight discovery (R&D) simultaneously.
- Increased focus on maintainability and flexibility as development cycle and processing times shorten.
- Increased need for on-demand scalability.
- Flexible cost models
A case for the Cloud

**Product:** Matillion ETL for Amazon Redshift

**Use Cases:** Data Warehouse Modernization, Reporting Analytics, Security

**Establishing workload/cost centers across clients to effectively:**
- Understand how much effort a client costs in ETL development
- Discover and separate customized ETL code that exists for a client
- Scale workloads appropriately

The ability to materialize complex business and data processing rules within Matillion allows Clutch to tackle customer challenges that were previously viewed as too time-consuming or costly.
A case for the Cloud

**Product:** Matillion ETL for Amazon Redshift

**Use Cases:** Data Warehouse Modernization, Reporting Analytics, Security

The Power and Gas division at Siemens used Microsoft SQL and a few different BI tools for reporting. In order to scale their data efforts, their team needed to move to AWS Redshift to consolidate different data sources.

It was implemented and providing insights **within an hour** providing the speed of data analysis that they needed.
A case for the Cloud

Product: Matillion ETL for Amazon Redshift

Use Cases: Data Warehouse Migration

GE’s Power and Water division needed to raise the bar on data warehouse scalability, integration, stability and development velocity while maintaining their current compute capabilities.

Moved to a managed Redshift solution in the cloud. This project only took 6 months from ideation to go-live with a PoC pilot cost of less than $100.

Results:
● 45% reduction in their operating cost
● Preserved performance and added stability
● Simplified operations
● OPEX savings. No CAPEX required.
● Fast development from concept to reality
● Highly resilient, seamless scaling
A case for the Cloud

**Product:** Matillion ETL for Snowflake

**Use Cases:** Data Warehouse Modernization, Reporting Analytics, Security

Legacy software too costly to maintain and needed a more flexible option. Moving to Snowflake was a key piece but they needed a solution that could also:

- Move data from its SQL Server environment into the chosen CDW.
- Quickly process large amounts of data to meet performance objectives.
- Ensure that data in transit remained secure.

The Results:
- Launch in minutes, with entire infrastructure running in Matillion in 2 days.
- Rapidly onboard data engineering team, becoming proficient in Matillion in less than 14-days.
- Greater extensibility, allowing for more integrations, dimensional modeling, and dispersed analytical modeling approach.
- Decrease latency to improve bandwidth and performance.
- Greater value for IT spend, relative to its ETL processes.
- Reduce time needed for long-running jobs - over 22 hours to 6 hours.
Modernizing Do’s and Don’ts

Do:
- Watch consumption closely
- Remember that messy data is *still* messy.

Don’t:
- Treat it like legacy tech
- Expect an easy button
- Forget the fundamentals
Be Bold with a Modern Data Warehouse

- Try all the things
- Test the limits
- Leverage the technology
- Be agile
Simplicity. 
Speed. 
Scalability. 
Savings.

Visit Us at Booth 1455

Purpose-built data transformation for cloud data warehouses

- Our intuitive UI and approach to data transformation makes complex tasks simple
- We deliver the fastest time to value, from launch to development to production
- Built to take advantage of the power and features of Amazon Redshift, Snowflake, and Google BigQuery
- Pay-as-you-go with no long term commitments

matillion.com/get-a-demo/

© 2019 Matillon. All rights reserved
Rate today’s session