Integrating Data in Teradata Vantage – A Case Study on Delivering Business Insights During COVID-19

Liza Duffy, Teradata
Phil Bangayan, Teradata
Key Takeaways

1. Data silos create chaos - success lies in good data integration, orchestration and governance.

2. Data Science success begins with translating vague requests into valuable insights.

3. Integrating external data sources are an excellent way to boost analytics and insights.
Teradata At a Glance

Corporate Purpose
We transform how businesses work and people live through the power of data

We Are Connected
Teradata is the connected multi-cloud data platform from start to scale

Culture & Community
Doing Good With Data™

Gartner
Cloud Leader in 2020 Gartner Magic Quadrant(1)*

Forrester®
Cloud Leader in 2021 The Forrester Wave™(2)*

19 of top 20 telcos
18 of top 20 global banks
15 of top 20 manufacturers
100+ technology partners
Financially strong and profitable with ~$2.0 billion revenue
CSR Leader: Dow Jones Sustainability Index for 11 years
S&P 400 constituent

1 The connected multi-cloud data platform for enterprise analytics
2 The most effective platform for driving business outcomes
3 The most open partnering approach to embrace the modern ecosystem

Gartner Magic Quadrant for Cloud Database Management Systems
The Forrester Wave™: Cloud Data Warehouse, Q1 2021
Our customers (sample)
“Using data and analytics, we looked at those that attended the game in the upper bowl, and we found two personas. Those that bought season tickets and were loyal fans who come to almost every game or shared the tickets with friends or family. And then there are those that buy a lot of season tickets that go into a secondary, and potentially even a third, market. We look to see the actual end-consumer that came to the game, what they actually paid, and what their overall experience was. We want to cut out a little bit of the middleman and we’ve strategically raised prices for those that were buying to resell while holding true on lower prices for those customers who are true and loyal fans.”

Dustin Spangler, Vice President, Data and Analytics
Larry H. Miller Sports & Entertainment
Businesses Use Data to Accelerate on Strategy

Data is the fuel

- How to best manage and orchestrate data?
- Who should see/interact with data across the organization?
- What are the considerations in bringing in external data sources?
How to Best Manage and Orchestrate Data?

Data Orchestration is the **automation of data-driven processes** from end-to-end, including **preparing data**, making **decisions** based on that data, and **taking actions** based on those decisions. It’s a process that often spans across many different systems, departments, and types of data.

- Data size and expansion
- Cost – Technical and People
  - Data Lake
  - Cloud
  - Traditional Data Warehouse
- Speed to Value

https://www.openprisetech.com/blog/what-is-data-orchestration/#:~:text=Data%20Orchestration%20is%20the%20automation,departments%2C%20and%20types%20of%20data.
Who should see/interact with data across the organization?

Data governance (DG) is the process of managing the availability, usability, integrity and security of the data in enterprise systems, based on internal data standards and policies that also control data usage. Effective data governance ensures that data is consistent and trustworthy and doesn't get misused.

- The larger and more complex the company, the greater the need for data governance.
- Decentralization aka “silos”
- Can you access the data you need when you need it?
- Do you need lengthy approvals or is there self service?

https://searchdatamanagement.techtarget.com/definition/data-governance#:~:text=Data%20governance%20(DG)%20is%20the,and%20doesn't%20get%20misused.
A Data Silo Story

A large US Bank wants to analyze consumer complaints during COVID

While they have access to the complaints, they also want to understand the customer experience with credit card late fees.

They search around their database, they ask others for where the data resides, they request permission to access that data.

This all takes time. Too much time.
Affinity Analysis of Platforms
External Data is Valuable

Considerations

There are many open data sources available today to enhance your analysis or predictive model:

- weather data
- event data (severe weather, covid 19)
- public health data
- crime
- GIS

Considerations:
- data quality and quantity
- IT department restrictions
- value to your analysis/model
So… What About COVID?

Business Impacts in 2020-2021
- travel restrictions
- supply chain interruptions
- work from home, learn from home
- health care alignment
- stimulus checks
- K shaped recovery

What's an Analyst/Data Scientist to do?
Background

• Teradata services many Fortune 500 companies across many industries
• Many media & entertainment companies’ business model relied on in-person gatherings

Sources:
Disruption due to COVID-19

• Coronavirus shut down in-person events, hurting media & entertainment industry

Sources:
Pressing Issue

• Studios had already planned release schedule for movies
• Release window only lasts few weeks
• Don’t want to release to theaters if regulators disallow or customers not ready

Sources:
Big Question

“How do we adjust our release schedule?”

• Note: very ambiguous and broad question

Source:
https://proofthatblog.com/2013/07/01/the-question-is-what-happened-to-the-question-mark/
Existing Data

- Most data (at the time) focused exclusively on health information
- Not necessarily focused on business questions

Source: https://coronavirus.jhu.edu/map.html
Data to Make Informed Decision

- Determined need for three sets of data to answer business question
- Behavioral measures demand based on actions taken
- Power comes when combined with first-party data
- Predictions on these data
Health Data

- Incorporate data on cases, deaths, and vaccinations

Sources:
- https://apnews.com/article/coronavirus-vaccine-2nd-shot-us-af9bc295faa04e204e4db62e336
- http://www.healthdata.org/covid
Behavioral Data

- What people searching for (Google Trends)
- Where people are moving (Google Mobility)
- On which items people are spending money (Bureau of Economic Analysis)

Sources:
https://www.google.com/
https://www.gpsworld.com/19-countries-track-mobile-locations-to-fight-covid-19/
https://www.bea.gov/data/consumer-spending/main
First Party

• Combining with first-party data makes the information relevant to solving business problem

Source:
https://www.fandango.com/movie-theaters
Ingesting Data

Sources

- HTML
- XML
- JSON
- CSV
- X

Teradata Staging (Python)
- CUST_RTN_ETL_STATIC_DATA.py
- CUST_RTN_ETL_SKC_TO_STG.py
- CUST_RTN_ETL_STG_TO_CORE.py

Teradata FTP (Python)

Teradata Staging

Audit
- ETL_Indicator_Proj_Audit

Audit (Blended Procedure)
*ETL_CUST_DATA_CORE
- ETL_LOOKUP_CORE
- ETL_BEA_CORE
- ETL_CENSUS_DATA_CORE
- ETL_CONSUMER_SENTIMENT_CORE
- ETL_COVID_CASES_CORE
- ETL_COVID_MODEL_CORE
- ETL_COVID19_DATAHUB_CORE
- ETL_FUEL_PROD_CORE
- ETL_GOOGLE_MOBILITY_CORE
- ETL_GOOGLE_TREND_CORE
- ETL_LABOR_STATS_CORE
- ETL_TSA_TRAVEL_CORE
- ETL_POST_LOAD_CORE

Exception
- ETL_Proc_Error_Logs

Teradata Core

Teradata Semantic (Views)

- F_IND_DASH_Covid_NAT_ESTIMATES_V
- F_IND_DASH_DataHub_V
- F_IND_DASH_Google_Search_Trends_V2_V
- F_IND_DASH_Health_V2_V
- F_IND_DASH_MACROECONOMICS_GEO_MONTHLY_V
- F_IND_DASH_MACROECONOMICS_V2_V
- F_IND_DASH_MOBILITY_GEO_V2_V
- F_IND_DASH_MOBILITY_GEO_WEEKLY_V
- F_IND_DASH_NYT_COVID19_DATAHUB_TD_EMP_LOG_V
- F_IND_DASH_NYT_COVID19_GEO_7MAVG_WEEKLY_GNSHT_V
- F_IND_DASH_Timeline_to_safety_V
- F_IND_DASH_COVID19_DATAHUB_STATE_V

Tableau

Teradata Semantic (Views)
Visualizing Data: Executive Dashboard

National Key Readiness Indicators

- New Cases: -16%
- New Deaths: -20%
- Retail & Rec. Mobility: 1.1 PPT
- Workplace Mobility: +1 PPT
- Unemployment Rate: -3%
- Consumer Sentiment: +4%
- Clothing & Footwear: +16%
- Food Service & Housing: +12%
- Motor Gasoline: +3%
- Jet Fuel: +17%

Back to Normal Snapshot

% of Population in Counties by COVID Status (Saturday):

- Safe: 58%
- Passing: 40%
- Surging: 2%
- Flat/Other: 0%

National Macroeconomic Outlook

- Unemployed: 9.71M
- Unemployment Rate: 6.00%
- % Change from Previous: -3%
- % Change from LY: -26%

Timeline to Safety, by State

- Number of states: 28
- Location: Florida, Nevada, New York, New Mexico, Oregon, Washington, Arizona, California, and others.

Number of states: 22

- Location: Florida, New Mexico, Arizona, California, and others.
Visualizing Data: Drill Down to Counties
Visualizing Data: Applied to First-Party Data
Predictive Modeling on Data

County-level cases and deaths

Economic factors

Blue indicates historical values, and orange represents Vantage-powered predictions.
Doing Good

- Made all code and templates available for free on github
- Framework for others to make their own solutions
- Installed for free at over 30 Teradata clients
Implementations

Source: https://stock.adobe.com
Helping You

- What challenges does your industry face as a result of the current situation?

- How can utilizing publicly available health and behavioral information improve your decisions?

- Where can the themes of this solution help you?

Source: https://stock.adobe.com
Hands On (Optional)

- Download Tableau Reader
  - https://www.tableau.com/products/reader

- Download Resiliency Dashboard
  - Main github site: https://github.com/Teradata/RTN
  - Click on Tableau folder
  - Select WhartonDashboard.twbx
  - Click on Download