Maximizing the Data Analytics Potential from a Data Warehouse

Andrew Herd

Software Development Manager - Kimbel Mechanical Systems, Inc

10 years experience

BS in Mathematics

NWA PMI Chapter

TechnoHerder.com

1

Lessons Learned & Value Gained

Lessons learned and values gained after creating a data warehouse.

There are four separate and distinct components to consider in the DW/BI environment: operational source systems, ETL system, data presentation area, and business intelligence applications.

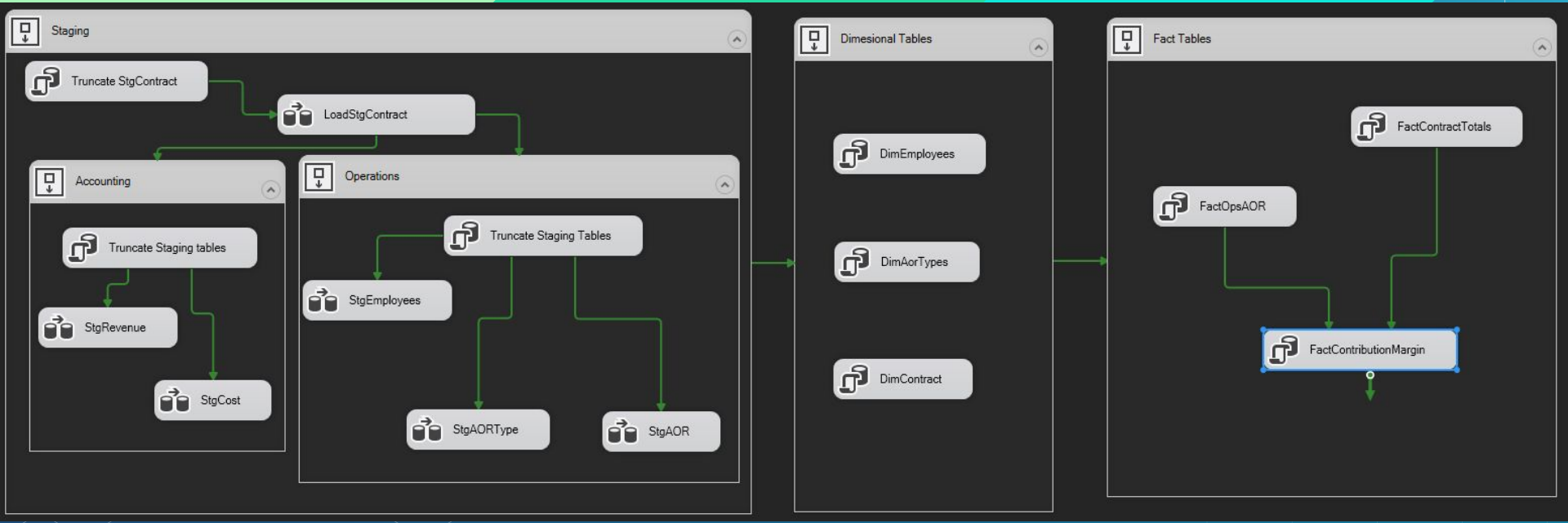
— Ralph Kimball, The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling

Data Warehouse ETL

Staging

Dimensions

Facts



Lessons Learned

- Start with the End in mind
- Denormalize the Data Warehouse
- Alert and process Bad data
- Track and Optimize reporting Queries
- Create Power User IDs for best security practices
- Use the DW for Analytical processing

Value Gained

- One source of Truth
- Quickly recognize trendlines and outliers
- Create Predictive Analytics
- Analyze profits per Operational grouping
- Compare trends with other sources (FRED API)
- Build datamarts to make Report building easier
- Free up resources on Production server

COMPARTMENTALIZE SERVERS

- Separate the Data Warehouse from the other Production servers.
- Balance the workload between servers.
- Let each server instance perform optimally for its one purpose.

2

Data Sets

Identify Data Sets that bring back value and can be leveraged for Artificial Intelligence and Machine Learning.

It’s all in the Data

- Start with the business; no data set can replace experience. Produce meaningful data
- What is behind the data they currently use?
- Do other markets trend closely with the business? What were ‘bad years’ and why?
- Look inward, what trends between regions?
- Continual improvement by adding and refining data sets

Prepare Data for AI & ML

Data Collection

Collect relevant data, quality over quantity. Automate processes for data. Design relevant attributes per data object. Use APIs to collect data.

Format & Clean

Format for consistency, this greatly aids aggregating the data in different ways. Use business friendly naming conventions.

Missing Data

Will you delete the record, exptrolate the value, forward/backfilling, regression model, or use mode/median/average?

Data Features

Transform raw data into features that better represent a pattern to the learning algorithms. Cold hard facts that can be decomposed into multiple parts.

Key Factors

Use common sense and ignore features completely irrelevant to the output. Get expert insight from the business.

Split the Data

Split the data into three sets: training, validation, and test. Training teaches the model, validation optimizes the model, and test to determine the quality of the model's output.

DATA SCIENCE

3

Data Science

Identify meaningful trends and patterns for strategic insight and leveraging algorithm creation to identify metrics and shift through data to recognize potential issues.

Meaningful Data

- Explainable data - build a report that tells a story that is easy to understand.
- Research historical outliers to understand causations and test theories with data sets.
- Re-evaluate models in a timely manner to ensure data quality with validation.
- Keep the business in the loop; use the business for Expert opinion.

Considerations

- What are calculated warnings levels?
- What effects are from the global pandemic?
- How would you add this to the model?
- What are the business experts seeing?
- Updating processes to generate new features.
- Are the reports easy to read?

4

New Outlook

Unexpected metrics developed from creating the data warehouse and how that changed strategy.

Looking inward

Analysis

We have the most data on ourselves. What are we unsure of? Looked at various timelines and organizational groupings to visualize what strategies work best. Predict operational costs and revenue.

Detailed Reporting

Created new dimensions to create finer detail in reporting. These new levels of detail identified areas of waste and increased profitability.

New Ideas

More Features

We need more data. We automate more processes to create more data collection points. We created more reporting and detailed analysis, never ending loop. Close absence of facts

Optimize Processes

New reports to check assumptions. Compare Bids to the Project's actual Contribution Margin. Optimize business estimations and processes.

Chat Bot

NLP in Power BI to generate data. Build a QnA Bot built on a knowledge base for internal jargon, processes, helpful links, etc. Chat bot for external facing groups, answer the easy questions.

THANKS!

ANY QUESTIONS?

You can find me at:

- TechnoHerder.com
- andrew.herd@kimbelmechanical.com

Supporting Material

Write up based from this presentation can be found here:

https://TechnoHerder.com/dataWarehouse.php

Download the project supporting that write-up from GitHub:

https://github.com/aherd2985/dataWarehouseSession

Used a template from