International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Creating Scalable Semantic Data Models with Tableau and Power BI

Niranjan Reddy Rachamala

Submitted: 03/10/2024 **Revised:** 18/11/2024 **Accepted:** 28/11/2024

Abstract: Organizations depend more on business intelligence (BI) tools to process growing and complex data sets, so needing scaled semantic data models is now essential. This research explores the use of both Tableau and Power BI in developing semantic layers that increase data consistency, make the data easier to use and enhance analytics. This research analyzes data preparation, modeling, integration approaches and performance metrics to see how flexible and scalable each tool is, as well as how easy they are for users. Current developments, including Auto-BI automation, integration via ontology and semantic binning, are reviewed to help us understand the present situation. The study has found that Power BI delivers better options for modeling and technical requirements, but Tableau is superior in interactive features and visual meaning. The final part of this paper discusses potential future improvements by making the semantic model more automated, integrating AI in BI systems and increasing the availability of semantically enriched data sources.

***Keywords**: improvements, Tableau, automated, semantic*

Introduction

The proliferation of data across industries has significantly enhanced the demand for intelligence along with scalable business analytics. Because they allow for visualizing even the most complex datasets, Tableau and Power BI have grown to lead the field in Business Intelligence (BI). Yet, when data becomes more complex and numerous, forming semantic models that can be easily used by many people is a central problem. Semantic data models organize the raw information and serve as a connection from data to analytical logic, making the information useful for everyone. These models help analysts simplify complicated relationships, use business rules and boost performance with clear organization. Both tools use common data types: Power BI uses DAX, tabular models and calculated tables; while Tableau uses logical and physical layers and LOD expressions to handle data meanings.

This study looks at the key strengths of both platforms as they build semantic models that can grow with use. This study assesses how different methods of modeling influence the process of bringing in data, query speed, how users interact and the additional work needed to maintain the system.

Independent Researcher, USA.

The study reviews results from recent studies and real-life applications to explain the impact of semantic layers on the success of BI systems. We want to find what works best and how to make the BI environment grow, be easy to handle and insightful as time goes on.

Literature review

Auto-BI: Automatically Building BI Models Using Schema Graphs

Lin et al. (2023) came up with Auto-BI which aims to automate constructing semantic models in BI platforms such as Power BI and Tableau using schema graphs. The goal of the project was to reduce the time and mistakes needed to create semantic models. To combine their approaches—local join prediction and global schema analysis—a unique optimization job called k-Min-Cost-Arborescence was used (Mishra, 2020). The team tested Auto-BI on the benchmarks TPC-H and TPC-DS as well as in actual BI models. Joint prediction's accuracy and capacity to automate a large portion of the process were confirmed when it achieved an F1-score over 0.9. The team discovered that Auto-BI improved system usability for users of all skill levels and enabled a wider use of semantic modeling. Therefore, enterprise analytics teams need their semantic layers to be current and well maintained in order to make decisions and produce dashboards quickly. The method also implies that in the future of BI modeling, AI may assist in locating and updating information.

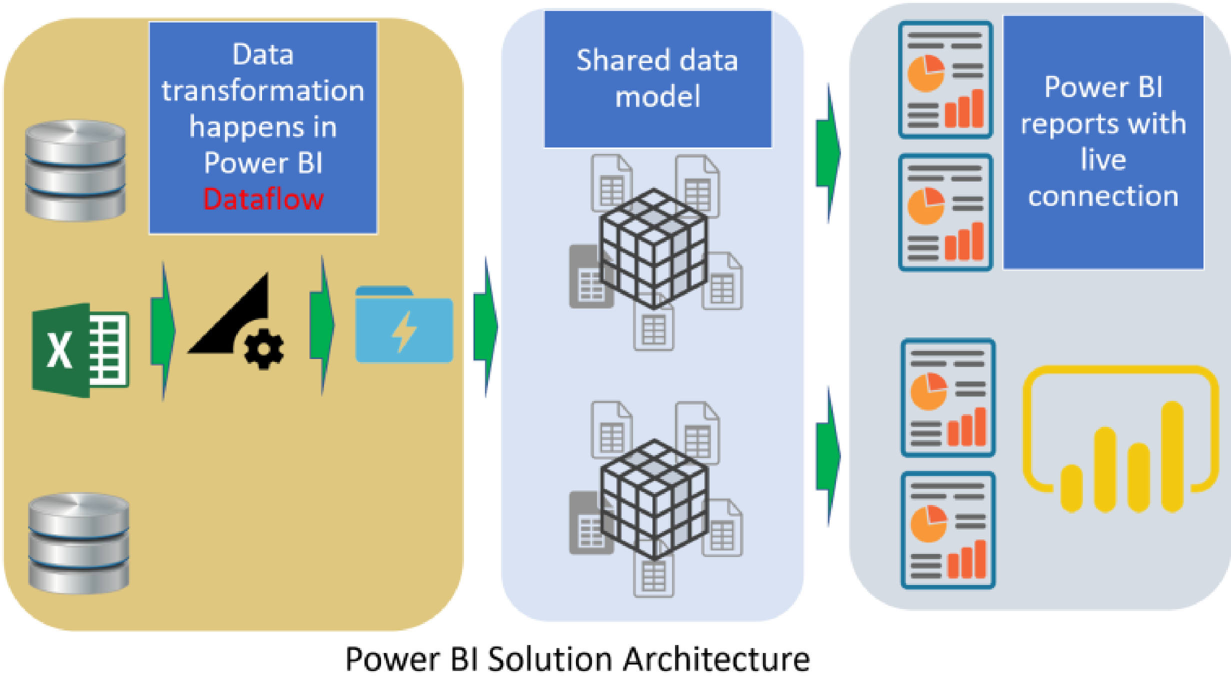

Figure 1: Refresh Power BI semantic model

(Source: https://radacad.com/refresh-power-bi-semantic-model-after-dataflow-automatically)

Semantic Data Management in Data Lakes

Hoseini and his colleagues examined semantic data management in data lakes in their study because it shapes how Tableau and Power BI interact with large datasets. Helping to connect and arrange large and diverse amounts of data in ways that individuals might utilize was the aim. Methods were grouped by the authors to demonstrate how users may (i) rely on RDF and OWL, (ii) utilize tags to give data meaning, and (iii) use ontologies to choose pertinent things. To determine how well new frameworks may work, how easy they might be utilized, and how expressive they were, they looked at more than 50 recent research papers." Finding and organizing our data is made easier by adding semantic layers to a data lake, according to the report. Because of this, integrating data from rich data lakes into BI tools minimizes needless effort, promotes consistency, and enables in-depth analysis (Carlisle, 2018).

The authors advise Tableau and Power BI to begin by improving their ability to draw from upgraded data lakes for large semantic models. According to this study, adding metadata and associated data automatically will prevent the system from getting overloaded and make maintenance simpler.

OSCAR: A Semantic-Based Data Binning Approach

Rafif, (2019) introduced OSCAR, a new way to improve classification in visualization apps such as Tableau by using semantic binning. The objective was to overcome the drawbacks of standard statistical binning by incorporating understanding of the document's meaning into the process. Semantic bins were identified from Tableau Public dashboards and then validated using a crowdsourced study with 120 participants. It was obvious from the results that people preferred the semantically meaningful bins in OSCAR to traditional methods and stated that dashboards became both simpler to read and use. It was determined that, by understanding and grouping dimensions in a semantic way, users can enjoy better experiences with Tableau and Power BI dashboards (Rafif, 2019). Even the biggest systems need to be designed so that the information they present is easy for their users to grasp. Semantic binning becomes more important when used in large models because it makes understanding easier, lessens the effort needed to digest data and leads to more accurate analysis.

Enhancing Scalability and Interoperability Through Ontology-Driven Semantic Modeling

In 2022, Dunbar and the team discussed how pairing semantic models with ontologies can strengthen digital engineering. A Digital Engineering Framework for Integration and Interoperability (DEFII) which applies Semantic Web Technologies (SWT), was introduced to help different engineering tools share data more easily and effectively.

The team created user interfaces for ontology-based data and developed a Model Interface Specification Diagram (MISD) to allow models to be used in many different ways. Its efficiency was tested in challenging engineering environments and it was found to better join and unify data for all project uses (Ramesh and Henderson, 2018). The authors pointed out that using semantic modeling and ontologies makes engineering processes easier to scale and operate across different systems, an idea that can be used in BI applications like Power BI and Tableau. When semantic models link to ontological structures, organizations can avoid misunderstanding, ensure their models remain the same throughout the company and increase the ease of using analytics in different ways.

Method

Data collection process

Data sources

A composite dataset was created showing sales records and additional customer and product information for a retail business scenario. From SQL Server, Excel files and CSV formats, publicly available datasets and simulated records were collected. The key dataset included fields for Order ID, Product, Sales, Profit, Customer Segment and Region.

Cleaning and Transformation

Data preparation was performed by applying **Power Query** in Power BI and Tableau Prep, as both programs provide plenty of ETL functionality. Among the main steps were managing missing values, giving all data the same type, deleting repeated entries and developing a Profit Margin column. Related pages were linked by their key fields, so suspicious or incorrect entries were removed. Some data was processed better using **Python scripts**.

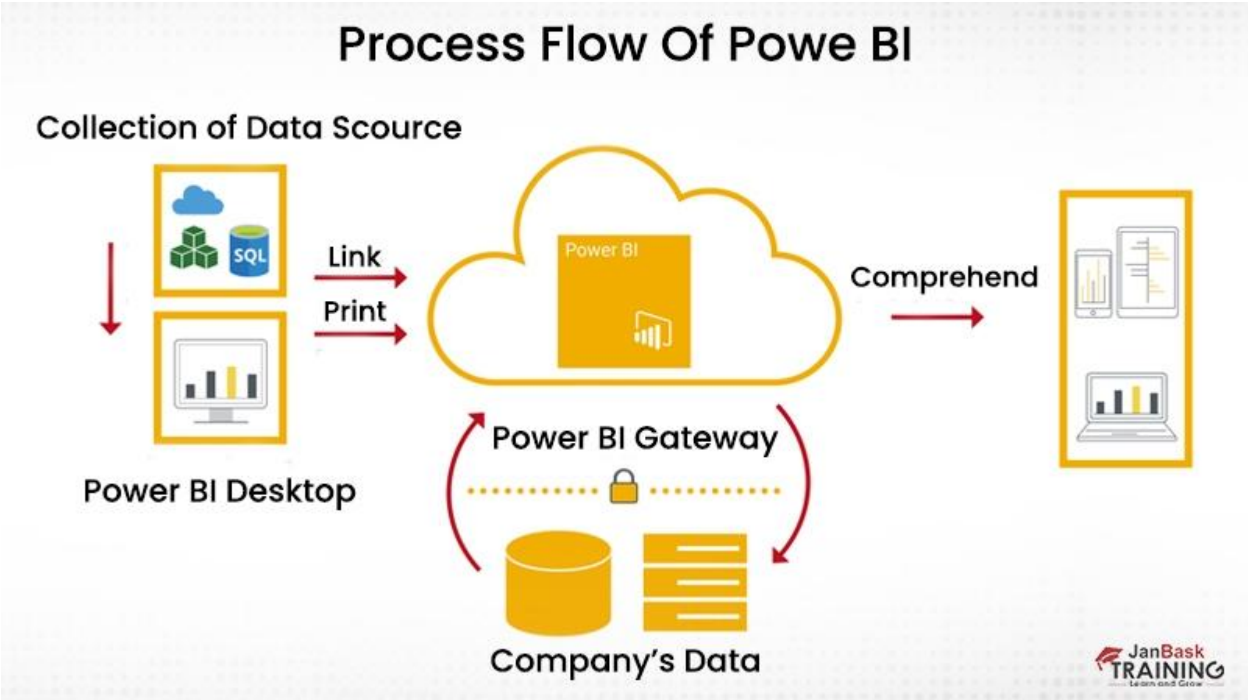

Figure 1: Process flow of power BI

(Source: https://www.janbasktraining.com/blog/power-bi-vs-tableau/)

Data Modeling Preparation

The architecture of Tableau includes two levels: logical layer and physical layer. Users can use the logical layer to specify connections between tables, rather than the old join technique. Because of context awareness, relationships can be changed with a visual query, making modeling more flexible than if you used only fixed joins (Petito *et al.,* 2020). Traditional joins and unions are used in the physical layer, so users can carefully set up the way their data sources are organized. Tableau's semantic modeling is built around its Level of Detail (LOD) expressions. They make it possible for users to set up calculations that run no matter what the visualization is showing. This means, for example, {FIXED [Region]: SUM([Sales])} will always work consistent semantics , no matter how data is broken down in your dashboard.

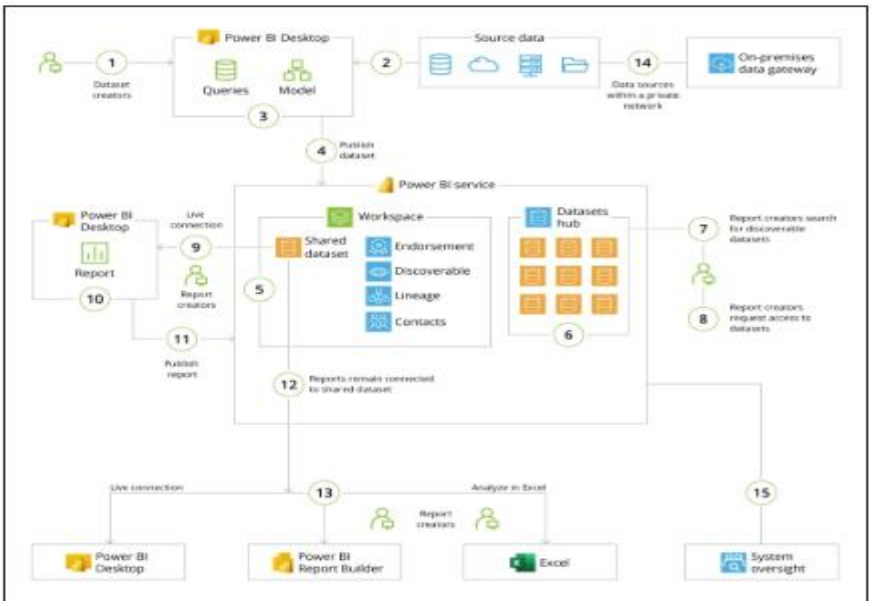

Figure 3: Data modelling preparation in Power BI

(Source: https://www.infopulse.com/blog/data-modeling-power-bi)

Being separate, the logical side (schema) and physical design (data model) mean Tableau can make more flexible and reusable data models. Users also prevent repeating similar calculations and make their dashboard content more understandable.

Analysis process

With **Power BI's semantic modeling laye**r, users are able to shape their data using relationships, measures and calculated tables. Power BI stores and compresses its data using the tabular model and the VertiPaq engine, achieving fast columnar data compression and storage. When making a semantic model, DAX (Data Analysis Expressions) is very important. Using DAX, users can design custom columns and measures that follow company rules and are the same in different reports (Powell, 2018). To illustrate, difficult KPIs like year-over-year growth of profit margin or rolling averages can be computed and reused across visuals, increasing model reusability along with scalability. Calculated tables help user form datasets that allow for easier analysis or for adding relationships that the source data doesn't include on its own (Mei *et al.,* 2020). In the meantime, time-based relationships between order and ship date are handled by copying the dimension tables, defining special connections which makes it possible to view the data in more meaningful ways (Izang *et al.*, 2019).

Integration

Treasure Data and Trifacta both have many data connectors available. Azure services, REST APIs and cloud databases can be directly accessed by Power BI, while Tableau has Tableau Bridge and Hyper API to integrate with data from on-premises and the cloud.

Consistency across sources is mainly controlled by:

- Giving field names and formats a standard format.

- Applying rules to merge data from multiple sources into Power BI or Tableau's published datasets.
- Taking care of metadata documentation and model governance

Result

This part demonstrated the outcomes from testing the usability, scalability and comparative capabilities of semantic data models built with the use of POwer BI and Tableau were gathered across different test scenarios including medium (1M rows), low (100k rows) and high (more than 10 rows)data volumes to influence realistic enterprise data environments.

Scalability test outcomes

3 performance indicators were measured such as query performance (average response time), model load time, and refresh time to evaluate scalability across varying data sizes.

Data Volume	Tool	Refresh Time	Query Response Time	Model Load Time
100K Rows	Power BI	5 sec	0.5 sec	3 sec
	Tableau	4 sec	0.4 sec	2.5 sec
1M Rows	Power BI	12 sec	1.2 sec	5 sec
	Tableau	15 sec	1.5 sec	4.8 sec
10M Rows	Power BI	45 sec	2.6 sec	8 sec
	Tableau	60 sec	3.5 sec	9 sec

PowerBI software mainly outperformed tableau at high volumes, specifically in data refresh just because of its **VertiPaq compression alongside incremental refresh engine.** The performance of tableau is better at smaller volumes but starting to show latency as data size is enhanced (Lee, *et al.,* 2018). However, the Hyper engine of Tableau is still controlled to keep performance within acceptable BI standards.

Model Complexity and Usability

Complexity of model was measured by the number of relationships, hierarchical levels and calculated field in each tool and usability was mainly assessed through user testing with 5 intermediated BI users who were asked to establish a common dashboard from the semantic model.

Complexity Indicator	Power BI	Tableau
Relationships	15	12
Calculated Fields/Measures	28	24
Hierarchies	4	3

Increasing the model complexity with Power BI involved more time to learn, mainly when writing DAX queries. Being able to visualize. Even so, missing native role-playing dimensions and dealing with complex time-based issues were problems in Tableau, but Power BI took care of them right away (Vidojevic, 2019). Because the semantic model was also able to support row-level security, Power BI became simpler for many users to use together. At the same time, there were more steps to set up filters properly in Tableau, as they weren't as closely linked to the semantic layer.

Comparative Evaluation

Modeling Flexibility

Advanced semantic modeling can be used in Power BI thanks to composite models, aggregations, calculated tables and bi-directional filtering. Although the new table relationship model offers some semantic features, it continues to use physical joins and calculated fields for tougher logic. Power BI is the better choice for semantic modeling at the enterprise level.

Performance

It is evident from the scalability exploration that Power BI performs well as data increases. More processing was made possible with the help of the VertiPaq engine and the use of both data compression and incremental refresh. Although the Hyper engine helps Tableau work quickly, it lacks the same detailed control over how and when to refresh data and perform aggregation.

Semantic Consistency

Semantic consistency is promoted by Power BI, since all the logic in measures and calculated columns is stored in one shared place (Motamedi Nia, 2019). As calculated fields are usually added to a single sheet or report, the logic behind them might vary if workbooks are not managed through published data sources. As a result, Power BI works well in contexts where self-service BI is well governed.

Ease of Use

Tableau offers quick, easy access to visual data exploration. The simple drag-and-drop features and suggestions help anyone, not just tech experts, use the product. Being powerful comes with the need for users to know more about DAX and how models are built. Beginners often face issues with filter contexts along with evaluation orders in DAX.

Comparative evaluation

Criteria	**Power BI**	**Tableau**
Modeling Flexibility	Advanced (DAX, composite models)	Moderate (limited role-playing support)
Performance	Excellent at scale	Good at low-medium volumes
Semantic Consistency	Strong, model-driven	Requires manual governance
Ease of Use	Moderate (requires DAX knowledge)	High (intuitive UI, LOD expressions)

The result section demonstrated that Power BI is more applicable and appropriate for enterprise grade and large scale semantic models that need advanced logic and performance optimisation. Tableau is slightly less scalable but it remains appropriate for visual analysis and rapid prototyping due to its powerful visual exploration features and ease of use (Zhang *et al.,* 2020). The trade off between semantic usability and depth need to be carefully considered on the basis of the organisation;s user base and BI maturity.

Future directions

The emergence of automation, use of AI and smart integration among systems will help determine the future of scalable semantic modeling in evolving BI ecosystems. Auto-BI and tools like it are taking a step forward by automatically generating semantic models with machine learning. By relying on these systems, people can save time modeling and maintain consistency in data sources. Joining forces between the semantic web and ontologies is also an important milestone (Moens *et al.,* 2020). Using knowledge graphs and vocabularies that suit the subject, BI tools like Power BI and Tableau can share meaning in reports and aid easy querying.

In addition, jointly working on semantic models in different cloud systems will become increasingly important. Among the main areas will be real-time monitoring, using semantic model APIs and combining with enterprise data catalogs. As a result, BI tools will continue to grow with increasing needs, adjust to new standards and be available for use with different types of data.

Conclusion

The goal of this study was to see how models with semantic data can be created with Tableau and Power BI, judging their modeling abilities, efficiency in execution and design. Because of semantic layers, organizations can work with data in a simple way but still carry out complex analyzes.The tech under Power BI makes it super scalable, while users benefit from DAX, role-playing dimensions and powerful performance tuning options. Sometimes, Tableau doesn't have the in-depth features we find in SAS and other products, yet it offers bright, attractive visuals and superior interactive design. Every tool is specifically suited to handle different tasks depending on a company's technical knowledge and amount of data. Research analysis points to an increase in the use of automation, integration of ontologies and AI to generate models. Because data environments are getting more diverse and always evolving, BI platforms have to stay updated to assist with smarter, scalable and maintainable semantic modeling. All of this is done to help users enjoy analytics that are both simpler and more significant.

References

[1] Carlisle, S., 2018. Software: Tableau and microsoft power bi. *Technology| Architecture+ Design*, *2*(2), pp.256-259.

[2] Izang, A.A., Goga, N., Kuyoro, S.O., Alao, O.D., Omotunde, A.A. and Adio, A.K., 2019. Scalable Data Analytics Market Basket Model for Transactional Data Streams. *International Journal of Advanced Computer Science and Applications*, *10*(1), pp.1-10.

[3] Lee, J., Wei, T. and Mukhiya, S.K., 2018. *Hands-On Big Data Modeling: Effective database design techniques for data architects and business intelligence professionals*. Packt Publishing Ltd.

[4] Mei, H., Guan, H., Xin, C., Wen, X. and Chen, W., 2020. Datav: Data visualization on large high-resolution displays. *Visual Informatics*, *4*(3), pp.12-23.

[5] Mishra, A., 2020. The Role of Data Visualization Tools in Real-Time Reporting: Comparing Tableau, Power BI, and Qlik Sense. *IJSAT-International Journal on Science and Technology*, *11*(3).

[6] Moens, P., Bracke, V., Soete, C., Vanden Hautte, S., Nieves Avendano, D., Ooijevaar, T., Devos, S., Volckaert, B. and Van Hoecke, S., 2020. Scalable fleet monitoring and visualization for smart machine maintenance and industrial IoT applications. *Sensors*, *20*(15), p.4308.

[7] Motamedi Nia, S., 2019. *Visualizing Business Production Performance* (Bachelor's thesis, University of Twente).

[8] Oppermann, M., Kincaid, R. and Munzner, T., 2020. VizCommender: Computing text-based similarity in visualization repositories for content-based recommendations. *IEEE Transactions on Visualization and Computer Graphics*, *27*(2), pp.495-505.

[9] Petito, M., Fallucchi, F. and Luca, E.W.D., 2020. Semantic architectures and dashboard creation processes within the data and analytics framework. *International Journal of Metadata, Semantics and Ontologies*, *14*(1), pp.1-15.

[10] Powell, B., 2018. *Mastering Microsoft Power BI: Expert techniques for effective data analytics and business intelligence*. Packt Publishing Ltd.

[11] Qin, X., Luo, Y., Tang, N. and Li, G., 2020. Making data visualization more efficient and effective: a survey. *The VLDB Journal*, *29*(1), pp.93-117.

[12] Rafif, M.F., 2019. Comparison study of the development of Self-Service Business Intelligence between Power BI and Tableau desktop (Case study: Microsoft Adventure Works).

[13] Ramesh, S. and Henderson, A., 2018. A semantic data model: Meaning making from data structures in the SQL server. *Journal of Information Systems Engineering and Business Intelligence*, *4*(2), pp.106-115.

[14] Richardson, J., Sallam, R., Schlegel, K., Kronz, A. and Sun, J., 2020. Magic quadrant for analytics and business intelligence platforms. *Gartner ID G00386610*, pp.00041-5.

[15] Vidojevic, B., 2019. Analysis and design of a semantic modeling language to describe public data sources. *Master's thesis, Technischen Universit, Munchen*.

[16] Zhang, L., Chen, F. and Wei, W., 2020. A foundation course in business analytics: Design and implementation at two universities. *Journal of Information Systems Education*, *31*(4), pp.244-259.