

Tableau Dashboards Sourced from OLAP vs. RDB: An Analysis

Seattle-based Tableau Software is widely recognized as one of the leading providers of data visualization products that focuses on Business Intelligence (BI) and Business Analytics (BA). Their systems help corporate users analyze, visualize and share information.

Tableau has been ranked by Gartner and IDC as one of the world's fastest growing software companies. With a customer base of over 10,000 companies, and tens of thousands of people using Tableau to share data in their blogs and websites, they are a force in the BI world. There are other data visualization systems in the business intelligence market, but the purpose of this article is to specifically discuss Tableau dashboards sourced from Online Analytical Processing (OLAP) vs. Relational Database (RDB) sources.

Reasons for Using Tableau

Tableau is a popular BI and Analytics tool among technical and non-technical users for several reasons.

- **Self Service BI.** Users of corporate BI and BA systems can create their own reports and dashboards without putting additional strain on the BI development team.
- **Quick Development.** Tableau is designed with quick and straight forward design and drag/drop features that allow non-technical users to quickly build effective dashboards.
- **Data Visualization.** Tableau allows for a highly interactive and effective analytical experience for users and provides professional layout templates and features for meaningful reporting and dashboarding.

A Typical BI and Analytics User Profile

While BI and Analytics users come in many varieties (finance, accounting, operations, marketing, sales, and more), here is a classic profile for a typical user. The user:

- Has very little, if any, database or technical knowledge outside Microsoft Excel—where they are often quite knowledgeable
- Is not educated on relational database concepts such as tables and metrics
- Has different formulas, calculations and requirements for corporate metrics—based on their own needs or departmental needs and procedures
- Needs required data to be updated frequently throughout the day in order to make the best business decisions

Because of this typical profile, the need for a BI dashboarding system that is easy to create, navigate and use, is very important. The platform needs to provide near-instantaneous data query results and with data that is refreshed frequently.

Whether they realize it or not, most users also need a system that is reusable and maintainable. All underlying components need to be reused by other tools and users. Dimension hierarchies and business metrics should be standardized across all business units and maintained in a central location to avoid data errors. Based on these factors, below, you will see the pros and cons of users obtaining their reports and dashboards from an OLAP source vs. a RDB source.

Simplified Tableau Dashboard Creation

Description	Best Source
<p>Users do not have the expertise to build database relationships or join database tables from a back end database.</p> <p><u>OLAP</u></p> <ul style="list-style-type: none"> • Data relationships already exist • Defined hierarchies and metrics exist <p><u>RDB</u></p> <ul style="list-style-type: none"> • User needs to re-create data relationships • All metrics and formulas need to be re-created • User needs much more technical expertise <p><u>Risk</u></p> <p>If users are given the opportunity to create their own data relationships and metrics, they could become confused, build the wrong data relationships, and ultimately lose confidence in, and not use, the system.</p>	<p>OLAP <input checked="" type="checkbox"/></p> <p>RDB</p>

Fast Data Query Performance

Description	Best Source
<p>When comparing the performance of OLAP vs. RDB, it is clear that OLAP outperforms a RDB when querying data.</p> <p><u>OLAP</u></p> <ul style="list-style-type: none"> Created specifically to aggregate data Hierarchical data structures well-defined Data retrieval performance exceeds that of a RDB <p><u>RDB</u></p> <ul style="list-style-type: none"> Fact tables can contain millions of rows which is time-consuming for a RDB to aggregate User needs to understand table relationships to get proper performance gains <p><u>Risk</u></p> <p>Users do not understand the design of the RDB warehouse. If the users are given access to build their own relationships, then it is possible for the queries to run poorly due to invalid associations. The BI development team has typically spent many weeks or months building a fast and robust OLAP design. The performance of the OLAP cube will exceed querying directly against the database.</p>	<p>OLAP <input checked="" type="checkbox"/></p> <p>RDB</p>

Low Latency Data

Description	Best Source
<p>The data warehouse is continuously updated throughout the day. Often every 15 minutes for some fact tables.</p> <p><u>OLAP</u></p> <ul style="list-style-type: none"> Typically scheduled to incrementally process every 15 minutes Typically scheduled to fully reprocess every hour Small margin of time for data latency <p><u>RDB</u></p> <ul style="list-style-type: none"> Fact tables typically updated every 15 minutes/one hour Dirty reads or locking of tables when refreshing <p><u>Risk</u></p> <p>If Tableau is sourced to a RDB, it would need to continuously retrieve data from the fact tables and put the new records into memory. That can create a risk that the data is not updated frequently enough, especially with large fact tables. If Tableau is sourced to OLAP, the user will always get the most recent dataset and without dirty reads.</p>	<p>OLAP <input checked="" type="checkbox"/></p> <p>RDB</p>

Components Exist in One Location for Reusability and Maintainability

Description	Best Source
<p>An OLAP cube has already been created and is being utilized for other reports within the business unit(s).</p> <p><u>OLAP</u></p> <ul style="list-style-type: none"> • Allows for one version of “the truth”—one place for metric calculations and one place for hierarchical structures • OLAP can be utilized by multiple user interfaces providing the same data, e.g. Excel, SSRS, Tableau • Any changes to the cube will be immediately reflected in all user interfaces <p><u>RDB</u></p> <ul style="list-style-type: none"> • All OLAP metrics and hierarchies would need to be recreated for use by Tableau • Two versions of “the truth” would exist: OLAP and the RDB <p><u>Risk</u></p> <p>The BI team will lose the ability to reuse OLAP components if Tableau was sourcing the RDB. This would cause redundancy, possibly different values for metrics, and increased development time due to the need of making changes in two places. The chance of mistakes goes up.</p>	<p>OLAP <input checked="" type="checkbox"/></p> <p>RDB</p>

Summary

If your organization has built an OLAP cube for reporting, Tableau should be sourced from the OLAP cube rather than a relational database source due to query performance, less user knowledge required, reusability of the OLAP components across multiple interfaces, and the maintainability of hierarchies, metrics and calculations.

Technical References

Reference 1

<http://www.biprofessional.com/2012/04/a-quick-look-at-tableau/>

Conclusions

When using a RDB:

1. It takes a couple of hours to retrieve 100M records before the user can see data. With OLAP, data will exist on demand. **Advantage: OLAP**
2. It seemed “at least as fast” as an OLAP cube. **Advantage: OLAP**

3. Users are required to be comfortable with arranging hierarchies themselves.
Advantage: OLAP
-

Reference 2

<http://kb.tableausoftware.com/articles/knowledgebase/functional-differences-olap-relational>

Conclusion

When using a RDB, many features and functionalities are exposed to the user versus using OLAP. This is due to the user needing to recreate every single object found in OLAP which causes extra work and redundancy. **Advantage: OLAP**

Reference 3

<http://www.practicaldb.com/data-visualization-consulting/tableau-vs-qlikview/>

Conclusion

Tableau is the ONLY advanced data visualization tool that can use SSAS cubes which makes it a much more attractive reason to use it. **Advantage: OLAP.**

Reference 4

<http://www.interworks.com/blogs/bbausili/2011/12/07/three-points-remember-success-tableau-and-olap-cubes>

Conclusions

1. By using OLAP cubes, a user will have analytical features not found when using RDB.
Advantage: OLAP
 2. Users/development team would need to re-create all the formulas and metrics if not sourced from cube which would require more work. In most environments, this is not feasible. **Advantage: OLAP**
 3. The BI development team knows how to code in MDX, but users do not. This would require the development team to recreate formulas and calculations that already exist. **Advantage: OLAP**
-

Reference 5

http://onlinehelp.tableausoftware.com/current/pro/online/en-us/performance_multidimensional.html

Conclusion

The above article contains four bullet points about improving OLAP performance. The main concern is that with large dimension members, the user should attempt to filter the members to speed up performance. **Advantage: OLAP**

http://onlinehelp.tableausoftware.com/current/pro/online/en-us/performance_relational.html

Conclusion

The above article contains eight bullet points about improving RDB performance. The main concerns include: needing to aggregate metrics (not a default), creating an extract so Tableau contains to an offline source (now stale data), and breaking up your data into smaller pieces for easier access (not feasible in the DW). **Advantage: OLAP**

By Justin Taylor, StatSlice Systems

Justin, a StatSlice Manager Consultant, (www.statslice.com), has been architecting, building and managing business intelligence projects since 1999, including solutions for Fortune 500 companies. When Justin joined StatSlice, he was looking for an opportunity to make a major contribution to the company and utilize his technology experience to help clients solve their business problems. He enjoys opportunities to speak at technology events, conferences and training. He holds several Microsoft certifications include MCT (trainer), MCTS and MCITP certifications in SQL Server BI, and Certified Scrum Master through ScrumAlliance.org.

For More Information

For more information about StatSlice Systems products and services, call (214) 206-9290 or email us at info@statslice.com. Please visit us at <http://www.statslice.com>