Retail Planning Systems

How to Really Measure Scalability

Business Value Research Series

Table of Contents

Overview

Background

During the past decade, several software vendors have emerged with the goal of providing automated tools that offer retailers a systematic approach to inventory planning and management. These vendors did a great job of providing the first real generation of top-down and bottom-up inventory planning tools that allowed retailers to have better visibility and centralized control of their businesses. Initially, early planning solutions were revered given the standard approach used by most retailers: hard-copy reports and silo Lotus™ spreadsheets linked together through manual processes.

However, it is often said, "retail is detail," and the need to incorporate more granular data into the planning process became the constant challenge. What started out as a major productivity tool for managing the business at higher echelons—typically department or class levels across the chain—soon became unacceptable. As retailers started planning at lower levels, it became evident that this new breed of planning solutions was limited in its ability to scale. As a result, planners had to choose between best-practice functionality and the ability to plan down to a desired level. As a result of this trade-off, the prevalent planning tool for most retailers became (and still is) Excel™ and Lotus™ spreadsheets, while these first-generation planning tools were relegated to supporting only discrete parts of the planning process.

Why Scalability Matters

Planning is far from a one-time, sequential, consistent process. Rather, planning— especially in-season—can best be described as an iterative, optimization odyssey where art meets science. In a world where each planner requires maximum flexibility to continually reconcile top-down with bottom-up and middle-out planning analyses and must frantically simulate planning scenarios to compare this year, last year, current forecast and several "what-if" reforecast models to determine the right plan (or at least the right plan for today), there are infinite combinations of activities that a planning system must support to maximize benefit and drive ROI for the planner.

Further complicating this need for flexibility is the need to manage significant volumes of data in a non-linear manner. Whether you are a specialty retailer or department store driving GMROI per square foot or a mass merchant or grocery chain collaborating with your suppliers to manage the store shelf, the level of granularity at which you drive your business has become the key differentiator among all retailers. The challenge is how to process and uncover the potential opportunities to more effectively merchandise across the millions of SKU/location combinations within your chain and across your channels in a manageable way. Current planning systems' shortcomings consistently fail at the intersection of plan data volumes and flexible functionality.

The Bottom Line

Retailers must understand the truth about measuring scalability and assessing technology for planning systems to ensure that they are selecting a planning platform that will support both the current standard and potential future vision for planning best practices. Next-generation technologies and new software design techniques are allowing traditional planning metaphors to be redefined. A glimpse at the future—where the planner has real-time access to more granular customer information and causal factors about consumer behavior, as well as the ability to apply deeper science across millions of SKU/locations—reveals a new set of opportunities for retailers to plan in ways never before envisioned. Current retail planning systems have not re-architected their solutions to support traditional planning processes and do not provide a next-generation architecture designed to support retailers as they incorporate more complex planning techniques. Planning and IT organizations must work in partnership to analyze planning vendor performance claims and use the techniques outlined in this white paper to assess scalability.

This white paper includes the following sections:

- Measuring Scalability: Describes the metrics that should be used to measure scalability;
- Assessing Architecture: Describes the key architecture components and design techniques that should be considered when selecting a scalable technology platform.

Measuring Scalability

There are two measurement criteria that must be accurately defined to measure scalability:

- Data requirements as measured by cell counts;
- Online performance requirements as measured by seconds per screen.

Data Requirements

Plan data volumes are measured by cell counts. Cell counts represent the number of possible combinations of plan dimensions for each plan measure. There are two cell counts necessary to understand potential scalability:

- Per plan workable cells;
- Per company workable cells.

Creating the basis for cell count calculations

Cell count calculations are simply determined by multiplying the maximum number of values for each dimension for a retailer's specific planning methodology. For most retailers, there is a minimum of five dimensions of data that must be considered to adequately measure the quantity of data that must be managed by the planning system to be effective for the planner. These include: *Product x Time x Location x Measures x Plan Versions*

A Representative Example for a Typical Retailer

The example below represents the scope of dimensions faced by a typical retail planner performing assortment planning at the item level.

- Product
  - SKU level
  - Whether you are a department store with a large number of department/class combinations and high SKU count with fewer locations, or a specialty retailer with a smaller SKU count but more locations, a conservative method for measuring scalability should be based on a reasonable number of key items (SKUs) that would typically be managed by a planner. Let's assume, on average, that there are 125 items in a class and that all items are being planned.
- Time
  - Year, season, quarters, months and weeks in a single plan or company plan translating into 72 periods in a fiscal planning period.
- Location
  - Most retailers cluster stores/locations. We have assumed 5% of store count (i.e., 800 store count; 40 clusters) for this example.
- Measures
  - Reflects the planning calculations. For assortment planning, there are minimally 10 KPIs, including:

- Beginning inventory
- Sales—regular
- Sales—promotional
- Sales—markdown (close-out)
- Receipts
- Open order
- Adjustments (sometimes split into inventory, employee discount, shrinkage)
- Ending inventory
- Ratios to get inventory (forward weeks of supply, stock to sales, inventory turn, sell through %)
- Sales, inventory and receipt dollar variances; plan to forecast and plan to actual
  - Assuming that the above KPIs are all unit measures, most retailers add Average Unit Retail (AUR) and their derived complements for each of the above measures—essentially doubling the set of KPIs in a plan.

- Versions
  - Provides integrated comparisons of a plan for simulation and comparison purposes. "Integrated" specifically means the ability to line up each measure's version side by side within the same plan to be effective.

Most planners use LY, forecast and several versions of in-season "what-if"/re-forecast models associated with each of their plans.

The following matrix quantifies the workable cell counts that must be stored in a plan to support a planner's typical process for a typical retailer for item-level planning.

Product	Time	Location	Measures	Versions	Per Plan Cell Count	Minimum No. of Plans	Total Company Cell Count
125	72	40	10	5	21,600,000	25	540,000,000

Online Performance Requirements

Planning systems must also provide functionality that meets acceptable performance measures for the stated data requirements in the figure on page 3. On page 4 is a guideline that identifies key planning system functions that typically limit retailers from implementing best practices due to performance constraints. The third column details goals for function performance time, based on current peak system capabilities.

Function Task	Description	"Not To Exceed" Based on Average per user plan size (20-25 Million Cells)
Creation of Initial Plans	• Typically, a system administration function one-time setup for each fiscal year or season • Creating a single plan and defining its dimensions • Includes seeding plans	< 15 minutes.
Opening Plans	User selects a plan from the "find plan" list and opens it.	< 10 seconds
Saving Plans	User is on a plan worksheet and saves the plan data to the database; there should be no impact based on number of changes since the last save.	< 10 seconds
Automatic Prorate and Summate/Calculating	User enters values at different levels of the merchandise/ location hierarchy and calculates new values. • Sales changes • Receipt/purchases changes • Shipment changes • Any other changes	< 5 seconds
General UI Client User Interaction on Plan Views • **Rotate/Rotation of Plan Dimensions** • **Expand Collapse of Dimensions** • **Displaying Views** • **Drag & Drop Dimensions** • **Displaying Plan Versions**	User is on a plan worksheet and is performaing UI functions. Times vary based on number of visible cells in a single view.	• Rotate < 2 seconds • Expand/collapse < 2 seconds • Customized views < 5 seconds • Drag/drop < seconds • Create/integrate plan version (TY, LY, What-if) < 2 seconds
Lock/Unlock	User performs the lock/unlock feature to sales values.	< 10 seconds
Plan Groups	• Plan groups are used to build plan groupings from single plans. For example, a group plan is used to create a sum of class plans in department 10 of the sum of class plans in division 1. • Each template (KI & FP) is a separate group of plans.	< 2 minutes
Plan Seeding – Initialize a plan	• Seeding allows the plan to be initialized with last year or another plan version	< 2 minutes

Assessing Architecture

IT organizations must not look at benchmarks alone to assess the scalability of a planning system. They must look at the underlying architecture to understand the platform's ability to not only support current planning processes and performance criteria, but to also continue meeting scalability requirements as retailers push the limit on more granular planning. To identify those planning products that will be platforms of the future, there is one simple question to explore:

Does the planning system architecture have a multi-dimensional cube or a relational database-centric design?

Though great marketers including the likes of Makoro and Arthur used the term "cube" when describing their planning systems' underlying architecture, it was really a way to illustrate the concept of three dimensions once time was added to product and location.

It explained how plan views could be flipped around from the top down. This was a tremendous stride forward over two-dimensional analysis provided by Lotus and Excel at the time.

However, the "Achilles heel" for all of the vendors offering planning solutions was the resulting performance limitations that came with offering this flexibility. Planners faced a constant trade-off between level of data and functionality, forcing planners to carve up their planning processes into smaller discrete components. This translated into a proliferation of disparate plans across the merchandising organization, often forcing the same planner to manage multiple plans within the same departments and classes that they owned.

There is one common explanation for this occurrence: non-scalable architecture. All of these planning systems were predicated on a relational database design where planning application logic needed to access the database via traditional methods for real-time bottom-up aggregation and top-down cascading. All technologists know the significant limitations caused by I/O intensive architectures, especially for interactive, real-time applications. In a less-than-successful attempt to improve performance, vendors delivered "heavy-client"-designed applications where much of the logic and data were brought to the desktop. Then, exacerbating the problem, vendors tried to deliver the benefits of "browser-based" applications by Web-enabling their user interfaces.

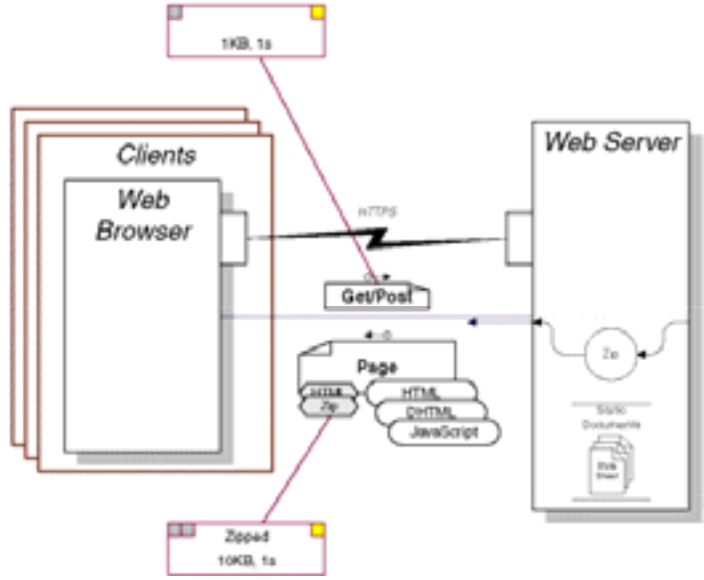

Figure 1: Thin Web application client

Unfortunately, they did so without fundamentally redesigning their underlying architecture, resulting in both increased performance problems and decreased friendly user interfaces.

Given this performance challenge and the technical design limitations of the retail market's current planning vendors, successful planning systems should have architecture analogous to the cube designs used by the most scalable relational online analytical processing (ROLAP) engines in the market. However, ROLAP technology is read-only. Therefore, a scalable planning platform should employ a variety of innovative technical design techniques to build a ROLAP-like engine that is both READ and WRITE.

There are two key architecture areas that must be employed by planning systems to overcome performance challenges of the past:

- User Interface
- Data Architecture
  - Data storage
  - Multi-dimensional design
  - Data access

User Interface

Planning solutions must deliver a very robust user interface to support the usability needs of planners. Early "client-server" architectures based on heavy-client designs were able to deliver rich user interfaces that adequately supported the wide variety of functional capabilities demanded by planners. Of course, we know this functionality was a trade-off with plan granularity. As the browser-based user interface (UI) became the preferred method of delivering software, planning vendors struggled to deliver even comparable functionality to the heavy-client model. Today's ideal planning system should deliver a rich user interface that exceeds prior UI paradigms while still maintaining a browser-based platform.

The design challenge is to minimize network traffic between the browser and the application server. This is accomplished through three key methods:

- Avoid applets or plug-ins
- Utilize data compression
- Reduce "chattiness"

A very "thin" true Web application client that does not include applets or plug-ins delivers rich user interface functionality. The user interface can be delivered on any standard Web browser that supports Microsoft Internet Explorer 5.5 and higher. HTML/HTTP is used to transfer presentation information to the user's display terminal.

If a planning system employs standard Web client abilities and does not include applets or plug-ins, the application is able to "compress" the data sent to the user, shrinking even extremely sophisticated 2 MB pages down to 30 KB of data. This enables a high-quality UI, very little network traffic and fast performance.

Further, a thin-client application includes less frequent and smaller interactions with the application server (i.e., reduced "chattiness"). Unlike applications that use applets or plug-ins, which stress the network's bandwidth because of the high volume of requests between the applet and the server, a thin-client UI has a core rhythm of interaction in a second or two instead of a slew of interactions that ultimately transport the necessary information to the client.

Manhattan Associates' market research shows that all current planning solutions, other than its own, that have attempted to migrate from client-server to browser-based user interfaces use Java applets or other plug-ins on the desktop, leading to numerous issues as they scale both the number of users and plan granularity. Additionally, Java applets require larger PCs, as well as increased memory and processing speed, thereby increasing total cost of ownership—something that should be lower in a

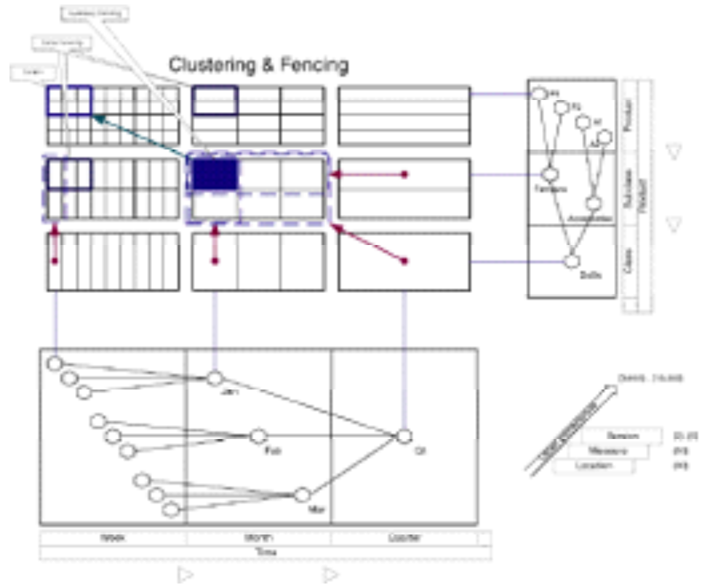

Figure 2: Clustering optimizes user experience

browser-based environment. Overall, current planning vendors must redesign their underlying architecture to reduce the use of applets and plug-ins to deliver a high-performance and robust functional user interface.

Data Architecture

Planning system architecture should organize internal data in an intelligent and performance-oriented manner. Although traditional relational database-centric applications may have database designs for optimal performance, they typically are purpose-built for a static set of functional requirements. Truly scalable planning systems rely on data architecture that enables continued increases in planning functionality sophistication—such as future increases in planning granularity—while maintaining high performance levels. One key design principle is to de-couple the data architecture from planning functionality, thereby allowing additional functional requirements to be added without whole design changes to the database—for example, easily add new planning dimensions such as customer, channel and/or vendor without having to redesign the rows and columns in the underlying database tables while meeting expected performance levels. Key data architecture components for maximum scalability include:

- Non-relational data storage
- Multi-dimensional plan management

Data Storage

Sophisticated plan data storage should use a technique that, for the purposes of this paper, we'll term "clustering." "Clusters" are objects responsible for retrieving and storing plan data in appropriate units of granularity. Clusters are initially created with the definition of a plan, but can be further re-tuned based upon plan usage characteristics.

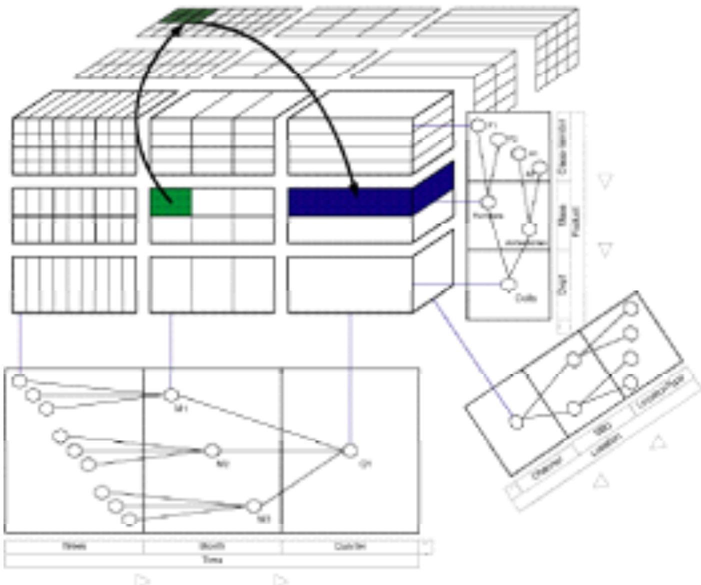

Figure 3: Multi-dimensional design

In addition to the ability to track information state in detail, each cluster leverages a set of CLOBs that stores plan data in a large-grain, compressed and even human-readable form. Clusters can interact with the database in ways to limit conversations and to make sure data is always available, without overloading memory, within the application server itself.

In order to achieve acceptable performance levels, traditional planning solutions typically store totals within the relational database at higher levels of product, location and time. Although this common technique may generate acceptable interactive performance levels, there is a significant performance impact on real-time plan changes that necessitates recalculations of totals—often requiring batch processing for plan totals to recalculate from the lowest level of the plan.

Multi-Dimensional Design

Planning architecture should have intelligent business rules that define the relationships among measures and cells. As a result, a planning system should use this information to efficiently process calculations and access data for only what is necessary to maintain plan integrity. By using this design technique, planning systems can optimize performance areas, including information retrieval and expansion, data updates and calculation staleness. For example, if a planner changes his planned receipts, the application knows that EOP must be recalculated, but all measures related to sales can remain the same.

Multi-dimensional analysis cleanly identifies base knowledge from derivable information and does not require reading, updating, locking and otherwise working with aggregate cells as part of normal behavior. The count of aggregate cells can grow to be larger than base vertices, so by removing these superfluous interactions and their intertwining relationships, the application avoids overhead of unnecessary aggregation.

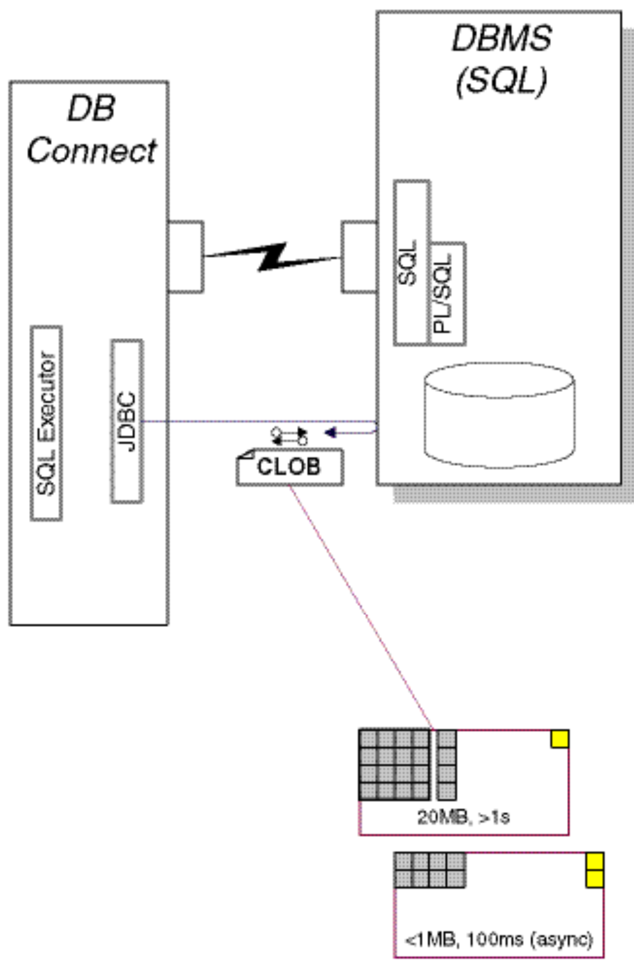

Figure 4: Data access–limited I/O interaction

Data Access

Manhattan Associates' market research indicates that current planning systems not only read the database during the opening of a plan, they also access it and recalculate the view every time a user rotates information on the screen. Therefore, there is a continuous interaction between the application and the database, leading to this "chatty" behavior and slower performance.

Therefore, reducing I/O between the application and database is critical to optimizing performance. To accomplish this, when plans are presented to the user for general use, the system reads the database once during the open plan function. The plan is then housed in memory in its compressed CLOB form, and expanded where needed as user access is requested. Since the plan is resident in memory, no additional database reads are required to access the plan view. The only other interaction with the database occurs during the plan save function. Any other interaction with the plan view, such as rotate, calculate, drop-down selections, custom view or vendor view, accesses the data directly from memory. This makes plan interaction extremely fast.

Conclusion

This white paper should be used as a guideline to help retailers understand planning vendor performance claims and accurately assess the scalability of planning systems when selecting a solution. Current planning systems consistently fail at the intersection of plan data volumes and flexible functionality. Planners and technologists must work in partnership to fully assess a planning system's scalability. The key tenets of a successful scalability assessment include defining and measuring the right performance criteria and assessing the underlying technical architecture.

Measure Accurately

Retailers must ensure that correct measurements are used to assess the quantitative aspects of measuring scalability. These include:

- Per plan workable cells
- Per company workable cells
- Online screen response times

Architecture Matters

Retailers must also consider technical architecture when assessing scalability. Planning solutions must not only support traditional planning processes, but also be capable of providing a long-term platform that will be able to support how planning will be conducted in the future. The technical architecture must include proven scalability principles combined with sophisticated design techniques in the three key architecture areas as described within this white paper to be called a scalable platform.

Scalability within your planning system is key to a positive return on your investment – in the immediate future and for years to come. Whether you are looking to increase your number of stores, diversify your product line, add distribution centers, or expand across business channels, a truly scalable planning solution will enable you to respond quickly to changes in your business, improving productivity, service levels, and profitability.

About Manhattan Associates, Inc.

Manhattan Associates is a leading supply chain solutions provider. The company's supply chain planning, supply chain execution, business intelligence and business process platform capabilities enable its more than 1200 customers worldwide to enhance profitability, performance and competitive advantage. For more information, please visit www.manh.com.

Offices in:

Australia
China
France
India
Japan
Mexico
Netherlands
Singapore
United Kingdom
United States

Global Headquarters
2300 Windy Ridge Parkway
Seventh Floor
Atlanta, GA 30339
tel +1 770.955.7070
fax +1 770.955.0302
www.manh.com