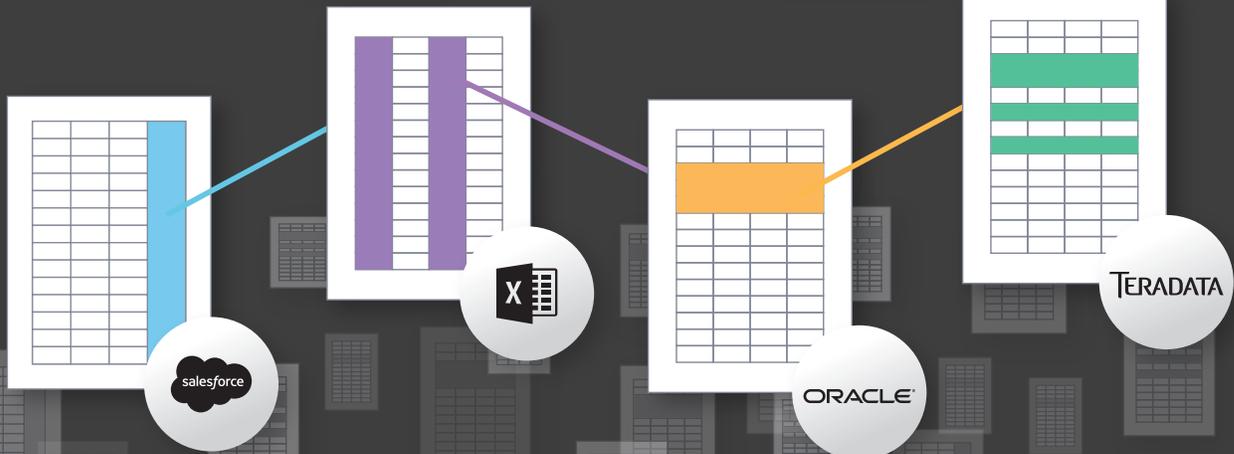
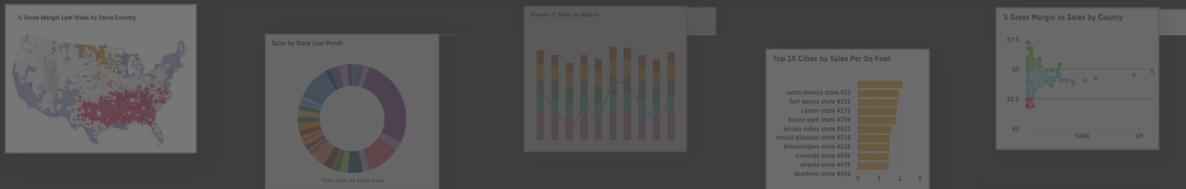


WHITE PAPER

# Relational Search: A New Paradigm for Data Analytics

 sales product last month california



# Table of Contents

The Third Wave of BI 3

---

The Power of Search-Driven Analytics 4

---

Why Relational Search is Different 5

---

A Next Generation Data Analytics Platform 7

- Relational Search Engine
  - BI & Visualization Server
  - Collaboration
  - In-Memory Relational Cache
  - Distributed Cluster Manager
  - Data Connectors & APIs
  - Enterprise Security & Governance
- 

Conclusion 21

---

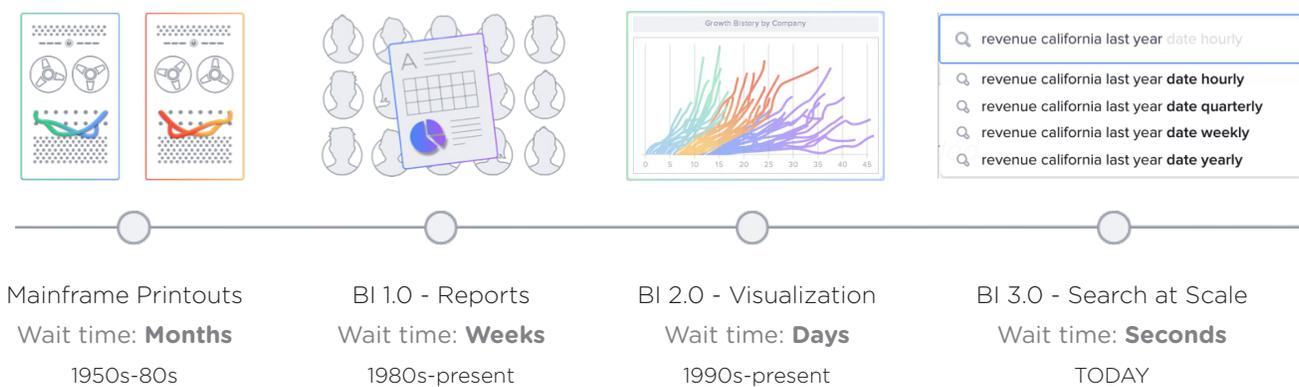
# The Third Wave of BI

Data is rapidly becoming every company’s most valuable asset. For those that quickly translate it into meaningful insights, there is a wealth of opportunities waiting. Yet for most organizations, fast and easy insights are the exception, not the rule. Despite billions spent on Business Intelligence (BI), adoption hovers at an abysmal 22 percent, according to BI Scorecard, leaving millions of business people without access to the information they need to make smarter decisions.

The reason? Legacy BI technology just wasn’t designed to handle the scale, speed and complexity of today’s data landscape. As a result, BI teams spend countless hours and costly resources just to deliver one report or dashboard, while business people wait in frustration. This paradigm has created a massive divide between the creators and consumers of analytics and neither side is happy.

To bridge this divide, a second wave of “self-service” desktop visualization tools came out in the ‘90s to try to meet the needs of data-hungry business users. Unfortunately these tools also created data sprawl and a governance nightmare. And despite their promise, they were still too complicated for the average business person to use without training or hours of BI support.

Now, a third wave of modern BI and analytics solutions is changing the paradigm for the BI industry and creating an entirely new world of possibilities. This third wave brings together the needs of both business users and BI teams, and provides true self-service analytics with enterprise-class governance and security. Business users get fast, easy, self-service analytics that actually works, so they can get the answers they need, when they need them. And BI teams get a governed, modern BI platform built to handle the scale, speed, and schema complexity of modern data architectures, so they don’t require massive amounts of upfront data modeling or performance optimizations - saving BI teams countless hours of work and resources.



## The Power of Search-Driven Analytics

In our personal lives, two billion humans use search every day to get instant access to massive amounts of information—no training or technical expertise needed. This ease-of-use at scale has inspired a new wave of analytics vendors that are using search to make it easy for business people to build their own reports and dashboards in seconds. Yet, as we all know from our consumer lives, not all search is created equal. It's important to understand the difference between various forms of search-based BI solutions.

The first kind of search is simply traditional web search, such as Google, Yahoo, or Amazon. This form of search, also called Object Search, enables users to search across data and return a list of indexed results. This is the most common form of search found in BI tools today. Object search makes finding existing dashboards and charts easier, but it does little to alleviate the burden on BI teams to constantly produce new reports.

A second form of search is natural language processing (NLP) which was made popular by Jeopardy-winning IBM Watson. NLP uses complex algorithms to interpret the intent of a user's search queries and return a probabilistic result. Unlike object search, with NLP, end users can actually begin to use search to build their own reports. However it too has some draw backs. For example, if a user searches "how many McDonalds are there in San Francisco?" an NLP engine will have to guess whether you mean "McDonalds" the restaurant chain, the street name, or the family name, and give you back an answer with an error rate as high as 20 to 30 percent. If you're running a business and looking for exactly how many units you sold last week or which products grew more than 10 percent last quarter, any answer that's not 100 percent accurate is unacceptable. And If you misspell or partially type a word by accident, your intent can be lost and with it the accuracy of your answer.

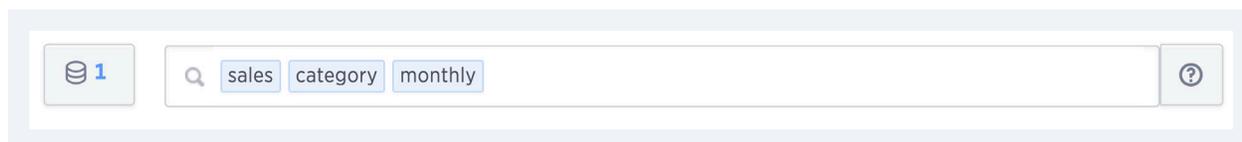
## Why Relational Search is Different

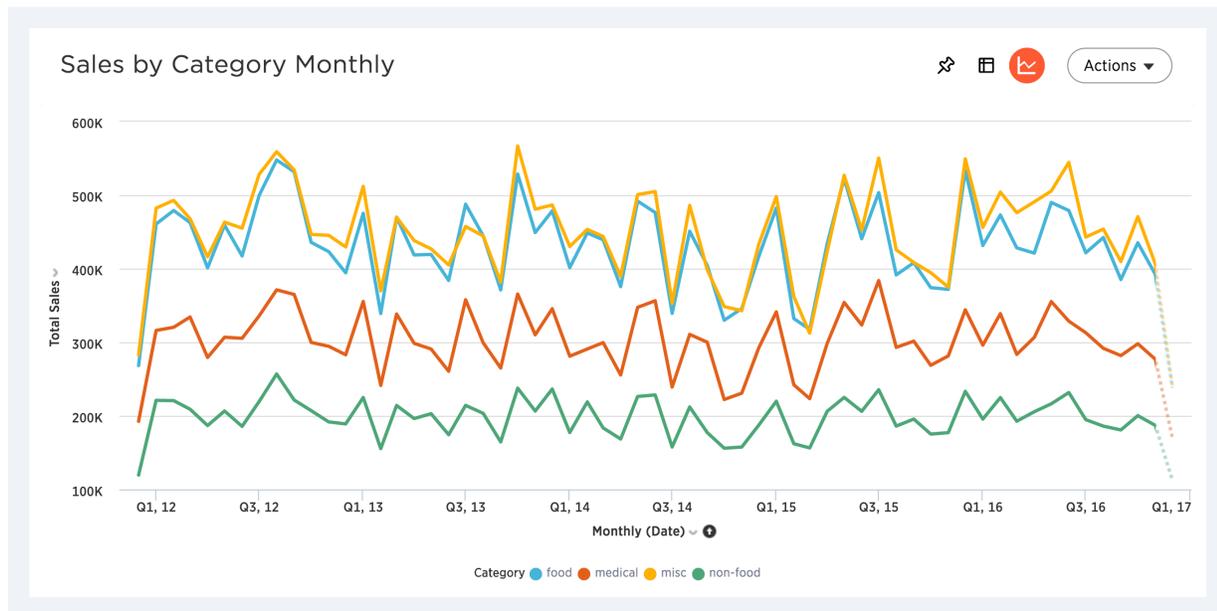
Relational search is a new breed of search-driven analytics solutions, unlike object search or natural language processing. A relational search engine allows anyone in an organization to use search to build 100 percent accurate charts and dashboards out of their relational data—on-the-fly, in a matter of seconds.

Relational data is any data that has a defined relationship or structure. The simplest form of relational data is a table in Microsoft Excel. Imagine a table showing all sales transactions for the past year, with columns showing dates, product category, and customer for each transaction.

TRANSACTION	PRICE	CATEGORY	CUSTOMER ID	DATE
134	\$500.26	Food	113114	6/27/16
101	\$140.40	Medical	113155	6/13/16
515	\$273.23	Misc	113679	4/1/16
478	\$111.72	Food	124890	3/22/16
319	\$314.98	Not-food	125689	5/3/16

With ThoughtSpot's Relational Search Engine anyone can connect to that raw dataset and type in "sales by category name monthly" and the engine will find all sales transactions for that product category, aggregate them by month in real time, and deliver a best-fit chart based on the characteristics of the data. In this case, ThoughtSpot produced a line chart.





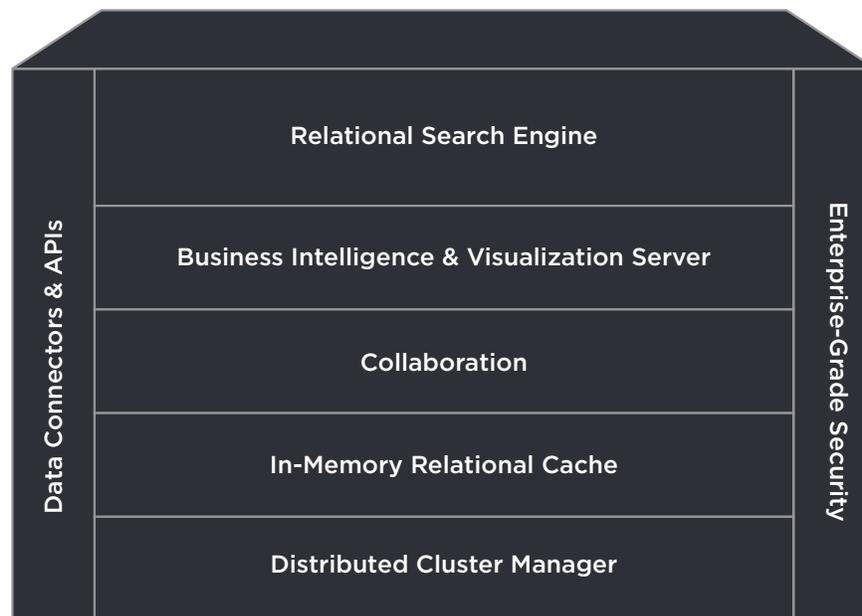
Visualization showing monthly sales by product category

Unlike object search, relational search engines don't require BI teams to pre-build dashboards or reports. The engine literally calculates the answers in real time as you type. And unlike NLP search engines, a relational search engine does not interpret your search query and then guess at an answer. Relational search deterministically takes each input and calculates a single, precise answer based on each letter you type. The result is an easy-to-use, penalty-free solution that lets non-technical business people truly do their own self-service analyses.

# A Next Generation Data Analytics Platform

A new breed of BI architecture built from the ground-up

Providing a simple search experience that can handle the data scale and schema complexities of the enterprise is not an easy problem to solve. ThoughtSpot's data analytics platform combines the precision of a relational search engine with the intelligence of a robust BI and visualization engine and the scale of an in-memory relational cache. The result is a radically different analytics experience that is lightning fast and scales to billions of rows of data from multiple sources—all with centrally managed security and governance. To build this highly scalable number-crunching machine, ThoughtSpot needed to create an entirely new breed of integrated architecture from the ground up. Open source technologies strung together are limited by the specifications they were built for and sacrifice performance at scale. With ThoughtSpot's architecture, each layer was custom built for search at scale, creating a high performance engine fine-tuned for search-driven analytics.



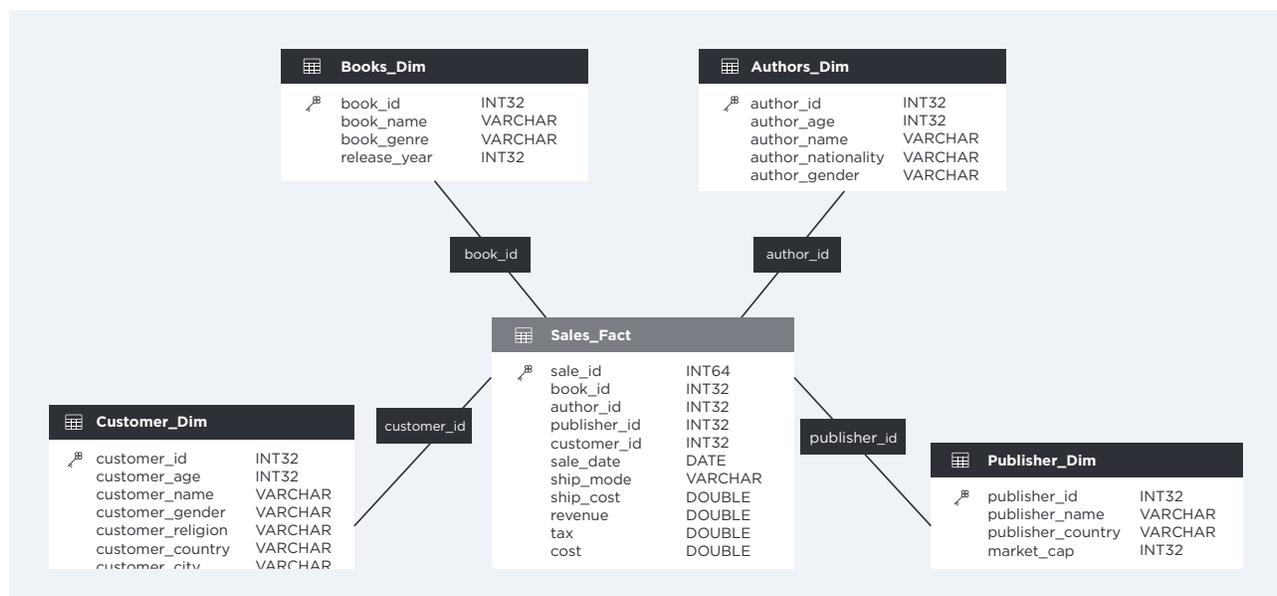
ThoughtSpot Product Architecture

# The Relational Search Engine

Anyone can build reports and dashboards in seconds

A third-wave BI platform must provide true self service to meet the needs of today's data-hungry business users. ThoughtSpot's Relational Search Engine lets anyone use search to build their own charts and dashboards in seconds, with no technical degrees or SQL coding required.

There are several key features of the Relational Search Engine that make it both powerful and easy to use. First, when you connect your data to ThoughtSpot, the Relational Search Engine caches both your data and your data schema in-memory. Once cached, the engine automatically creates a search index, or map, of both your data and schema. This means that ThoughtSpot indexes the column names, the tables each column belongs to, the relationships between those tables, and all of the underlying data values.

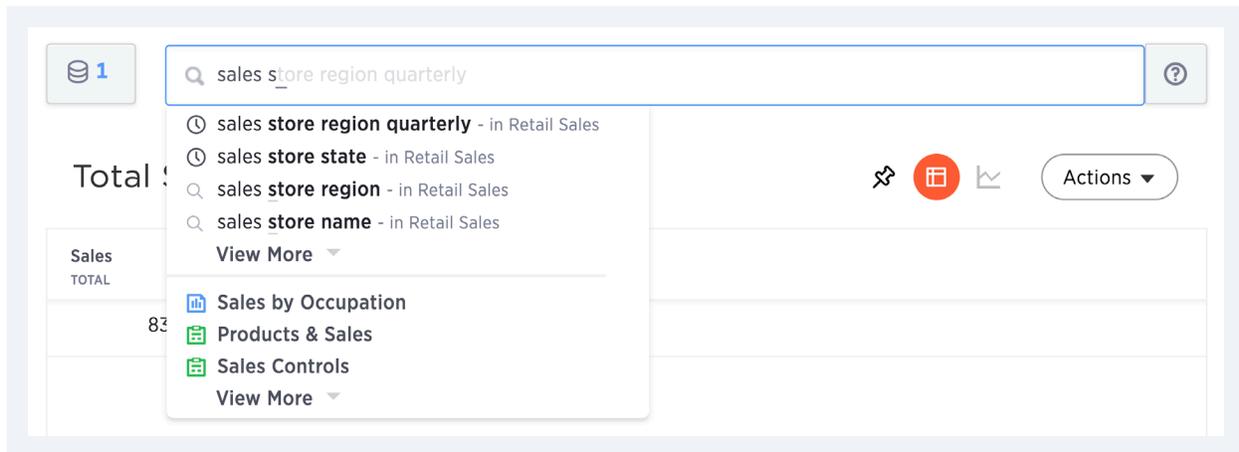


ThoughtSpot in-app schema diagram

This index gives ThoughtSpot a deep understanding of your data, eliminating the modeling work required for legacy BI tools, and allowing the product to return queries at lightning speed, just like Google or Amazon.

Next, similar to consumer search engines, ThoughtSpot's Relational Search Engine delivers a guided search experience to help end users get to the right answer, faster. As soon as a user starts to type into the search bar, ThoughtSpot immediately starts to suggest the next best search term right below the search bar, using a machine-learning algorithm called DataRank.

Just as when you type “wea” into Google and their type-ahead technology suggests “weather in San Francisco” or wherever you’re located, DataRank might suggest “sales by store region last week” after you have only typed “sal.”

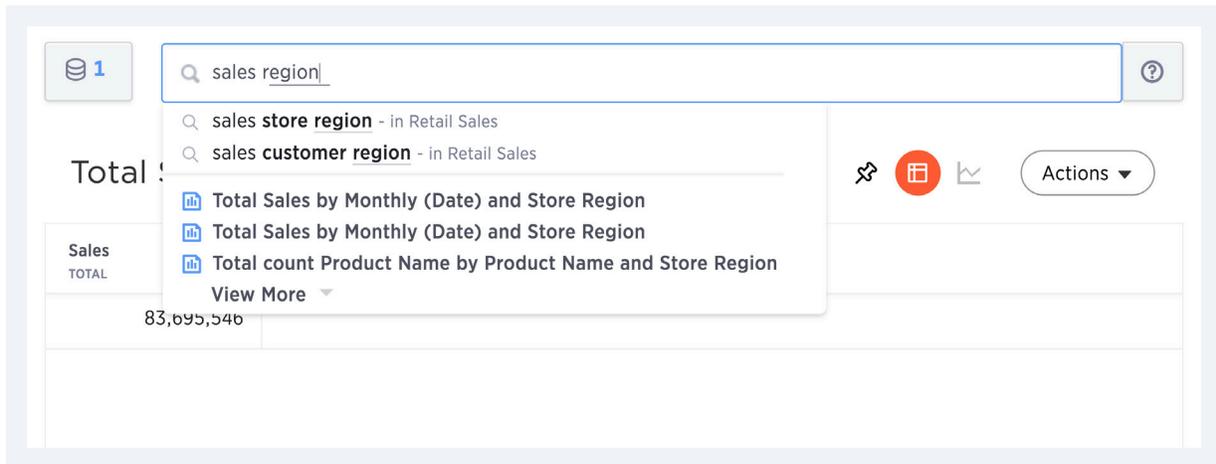


Search suggestions powered by DataRank

DataRank determines the order of suggestions based on data characteristics, user usage patterns, and role-based permissions in order to present the most relevant recommendations in real time as the user types. Suggestions include both raw data and previously saved charts and pinboards (dashboards) that use the search term. A user searching for “sales,” for example, will be able to see not only the next best search term and which data source it’s coming from, but also the most relevant charts and pinboards built by others in the company.

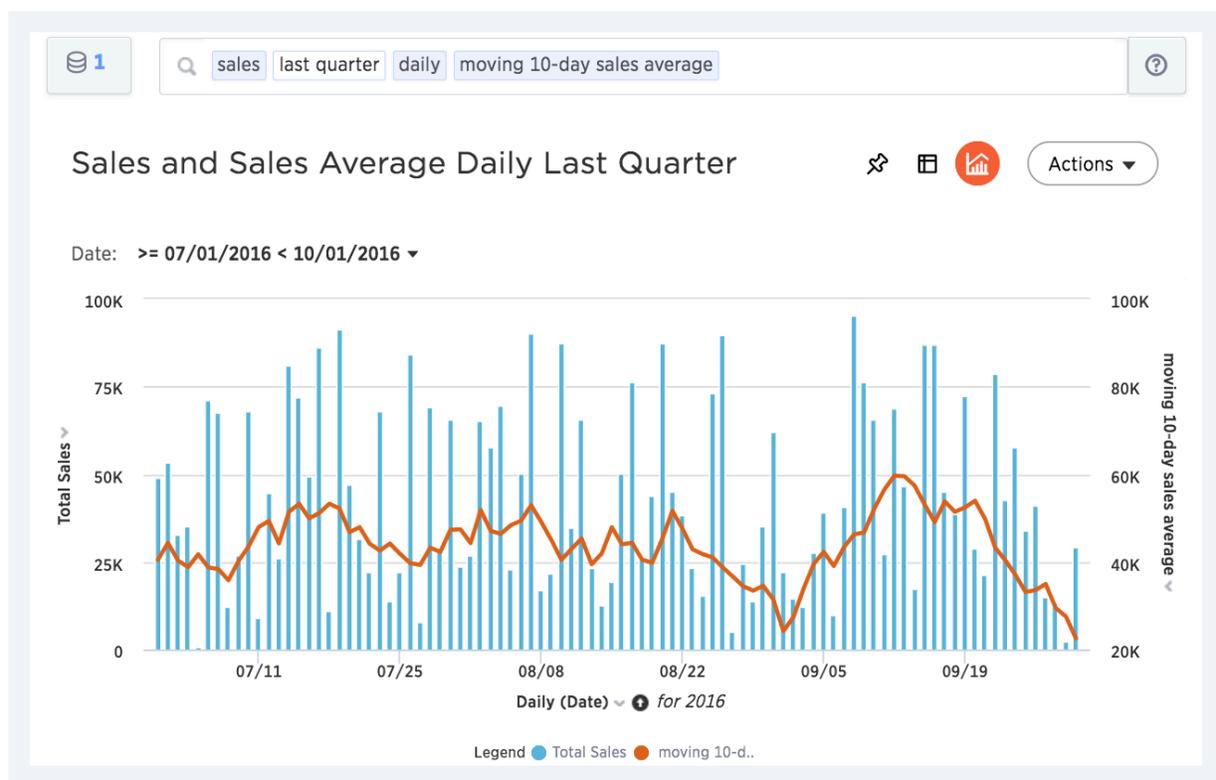
Unlike natural language processing, if a user searches for something that shows up in multiple places in the data, ThoughtSpot will never guess at what the user means. Instead the Relational Search Engine will prompt users to choose exactly what they want, highlighting the data lineage of the search term for clarification. In the example below, when the user types in “region” she is prompted to choose between “store region” and “customer region” since both terms exist in the data set.

Finally, ThoughtSpot’s Relational Search Engine comes with support for customizable synonyms, real-time keyword validation, and spell check right out of the box, as well as the ability to match substrings, synonyms, and formulas within the data. Customers can even create custom synonyms to match their own company vernacular, allowing users to type “sales,” “revenue,” or “bookings” for the same metric depending on how the user’s department might refer to it. If a word is misspelled, the search engine will suggest alternatives for the user to choose from.



Search suggestions across data sets with duplicate terms

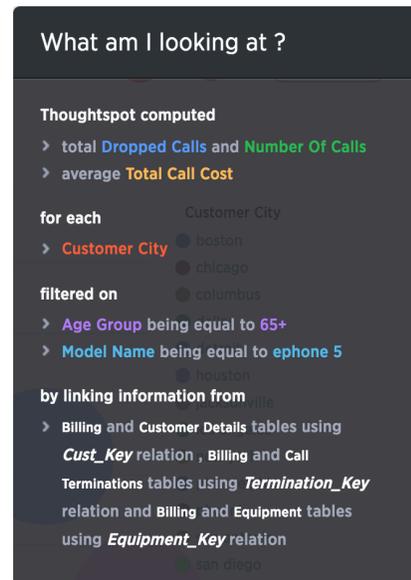
Users of ThoughtSpot also have the ability to incorporate advanced formulas, group sums, and nested searches, and users can even combine data from multiple source systems right within their search.



Visualization showing daily sales by store region today compared with a moving 10 day average

ThoughtSpot even has support for chasm trap schemas—also known as many-to-many joins—where two fact tables are joined through a shared dimension and can lead to over-counting in your analysis. For example, analyzing order details with customer details for a 360-degree view of customer activity is a traditional chasm trap scenario where many products are purchased by a given customer in a month and many customers purchased a given product in a month. ThoughtSpot's Relational Search Engine offers out-of-the-box support for these complex schema relationships so you can easily search across fact tables and get accurate answers every time.

Data lineage is tightly integrated throughout the entire search experience, helping build trust and improve understanding of the data for the end user. End users can easily confirm how ThoughtSpot is computing each answer, as well as the underlying data tables that are being accessed, by using the Search Inspector for complete transparency in every search.



ThoughtSpot Search Inspector reveals underlying query

# The BI & Visualization Server

Automatically generates queries and best fit charts

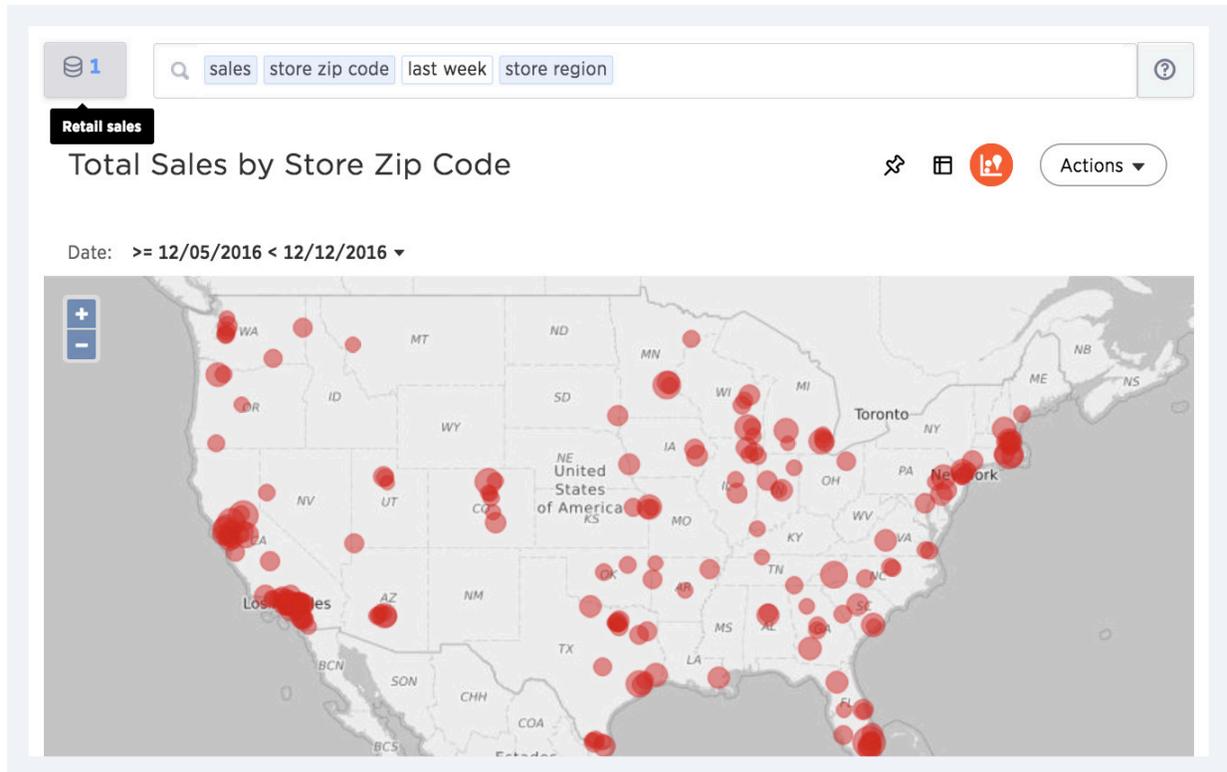
There have been tremendous advancements in visualizations over the last decade, so many in fact, that the visualization market is now commoditized. But when it comes to getting millions of non-technical people comfortable building their own analyses, the main obstacle that slows them down is the sea of chart choices, buttons, and drag-and-drop commands to learn. To identify meaningful insights from any search or query, users need to be able to quickly create powerful visualizations. In ThoughtSpot, this is done by our robust BI & Visualization Server. As a user types a search into ThoughtSpot, the BI & Visualization Server immediately goes to work, calculating an answer and producing an automatic best-fit chart from the data. These automatic charts take the guess work out of analysis, helping non technical users get started and get to their answers, faster.

Once a user types in a question, the BI Server does the work of intelligently translating the search into SQL to calculate the answer. The BI Server understands complex schemas and is able to deliver a robust search experience that can handle the complexities of multiple fact tables, formulas, sub-queries, and alternate join paths. A built-in query optimizer leverages DataRank's intelligence to predict queries as they are typed and run them in real-time to return answers in sub seconds. The most frequent queries are then cached for maximum query performance with zero tuning needed. As a result, queries that would take hours to run in legacy BI tools return answers in milliseconds.

Answers are then calculated in real time as the user types his search, right from the raw transaction-level data. As a result, end users can make ad-hoc changes to the grain or aggregation of a search in the moment, with no need for summary tables or aggregate views.

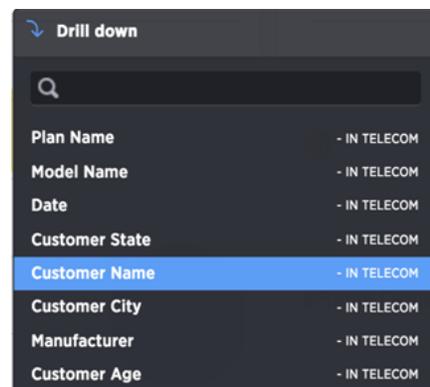
Once an answer is calculated, the Visualization Server analyzes the characteristics of the resulting dataset, intelligently determines the best-fit visualization for the analysis and presents back an interactive chart to the end user. The number of measures and attributes in a search will determine the chart type, and geocoded data such as state, county or zipcode will automatically return a map.

ThoughtSpot has over a dozen different chart types that users can choose from, including heat maps, tree maps, combined line and bar charts and pivot tables. And if a user has a preference for a chart that is different from ThoughtSpot's selection, he can easily make the switch in just two clicks. The charts can be further customized by changing the design across a number of parameters including colors, layout, zoom, labeling, sorting, and more.



Visualization showing total population

Charts and pinboards are also interactive, giving users the ability to change the underlying data right from within the Visualization Server. Users can add filters to the dataset or even drill down, up or across the data. Best of all, this functionality is enabled without needing a BI team or analyst to pre-define any drill paths or hierarchies. This saves BI teams countless hours of modeling, and days of tweaking reports and charts for the business.



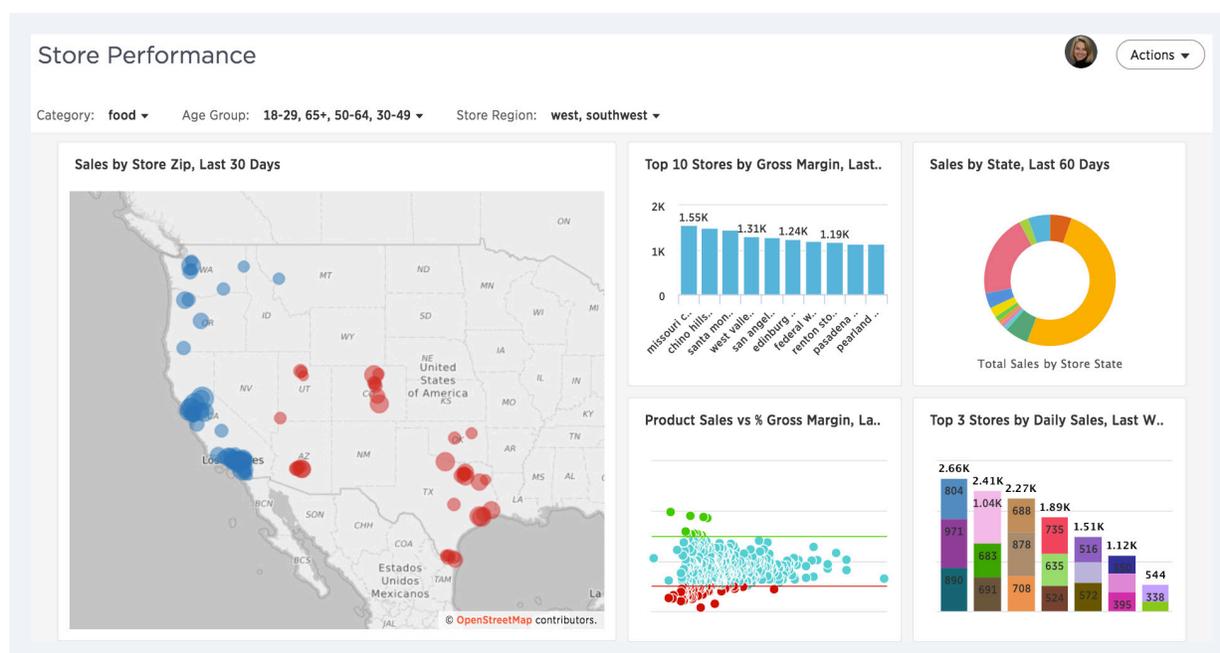
Drill down by customer name

# Collaboration

Real-time collaboration so you can stay in sync across the company

Analysis should never happen in a vacuum. Today it's more important than ever to leverage the collective wisdom within your company to accelerate consistent, data-driven decision making. With ThoughtSpot's enhanced collaboration features, this process couldn't be easier. Users can save their charts to a pinboard with a single click to build out their story, and then easily share and collaborate on that pinboard with any user or group in the company. And pinboards conveniently inherit the security settings of the viewer, displaying only the data the end user is allowed to see.

Pinboards are also dynamic, which means they are always displaying the most up-to-date information—no need to refresh a report or wait for it to load. And they include a full screen, in-app presentation mode, so you never have to bother with copying and pasting charts into another presentation. You can simply present directly from within ThoughtSpot.



Store performance pinboard

Now you can always know when something important changes in your data with KPI alerts and scheduled notifications that can easily be configured and customized for each user, such as a notification every time a deal closes. Get insights on the go, on your mobile device or email, without being tied to a dashboard or report so you can take action right when you need to.

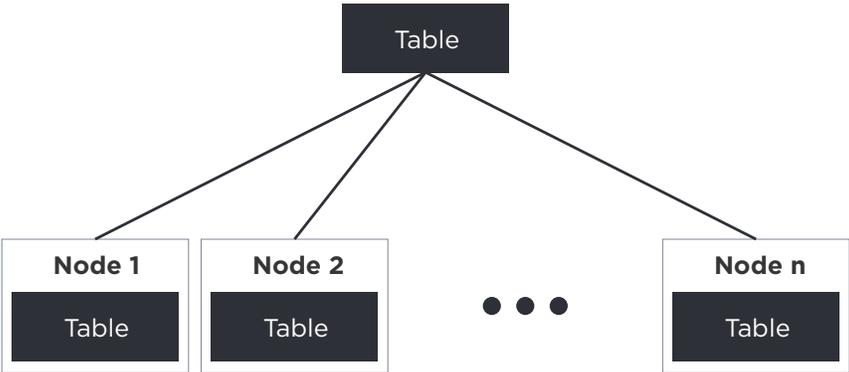
# In-Memory Relational Cache

Crunching billions of rows of data in milliseconds

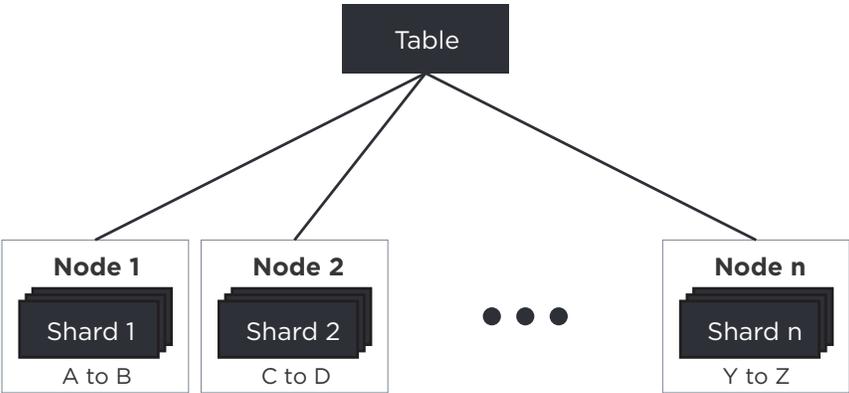
Today's organizations house large amounts of complex data from a number of different sources including the cloud, data warehouses or marts, semi-structured data from Hadoop, and of course flat files or spreadsheets. Modern BI platforms must be able to quickly scale to analyze large datasets with intricate schemas without sacrificing performance. Unlike desktop or cloud-only BI solutions, ThoughtSpot is capable of calculating results across billions of rows and terabytes of data in seconds. In fact, a recent performance test returned a two-billion row search query in 200 milliseconds.

Built from the ground up and optimized for relational search, ThoughtSpot's in-memory data cache can store billions of rows of data from any data source. Automatic data sharding and replication allows for the intelligent distribution of data, based on the schema and table sizes to minimize join time while maintaining high availability.

SCENARIO 1:  
Replication of tables  
across all nodes.

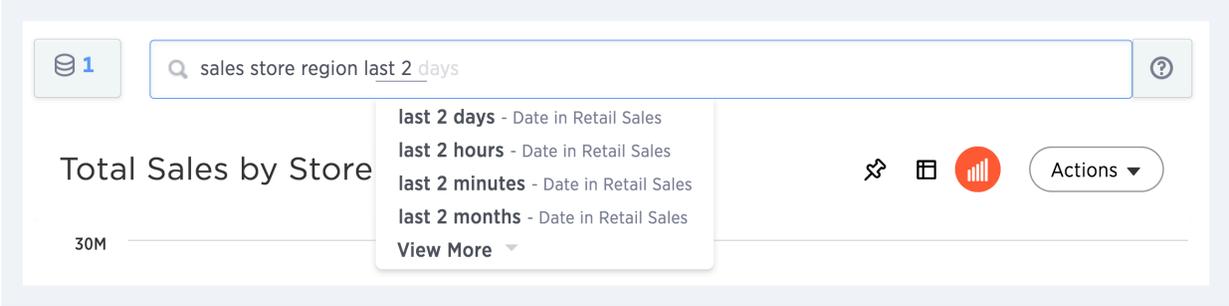


SCENARIO 2:  
Intelligent division of  
tables across nodes.



Data sharding and replication

The in-memory relational cache takes queries produced by the BI Server and calculates results against the data cache. Parallel computing allows the calculation engine to execute queries and perform joins and aggregations simultaneously, further reducing the time needed to produce an answer. Aggregations can be performed instantly without impacting performance. And data can be loaded into ThoughtSpot at the most granular level —no summary tables, aggregates or cubes needed. As a result, end users have the flexibility to move between time dimensions as they search from a summary view all the way down to the most granular level of data.



Slicing of data to any time dimension

# Distributed Cluster Manager

Monitor, manage, and scale out with ease

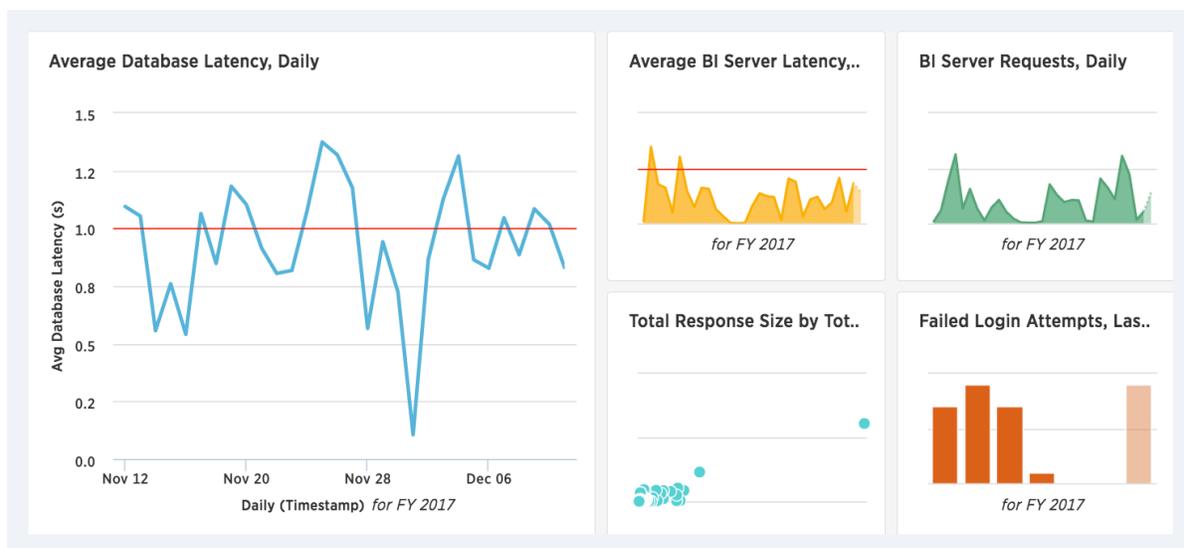
ThoughtSpot's Distributed Cluster Manager lets BI teams rest easy knowing that ThoughtSpot will always be ready when needed. ThoughtSpot sits on top of a multi-node Hadoop cluster that can be scaled out infinitely to support hundreds of terabytes of data. The Distributed Cluster Manager provides optimal distribution of workload for fault tolerance, redundancy, and fail-over options.



Distributed cluster manager workload

ThoughtSpot provides data replication and automatic node fail-over by persisting data on disk so that if a node goes offline, data can be pulled into memory on surviving nodes with zero downtime. Additionally, disaster recovery features automatically create backups and snapshots of your data to restore your cluster.

ThoughtSpot also comes with a built-in proactive monitoring and notification system for easy administration so you don't need to have a team of experts on-call 24 hours a day. The system continuously monitors the health of the cluster and intelligently repairs itself when needed, reducing administration overhead.

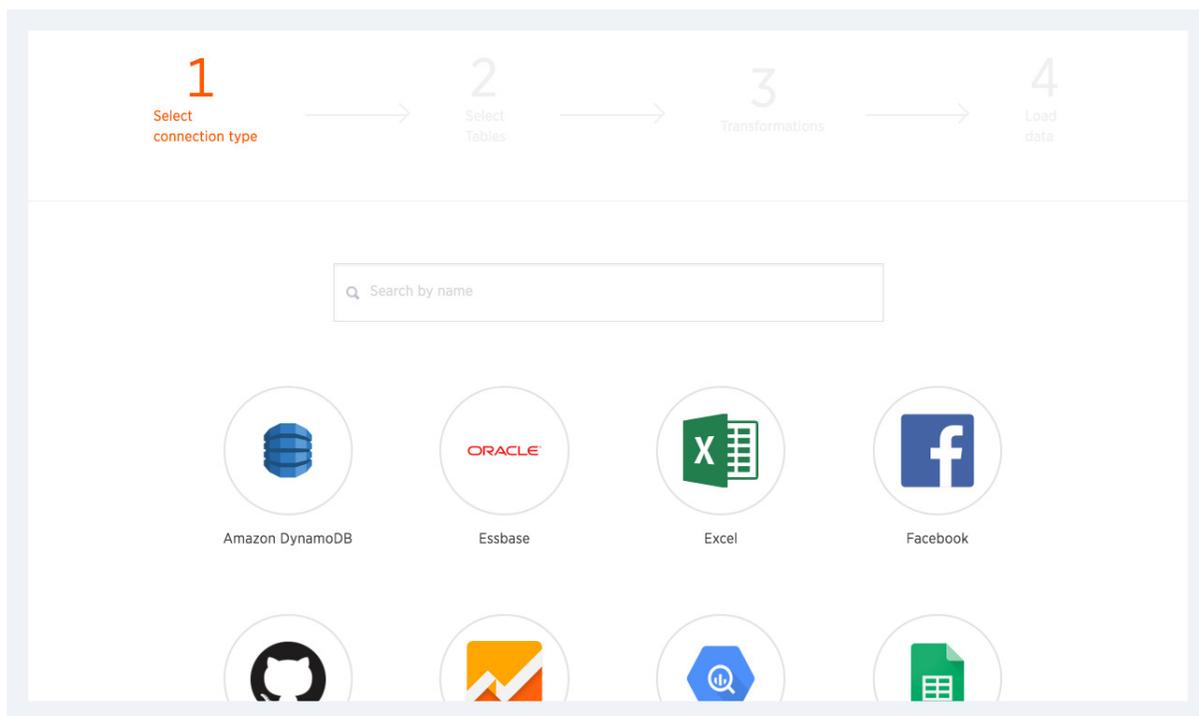


Usage monitoring pinboard

# Data Connectors & APIs

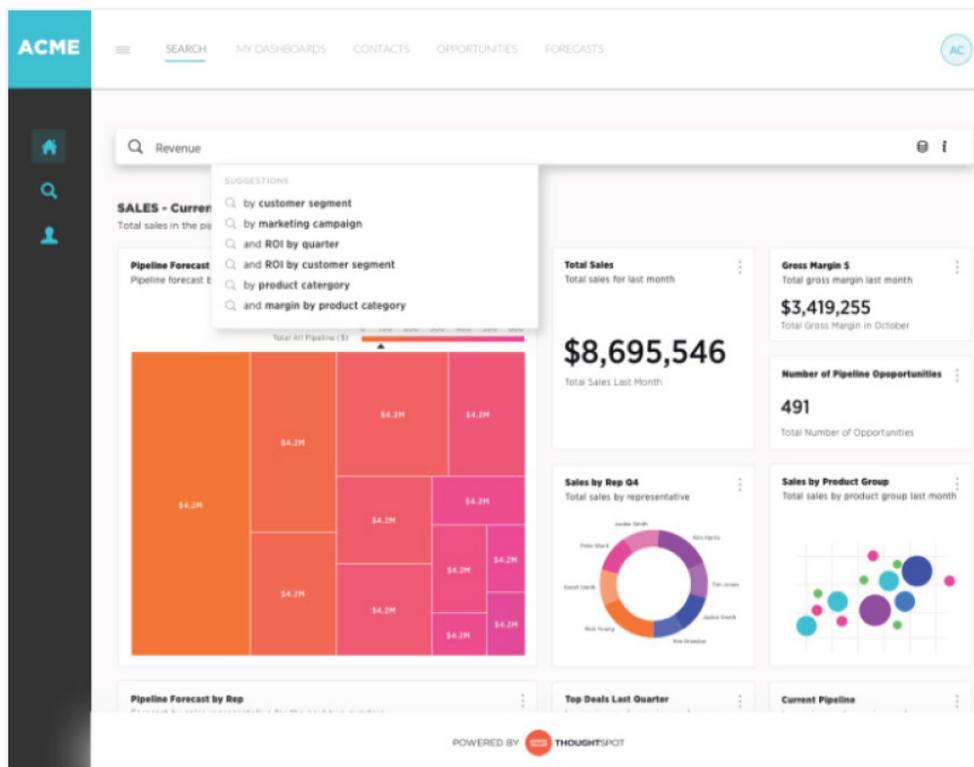
Securely connect to any data source in minutes

With the explosion of data has come a tidal wave of new data sources that BI teams and data architects need to manage and integrate into their analytics workflows. Connecting data to ThoughtSpot is simple because it can connect to any data source, from on-premise databases and data warehouses, to cloud-based data sources, flat files, social media data, and even semi-structured data from sources such as Hadoop. Data is ingested at speeds up to 1TB an hour so users can quickly start searching for answers from their data. To further accelerate the data pipeline ThoughtSpot Data Connect, with its hundreds of out-of-the-box data connectors offers robust self-service data preparation, transformation, and data-blending capabilities. Organizations who wish to leverage their existing investments in ETL solutions can also integrate their current ETL pipelines with ThoughtSpot via industry-standard ODBC and JDBC interfaces. All datasets are managed in a centralized data catalog in ThoughtSpot, and refreshed at scheduled intervals, promoting trusted and governed usage across the entire organization.



ThoughtSpot Data Connect

For organizations looking to share data externally with customers, partners, and distributors, ThoughtSpot offers the same lightning-fast and scalable functionality through our embedded analytics platform and REST API. To do this we offer two different options. The first is complete iFrame support for embedding individual charts and pinboards via the Object API. The second, the Data API, allows direct access to data from a ThoughtSpot answer as a JSON file that can be embedded directly into a custom website app or portal. As a result, companies can easily share ThoughtSpot-powered custom analytics with their entire ecosystem.



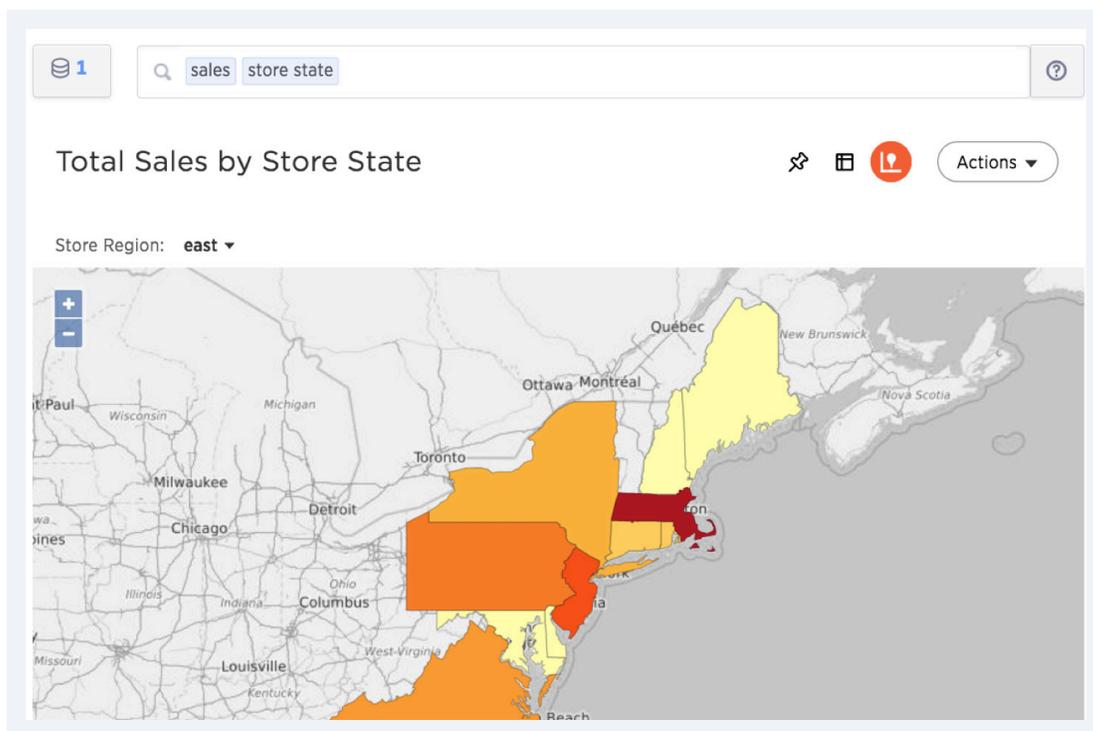
Embedded analytics customer portal

# Enterprise Security & Governance

Centrally managed access for thousands of users

The challenge with most self-service solutions is that they were not built with the modern enterprise in mind. These desktop tools create pockets of silo-ed analytics that grow with little regard for data security concerns. Meanwhile maintaining consistency across all these different reports and dashboards is nearly impossible—creating a governance headache and more work for the BI team.

With ThoughtSpot, security and governance were built into the architecture from the start. ThoughtSpot supports secure data discovery for thousands of users out-of-the-box with row, column, and object-level security. Regional sales managers can each access the same “Sales Pinboard,” and row-level security will determine what data shows up when the charts are presented.



Visualization of sales by store state for east regional sales manager

Customers also benefit from one shared data model and common metrics across the business, complete with usage monitoring and audit trail.

## Conclusion

ThoughtSpot's Relational Search Engine is changing the paradigm for business people and data professionals, creating a win-win. Non-technical business people get an easy-to-use solution to build their own reports and dashboards in minutes, instead of waiting days or weeks for answers from their BI team. They can be confident that no matter when data was analyzed or reports were built, everyone is always working off one version of the truth.

For data professionals, ThoughtSpot provides true self service, which means they can finally say goodbye to their massive BI backlogs of report requests and focus on the strategic data initiatives they were trained for. The speed and performance of the platform across billions of rows of data means BI teams no longer need to spend hours tuning and optimizing for performance with cubes, aggregate views, summary tables, or marts. And finally, data professionals also have centrally managed control and visibility, with granular security access to a single shared data model across groups and users.

For organizations that are ready to make the leap, the third wave of BI finally offers a way to deliver true self service for the business, within a governed and secure environment.

## About ThoughtSpot

ThoughtSpot is a next generation analytics platform powered by the World's first Relational Search engine. ThoughtSpot's search-driven analytics lets business people build reports and dashboards in seconds. ThoughtSpot's new breed of BI architecture and in-memory calculation engine was built from the ground up to make it easy to analyze billions of rows of data across multiple data sources, while delivering sub-second performance and enterprise-wide governance. ThoughtSpot connects with any on-premise, cloud, big data, or desktop data source and deploys 85% faster than legacy technologies. With ThoughtSpot, BI & Analytics teams have cut their reporting backlogs by over 90% and enabled thousands of daily AI-powered decisions throughout their organizations. ThoughtSpot has built the world's most advanced, yet easy-to-use number-crunching machine with a singular mission - to deliver access to data and insights at "human scale."

For more information on ThoughtSpot, please visit [thoughtspot.com](http://thoughtspot.com)