ThoughtSpot – For Near Instant Analytics Gratification
       By Kirsten Kissmeyer, July 11 2016

ThoughtSpot ups the ante when it comes to rapidly and effortlessly delivering insightful and completely ad-hoc data analytics and visuals to your business, even for large many TB data sets.

ThoughtSpot has trail blazed a new area of BI called Search BI. This type of BI differs from the current genre of more established BI tools such as Tableau in that it embeds and applies knowledge about how data of different categories is generally analyzed and most effectively visualized. This knowledge is then mapped onto your business's specific domain data.   The alignment and cataloguing of the business domain data and Meta data is then used to provide an optimized, intelligent and guided search capability through it. A business user simply begins typing what they are looking for into the search box and then ThoughtSpot offers completions of the search as the user types. The suggested completions are offered in the order that similar information is searched (built in, out of the box, ThoughtSpot knowledge) and that this particular business data has been searched previously (ThoughtSpot learned knowledge about your business's implementation and use.) Some sample potential searches that a business user might issue include: New customers by region and month, outstanding payments due by customer profile, trended average for sales by quarter, non-renewals by month in zip code 976677.

Once the business user has finalized and executed a search, ThoughtSpot automatically generates both the underlying data plan for retrieving the result data and the visuals to render the information in a way that is deemed most effective. These visuals are called PopCharts and include line charts, bubble charts, geomaps, tree maps, pie charts, and many more. The type of chart and other properties such as colors may be customized by the user. Your favorite visuals can be pinned to themed pin boards you design as well as shared with colleagues. The data retrieval plan for the search results can also be reviewed and modified to trouble shoot any issues that you may see in the results.

With its embedded data analysis knowledge and automation in rendering visuals, ThoughtSpot can deliver analytics far faster than traditional BI tools that require a lot more manual development at both the data and presentation layers.

This all sounds too good to be true– doesn't it? Like magic. So, I am sure you have many questions on how it is actually accomplished. Like -

What is involved in making your business data available to ThoughtSpot to search?

Data from your business data stores are loaded into ThoughtSpot distributed in-memory and analytically optimized data cache using ThoughtSpot Data Connect technologies that allows you to connect to, and select tables from any SQL Server, Oracle, or MySQL or any JDBC enabled database, Hadoop, any supported cloud-based application stores such as SalesForce, Marketo, Workday, or JIRA or any Excel or CSV flat file. You can then pull data from these connected tables, applying filters and expressions for new columns as needed to limit and enrich the data. The filters and expressions provide a mini ETL type capability. New data can be upserted or truncate and loaded. The data loads specified can then be named and run on demand or scheduled to run at certain days, dates, times, and/or intervals.

How is your business data interpreted and optimized for immediate and meaningful search by ThoughSpot?

 In the process of specifying what business data is made available for search, ThoughtSpot identifies the dimensions and facts of the data for alignment and incorporates this information into its search base. The process is a mixed-initiative one and the ThoughtSpot engineer can override and change the analytical search configuration. This analytical search configuration includes providing synonyms for both meta data and data values.   The information is then optimized for search using similar approaches and algorithms applied by web search engines.

Scalability - How large a Data Set can ThoughtSpot handle, and how many simultaneous users?

ThoughtSpot is able to handle infinitely large data sets and user communities as it is designed to scale horizontally.

A ThoughtSpot implementation can be deployed as on-premise hardware in your data centers, on servers provided by ThoughtSpot, or on-premise VMWare, on VMware you provide in your data centers. You

may also utilize ThoughtSpot as a completely cloud-based solution, on AWS or Microsoft Azure, where you pay for data storage and user capacity based on your subscription and/or metered usage.

For an on-premise Thoughtspot provided server configuration, ThoughtSpot data and user capacity may be increased by simply adding more ThoughtSpot "bricks" to your configuration.

One brick fits in a 2U form factor and is installed into a standard size rack. Each brick contains 4 nodes where each node has 512GB RAM, 20 cores (40 logical cores), (5) 1 TB disks, (1) 128 GB SSD (for the OS), (1) 10 GIGE port, and (1) 100 MB management port.

A minimum initial configuration of 1 brick with 4 nodes is recommended to provide full High Availability capabilities. For optimizations based on large data set size, ThoughtSpot can also compress and shard the dimensions and facts across nodes. Maximum data capacity varies depending on the data model, and on how the data is sharded and how much data needs to be replicated and at what compression ratio. Capacity planning to utilize 700GB of each 1 TB disk for actual data storage is a good rule of thumb.   A single 1 TB unit can handle average width tables that have upwards of 1 billion rows of data.

As far as user capacity goes, each brick is estimated to support 3000 users @ 10% concurrency, which = 300 concurrent users (logged in), which typically = 30 active users, which typically = 3 users sending queries at any point in time.

ThoughtSpot software may also be installed into a virtual OS. However, the specifics for the capacity of this virtual deployment is dependent on the configuration of the virtual OS and the hardware it resides on.

How much time and effort will it take to implement ThoughtSpot for your business?

Once permissions and specifications to access necessary business data stores have been provided, the typical ThoughtSpot implementation, to the point where searches may be issued with meaningful results returned, takes only a few hours to a couple of days. This initial implementation can then typically be productionized in just 2-4 weeks. The amount of effort and time involved in setting up a ThoughtSpot

application depends largely on the number of data sources and the dimensional and alignment complexity across them.

Where does ThoughtSpot fit into Your BI Enterprise Landscape?

ThoughSpot can sit on top of your organization's existing Data Warehouse, Master Data Management, and ETL stack. Although ThoughtSpot does not require the information that it searches to be pre-aggregated, availability of a pre-aggregated data store does simplify implementation for some searches. However, it is necessary to make available transaction level data as well to allow more kinds of ad-hoc searches.

ThoughtSpot works alongside more traditional BI tools such as Tableau or Business Objects or SSRS.

ThoughtSpot is being used at many organizations now to provide a more dynamic and ad-hoc analytics capability directly to Business users that is not as dependent on the pre-aggregation of data for optimized performance. Business users can independently explore different ways of looking at their business data that may provide unexpected insights. They can, to a far greater extent, innovate on the way they analyze the information they have without the need to involve IT up front. More traditional BI tools like Tableau that are already in place are then often used to implement reports on useful new analytics that have been discovered via ThoughtSpot.