White Paper

The Ultimate Guide to Real-Time Data Visualization

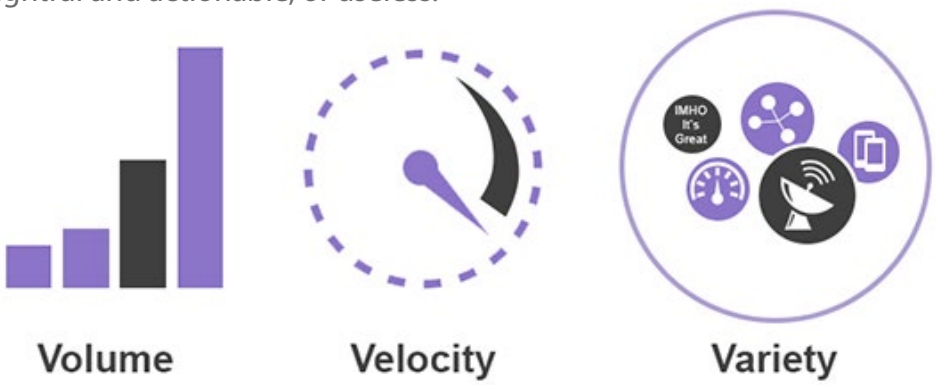
Abstract

With the explosive growth of data in recent years, and the opportunities afforded by big data, the need for real-time business intelligence is growing by the day. IBM estimates that 90% of all the world's data has been created in the past two years alone[1]. However, while this firehose of data hits us at ever increasing speeds, IDC observes that less than 1% of all available data is analyzed[2]. Businesses face the problem of being rich on data, but poor on information.

In this white paper, we discuss how organizations can gear up to handle the big data revolution, and more specifically, how to approach real-time business intelligence in a way that empowers their decisions. We'll take a look at the various functional pieces of a real-time business intelligence system, with an emphasis on the data visualization aspect.

At the Speed of Data

Gartner draws a contrast between data as we knew it and big data based on three characteristics - volume, velocity, and variety[3]. When discussing real-time analytics, while none of the three can be left out, velocity is of prime importance. This is because the speed of capturing, storing, processing, and visualizing data in real-time business intelligence systems makes all the difference between the information being insightful and actionable, or useless.



From Relational Databases to Big Data

Real-time analytics wasn't possible a decade back. It was enough for a store owner to know how many items of a product were sold each day, and which were the fastest selling. This information aggregated over a quarter would assist with making strategic decisions, for example, about which product lines were increasingly in demand and showed the most growth potential. Businesses were happy to store data in relational databases, which accepted data only in a pre-defined structure. The stored data was queried weekly or monthly using SQL queries that were processed in sequential order, making it difficult to run complex, large-scale queries. Also, these queries would return results successfully only if all databases involved were functional, and the data was stored in a consistent manner. This infrastructure couldn't scale up when there was terabytes of data being generated in a single organization on a daily basis.

Fast forward to today, and the same store owner is not satisfied with getting just the history of last quarter's sales. On an operational level, they want a real-time BI system that knows what a particular customer purchased today, what they're most likely to buy next, and which add-ons to suggest right now, before the customer's attention fades. This sort of real-time, predictive analytics wasn't possible back in the days of traditional data warehousing. But today, with vastly reduced storage prices, increased RAM capabilities, powerful querying tools, and capable visualization components, real-time analytics is becoming a reality in more and more business scenarios. It is now not only possible, but imperative that enterprises master the art of real-time business intelligence for both strategic and tactical purposes.

Real World Applications of Real-time Analytics

Real-time analytics is being used ever increasingly today in across all industries, in both B2B and B2C applications.

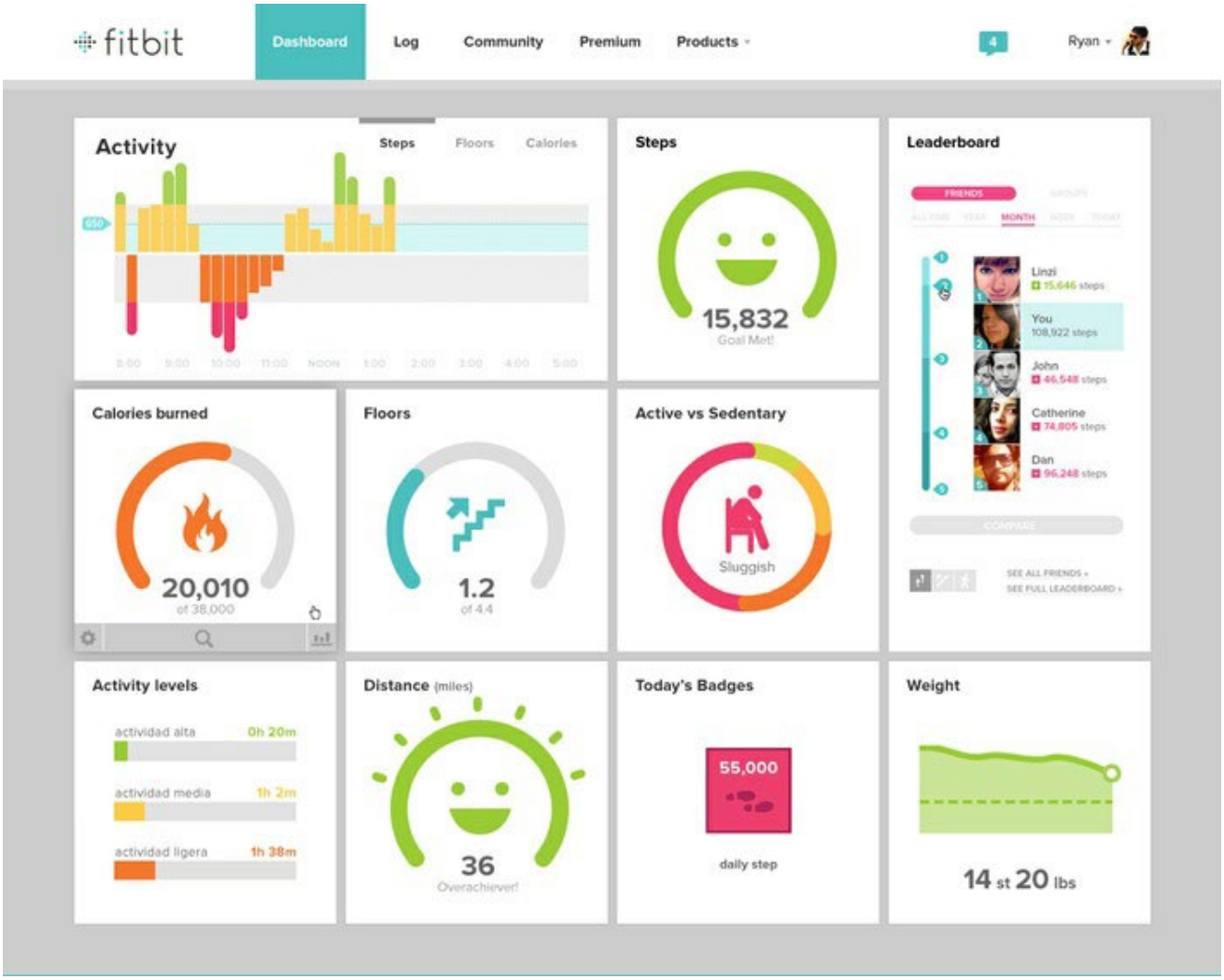
- In marketing automation products like Marketo, real-time analytics can process up-to-the-minute information about an enterprise's prospects. This enables Marketing and Sales teams to collaborate internally, communicate more effectively with prospects, and close deals faster.

- Social media monitoring applications like Radian6 and Klout crawl various social media sites to identify mentions of a particular brand. The data is in the form of text content, and meta-data like status updates, tweets, comments, likes, and shares. This data is augmented with a sentiment score based on how positive or negative the comment is, and an influencer score to prioritize users with more visibility higher. This informs marketers in a timely way about the pulse of conversations around their brand.

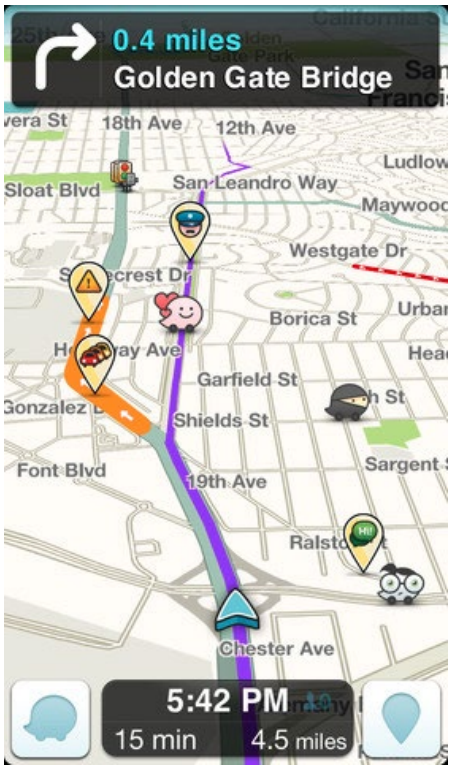
- In the healthcare industry, clinical decision support systems (CDSS) like VisualDx assist physicians by allowing them to enter a patient's data into the CDSS, which is then compiled with information from multiple external knowledge bases, to give the physician an accurate diagnosis of the sickness, and case-specific suggestions.

- In the financial services sector, startups like Flyberry Capital look for patterns in news updates, weather data, and even social media updates. They then use these patterns to make data-driven decisions on their investment strategies for large portfolios.

- With the "Internet of Things", and passively shared data becoming more pervasive, gadgets like Fitbit are becoming mainstream. Fitbit measures personal fitness data, compares it to the ideal numbers and awards badges accordingly.



- Location-based applications seek to make GPS navigation smarter and more timely. A popular GPS-based mobile app, Waze, captures data about travel time, and routes from commuters, processes this data, and provides on-the-go suggestions for users to beat the traffic.



What's common across all these use cases is that the data analyzed is extremely time-bound, and in some cases, mission critical. Even businesses that don't entirely depend on real-time analytics leverage it to enhance some of their vital business functions, or risk of losing their competitive edge.

With this overview of real-time business intelligence in enterprises today, we are now ready to dive deeper into the finer details.

Nuances of Latency

All real-time business intelligence systems have some latency. The duration of the latency depends on various factors such as the volume and variety of data being collected, the speed at which the connectors store and move data within the DBMS, the ability of the querying tools to run complex queries, and finally, on the visualization tool's ability to transform the data into shapes. It is important to understand the nuances of latency, as the primary goal of a real-time business intelligence system is to minimize the time between an event happening and an action being taken in response.

Analyst Richard Hackathorn describes three types of latency[4]:

1. Data latency: The time taken to collect and store the data
2. Analysis latency: The time taken to analyze the data and turn it into actionable information
3. Action latency: The time taken to react to the information and take action

For the purposes of this white paper, we simply touch upon aspects of data latency, and action latency, but elaborate on analysis latency as it is most relevant when talking about real-time data visualization.

Demystifying Real-time Analytics

There's a wide range of applications, and use cases for real-time analytics in every industry, and at every level of business. Across this spectrum, the time interval between capture of information and action taken on it, can vary from as short as milliseconds to as long as minutes or even hours. This warrants a clear definition of what we understand by the phrase "real-time." There are two categories of real-time processing:

1. Real-Time (In-stream) Processing

This is the more authentic of the two categories. In this method, the time interval between the occurrence of an event and response to it is in the millisecond range. The financial services segment is a prime example of this type of analysis. Stock trading applications use complex event processing tools to analyze continuously streaming stock prices, and trigger automated actions or alert people to patterns and trends.

2. Near Real-Time (Ad-hoc) Processing

This category works on "as fast as disk" processing speeds, or "near real time." the time delay is introduced by latency in data processing or in network transmission. It involves fast ad-hoc and interactive queries to capture data, query it, and present it in seconds to minutes. While it isn't clearly defined, the delay in near real-time processing can be as high as 15 to 20 minutes in some applications. An example of this would be the Chinese telecom company, CMG, partnering with Intel to allow its customers to view their mobile call logs, and billing details with a delay of a few minutes.

How Real-Time Analytics Works

Despite the extremely short time duration from end to end, there are three broad steps to how data is visualized in real-time.

1. Streaming data is captured

Live streaming data is captured using scrapers, collectors, agents, listeners, and is stored in a database. This is the stage where the volume, velocity, and variety of data is defined.

2. The data is stream processed

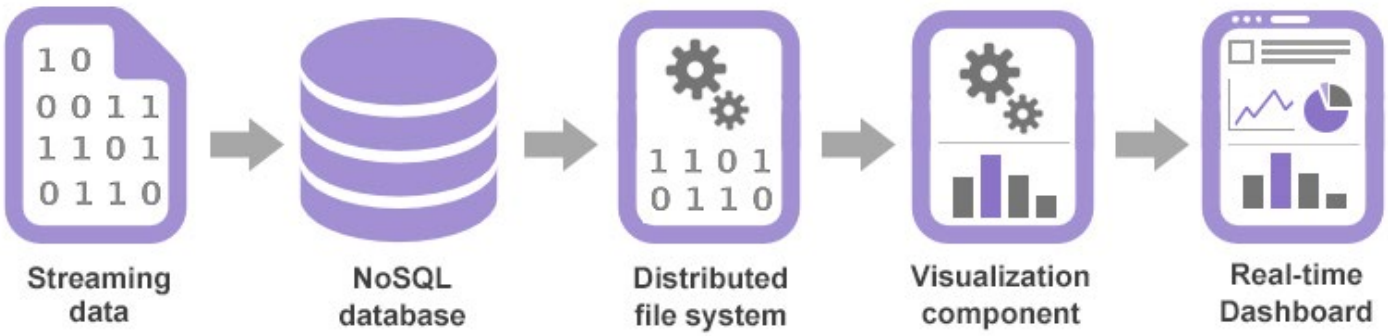
The streaming data is processed in many ways like splitting, merging, doing calculations, and connecting it with outside data sources. This is done by fault-tolerant, distributed database systems like Hadoop, HBase, Cassandra, and Storm. After processing, the data is ready to be read by the visualization component.

3. The processed data is read by the visualization component

The processed data is stored in a structured format, like JSON or XML, which is read by the visualization component. The frequency at which processed data is passed from the database to the JSON/XML file is termed as the update interval.

4. The visualization component updates the real-time dashboard

The visualization component then reads the data from the structured data file (JSON/XML), and draws a chart, gauge, or other visualization in the reporting interface. The frequency at which processed data is drawn on client-side is called the refresh interval.



All this happens in seconds, or even milliseconds, where immediately after the data is collected, it is processed and visualized in dashboards and reporting interfaces. The most commonly used client-side application today is the web browser, but other clients include on-premise applications, and embedded systems like routers, and patient monitoring devices.

With this understanding of the various components of a real-time business intelligence system, we can now dive deeper into how real-time data is visualized. While visualization is often the last leg in the entire process, it's often the starting point when defining business goals and objectives. Without the right visualization, the data, no matter how valuable in itself, will be of little or no effect. Considering the amount of effort that goes into capturing, storing, and querying the real-time data, the vital task of visualizing that data deserves equal, if not more attention to do justice to the entire project. A well thought-out dashboard can make all the difference between the real-time data being actionable or useless.

Choosing a visualization solution for real-time data isn't easy. There are tons of complexities across the entire real-time business intelligence system, which need to be factored in. Add to that the critical nature of the data being handled, and it can be a long, daunting task to zero in on a solution that's just right. To make this task easier, let's look at a list of requirements that are crucial for today's enterprises that need a visualization solution, or charting component for real-time data.

Choosing a Real-Time Charting Solution the Right Way

Here are the top 4 factors that can't be ignored when choosing a real-time charting solution:

1. Update & Refresh Intervals

Considering the sequence of updating a structured file (JSON/XML) first, and then refreshing the client, a real-time visualization component should be capable of extremely low update and refresh intervals. If the speed of the streaming data is faster than the visualization component can handle, the chart either loads weirdly, or doesn't load at all. To get around this, charting components may suggest to aggregate data in longer time periods, and add a delay to the refresh interval. While some make do with this, the ideal solution is to have a component that's capable of both update and refresh intervals as short as a second, and can handle extremely high volumes of streaming data.

2. Persistent Storing of Historical Data

While accessing immediate information is a given with real-time data visualization, another important requirement to consider is the viewing and analyzing of historical information. For example, when viewing a stock ticker, while a day trader would want to frequently check the changes in a particular stock, they may also want to go back a day or earlier to see the historic trend of the stock. In this case, the visualization component should not only plot real-time data, but also allow for loading of unlimited historical data points. Charting solutions address this using a history buffer, which is based on a particular number of historical data points, or a historical time range. This historical data facilitates a richer analysis, and more informed decisions.

3. Large Number of Data Sets

In social media monitoring apps that track hundreds of thousands of mentions of a brand across diverse sources, the volume of data captured is enormous. This takes tremendous processing power on the part of the visualization component to keep up with the pace of updates. It becomes imperative to be able to pass a large number of data points per update. This will ensure your data-heavy dashboard doesn't break under the load.

4. Timestamps for More Control

While it's expected to see high consumption of bandwidth when working with real-time charts, it could sometimes get overwhelming. For example, if data is streaming at 100 data points per second, it could consume too much bandwidth, clogging the entire network with 100 updates, and 100 refreshes per second. In this case, it's important to record all the streaming data for historical analysis, and still find a way to conserve bandwidth. In this situation, it's recommended to use a timestamp. This way, the visualization component can receive 100 updates per second, but refresh the chart once every second, for example, rather than every 10 milliseconds. The 100 data points are still plotted, but not in the same frequency that they're received. This feature offers more control over the performance of your entire system, and shouldn't be overlooked.

The Future of Real-time Data Visualization

While we continue to fully understand the applications of big data in our lives and organizations, the volume, velocity, and variety of data being generated is only on the rise. It's clear that companies that recognize the opportunity in real-time business intelligence, will come out on top.

Increasingly, CIOs don't want to jump through hoops to find actionable insights in their real-time dashboards. Instead, they want to understand significant correlations without having to think about the

methods used to uncover those correlations. In fact, moving from correlation to prediction is the order of the day. If startups like Flyberry Capital, and Recorded Future (another predictive modeling startup) make progress in this space, it has the potential to disrupt entire industries.

Finally, as mobile devices become even more pervasive, the amount of passively shared real-time data is going to change the way we live. Companies like Fitbit and Sensaris have already made considerable progress in this direction with wearable devices that track fitness, health, and even ecological data.

Despite all the advancement we've made in visualizing real-time data, we're still at the very beginning of what's possible. Makes you wonder - what next!

About FusionCharts

FusionCharts has been a pioneer in the industry of data visualization over the past decade. FusionCharts Suite XT brings you JavaScript charts for the grown-ups. It features charts gauges, and maps that are ideally suited for real-time dashboards in web and mobile applications.

Sources -

1. http://www-01.ibm.com/software/in/data/bigdata/
2. http://www.emc.com/about/news/press/2012/20121211-01.htm
3. http://www.gartner.com/technology/topics/big-data.jsp
4. http://www.bolder.com/pubs/DMR200401-Real-Time%20to%20Real-Value.pdf