



*“..In 2009, Zynga purchased the Vertica Analytics Platform and adopted it as their enterprise data warehouse platform. Today, Zynga runs one of the largest Vertica clusters in the world.”*

# The Impact of Social Graphing Analysis on the Bottom Line: How Zynga Performs Graph Analysis with the Vertica Analytics Platform

## Introduction

Today, social networking impacts every type of business, from restaurants and recruiting to consumer goods and retailing. Social networking platforms like Facebook and Twitter, and social gaming companies like Zynga have changed the way many organizations are marketing and positioning their products and services.

The value of these platforms lies not only in understanding the information that is generated by users, but also in understanding the information flow and linkages between users and their own unique social networks. With this understanding comes the ability to interact with the mavens – those individuals who shape opinion within these communities – thereby gaining their support and influence. While conceptually simple, the challenge lies in loading and analyzing significant volumes of data in near real time to determine these relationships and influencers in a manner that not only drives revenue and improves the bottom line but also optimizes a customer’s experience.

You may be surprised to find that Zynga is performing some of the most advanced analytics anywhere. A lot can be learned from the world’s largest social gaming company, as Zynga’s user community offers a microcosm of several real-world social graphing scenarios. This white paper describes how Zynga uses the Vertica Analytics Platform to improve its business and game features.

## About Zynga

Zynga is the world’s largest social games developer, with a stated mission to connect the world through games. Every day, millions of people interact with their friends and express their unique personalities through Zynga games, which range from harvesting crops to slicing apples to playing poker. From its founding in July of 2007, Zynga has built a user base of over 250 million monthly active users through a combination of vision and the ability to rapidly deploy new games and in-game features. Zynga very quickly discovered that while “seat of the pants” intuition about how to create a strong game experience was a great start, it was not a sustainable model for a large-scale operation. In 2009, Zynga purchased the Vertica Analytics Platform and adopted it as their enterprise data warehouse platform. Today, Zynga runs one of the largest Vertica clusters in the world.

## The Challenge

While the concept of “social gaming” is not new – games are generally about being social – today we use the term to describe games that run on social networking platforms such as Facebook and MySpace, as well as on mobile devices like smart phones and tablets. These games are characterized by the fact that they run in a very lightweight client (typically a web browser), are engaging and with a modest learning curve, typically have open-ended goals and features, and rely heavily on the interaction between players for the best in-game experience.

There are three key metrics that drive the economics of social gaming; churn rates, the viral coefficient, and revenue per user.

### Churn

In this context, churn is defined as the loss rate of game players. Social gaming, because of its very nature, can have an extraordinarily high churn rate. Typical estimates are that, on average, social games have a churn rate of 50% per month – meaning that half of the new players signing up today will be gone in a month.

### Viral Coefficient

The viral coefficient is a measure of how effective current game players are at drawing new players – a key capability enabled by social network platforms. For example, if 100 Mafia Wars users are likely to cause five of their friends to sign up in a given month, that would be a viral coefficient of 1.05.

### Revenue Per User

Finally, there is expected revenue per user. This is an estimate of the lifetime revenue that a game player will generate, based on an estimate of monthly revenue per user and the churn. For example, if the average monthly revenue is \$5 per user, and churn is 50%, the expected revenue can be estimated as (\$5 (the first month) + \$2.50 (the second month) + \$1.25, etc.) or approximately \$10.

Understanding these characteristics is a major driver for a capable business intelligence platform in the social gaming space, these three key metrics drive investment in all areas of the business. Changing these numbers in even a small way on a per-user basis can make the difference between a game which earns \$5 profit per user, and \$50 profit per user. So understanding these numbers, and influencing them in their favor, is a top priority for social gaming firms.

But this is easier said than done. The first wave of social gaming applications on Facebook attempted to increase the viral coefficient with things like wall postings advertising in-game actions by players. But the popularity of this approach very quickly resulted in large amounts of “wall spam” – or game related wall postings that made it difficult for social network users to identify posts by friends. This caused the social network platforms such as Facebook to push back on the gaming firms and demand that they reduce wall spam.

This was the challenge that ignited the need for the Zynga team to perform graph analysis. It turns out that social games have a somewhat different social “graph” – or relationships between friends – than the social network platform itself. For example, in the Mafia Wars game people might have two types of friends, among others – those who actively play the game, and a more passive group who may have signed on to help expand a friend’s mafia organization but then left the game or play very infrequently. An important characteristic to note - players don’t always interact the same way with these two groups; within the active group, things such as gifts, offers of help, etc., are welcomed. Conversely, within the passive group, messages from the game, other players, etc. may be poorly received, even to the point of driving players away from the game.

Improving these interactions by guiding players to communicate appropriately with these two different types of relationships helps to increase revenue, reduce churn, and increase virality. In other words, to make every aspect of the game more profitable by improving the player experience significantly.

## The Solution

The first thing the Zynga team did was evaluate graph engines (dedicated software for graph analysis), however, none of the solutions they evaluated would operate at the necessary scale or performance. They quickly realized that Vertica would meet their needs, in the words of Dan McCaffrey – Director of Analytics Engineering at Zynga, “Vertica has MPP and scale solved. We can process data daily to produce an optimized graph.”

The Vertica Analytics Platform utilizes a native columnar storage and execution engine that offers a number of unique benefits over other data warehouse platforms. Because it is columnar, the Vertica Analytics Platform immediately delivers performance improvements due to the reduction in I/O necessary to service queries. While other columnar platforms such as “hybrid columnar” claim to offer the same benefit, Vertica is built from the ground up to push the I/O reduction to the limit. Thanks to Vertica’s advanced columnar compression technology, customers report data compression rates of 10-60x.

Next, the Vertica columnar execution engine delivers a number of further performance enhancements. These include the “just-in-time decoding” capability, which when processing queries, keeps the data in encoded form for as long as possible, thereby significantly reducing CPU and disk I/O. When combined with the optimized query plans produced by Vertica’s native columnar optimizer, the above features translate into dramatic performance benefits.

Furthermore, Vertica’s Massively Parallel Processing (MPP) architecture allows customers to deploy and run Vertica using industry-standard hardware or cloud solutions as building blocks called “nodes” – so users can build clusters with as many resources as they need. This means a small startup can deploy Vertica on a single node, and then simply add new nodes as needed. One, ten, a hundred or more nodes can be assembled into a Vertica cluster, putting thousands of processors, terabytes of RAM, and petabytes of disk to work as a single, parallel cluster. These attributes enable Vertica to deliver dramatic performance improvements on lower cost hardware, at scale and performance unparalleled in the industry.

In addition, Vertica’s rich set of in-database analytic functions, such as conditional and event-based windows, window-based lead, lag, etc., greatly simplify the task of preparing the data. Vertica also performs exceedingly well at the task of in-database transformations of data – a key task in preparing the graph data. For each user, all their interactions with other users are transposed to a single row – enabling the identification of the graph.

Graph analysis has traditionally been a showstopper for relational databases. While the table structure is simple – typically just a few columns – to express the graph and analyze it requires self-joins to express the connections in the graph, which result in an explosion in the number of rows in the result set due to combinatorial factors. Row-oriented database platforms are simply unable to cope with the massive volume of data created during the execution of graph queries.

Vertica, on the other hand, handles this quite well. Vertica’s columnar features deliver extremely high levels of performance on the graph data, and its MPP architecture allows the system to scale as needed. On this topic, Dan tells us that “...being able to run this on tables with tens of billions of rows of data with a fast turnaround is amazing.”

Zynga captures real-time game event data to several Vertica clusters. Social graph related data is separately streamed in real time to a dedicated cluster where the graph is generated on a daily basis. Each night, the models resulting from this graph are fed back into the game for use the next day. This has been successful enough that the Zynga team plans to expand its Vertica social cluster by 4x.

With this solution, the Zynga team has been able to improve the targeting of items such as gifts – an important type of game interaction – effectively increasing the level of interaction between the active players while minimizing spam to the passive players. Also, they have begun to build an item of tremendous value – a graph of their active social gaming network.

Having mastered the basic graph analysis challenge, the Zynga team is already looking towards the next challenge. As they expand their graph cluster, they will begin to enable capabilities to identify clusters of users with like behavior or common paths – which will enable even better targeting of game related promotions and activities.

## Conclusion

If social networking represents a major advance in the way businesses operate, social graph analysis represents an equally important advance in the way social networks operate. As the Zynga team discovered, graphing the relationships of their customers – the players of their games – enables them to improve the player experience as well as their own bottom line. By combining a native columnar engine with a massively parallel architecture, standard SQL capability as well as a rich set of in-database analytic functions, the Vertica Analytics Platform is uniquely suited to fulfill Zynga’s current and future needs, both as an enterprise data warehouse and as an advanced analytics platform.

## ABOUT VERTICA

Vertica Systems, an HP Company, is the leading provider of next-generation analytics platforms enabling companies to monetize their data at the speed and scale necessary to deliver significant value to customers and shareholders. Vertica’s scalability and flexibility are unmatched in the industry delivering 50x-1000x faster performance at 30% the cost of traditional solutions. Vertica is used by more than 300 customers across a variety of industries worldwide including Groupon, Twitter, Verizon, AOL, Guess?, Zynga, Playdom, BlueCross BlueShield, AdMeld, Sunoco, Mozilla and Comcast.

For more information or to request a demonstration, contact Vertica at (978) 600-1000 or visit the company’s website at [www.vertica.com](http://www.vertica.com).