



**Grid Dynamics**

# Grid Dynamics Powers Business Analytics Solution for Clickstream Data

Global digital marketing agency is able to collect and analyze 100 million events every day

## *Problem*

- Rich, interactive websites generate ~500 GB per day of clickstream data from thousands of sources and need to be analyzed in real-time for improved marketing performance
- Siloed third-party BI tools are unable to provide complete, integrated, and flexible insight into user behavior
- The agency's clients are unable to properly measure marketing campaign performance due to lack of insight

## *Why Grid Dynamics?*

### *Experience and expertise in*

- delivering integrated solutions that include ETL processing, data warehouse, and reporting across large data volumes
- implementing a single, consolidated web-based console for data analysis
- working with multiple third-party data warehousing frameworks

## *Business Results and Technical Benefits*

- Ability to turn 100 million events per day into actionable information
- Ability to make changes to campaigns on the fly based on user behavior
- Ability to meet SLAs with measurable results

There's no doubt that more and more of today's business is conducted online. Rich media websites, robust eCommerce platforms, and responsive mobile apps allow consumers around the world to view content, compare products, purchase goods and services, and interact with customer service and support. This trend is generating mountains of valuable clickstream data that can be analyzed by businesses to improve their insights into user behavior. This enables on-the-fly optimization of marketing campaigns to improve conversion rates.

## *Case Study*

Business intelligence into clickstream data is extremely important to a large digital marketing agency that works with clients to enhance their web presence through online advertising and marketing. The agency built and manages a 200 TB database environment to collect 100 million events per day that occur on its clients' websites and mobile apps. These enriched and aggregated events add 70 GB of compressed data to the database environment every day, giving data scientists valuable insight into user behavior.

After developing a series of in-house tools, the agency decided that the increasing complexity and growth of clickstream data would require a more robust solution that was more scalable and flexible and easier to use. Initially, the management team looked into third-party data providers and data services to mine the data, however, each tool provided insight into just a single data stream



rather than giving the agency a holistic view into all clickstream data. As a result, the information was siloed, preventing the agency from generating business intelligence that took all data sources into account. This lack of visibility and integration made it clear that the agency needed a holistic solution that would reduce complexity and deliver the ability to drill down to specific data streams for in-depth analysis.

“We work with our clients to collect all sorts of valuable information on user behavior, but the data is only as valuable to clients as the agency’s ability to discover insights and create actionable information,” said the Vice President of Engineering for the digital marketing agency. “A new advertising campaign that we create for our clients may experience an increase in users or an improving conversion rate, but the maddening thing is that these companies will never know. We need valuable insight into clickstream data that we can share with them and make informed decisions quickly.”

### A Powerful, Yet Intuitive Solution

The global digital marketing agency engaged with Grid Dynamics, a leading provider of open, scalable, next-generation big data solutions, to help build a data warehouse environment to handle the large volume of clickstream information. The Grid Dynamics professional services team helped the agency enhance its existing data warehousing environment to make sure that the appropriate clickstream data is being collected, stored, and made available for analysis.

The resulting solution is made up of a powerful yet flexible ETL processing environment that captures 100 million events from dozens of the agency’s customers’ websites and applications every day.

Clickstream data is collected from websites, social media, and mobile apps, requiring the environment to pull information from multiple channels. The 70 GB of clickstream data generated each day is stored in a robust and scalable Apache Hadoop cluster with 25 nodes, and a consistency checking mechanism ensures that each of the 100 million transactions is captured, catalogued, and made available for further analysis.

The data is then fed through an aggregation pipeline where it can be viewed and analyzed using out-of-the-box and custom reports through a number of data warehousing frameworks such as Hive, Hue, Pig, and Cascading. Grid Dynamics ensures that thousands of data streams are consolidated in a single environment, providing accessibility to both data scientists and front-line users for easy analysis. Users can access the data through these third-party interfaces, allowing the data to be analyzed on the fly or through scheduled reports.

### Meeting SLAs with Measurable Results

The data warehousing solution enhanced by Grid Dynamics allows the global digital marketing agency to analyze large volumes of clickstream data for its customers, enabling quick, informed decisions on marketing and sales strategies based on valuable insight into user behavior. This allows the agency to adjust advertising campaigns and media structure on the fly, enhancing the impact of each campaign. For a company that enters into service level agreements with its customers and relies on getting measurable results the benefits are mission critical.

“Ultimately our goal is to create and manage digital marketing campaigns that drive results for our customers,” said the Vice President of Engineering. “Grid Dynamics ensures we are doing everything we

can to deliver smart, targeted campaigns based on real user behavior. We'd be twisting in the wind with no way to know what was working and what was not working without the valuable insight the solution provides."

## About Grid Dynamics

Grid Dynamics is a leading provider of open, scalable, next-generation commerce technology solutions for Tier 1 retail. Grid Dynamics has in-depth expertise in commerce technologies and wide involvement

in the open source community. Great companies, partnered with Grid Dynamics, gain a sustainable business advantage by implementing and managing solutions in the areas of omnichannel platforms, product search and personalization, and continuous delivery. To learn more about Grid Dynamics, find us at [www.griddynamics.com](http://www.griddynamics.com) or by following us on [Twitter@GridDynamics](https://twitter.com/GridDynamics).



**Grid Dynamics**

4600 Bohannon Dr., Suite 220 • Menlo Park, California 94025 • 650-523-5000 • [sales@griddynamics.com](mailto:sales@griddynamics.com) • [www.griddynamics.com](http://www.griddynamics.com)

© 2014 Grid Dynamics International, Inc. All rights reserved. Other product names mentioned herein may be trademarks or registered trademarks of their respective companies.