



# ECOMMERCE WEB ANALYTICS

*The client, the nation's leading online wedding marketplace, was undergoing tremendous growth, leading to increased web traffic and ever expanding volumes of customer data. As a result, the client's data was becoming too large for their local management system, and they were seeking a cost-effective solution that could be scaled easily as their data volumes grew.*



## BUSINESS CHALLENGE

- Client required processing of huge unstructured web-click data in legacy data warehouse
- Analytics of web-click data was key to customer intelligence



## THE DISYS SOLUTION

- Solution framework involved automated raw-data parsing, distributed loading to setup of analytics system on AWS cloud
- Services span automation, infra setup and management, data transformation and loading
- Highly skilled Cloud services team executed project in rapid fashion, using public cloud infrastructure and programming tools
- Flexible model



## RESULTS

- Solution provides new analytics for client on web data on visitors and prospects on their online marketplace
- Scalable, On-demand infrastructure usage reduces cost and aligns with dynamic requirement
- Automated solution decreases ongoing service costs in data transfer, data transformation and update

## BIG DATA ON AWS CLOUD

After careful consideration, DISYS recommended Amazon Redshift as the optimal solution due to its powerful query times, security functionality, and ability to seamlessly scale up and down depending on data size. This solution included setting up an appropriate data scheme on Redshift, processing a legacy data warehouse and raw web log files, and transforming and loading the files to S3 both in batch mode and continuous loading mode using popular ETL platforms. The ability to optimize loading from S3 to Redshift, and Redshift cluster ability to be scaled up and down were key parameters in helping process hundreds of gigabytes of data.

A solution framework was designed and developed, incorporating server, storage, DB, warehouse, and analytics components. The key tasks of automating the data parsing, transformation, loading, and scalable cluster provisioning in batch mode were performed by the DISYS Cloud Services Team. The underlying infrastructure components enabled every step of the process to be reliable, secure, fast, scalable, and easily accessible over the web.

DISYS performed AWS cloud infrastructure services including consulting, design and architecture, migration and deployment, and managed services. DISYS' Cloud Services Team has experience with all AWS components including Redshift, EMR and Data pipeline. These services are underpinned by DISYS' Sirro platform, providing users with a complete cloud management solution

including provisioning, user management, dashboard information, resource health monitoring, and an intelligent billing engine.

## AWS COMPONENTS

AWS Redshift was used for a cost-effective, scalable and powerful data warehouse cluster solution. This was key because the bulk of the data load was performed in batches from S3, making a dynamic cluster a cost-effective solution. All data was distributed using keys, and loaded into the reliable S3 repository. EC2-AMI-based Microstrategy was used to deliver real-time business intelligence.

## DISYS EXPERTISE IN OPTIMIZING THE BUSINESS SOLUTION

Prior to contacting DISYS, the client did not process the data, and allowed it to accumulate in the log file and warehouse. Processing of such data using legacy RDBMS was slow and daunting. DISYS' solution framework included a single front-end repository for batch data collation, rapid loading into scalable Redshift cluster, and using an ETL platform for continuous connection and loading into Redshift. DISYS' team also set up the SQL front-end for simple queries, and the Microstrategy instance for full analytics.

DISYS' team built a flexible automation engine application using extensible programming frameworks. It was orchestrated by using the AWS SDK, and integrating the AWS components at the backend. The Sirro platform monitored the load

## OUR OFFERINGS

DISYS provides end-to-end, project-based, vendor-independent Infrastructure Support Services. DISYS supports global enterprises, government agencies, and organizations of any size in managing and operating their IT systems and services to optimize efficiency, security, and economy.

DISYS' comprehensive, solutions-based, Infrastructure Support Services include:

- Data Center Services
- Messaging Services
- Networking & Telecommunication Services
- Security Services
- Storage Services
- End User Computing (EUC) Services

“ DISYS proved that a complete ‘analytics-as-a-cloud-service’ could be provided that met speed, security, flexibility and dynamic reporting requirements by using a strategic, scalable infrastructure. ”



on the Redshift cluster, thus triggering scaling up and down of the cluster. DISYS’ team also enabled key-based parsing of inbound data and distributed loading, greatly increasing the loading speed of the voluminous data.

## **SIMPLE, TURNKEY SOLUTION**

The data store and processing requirement for the web-clicks was not an easy task. DISYS proved that a complete ‘analytics-as-a-cloud-service’ could be provided that met speed, security, flexibility and dynamic reporting requirements by using a strategic, scalable infrastructure. The client obtained the solution as a service from DISYS and AWS, provided by highly skilled engineers, using powerful infrastructure at low cost.

---

DISYS, LLC is an IT staffing and consulting company serving Fortune 500 and other global-scale enterprises worldwide. DISYS delivers strategic value by understanding and responding to a client’s environment, problems, and challenges either by assembling the most highly talented team for any job or delivering more comprehensive, cost-effective IT solutions. Incorporated in 1994 as a certified Minority Business Enterprise, DISYS is headquartered in McLean, Virginia, with offices and delivery locations worldwide.

© 2015 Digital Intelligence Systems, LLC. All trademarks are property of their respective owners.

