

# Avito

## Big Data Analytics: at the heart of Avito

### Overview

Millions of people every day go to the Avito portal to sell or buy almost anything they want: from digital technology and clothing to cars or real estate. Advertisements are posted by individuals, entrepreneurs and companies, so people can buy both used goods and brand new products.

### Challenge

Avito, the number-one classifieds site in Russia, wanted to collect and save all its business-related data. Big Data has become part of day-to-day reality for the company, and is instrumental to its success. Avito employees handle both gigantic volumes of information and endless streams of many types of complex data on a daily basis.

Data is currently sourced from 26 different systems, including the main website and other

**"We follow an approach whereby top managers carefully study data and test various hypotheses before making a decision."**

#### IVAN GUZ

BI Director  
Avito

online projects (such as Domofond.ru), the CRM system, software solutions for managing banners, newsletter and SMS mail outs, mobile app components and much more. A graphic example of Big Data entering the corporate network at a very high rate is clickstream.

It is a stream of events reflecting all user activity on websites: clicks, page jumps and other activities. Almost one billion events are recorded on a daily basis, and sometimes as many as two to three million events per minute. Data from the company's back office systems storing customer classifieds are the most diverse: their structure is becoming progressively more complex and multi-tiered.

The company also accumulates external data from such context advertising systems as Yandex. Direct, Google Adwords and Rubicon, from a number of other marketing data sources, from social networks and also foreign exchange rates from the European Central Bank.

However, the tools based on PostgreSQL and Excel used prior to 2013 did not allow Avito to collect and process this valuable clickstream data. They were good enough only for preparing simple reports.

Avito formed a Business Intelligence (BI) unit in mid-2013, shortly after gaining leadership in



### At a Glance

#### Industry

Computer Services

#### Location

Moscow, Russia

#### Challenge

Avito wanted to enable the versatile use of available data to develop various business streams and generate additional benefits.

#### Products and Services

Big Data Analytics

#### Results

- + Created a single corporate analytics system based on the BI platform.
- + This transforms data into an asset that can successfully generate revenue for the company.
- + Maximizes the benefits of available data and increases revenue from data.

its market segment following a takeover of two competitors: OLX.ru and Slando.ru. Company management then tasked the BI unit with monetizing corporate data through in-depth analysis of incoming information.

### Solution

To achieve data monetization by transforming data into a successful revenue-generating asset, the BI unit decided to focus its efforts on optimization of pricing models for commercial services, banner ads, A/B testing of user interfaces of the website, automatic moderation of classifieds and the preparation of financial statements.

"These plans called for using a powerful tool that would allow both to collect, store and process large volumes of complex data and to scale up capacities to meet the requirements of future tasks," explains Ivan Guz, BI director, Avito.

A key requirement for the BI platform is its virtually unlimited scalability: both in terms of the volumes of data and the speed with which it is collected, stored and processed.

"At the time when the platform was being chosen, it was impossible to predict the number of data sources (only two systems were used originally), the volume of data, the rate at which information would be received, the specific analytical tasks to be solved or the approaches to solving them," adds Nikolay Golov, corporate data storage architect, Avito.

Notably, Avito required full scalability of the BI platform not only in terms of performance and data volumes but also in terms of analytical coverage of the progressively more complex data. Another key parameter of selection was support of SQL queries, while the third key requirement was system scalability based on standard x86 servers purchased from different vendors.

The platform was chosen based on the outcome of tests conducted by Avito experts. Actual company data was used as test data to verify that the platform was capable of loading them quickly.

Three bidders made it to the final stage of selection: the platforms Micro Focus®, Vertica, Greenplum, and Oracle Exadata. Avito experts did not like the capabilities of using standard servers and their scalability features offered by Oracle; Greenplum turned out to be slower than Vertica in terms of data loading speed.

"We knew that our project was at the cutting edge of technological innovation, which is why one of the vendor selection criteria was the ability to request assistance from colleagues at other Russian companies in case of major problems. By that time Vertica had already been deployed at Yota, so ultimately we opted for this particular product," recalls Golov.

The company deployed the BI platform using in-house resources. Avito experts contacted Micro Focus experts only when the most challenging issues arose. Prior to the project launch, the company's entire server infrastructure was based in Sweden, purchased under a leasing arrangement from a data center owner. The Vertica platform was deployed on a cluster of three servers, and two more servers were used for data extraction, conversion and loading, more commonly known as ETL. The integration with clickstream was then achieved using an arrangement with intermediate caching of data describing the history of events over the past three to four days on a MongoDB Database Management System.

The biggest challenges were encountered when the time came to expand the BI system. Once the data volume exceeded a certain threshold, the company had to give up the old approaches to data management and find new, more effective ways of data processing. The

company also faced difficulties when expanding the cluster that hosted the Vertica platform.

Avito experts believe that all of these complications resulted from insufficient experience at the time of platform deployment.

### Results

The BI system based on Vertica has become an integral part of the Avito business model, without which the company would be unable to succeed because all information received by the company trickles into the BI system. Since mid-2016, BI director Ivan Guz has become one of the company's top managers, making key decisions relating to Avito business growth.

The Vertica platform has since been deployed on a cluster of 14 servers at one of the Moscow-based data centers (Avito is required by law to use a local data center because it handles personal data); three more servers are available in cold standby mode; one server is reserved for ETL procedures; and eight more MongoDB servers operate as part of a cluster that caches clickstream events. The number of Vertica servers will be doubled (to 28) in the near future; one additional server will be added to the MongoDB cluster and one more backup server for ETL will be installed.

Efforts to expand system capabilities are ongoing; every time Avito launches functional modules supporting new business streams for the company, they are instantly integrated with Vertica. Tools for analyzing new data are also added, including the latest data analysis techniques such as deep learning and computer vision.

The system is primarily used by Avito employees at the BI unit; of the three dozen employees, six specialize in maintenance and expansion of the system based on the Vertica platform, while others focus on data analytics for various applied tasks.

---

They include optimization of pricing of commercial services depending on their category, user geography and other parameters, moderation (identification of duplicate classifications, category violations, fraudulent classifieds), optimization of banner advertising (including CTR auctions), targeted mail outs of newsletters and SMS, and CRM analytics. Vertica is also used in preparing reports for investors with financial performance indicators, traffic volumes and other parameters.

The company's top managers also use the analytical features of Vertica: "We follow an approach whereby top managers carefully study data and test various hypotheses before making a decision," says Ivan Guz, BI director, Avito.

Data analysis is performed by analysts in the company's key business units. There are immediate plans to grant access to Vertica to any Avito employee whose job requires analyzing specific data.

The BI unit operates according to the Data Lab concept, which is based on finding new ideas

and approaches to using data and analytics with tangible, monetizeable benefits for business.

"For example, we realized a need to constantly improve our fraudulent classifieds detection algorithms to make them more effective," comments Golov.

All new Avito business projects are currently implemented with the involvement of the BI unit which helps them work out every possible way to manage data: from the integration of new functional modules on the Vertica platform to data analysis and usage. All new data received by the company is directed into the BI system, including both the latest functional modules and new external data sources that Avito chooses to use.

Many BI tasks are solved through open data analysis competitions in which Avito invites input from external professionals and then selects the most promising ideas and approaches. The company often hires the authors of the most interesting ideas.

"Avito has done a colossal amount of work not only deploying and mastering the Vertica analytics platform but also cultivating a culture of using data and business intelligence tools for accomplishing all sorts of business tasks," says Evgeny Stepanov, who leads the Micro Focus Big Data Platform business in Russia, adding: "This customer's projects highlight the most notable advantages of Vertica. Avito does a spectacular job using the features of this platform in line with international Big Data best practices."

"We measure how features are used and if they are not popular, we can drop them. Customized voicemail is one such example—customers weren't using it and it wasn't adding value, so we removed it from the app," concludes Medrano.

"Vertica is helping us to roll out three or four releases of the app every month because we can analyze the user data quickly and can respond to user demands. Additionally, this data is available to all of our 150 internal developers, not just the three-man analyst team. We call it democratizing data for better business."

**“At the time when the platform was being chosen, it was impossible to predict the number of data sources, the volume of data, the rate at which information would be received, the tasks to be solved or the approaches to solving them.”**

**NIKOLAY GOLOV**

Corporate Data Storage Architect  
Avito

Contact us at:  
[www.vertica.com](http://www.vertica.com)