



pandora
big data solutions

PANDORA INSIGHT BIG DATA SOLUTION

www.pandora-insight.com

Big Data analytics is on top of business imperatives in many important industries oriented to retail consumers, including communications, banking and insurance, retail and FMCG. Market leading vendors are trying to collect feedback about user experience related to their brands and services and to compare it to user experience of competitive brands. Conclusions based on relevant information are used for planning their market activities.

Most valuable data is spread on various review web sites, different discussion forums and social networks such as Facebook and Twitter, unstructured and unformatted and in different languages. To gather relevant feedback from customers all the data will be collected, filtered, stored, aggregated and analyzed.

Requirements for such analysis can be clearly defined:

- A Big Data analytical system must deliver information in real-time or as near to real-time as possible.
- Data sources must be as complete as possible. Defining a new data source and mapping the data from new data source must be easy and fast to implement and should not require technical skills.
- Semantic analysis, filtering of non-relevant content, relevancy rating and rules for scoring of sentiment must be clearly defined and tested.
- Information dissemination and analysis must be simple to use but powerful, with set of predefined intuitive visualizations of relations between relevant variables and possibility for the end-user to create new reports and analysis.

There are several ways to perform such analysis and each of the standard ways has some advantages and disadvantages. Market research agencies can deliver on-demand research study, but the relevancy of such study is always questionable and in most cases results are outdated due to Volume-Velocity-Variety nature of Big Data. Another option is to implement internal Big Data analytics solution that will assure relevant data, but with high initial investment in technology and development and long time to deliver first results.

Our unique **Big Data Analytics as a Service Solution** is the ideal solution for businesses that need such information. You can have a powerful and meaningful set of analytics within a few days, without large investment in technology infrastructure and skilled technology resources, based on our robust, scalable and comprehensive service architecture.*

* Service is currently available for test users.

Functionality

Pandora project deals with the development of systems for collection of web reviews, evaluation and aggregation of evaluated results, as well as packaging the evaluated results for the purpose of presentation to end customer.

The system also includes administration functionalities, such as job scheduling or manual job initiation, gathering information about job runs and statistics about different processing segments in order to identify and follow correctness of system functioning and to gain insight into evaluation accuracy.

For example, a user defines mapping to crawl data from a traveller website about hotels. Data is crawled from the website and extracted to a database based on defined mappings. Sentiment analysis, data transformation and aggregation are applied to collected data. The result is a graph where end users can see product grades (based on text reviews, evaluated by sentiment analysis process) for products they are interested in.

The value for end users is in the fast and simple evaluation of their product of interest based on other user opinions.

Our Solution has three main Tiers - Storage, Engine and Access Tier. Storage tier is based on standard open source Hadoop technologies, including NoSQL HBase and Impala Massively Parallel Processing (MPP) query engine.

Storage tier ensures required scalability and performance. It is tightly integrated with Data Access Interface Framework.

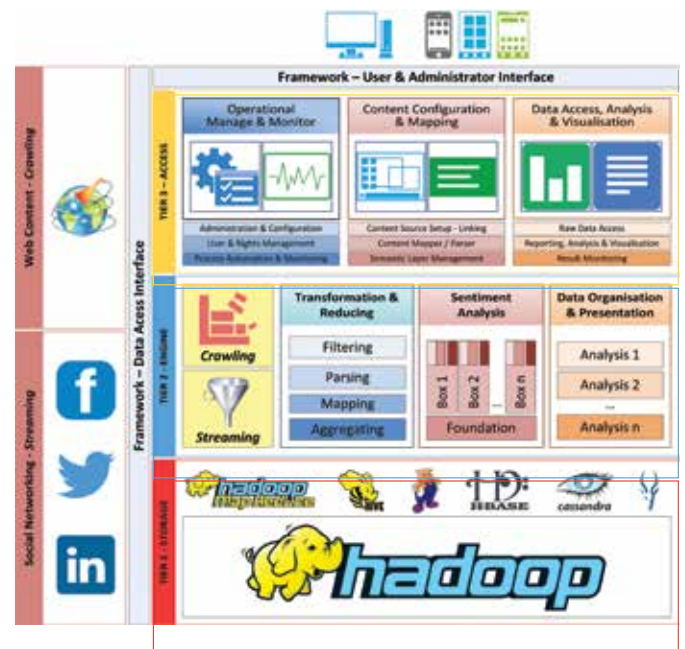


Figure 1. Big Data Analytics as a Service Solution Architecture

The **Engine Tier** is the heart of our solution. Defined content from web sites and social networks is streamed and crawled, and then parsed, filtered, mapped to a set of predefined attributes, aggregated and stored for each Product Type. Modular Sentiment Analysis is the next step in the process. Dictionaries are organized hierarchically, with Foundation level applicable for all Products. The third level is made of Blade dictionaries that are Product Type specific and have the best score, like Smartphone or Tablet Blade in Technology Box. Blades and Boxes for additional Products Groups and Types and content languages can be added easily. On top of the scored content we have a set of powerful analyses and visualizations.

The **Access Tier** handles Administration & Configuration, Management & Monitoring and End User requests through HTML5 web and mobile access interfaces. Users can do all required tasks from one simple interface without knowing the complexity of the underlying technologies. The process starts by opening the account and defining users within the account with different privileges, choosing of the already mapped sources or mapping the content of the web-site that you want to include in your analysis, defining the keywords and brands for analysis, scheduling and monitoring data processing, using existing visualizations and analysis or defining new visualizations.

Visualization

Visualization, presentation and correlation of extracted information is of vital importance in every decision making process. In order to be useful, visualizations need to be simple and easy to understand but also very clear and informative. Comparison of one's own products or comparison with similar products of the competitors according to different categories based on user opinions is of great value in market positioning and in planning penetration into new market segments. Our solution offers a variety of pre-defined visualizations where users can choose products for comparison, as well as categories on which comparison will be based. Two examples of product comparisons are described in detail.

Product comparison with a radar chart (Chart 1.) allows the user to gain insight into product analysis for all categories defined for respective product type and is very valuable for comparison of fewer products. All products are analysed by categories defined specifically for respective product type and graded according to 0 to 10 rating scale.

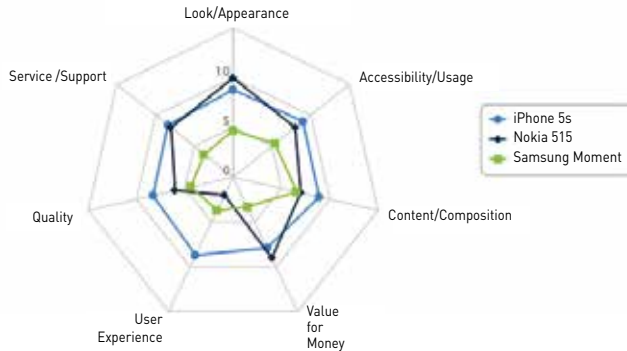


Chart 1. Radar chart showing comparison of analysis results for chosen types of smartphones.

Product comparison with a bubble chart (Chart 2.) enables insight and comparison of a larger number of products by three categories – two of which are presented on chart's axes while the third one is presented by bubble size. Comparison includes three categories from the radar chart from Chart 1. Two categories (Quality and User Experience) are shown as chart axes, while the third category (Value for Money) is shown as bubble size. On mouseover bubbles exact grades for three chosen categories for respective product are shown.

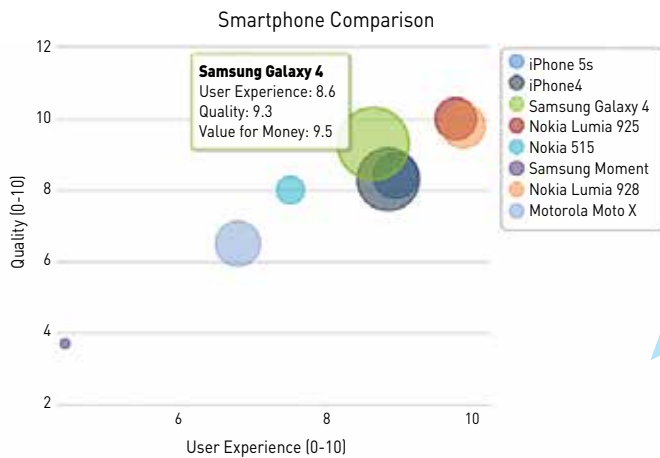


Chart 2. Bubble chart representing comparison of analysis results for a larger number of smartphones.

About Bird Consulting and Poslovna Inteligencija

Bird Consulting (Business Intelligence Research and Development) is a UK branch office of Poslovna inteligencija. Poslovna inteligencija is a regional leader in Southern and Eastern Europe in implementation of intelligent information systems for decision support with over 70 corporate customers in 20 countries and over 160 successfully completed projects. The company is focused on areas of strategic ICT consulting, implementation of Big Data analytics, business intelligence and data warehousing systems, data integration and data quality management systems, performance management systems, systems for planning and budgeting, risk management systems (Basel III, Solvency II), Balanced Scorecard (BSC) and Master Data Management (MDM) systems, based on all world leading technologies. In offices in Zagreb, Ljubljana, Belgrade, Sarajevo and Podgorica, the company employs more than 80 certified professionals skilled in implementation of these systems.

Contacts:



BIRD Consulting (UK branch office)
Business Intelligence Research and Development
88 Wood Street
London, United Kingdom
E-mail: london@birdconsulting.eu

poslovna
inteligencija
Poslovna means Business

Poslovna Inteligencija d.o.o.
Zagreb, Stubička 50b
10000 Zagreb, Croatia
Phone: + 385 (1) 4617 945
Fax: + 385 (1) 4617 946



pandora
big data solutions

BIG DATA SERVICES

We are able to address continuously growing demands for big data analytics. Big data analytics is the concept of data analysis and derivation of useful information from large volumes of data that are rapidly being generated in modern business – from conventional database systems to unstructured data gathered from location and sensor sources and social networks. We provide services needed for successful preparation, design, development and implementation of Big Data Processing and Analytics systems based on standard platforms. We offer two specific service packages, described in more detail – one for fast kick-off of your Big Data initiatives and the other for fast and efficient solution development. Contact us for more information!

BIG DATA READINESS ASSESSMENT

You would like to start your first Big Data initiative, but you are not sure what to do and how to create value for your organization? You don't have a clear picture of investment and resources you need to secure and what benefits you can expect? You would like to protect your existing investment in IT infrastructure? Our unique Big Data Readiness Assessment package, based on our Information in Action methodology, will help you to move in the right direction fast and efficient, by identifying most lucrative use cases for your business and optimal architecture, technologies and models of implementation, where you will be able to find and use the hidden value of your data in shortest possible timeframe.

BIG DATA EDUCATION AND TRAINING COURSES

Big Data training covers objectives of every data professional: developers, administrators and data analysts. Information provided on the training ranges from pure and fundamental description of Big Data ecosystem, paradigm and component overview to fully practical hands-on training based on specialized technologies that run on top or within selected Big Data platform stack. We have defined three different levels of courses - beginners, technical fundamentals and analysts, that can be combined based on customer preference. Education combines basic and advanced understanding of Big Data topics (with focus on Hadoop) and sharpens skills with courses in data visualization, data mining and statistics.

BIG DATA DEVELOPMENT

You are faced with a lack of resources for your Big Data implementation and you need more Data Scientists and Hadoop developers? We have a pool of experts experienced in different areas of Big Data implementations, ranging from MapReduce and Pig scripting, data streaming and collection with Apache Nutch and Flume, data transformation and integration with Hive, Pentaho Kettle and Informatica, MPP, columnar and in-memory databases including Cloudera Impala and SAP HANA, to semantic analysis using Knime and SPSS and visualizations using Pentaho, Tableau and various other front end platforms.