FRESHCODE
We make ideas happen

Skype: alex.freshcode
Email: alex.freshcode@gmail.com
Phone: +38 (063) 256 63 78

Data Parcer

Client: Data Parser

Site: under construction

Industry: Big Data

Country: USA

Size of the Team: 4 people

Duration: 9 months

Technologies Used:

JAVA
MONGODB
ELASTIC SEARCH
RABBITMQ
SPRING
HIBERNATE
ANGULAR
POSTGRES
BOOTSTRAP
REQUIREJS

OUR CLIENT

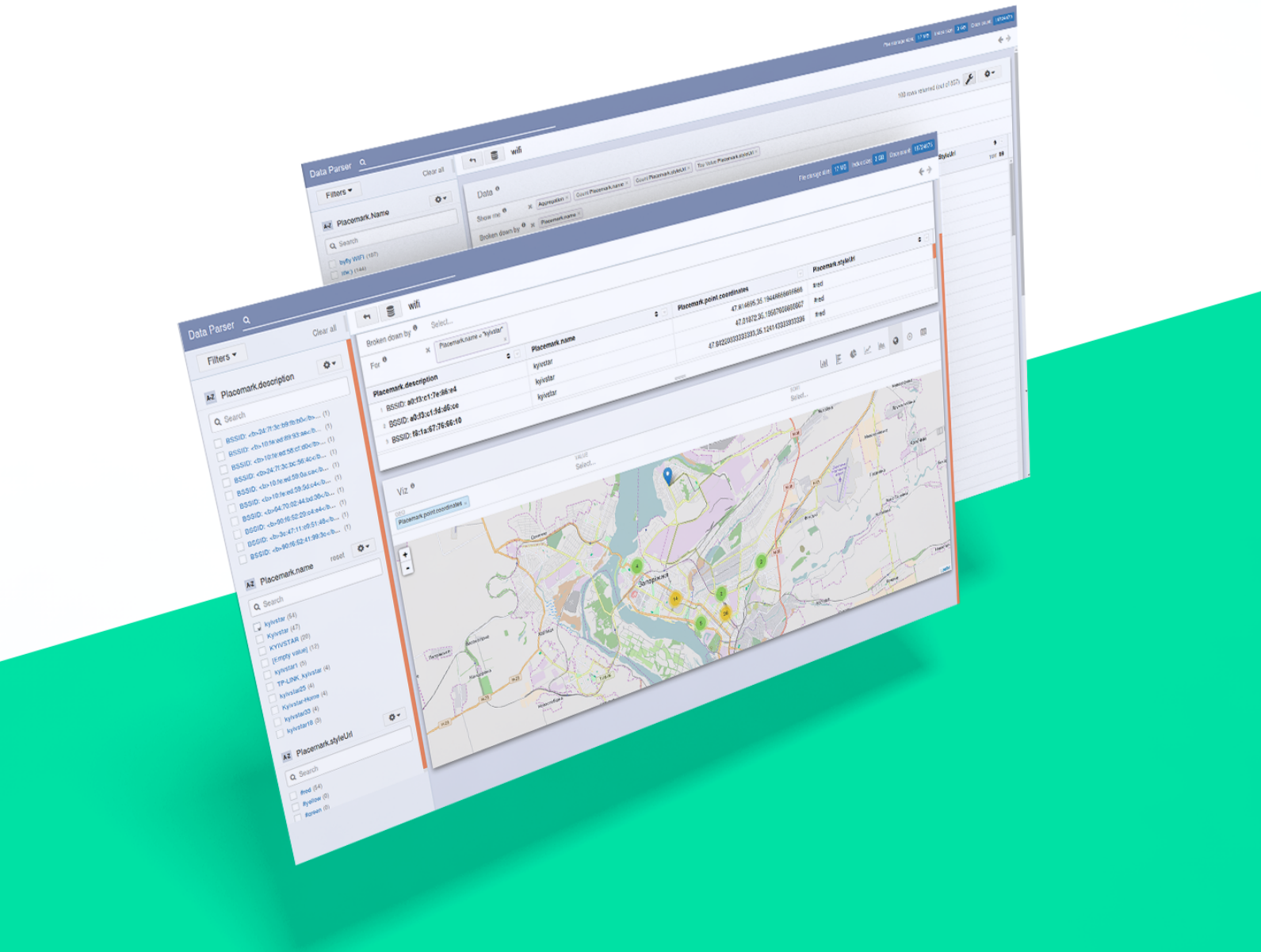
Data analysis is a highly competitive industry and only those who offer the most effective and user-friendly services can survive in this cutthroat business. Our client needed a cloud-based ETL (extract, transform, load) application for unstructured data from any number of sources with advanced search and data aggregation possibilities.

Skype: alex.freshcode
Email: alex.freshcode@gmail.com
Phone: +38 (063) 256 63 78

BUSINESS ISSUES

While creating Big Data analysis application for the client our team of developers faced some challenges, including:

- JSON files compatibility with unpredictable unstructured data.
- Possibilities of scaling up the structure for new clients.
- Large files support (up to 4 GB).
- Complicated data aggregation and complex search functions.
- Graphic user interface development for entry-level users.
- Providing API for third-party services.
- Regular index synchronization with external sources.



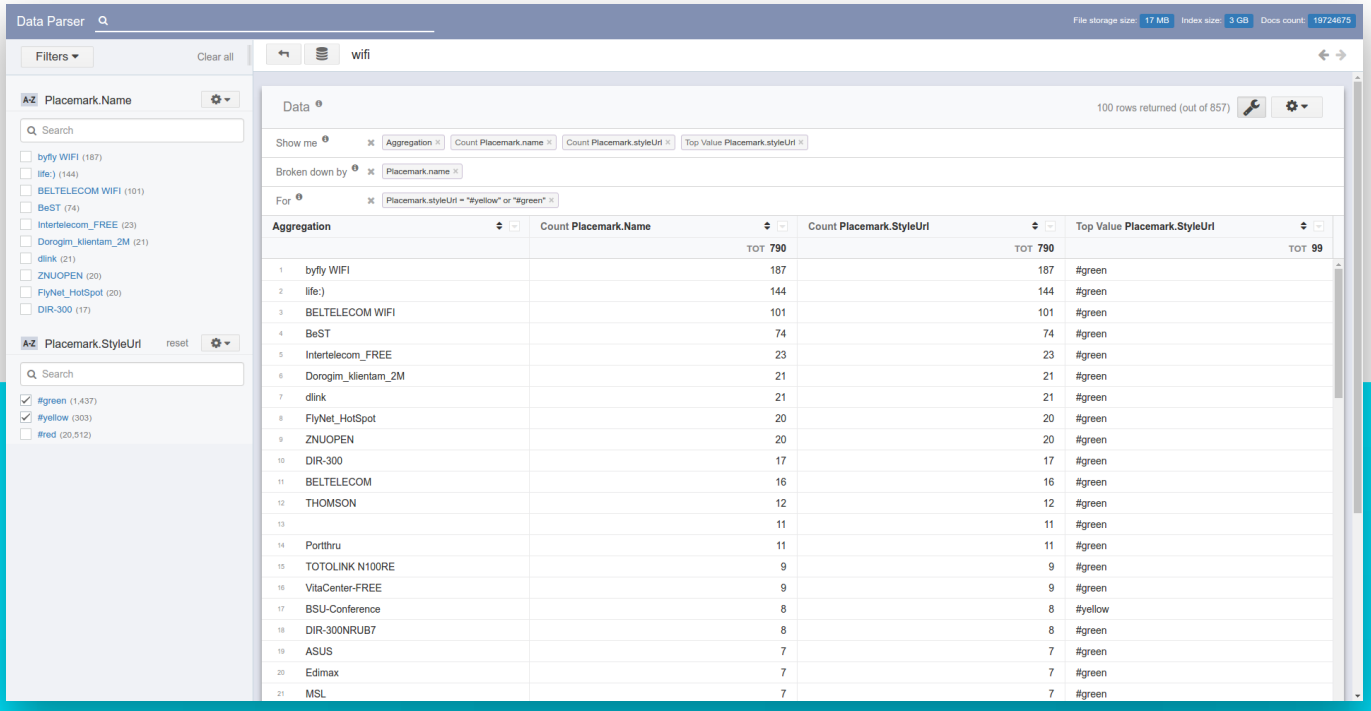
Skype: alex.freshcode
Email: alex.freshcode@gmail.com Phone: +38 (063) 256 63 78

SOLUTIONS

Our developers created the unique pre-processor to facilitate the work of Elastic Search with unstructured data. The self-adapting algorithm was implemented for join-type operations. This algorithm determines data type and relevant join strategy for different data sources in order to accelerate processing speed and improve system's productivity.

For quick and secure data exchange between allocated servers, message broker was created based on RabbitMQ and MongoDB. This solution synchronizes servers' internal workings and divides problems among allocated nodes of a block.

Our team developed the user-friendly ETL editor (GUI) for entry-level customers. Data model is easy to manipulate by moving and joining graphical representations of data blocks and operations. Web-interface synchronization prevents collisions and conflicts caused by different users' actions. The number of users is defined by allocation block parameters and can be adapted at any time.



BUSINESS VALUE

Our client is going to provide Big Data services to IT giants such as Google and IBM. Data Parser customers will be able to choose between two subscription plans. Large companies with a huge amount of data will get access to dedicated large storage servers for quick calculations. Shared services with lower productivity will be provided to smaller customers with simpler needs.