

# RTBHOUSE =

## Big Data Engineer

### Team Core

#### TWOJE ZADANIA:

- Rozwój i utrzymanie rozproszonych systemów przetwarzających duże ilości danych (większość real-time) na potrzeby naszej platformy RTB
- Optymalizacja tworzonego oprogramowania pod kątem wydajności i zużycia zasobów
- Zapewnienie niezawodności i skalowalności budowanych rozwiązań
- Tworzenie testów wydajnościowych i poprawnościowych dla nowych elementów systemu
- Analiza nowych technologii pod kątem możliwości ich zastosowania w warunkach produkcyjnych
- Rozwój narzędzi monitorujących i analizujących pracę produkcyjnego systemu
- Ciągła optymalizacja istniejących narzędzi i procesów

#### WYBRANE WYKORZYSTYWANE TECHNOLOGIE:

- **Java, Python**
- **Hadoop, Kafka**
- **Kafka Streams, Flume, Logstash**
- **Docker, Jenkins, Graphite**
- **Aerospike, PostgreSQL**
- **Google Big Query, Elastic**

#### PRZYKŁADOWE TEMATY:

##### Wybrane zagadnienia, którymi zajmowaliśmy się w ostatnim czasie:

- Wymiana frameworka w komponencie przetwarzającym dane (przejście ze Storm-a na Kafka Streams)
- Stworzenie mergera strumieni danych w oparciu o Kafka Client API
- Stworzenie synchronizatora profili użytkowników między DC-kami w oparciu o Kafka Streams
- Stworzenie komponentu wyliczającego agregaty w oparciu o Kafka Client API i filtry Blooma
- Wdrożenie Logstash-a do ładowania i Elastic-a do odpytywania indeksowanych danych (przejście z Flume + Solr)
- Stworzenie end-to-end monitoringu poprawności i opóźnienia danych
- Wymiana komponentu streamującego dane do BigQuery i HDFS (z Flume na własne rozwiązanie oparte o Kafka Client API) – w trakcie
- Ciągłe utrzymanie systemu, wykrywanie i rozwiązywanie problemów wydajnościowych, a także skalowanie ze względu na rosnącą ilość danych

**NASZE OCZEKIWANIA:**

- Biegłość w programowaniu
- Doskonałe rozumienie tego, jak działają złożone systemy informatyczne (od poziomu sprzętu, poprzez software, aż po algorytmikę)
- Dobra znajomość podstawowych metod tworzenia programów współbieżnych i systemów rozproszonych (od poziomu wątków do poziomu kontynerów)
- Praktyczna umiejętność obserwowania, monitorowania i analizowania działania produkcyjnych systemów (oraz wyciągania z tego wartościowych wniosków)
- Umiejętność krytycznej analizy tworzonych rozwiązań pod kątem wydajności (od szacowania teoretycznej wydajności projektowanych systemów do wykrywania i usuwania faktycznych problemów wydajnościowych na produkcji)
- Gotowość do pracy w modelu DevOps

**DODATKOWYMI ATUTAMI BĘDĄ:**

- Doświadczenie w tworzeniu systemów rozproszonych
- Dobra znajomość wybranych technologii Big Data takich jak Hadoop, Kafka, Storm, Spark lub Flink
- Znajomość metod i narzędzi profilowania aplikacji (najlepiej Java, zarówno z poziomu JVM-a jak i Linux-a)

**OFERUJEMY:**

- Atrakcyjne wynagrodzenie
- Pracę w zespole pasjonatów, którzy chętnie dzielą się swoją wiedzą i doświadczeniem
- Wyjątkowo elastyczne warunki pracy – nie mamy core hours, nie mamy limitów na urlopy, można w dużej części pracować zdalnie
- Dostęp do najnowocześniejszych technologii i możliwość realnego wykorzystywania ich w projekcie o dużej skali i wysokiej dynamice
- Sprzęt i oprogramowanie, jakiego potrzebujesz

Więcej informacji: [techblog.rtbhouse.com/jobs/](https://techblog.rtbhouse.com/jobs/)

Kontakt: [it-jobs@rtbhouse.com](mailto:it-jobs@rtbhouse.com)