BO·TT·EGA
IT minds

Program szkolenia:

Elasticsearch - architektura i zagadnienia zaawansowane

Informacje:

Nazwa:	**Elasticsearch - architektura i zagadnienia zaawansowane**
Kod:	**BigDataML-elastic**
Kategoria:	BigData, streaming i Machine Learning
Odbiorcy:	developerzy, architekci
Czas trwania:	3 dni
Forma:	30% wykłady / 70 % warsztaty

Szkolenie ma na celu poszerzenie wiedzy na temat technologii wyszukiwania pełnotekstowego z wykorzystaniem Elasticsearch oraz dobór właściwej architektury danych ze szczególnym uwzględnieniem modelowania danych dostosowanych do potrzeb.

Dodatkowym atutem szkolenia jest część administracyjna związana z konfiguracją, utrzymaniem, monitoringiem i skalowaniem Elasticsearch. Będzie ona niezwykle cenna dla osób zajmujących się administracją systemów oraz DevOps-ów.

W trakcie szkolenia omawiane są również typowe problemy, z którymi na co dzień spotykają się użytkownicy. Owe problemy to zbiór przepisów zbieranych od 2011 roku, czyli praktycznie od początku istnienia produktu Elasticsearch. Odniesienie do źródeł problemów, a nie tylko sposobów ich “łatania” pozwoli na zdobycie solidnych fundamentów do analizy tych które nie zostaną omówione w trakcie szkolenia.

Zalety szkolenia:

- Sprawdzone receptury
- Tuning wydajności
- Integracja z innymi systemami

BO·TT·EGA
IT minds

Szczegółowy program:

1. Architektura

1.1. Wprowadzenie do skalowalnych baz danych NoSQL

1.2. Wyzwania wynikające ze stosowania baz rozproszonych

1.2.1. Eventual consistency i teoria CAP

1.2.2. Zarządzanie infrastrukturą wielowęzłową

1.2.3. Problemy sieciowe

1.2.4. Rozdwojenie jaźni (split-brain)

2. Sposoby modelowania w dokumentowych bazach danych

2.1. Płaskie agregaty

2.2. Zagnieżdżone dokumenty

2.3. Miękkie relacje pomiędzy encjami

3. Wprowadzenie do wyszukiwania pełnotekstowego (full-text search)

3.1. Algorytmy stosowane w wyszukiwaniu pełnotekstowym

3.2. Możliwości rozwiązań umożliwiających FTS

3.2.1. Wyszukiwanie za pomocą jednego pola

3.2.2. Uwzględnianie literówek (fuziness)

3.2.3. Pomijanie nieznaczących znaków

3.2.4. Zakreślanie pasujących fragmentów tekstu (highlighting)

4. Dlaczego Elasticsearch?

4.1. Omówienie alternatywnych rozwiązań i porównanie możliwości

4.2. Ekosystem Elastic Stack

5. Korzystanie z Elasticsearch

5.1. Instalacja i konfiguracja Elasticsearch

BO·TT·EGA
IT minds

- 5.2. Metody modyfikacja dokumentów
  - 5.2.1. Indeksowanie
  - 5.2.2. Aktualizacja
  - 5.2.3. Usuwanie
  - 5.2.4. Reindeksacja danych
- 5.3. Wyszukiwanie danych
- 5.4. Komunikacja z poziomu aplikacji
- 5.5. Wady i zalety wykorzystania Spring Data w warstwie dostępu do danych Elasticsearch
- 5.6. Testy jednostkowe i integracyjne mechanizmu wyszukiwania

6. Modelowanie danych w Elasticsearch

- 6.1. Dobór właściwej architektury składowania danych do problemu
  - 6.1.1. Przechowywanie danych w jednym indeksie
  - 6.1.2. Rozbicie danych pomiędzy indeksami
  - 6.1.3. Rozbicie danych pomiędzy shardami
  - 6.1.4. Przechowywanie danych określonych czasem (time-series)
- 6.2. Dynamiczne tworzenie struktur vs. statyczna kontrola typów
- 6.3. Zaawansowane podejście do analizy danych tekstowych (analizatory tekstu)

7. Zaawansowane wyszukiwanie z użyciem Elasticsearch

- 7.1. Omówienie różnych sposobów implementacji mechanizmów Quick search
- 7.2. Agregacja danych
- 7.3. Kategoryzacja dokumentów na przykładzie Percolator API

8. Performance tuning

- 8.1. Konfiguracja Elasticsearch pod kątem wymagań stawianych przed systemem
- 8.2. Co robić, gdy indeksowanie jest zbyt wolne?
- 8.3. Jak radzić sobie ze zbyt wolnymi zapytaniami?

BO·TT·EGA
IT minds

8.4. Modyfikacja architektury klastra celu zwiększenia wydajności

9. Integracja Elasticsearch z obecną architekturą

9.1. ELK, jak podstawa szybkiej integracji z działającym systemem

9.2. Rozwój funkcjonalności istniejących systemów poprzez implementację wyszukiwania pełnotekstowego

9.3. Elasticsearch jako jedyne źródło danych aplikacji

9.4. Metody integracji z innymi rozwiązaniami składowania danych

10. Utrzymanie i rozwój infrastruktury klastra Elasticsearch

10.1. Diagnozowanie typowych problemów

10.1.1. Długo trwające zapytania

10.1.2. Rażący spadek wydajności usługi

10.1.3. Rozszczepienie klastra (split-brain) oraz niepożądane złączenie środowiska testowego i produkcyjnego

10.1.4. Konflikt typów w atrybutach dokumentów

10.1.5. Niepoprawne wyszukiwanie danych z powodu błędnej instalacji /konfiguracji klastra

10.1.6. Odrzucanie zapytań z powodu zbyt dużego obciążenia

10.2. Metody zabezpieczania klastra

10.3. Najlepsze metody aktualizacji oprogramowania

10.4. Najważniejsze narzędzia przydatne w codziennej pracy

10.4.1. Head

10.4.2. Kopf

10.4.3. Sense

10.4.4. Marvel

10.4.5. Curator

10.5. Monitorowanie klastra