ОНЛАЙН-ОБРАЗОВАНИЕ

Не забыть включить запись!

Меня хорошо слышно && видно?

Напишите в чат, если есть проблемы!

Ставьте + если все хорошо

Правила вебинара

Активно участвуем

Задаем вопрос в чат или голосом

slack

Off-topic обсуждаем в Slack #канал группы или #general

Вопросы вижу в чате, могу ответить не сразу

OTUS

Занятие 15. Elasticsearch.

Задорожный Антон

Цели вебинара

После занятия вы сможете:

1. познакомиться с основными компонентами стека ELK
2. разрабатывать простые запросы для Elasticsearch
3. использовать Elasticsearch для основных кейсов

Краткое содержание

- Введение в Elasticsearch и ELK стек
- Модель данных и базовый API
- Kibana
- Logstash
- Применимость / кейсы

Введение в Elasticsearch и ELK

Elasticsearch и стек ELK

- Elasticsearch, Java/Javascript, 2010
- Распределенная система поиска / документарная NoSQL СУБД
- Сложная лицензия

Часто используется со следующими компонентами:

- Logstash – средство доставки данных (data pipelines)
- Kibana – среда визуализации

Модель данных и базовый API

Модель данных и базовый API

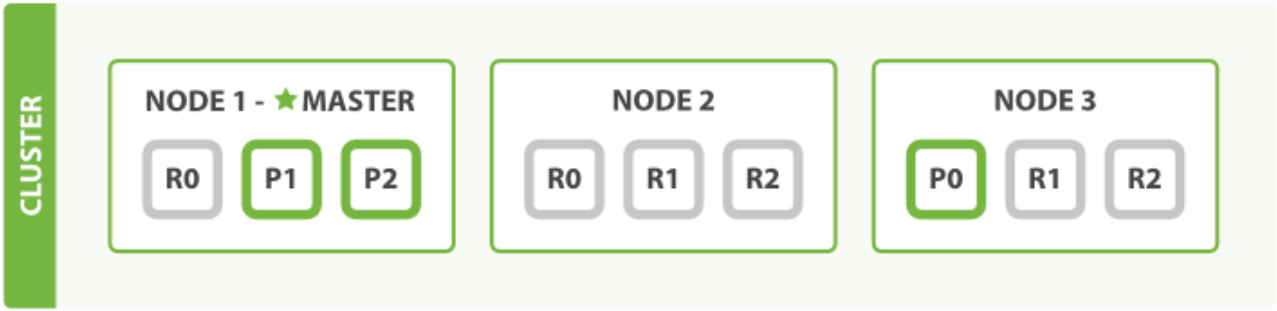
- Elasticsearch использует свою терминологию
- База данных -> Индекс (Index)
- Таблица -> Тип (Type)
- Строка -> Документ (Document)
- Колонка -> Поле (Field)

Два API для поиска:

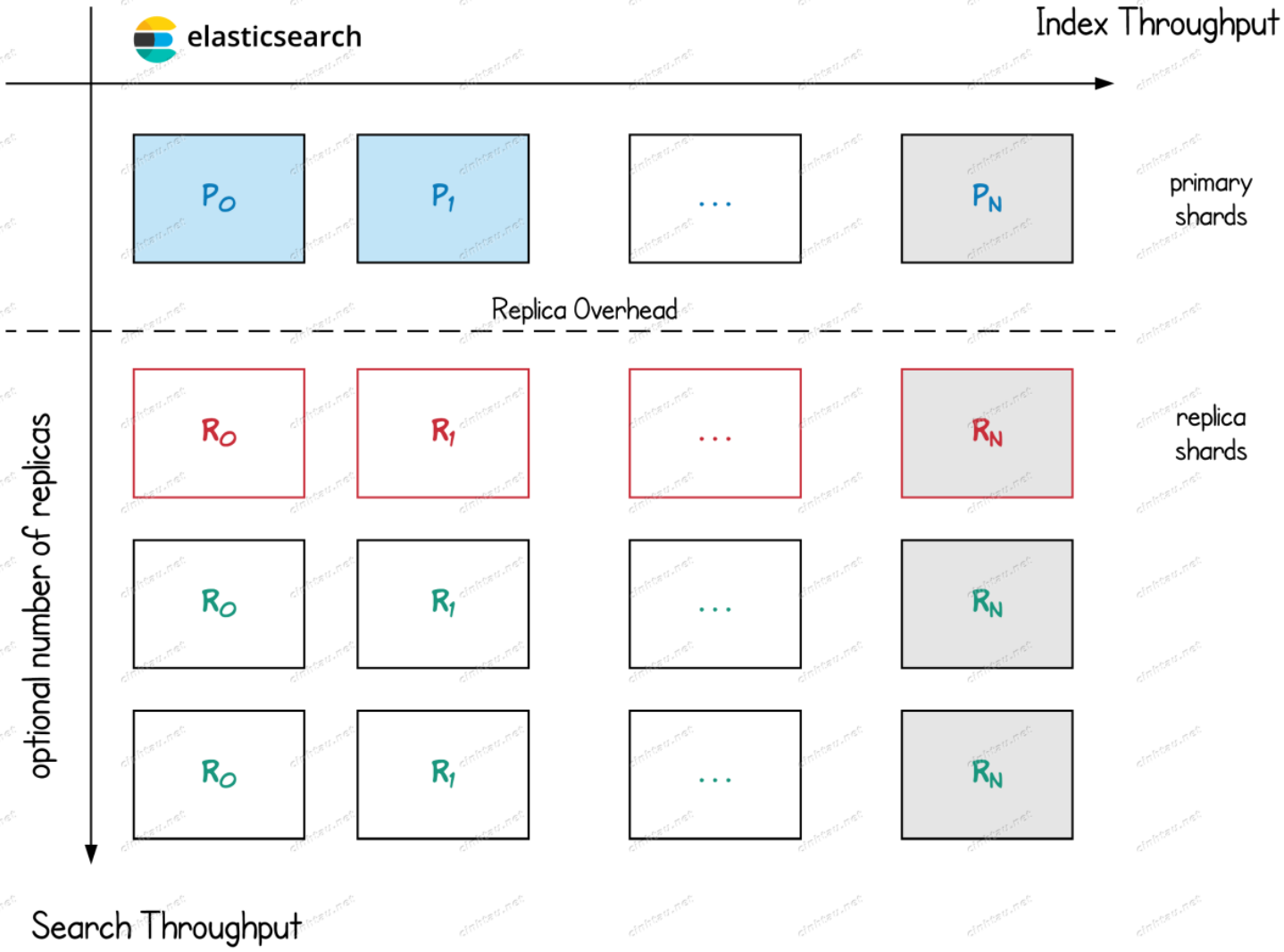
- Query-string: /index/type/_search?q=somefield:SomeValue
- DSL

Архитектура

- Распределенное хранение данных
- Данные делятся по хэшу от ключа _id между шардами на уровне индекса
- Шарды реплицируются между нодами, запись всегда идет через primary shard, чтение может идти и из replica shard
- Каждая шарда это один индекс lucene



Баланс производительности



Kibana

Kibana

- Прекрасный (лучший в классе) компонент визуализации
- Плотно связан с Elasticsearch, для других поисковых движков есть аналоги (Banana для Apache Solr)

Предоставляет следующие компоненты:

- Visualize – средства формирования запросов к Elasticsearch (аналог Tableau)
- Timelion – средство для разработки графиков
- Dashboard – средство разработки дашбордов (из визуализаций)
- Dev Tools – удобное интерактивное средство для отладки средств

Logstash

Logstash

- Средство поставки логов и других событий
- Изначально независимое (и медленное), сейчас развивается Elastic (и быстрое)
- Аналог Apache Flume, но еще больше заточено под логи

Logstash конфигурация

- Input – источник данных: file, syslog, nc, redis
- Filter – фильтр: grok, drop, geoip
- Output – приемник данных: elasticsearch, file, kafka
- Codec – сериализатор: plain, json, msgpack

Logstash пример

```
input {
  file {
    path => "/tmp/access_log»
    start_position => "beginning»
 }
}
filter {
  if [path] =~ "access" {
    mutate { replace => { "type" => "apache_access" } }
    grok {
      match => { "message" => "%{COMBINEDAPACHELOG}" }
    }
  }
  date {
    match => [ "timestamp" , "dd/MMM/yyyy:HH:mm:ss Z" ]
  }
}
output {
  elasticsearch {
    hosts => ["localhost:9200"]
  }
}
```

Применимость / кейсы

Применимость и кейсы

- Поиск
- ELK как система работы с логами / инфраструктурный компонент
- Операционное хранилище для machine learning

Поиск

- Дополнительное хранилище для актуальных документов или других данных
- Чаще всего не стоит хранить все данные
- Также часто не стоит индексировать все данные что хранятся в ES
- Процесс реиндексации должен быть стандартным и поставляться с первым релизом

ELK как система работы с логами

- В какой-то момент стало стандартом де-факто, особенно для on-premise решений
- Активно вытесняет Splunk
- Может комбинироваться для различных областей мониторинга:
- Инфраструктура и приложения
- Поддержка пользователей
- Бизнес-показатели

Операционное хранилище для machine learning

- Хранение векторов предикторов / фич
- Хранение моделей
- Интеграция с Apache Spark

Следующий вебинар

Тема: Apache Spark – 1 часть

Среда 2019.08.05 в 20.00

Ссылка на вебинар будет в ЛК за 15 минут

Заполните, пожалуйста, опрос о занятии

Спасибо
за внимание!

ОНЛАЙН-ОБРАЗОВАНИЕ