

**adaptTo()**

APACHE SLING & FRIENDS TECH MEETUP  
BERLIN, 25-27 SEPTEMBER 2017

Only POC

Continuous content updates with Kafka  
Jörg Hoh, Adobe Systems

# Motivation

- Lots of usecases require integration of content-based solutions with (message-based) backend-systems.
- Often „content delivery“ (a.k.a your website) is considered the end of the „data“ pipeline.
  - It's not!

# Typical approaches for data exchange

## Imports

- Periodic pull from backend system(s) and parse the data

## Exports

- Provide the data via a REST-like interface and let the other(s) pull and parse it.

# Apache Kafka?

- Real-time data streaming pipelines
- Highly scalable and production proven
- Based on the publish-subscribe pattern





# Data ingestion with Kafka

- Realtime, asynchronous
- Typed data (using Avro, Kryo, Jackson, ...)
- Just read data from a single queue
  - All sources push to this queue

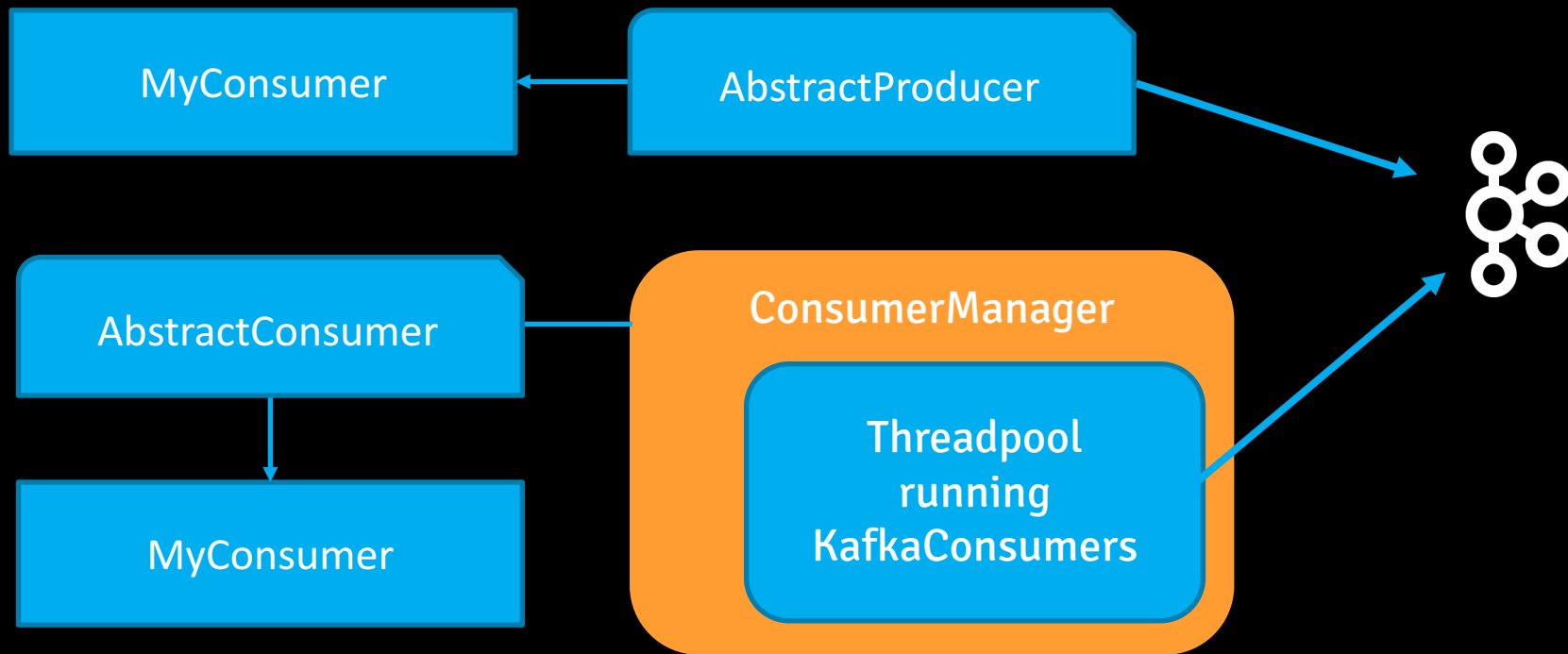


# Feeding back data

Write data to a queue and let one or more consumers read it

- Logs
- Audit information
- Content, UGC, form submissions, ...
- (AEM replication?)

# The design



# A simple Consumer

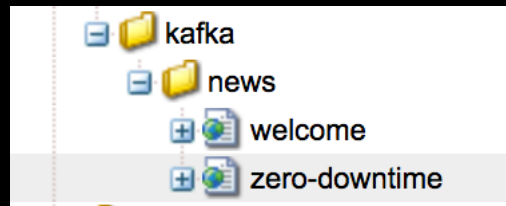
```
@Component(label="News importer",...)
@Service()
public class Importer extends AbstractConsumer<String, String> {

    public void handleRecords(ConsumerRecords<String, String> records) {
        Iterator<ConsumerRecord<String, String>> iter = records.iterator();
        while (iter.hasNext()) {
            ConsumerRecord<String,String> record = iter.next();
            String payload = record.value();
            ...
        }
        ...
    }
}
```

# Demo: Producer

```
{ "newsId" : "welcome",  
  "headline" : "Welcome to AdaptTo()",  
  "text" : "Lorem ipsum",  
  "author" : "joerg"  
}
```

```
cat data1.json | ~/kafka/kafka_2.11-0.10.2.0/bin/kafka-console-producer.sh  
--broker-list localhost:9092 -topic testdata
```



# Demo: Consumer

```
bin/kafka-console-consumer.sh --bootstrap-server localhost:9092 --topic page-audit
```

```
{"type":["PageDeleted"],"page":["/content/kafka/news/2"],"user":["admin"]}  
{"type":["PageDeleted"],"page":["/content/kafka/news/1"],"user":["admin"]}  
{"type":["PageCreated"],"page":["/content/kafka/news/welcome"],"user":["admin"]}  
{"type":["PageCreated"],"page":["/content/kafka/news/zero-downtime"],"user":["admin"]}
```



# Rethink your architecture!

- Even if AEM is only serving content, it's not disconnected from other systems.
- Having the chance to easily transfer data from and to other systems will change your architecture!

- Twitter: @joerghoh
- Sample code:  
<https://github.com/joerghoh/kafka-integration>