

Session 09 Assignment Project Exam Help

FIT5202 Big Data Processing

<https://powcoder.com>

Add WeChat powcoder

Data Streaming using Apache Kafka and Spark

Session 09 Agenda

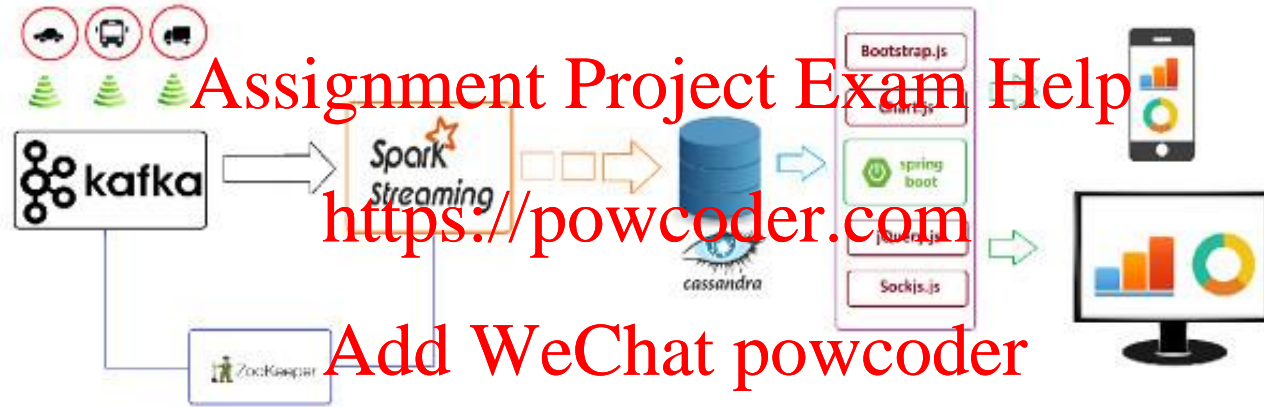
- Session 08 Review
 - Implicit vs Explicit Data
 - Matrix Factorization
 - Collaborative Filtering with ALS
- Spark Streaming Basics
 - Demo : word count
 - **Lab Task : Click Stream Analysis and Visualization**
- Streaming using Apache Kafka
 - Kafka Producer
 - Kafka Consumer
 - Visualizing in real-time
 - **Use case : Click stream visualization**

Assignment Project Exam Help

<https://powcoder.com>

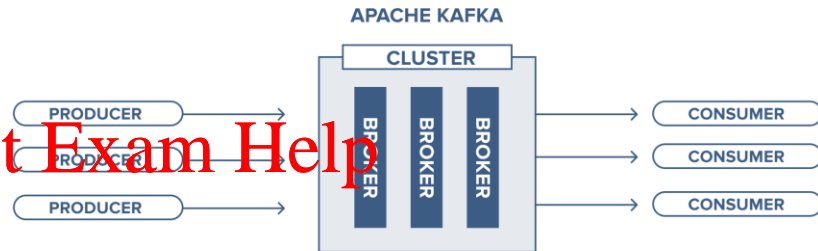
Add WeChat powcoder

Kafka Use Case (Traffic Data Monitoring)



What is Apache Kafka?

- Publish-subscribe messaging system
- Enables distributed applications
- Brokers utilize Apache ZooKeeper for management and coordination of the cluster
- Each broker instance is capable of handling read and write quantities reaching to the hundreds of thousands each second (and terabytes of messages) without any impact on performance.

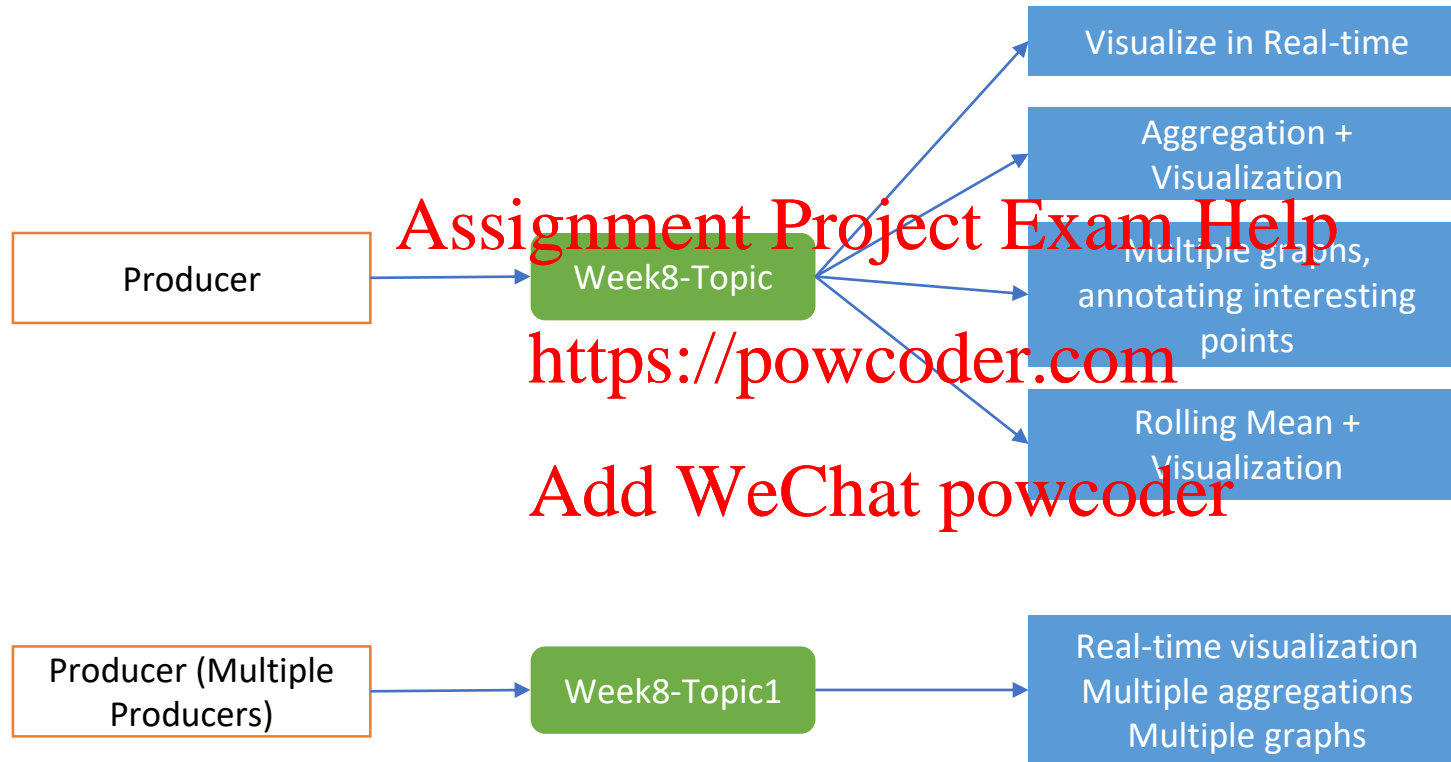


Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

DEMO Kafka Implementation Scenarios for Lab



Assignment Project Exam Help
<https://powcoder.com>
Add WeChat powcoder

Kafka Producer and Consumer Properties

▪ KafkaProducer

- Bootstrap_servers
- Value_serializer
- Api_version

▪ KafkaConsumer

- Consumer_timeout_ms
- Auto_offset_reset
- Bootstrap_servers
- Value_deserializer
- Api_version

Assignment Project Exam Help

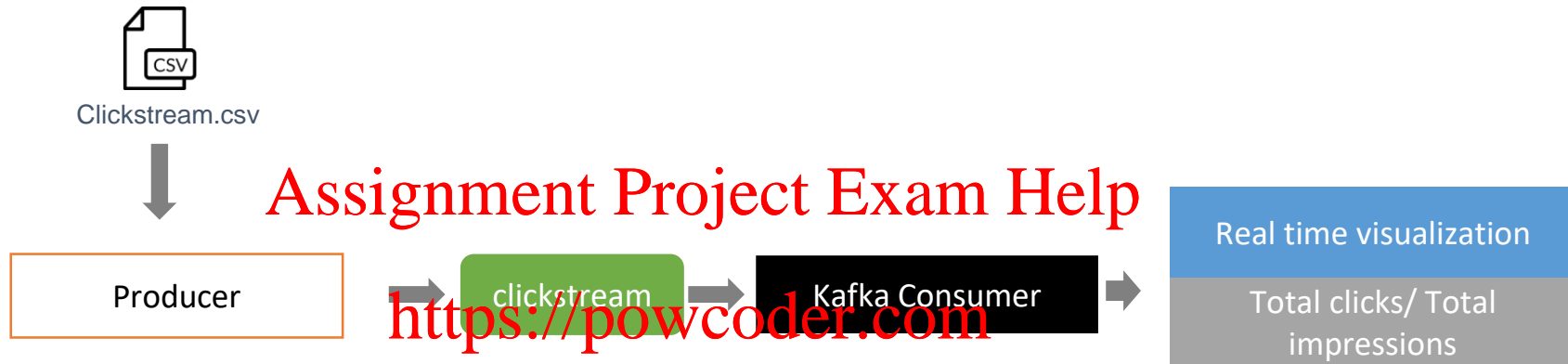
<https://powcoder.com>

Add WeChat powcoder

```
_producer = KafkaProducer(bootstrap_servers=['localhost:9092'],  
                           value_serializer=lambda v: json.dumps(v).encode('ascii'),  
                           api_version=(0, 10))
```

```
_consumer = KafkaConsumer(topic,  
                           consumer_timeout_ms=10000, # stop iteration if no message aft  
                           auto_offset_reset='latest', # comment this if you don't want  
                           bootstrap_servers=['localhost:9092'],  
                           value_deserializer=lambda x: loads(x.decode('ascii')),  
                           api_version=(0, 10))
```

Lab Task for Kafka



Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Spark Structured Streaming

To be covered in Session 10 Lecture

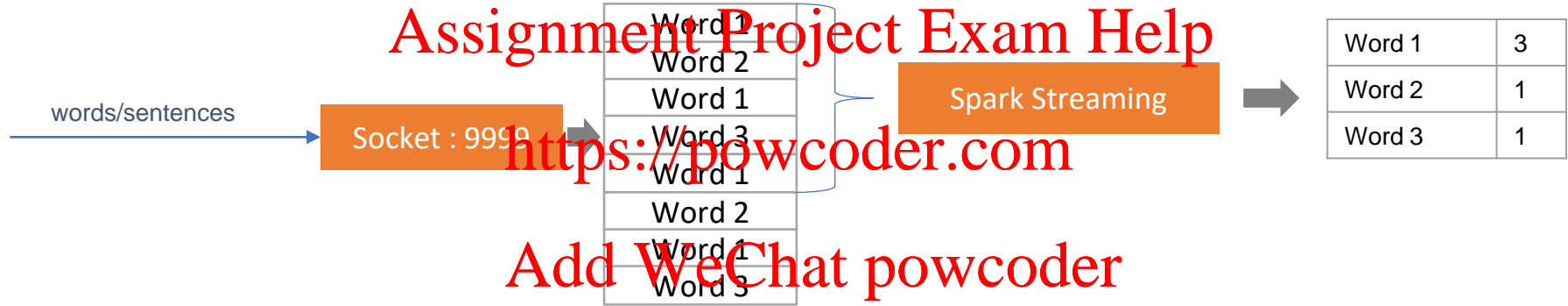
Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

DEMO Spark Structured Streaming

Word Count Demo



Lab Task for Spark Structured Streaming



Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Thank You!

See you next week.

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder