Pontifical Catholic University of Rio Grande do Sul
Graduate Program in Computer Science

Data Stream Processing with Apache Kafka and Spark Structured Streaming

Carlos Henrique Kayser
Email: carlos.kayser@edu.pucrs.br

Scalable Data Stream Processing
Prof. Dr. Dalvan Jair Griebler

May 14th, 2022

Apache Kafka

Apache Kafka Topic

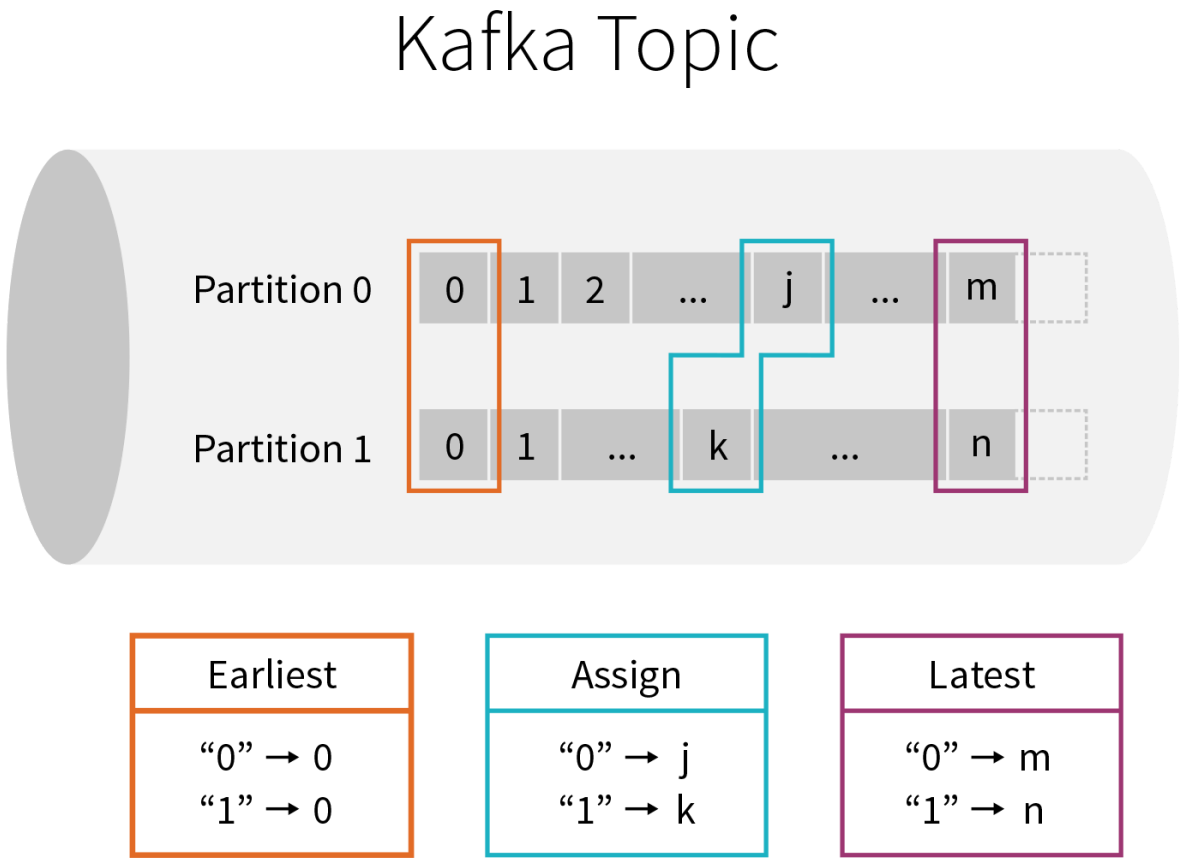
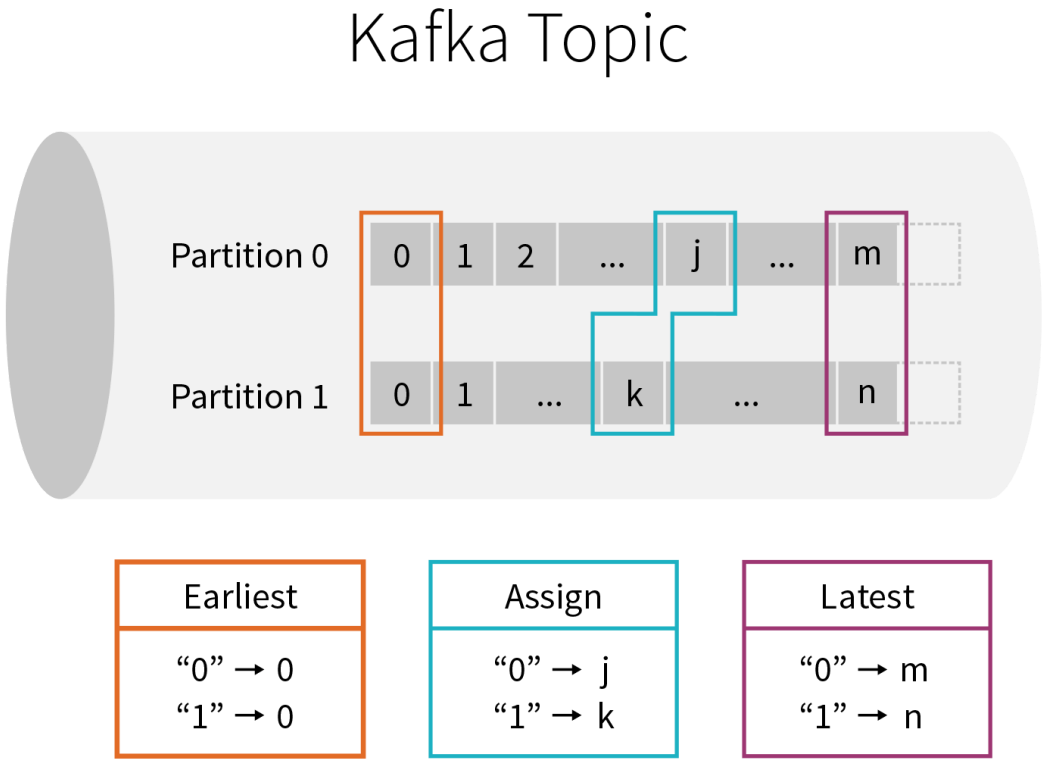
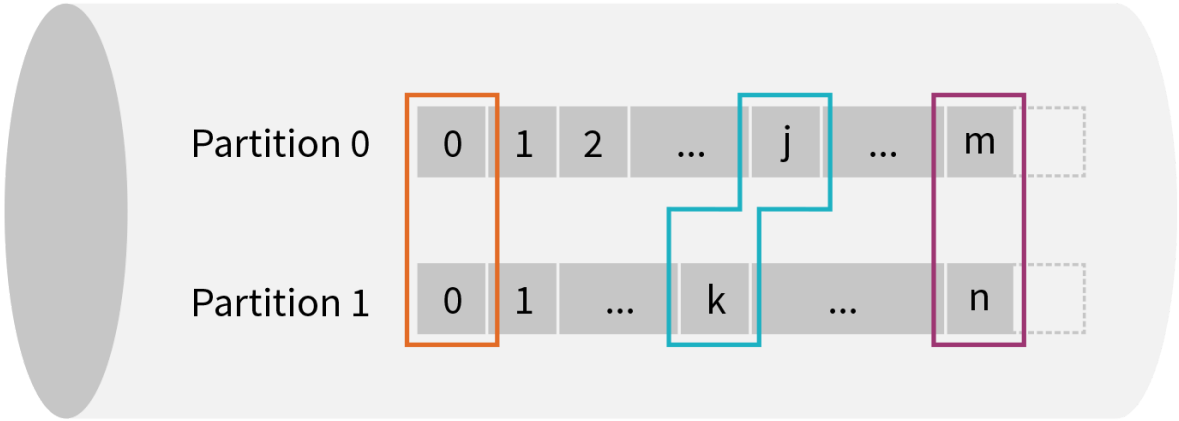


Fig.1 - 4K Mountains Wallpaper


Kafka Topic
Partition 0
0
1
2
...
j
...
m
Partition 1
0
1
...
k
...
n
Earliest
“0” → 0
“1” → 0
Assign
“0” → j
“1” → k
Latest
“0” → m
“1” → n

Kafka Topic

Kafka Topic



Earliest	Assign	Latest
“0” → 0 “1” → 0	“0” → j “1” → k	“0” → m “1” → n

Kafka Topic

Apache Spark Structured Streaming

Apache Spark Structured Streaming

- Starting Streaming Queries
  - Details of the output sink
  - Output mode
  - Trigger interval
  - Checkpoint location

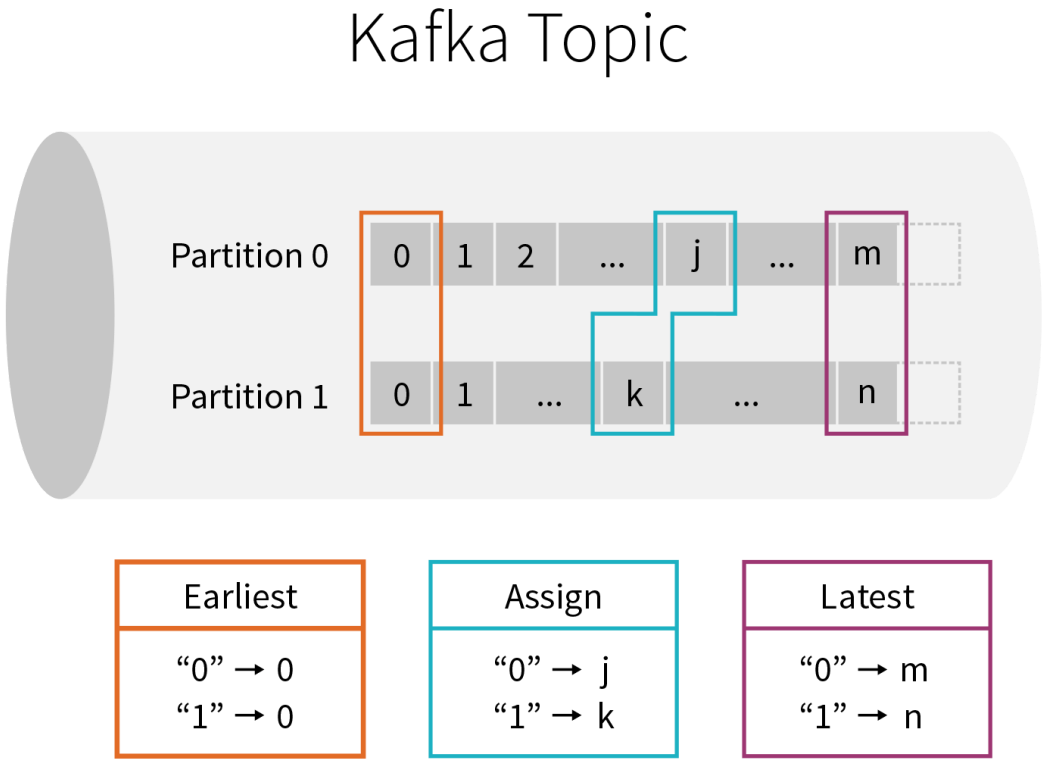


Fig.1 - 4K Mountains Wallpaper

Render inline math such as $ax^2 + bc + c$.

$$I_{xx} = \int \int_R y^2 f(x, y) \cdot dydx$$

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(\xi)\, e^{2\pi i \xi x}\, d\xi$$

I am a slide

```
query = results_kafka.writeStream \
    .format("kafka") \
    .option("kafka.bootstrap.servers", "localhost:9092") \
    .option("topic", "predictions") \
    .option("checkpointLocation", "/home/kayser/temp") \
    .outputMode("Append") \
    .start()

time.sleep(30)

query.stop()
```

Document Title

The usual Markdown Cheatsheet
does not cover some of the more advanced Markdown tricks, but here
is one. You can combine verbatim HTML with your Markdown.
This is particularly useful for tables.
Notice that with **empty separating lines** we can use Markdown inside HTML:

Json 1	Markdown

<pre>{
  "id": 1,
  "username": "joe",
  "email": "joe@example.com",
  "order_id": "3544fc0"
}</pre>
	<pre>{
	  "id": 5,
	  "username": "mary",
	  "email": "mary@example.com",
	  "order_id": "f7177da"
	}</pre>

I am slide

```
// I am code block
```

I am slide

```
// I am code block
```

Slide 1 title

Some super quickly created demo slides

- Do not need anything else than markdown
  - Slides title starts with # (also starts a new slide)
  - Bullet points, newlines, empty lines: all standard markdown
- However, can also use other stuff, e.g.:
  - Some HTML (e.g.)
  - When using pandoc beamer, can use latex commands (e.g. \center, \large, etc)\dots

Slide 2 title

\center The slide syntax is so simple that you can quickly create a handful of slides on basically any device in any editor. E.g. on your mobile on the way to the meeting where you need the slides. Right before the meeting starts you use pandoc to create the actual slides from your source.