

Streaming HDInsight Kafka Data into Azure Databricks



Michael Bender

AUTHOR EVANGELIST, PLURALSIGHT

@MichaelBender



Overview



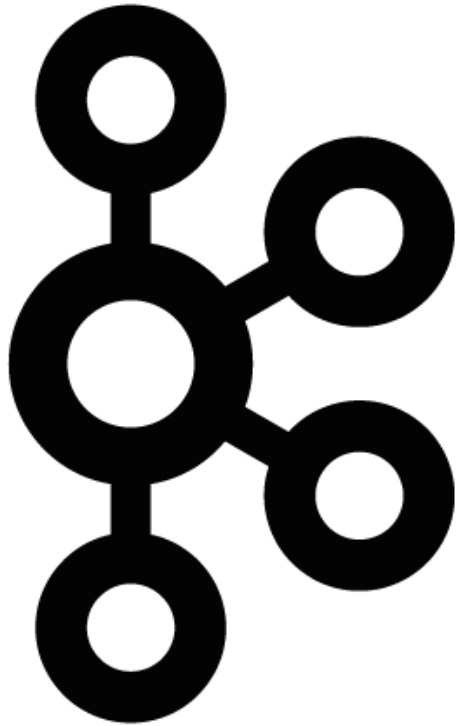
Apache Kafka on HDInsight

Streaming social media data

Azure Databricks streaming



Apache Kafka on Azure HDInsight



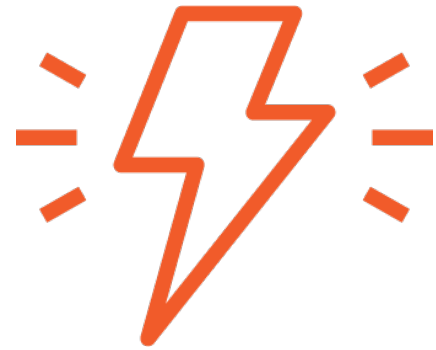
Event ingestion platform

Single source for consumers

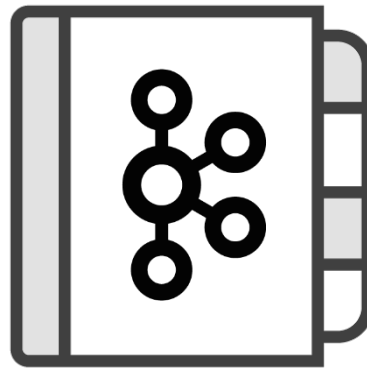
Azure HDInsight



Streaming with Kafka



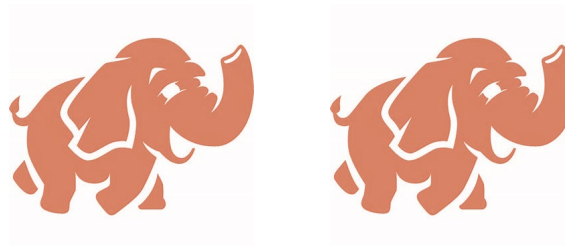
Event Producers



Topic



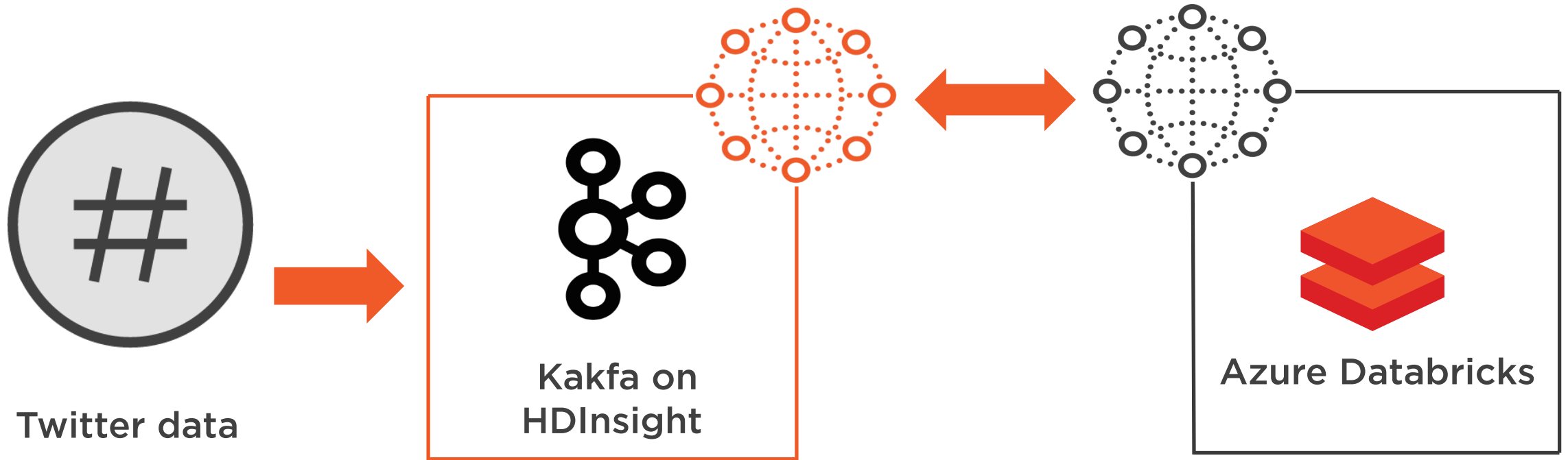
Event Consumer



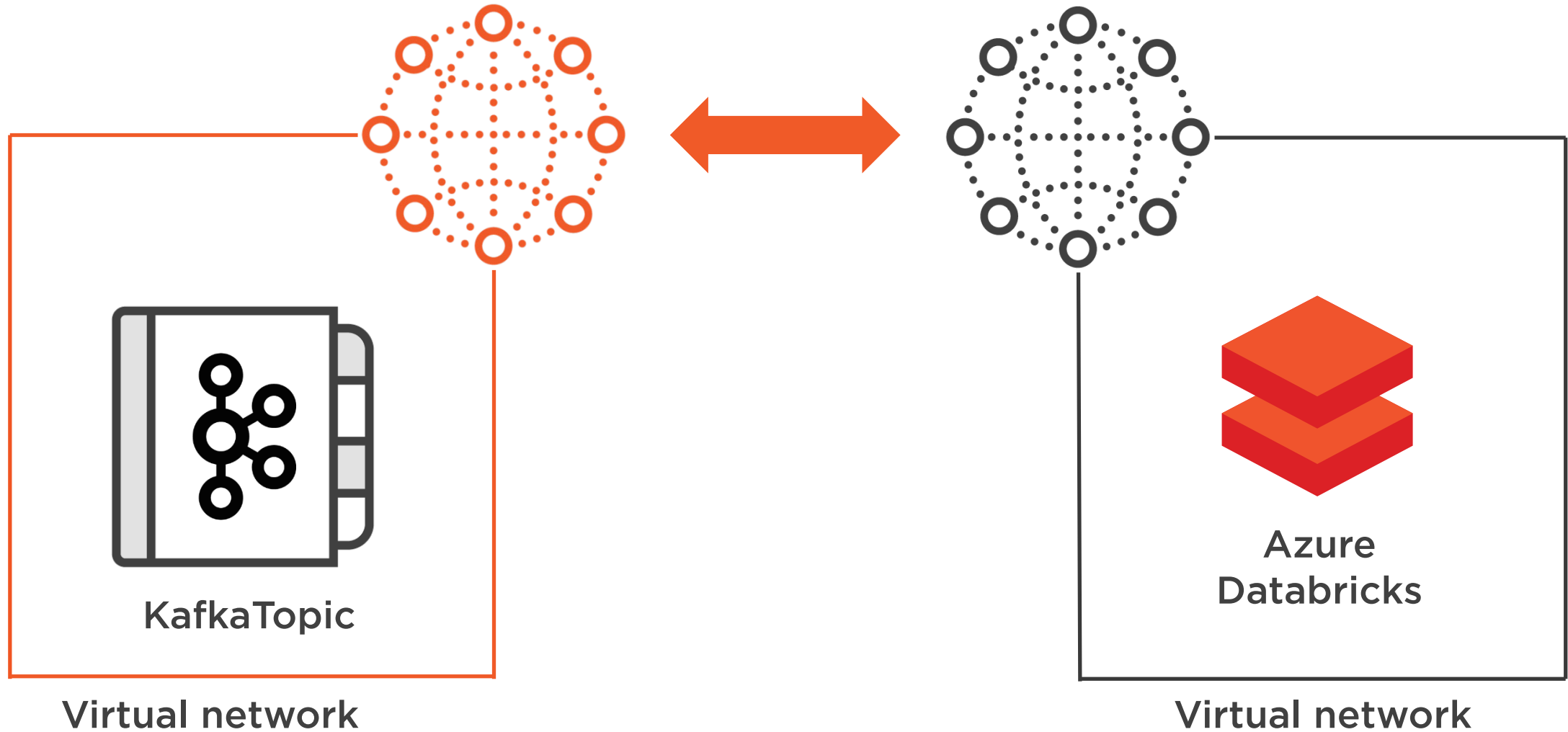
HDInsight Kafka
cluster and brokers



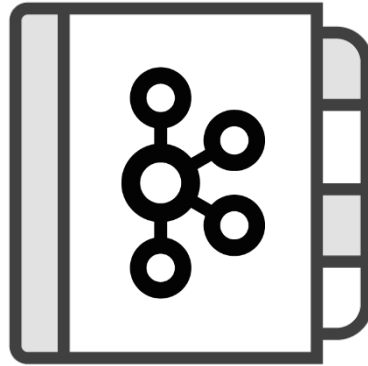
Streaming Twitter Data



Virtual Network Peering



Azure Databricks and Streaming Data



KafkaTopic



Producer
notebook



Consumer
notebook



Summary



Flexible and scalable streaming

Topics and brokers work both sides

Notebooks are key



For Further Learning

Azure Databricks Documentation at docs.microsoft.com
<https://docs.microsoft.com/en-us/azure/azure-databricks/>

Azure Databricks Documentation at
docs.azuredatabricks.net
<https://docs.azuredatabricks.net/user-guide/index.html>

*Specific links located in module exercise files on Pluralsight.com



Next Up:
Building Batch Data Processing Solutions
in Microsoft Azure
by Tim Warner

<http://bit.ly/timwbatch>

