🚀 Real-Time Data Streaming Workshop

Master Apache Kafka & Databricks Structured Streaming

Transform your understanding of real-time data through hands-on challenges with practical, beginner-friendly tutorials

🎯 What You'll Build

By the end of this workshop, you will have created:

- **Live data producers** that generate e-commerce events in real-time
- **Kafka topics** to reliably stream millions of messages
- **Python consumers** to process streaming data
- **Databricks pipelines** that transform streams into analytics-ready tables
- **Delta Lake tables** with ACID transactions and time travel capabilities

🌟 Why This Workshop Stands Out

Interactive Learning Platform

Our custom-built workshop application provides:

- **Step-by-step tutorials** with code examples you can copy and paste
- **Live data generation** - create realistic e-commerce events on demand
- **Progress tracking** - gamified learning experience with achievements
- **Instant feedback** - test your Kafka setup in real-time
- **Challenge mode** - test your knowledge with hands-on exercises

Real-World Scenarios

Learn using realistic examples:

🛒 E-commerce

Order processing and product analytics

👥 User Activity

Tracking and engagement metrics

📊 Analytics

Real-time dashboard updates

🔔 Notifications

Event-driven messaging

📚 Workshop Structure

Part 1: Apache Kafka Fundamentals (90 minutes)

Master the world's most popular event streaming platform:

- **Core Concepts:** Topics, Partitions, Producers, Consumers
- **Architecture:** Brokers, Clusters, and how Kafka scales
- **Security:** SASL/SSL authentication with Confluent Cloud
- **Python Integration:** Building producers and consumers
- **Consumer Groups:** Load balancing and fault tolerance

```
# You'll write code like this: producer.send('ecommerce-events', { 'user_id': 'user_123', 'action': 'purchase', 'product': {'id': 'prod_456', 'price': 99.99} })
```

Part 2: Databricks Structured Streaming (90 minutes)

Process infinite data streams with Apache Spark:

- **Streaming DataFrames:** Process data as it arrives
- **Transformations:** Parse JSON, filter, aggregate in real-time
- **Delta Lake:** ACID transactions for streaming data
- **Unity Catalog:** Organize data with modern governance
- **Fault Tolerance:** Checkpointing and exactly-once processing

```
# You'll build streaming pipelines like this: spark.readStream \ .format("kafka") \ .option("subscribe", "ecommerce-events") \ .load()
```

```
\ .writeStream \ .format("delta")
\ .table("ecommerce_analytics")
```

Part 3: Code Examples & Best Practices (60 minutes)

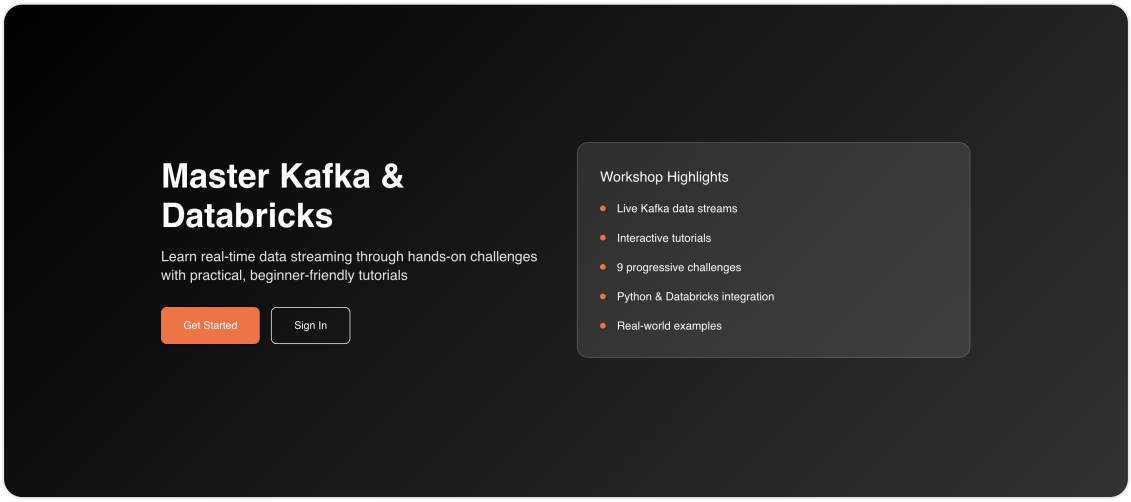
- Complete examples for common use cases
- Production patterns: Error handling, monitoring, optimization
- Testing strategies: Validating streaming applications
- Performance tuning: Throughput and latency optimization

🛠️ What You'll Use

Technology	Purpose	You'll Learn
Apache Kafka	Event Streaming	Producers, consumers, topics
Confluent Cloud	Managed Kafka	Cloud-native streaming
Python	Client Development	kafka-python library
Databricks	Spark Processing	Structured Streaming
Delta Lake	Data Storage	ACID transactions
Unity Catalog	Data Governance	Modern data organization

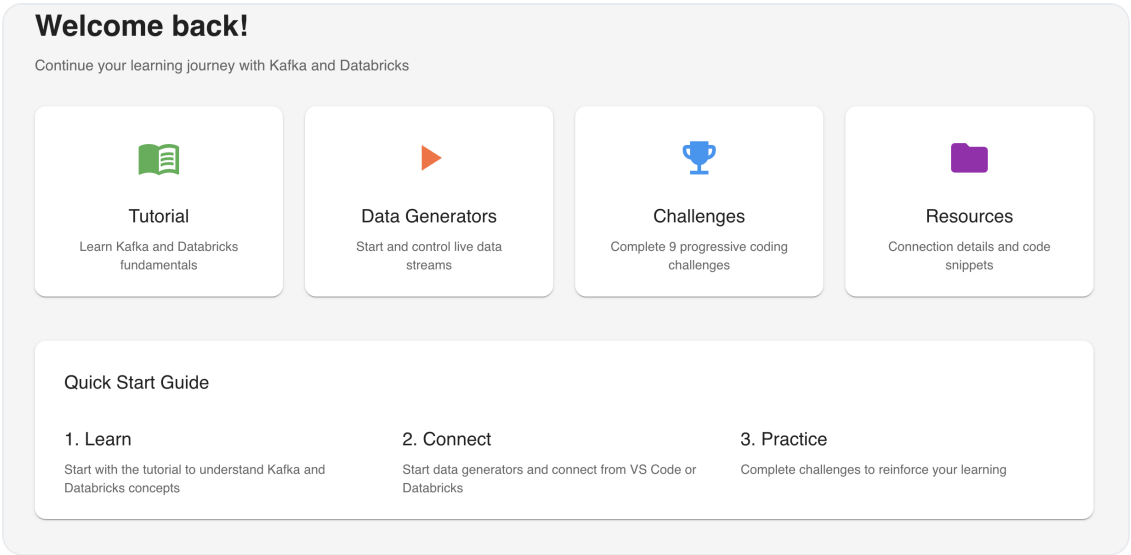
Workshop Platform Preview

1. Welcome Home



Master Kafka & Databricks - Your journey begins here! Our sleek landing page welcomes you with workshop highlights including live Kafka data streams, interactive tutorials, 9 progressive challenges, Python & Databricks integration, and real-world examples.

2. Dashboard Overview

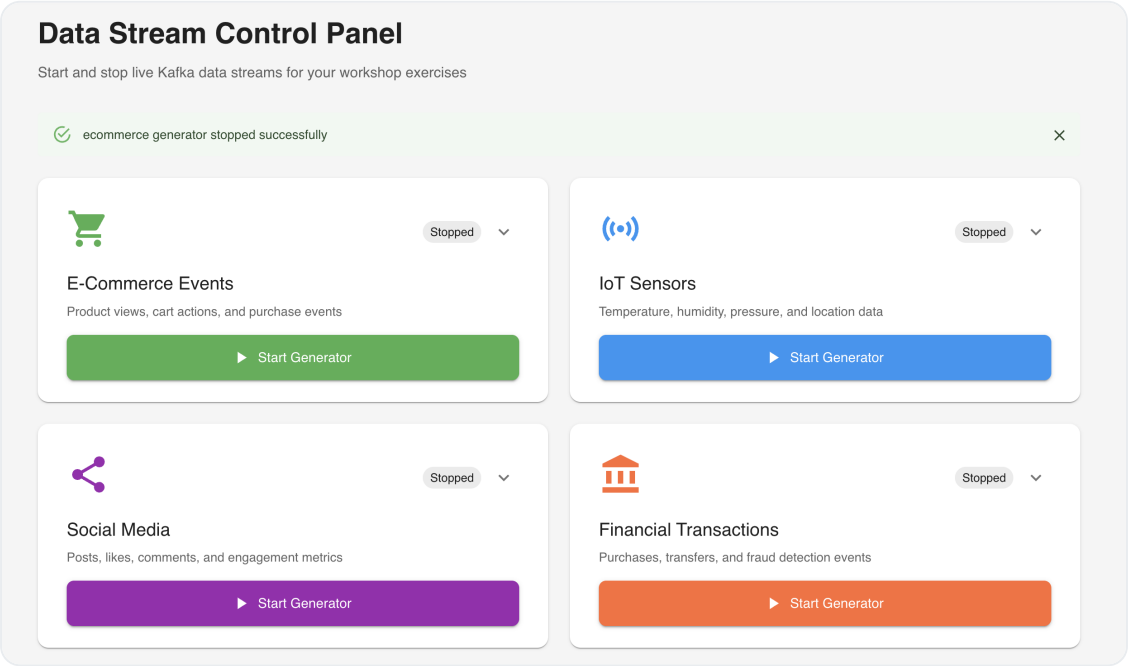


Welcome back! The dashboard features 4 main sections: Tutorial, Data Generators, Challenges, and Resources. Our Quick Start Guide provides a clear 3-step path: Learn, Connect, and Practice.

3. Interactive Tutorials

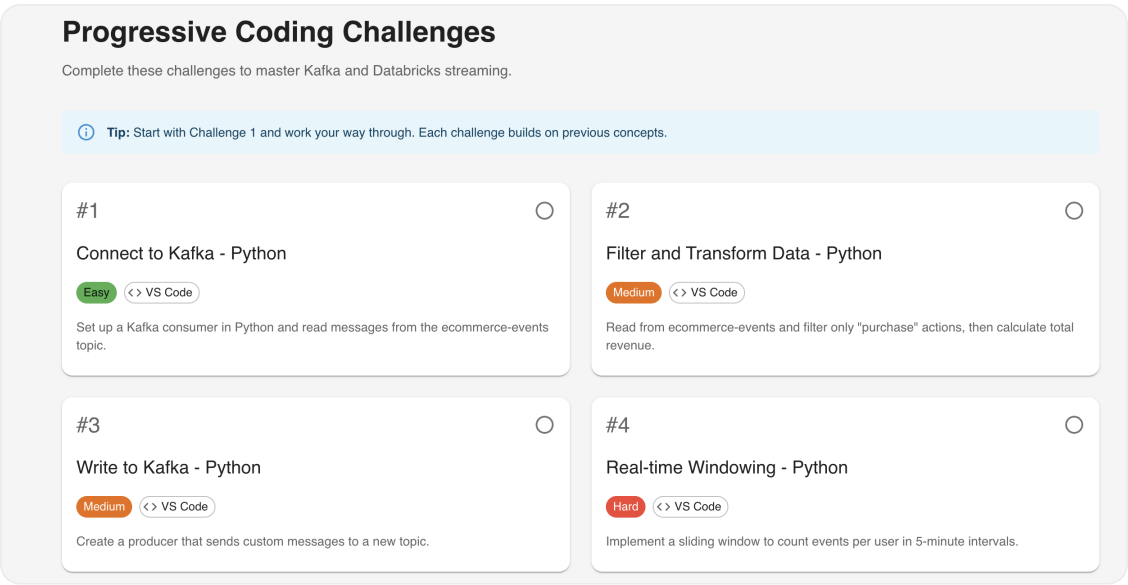
Kafka & Databricks Tutorial - Your comprehensive learning companion with progress tracking, structured content covering both Apache Kafka and Databricks, and a complete setup checklist. Every section includes syntax-highlighted code examples you can copy directly.

4. Live Data Generator



Data Stream Control Panel - Your streaming data laboratory with four realistic data generators: E-Commerce Events, IoT Sensors, Social Media, and Financial Transactions. Each generator has status indicators and easy start/stop controls.

5. Challenge Yourself



Progressive Coding Challenges - Test your mastery step by step with hands-on challenges ranging from connecting to Kafka, filtering and transforming data, writing to Kafka, and implementing real-time windowing. Each challenge shows difficulty level and clear objectives.

📋 Prerequisites

Required Knowledge

- Basic Python programming
- Understanding of JSON data format
- Familiarity with command line/terminal
- Basic SQL knowledge (helpful but not required)

Required Setup (We'll help you configure these)

- Python 3.8+ installed
- Confluent Cloud account (free tier available)
- Databricks Community Edition account (free)
- Internet connection

> **Good news!** You don't need to pay for cloud services. We use free tiers: Confluent Cloud free trial + Databricks Community Edition.

👥 Who Should Attend?

Perfect For:

- **Data Engineers** wanting to master streaming
- **Software Engineers** building real-time features
- **Data Analysts** ready to work with live data
- **Data Scientists** needing real-time ML pipelines
- **Students** exploring modern data engineering

You'll Succeed If You:

- Love hands-on learning (we code together!)
- Want to build real-world projects
- Are curious about how platforms like Uber, Netflix, and LinkedIn process data
- Enjoy problem-solving and debugging
- Are ready to ask questions and experiment

🎁 What You'll Take Home

Skills

- Ability to design and implement real-time data pipelines
- Hands-on experience with industry-standard tools
- Understanding of streaming vs. batch processing
- Production-ready code patterns and best practices

Resources

- **Complete code examples** - Production-ready templates
- **Comprehensive tutorial** - Available online forever
- **Reference materials** - Kafka and Spark documentation
- **Workshop platform access** - Practice anytime

> 🏅 **Certificate (Optional):** Complete the challenge section for a certificate of completion

🎓 Learning Outcomes

By completing this workshop, you will be able to:

- **Explain** the architecture and use cases for Apache Kafka
- **Design** scalable event-driven systems
- **Implement** Kafka producers and consumers in Python
- **Build** real-time data pipelines with Databricks
- **Process** streaming data with Spark Structured Streaming
- **Store** streaming results in Delta Lake with ACID guarantees
- **Troubleshoot** common Kafka and Spark streaming issues
- **Apply** best practices for production deployments

🚀 Ready to Start?

Before the Workshop:

- Check your email for workshop access link
- Create Confluent Cloud account (instructions provided)
- Sign up for Databricks Community Edition
- Install Python 3.8+ on your machine
- Visit the workshop platform to familiarize yourself

Workshop Day:

- Arrive 15 minutes early for setup help
- Bring a laptop with 4GB+ RAM
- Ensure stable internet connection

- Have a notepad for questions
- Bring water/coffee - it's an intensive session!

💬 Common Questions

Q: Is this workshop beginner-friendly?

A: Yes! If you know basic Python, you're ready. We explain everything step-by-step.

Q: Do I need to pay for cloud services?

A: No! We use free tiers: Confluent Cloud free trial + Databricks Community Edition.

Q: What if I get stuck during the workshop?

A: Instructors are available throughout. Plus, our platform has troubleshooting guides.

Q: Can I access the materials after the workshop?

A: Absolutely! The platform and all tutorials remain available online.

Q: Will there be breaks?

A: Yes! We have scheduled breaks and encourage asking questions anytime.

🌟 Join the Real-Time Revolution

The future of data is streaming. Companies like Uber, LinkedIn, Netflix, and Airbnb process billions of events per day using Kafka and Spark. This workshop gives you the exact skills they use.

Don't just learn about streaming data—build it!

This is not just a workshop—it's your gateway to mastering real-time data engineering.

📞 Contact & Support

Workshop Organizer: [Your Name/Organization]
Email: [contact@example.com]
Platform: [Workshop URL]

Technical Support: Available 24/7 via platform chat

See You at the Workshop! 🎉

#DataEngineering #ApacheKafka #Databricks #RealTime #Streaming #BigData