Apache Flink for Stream Processing (20 hrs)
By Dr. Vishwanath Rao

Module 1: Introduction to Apache Flink
- Overview of Stream Processing
  - Differences between batch and stream processing
  - Use cases and applications of Flink
- Introduction to Apache Flink
  - Architecture and key components
  - Flink ecosystem

Module 2: Flink Streaming Fundamentals
- Introduction to Flink Streaming
  - DataStream API vs. DataSet API
  - Creating and processing DataStreams
- Data Sources and Sinks
  - Reading data from sources (Kafka, files, etc.)
  - Writing data to sinks (databases, file systems)

Module 3: Windowing and Process Functions
- Understanding Windowing Concepts
  - Types of windows: Tumbling, Sliding, Session
  - Use cases for windowing
- Implementing Aggregations
  - Common aggregation functions
  - Custom aggregators
- Process Functions
  - Keyed and non-keyed process functions
  - Use cases and examples

Module 4: Watermarking and Handling Late-Coming Events
- Introduction to Watermarking
  - What are watermarks and their purpose?
  - Types of watermarks (periodic, punctuated)
- Handling Late Events
  - Strategies for dealing with late arrivals
  - Use cases for late-event handling

Module 5: State Management and Optimization
- Understanding State in Flink

- ○ Types of state: Keyed state vs. Operator state
  - ○ Use cases for stateful processing
- State Optimization Techniques
  - ○ Best practices for state management
  - ○ Memory management and tuning

Module 6: Observability and Monitoring
- Importance of Observability in Stream Processing
  - ○ Key metrics to monitor in Flink
- Tools and Techniques for Monitoring
  - ○ Using the Flink Web UI
  - ○ Integrating with external monitoring tools (Prometheus, Grafana)

Module 7: Checkpointing and Savepoints
- Understanding Checkpointing
  - ○ What are checkpoints and their purpose?
  - ○ Configuring checkpoints for fault tolerance
- Savepoints
  - ○ Difference between checkpoints and savepoints
  - ○ How to create and manage savepoints

Module 8: Connectors in Flink
- Overview of Flink Connectors
  - ○ Common connectors: Kafka, JDBC, Files, etc.
- Implementing Custom Connectors
  - ○ Creating and configuring custom connectors
  - ○ Best practices for connector development

Module 9: Parallelism in Flink
- Understanding Parallelism
  - ○ What is parallelism and why is it important?
  - ○ Configuring parallelism for tasks and operators
- Strategies for Optimizing Parallelism
  - ○ Load balancing and resource allocation
  - ○ Case studies of parallel processing

Module 10: Debugging and Metrics
- Debugging Flink Applications
  - ○ Common debugging techniques and tools
  - ○ Analyzing errors and exceptions

- Collecting and Analyzing Metrics
  - Instrumentation of Flink applications
  - Using metrics for performance tuning