



Advanced Cask Data Application Platform

The **Advanced Cask Data Application Platform** course will enable participants get deeper insights into CDAP building blocks and advanced features to build CDAP-optimized applications. This course is comprised of a combination of lectures and labs to reinforce the core **concepts**.

Audience & Prerequisites

This course is designed for developers and architects who are looking to use CDAP. Participants should have completed the Introduction to CDAP course as a prerequisite.

Materials Required

- Laptop with Mac, Windows or Linux on 64-bit OS.
- Virtual Box installed (<https://www.virtualbox.org/wiki/Downloads>)
- CDAP Standalone VM installed (Instructions provided separately)

Course Overview

- CDAP Architecture and internals
- Testing and Debugging Strategies
- Deeper understanding of transactions
- Metrics and Logging
- Preferences and Namespaces in CDAP
- Security in CDAP – Authentication, Keystore
- Security in CDAP – Authorization and Impersonation



Course Outline

<p>1. CDAP architecture deep dive Understanding</p> <ul style="list-style-type: none">• CDAP system components• Understanding data flow <p>2. Testing and Debugging</p> <ul style="list-style-type: none">• Unit testing CDAP applications• Integrating testing in CI• Debugging CDAP applications in clusters <p>3. Stream internals</p> <ul style="list-style-type: none">• Stream service design• A deeper look at Streamwriter and scaling <p>4. Advanced Datasets</p> <ul style="list-style-type: none">• Creating reusable data patterns with Custom Datasets• Dataset-Hive integration• Readless increments <p>5. Transactions deep-dive</p> <ul style="list-style-type: none">• The need for transactions• Optimistic concurrency control• Transaction manager• Tephra <p>6. Advanced topics in Tigon</p> <ul style="list-style-type: none">• Enabling exactly-once processing using transactions• Tigon partitioning strategies• Batching to optimize throughput• Generator flowlets <p>7. Metrics and logging</p> <ul style="list-style-type: none">• Metrics and logging programming APIs• Accessing metrics and logs using REST	<p>8. Preferences API</p> <ul style="list-style-type: none">• Introduction to preferences• Setting multi-level preferences• Scoped runtime arguments <p>9. Namespaces</p> <ul style="list-style-type: none">• Data and application isolation using Namespaces• Guaranteeing application resources in Namespaces <p>10. Security - Authentication</p> <ul style="list-style-type: none">• Introduction to CDAP security• Enabling perimeter security• Custom authentication <p>11. Security - Keystore</p> <ul style="list-style-type: none">• Encrypting configuration in secure keystore• Using keystore in programs• Keystore and Runtime arguments• Demo <p>12. Security - Authorization</p> <ul style="list-style-type: none">• Security extensions for authorization• Enabling authorization with Sentry• Demo <p>13. Security – Cross Namespace access and Impersonation</p> <ul style="list-style-type: none">• Enabling accessing datasets from different namespace• Secure impersonation• Demo
---	---