



StreamSets Data Collector Mentored Learning Workshop

Overview

The focus of Mentored Learning workshops is to Mentor/Train IT Professionals by transferring real-world knowledge and aiding in the Development Strategies in the advanced instruction of applicable technologies that are utilized within your environment by your in-house technology teams. Heavy concentration is placed on StreamSets Data Collector' Best Practices within Installation, Configuration, Design, Customization and Troubleshooting as it relates to your environment and industry standards.

Our sessions are facilitated by IT professionals with extensive real-world knowledge (15 year minimum consulting experience) and team driven personalities (5 year minimum training experience) that enable them to successfully transfer their knowledge and experience so that your internal organization can effectively manage and perform day to day tasks and mission critical projects.

The curriculum and mentored framework will include live exercises and demonstrations focusing on heavy participation with your internal staff. The exercises and demonstrations will allow the Mentor and the students to explore, test, and reinforce various theories/topics presented during the engagement. A list of set-up requirements will be sent to a contracted client at least 3 weeks prior to commencement.

Topics of Expertise for Training Framework

INSTALLATION

- Manual & Service Starts – Full Installation
- Stage Libraries
- Cloudera Manager
- Cloud Service Providers
- External Libraries
- SDC Edge

CONFIGURATION

- Data Collector
- Data Collector Environment
- Viewing Data Collector
- Pipeline – SSL/TLS
- Pipeline Expression
- Control Hub
- Guidelines for SDC RPC Pipelines
- Java options for Cli Command

DESIGN

- Pipeline Data Flow & Shortcut Keys
- Edge Properties & DC Receiving Pipelines

TROUBLESHOOTING

- Accessing Error Messages
- Pipeline & Data Preview
- Origins, Scripting Origins and Directory
- Hadoop FS
- Kafka Consumer
- Oracle & SQL Server CDC Client
- Salesforce
- Processors
- Encrypt and Decrypt Fields
- Destinations
- Cassandra and Hadoop FS
- HBase
- Kafka Producer
- MemSQL Fast Loader
- SDC RPC & Executors
- Hive Query
- JDBC Origins and Connections
- No Suitable Driver & DB Connections
- MySQL JDBC Driver and Time Values
- Performance & Cluster Execution Mode



PHIL-COSTANZO.COM

EDUCATION & MENTORING
SOCIAL MEDIA INTELLIGENCE
INTERNET INTELLIGENCE

StreamSets Data Collector Mentored Learning Workshop

CUSTOMIZATION

- Installation
- Configuration
- Upgrading
- Pipeline Concepts & Design
- Pipeline Configuration
- Data Formats
- Origins
- Processors
- Destinations
- Executions
- StreamSets Data Collector Edge
- StreamSets Control Hub
- Dataflow Triggers
- Drift Synchronization for Hive
- Drift Synchronization for PostgreSQL
- Multithreaded Pipelines
- Microserve Pipelines
- SDC RPC Pipelines
- Cluster Pipelines
- Data Preview
- Rules & Alerts
- Pipeline Monitoring & Maintenance