



# PROFESSIONAL SERVICES OFFERING ETL ONBOARDING QUICK START

ETL Onboarding Quick Start service will help you unlock the power of Hortonworks Data Platform (HDP®) in your organization to modernize the ETL process while reducing operational costs of your data warehouse. The ETL function is a relatively low-value computing workload that can be performed at a low cost in Apache Hadoop®. When onboarding to Hadoop, data is extracted, transformed and then the results are loaded into the data warehouse. The result: critical CPU cycles and storage space are freed for the truly high-value functions—analytics and operations—that best leverage advanced capabilities in the data architecture.

## HIGHLIGHTS

- Gain access to expertise and best practices from Hortonworks, the leader in Open Enterprise Hadoop
- Gain enterprise asset of functional Hortonworks Data Platform cluster
- Rapidly identify and build three ETL processing chains into HDP ecosystem
- Build resources and capabilities while creating a solution aimed at demonstrating success to senior leadership.

## SERVICE DESCRIPTION

Hortonworks to partner with members of customer's Data Source, HDP, and Business teams to manage data through the identification, extraction, transformation, loading, and validation phases of its lifecycle, resulting in data fully on-boarded and available within the customer's HDP cluster. Furthermore, Hortonworks will focus efforts within the engagement on training and mentoring of key customer resources to advance skills and capabilities.

## ENGAGEMENT DETAILS

Below is a list of key activities to be performed as part of the ETL Onboarding Quick Start project:

### HDP Cluster Quick Start

- Verify OS and file system configuration
- Install Ambari and HDP
- Run Benchmark Test and tune performance
- Setup NameNode HA
- Backup critical configurations
- Create operational runbook
- Knowledge transfer and mentoring
- Handoff to Hortonworks Support organization.

### Application Architecture and Design

- Conduct workshops
- Analyze existing workloads and requirements
- Assess existing data dictionaries
- Identify and develop global transformations
- Create ETL architecture and design.

### Application Development

- Acquire data from sources
- Develop data ingestion mechanisms
- Construct data cleansing routines
- Build Oozie processing flows
- Develop post-ingestions schemas for presentation
- Create metadata and lineage source to target maps.

### ETL Validation

- Execute count and sample validation

### Application Documentation

- Update data dictionary including Hive layouts, data flow diagrams, etc.
- Document code inventory
- Create data validation reports
- Create known issue and feature gap reports
- Create and document data ingestion templates.

### Training & Mentoring

- Deliver on-the-job training (OJT) and mentoring to customer resources
- Generate recommendations on training curriculum.

#### Deliverables

- Base HDP cluster
- Runbook defining installation and configuration of HDP
- Updated data dictionary
- Code inventory
- Data validation summary report
- Known issues and gaps inventory.

#### Project Assumptions

- Work will be performed on one (1) cluster
- Hortonworks will have direct access to critical resources
- HDP will run in a supported Linux environment
- Only Apache open source tools will be used
- Customer resources have completed core HDP training.

#### Customer tasks and dependencies

- Hardware and OS installation and configuration on 16+ node cluster are complete
- Dedicated resources will be provided for OJT for the duration of the project
- Data model and data dictionary will be provided for all in-scope data
- In-scope source teams are engaged and on-board before project initiation
- Security and compliance classification and management is addressed pre-engagement
- Identify and mitigate any pertinent policies or processes pre-engagement.

#### Recommended Customer Team

- Executive Sponsor
- Project Manager
- Systems Administrator
- ETL Developer.

#### Typical Hortonworks Team

- Platform Architect
- Applications Architect
- Engagement Manager.

#### Out of Scope

- Cluster security activities, including but not limited to:
  - Kerberos
  - Audit, authorization, and column restrictions
  - AD integration
- Streaming data ingestion
- ESB integration and/or alternate ETL tooling
- Data validation in excess of 95% accuracy
- Data validation issue mitigation on multi-byte character encoding issues
- Data validation issue mitigation on rounding variances
- Complex items including but limited to:
  - Stored procedures
  - Dynamic queries
  - Detailed data cleansing.

#### FOR MORE INFORMATION

This document does not constitute an offer. For more information about ETL Onboarding Quick Start or any other Hortonworks Professional Services offering, contact your Hortonworks representative or visit our website: [hortonworks.com/services](http://hortonworks.com/services)

## About Hortonworks

Hortonworks is a leading innovator at creating, distributing and supporting enterprise-ready open data platforms. Our mission is to manage the world's data. We have a single-minded focus on driving innovation in open source communities such as Apache Hadoop, NiFi, and Spark. Our open Connected Data Platforms power Modern Data Applications that deliver actionable intelligence from all data: data-in-motion and data-at-rest. Along with our 1600+ partners, we provide the expertise, training and services that allows our customers to unlock the transformational value of data across any line of business. We are Powering the Future of Data™.

#### Contact

For further information visit [hortonworks.com](http://hortonworks.com)

+1 408 675-0983  
+1 855 8-HORTON  
INTL: +44 (0) 20 3826 1405

