Services Included with Teradata IntelliCloud (Hadoop)

1	**Application Administration**: These set of services are geared to administer the file-system activities around the platform; this includes Hadoop Core including HDFS and YARN.
1.01	Maintain any scheduled cron jobs and administrative scripts on Hadoop NameNode, DataNodes (e.g., file system temporary area cleanup)
1.02	Review status of administrative cron-job and shell scripts (e.g., file system cleanup scripts, etc.)
1.03	Maintain Network Time Protocol (NTP) setting
1.04	Provide access to a Network Time Protocol server
1.05	Monitor Cron daemon service availability for scheduled tasks (e.g., clean temporary files and compress old logs)
1.06	Maintain postfix or sendmail settings for cron-based alerts
1.07	Provide access to Email routing gateway and enable firewall changes for sending out email monitoring alerts
1.08	Support system admin functions during maintenance windows
1.09	Coordinate with server maintenance team for planned and unplanned daemon / cluster startup or shut-down
1.10	Make available a dedicated DNS resolver for the Cluster
1.11	Maintain host file entries as needed on Hadoop Nodes
1.12	Maintain system-level log file, logging detail and log archive settings
1.13	Maintain daemon log files and log archives
1.14	Manage NameNode directory structure
1.15	Maintain Ambari/Cloudera Manager data collection processes, thresholds and alert settings (Maintain=maintenance, Configure=Install)
1.16	(If Applicable) Maintain Nagios alerts setting for Hadoop services
1.17	(If Applicable) Maintain oozie configuration settings
1.18	(If Applicable) Maintain Flume configuration settings
1.19	(If Applicable) Maintain WebHCat configuration settings
1.20	(If Applicable) Maintain WebHDFS configuration settings
1.21	Coordinate with Client Networking for Hadoop daemon communication and Web UI port maintenance
1.22	Maintain Data Integrity setting for cluster data replication
1.23	Run data node block scanner on periodic schedules to verify cluster data integrity
1.24	Run file system (HDFS) checks on periodic schedules to verify file system integrity
1.25	Maintain space allocation for logical (within same cluster) development, test and production directory hierarchies for Client
1.26	Allocate space for user directories and files based on authorized requests
1.27	Maintain memory swap space settings
1.28	Maintain HDFS Directory Quotas (includes replication space)
1.29	Maintain HDFS Name Quotas
1.30	Maintain HDFS Trash settings
1.31	Perform Trash expunge
1.32	Maintain HDFS Logs clean-up processes
1.33	Report, when needed, on HDFS quota and actual usage for production data folders
1.34	Report, when needed, on HDFS quota and actual usage of trash and log folders
1.35	Coordinate with CS to maintain system-wide compression codec parameters for HDFS core
1.36	Coordinate with CS to maintain system-wide compression codec parameters for YARN
1.37	Manage production HDFS files/objects within OS, HDFS
1.38	Manage production directory and directory hierarchies within OS, HDFS
1.39	Manage Hive query objects
1.40	Manage HBase data objects

1.41	Manage and Maintain Hmaster and HRegionServer daemons
1.42	Manage and Maintain Hmaster (active and backup) configuration settings
1.43	Manage and Maintain Hregion Server configuration settings
1.44	Coordinate with Administration for start/stop of HBase and Zookeeper services
1.45	Manage and Maintain HBase Tables and Hierarchies
1.46	Manage and Maintain HBase User Tables and Hierarchies
1.47	Manage and Maintain HBase table attributes and regions
1.48	Manage and Maintain HBase cluster load balancer services
1.49	Manage and Maintain HBase configuration with Zookeeper
1.50	Manage and Maintain HBase-site and environment configuration
1.51	Review and Analyze HBase and HDFS metadata issues
1.52	Manage and Maintain HBase logs
1.53	Report on HBase data space allocation
1.54	Review and Analyze HBase data skew across cluster
1.55	Manage HBase data access logs / settings
1.56	Manage and Maintain Hcatalog and Hive Metadata Database Tables and Hierarchies
1.57	Manage and Maintain Hcatalog and Hive User Tables and Hierarchies
1.58	Manage and Maintain Hcatalog and Hive User Views
1.59	Manage and Maintain Hcatalog and Hive Single / Dynamic Partitions
1.60	Manage and Maintain Hcatalog and Partitions notifications
1.61	Manage and Maintain Hcatalog and Hive Indexes
1.62	Manage and Maintain Hcatalog and Hive Functions
1.63	Review and Analyze Hcatalog and Hive Data Skew
1.64	Maintain Hcatalog metadata database settings
1.65	Review Hcatalog metadata database logs
1.66	Perform Hcatalog metadata periodic backups
1.67	Manage and Maintain Hcatalog and Hive data access
1.68	Respond to Hcatalog load or data access incidents
2	**Application Monitoring:** Services to provide core system monitoring functions including cluster availability; this will include a standardized monitoring structure through Viewpoint and/or Ambari. The service may additionally utilize alerts/ data collection from Nagios, Zookeeper and Ganglia to aid in issue isolation and troubleshooting.
2.01	Monitor HDFS cluster • Monitor NameNode heap consumption • Monitor NameNode upgrade and safemode transitions • Monitor HDFS disk usage • Need-based log analysis for NameNode • Need-based analysis of thread stacks for NameNode • Need-based analysis of JMX metrics for HDFS
2.02	Monitor YARN cluster • Monitor ResourceManager uptime • Monitor all the NodeManagers in your cluster (live, blacklist, graylist, excluded) • Monitor YARN container (average containers per node, container utilization) • Monitor total jobs and containers • Monitor ResourceManager scheduling information to analyze the queue usage • Need-based analysis of historical trends for Jobs • Need-based analysis log for ResourceManager • Need-based analysis of thread stacks for ResourceManager • Nee- based analysis of JMX metrics for YARN

2.03	Monitor HBase Cluster • Monitor all the RegionServers (live, dead, in transition) • Monitor HBase Master status • Monitor heap utilization for your HBase cluster • Need-based analysis of HBase cluster logs • Monitor key health metrics for Zookeeper using the Zookeeper info link • Nee- based analysis of debug dumps for HBase cluster
2.04	Monitor Hadoop Cluster Key Indicators • Bytes written to and read from HDFS cluster (aggregated across all the DataNodes) • Utilized containers vs total available containers for the YARN cluster • Combined YARN backlog
2.05	Review executing jobs for monitoring anomalies
2.06	Kill jobs, map or reduce tasks based on Client authorization
2.07	Report on key system metrics trends including Data Node and NameNode resource utilization (CPU, I/O, memory)
2.08	Identify dead node and investigate node issues
3	**Application Security**: This provides standardized services to provide OS level and HDFS roles and access rights administration based on access controls deployed by the customer.
3.01	Create, Modify, Delete Linux user ids and passwords POSIX user admin
3.02	Create, Modify, Delete Hue user admin
3.03	Maintain Hadoop cluster local Kerbros KDC and Default Domain configuration
3.04	Maintain user authentication settings on the NameNode using local and Kerberos
3.05	Maintain the linkage / configuration of NameNode with LDAP authentication mechanism
3.06	Maintain network linkage of NameNode with LDAP authentication server
3.07	Maintain Kerbros KDC authentication security configuration with HBase (Server side)
3.08	Maintain Kerbros KDC authentication security configuration with Zookeeper
3.09	Maintain authentication security configuration with Ambari
3.10	Maintain Kerbros KDC authentication security configuration with HCatalog
3.11	Maintain Hadoop HDFS user home directories and ownership permissions as required
3.12	Maintain security access to the Hcatalog / HBase metadata database
3.13	Grant, Revoke Privileges on file system objects to users, groups as required
3.14	Maintain user attributes as required
3.15	Maintain and Administer privileged user logins, passwords based on client security guidelines
3.16	Periodically review and Report access violations
3.17	Perform periodic security reporting including roles and user entitlements
3.18	Provide scheduled security compliance reports for evidence of compliance to customer defined security controls (e.g., access denial report, logon reports, etc.) on a predetermined schedule
3.19	Maintain Hadoop administrative user permissions for creating log files
4	**Configuration Management**: Participate as a stakeholder in CS configuration management changes for the Hadoop cluster and coordinate cluster restarts.
4.01	Coordinate with Cloud Ops for HDFS Namenode and Data node configuration file maintenance (HDFS-site.xml) (*e.g., DataNode storage locations, blocksizes and reserved storage space)*
4.02	HDFS log configuration file maintenance
4.03	Hadoop core configuration files maintenance *(core-site.xml)*
4.04	YARN ResourceManager and NodManager configuration files maintenance *(yarn-site.xml) (e.g., container memory, cpu, etc.)*
4.05	Maintain Hadoop default environment variables (e.g., log directory, heap size, etc.)
4.06	Distribute new configuration file changes *(HDFS safe mode, stop/ restart)*

Services Included with Teradata IntelliCloud (Hadoop)

5	**Cluster Change Management**: Participate in HDFS and Cluster hardware and software change activities supporting cluster administrative components.
5.01	Coordinate with Cloud Ops hardware support team on cluster hardware changes
5.02	Coordinate with Cloud Ops on cluster OS software changes management
5.03	Coordinate with Cloud Ops on Hadoop distribution change management
5.04	Cluster hardware / software / Hadoop distribution
5.05	Capture cluster configuration metrics prior to and after major hardware or software changes
5.06	Maintain data management settings during server, web server and network changes
5.07	Ensure Hadoop server functions properly and all services are online after cluster change management activities
6	**Performance & Capacity**: Maintain cluster performance and resource usage data. Report on key performance metrics including resources, and user query/workload performance trends on periodic basis. **Workload Management**: Performance workload prioritization through capacity or Fair scheduler techniques for user-batch workload. **Performance Analysis and Tuning**: Perform analysis for resource (CPU, I/O, memory) degradation, compression, data skews and long running workloads.
6.01	Collect data and report on % CPU Utilization and periodic (1, 5, 15 minutes) CPU load averages
6.02	Collect data and report on average disk data transfer rates (I/O bandwidth)
6.03	Collect data and report on number of disk I/O operations per second
6.04	Collect data and report on HDFS bytes written/read
6.05	Collect data and report on Average Memory utilization
6.06	Collect data and report on Average Swap space utilization
6.07	Collect data and report on Average data transfer rate (In-flow / Out-flow network traffic)
6.08	Collect data and report on Average network latency
6.09	Perform data collection and reporting on Map slot utilization trends
6.10	Perform data collection and reporting on Reduce slot utilization trends
6.11	Collect data and report on prioritized service operations (load jobs) under Teradata responsibility
6.12	Collect data and report on the average percentage container utilization for YARN
6.13	Collect data and report on the percentage storage capacity utilization for HDFS
6.14	Collect data and report on user and system data space utilization and growth trends
6.15	Collect data and report on canary user job performance trends; utilize performance canary jobs to identify variances from threshold
6.16	Review and document Client's existing Jobs / Workload requirements
6.17	Review and document Client's existing Jobs / Workload Fair / Capacity based scheduling settings
6.18	Review and document client jobs/workload categorization, session priorities and concurrency settings
6.19	Perform YARN Job trend analysis for pre-scheduling periodic Jobs
6.20	Maintain Capacity or Fair Scheduler based Workload prioritization settings based on approved change requests

6.21	Review and Analyze YARN cluster Key performance Indicators for performance improvement opportunities • Trending data for the YARN jobs - Arriving Job rate per performance period - Running Job rate per performance period - No. of Jobs submitted - No of Jobs completed - Failed Jobs - Hung Jobs - Aborted Jobs • Container Utilization • ResourceManager heartbeats • ResourceManager responsiveness • ResourceManager activity pattern • ResourceManager's key Garbage Collection statistics • ResourceManager's memory and heap utilization
6.22	Review and Analyze HDFS Daemon processes metrics for performance improvement opportunities • CPU Utilization • I/O Utilization • Memory Utilization • Heap-size Utilization • Node up/down trends • HDFS job trends
6.23	Review and Analyze HBase cluster Key performance Indicators for performance improvement opportunities • Number of requests received by HBase Master • Total number of read requests aggregated across all the RegionServers • Total number of write requests aggregated across all the RegionServers • Average number of regions served by each RegionServer • File system read latency aggregated for all the RegionServers • File system write latency aggregated for all the RegionServers • Key Garbage Collection statistics aggregated for all the RegionServers • Average responsiveness for all the RegionServers
6.24	Review and Analyze aggregated server utilization for performance improvement opportunities • CPU Utilization • I/O Utilization • Memory Utilization • Heap-size Utilization • Node up/down trends • Cluster job trends
6.25	Review and Analyze Cluster resource skew from YARN and HDFS jobs
6.26	Review and Analyze data compression opportunities for YARN and HDFS jobs
6.27	Analyze HDFS jobs for alternate distribution options to enhance job performance
6.28	Analyze YARN jobs for alternate distribution options to enhance job performance
6.29	Review performance incidents, troubleshoot and provide resolution options
6.30	Review performance improvement opportunities with Client for implementation
6.31	Coordinate with Client developers, data scientists and analysts for implementing performance changes

6.32	Recommend advanced performance optimization PS services as needed
6.33	Use data balancer settings and balance data blocks across cluster
6.34	Run periodic cluster performance benchmarks
6.35	Baseline cluster resource consumption prior to and after cluster hardware changes • IO Performance using benchmark tools like TestDFSIO • Shuffle phase performance across cluster using benchmark tools like TERASORT • Vital workload canary setup using benchmark tools like MRBench • Sample load performance test jobs • CPU, Memory, and Network performance statistic
7	**Support**
7.01	Maintain Call Register
7.02	Maintain Issues, Risks & Actions Register
7.03	Manages incidents or changes according to client's remediation / change management process
7.04	Communicate daily operational and incident status to appropriate groups as established in the operations manual
7.05	Review service change requests and provide feedback on work effort, implementation plan, estimated duration and back out plan for changes
7.06	Alert team members regarding operational anomalies and corrective actions required due to performance and/or capacity utilization issues based on run book escalation process
7.07	Submit incidents for hardware problems identified by Hadoop Administrator to Cloud Ops and follow through to resolution
7.08	Submit incidents for software problems identified by Hadoop Administrator to Cloud Ops and follow through to resolution
7.09	Prepare or contribute to monthly Teradata operations calendar
7.10	Monthly system & operations status report
7.11	Manage day-to-day contractual obligations and change requests
7.12	Schedule and manage resources in accordance with agreed-upon service coverage and delivery quality expectations
7.13	Respond to requests for extra shifts or changes in coverage
7.14	Participate in incident post mortem/root cause review
7.15	Serve as a focal point for management and escalations of high impact problems
7.16	Act as escalation point and facilitate cross organizational communication within Teradata when required to resolve issues
7.17	Conduct Quarterly Service Review