Business

Technical appendix to the Managed Applications Service Description
Managed Big Data

Table of contents

1 Description of the Big Data Managed service

Managed Big Data is a service that allows customers to generate value from their business data, and offer for example predictive maintenance, fraud detection or better customer knowledge.

This service consists of different solutions that are all managed by the Service Provider (Infrastructure and Big Data Components). These solutions work with the same philosophy: Collect data in batch or streaming mode, Store data, process data and visualize data.

The following Big Data software solutions are offered and managed by the Service Provider. Each solution can be selected according to customer needs

- Big Data with Cloudera CDP/ CDF
- Native Big Data Services with Flexible Engine
- Native Big Data Services with Google GCP
- Native Big Data Services with Microsoft Azure
- Native Big Data Services with AWS.
- Log As A Service Solution with Elastic

1.1 Cloudera Solutions and Native Big Data Solutions of Flexible Engine, GCP, AWS and Azure

The service includes some or all of the following:

- A secure administration portal provided with the Big Data platform,
- A dedicated tenant to ingest, store, process and visualize Customer data,
- A 24/7 monitoring and alert solution,
- Installation, configuration and operational maintenance of Big Data solution components by the Service Provider,
- The implementation by the Service Provider of the following services on request:
    - Audit evaluation of data and architecture
    - Migration of an existing Big Data solution to the Service Provider environment,
    - Development of business use cases with the Service Provider's internal partners,

For the Big Data Services offered in self-service by the Public Cloud Provider (Flexible Engine, GCP, Azure and AWS), the Provider manages the infrastructure and Big Data components for the Customer.

1.2 Log As A Service (LaaS) Solution

The Log As a Service solution managed with Elastic components is a service managed by the Service Provider. It is a complete end-to-end log analysis solution that helps in-depth research, analysis and visualization of logs generated by different machines.

The Log As A Service solution is offered in shared or dedicated mode according to the customer's choices:

- In the mutualized mode, the Service Provider implements a single and common platform for all customers. It is based on Elastic Cloud Enterprise (ECE) and runs on the IaaS Flexible Engine.
- In the dedicated mode, the Service Provider implements a platform dedicated to the customer. It is based on Elastic products (ECE version or standard version subject to Platinum license). The platform will be based on any type of IaaS (after feasibility study and technical validation by the Service Provider).

The service consists of some or all of the following:

- A secure administration platform, shared for all Clients (instantiated within the Log As A Service platform) or dedicated to each client on the Client's preferred IaaS.
- The instantiation of Customers in shared secure Allocators,
- Management of Elastic licenses provided by the Service Provider
- 24 x 7 supervision of deployed Elastic clusters
- Installation and maintenance in operational conditions of all or part of the components (Kafka, Logstash, APM, Machine Learning, Kibana) for the Customer
- Implementation of the following support services on request:
    - Custom data collection for:
        - Adding new data sources to your Elastic cluster
        - Support in updating the architecture (adding nodes) if necessary, according to the volume of data added
        - The creation of custom dashboards for the creation of dashboards adapted to the Client's use
    - Deployment of machine learning for the Client
    - Support in the use of Elastic and the solution for the Customer's teams on the use of the Elastic suite

2 Prerequisites for access to the service

- The Managed Big Data Service is based on a supported IaaS and its services, to which the Customer must first subscribe with the third-party cloud provider or the Service Provider. IaaS services are not part of the Service except for the Log As a Service solution in shared mode which includes a shared infrastructure.
- The Customer must subscribe the necessary network access to the Service of the selected IaaS
- The public cloud provider's native Big Data services (FE, GCP, Azure and AWS) can only run on its IaaS
- The tenant subscribed by the Client is a managed tenant, administered by the Service Provider.
- For Big Data with Cloudera, Cloudera CDP/CDF software licenses must be purchased by the Customer from the Service Provider.
- For the Log As a Service shared solution, the customer does not need to purchase Elastic licenses. They are already included in the Work Unit rate and are provided by the Service Provider.
- For the dedicated Log As a Service solution, the customer must purchase Elastic licenses. Elastic license management is provided by the Service Provider

3 Services proposed

The following phases are optional for the service: data and security assessment and migration.

The following phases are mandatory for the service: Installation, configuration, operation, supervision, monitoring, backup and change management.

The following table lists the services provided under the "Managed Big Data" services and all its solutions:

Table 1: Description of "Managed Big Data optional services"

Phase	Activities
Audit of data and security	**A project management phase led by the Service Provider to** ▪ Define the scope of work ▪ Evaluate the Client's architecture, data and processes (batchs) ▪ Evaluate data policies and network prerequisites **ETL Evaluation** ▪ Purpose of the data ▪ Volume and criticality of flows ▪ Description of the jobs ▪ Location and format of original data **Assessment of security constraints** ▪ Confidentiality and sensitivity of data ▪ Role of the IAM organisation ▪ VPN or interconnect solution
Migration	**A project management phase led by the Service Provider to** ▪ Manage the migration of an existing Big Data solution to the Service Provider's environment, ▪ Prerequisites: The data and security evaluation phase must be completed **Collect the Client's needs** **Definition of functional and technical architecture (macro and detailed)** **Establish migration plan and test plan taking into account** ▪ Integration of database assets and structures ▪ About security and encryption ▪ Current pipelines and treatments **Perform the migration** ▪ Install and configure the Big Data solution (detailed in the next sections) ▪ Configure the Customer's pipelines ▪ Migrate Customer data and processing ▪ Test migration

Table 2: Description of "Managed Big Data services"

Phase	Activities
Installation	**Installation tasks under the responsibility of the Service Provider** ▪ Definition of the design, ▪ Infrastructure configuration (tenant and VM) ▪ Big Data software installation according to vendor recommendations ▪ ITSM Software Installation ▪ Cloudera CDP/CDF and Logaas: software license installation Depending on the complexity and size of the clusters, big data experts (junior/senior) will carry out the installation. **Installation tasks requested by the Client and performed by the Service Provider** ▪ Additional VMs to run applications

Phase	Activities
Configuration	**Installation tasks under the responsibility of the Service Provider** ▪ Creation of initial user accounts and role policies ▪ Configuration of monitoring and supervision ▪ For "Native Big Data Services with Flexible Engine", when the "Advanced Security" option is subscribed, configure the Environment to use Kerberos and/or LDAP for user authentication ▪ Configure security and backup policies ▪ Install and configure third-party software if necessary. **Configuration tasks requested by the Client and performed by the Service Provider** ▪ Configuration and tuning of all VM and components depending on customers application
Operations	Only the Service Provider has root or administrator access on servers with a managed Tenant. The Customer will be granted access to the relevant applications as a User only, to perform the tasks under its responsibility. **Administrative tasks under the responsibility of the Service Provider** ▪ Review the performance of the nodes and perform the adjustment, ▪ Record, analyze alarms and apply remediation procedures, ▪ Manage the start and stop of Big Data services, ▪ Add/Remove nodes, ▪ Ensure the availability of big data processing and task logs ▪ Perform minor and major updates of Big Data components **Administrative tasks requested by the Client and performed by the Service Provider** ▪ Create scheduled processes and tasks ▪ Manage data retention policies ▪ Management of the Client's user accounts ▪ Run/implement minor or maintenance versions specific to a third-party component ▪ Complete processing that is running in Hadoop ▪ Management of ingestion flows and Kafka topics **Complex administrative tasks requested by the Client and performed by the Service Provider** ▪ The execution/implementation of major versions specific to a third-party component is considered an additional service and will be billed to the Customer as such.
Supervision (Alerting, Monitoring)	**Supervision tasks under the responsibility of the Service Provider** ▪ Preventive monitoring of metrics and KPIs ▪ Monitor availability and alert the Customer ▪ Monitor performance and alert the customer ▪ Monitor capacity (including disk) and alert customer ▪ Monitor the status of the service ▪ The alarms generated are sent to the central monitoring console, which is managed by the Service Provider's operating teams (incident management process). **Supervision tasks requested by the Client and performed by the Service Provider** Monitor customer specific jobs to detect long runs, errors, etc.
Backups	The Service Provider implements backup policies in order to protect the Customer's Platforms in the event of loss of configuration of a component, allowing the restoration of a Node or rollback. However, Slave Nodes are not backed up: in case of service failure, a new Slave Node is deployed by the Provider. Customer data is not included in backup policies. For all other VMs, the backup solutions implemented are: ▪ Hot file backup for master and device nodes ▪ Dumps for the databases ▪ Snapshots if possible

Phase	Activities
Change Management	The service offers the Client the possibility to request: - Standard changes - Custom edits See chapter below for an exhaustive list

4 Technical specifications

Managed Big Data includes the following solutions and components

4.1 Cloudera CDP/ CDF managed by the Service Provider

Cloudera CDP/CDF includes the components listed below.

Table 3: Specifications of the components of "Managed Big Data with Cloudera CDP/CDF"

Module	Components
Core	**CDP (Cloudera Data Platform)**
	Dataiku Studio DSS
	Airflow
Ingestion	**CDF (Cloudera Data Flow)**
Processing	**Spark**
	Hive
	Druid
Storage	**Druid**
	Hue
Visualization	**JupyterHub**
	Zeppelin
	Dataiku Studio DSS
	R Studio
	Blackboard

4.2 Native Big Data Services with Flexible Engine managed by the Provider

Flexible Engine offers a set of Big Data components in native services, managed by the Service Provider and listed below.

Table 4: Component specifications of "Big Data Native Services with Flexible Engine"

Modules	Components
Ingestion	**DIS (Data Ingestion Service)**

Modules	Components
	DMS (Distributed Message Service for Kafka)
Processing	CSS (Cloud Stream Service)
	MRS (MapReduce Service)
	DLI (Datalake Insight)
	MLS (Machine Learning Service)
	CS (Cloud Stream)
Storage	OBS (Object Storage Service)
	DWS (Data Warehouse Service)
Visualization	MRS (MapReduce Service)
	CSS (Cloud Stream Service)

4.3 Native Big Data Services with Google GCP managed by the Provider

Google GCP offers a set of Big Data components in native services, managed by the Service Provider and listed below.

Table 5: Component specifications of "Big Data Native Services with GCP"

Modules	GCP Components
Ingestion	Cloud Pub/Sub
Processing	Dataflow Cloud
	Cloud Dataprep
	Cloud Dataproc
	Datafusion Cloud
	Cloud Composer
Storage	Data Catalog
	SQL Cloud
	Cloud spanner
	Big Query
	Cloud Datastore
	Cloud Bigtable
Visualization	Data Studio

4.4 Native Big Data Services with AWS Managed by the Provider

AWS offers a set of Big Data components in native services, managed by the Provider and listed below.

Table 6: Component specifications for “Big Data Native Services with AWS”

Modules	AWS Components
Ingestion	Kinesis Streams
Processing	MRB
	Kinesis Analytics
Storage	S3
	Aurora
	RDS
	Redshift
	DynamoDB
	Elastic Search
	Kinesis Streams
Visualization	QuickSight

4.5 Native Big Data Services with Microsoft Azure managed by the Provider

Azure offers a set of Big Data components in native services, managed by the Provider and listed below.

Table 7: Component specifications for “Big Data Native Services with Azure”

Modules	Azure Components
Ingestion	Event Hub
Processing	HDInsight
	Data Factory
	Databricks
	Stream Analytics
	Analysis Services
	Machine Learning
Storage	Data Lake Storage
	Cosmos DB
	Synapse Analytics
Visualization	PowerBI

4.6 Log As A Service (LaaS) solution managed by the Service Provider

The Log As A Service solution based on Elastic offers a set of Big Data components in native services, managed by the Service Provider and listed below.

Table 8: “Solution Log As A Service” Component Specifications

Modules	Components
Ingestion	**Beats**
Processing	**Logstash**
	Observability
	SIEM
	Application Performance Monitoring (APM)
	Machine Learning (AIops)
Storage	**Elastic**
Visualization	**Kibana**

5 KPI & alerts

Monitoring and logging

Alarms are generated by the Vendor’s monitoring system, supplemented by Shinken agents for the Cloudera CDP/CDF solution.

Alarms are sent to the central monitoring console managed by Orange Business’s operational teams.

Metrics observed in the RUN phase

- Status of services through life tests
- Host status (system/network/storage resources, processes, components, configurations)
- Performance of services and hosts
- Queue Status
- Status of tasks/processes scheduled in streaming and batch

Alerts configured

- Interruption of service and components
- Performance of services and components
- Network Connectivity Outage
- Full queue
- Processing and requests in error and warning
- Errors/warning in component and application logs

6 Backup and restore

Data backup and restore

System data, access data and client data are backed up by dump or snapshot depending on the solutions.

Service restore

Detailed procedure to restore the configuration after a service reset.

7 Price conditions

The Service's pricing consists of:

- A fee for access to the service «Managed Big Data» including all the tasks mentioned in the implementation RACI and indexed on the number of servers and features to configure.
- A monthly recurring covering the activities related to the maintenance in operational condition of the service «Managed Big Data» indexed on the number of functionalities and servers or Work Unit depending on the solution implemented

Service rates do not include:

- The price of the infrastructure that you must otherwise subscribe with the IaaS provider according to the rates in force (except for the «Log As A Service» solution in shared mode and on IaaS Flexible Engine)
- The requests for change.
- Vendor licenses for the «Cloudera CDP/CDF» solution
- Vendor licenses for the «Log As a Service» solution in dedicated mode.

8 Change Requests

We prepare the implementation of a change in consultation with you. Once the request is processed, you will be notified to validate and close the request.

8.1 Big Data with Cloudera CDP/ CDF

Component	Description	Number of Tokens
Big Data	Change a component configuration setting	2 tokens
Big Data	Stop/Restart a service (in maintenance mode)	2 tokens
Big Data	Manage rights and quotas: users, directories	2 tokens
Big Data	Run a system command (root)	2 tokens
Big Data	Request information (logs, status)	2 tokens
Oozie	Create/Edit/Delete one or more job(s)	2 tokens
Oozie	Suspend/Reactivate/Reschedule a job	2 tokens
Oozie	Start/Stop a job	2 tokens
Oozie	Request information (logs, status)	2 tokens
KAFKA	Create/Delete a topic	2 tokens
KAFKA	Modify parameters of a topic (retention, partitions,.)	1 token
KAFKA	Request information (log, status)	2 tokens
Hive	Create/Delete a Hive Database	2 tokens
Hive	Create/Delete a Hive table	2 tokens
Hive	Request information (log, status)	2 tokens
YARN	Create/Remove a new Yarn queue	2 tokens
YARN	Change parameters of a Yarn queue (name, % of uses,..)	1 token
YARN	Request information (log, status)	2 tokens
HDFS	Create/Delete a Directory	2 tokens
HDFS	Change a directory replication factor	2 tokens
HDFS	Request information (log, status)	2 tokens

IPA	Create/Edit/Delete a user	2 tokens
IPA	Create/Edit/Delete a group	2 tokens
IPA	Set up a matrix	3 tokens
SPARK	Set up new Python libraries	2 tokens
SPARK	Request information (log, status)	2 tokens
Ranger - HDFS	Permission on a directory (add/remove/edit)	2 tokens
Ranger - HDFS	Permission on a user directory (add/delete/edit)	2 tokens
Ranger - HIVE	Permission on a base (add/remove/edit)	2 tokens
Ranger - HIVE	Permission on a table (add/remove/edit)	2 tokens
Ranger - KAFKA	Permission on a topic (add/delete/edit)	2 tokens
Ranger - YARN	Permission on a queue (add/remove/edit)	2 tokens
NoSQL - ElasticSearch	Add access for a user/group	2 tokens
NoSQL - ElasticSearch	Set up a matrix	3 tokens
NoSQL - ElasticSearch	Request information (log, status)	2 tokens
NoSQL - Kibana	Add access for a user/group	2 tokens
NoSQL - Kibana	Request information (log, status)	2 tokens
NIFI	Add/Remove/Change permissions	2 tokens
NIFI	Add/Delete/Edit a user/group	2 tokens
NIFI	Request information (log, status)	2 tokens
InfraSoLR	Add/Remove/Change a configuration	2 tokens
InfraSoLR	Add new InfraSolr instances	2 tokens
InfraSoLR	Request information (log, status)	2 tokens
DRUID	Create/Edit/Delete a data source	2 tokens
DRUID	Change the number of simultaneous requests	3 tokens
DRUID	Modify the configuration (LLAP, druid, coordinator, ...)	3 tokens
DRUID	Request information (log, status)	2 tokens
Hbase	Create/Delete an Hbase database	2 tokens
Hbase	Create/Remove an Hbase table	2 tokens
Hbase	Request information (log, status)	2 tokens
RANGERKMS	Encrypt a directory	2 tokens
RANGERKMS	Create a key	2 tokens
RANGERKMS	Request information (log, status)	2 tokens
Others	Other request	3 tokens

8.2 Native Big Data Services with Flexible Engine

Component	Description	Number of Tokens
Big Data	Change a component configuration setting	2 tokens
Big Data	Stop/Restart a service (in maintenance mode)	2 tokens
Big Data	Manage rights and quotas: users, directories	2 tokens

Big Data	Run a system command (root)	2 tokens
Big Data	Request information (logs, status)	2 tokens
MRS - Oozie	Create/Edit/Delete one or more job(s)	2 tokens
MRS - Oozie	Suspend/Reactivate/Reschedule a job	2 tokens
MRS - Oozie	Start/Stop a job	2 tokens
MRS - Oozie	Request information (logs, status)	2 tokens
MRS - KAFKA	Create/Delete a topic	2 tokens
MRS - KAFKA	Modify parameters of a topic (retention, partitions,.)	1 token
MRS - KAFKA	Request information (log, status)	2 tokens
MRS - Hive	Create/Delete a Hive Database	2 tokens
MRS - Hive	Create/Delete a Hive table	2 tokens
MRS - Hive	Request information (log, status)	2 tokens
MRS - YARN	Create/Remove a new Yarn queue	2 tokens
MRS - YARN	Change parameters of a Yarn queue (name, % of uses,..)	1 token
MRS - YARN	Request information (log, status)	2 tokens
HDFS	Create/Delete a Directory	2 tokens
HDFS	Change a directory replication factor	2 tokens
HDFS	Request information (log, status)	2 tokens
MRS - SPARK	Set up new Python libraries	2 tokens
MRS - SPARK	Request information (log, status)	2 tokens
MRS - Ranger - HDFS	Permission on a directory (add/remove/edit)	2 tokens
MRS - Ranger - HDFS	Permission on a user directory (add/delete/edit)	2 tokens
MRS - Ranger - HIVE	Permission on a base (add/remove/edit)	2 tokens
MRS - Ranger - HIVE	Permission on a table (add/remove/edit)	2 tokens
MRS - Ranger - KAFKA	Permission on a topic (add/delete/edit)	2 tokens
MRS - Ranger - YARN	Permission on a queue (add/remove/edit)	2 tokens
NIFI	Create/Edit/Delete one or more job(s)	2 tokens
NIFI	Add/Remove/Change permissions	2 tokens
NIFI	Add/Delete/Edit a user/group	2 tokens
NIFI	Request information (log, status)	2 tokens
MRS - Hbase	Create/Delete an Hbase database	2 tokens
MRS - Hbase	Create/Remove an Hbase table	2 tokens
MRS - Hbase	Request information (log, status)	2 tokens
Others	Other request	3 tokens

8.3 Native Big Data Services with Google GCP

Component	Description	Number of Tokens
Big Query	Create table/modify table/delete table Add/modify/update/delete user with policies Copy table	1 token
Big Query	Charge data from a bucket	2 tokens
Big Query	Other exchange	Estimation in tokens based on time spent
Pub Sub	Create/modify/delete instance	1 token

Pub Sub	Create snapshot msg	2 tokens
Pub Sub	Other exchange	Estimation in tokens based on time spent
Pub Sub lite	Create/modify/delete instance Reservation management Throughput capacity	1 token
Pub Sub lite	Create snapshot msg	2 tokens
Pub Sub lite	Other exchange	Estimation in tokens based on time spent
Dataproc	Create/delete cluster	1 token
Dataproc	Bench/config cluster	4 tokens
Dataproc	Other exchange	Estimation in tokens based on time spent
Dataflow	Delete Job	1 token
Dataflow	Deploy/Create Job	1 Business Hour day
Dataflow	Other exchange	Estimation in tokens based on time spent
Cloud composer	Create/modify/delete GKE instance	1 token
Cloud composer	Add node	2 tokens
Cloud composer	Other exchange	Estimation in tokens based on time spent

8.4 Native Big Data Services with Azure

8.5 Native Big Data Services with AWS.

9 Custom requests

You can make a request out of catalog and provide the details of your need. We will set up an half hour phone call with you to ensure a good understanding of the need. 2 cases then arise:

- If the functional requirement is immediately qualified as simple, medium or complex tasks as defined in the catalogue, the Change request is finally reclassified as a request in the catalogue and can be processed by the operational teams.

- If the functional need is not immediately translatable into simple or complex tasks and this will require an in-depth study with a time frame and completion time, an estimate of the number of Tokens needed for the study will be made. This study is without guarantee of results given the wide variety of functional needs that can be expressed. In case of agreement, the study is carried out and leads to a feasibility or not. In case of feasibility, this is accompanied by an assessment of the costs related to its implementation. These expenses will be qualified as simple or complex change requests according to the criteria stated above.

10 Limitations

The Customer is responsible for the following activities:

- Sizing of the environment in accordance with the components requested by the Customer,
- Perform business tasks related to customer applications and use cases,
- Check the proper functioning of its business applications and use cases, in addition to Big Data software
- Produce and maintain documentation for the installation and configuration of its applications and use cases.
- Make the decision to restore Big Data nodes and databases,

- Project management, unless otherwise specified in the Technical and Financial Proposal

Execution/implementation of major versions specific to a third-party component is not included, available through additional professional services only and a specific project.

To benefit from the various services "Log As A Service" the Customer must perform the following actions:

- Adding logs in the cluster, creating indexes, managing Kibana spaces, managing users and roles, querying logs, reading access logs, creating a lifecycle policy, creating pipelines «beats», from the creation of dashboard on Kibana,
- Installing and configuring software/agent on application servers to transfer logs to the ECE cluster using for example filebeats and https.
- Installing Beats for data injection.