ukcloud

Making transformation happen.

Service Proposition

Big Data & Hadoop

IaaS

Overview

This proposition outlines UKCloud's positioning around the delivery of Hadoop/Big Data solutions.

Using our 'UKCloud for OpenStack' service coupled with specialised skills from Bluemetrix, we now have a proven mechanism to successfully deliver Hadoop to our customers on a consumption basis without having to undertake the complexities of Hadoop cluster design, implementation or support.

What is Big Data and how does Hadoop assist?

Organisations are creating data at an unprecedented rate. New technologies such as Internet of Things (IoT) sensors, 'Smart' devices & cities and Artificial Intelligence generate data with such fierce velocity and volumes that traditional processing is unable to provide information from this data quick enough to provide the fullest business value.

Hadoop is an open source collection of tools designed specifically to capture, store and process this new wave of data in a cost-effective way, whilst also facilitating high speed data aggregation and analytics that provide organisations with valuable insights.

How does UKCloud's Hadoop proposition assist customers?

In providing a 'one-stop shop' for customers to procure both Infrastructure as a Service and Hadoop expertise as a single procurement exercise, the following benefits can be realised;

- Removes the burden of complex Hadoop cluster design and implementation from organisations
- Reduces the support overhead traditionally associated with operating a Hadoop cluster
- Reduce the capital expenditure risk typically associated with exploring big data by using the flexibility of an OPEX-based cloud service
- Reduce storage cost and complexity by consolidating disparate data stores into a central location, while facilitating holistic analytics across the widest possible dataset
- Enable analytics across multiple organisations' datasets by creating a shared data repository
- Create a data store capable of handling high-volume, high-variety and high-velocity data (for example, social media sentiment analysis, IoT sensor data, large machine – generated/sensor data) that exceeds the capabilities of traditional database technologies

How is the proposition packaged?

UKCloud has packaged our Hadoop proposition in the following three simple steps;

> Design and implement Hadoop solution → Choose your OpenStack Infrastructure → Engage enhanced Hadoop support (optional)

UKCloud
A8, Cody Technology Park
Ively Road, Farnborough,
Hampshire, GU14 0LX

Telephone
01252 303 300

Visit the website
www.ukcloud.com

Email
info@ukcloud.com

Follow us on Twitter
@ukcloudltd

1. Hadoop design and implementation

Although there are many companies offering Hadoop consultancy and support, UKCloud has an established relationship with Bluemetrix who have proven their Hadoop design, delivery and support expertise.

UKCloud will work with Bluemetrix to design, build, deploy and test a customer's Hadoop cluster before handing it over to the customer ready to start ingesting and processing data.

Generally, the delivery of the cluster design and implementation comes in the choice of the following two options;

1.1 Hadoop deployment without Kerberos security

This is where a standard Hadoop cluster is deployed without any of the advance security or auditing Kerberos security provides. Ideal use cases would be proof of concept/pre-production environments, or production environments where customers have existing data access protection controls.

An indication as to delivery time and cost would be 5 days at £6,500.

1.2 Hadoop deployment with Kerberos security

This is where a standard Hadoop cluster is deployed with Kerberos security to provide advance security and auditing capabilities from day one. An ideal use case would be a production environment where customers have no existing data access protection controls.

An indication as to delivery time and cost would be 10 days at £13,000.

Note: The options stated above are based upon general Hadoop usage. More specific use cases may require more advanced design and implementation considerations. Customers unsure of their requirements are advised to engage with Bluemetrix for a one-day design workshop prior to entering the implementation phase. Indicative cost for this workshop is £1,600.

2. Choose you OpenStack Infrastructure

UKCloud's OpenStack Infrastructure can be consumed from a mix of the following options;

2.1 Standard OpenStack Instances

These are our standard public OpenStack instances (t, m and r series), engineered to run standard cloud native workloads at a cloud economic price-point. Full details of these instances can be found in our [UKCloud for OpenStack](#) Service Definition; an indication of pricing as of G10 shown in Appendix B.

2.2 Big Data/Analytic optimised OpenStack Instances

An extension of our public OpenStack service, our b-series instances provide a bare-metal style Infrastructure as a Service solution. Our b-series instances provide high-capacity ephemeral storage from drives that reside in the same host as the compute workload, making them ideal for more data intensive workloads requiring consistent

UKCloud
A8, Cody Technology Park
Ively Road, Farnborough,
Hampshire, GU14 0LX

Telephone
01252 303 300

Visit the website
www.ukcloud.com

Email
info@ukcloud.com

Follow us on Twitter
@ukcloudltd

performance. Full details of these instances can be found in our [UKCloud for OpenStack](#) Service Definition an indication of pricing as of G10 shown in Appendix B,

2.3 Dedicated, Private Compute OpenStack Instances

For organisations wanting an extra layer of data isolation from a security and performance perspective, UKCloud also offers a Private Compute option that can be used as the infrastructure foundation for your Hadoop Cluster. Full details can be found in our [Private Cloud for Compute](#) Service Definition,

3. Engage enhanced Hadoop support (optional)

By default, UKCloud will provide our awarding winning support for your Hadoop infrastructure along with providing reactive support services to your Hadoop service as defined in our '[Hadoop management Roles & Responsibilities (R&R)](#)' service matrix, all at no additional cost.

Once the Hadoop cluster has been delivered to the customer, it is assumed that the customer will assume the roles and responsibilities defined in Appendix A for on-going pro-active support and maintenance. Customers unable or unwilling to undertake the responsibilities defined in Appendix A may want to consider extending their Hadoop support via one the following support arrangements:

3.1 Standard reactive support

UKCloud can provide enhanced Hadoop support as a value-added reactive service, charged at an indicative rate of £200 per hour that an incident or request is actively being worked on.

3.2 Proactive support contract

This option provides a higher level of managed service to the customer's Hadoop cluster. Beyond standard reactive support situations, this pro-active support contract offers;

- The operation and maintenance of the Hadoop Cluster
- Monitor the cluster and carry out preventative daily maintenance (or as required)
- Optimise clusters performance and security
- Resolve service and change requests as submitted by client

As an indication of cost, proactive support is charged monthly at £200 per deployed Hadoop node (virtual or physical).

Why UKCloud

- Our multi-cloud platform offers choice, scalability and flexibility for our public sector customers, without locking them into a single proprietary technology stack

- We are UK sovereign, with secure and resilient government-grade UK data centres separated by over 100 kilometres, located in Crown Campus. We give 2% of our pre-tax profits to charity, and our values, beliefs and behaviours are aligned to the needs of the UK Public Sector

UKCloud
A8, Cody Technology Park
Ively Road, Farnborough,
Hampshire, GU14 0LX

Telephone
01252 303 300

Visit the website
www.ukcloud.com

Email
info@ukcloud.com

Follow us on Twitter
@ukcloudltd

- Secure cloud platforms optimised for public sector workloads, fully aligned to the NCSC 14 Cloud Security Principles and subject to regular technical CHECK tests - information security and data governance comes as standard. We're compliant with GDPR/UK DPA 2018 (CISPE certified), the Network and Information Security Directive (NIS) and the new NHS Data Security and Protection Toolkit. These are supported by our wide portfolio of industry and public sector accreditations and certifications

- UK-based telephone service desk providing 24/7 support for P1 critical incidents, and a dedicated UK 24/7 Network Operations Centre (NOC) using industry-leading monitoring solutions on our Infrastructure. Access to UKCloud's supportive resources including Technical Account Managers, Cloud Delivery Managers, and Customer Success Managers to help customers through the onboarding and ongoing delivery of their solutions

- We work with a community of over 240 partners, including Independent Software Vendors, System Integrators and Managed Service Providers, to deliver end-to-end solutions specifically for the UK Public Sector

Why Bluemetrix

Bluemetrix have been working with Hadoop since 2009 and have been providing professional services since 2015 to our customers to help them design, deploy and operate their Hadoop deployments.

Since 2015 they have worked on approximately 400 projects across the EMEA and APAC regions for Tier 1 Enterprise level companies across all major sectors, such as financial, industrial, government, retail and automotive.

We provide a full end to end solution around the Hadoop stack, dealing with everything from Architecture to Deployment to Operations and all steps in between.

How do customers engage this proposition?

Initially a UKCloud Solutions Architect should be engaged with the end-customer to understand their requirements and validate whether a big/data/Hadoop solution may be appropriate. UKCloud's Solutions Architects can then guide customers through the options outlined in this document to further verify whether our proposition fits the customer's use case and start to define which of the options stated above should be applied.

UKCloud's Product Management team in collaboration with operations and the customer's Technical Account/Customer Success Manager will then work to co-ordinate the design and implementation phases to deliver a fully operational cluster to the customer.

Frequently Asked Questions

Q. Is there a preferred distribution of Hadoop?

Both UKCloud and Bluemetrix has experience on Hortonworks Data Platform (HDP) and Cloudera Data Hub (CDH). With the recent merger between Hortonworks and Cloudera, creating a new single Cloudera company, we recommend deploying any of the latest version of either HDP or CDH, until the release of the unified version.

UKCloud
A8, Cody Technology Park
Ively Road, Farnborough,
Hampshire, GU14 0LX

Telephone
01252 303 300

Visit the website
www.ukcloud.com

Email
info@ukcloud.com

Follow us on Twitter
@ukcloudltd

Q. Do customers need to use UKCloud's preferred Hadoop delivery partner?

No, customers are free to use whatever delivery partner they prefer or build Hadoop themselves on top of any of UKCloud's multi-cloud technologies. This proposal has been created to help simplify the complexities of deploying a Hadoop cluster for those customers with no existing capability.

Q. Do I have to use OpenStack for Hadoop?

For the purposes of this proposition the answer is yes, however customers with existing Hadoop capabilities are welcome to build their Hadoop cluster on any of UKCloud's multi-cloud technologies.

Q. Is a free trial available?

Free trials are available for our UKCloud for OpenStack and all other IaaS services, but due to the complex nature of Hadoop deployments we are unable to offer a free trial of this Hadoop proposition.

UKCloud
A8, Cody Technology Park
Ively Road, Farnborough,
Hampshire, GU14 0LX

Telephone
01252 303 300

Visit the website
www.ukcloud.com

Email
info@ukcloud.com

Follow us on Twitter
@ukcloudltd

Appendix A – Hadoop management Roles & Responsibilities (R&R)

The scope of the Hadoop proposition is split in to three distinct layers; IaaS, Hadoop Core Service and Hadoop Extended Services. The roles and responsibilities for these layers are defined below;

1.0 IaaS Layer

This is the OpenStack cloud infrastructure that will be used to support the Hadoop cluster. The R&R of services delivered at this layer are as follows;

	UKCloud	Customer/Partner
Deployment		
Virtual resources (instances, network, etc.)	X	
OS install and config	X	
Management		
Critical OS Patching on Hadoop nodes	X	
Increase capacity (add worker nodes)	X	
Instance back-ups	N/A	N/A
Monitoring		
OpenStack infrastructure/services	X	
OpenStack virtual resources (for example, instance ping test)	X	
Support		
OpenStack Infrastructure	X	
Restart of hung instances		X
Core Connectivity (Internet, N3/HSCN, PSN, Janet)	X	
OS on Hadoop nodes (basic services to support Hadoop)	X	
External Services (for example, Active Directory, LDAP integration)		X

2.0 Hadoop Core Service Layer

This is a composite of the minimal key services required to run a Hadoop Cluster. These components consist of the following;

- **Hadoop Distributed File System (HDFS)** – a scalable, fault-tolerant, distributed storage system that works closely with a wide variety of concurrent data access applications, coordinated by YARN

- **YARN** – the Hadoop resource negotiator that allows multiple data processing engines such as interactive SQL, real-time streaming, data science and batch processing

- **Ranger/Sentry** – define, administer and manage security policies across the cluster

- **Cloudera Manager** – the management platform for provisioning, managing, monitoring and securing Hadoop clusters

UKCloud
A8, Cody Technology Park
Ively Road, Farnborough,
Hampshire, GU14 0LX

Telephone
01252 303 300

Visit the website
www.ukcloud.com

Email
info@ukcloud.com

Follow us on Twitter
@ukcloudltd

- **ZooKeeper** – provides a distributed configuration service, a synchronization service and a naming registry for distributed systems

- **MapReduce (v2)** - the original framework for writing applications that process large amounts of structured and unstructured data stored in HDFS

- **Spark2** – a fast, in-memory data processing engine and expressive development APIs to allow Hadoop data workers to efficiently execute streaming, machine learning or SQL workloads that require fast iterative access to datasets

The R&R of services delivered at this layer are as follows;

	UKCloud	Customer/Partner
Deployment		
Installation of Hadoop Core services		X
Definition of Hadoop Core services configurations		X
Installation of additional services (non Hadoop Extended)*		X
Configuration of additional services (non Hadoop Extended)*		X
Management		
Hadoop Patching (in-line with Hadoop patches)		X
Patching (outside of Hadoop patches)*		X
Hadoop Upgrades (in-line with Hadoop releases)		X
Hadoop Upgrades (outside of Hadoop releases)*		X
Increase capacity (configure Core services on worker nodes)	X	
Performance Tuning		X
HDFS backups/restorations		X
Hadoop Core Services config backup/restoration		X
Monitoring		
HDFS capacity	X	
Compute capacity		X
HDFS Performance (customer defines thresholds)	X	
Hadoop Core Services *Service Status (P2) *Error Events (P1) (Email alert will also be sent from Ambari to customer nominated contact)	X	
Hadoop compute performance		X
Support		
Hadoop Core Services Triage, 1st line support	X	
Hadoop Core Services troubleshoot/restart based upon monitoring defined above	X	
General usage queries (beyond getting started)		X
Hadoop design/application consultancy		X
External Services (e.g. Analytics, BI, OLAP)		X

* May result in platform becoming unsupported

UKCloud
A8, Cody Technology Park
Ively Road, Farnborough,
Hampshire, GU14 0LX

Telephone
01252 303 300

Visit the website
www.ukcloud.com

Email
info@ukcloud.com

Follow us on Twitter
@ukcloudltd

3.0 Hadoop Extended Services Layer

An extended set of Hadoop services that provide additional data streaming, storage and processing beyond the Hadoop Core Services. These extended services include;

- **Pig** – allows users to write complex MapReduce transformations using a simple scripting language called Pig Latin

- **Hive/Impala** – a SQL-based data warehouse store native to Hadoop

- **HBase** – a non-relational (NoSQL) database that runs on top of HDFS

- **Phoenix –** a massively parallel, relational database engine supporting OLTP for Hadoop using Apache HBase as its backing store

- **Zeppelin** – a multi-purposed web-based notebook which brings data ingestion, data exploration, visualization, sharing and collaboration features to Hadoop and Spark

- **Kafka** – a high performance fault tolerant messaging system/streaming platform

- **Tez** – a framework for building high performance batch and interactive data processing applications, coordinated by YARN

The R&R of services delivered at this layer are as follows;

	Hadoop Extended Services		All other Hadoop Stack Components	
	UKCloud	**Customer/Partner**	**UKCloud**	**Customer/Partner**
Deployment				
Installation of Extended Hadoop Components		X		X
Configuration of Extended Hadoop Components		X		X
Management				
Hadoop Patching (in-line with official Hadoop distribution patches)		X		X
Patching (outside of Hadoop patches)*		X		X
Performance Tuning		X		X
Hadoop Extended Components config backup/restoration		X		X
Monitoring				
Hadoop Extended Components (monitored via Ambari) • Service Status (P2) • Error Events (P1)	X			X
Support				
Extended Hadoop Components - Triage, 1st & 2nd line support		X		X
Service troubleshooting/restarts based upon monitoring defined above		X		X
General usage queries (beyond getting started)		X		X
Hadoop Component design/application consultancy		X		X
External Services (for example, Analytics, BI, OLAP)		X		X

* May result in platform becoming unsupported

UKCloud
A8, Cody Technology Park
Ively Road, Farnborough,
Hampshire, GU14 0LX

Telephone
01252 303 300

Visit the website
www.ukcloud.com

Email
info@ukcloud.com

Follow us on Twitter
@ukcloudltd

Appendix B – OpenStack infrastructure pricing

These are our standard public OpenStack instances (t, m and r series), engineered to run standard cloud native workloads at a cloud economic price-point. Full details of these instances can be found in our UKCloud for OpenStack Service Definition, an indication of pricing as of G10 shown below;

Instance flavour	vCPU	RAM (GiB)	Ephemeral HDD (GiB)	Security domain	
				Assured £/hr	Elevated £/hr
t1.nano	1	0.5	10	£0.01	£0.02
t1.tiny	1	1	10	£0.015	£0.03
t1.small	1	2	20	£0.03	£0.05
t1.medium	2	2	40	£0.035	£0.055
t1.large	2	4	40	£0.04	£0.06
m1.small	4	8	60	£0.06	£0.10
m1.medium	4	16	60	£0.14	£0.18
m1.large	4	32	60	£0.30	£0.36
r1.small	8	16	60	£0.16	£0.22
r1.medium	8	32	60	£0.35	£0.40
r1.large	8	64	60	£0.70	£0.80

Note: All storage figures shown are raw and do not account for RAID, HDFS or system overheads.

1.2 Big Data/Analytic optimised OpenStack Instances

An extension of our public OpenStack service, our b-series instances provide a bare-metal style Infrastructure as a Service solution. Our b-series instances provide high-capacity ephemeral storage from drives that reside in the same host as the compute workload, making them ideal for more data intensive workloads requiring consistent performance. Full details of these instances can be found in our UKCloud for OpenStack Service Definition, an indication of pricing as of G10 shown below;

Instance flavour	vCPU	RAM (GiB)	Ephemeral HDD (GiB)	Security domain	
				Assured £/hr	Elevated £/hr
b1.medium	28	220	7200	£3.50	£4.00
b1.large	56	440	14400	£5.00	£5.50

Note: All storage figures shown are raw and do not account for RAID, HDFS or system overheads

UKCloud
A8, Cody Technology Park
Ively Road, Farnborough,
Hampshire, GU14 0LX

Telephone
01252 303 300

Visit the website
www.ukcloud.com

Email
info@ukcloud.com

Follow us on Twitter
@ukcloudltd