The Server Labs G-Cloud 15 Service Description

G-Cloud 15 Service Definition - Cloud

OpenCGA Cloud Services

	Ref: TSL/GCLOUD15/SERVICE_DESC
	Issue : 1.0
	Date: Jan 2025 For: **G-Cloud-15**

Table of Contents

.

1 What is OpenCGA

OpenCGA is an open-source project that implements a high-performance, scalable and secure platform for Genomic data analysis and visualisation.

OpenCGA provides the most advanced and complete genomic data platform. The performance, scalability and huge number features makes of OpenCGA an unique full-stack solution today. OpenCGA takes care of security and implements a high-performance query engine and analysis frameworks for *Big Data* analysis and visualisation in current genomics. OpenCGA uses the most modern and advanced technologies, and has been designed and implemented to scale to hundreds of thousands if genomes accounting for petabytes of variant data. It is built on top of three main components: *Catalog Metadata Database*, *Variant Storage Engine* and *Analysis Framework*.

○ OpencGA Main Features

- **Authenticated** and **secure** platform to query and visualise data. An advanced **permission** system has been implemented to ensure data privacy.
- A **metadata database** to keep track of registered users, projects, studies, files, samples, families, jobs, ...
- **Advanced Clinical Data** database implemented, users can define their data models for samples, patients or families.
- **Alignment storage** allows to index BAM/CRAM, calculate index and query data and coverage
- The most advanced, high-performance and scalable **Variant Storage Engine** solution today. Variant Storage Engine can normalise, load, index, aggregate, annotate and precompute variant stats for hundreds of thousands of whole genomes.
- **Analysis Framework** implemented on top of variant and alignment storage engines. OpenCGA comes with many analysis already implemented such as GWAS. Users can easily extend OpenCGA functionality by implementing a plugin or connecting to a external binary.
- Real **Big Data Analytics** supported, you can use different computing frameworks such as MapReduce or Spark on top HBase or Parquet files.
- Full **Clinical Analysis Solution** implemented, you can create the cases and run different clinical interpretations algorithms from your scripts or from a web application.
- Rich and comprehensive **RESTful Web Services API** with more than 160 endpoints to manage, query and analyse metadata, variants, alignments and clinical data.

.

- Easy **programmatic access** and **pipeline integration** thanks to the four different **client libraries** developed in **Java, Python, R** and **Javascript**
- Interactive **web-based application** to query, analyse and visualise variants, alignments and clinical data

2 What are the benefits of TSL Cloud Services

Running OpenCGA in the cloud requires the deployment of a large and complex cloud Infrastructure stack. Our expertise not only in the individual technologies of this stack, but in relation to how the complex OpenCGA application works and interacts with the stack, provides a rapid fast track deployment for Life Sciences organisations.
Added to this, support across the major cloud providers. Our fast track solution provides ready made, out of the box configuration to support the major clouds implementation of the stack technologies. This provides a huge time saving and trial and error for the customer.

3 What included in the TSL Cloud Services for OpenCGA

○ Deployment of OpenCGA Infrastructure Stack

The following stages are completed and components are deployed and configured out of the box.

- AWS account with users and permissions
- S3 terraform state bucket and terraform dynamodb table
- VPC cidr range for OpenCGA
- Opencga DNS name, certificates
- Deployment of AWS terraform stack
- Helm chart values configuration
- Security hardening/blocking of ingress NLB
- Helm deploy of OpenCGA and IVA charts

NOTE: This full deployment process can customised, depending upon the customers local cloud maturity and expertise. eg some customers may have pre-created accounts or VPCs to receive the deployment.

○ Testing of OpenCGA Infrastructure Stack

The OpenCGA deployment is signed-off and handed over to the customer after pre-defined OpenCGA smoke tests have been completed. These include data ingestion and query to confirm the operation of the solution.

.

- ## OpenCGA Infrastructure runbook

A comprehensive and detailed runbook is included which will enable the customer engineer to repeat all the tasks completed and needed to fully rebuild and operate the deployment.

- ## OpenCGA Infrastructure Sizing

The service provides guidance to sizing deployments for

1. build and daily running of OpenCGA, for differing sizes of deployment
2. once a system is running, how to change the deployment to optimise performance, during the ingestion of various volumes of data e.g for ingestion of samples from 10k to 100k

There is not one way to process sample ingestion as this varies based on how the ingestion will be done. That is parallelisation of loading, loading mix, other processing going on, etc. We describe different loading scenarios, and how this affects the various OpenCGA components in AWS. Understanding this will allow the right components to be scaled accordingly.

Also there may be times when large loading needs to be done (eg initial system deployment), which will be reduced after some period, and thus the environment can be scaled down to manage costs. The reverse is true; there may be a need to scale up the environment when a large amount of processing is expected; eg when a new study comes along.

.

4 The Server Labs

The Server Labs (TSL) is a 100% privately founded IT Consultancy and Software Development Company with headquarters in the UK and offices in Germany and Spain and now established as a leader in Cloud Computing services. The Server Labs focuses on the design and implementation of IT architectures and advanced software engineering projects working with the most advanced technologies to provide its clients cost-effective, scalable and high performance solutions. The Server Labs has been using the Cloud since 2006 and working with its customers in the cloud since 2008 and was one of the first European partner's of Amazon Web Services.

The Server Labs has clients in many different industry areas such as space, finance and telecoms. We collaborate with our clients to obtain success, committed to innovation, enjoying what we do every day and growing with every challenge.

The Company's mission is:

- To provide ***expert services*** in the field of IT architectures and advanced software engineering
- To ***improve radically the software development process***
- **To help organisations achieve better business results through the correct use of latest technologies**
- To have ***100% satisfied clients***
- To create high quality ***innovative software solutions***, providing added value to our customers

The specific value, experience and expertise that The Server Labs can provide for e-LfH are:

1. **Technical excellence** and capability to act as lead on architectural decisions and as technology expert in software and system subjects.
2. **Architecture experience at software and system level.**
3. **Proven experience in HPC and Big Data Projects**
4. **Real cloud computing experience, at IaaS, PaaS and SaaS levels, and for both compute power as well as storage solutions in different clouds.**
5. **Quality control based on ISO9001 for all software systems developed**
6. **Technological excellence**, especially in the main technologies required for the project, including HPC, web services and security technologies.

All our architects and engineers are experts with an average of 10 years' experience in the planning, design and development of complex software systems. Our multinational team has been a pioneer in Java technologies, Object Oriented Analysis and Design and distributed architectures, and has the required hands-on experience in many state of the art technologies. In the last few years, The Server

.

Labs has positioned itself as a leader in Cloud computing services, helping organisations move to the Cloud at all levels. For more information on current projects being undertaken by The Server Labs in Cloud computing, please see Appendix B.

Our experience working across several industries has given us a good understanding of the different requirements so we are able to provide the solution that best suits each particular business and reuse the lessons learnt in the other industry sectors when applicable. Our clients span organisations such as Banks (BNP Paribas, BBVA, Caja Madrid) , the European Space Agency (ESA), Madrid Underground (Metro de Madrid), ICCAT, Amadeus, TRAGSA, TIBCO (a leader in Messaging and Service Bus architecture systems), O2 and Telefonica, Vodafone, ORACLE, several, Sun Microsystems, TUI and Marsans travel, etc.

.