



BAHIA SOFTWARE

CTS-2022-0054

Análisis de requisitos

Miguel Boubeta
Sergio Figueiras
Román López
Luis Lorenzo

25/04/2023



*Unha maneira
de facer Europa.*



Fondos Europeos



Despregamento dunha infraestrutura baseada en tecnoloxías cuánticas da información que permita impulsar a I+D+i en Galicia.

Apoiar a transición cara a una economía dixital.

Operación financiada pola Unión Europea, a través do FONDO EUROPEO DE DESENVOLVEMENTO REXIONAL (FEDER) como parte da resposta da Unión á pandemia da COVID-19

Tabla de contenidos

1. Introducción	6
2. Infraestructura del CESGA.....	8
2.1 Finisterrae III.....	10
2.2 Entorno cloud	14
2.3 Sistema Big Data.....	15
2.4 Ordenador cuántico	17
3. Análisis de requerimientos	21
3.1 Control de acceso.....	21
3.2 Gestión de colas.....	23
3.3 Interfaces de comunicación	24
3.4 Integración con recursos externos	27
3.5 Representación de los circuitos cuánticos.....	28
3.6 Representación de los resultados.....	30
3.7 Gestión de la contabilidad	32
4. Conclusiones	36

Lista de figuras

Figura 1: Finisterrae III.....	11
Figura 2: Integración del nodo QLM en Finisterrae III.....	18
Figura 3: Visión global de la oferta de Fujitsu.	20
Figura 4: Plataforma del Finisterrae III.....	21
Figura 5: Arquitectura de una instalación de GridWay Metascheduler en dos entornos.....	24
Figura 6: Representación de las distintas capas del Modelo OSI	26
Figura 7: Pila de llamadas de del controlador PCI de la QPU.	27
Figura 8: Diagrama de integración con recursos de computación clásica y cuántica de proveedores externos (ejemplo de integración con la plataforma de AWS).	28
Figura 9: Representación de un circuito cuántico.....	29
Figura 10: Representación de un circuito mediante "quantum circuit cutting"	29
Figura 11: Pantalla de trabajo del Logisim.	30
Figura 12: Orquestación de tareas mediante grafos.	30
Figura 13: Ejemplo de representación de los resultados de un circuito cuántico.	31
Figura 14: Métodos alternativos de representación de la información.....	31

Lista de tablas

Tabla 1: Capacidad de almacenamiento (TB) desglosado por tipo. 9

Tabla 2: Relación de la evolución del input de cúbits con la salida en bits. 25

Lista de acrónimos

- CPD *Centro de procesamiento de datos*
- DAG *Grafo acíclico dirigido*
- GPU *Unidad de procesamiento gráfico*
- HPC *Computación de alto rendimiento*
- ICTS *Instalación científico técnica singular*
- PaaS *Plataforma como servicio*
- QPU *Unidad de proceso cuántica*
- QRNG *Generador cuántico de números aleatorios*
- QVM *Máquina virtual cuántica*
- RES *Red española de supercomputación*
- SSH *Shell segura*
- VM *Máquina virtual*
- VNC *Computación virtual en red*

1. Introducción

La computación cuántica constituye una de las disciplinas más disruptivas del siglo XXI. A diferencia de la computación clásica que utiliza los bits para representar la información, en computación cuántica se utilizan los **bits cuánticos** o **cúbits**. Estos pueden estar en múltiples estados de forma simultánea, lo que permite realizar cálculos de una manera mucho más eficiente y rápida. No obstante, a pesar de los significativos avances logrados en este ámbito, esta disciplina se encuentra aún en sus primeras etapas y todavía son muchos los desafíos que deben superarse para que la computación cuántica pueda alcanzar todo su potencial.

Uno de los principales desafíos de la computación cuántica es la **escalabilidad**. La mayoría de las computadoras cuánticas actuales solo pueden manejar un número reducido de cúbits, lo que constituye una importante limitación para resolver problemas complejos que pueden requerir cientos o incluso miles de cúbits. Sin embargo, la construcción de una computadora cuántica escalable es una tarea ardua, ya que a medida que el número de cúbits aumenta, también lo hace la complejidad del sistema. Además, los cúbits son extremadamente **sensibles** a su entorno (temperatura, radiación, interferencias electromagnéticas, etc.), por lo que cualquier interferencia externa puede afectar a su estado, provocando que los resultados obtenidos sean inexactos. Por ello, son muchos los investigadores que están trabajando en la identificación de materiales y diseños que permitan mejorar la estabilidad de los cúbits para la **corrección de errores**. Dado que los sistemas cuánticos son propensos a errores debido a la naturaleza probabilística intrínseca de los cúbits, es esencial desarrollar técnicas que permitan detectar y corregir errores en el procesamiento de la información. El control de los errores es un campo de investigación activo y en constante evolución. Su objetivo se centra mitigar efectos indeseados y garantizar una buena precisión. Otro aspecto importante a tener en cuenta para garantizar sistemas cuánticos robustos y confiables es el **ruido** y la **decoherencia**. La decoherencia es el proceso en el que los cúbits pierden su capacidad cuántica y se comportan como bits clásicos. Este comportamiento indeseado puede ser causado por factores externos, como la interferencia electromagnética, o internos como la interacción entre cúbits.

Otro desafío importante es la necesidad de **desarrollo de algoritmos y programas específicos**. A diferencia de la computación clásica que se basa en la lógica booleana, la computación cuántica usa algoritmos cuánticos que son significativamente distintos tanto en términos de complejidad como de

funcionalidad y, aunque en los últimos años se han conseguido importantes avances, la mayoría de los algoritmos cuánticos se encuentran aún en una etapa muy temprana de desarrollo y requieren de un importante conocimiento de la física cuántica para su desarrollo y aplicación. Por otra parte, la representación de circuitos cuánticos con un gran número de cúbits presenta también un desafío importante. A medida que aumenta el número de cúbits, la dimensión del espacio de Hilbert correspondiente aumenta de forma exponencial. Así, por ejemplo, un sistema de 41 cúbits tendría asociado un espacio de Hilbert de 2^{41} estados cuánticos posibles. Esto hace que sea imposible representar el estado completo del sistema de forma explícita.

La forma más común de representar los circuitos cuánticos es mediante diagramas. En ellos, los cúbits se representan con líneas horizontales y las puertas cuánticas mediante cajas. Aunque se trata de un tipo de representación intuitiva para circuitos cuánticos de pocos cúbits, esta se vuelve impracticable cuando el número de cúbits es elevado. En este caso, los diagramas requieren una cantidad significativa de espacio y se vuelven complejos para la interpretación y depuración. Para superar estos desafíos, los investigadores están desarrollando técnicas y herramientas de visualización avanzadas que faciliten su comprensión (por ejemplo, mediante técnicas de visualización interactivas).

Además de todos estos retos técnicos, también existen importantes desafíos sociales y éticos asociados a la computación cuántica. Uno de los principales problemas es la **seguridad de la información**. A medida que la computación cuántica se vuelve más poderosa, también lo hacen las amenazas a la seguridad. Los algoritmos cuánticos pueden romper fácilmente muchos de los sistemas criptográficos actuales. Por ello, son necesarias nuevas soluciones de seguridad que permitan proteger la información confidencial. Otro aspecto a tener en cuenta también es la **brecha digital**. Dado que la computación cuántica es una disciplina todavía emergente, su costo es muy elevado y requiere de recursos especializados tanto para su desarrollo como su uso. Esto puede producir que determinados países u organizaciones acaben monopolizando el desarrollo de tecnologías cuánticas, produciendo importantes desigualdades tanto en la economía global como en la sociedad en general.

En las siguientes secciones se realiza un análisis de requisitos de los principales desafíos y retos a abordar en la integración de una computadora cuántica en una infraestructura de supercomputación como la del Centro de Supercomputación de

Galicia (CESGA). Para poder ofrecer una propuesta de valor es preciso, en primer lugar, tener un conocimiento detallado de las infraestructuras de supercomputación y sus necesidades, tanto actuales como futuras. Por este motivo, en la Sección 2 se realiza una descripción detallada de la infraestructura del CESGA, incluyendo sus principales entornos (el Supercomputador Finisterrae III, el entorno cloud, el sistema big data y el ordenador cuántico). Por otra parte, en la Sección 3 se recogen los principales desafíos a tener en cuenta en la integración de una computadora cuántica en una infraestructura como la del CESGA. Más concretamente, se realiza un análisis exhaustivo de aspectos claves como el control de acceso, la gestión de colas, las interfaces de comunicación, la integración con recursos externos, la gestión de la contabilidad y la representación tanto de los circuitos cuánticos como de los resultados. Finalmente, en la Sección 4 se recopilan las principales conclusiones del estudio.

2. Infraestructura del CESGA

El CESGA es un referente de supercomputación reconocido y consolidado tanto en España como a nivel internacional. Actualmente ya cuenta con una infraestructura avanzada de gran capacidad de cálculo que la hace atractiva para resolver problemas complejos en el ámbito científico. Además, bajo la iniciativa del Polo de Tecnologías Cuánticas, el CESGA está centrando los esfuerzos para posicionar a Galicia como un referente en computación cuántica.

La infraestructura del CESGA se compone de varios entornos que, de forma conjunta, se articulan para ofrecer a la comunidad científica una plataforma de cómputo de alto rendimiento y servicios de computación avanzada. Más concretamente, la infraestructura del CESGA cuenta con:

- **Supercomputadores:** servidores de última generación y diferentes arquitecturas capaces de llevar a cabo cálculos complejos y de alta dimensión.
- **Red de almacenamiento:** red de almacenamiento de alta velocidad que permite a los usuarios almacenar grandes volúmenes de datos.
- **Red de comunicaciones:** red de comunicaciones de alta velocidad y baja latencia que interconecta los diferentes elementos de la infraestructura.

- **Computación cuántica:** a finales del 2021 se ha desplegado en el Finisterrae III la primera infraestructura cuántica del CESGA, un emulador de circuitos cuánticos de 30 cúbits.

Los principales sistemas de computación del CESGA incluyen: el superordenador Finisterrae, la plataforma cloud y el entorno Big Data. El Finisterrae es el nombre genérico para designar las distintas generaciones de superordenadores del centro (en marzo de 2023 se apagó el Finisterrae II, por lo que la versión actual del superordenador es Finisterrae III). Este superordenador está especialmente pensado para ejecutar simulaciones u operaciones matemáticas complejas de altas prestaciones sobre grandes conjuntos de datos en paralelo o que requieran del uso de GPUs. La plataforma cloud proporciona un entorno flexible para la virtualización de aplicaciones, simulaciones bajo demanda o que requieran cierta interacción con el usuario. Finalmente, el entorno Big Data permite el procesado de grandes volúmenes de información en paralelo mediante el uso de herramientas específicas del ámbito (por ejemplo, Hadoop).

Por otra parte, el CESGA dispone de diferentes soluciones de almacenamiento para poder satisfacer las crecientes necesidades de la comunidad científica. En la Tabla 1 se detalla la capacidad de almacenamiento por tipo¹. A nivel global, la capacidad de almacenamiento disponible es de 26.186 TB.

TIPO DE ALMACENAMIENTO	CAPACIDAD (TB)
BIG DATA	816
SCRATCH	370
PARALLEL SCRATCH	1.000
HOME + PERMANENT DISK STORAGE	4.000
TAPE	20.000
ALMACENAMIENTO TOTAL	26.186

Tabla 1: Capacidad de almacenamiento (TB) desglosado por tipo.

¹ Almacenamiento CESGA. Consultado con fecha de 22/03/2023 en <https://www.cesga.es/infraestructuras/almacenamiento/>

2.1 Finisterrae III

El supercomputador **Finisterrae III** posee una capacidad de cálculo 12 veces mayor a la de su predecesor, proporcionando una potencia de cálculo de 4.36 PetaFLOPS que permiten resolver 4 mil billones de operaciones matemáticas por segundo. Al igual que en el caso de sus predecesores, el Finisterrae III se ha incorporado al conjunto de infraestructuras que componen la Instalación Científico Técnica Singular (ICTS) de la Red Española de Supercomputación (RES).

El Finisterrae III está compuesto por 357 nodos interconectados a través de una red Infiniband HDR 100². A nivel global, el Finisterra III está compuesto por:

- 714 procesadores Intel Xeon Ice Lake 8352Y de última generación con 32 cores a 2.2 Ghz.
- 22.848 núcleos de procesado.
- 157 aceleradoras GPUs (141 Nvidia A100 y 16 Nvidia T4).
- 126 TB memoria agregada.
- 355 TB de capacidad de almacenamiento en discos en estado sólido SSD NVMe.
- Red Infiniband HDR 100 network.
- Potencia de cómputo de 4.36 PetaFLOPS.

La Figura 1 recoge las principales características que conforman el supercomputador de una forma más detallada por grupos de nodos de cómputo.

² Finisterrae III. Consultado con fecha de 20/03/2023 en <https://cesga-docs.gitlab.io/ft3-user-guide/overview.html>

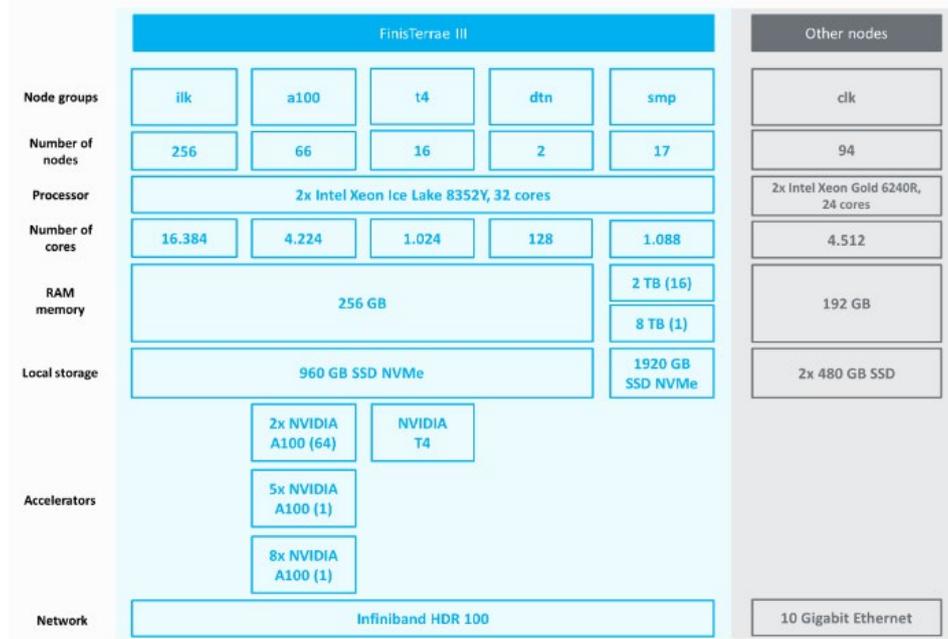


Figura 1: Finisterrae III.

El superordenador Finisterrae III está compuesto por:

- 256 nodos ILK: 16.384 cores.
 - 2x Intel Xeon Ice Lake 8352Y con 32 cores cada uno (64 cores por nodo).
 - 256 GB de memoria RAM (247 GB para uso real).
 - 960 GB SSD NVMe de almacenamiento local.
- 64 nodos A100 con aceleradoras GPU: 4096 cores.
 - 2x Intel Xeon Ice Lake 8352Y con 32 cores cada uno (64 cores por nodo).
 - 256 GB de memoria RAM (247 GB para uso real).
 - 960 GB SSD NVMe de almacenamiento local.
 - 2x GPUs NVIDIA A100.
- 1 nodo A100 con 5 aceleradoras GPUs: 64 cores.
 - 2x Intel Xeon Ice Lake 8352Y con 32 cores cada uno.
 - 256 GB de memoria RAM (247 GB para uso real).
 - 960 GB SSD NVMe de almacenamiento local.

- 5x GPUs NVIDIA A100.
- 1 nodo A100 con 8 aceleradoras GPUs: 64 cores.
 - 2x Intel Xeon Ice Lake 8352Y con 32 cores cada uno.
 - 256 GB de memoria RAM (247 GB para uso real).
 - 960 GB SSD NVMe de almacenamiento local.
 - 8x GPUs NVIDIA A100.
- 16 nodos T4 con aceleradoras GPU: 1024 cores).
 - 2x Intel Xeon Ice Lake 8352Y con 32 cores cada uno.
 - 256 GB de memoria RAM (247 GB para uso real).
 - 960 GB SSD NVMe de almacenamiento local.
 - 1x GPU NVIDIA T4.
- 2 nodos DTN: 128 cores.
 - 2x Intel Xeon Ice Lake 8352Y con 32 cores cada uno.
 - 256 GB de memoria RAM (247 GB para uso real).
 - 960 GB SSD NVMe de almacenamiento local.
- 16 nodos SMP: 1024 cores.
 - 2x Intel Xeon Ice Lake 8352Y con 32 cores cada uno.
 - 2TB de memoria RAM (2011GB para uso real).
 - 1920GB SSD NVMe de almacenamiento local.
- 1 nodo SMP: 64 cores.
 - 2x Intel Xeon Ice Lake 8352Y con 32 cores cada uno.
 - 8TB de memoria RAM (7975GB para uso real).
 - 1920GB SSD NVMe de almacenamiento local.

Además, se han recuperado también un conjunto de nodos de los clusters previos. En este caso, los nodos se conectan a través de una red 10 Gigabit Ethernet. Más concretamente, los nodos que se han recuperado son:

- 94 nodos CLK: 4512 cores.
 - 2 Intel Xeon Gold 6240R (Cascade Lake) con 24 cores (48 cores por nodo).

- 192GB de memoria RAM (180GB para uso real).
- 2x 480GB SSD de almacenamiento local.
- 20 de esos nodos tienen especial prioridad, así que no están disponibles todo el tiempo para uso general.
- 18 nodos AMD EPYC: 1152 cores.
 - 2x AMD EPYC 7452 @ 2.35GHz con 32 cores (64 cores por nodo).
 - 256GB de memoria RAM.
 - 2TB HDD de almacenamiento local.

Para garantizar la seguridad, el acceso a los servidores del Finisterrae III está acotado a un conjunto de centros autorizados. Fuera de estos centros autorizados, el acceso debe realizarse a través de una VPN o mediante la terminal SSH (Secure Shell) que se ofrece en la Sección de Herramientas dentro del Portal de Usuario para securizar y encriptar las comunicaciones. Para la transferencia de datos con el Finisterrae III es necesario un cliente *scp* o *sftp* (o el comando *rsync*). En el caso de querer transferir grandes volúmenes de datos (GB o incluso TB), el supercomputador cuenta con 2 nodos especiales DTN para agilizar el proceso.

En el Finisterrae III, existen dos tipos de almacenamiento: uno permanente y otro temporal. A su vez, el sistema de almacenamiento permanente se divide en 3 particiones principales: HOME, STORE y LUSTRE. La partición HOME es de 10 GB y sirve para almacenar ficheros de código. Por otra parte, las particiones de STORE y LUSTRE permiten almacenar los resultados de las simulaciones y poseen unas capacidades de 500 GB y 3 TB, respectivamente. La velocidad de acceso es baja en los dos primeros casos (HOME y STORE) y alta en el caso de LUSTRE.

Para abordar procesos de cómputo complejos e intensos, el Finisterrae III pone a disposición de la comunidad científica diferentes estrategias para paralelizar los procesos: paralelismo en un único nodo (OpenMP), multi-nodo (MPI) o híbrida (MPI/OpenMP).

El sistema de colas del Finisterrae III está basado en SLURM, un sistema de gestión de tareas altamente escalable y tolerante fallos para clusters Linux. Entre sus principales funcionalidades destacan:

- Asignar acceso exclusivo y/o no exclusivo a los recursos durante un periodo de tiempo.

- Proporciona un marco para iniciar, ejecutar y monitorear trabajos en los nodos asignados.
- Arbitra la gestión de los recursos en base a la cola de trabajos pendiente.

Para el envío de trabajos al CESGA es obligatorio especificar la memoria y el tiempo requerido.

2.2 Entorno cloud

El CESGA cuenta también con un **entorno cloud** compuesto por un conjunto de servidores y sistemas de almacenamiento, gestionados a través del software OpenNebula, y dedicados a prestar una infraestructura de computación virtual configurable para la investigación e innovación. Se trata de un sistema versátil que cuenta con una línea de comandos orientada a los usuarios más avanzados y también una interfaz gráfica para facilitar el uso a perfiles de usuarios con conocimientos más básicos en informática. Con este sistema se pretende atender a las crecientes demandas de la comunidad científica y, muy especialmente a aquellas que requieran de cálculos complejos y de alta productividad que puedan ser fácilmente distribuidos entre un conjunto de nodos. A nivel global, la infraestructura cloud³ está compuesta por:

- 6240 núcleos de proceso.
- 24 TB de memoria agregada.
- 3 PB de capacidad de almacenamiento.

Más concretamente, el clúster está compuesto por un conjunto de 114 nodos de computación con 48 núcleos de proceso. Cada nodo posee 192 GB de memoria y 960 GB de almacenamiento en disco. Adicionalmente, cuenta con 16 nodos de almacenamiento con 48 núcleos de proceso. Cada nodo cuenta con 192 GB de memoria y 193 TB de almacenamiento en disco.

Buena parte de la infraestructura cloud está integrada en un entorno de cálculo común con el supercomputador Finisterrae. Esta fórmula dota a la infraestructura de una gran flexibilidad, permitiendo que los usuarios puedan delegar los trabajos computacionalmente exigentes en máquinas desplegadas en el entorno CLOUD-HPC-CESGA mientras mantienen sus equipos locales para ejecutar otros trabajos

³ Infraestructura de servicios cloud. Consultado con fecha de 20/03/2023 en <https://www.cesga.es/cesga-renueva-infraestructura-de-servicios-cloud/>

más livianos de forma simultánea. Por defecto, los recursos máximos para proyectos de usuarios de HPC son:

- 4 instancias.
- 8 CPUs.
- 8 GBs de memoria RAM.
- 4 volúmenes.
- 12 snapshots de los volúmenes.
- 500 GB de almacenamiento en disco.
- 1 dirección IP pública flotante.
- 10 grupos de seguridad.
- 100 reglas de grupos de seguridad.
- 100 redes.
- 500 puertos.
- 10 routers.

No obstante, estos recursos máximos se pueden dilatar en base a las exigencias del proyecto si se realiza una petición formal al Departamento de Sistemas del CESGA. Antes de desplegar una máquina virtual (VM por sus siglas en inglés), es necesario realizar unos pasos de configuración previos. Más concretamente, se necesita:

- Crear una clave pública.
- Establecer los grupos de seguridad de la VM.
- Crear y añadir los volúmenes de almacenamiento de la VM.

2.3 Sistema Big Data

La infraestructura Big Data del CESGA cuenta con dos entornos para el procesamiento de grandes volúmenes de datos en paralelo:

- Entorno CDH 6.1.1: plataforma basada en Hadoop 3.
- Entorno Platform as a service (PaaS): plataforma que ofrece una mayor versatilidad para aquellos casos en los que se deseen herramientas alternativas a las ofrecidas en el ecosistema Hadoop, como pueden ser, por ejemplo, Cassandra, MongoDB, Redis, etc.

Este sistema está compuesto por un clúster dedicado con las siguientes características:

- 38 nodos de los cuales 4 de ellos son nodos maestros y 34 nodos esclavo.
- Capacidad bruta de 816 TB.
- Capacidad de E/S agregada de más de 30 GB/s.
- Conexión de 10 GbE entre los nodos del clúster.

Se trata de un entorno versátil pensado tanto para perfiles investigadores y científicos con conocimientos básicos en informática como usuarios avanzados. Para conectarse al sistema Big Data, es necesario estar conectado a la VPN o conectarse a través de un escritorio remoto. El cluster dispone de varias interfaces web para conectarse al cluster Hadoop (BD|CESGA WebUI) y realizar algunas de las tareas más comunes como enviar y monitorear trabajos, ejecutar consultas en Hive o explorar el sistema de ficheros HDFS (HUE). Esta interfaz sirve como complemento y está especialmente pensada para aquellos usuarios que no tengan tanta soltura con la línea de comandos.

Para cargar datos en el entorno Big Data existen varias opciones:

- *scp* para cantidades pequeñas de datos (< 10 GB).
- Servicio *GridFTP* de Globus para grandes cantidades de datos.

En cualquiera de los casos, se recomienda el uso de un servidor DTN para transferir los datos ya que mejora notablemente el rendimiento. Para la gestión de los recursos en el ecosistema Hadoop se usa YARN. Este sistema juega un papel análogo a SLURM en el contexto del Supercomputador Finisterrae III. YARN también posee una interfaz web (YARN Web UI), desde la que es fácil hacer un seguimiento del estado y progreso de las tareas enviadas. Por defecto, los trabajos enviados entrarán en un sistema de colas que gestionará las prioridades del resto de usuarios mediante un reparto justo en base a diferentes características del trabajo como la capacidad de cómputo o la cantidad de memoria requeridas. No obstante, existen ciertas peculiaridades como el caso de los trabajos interactivos (por ejemplo, Jupyter Notebooks) o los trabajos pequeños urgentes que se ejecutan en colas dedicadas.

El entorno Big Data del CESGA pone a disposición de los usuarios un conjunto de herramientas abarcando las alternativas más representativas para el procesamiento de grandes volúmenes de datos. Entre ellas se incluyen:

- **Sistema de ficheros de Hadoop (HDFS).**
- **Spark.** Se trata probablemente de una de las herramientas más populares del ecosistema Hadoop. Proporciona importantes mejoras de rendimiento sobre los trabajos estándar de MapReduce y simplifica el proceso de desarrollo al ofrecer APIs de alto nivel como la API de DataFrames. Actualmente, la versión disponible es Spark 2.4.0⁴ y se puede lanzar a través de una shell o a través de sesiones interactivas tanto de Python (preferentemente a través de Anaconda) como de R (Sparklyr).
- **Jupyter Notebooks.** Los notebooks de Jupyter ofrecen una interfaz web atractiva para el desarrollo y ejecución de aplicaciones. El entorno Big Data dispone JupyterLab, un entorno de desarrollo que permite tanto la ejecución de notebooks de Jupyter como proyectos Python convencionales.
- **Hive.** Permite realizar consultas SQL a datos en formato HDFS. Para ello, internamente Hive se encarga de traducir las sentencias SQL a tareas MapReduce.
- **Impala.** Sigue una filosofía similar a Hive. La principal diferencia radica en que el tiempo de inicio de Impala es mucho más rápido por lo que es altamente recomendable para realizar pequeñas consultas interactivas de manera ágil. Por la contra, para grandes consultas, es más recomendable usar Hive.
- **Sqoop.** Permite importar los datos de un modelo relacional a formato HDFS. El CESGA dispone de diferentes conectores incluyendo MySQL / MariaDB, PostgreSQL, Microsoft SQL Server u Oracle 18c.

Además, aunque no se ofrece de forma predeterminada, el CESGA dispone de servicios de software adicionales que los usuarios pueden usar si así lo requieren. Entre ellos se incluyen Anaconda, Maven o sbt.

2.4 Ordenador cuántico

Desde finales del 2021, y como fruto de la iniciativa del Polo de Tecnologías cuánticas, el CESGA ha incorporado a la infraestructura del Finisterrae III el primer **emulador cuántico** instalado en España, un Atos QLM con el que se pretende

⁴ Infraestructura de servicios Big Data. Consultado con fecha de 28/03/2023 en <https://cesga-docs.gitlab.io/bigdata-user-guide/index.html>

posicionar al CESGA como un referente en el ámbito de la computación cuántica. El equipamiento está diseñado para investigar y explorar aplicaciones que requieran de la generación de números aleatorios (ver Figura 2 para más detalles).

	Thin nodes	GPU nodes	Vis nodes	Transfer nodes	Fat nodes	OLM node
CANTIDADE	256	66	16	2	17	Simulador de circuitos cuánticos Atos 30 qbits
PROC.	2x Intel Xeon Ice Lake 8352Y 32 cores 2.2 GHz					
MEMORIA	256 GB				2 TB (16) 8 TB (1)	
DISCO LOCAL	960 GB SSD NVMe				1920 GB SSD NVMe	
GPU	2x Nvidia A100 (64)		Nvidia T4			
	5x Nvidia A100 (1) + 256 GB RAM					
	8x Nvidia A100 (1) + 256 GB RAM					
REDE	Infiniband HDR 100					

Figura 2: Integración del nodo QLM en Finisterrae III.

Entre los potenciales casos de uso se tienen:

- Soluciones de ciberseguridad.
- Criptografía y soluciones de firma electrónica.
- Soluciones científicas que requieran de técnicas de generación de números aleatorios como los algoritmos de Monte Carlo, algoritmos de optimización, etc.
- Soluciones que requieran de generación de números aleatorios de manera fiable e impredecible.

En enero de 2023 se ha actualizado la versión de este Generador Cuántico de Números Aleatorios (QRNG por sus siglas en inglés) para mejorar el rendimiento de las aplicaciones de usuario.

El QRNG instalado es el modelo FMC 400 y entre sus principales características se incluyen:

- Entropía típica de al menos 0.94.
- Tasa de aleatoriedad sin procesar de 400 Mb/s.

- Tasa de salida mínima de bits aleatorios de 100 Mb/s.
- Actualmente no hay limitaciones del número máximo de números aleatorios que se pueden solicitar más allá de las propias limitaciones impuestas por la máquina de host.

El sistema está instalado en el centro de datos del CESGA y está conectado a la red Ethernet. El acceso se puede realizar mediante APIs (Python y C). Actualmente, el QRNG tiene una restricción del número de usuarios, siendo posible únicamente **un cliente de forma simultánea**.

Para la gestión del QLM se siguen las mismas políticas que las de las infraestructuras de computación y almacenamiento. En estos momentos el QLM se encuentra en un ambiente aislado dedicado a investigación y, cualquier investigador público o privado que lo desee, puede emitir una solicitud incluyendo una pequeña memoria técnica describiendo el propósito de la investigación. Por otra parte, debido a las limitaciones y novedad de la infraestructura del emulador cuántico, la duración total de acceso no podrá superar los 3 meses. El acceso al servicio será gratuito hasta 2023 (incluido) y para la asignación del acceso, se priorizarán las solicitudes de las entidades afiliadas con ubicación en Galicia.

Además, en el marco del Polo de Tecnologías Cuánticas de Galicia, el CESGA ha contratado a Fujitsu, proveedor del equipo cuántico Oxford Quantum Circuits, uno de los primeros **ordenadores cuánticos** de Europa y el más potente de España hasta la fecha por 13,9 millones de euros⁵. Más concretamente, la oferta presentada por Fujitsu (ver Figura 3 para más detalles⁶), incluye:

- Un **computador cuántico**.
- Un **computador de altas prestaciones** (HPC).
- Un **emulador cuántico**.
- Un **sistema de almacenamiento** para alojar los resultados de los algoritmos cuánticos para su posterior análisis y validación.

⁵ Ordenador cuántico CESGA. Consultado con fecha de 28/03/2023 en <https://www.cesga.es/computador-cuatico-de-32-cubits-para-el-cesga/>

⁶ Visión global de la oferta de Fujitsu. Consultado con fecha de 20/04/2023 en <https://www.cesga.es/download/dixitos-abril-2023/>

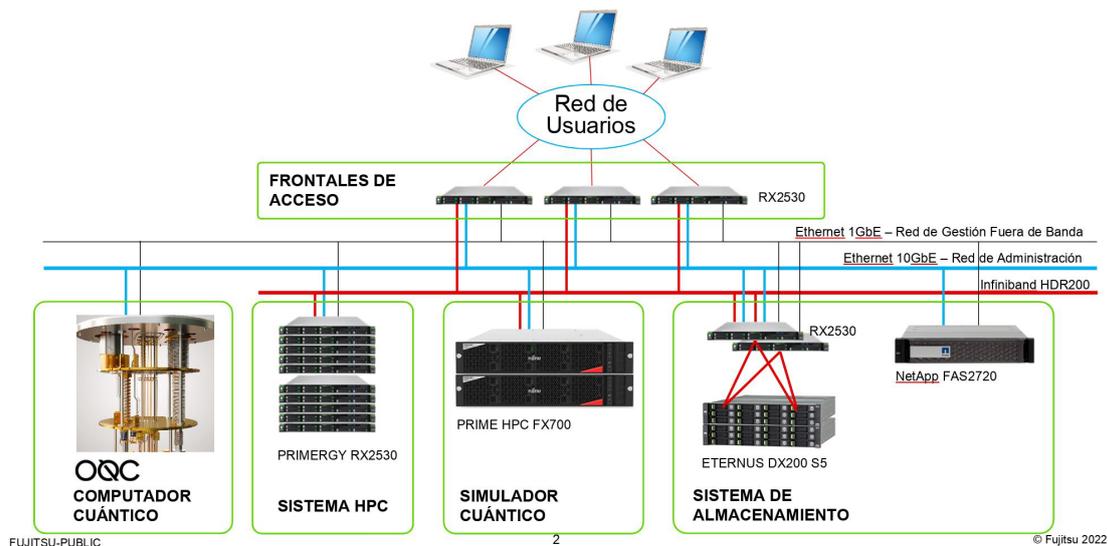


Figura 3: Visión global de la oferta de Fujitsu.

El ordenador cuántico tendrá 32 cúbits y se instalará a finales del 2023, permitiendo a los investigadores, tanto públicos como privados, estudiar las ventajas de la computación cuántica en diferentes áreas como la física, la química, la biología, la inteligencia artificial, la investigación operativa o la criptografía.

Con esta inversión del gobierno gallego, se busca posicionar a Galicia en la vanguardia de la computación cuántica, fomentar la innovación y atraer el talento investigador. La oferta presentada por Fujitsu, además del ordenador cuántico de 32 cúbits basado en la tecnología de *superconductores*, incluye una serie de infraestructuras complementarias y un plan de colaboración que facilitará el impulso del Polo de Tecnologías Cuánticas de Galicia mediante la creación del Centro Internacional de Excelencia en Tecnologías Cuánticas en la Comunidad.

Se trata pues de un sistema flexible compuesto por diferentes piezas en las que el ordenador cuántico constituye el elemento central, y donde éste se integra y comunica con los otros elementos que componen el entorno para hacer de esta nueva infraestructura cuántica un referente tanto nacional como internacional.

Fujitsu colaborará con las universidades gallegas para la formación y retención del talento, desarrollando planes de formación específicos orientados tanto a universidades como a empresas y elaborando proyectos de I+D en el ámbito de la computación cuántica conjuntamente con otras entidades gallegas. Además, también se impulsarán bolsas formativas y se llevará a cabo un proyecto de consultoría para desarrollar casos de uso con empresas e identificar los sectores que más se pueden beneficiar de esta tecnología.

3. Análisis de requerimientos

3.1 Control de acceso

Para garantizar la seguridad, el acceso a los servidores de un supercomputador como el del CESGA está limitado a un subconjunto de centros autorizados entre los que se encuentran los grupos de investigación de las tres universidades gallegas y del CSIC u otras empresas o instituciones con acuerdos especiales. El acceso fuera de estos centros autorizados debe realizarse mediante VPN.

Dada la alta demanda de recursos del CESGA, especialmente a los sistemas de computación cuántica, la organización necesita un sistema que permita controlar tanto el acceso de los usuarios como los recursos mencionados. Por ello, para poder acceder a los diferentes sistemas de computación cuántica, es preciso ampliar y/o actualizar los sistemas de *login* internos.

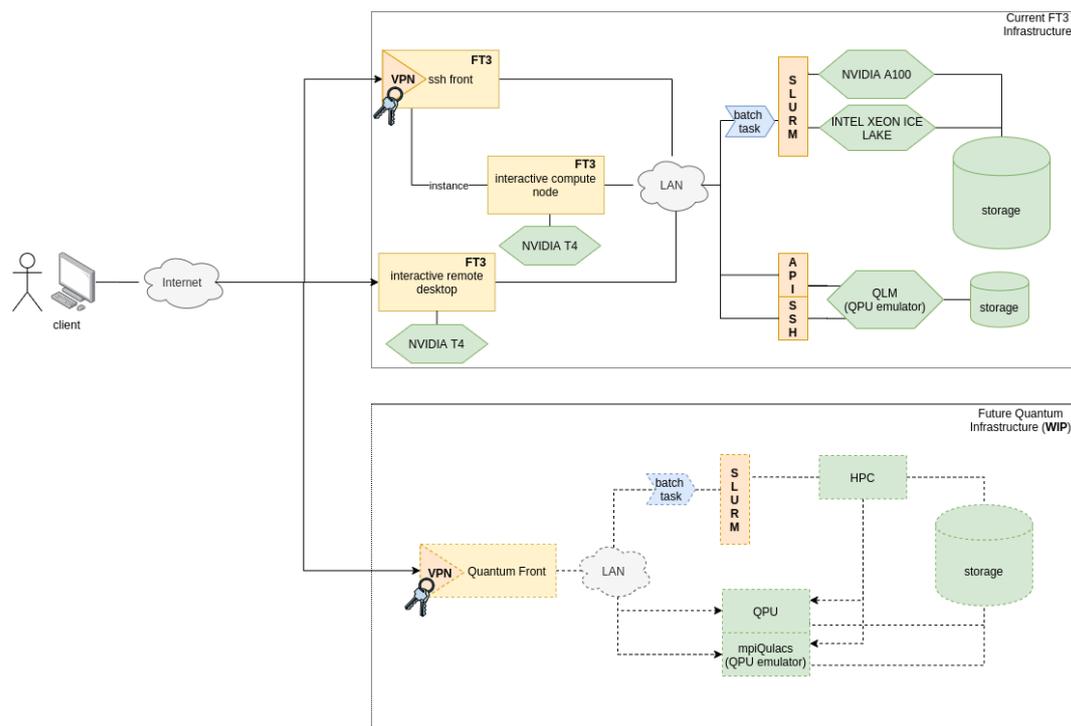


Figura 4: Plataforma del Finisterrae III.

Actualmente, el CESGA dispone de dos maneras de ingresar dentro de sus sistemas (ver Figura 4 para más detalles):

- **Terminal SSH.** Se puede acceder a una terminal SSH desde la sección de herramientas dentro del portal de usuario que abre una línea de comandos directamente en el Finisterrae III.
- **Remote desktop (VNC).** Proporciona un escritorio Linux con acceso directo a los directorios y terminal de comandos del Finisterrae III. Estos escritorios remotos se destruyen pasadas las 36 horas, perdiendo cualquier dato que no se haya guardado, a menos que se cierre la sesión en el escritorio remoto antes de alcanzar el límite o se extienda la duración.

El CESGA ofrece 4 formas de acceder a su equipamiento en base a los recursos requeridos y el nivel de interactividad deseado. Más concretamente, las alternativas disponibles son:

- **Uso interactivo:** Proporciona acceso a un entorno interactivo con recursos limitados (memoria virtual máxima de 8 GB) que, aunque es insuficiente para realizar simulaciones complejas, si es válido para trabajar con ficheros e interactuar con el sistema.
- **Remote desktops:** Proporcionan acceso a las aplicaciones mediante una interfaz gráfica. La memoria máxima disponible en este caso es de 32 GB.
- **Nodos interactivos dedicados:** Los usuarios pueden iniciar sesiones interactivas con recursos dedicados. Los recursos máximos disponibles son 64 cores y 247 GB de RAM durante un periodo máximo de 8 horas.
- **Sistema de colas:** Están pensadas para lanzar simulaciones pesadas que requieren múltiples cores y cantidades importantes tanto de memoria como de tiempo. En este caso es obligatorio pasar por el sistema de colas para garantizar un reparto eficiente de los recursos para cada usuario.

El acceso a los recursos actuales no debería ser un limitante a la hora de solicitar los recursos cuánticos disponibles. Por otra parte, desde estas interfaces no se tiene acceso a los recursos cuánticos, ya sean emuladores o el procesador cuántico (QPU, por sus siglas en inglés). Estos solo son accesibles cuando se emiten trabajos en *batch*, en sus distintas modalidades, tal y como se ve en el diagrama de la Figura 4.

3.2 Gestión de colas

El acceso a los distintos recursos de computación que proveen los centros de supercomputación como el CESGA, ya sean QPUs, emuladores o los recursos de computación clásica, los usuarios tienen que emitir una solicitud mediante un sistema de colas para poder orquestar y optimizar el tiempo de los recursos.

Para el cumplimiento de este requerimiento, el CESGA dispone de un sistema de gestión de recursos enfocado a los entornos clúster, SLURM (Simple Linux Utility for Resource Management), de esta forma, los usuarios aplican, desde unos nodos de *login*, la ejecución de sus tareas bajo una serie de premisas (e.g., tiempo máximo de ejecución de dicha tarea) y un entorno determinado: lenguaje, librerías requeridas, el posible uso de aceleradores (GPUs, coprocesadores, etc.).

Con la aparición de estas nuevas tecnologías de aceleración, se tienen que incorporar nuevas restricciones y reglas para la emisión de trabajos cuánticos. Además, se deben tener en cuenta una serie de casuísticas que se pueden dar en este nuevo paradigma:

- Ejecución sobre un emulador/QPU.
- Tipo de emulador y/o QPU (dados los niveles de ruido).

Además de las posibilidades anteriormente citadas, se debe contemplar que el circuito cuántico se ejecute en la infraestructura del centro de supercomputación o fuera de ella. Por tanto, se tienen que estudiar e integrar los sistemas de gestión de recursos con proveedores de recursos cuánticos externos, como por ejemplo AWS Braket o Azure Quantum. Para acometer este otro requerimiento se contempla el estudio de herramientas de *metascheduling* como por ejemplo GridWay Metascheduler⁷.

⁷ GridWay Metascheduler. Consultado con fecha de 20/04/2023 en <https://gridway.ucm.es/doku.php>

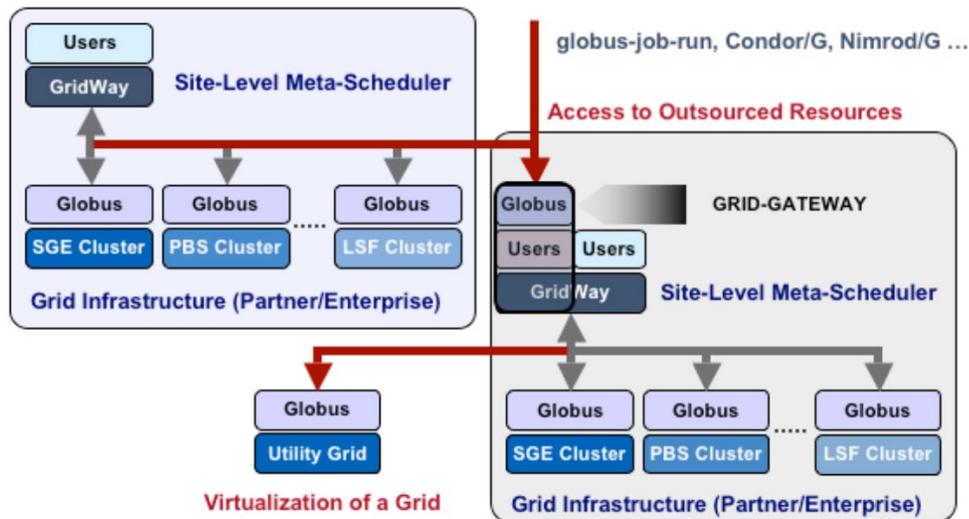


Figura 5: Arquitectura de una instalación de GridWay Metascheduler en dos entornos.

Tal y como se ve en la Figura 5, esta clase de software permite poner a disposición de los usuarios los recursos de diferentes infraestructuras. Esta aproximación sería de utilidad para realizar análisis de benchmarking entre las diferentes tecnologías de computación cuántica o en caso de formar parte de la red Quantum Spain y querer integrarse con otros Centros de Procesamiento de Datos (CPD). No obstante, un caso de estudio sería la integración con proveedores cloud tales como: AWS, Azure, etc., donde además del balanceo de la carga de cómputo, se tiene que poder delimitar los recursos a los que puede acceder el usuario final y el tiempo/coste de dichos servicios.

En caso de que este tipo de herramientas no se adapten a los diversos productos de computación cuántica que hay en el mercado, se tendrá que realizar un estudio de los requisitos y un análisis técnico sobre productos software que permitan orquestar y gestionar los distintos recursos cuánticos que se deban tener en cuenta.

3.3 Interfaces de comunicación

Los protocolos de comunicación entre sistemas de computación clásicos (SOAP, REST, RPC, etc.) no son fácilmente adaptables al paradigma de la computación cuántica. A medida que esta nueva disciplina evoluciona, la complejidad de los

algoritmos aumenta, requiriendo un número mayor de entrada de datos, y, por ende, de salida.

La medición de los resultados en los circuitos cuánticos posee un comportamiento exponencial en base al número de cúbits, lo que imposibilita el uso de los protocolos previamente citados para el intercambio de información entre los entornos de usuarios y los sistemas QPU. Esta problemática queda patente en la Tabla 2 en donde se recoge el tamaño de la respuesta en función del número de cúbits. La tabla sugiere que a partir de 40 cúbits, el tamaño de la respuesta se vuelve inmanejable desde un ordenador convencional.

NÚMERO CÚBITS	TAMAÑO RESPUESTA (BITS)	TAMAÑO RESPUESTA
10	1024	≈ 128 Bytes
20	1048576	≈ 131 KB
30	1073741824	≈ 134 MB
40	1099511627776	≈ 137 GB
50	1125899906842624	≈ 140 TB

Tabla 2: Relación de la evolución del input de cúbits con la salida en bits.

Las citadas tecnologías se encuentran en la **capa de aplicación** (REST y SOAP), donde se establecen los mecanismos sobre cómo se procesan los datos. Estas a su vez, se ejecutan sobre la **capa de presentación** HTTP. Esta capa es la encargada de mostrar la información correctamente mediante las diferentes opciones de representación.

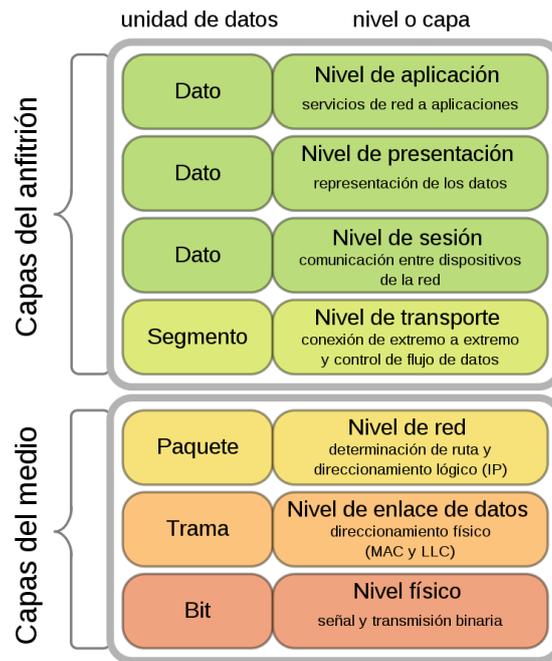


Figura 6: Representación de las distintas capas del Modelo OSI.

Otros mecanismos de intercambio de información, como RPC, funcionan más abajo en el Modelo OSI, en la **capa de sesión**, donde se proporcionan los mecanismos pertinentes para controlar el intercambio de información entre los sistemas que intervienen en dicha comunicación (ver Figura 6 para una explicación más detallada de las diferentes capas del Modelo OSI⁸).

Para completar este requisito, es necesario realizar un estudio en el cual se haga un balance de a qué nivel es más interesante la implementación del protocolo de comunicaciones entre los dispositivos clásicos y las QPUs, ponderando la versatilidad de dicho protocolo, para adaptarlo a diferentes escenarios, la dificultad de su implementación para su posterior mantenimiento y su escalabilidad para futuros escenarios donde aumente la capacidad de cómputo.

Además de un sistema de comunicación vía interfaces de red. Es necesario una comunicación vía PCI, como si fuese un componente periférico. Para poder disponer de este sistema se tiene que implementar un driver para el sistema de comunicaciones. Dicho driver debe hacer de intermediario entre el kernel del sistema operativo y la interfaz del periférico, en este caso, la QPU, tal y como se muestra en la Figura 7. De esta forma, se permite el intercambio de información

⁸ Modelo OSI. Consultado con fecha de 23/04/2023 en <https://programacionfacil.org/blog/el-modelo-osi-explicacion-sencilla/>
CTS-2022-0054

entre cualquier producto software que haga llamadas al sistema siempre y cuando esté instalado el driver que permite dichas comunicaciones entre el sistema y la interfaz del dispositivo.

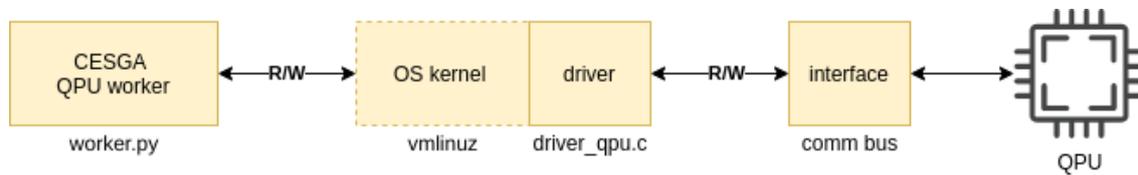


Figura 7: Pila de llamadas de del controlador PCI de la QPU.

3.4 Integración con recursos externos

Tal y como se ha citado en la Sección 3.2, uno de los requisitos de esta nueva infraestructura es poder satisfacer la creciente demanda de los investigadores en computación cuántica. Dada la limitación del número de cúbits de los QPUs actuales, los centros de supercomputación pueden buscar ampliar la oferta con recursos externos para que, los investigadores que así lo deseen, puedan ejecutar experimentos que sobrepasen la capacidad de los ordenadores cuánticos estándares actuales o realizar un benchmarking de las principales soluciones de computación cuántica del mercado a nivel global o por tipo de tecnología (iones atrapados, semiconductores, etc.).

Para la integración con estos proveedores de recursos externos, al igual que en la gestión de colas, se debe contrastar el uso del **GridWay Metascheduler**, no solo para el uso de recursos de computación cuántica (e.g. AWS Braket) sino también para el uso de la computación clásica (e.g. AWS EC2). Nuevamente, si el citado producto software no cumple con los requerimientos de integración es necesario el diseño, implementación, validación y puesta en producción de un componente software que aúne los mismos casos de uso que el GridWay Metascheduler, y, además, incorpore la iteración con los proveedores cloud. En la Figura 8 se muestra un diagrama de la integración con recursos de computación clásica y cuántica de proveedores externos. En este caso se muestra un ejemplo de integración con recursos de la plataforma de AWS.

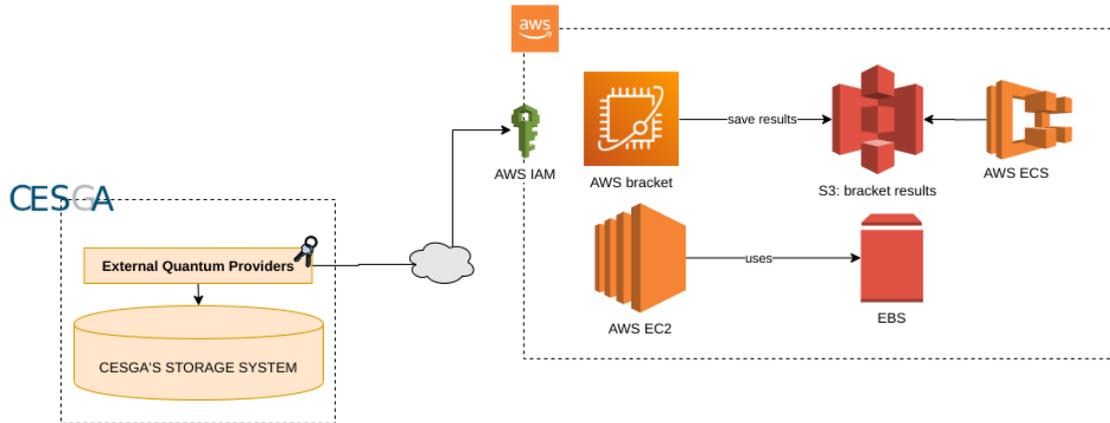


Figura 8: Diagrama de integración con recursos de computación clásica y cuántica de proveedores externos (ejemplo de integración con la plataforma de AWS).

De esta forma, dicho servicio haría uso de los servicios de autenticación (en el ejemplo anterior AWS IAM), para gestionar los recursos internos con los diferentes usuarios del CESGA. Otro aspecto a evaluar ya sea con GridWay Metascheduler o con otra solución desarrollada de forma interna, es la administración de cuentas maestras y subcuentas de los distintos proveedores cloud. Esto es necesario para poder delimitar los costos de cada usuario y/o cuenta que se gestione desde la infraestructura.

3.5 Representación de los circuitos cuánticos

Una de las principales diferencias entre la computación clásica y la cuántica es que en la primera se ejecutan algoritmos y en la segunda circuitos. En la Figura 9 se muestra un ejemplo de un circuito cuántico. Las entradas se representan en la parte izquierda y la salida (resultado de la medida) en la parte inferior.

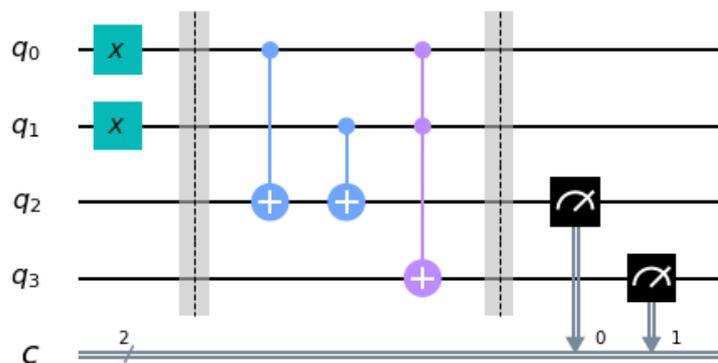


Figura 9: Representación de un circuito cuántico.

Esta clase de representaciones tienen problemas de escalabilidad ya que para algoritmos muy complejos donde se combinen varias entradas y salidas de distintas puertas hará que la longitud del circuito rebase el ancho de la pantalla. Otro problema añadido a esta clase de representaciones es la trazabilidad de los inputs y outputs cuando se encadenan varios operadores. Para combatir estas deficiencias, se han considerado varias aproximaciones, como el *quantum circuit cutting* (ver Figura 10 para más detalles), una técnica que permite descomponer un circuito en varios subcircuitos más pequeños que se pueden ejecutar en computadoras cuánticas más pequeñas.

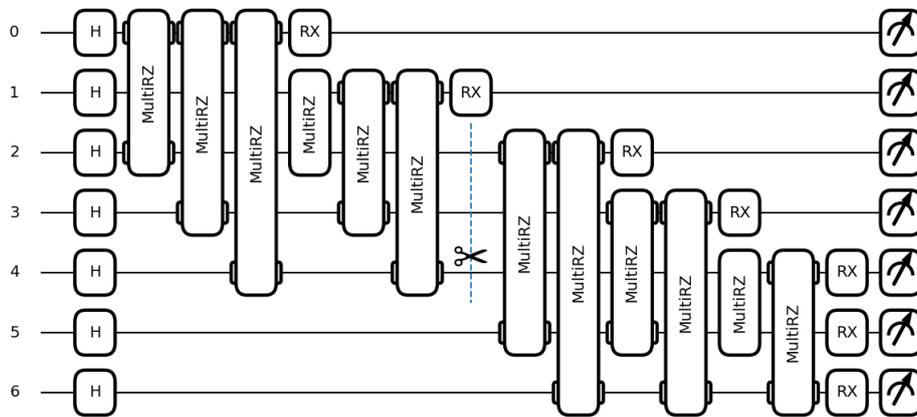


Figura 10: Representación de un circuito mediante "quantum circuit cutting".

Para mejorar la comprensión de los circuitos cuánticos sería deseable combinar una estrategia de quantum circuit cutting permitiendo la anidación de varios operadores elementales en un único operador global. Para ello, habría que investigar cómo lo realizan otras soluciones actuales, ya sean del ámbito de la computación cuántica o clásica. En la Figura 11 se muestra un ejemplo con Logisim, una herramienta que permite diseñar y simular circuitos lógicos digitales y obtener operadores anidados agregando varios operadores elementales.

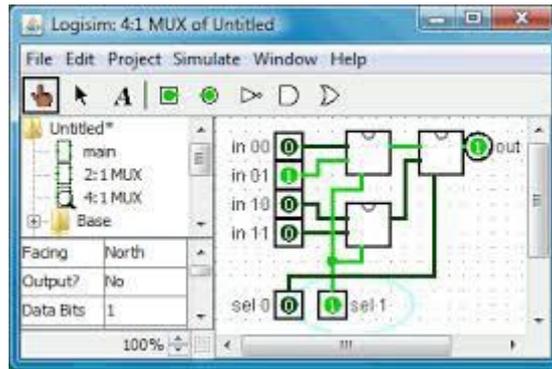


Figura 11: Pantalla de trabajo del Logisim.

Otra aproximación interesante es poder ver en formato grafo la ejecución de los diferentes steps que se ejecutan en las tareas tanto híbridas como cuánticas, pudiendo navegar entre los distintos operadores que las conforman y viendo las dependencias y relaciones entre ellas, tal y como hacen otros softwares. La Figura 12 muestra dos ejemplos de orquestación de tareas mediante representaciones en formato grafo (DAGs) obtenidas a través de la interfaz web de Spark (izquierda) y de Apache Airflow (derecha).

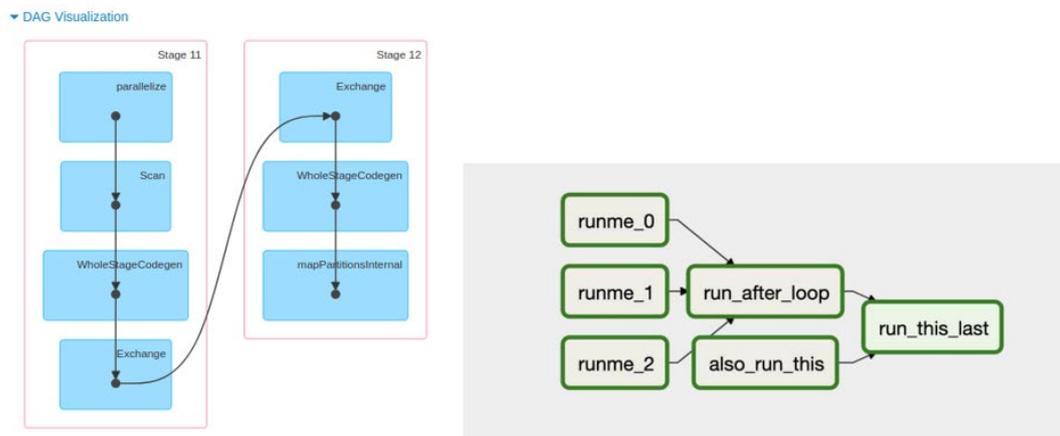


Figura 12: Orquestación de tareas mediante grafos.

3.6 Representación de los resultados

Cuando se ejecuta un circuito cuántico se obtiene una distribución de la combinación de todos los posibles cúbits de entrada, es decir, 2^n posibilidades, siendo n el número de cúbits de entrada. En la Figura 13 (izquierda) se reflejan las limitaciones de los métodos de visualización ordinarios cuando el número de cúbits es relativamente alto.

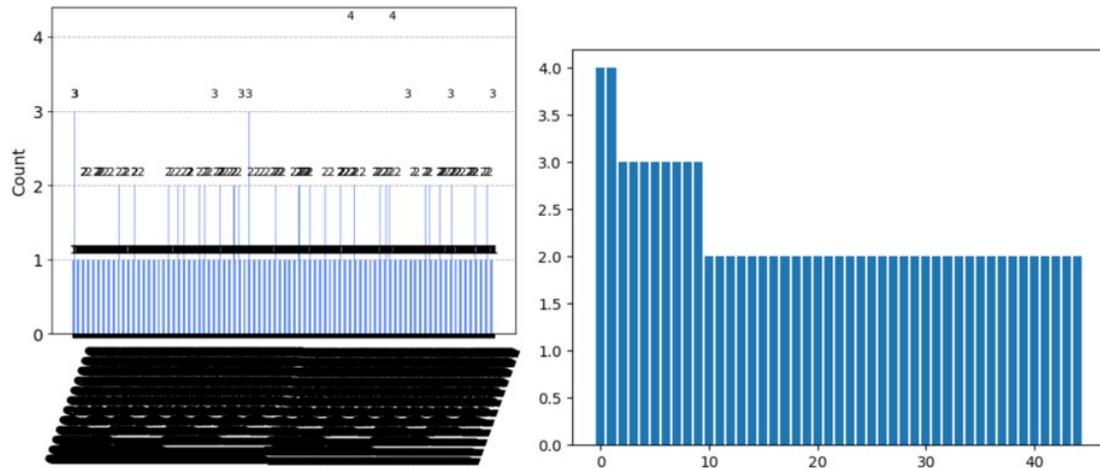


Figura 13: Ejemplo de representación de los resultados de un circuito cuántico.

En el ejemplo se muestra la salida de un circuito cuántico con 10 cúbits. Tal y como se puede apreciar en la Figura, la distribución de los resultados no presenta un comportamiento uniforme. Además, las barras y las etiquetas del barplot se solapan y no permiten ver con claridad la información relevante del gráfico. Por ello, sería interesante incorporar herramientas adicionales que permitan extraer la información relevante (por ejemplo, mediante charts dinámicos) para centrarse en los estados más probables (ver gráfica derecha en Figura 13).

Otra aproximación interesante, sería el estudio e investigación de la representación de datos en otras disciplinas (por ejemplo, business intelligence, química, física, etc.) para lograr extraer la información relevante, despreciando aquella más redundante que no aporta valor (ver una muestra de métodos alternativos de representar la información en la Figura 14).

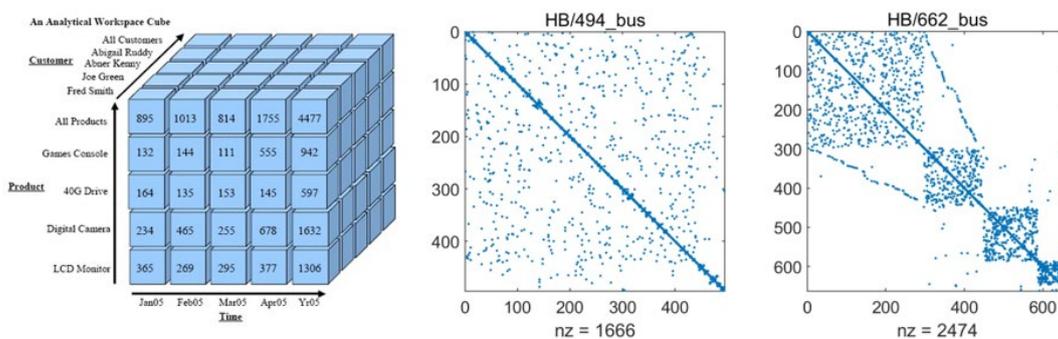


Figura 14: Métodos alternativos de representación de la información.

3.7 Gestión de la contabilidad

Una vez habilitados los servicios de computación cuántica y abiertos a la comunidad científica, los centros de supercomputación como el CESGA necesitan recabar información sobre los trabajos ejecutados en la infraestructura, tanto desde el punto de vista de la **contabilidad** como de la **monitorización**. La información recogida podrá ser usada para emitir los informes pertinentes sobre el uso de la infraestructura cuántica en donde, además de incluir información relativa a los procesos y tareas ejecutadas, se podrá incorporar también información sobre el consumo o la huella energética asociada. Con estos pasos, se busca analizar el impacto energético asociado al uso de la infraestructura en computación cuántica y, fomentar y avanzar hacia el uso de las tecnologías verdes. Por otra parte, la monitorización de los procesos puede ser útil para alimentar algoritmos de analítica avanzada enfocados a mejorar la gestión de los recursos disponibles.

Para la gestión la contabilidad es necesario analizar qué métricas o KPIs son los más adecuados estudiar para este nuevo servicio de computación. Para dar un poco de perspectiva de cómo se debería enfocar la gestión de la contabilidad, se deberían estudiar dos verticales:

- El modo de gestión de la contabilidad de los servicios actuales de un centro de supercomputación como el CESGA.
- El modo de gestión de la contabilidad de los principales proveedores de servicios de computación cuántica.

Actualmente, el CESGA dispone de 2 modelos de computación: el sistema de HPC del Finisterrae III y el entorno cloud. El entorno HPC dispone de tres modalidades:

- Sin prioridad.
- Con prioridad.
- En reserva.

En el caso de las GPUs, solo están disponibles bajo la modalidad de reserva y cada proceso que se ejecuta sobre este hardware consume 10 cores. En el caso de la plataforma cloud hay una pequeña diferencia respecto al caso anterior. A diferencia de la gestión de la contabilidad del sistema HPC que está basado en el consumo core/hora, en el entorno cloud también se tiene en cuenta la memoria RAM y la capacidad de almacenamiento (HD por sus siglas en inglés).

En cuanto al segundo vertical, se han encontrado opciones muy dispares en el modo en que los grandes proveedores de servicios cuánticos gestionan la contabilidad. Más concretamente, se ha realizado un análisis de la gestión de la contabilidad de Amazon Braket, IBM, Azure Quantum, Google Quantum AI, D-Wave y Pasqal.

En el caso de AWS, el servicio de computación cuántica AWS Braket⁹ se contabiliza en base al proveedor de hardware / tipo de QPU, haciendo distinción entre:

- Procesado de una tarea, es decir, la ejecución del circuito cuántico.
- Re-ejecución del circuito ("shot").

Los precios por tarea no varían en función del proveedor de hardware, manteniéndose en un precio fijo de 0.3\$. No obstante, en el precio por shot si se observan importantes diferencias respecto al proveedor. La opción más económica se corresponde con el proveedor Xanadu con la familia de QPU Borealis mientras que el precio máximo lo alcanzan los proveedores IonQ y Quera con un total de 0.01\$ por shot. En el caso de ejecución de circuitos cuánticos sobre emuladores solo se tiene en cuenta el tiempo de uso de la máquina por minuto, de la misma forma que se hace con las AWS EC2.

IBM dispone de 3 planes de acceso a sus servicios de computación cuántica: Open, pay-as-you-go y premium¹⁰. El primero proporciona acceso a una QPU de 5 o 7 cúbits y es un plan gratuito mientras que los dos planes restantes son de pago. A diferencia de Amazon Braket, IBM posee una facturación en base al tiempo de cómputo. El precio del plan pay-as-you-go es de 1.60\$ por segundo y da acceso al QPU Falcon de 27 cúbits. Por otra parte, las tarifas del plan premium dan acceso a Osprey (próximamente), Eagle y Falcon de 433, 127 y 27 cúbits respectivamente, y solo están accesibles mediante consulta directa con el proveedor.

Cualquiera de los planes anteriores da acceso también a los emuladores cuánticos. Además, con el plan premium se dispone de servicios de soporte y acceso a los sistemas exploratorios. Este procesador usa una nueva arquitectura en lugar de resonancia cruzada, lo que permite mejorar la calidad respecto a las iteraciones anteriores.

Por otro lado, la gestión de la contabilidad de los servicios de Azure Quantum y Google Quantum AI siguen una política de mayor complejidad. Azure Quantum

⁹ Gestión de la contabilidad de Amazon Braket. Consultado con fecha de 31/03/2023 en <https://aws.amazon.com/es/braket/pricing/>

¹⁰ Gestión de la contabilidad de IBM. Consultado con fecha de 31/03/2023 en <https://www.ibm.com/quantum/access-plans>

dispone de un amplio espectro de servicios de computación cuántica, tanto propios como de sus partners, y ofrece un crédito gratuito de 500\$ por proveedor. En el caso de los proveedores externos, son estos los que asumen tanto la definición como el control de la gestión de la contabilidad¹¹.

Los servicios de computación cuántica de Azure ofrecen dos niveles (Learn & develop y Performance at scale), sin compromiso mínimo y con capacidad de escalado en base a la complejidad de los problemas. En ambas opciones, la primera hora es gratuita y aunque la gestión de la contabilidad se realiza por hora de cómputo, la facturación se realiza a nivel de segundo. Los partners de Azure Quantum poseen tarifas individualizadas. En el portal de Microsoft se detallan las tarifas de IonQ, Quantinuum y Rigetti.

IonQ factura en función del número de puertas del programa, la complejidad y el número de lanzamientos (gate-shots). Asumiendo que cada programa cuántico consta de N puertas lógicas de uno o más cúbits, IonQ basa la gestión de la contabilidad en el número de capturas, definiéndose estas en base a la fórmula

$$QGS=NC,$$

donde

- N representa el número de puertas de uno o dos cúbits.
- C es el número de capturas de ejecución solicitadas.

Las puertas de dos cúbits con varios controles se facturan como $6(N-2)$ puertas de dos cúbits, donde N es el número de cúbits implicados en la puerta. Así, por ejemplo, una puerta NOT con tres controles se contabilizaría como $6(4-2)$ o 12 puertas de dos cúbits.

Más concretamente, IonQ dispone de dos planes: pago por uso y el plan Aria¹². Ambos planes incluyen acceso al simulador de IonQ de forma gratuita. El plan de pago por uso proporciona acceso a un equipo cuántico de iones atrapados Harmony de 11 cúbits y la gestión de la contabilidad se realiza en función del número de gate-shots cuánticas ejecutadas (más los costos de la infraestructura de Azure) con un mínimo de 1\$ por ejecución del programa, mientras que el plan Aria se trata de un

¹¹ Gestión de la contabilidad de Azure Quantum. Consultado con fecha de 31/03/2023 en <https://azure.microsoft.com/en-us/pricing/details/azure-quantum/>

¹² Gestión de la contabilidad de IonQ en Azure. Consultado con fecha de 31/03/2023 en <https://learn.microsoft.com/es-es/azure/quantum/pricing?tabs=tabid-paygo%2Ctabid-AQcredits0%2Ctabid-AQcreditsRigetti%2Ctabid-payasgo%2Ctabid-learndevelop&pivots=ide-computing>

plan de suscripción mensual que da acceso al equipo cuántico de IonQ Aria de 23 cúbits y también a la QPU Harmony. En este caso el precio de la suscripción es de 25.000\$ / mes a lo que habría que sumar los costos de la infraestructura de Azure.

Quantinuum usa un sistema de crédito que cobra por cada tarea en base al número de operaciones y el número de capturas ejecutadas. Las unidades de uso se basan en Créditos Cuánticos del Sistema H (HQC por sus siglas en inglés). Estos se definen a través de la siguiente expresión matemática,

$$HQC = 5 + C(N_{1q} + 10N_{2q} + 5N_m)/5000,$$

donde,

- N_{1q} es el número de operaciones de un solo cúbit en un circuito.
- N_{2q} es el número de operaciones nativas de dos cúbits en un circuito. La puerta nativa equivale a CNOT de hasta varias puertas de un solo cúbit.
- N_m es el número de operaciones de preparación y medida del estado (SPAM) en un circuito, incluida la preparación del estado implícito inicial, las medidas intermedias y finales, y los restablecimientos de estado.
- C es el número de disparos.

Quantinuum cuenta con dos tipos de suscripción: una estándar y una premium¹³. El precio de la suscripción estándar es de 125000\$ / mes mientras que para la premium es de 125000\$ / mes. A ambas suscripciones habría que añadirle los costos derivados de la infraestructura de Azure.

Por otra parte, la política de gestión de la contabilidad de Rigetti se centra únicamente en el tiempo de ejecución en sus procesadores cuánticos Aspen-M-2 de 80 cúbits y Aspen-M-3. No hay cargos adicionales por trabajo, por captura o por puerta, únicamente se factura 0.02\$ por incremento de 10 milisegundos de tiempo de ejecución. Además, Rigetti ofrece de forma gratuita para todos los usuarios acceso al simulador QVM.

En el caso de Google, ofrece acceso al QPU de IonQ desde su infraestructura cloud¹⁴. En este caso, la política de gestión de la contabilidad es similar a la que el

¹³ Gestión de la contabilidad de Quantinuum para Azure. Consultado con fecha de 31/03/2023 en <https://learn.microsoft.com/es-es/azure/quantum/pricing?tabs=tabid-aria%2Ctabid-standard%2Ctabid-AQcreditsRigetti%2Ctabid-payasgo%2Ctabid-learndevelop&pivots=ide-computing>

¹⁴ Gestión de la contabilidad de IonQ en Google Cloud Platform. Consultado con fecha de 31/03/2023 en <https://console.cloud.google.com/marketplace/product/ionq-public/ionq?q=search&referrer=search&hl=es>

proveedor de servicios cuánticos ha establecido para Azure. Por otra parte, en el caso de D-Wave y Pasqal, la gestión de la contabilidad se realiza en base al tiempo de cómputo.

4. Conclusiones

La computación cuántica constituye una de las disciplinas más disruptivas del presente siglo. En los últimos años, los investigadores han invertido grandes esfuerzos y han conseguido hallazgos importantes que permiten asentar las bases para trasladar el conocimiento adquirido al ámbito de la industria. La computación cuántica se basa en las propiedades de los cúbits. Estos pueden estar en múltiples estados simultáneamente lo que permite realizar cálculos de una manera mucho más rápida que con la metodología clásica. Las aplicaciones de esta disciplina abarcan un amplio abanico de posibilidades. Algunas de las más comunes son: optimización logística, modelos financieros complejos, diseño y descubrimiento de fármacos, genómica, ciberseguridad y criptografía, modelado de moléculas, diseño y modelado de materiales, física aeroespacial, simulación cuántica o simulación de fenómenos nucleares.

Al margen de las aplicaciones técnicas anteriores, esta disciplina también plantea nuevos retos éticos, como la seguridad de la información, ya que los algoritmos cuánticos pueden romper los sistemas de cifrado más comunes de la actualidad.

Bajo la iniciativa del Polo de Tecnologías Cuánticas, el CESGA está centrando sus esfuerzos para posicionar a Galicia como un referente en computación cuántica tanto a nivel nacional como internacional. Para ello, dispone ya de un emulador cuántico y va a ampliar en breve la infraestructura cuántica incorporando un ordenador cuántico, un emulador cuántico, un computador de altas prestaciones y un sistema de almacenamiento para alojar los resultados de los circuitos cuánticos.

Para poder satisfacer las necesidades actuales y futuras de un centro de supercomputación como CESGA, tanto de integración de la infraestructura cuántica como de apertura a la comunidad científica es preciso abordar una serie de desafíos. Para ello, se ha realizado un análisis de requerimientos, tanto de la infraestructura actual del CESGA como de los retos que es necesario acometer. Más concretamente, se ha realizado un análisis de requerimientos para el control de

acceso, la gestión de colas, las interfaces de comunicación, la integración con recursos de computación externos (por ejemplo, Amazon Braket, Azure Quantum o Google Cloud Platform), útil tanto para análisis de benchmarking como para ejecutar circuitos cuánticos de más de 32 cúbits.

Por otra parte, teniendo en cuenta el carácter exponencial de las salidas de los circuitos cuánticos y, dado que los sistemas de visualización tradicionales colapsan con el aumento del número de cúbits, se ha realizado también un análisis de los requerimientos en este ámbito, abarcando tanto la representación de los circuitos cuánticos como las salidas.

Finalmente, se ha realizado un análisis de la gestión de la contabilidad de los principales proveedores de servicios de computación cuántica en donde se observan desde políticas basadas en planes hasta métodos de mayor complejidad que tienen en cuenta tanto el número como la complejidad de las tareas.