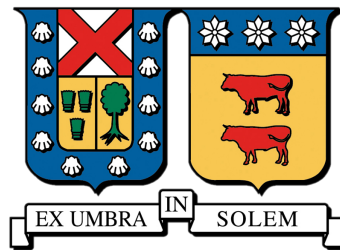


UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INFORMÁTICA
SANTIAGO - CHILE



**ARQUITECTURA DE REQUERIMIENTOS Y
DISEÑO PARA EL OBSERVATORIO
VIRTUAL CHILENO**

CAMILO JAVIER ESPINOZA MOLINA

**MEMORIA DE TITULACIÓN PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INFORMÁTICO**

PROFESOR GUÍA: MAURICIO SOLAR
PROFESOR CORREFERENTE: MAURICIO ARAYA

Junio - 2017

Agradecimientos

En estos párrafos de agradecimientos y buenos deseos para las personas que han sido partícipes de mi proceso de crecimiento durante, no sólo el desarrollo de este trabajo, ni de mi estadía en la universidad sino más bien (mucho más allá)... ¡de la vida! deseo comentar con gran entusiasmo lo importante que es para mí cerrar este ciclo lleno de altos y de bajos. Realmente deseo atribuirles con mucho cariño, en este pequeño espacio, este pequeño (gran) logro personal. El cual siempre estuvo marcado por diversas vacilaciones que compartimos, que nos incitaron a incursionar no sólo en los rumbos categorizados de la sociedad, sino (quizás exagerando) en la vorágine de ideas cuestionadoras, en el disfrute de la música y el arte, en la búsqueda por conformarnos como seres vivientes integrales y conscientes. Es muy agradable avanzar, cuando durante en cada momento de incertidumbre se tiene el soporte de cada uno de ustedes, sus sueños, su intriga y la contagiosa energía que ponen en sus vidas. Desde acá, opte si bien por no colocar sus nombres, queridos amigos, siéntanse con el derecho de atribuirse este sincero párrafo.

Por lo que sigue, deseo redactar un par de saludos especializados a mis padres y su inspiradora lucha por querer brindar siempre lo mejor para sus hijos, en apoyar cada idea vaga de uno con el único propósito de ver nuestro desarrollo pleno; a mi hermano Cristóbal y Makita, cómplices de mis locuras, compañeros cercanos y de confianza; a mi sobrinita Amanda que nutre el hogar con su alegría y energía. Cuando sepas leer, será muy lindo que veas estas líneas que buscan reflejar la felicidad que nos otorgas, a cada uno, sólo con tu presencia; y por último a Cami, compañera de vida, amiga y pareja, increíblemente acá estamos después de fugaces tres años, juntos, adaptándonos, escuchándonos y persiguiendo un sueño conjunto. Muchas gracias por tu apoyo infinito.



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

Esta página ha sido intencionalmente dejada en blanco.

Resumen

En el presente documento se detalla el procedimiento de investigación e implementación que permitió definir la arquitectura de requerimientos y el diseño correspondiente para el observatorio virtual chileno (ChiVO). Todo esto, de acuerdo a los desafíos actuales que se presentan en el área de la astronomía en términos del almacenamiento y procesamiento de los grandes volúmenes de datos procedentes de distintos observatorios astronómicos.

Mediante esta memoria el lector podrá evidenciar que la implantación de esta plataforma astroinformática chilena sustentada bajo un arquitectura “Cloud Computing” brinda una solución tangible capaz de sopesar en cierta medida el problema de la “avalancha de datos” para su uso dinámico y efectivo. Por otra parte podrá distinguir mediante un análisis sistémico, las funciones y relaciones que cumplen cada uno de los componentes que conforman el sistema de ChiVO, considerando en todo momento los protocolos y estándares recomendados para la implementación de un observatorio virtual.

Finalmente, es menester señalar que este trabajo está orientado para astrónomos que deseen utilizar los servicios que dispondrá este observatorio virtual (algunos ya disponibles), para ingenieros informáticos que busquen una base teórica para la fundamentación de arquitecturas “Cloud”, y en definitiva para cualquier persona con interés de dimensionar lo que conlleva el desarrollo de una plataforma astroinformática.

Abstract

This particular document expresses the investigation and implementation procedures, which allows the definition of an architecture for requirements and the corresponding design for the Chilean Virtual Observatory (ChiVO). All this, according to the current challenges that arise in the area of astronomy in terms of storage and processing of large volumes of data from various astronomical observatories.

Through this memory, the reader could witness that the implementation of this Chilean astro-informatics platform, supported by an architecture “Cloud Computing” manifests a tangible solution capable of weighing up (to some extent) the problem of “Data Avalanche” for its dynamic and effective use. On the other hand the reader will be able to distinguish through a systemic analysis, the functions and relations that fulfill each of the components that make up the ChiVO system, considering at all times the protocols and standards recommended for the implementation of a virtual observatory.

Finally, it is necessary to point out that this work is oriented towards astronomers who wish to use the services that the Chilean Virtual Observatory will have available (some of which already are), for computer engineers seeking a theoretical basis for the foundation of “Cloud” architectures, and ultimately, for any person with interest on what is involved in the development of an astro-informatics platform.

Glosario

A continuación se lista una breve descripción de las siglas y términos utilizados en esta memoria:

Definición de siglas y términos Generales :

1. **VO:** Observatorio Virtual. Manifiesta una plataforma informática que propicia un acceso sencillo y homogéneo a bases de datos de observaciones astronómicas; además, puede proveer herramientas para el análisis y procesamiento de la ingente cantidad de observaciones descritas.
2. **ChiVO:** Observatorio Virtual Chileno. Manifiesta una plataforma VO en desarrollo que busca (además de un acceso homogéneo a los datos) incorporar poderosos algoritmos para el procesamiento de los datos, obtenidos del observatorio astronómico ALMA.
3. **ALMA:** Atacama Large Millimeter/submillimeter Array. Interferómetro que comprende un conjunto de 66 antenas destinadas a la observación de longitudes de onda milimétricas y submilimétricas.
4. **IVOA:** International Virtual Observatory Alliance. Esta organización fomenta la colaboración entre observatorios virtuales de todo el mundo, mediante la prescripción de estándares para el funcionamiento efectivo de un VO.
5. **FITS:** Flexible Image Transport System. Representa el formato estándar para imágenes y tablas astronómicas.
6. **TI:** Tecnologías de la información. Conjunto de servicios tecnológicos alineados con el almacenamiento, procesamiento, transmisión y seguridad de

la información. En definitiva, resume todo lo relacionado con la informática, la electrónica y las telecomunicaciones.

7. **TAR:** Representa un formato de archivos usado ampliamente en entornos UNIX para la contención de varios archivos.
8. **GB, TB , PB:** Terabyte y PetaByte respectivamente. Unidades de medición utilizadas para describir el espacio de almacenamiento de datos tanto en discos como en memoria.

Definición de siglas y términos para el capítulo dos :

1. **SOA:** Service Oriented Architecture. Manifiesta la representación de una estructura abierta, flexible y federada basada principalmente en la combinación de distintos servicios interoperables.
2. **IaaS:** Infrastructure as a Service. Entorno “Cloud” que engloba recursos físicos TI (centrados en la infraestructura), los cuales podrán ser accedidos a través de herramientas e interfaces basadas en la “nube”.
3. **PaaS:** Platform as a Service. Entorno “Cloud” que representa a un conjunto de herramientas desplegadas y preconfiguradas dispuesta a ser utilizada para la creación de nuevas herramientas informáticas. Por lo tanto, en este dominio se condensan principalmente interfaces de programación y/o entornos de desarrollo.
4. **SaaS:** Software as a Service. Entorno “Cloud” que otorga la posibilidad de posicionar una aplicación como un servicio compartido en la “nube” (Esto es, bajo el alero de una arquitectura “Cloud Computing”).
5. **BCM:** Bright Cluster Manager[®]. Administrador de clústeres para HPC (High Performance Computing).

Definición de siglas y términos para el capítulo tres :

1. **XML:** Extensible Markup Language. Formato de texto simple y flexible que desempeña un papel importante en el intercambio de datos en la red.
2. **VOTABLE:** Formato XML flexible para el almacenamiento e intercambio de datos tabulares astronómicos.

Definición de siglas y términos para el capítulo cuatro :

1. **NIST:** Instituto nacional de estándares y tecnología. Organización que promueve normas para la mejora sustancial de la calidad de vida. Por medio del refinamiento (normalización) de unidades de medición ligadas a la ciencia (metrología).
2. **HPC:** Computación de alto rendimiento. Tecnología que provee los mecanismos para el desarrollo efectivo de simulaciones computacionales a problemas complejos.
3. **Jupyter Notebook:** Aplicación web que permite crear y compartir documentos que

El Jupyter Notebook es una aplicación web de código abierto que le permite crear y compartir documentos que contienen código en vivo, ecuaciones, visualizaciones y texto explicativo. Los usos incluyen: limpieza y transformación de datos, simulación numérica, modelado estadístico, aprendizaje automático y mucho más.

La Jupyter Notebook App es una aplicación cliente-servidor que permite editar y ejecutar documentos de portátiles a través de un navegador web. La aplicación Jupyter Notebook puede ejecutarse en un escritorio local que no requiera acceso a Internet (como se describe en este documento) o se puede instalar en un servidor remoto y acceder a través de Internet .

4. **ACALIB:** Biblioteca avanzada de cómputo para astronomía.
5. **JOVIAL:** “Jupyter override for astronomical libraries”. Plataforma web compuesta por distintas librerías astronómicas (ACALIB es una) para el procesamiento y posterior análisis de datos.
6. **DOCKER:** Herramienta que permite a desarrolladores y administradores de sistema desplegar sus aplicaciones en un contenedor ligero.
7. **NFS:** Protocolo que permite el acceso remoto a un sistema de archivos a través de la red.

Índice general

Agradecimientos	3
Resumen	5
Abstract	6
Glosario	7
1. Introducción	15
1.1. Objetivos	18
1.2. Alcance	18
1.3. Estructura del documento	19
2. Marco teórico paradigma “Cloud Computing”	20
2.1. Definición de Cloud Computing	20
2.2. Modelos para una arquitectura “Cloud”.	22
2.2.1. Modelos de servicios.	22
2.2.2. Modelos de despliegue	24
2.3. Arquitectura orientada a servicios (SOA)	25
2.3.1. ¿Qué es una arquitectura orientada a servicios?	26
2.3.2. Relación Cloud Computing - SOA	26
3. Arquitectura genérica de un observatorio virtual	28
3.1. Definición Observatorio Virtual	28
3.2. Estándares de interoperabilidad de IVOA	29
3.2.1. Capa intermedia: Subsección REGISTRY.	31
3.2.2. Capa intermedia: Subsección DATA ACCESS PROTOCOL STANDARDS.	33

3.2.3. Capa intermedia: Subsección VO CORE.	34
3.2.4. Capa intermedia: Subsección SHARING.	40
4. Arquitectura Cloud Computing para ChiVO	41
4.1. Definición modelos de servicios para ChiVO.	42
4.1.1. Infraestructura como servicio ChiVO	43
4.1.2. Plataforma como servicio ChiVO	53
4.1.3. Software como servicio ChiVO	64
4.2. Arquitectura de servicios para ChiVO	66
4.2.1. Consideraciones técnicas generales	67
4.3. Modelo de implantación para ChiVO	68
4.3.1. ChiVO PRIVATE CLOUD	70
4.3.2. ChiVO COMMUNITY CLOUD	71
Conclusiones	73
Bibliografía	76

Índice de figuras

2.1. Entorno PaaS basado en los recursos TI proporcionados por un entorno IaaS subyacente.	23
2.2. Una simple vista compuesta de entornos IaaS y PaaS que albergan tres implementaciones de servicios del entorno SaaS.	24
2.3. Elementos esenciales arquitectura “Cloud Computing”.	27
3.1. Arquitectura nivel 2 de IVOA.	30
4.1. Diagrama alto nivel arquitectura ChiVO.	42
4.2. IaaS ChiVO Alto Nivel.	43
4.3. Storage Pool CHIVO.	44
4.4. Estructura Lustre estándar para almacenamiento escalable.	45
4.5. Conjunto de servidores físicos de ChiVO.	47
4.6. Bright Cluster Manager®.	48
4.7. CHIVO Bright Cluster Manager	49
4.8. Arquitectura estandar de oVirt.	50
4.9. Nodos de virtualización de ChiVO	52
4.10. Plataforma de administración de virtualización de ChiVO	53
4.11. Diagrama PaaS (Alto nivel)	54
4.12. Diagrama Data Services ChiVO.	55
4.13. Diagrama Processing Services ChiVO.	57
4.14. Diagrama Jovial Services ChiVO	58
4.15. Diagrama Jovial Services extensión UPI ChiVO	60
4.16. Diagrama Apache Spark ChiVO	62
4.17. Arquitectura estandar de plataforma Apache Spark	63
4.18. Muestra de Jupyter Notebook en SaaS	64
4.19. vo.chivo.cl	65

4.20. Arquitectura de servicios de ChiVO	66
4.21. Diagrama de una organización usando una nube híbrida que utiliza nubes tanto privadas como públicas [5].	69
4.22. ChiVO Virtual Private Cloud	70
4.23. ChiVO Virtual Community Cloud	71

Índice de tablas

3.1. Tabla descriptiva para estándar “Resource Metadata”.	32
3.2. Campos modelo ObscoreDM. Fuente [12]	36
4.1. Detalles “Filesystem Lustre”	45
4.2. Descripción de elementos arquitectura oVIRT	51

Capítulo 1

Introducción

“El desarrollo de cualquier disciplina científica involucra el manejo de datos. En el caso particular de la astronomía, el incremento de la cantidad y tamaño de los archivos ha ido creciendo con el paso de los años, considerando así a ésta una ciencia de datos intensivos”. [8]

Históricamente la astronomía ha sido considerada una ciencia de carácter visual, en donde la primera imagen que se viene a la mente al pensar en esta área científica es la de un investigador observando el espacio con un gran telescopio. Lo cierto es que con el avance de la tecnología el procedimiento de recolección de datos, y de manera específica, de imágenes espectrales, ha ido cambiando drásticamente desde el año 1976 cuando se adosó un dispositivo de carga acoplada (sensor con células fotoeléctricas que registran una imagen; pieza fundamental de cámaras fotográficas y vídeos digitales) a un telescopio de 155 centímetros de diámetro, y con el cual se obtuvo útiles imágenes de Urano, Jupiter y Saturno [8]. En síntesis, el desarrollo de tecnologías asociadas a la obtención de imágenes impulsó con el tiempo mejoras sustentables en la cobertura espectral, en la sensibilidad fotométrica, en la resolución espacial y en definitiva, en la proliferación de los datos.

En la actualidad, el continuo aumento de los datos experimentales que proveen los observatorios astronómicos y distintos centros de investigación, tanto en cantidad como en tamaño, ha provocado numerosos inconvenientes a la hora de su administración en plataformas computacionales. Adicionalmente, el problema se acrecienta cuando ya no es importante sólo acceder a estos, sino poder procesarlos

(distribuir flujos de trabajo), analizarlos y visualizarlos en tiempos considerables de ejecución.

“The explosion of digital surveys in the last decade has ushered in a new era of data richness and complexity. About 3 years ago, it was estimated that more than 1 PB (petabyte) of public data were electronically accessible, and that the data volume was growing at rate of about 0.5 PB per year. Projections indicated that by 2020, 60 PB of archived data should be accesible to astronomers. The growth of data in size is going to accelerate dramatically in the next years as new project move into operation (e.g. LSST, ALMA, Euclid, SKA, just to name a few) will likely turn out that the numbers given above are under-estimated”. [10]

Se destaca del fragmento anterior extraído del libro “The universe of digital sky surveys”, que el crecimiento estimativo del volumen de datos públicos es de 0.5 PB por año y también que la proyección de estos para el 2020 será de 60 PB. Desde acá, se entiende la necesidad de conformar arquitecturas computacionales aptas para albergar la cantidad de datos emergentes y junto a ello, mejorar sustancialmente los servicios existentes para el desarrollo de la astronomía.

En relación a la implementación de recursos tecnológicos físicos, el auge de los denominados centros de datos (“datacenter”) y paradigmas tecnológicos como “Cloud Computing” y/o “microservicios” ha impulsado interesantes desafíos en el área que permitirían soportar no sólo la demanda futura de datos sino también su oportuna gestión mediante el uso de programas informáticos adaptados para estos. Así, disponer de una infraestructura que permite aprovisionar los dispositivos TI centralizados (servidores, bases de datos, dispositivos en redes, etc.) mediante la instalación de aplicaciones especializadas genera la posibilidad de sopesar el problema de la avalancha de datos actuales y futuros en el campo astronómico.

En relación a los programas informáticos especializados para los centros de datos se alude a nuevas soluciones tecnológicas que vienen a expandir las funcionalidades de este conjunto de recursos por medio de la distribución y procesamiento de trabajos en nodos computacionales (HPC) o virtualización de sistemas que propicien el despliegue de servicios para cada usuario que lo necesite. En este documento se profundizará sobre algunas aplicaciones de este tipo que son el soporte

principal para la implantación de entornos de desarrollo, librerías y en definitiva, de software centrado en el acceso, análisis y procesamiento de datos astronómicos.

Cabe destacar que en el campo astroinformático han surgido diversas aplicaciones diestras en la administración de los datos públicos precedentes de los distintos centros de observación e investigación; de manera particular, se denominan **Observatorios virtuales (VO)** y se definen como *“un gran repositorio donde es posible obtener datos que posibilitan estudios sistemáticos, pancromáticos y estadísticamente significativos de la evolución de sistemas astronómicos”* [8]. Estos servicios proporcionan los mecanismos bases para una plataforma de esta área y se posicionan como la cara visible dentro de grandes infraestructuras de almacenamientos con altas velocidades de transferencia. No obstante, la meta actual para este tipo de sistemas va más allá de ser un simple almacén de datos, se requiere ahora profundizar en el desarrollo de innovadoras herramientas que propicien la visualización y análisis ágil de estos mediante la implantación de plataformas dinámicas y efectivas.

Finalmente, este documento se centra en el observatorio virtual Chileno (**ChiVO**) que es uno de los tantos proyectos VO que se están desarrollando en el mundo. Por lo tanto, persigue como objetivo solventar los grandes desafíos que se tienen en cuanto al almacenamiento de los grandes volúmenes de datos y a su vez, brindar nuevas herramientas para el análisis científico mediante la incorporación de poderosos algoritmos de procesamiento. El observatorio ALMA que se estima generará más de 1TB de datos por día, es el principal suministrador de archivos al VO Chileno. Gracias a esta plataforma y los datos entregados es posible almacenar estos grandes volúmenes en el territorio local y así optimizar el tiempo de observación de los astrónomos Chilenos. Es necesario entonces detallar qué tipo de información entregan los datos de este observatorio astronómico y junto a ello el tamaño total en cada ciclo de acuerdo al año de publicación.

ALMA entrega a los usuario archivos .tar que contienen en su interior los diversos datos astronómicos recolectados siguiendo una estructura de directorios definida y jerárquica determinada por la organización. Estos directorios contienen en su interior archivos de imágenes FITS, tablas de calibración instrumental, reportes y manuales, scripts, entre otros. Los tamaños de cada ciclo de datos públi-

cos de ALMA alojados en los dispositivos de almacenamiento lógico de ChiVO son actualmente:

- Periodo: 2011 - (Cycle0): 4.62 TB.
- Periodo: 2012 -(Cycle1): 20.11 TB.
- Periodo: 2013 - (Cycle2): 23.06 TB.
- Periodo: 2015 - (Cycle3): 59.18 TB.

Tras la realización de diversos prototipos para la ingesta de los datos descritos, ChiVO se encuentra ahora profundizando en el desarrollo de modernos algoritmos para el procesamiento y visualización de los mismos. Ejemplo de esto, es la librería “ACALib” que provee una interfaz especializada para el análisis de datos astronómicos, la cual se unirá a una serie de servicios adicionales (descritos en el documento) que darán forma a la arquitectura general de la plataforma astroinformática chilena.

1.1. Objetivos

Se describe como objetivo principal la exposición de los requerimientos y diseños para el observatorio virtual Chileno, incluyendo los servicios disponibles y los que están en proceso de desarrollo. Todo esto, con el fin de determinar una arquitectura general para la plataforma astronómica. Con esto, el lector podrá apreciar que el observatorio virtual Chileno se sustentará bajo un modelo “Cloud Computing”, lo cual genera la necesidad de ajustarse a ciertas características esenciales y modelos genéricos propuestos para este paradigma tecnológico. En definitiva, la representación bajo estos esquemas será lo que guiará definitivamente a brindar una estructura ordenada para el despliegue de los servicios y con ello cumplir el propósito central indicado.

1.2. Alcance

Este documento define en detalle cada paradigma, concepto y servicio presente en ChiVO por lo que su orientación va dirigido a cualquier persona con interés de

dimensionar lo que conlleva el desarrollo actual de una plataforma astroinformática, también se dirige a astrónomos que deseen utilizar los servicios que dispondrá ChiVO y por último a ingenieros informáticos que busquen una base teórica para la fundamentación de arquitecturas “Cloud”.

1.3. Estructura del documento

Este documento se compone por cinco capítulos detallados a continuación:

- El presente capítulo corresponde a esta sección introductoria que busca denotar uno de los principales desafíos a los que se enfrenta en la actualidad el campo de la astronomía en relación a la administración de los datos procesados por los observatorios. Junto a esto se brindan pequeñas definiciones de los conceptos que serán abordados en esta memoria con el fin de interiorizar los capítulos que lo componen.
- El segundo capítulo introduce los conceptos necesarios para la implementación de una arquitectura genérica de “Cloud Computing”. Junto a esto, se presenta la definición de este paradigma tecnológico y cómo es posible posicionar los servicios que se desean disponer en la plataforma ChiVO mediante la consideración de ciertas características y modelos generales.
- En el tercer capítulo se describen los estándares de interoperabilidad de IVOA (alianza internacional de observatorios virtuales) y otros términos asociados con el fin de sustentar los conceptos que aparecerán en la siguiente sección.
- El cuarto capítulo dirige el proceso de formación de la arquitectura de ChiVO. Todo esto mediante el desarrollo progresivo de los conceptos entregados en un diagrama de alto nivel al principio de la sección. Como resultado final se entregarán los modelos proyectados para la plataforma Chilena de acuerdo a los desafíos que busca alcanzar.
- La conclusión brinda comentarios resolutivos del trabajo realizado en este documento en términos de las herramientas tecnológicas utilizadas y por qué fue conveniente escoger este tipo de arquitectura para modelar la plataforma ChiVO.

Capítulo 2

Marco teórico paradigma “Cloud Computing”

En este capítulo se describe la base conceptual que sostiene la formulación de una arquitectura “Cloud Computing”. Por lo tanto, es menester profundizar en los conceptos, la tecnología y la forma de modelar una arquitectura que precede a un sistema de este tipo.

2.1. Definición de Cloud Computing

Se entiende por “Cloud Computing” un paradigma que promueve a las organizaciones una nueva alternativa para la implantación de un modelo de negocio innovador. En términos prácticos, esto implica que de acuerdo a las necesidades que se tengan dichas administraciones podrían aumentar su capacidad de cómputo y almacenamiento de datos en tiempo real, generando así buenas oportunidades para las competencias de mercado. ¿En qué sentido influye entonces la implantación de este paradigma? Ciertamente es que la instauración de esta arquitectura contribuye en variados aspectos característicos en torno a la automatización de los requerimientos de negocios. Sólo por nombrar algunos se tiene: La **capacidad de planificación** que conlleva a determinar y cumplir con las futuras demandas de las organizaciones TI en términos de recursos, servicios y productos. **La reducción de costos** al prescindir de los gastos de infraestructura, administración, seguridad, personal técnico, etc; Y **la agilidad organizacional** permitiendo la adaptación ante cambios causados por problemas internos y/o factores exter-

nos. Es importante destacar que este concepto no solo responde a una lógica de negocios y su posición en el mercado, sino también extiende su usabilidad a organizaciones de carácter científico que buscan brindar nuevas herramientas para la investigación y el avance de un área específica.

A continuación se plantean dos definiciones que abarcan este concepto de manera genérica. La primera, que proviene del instituto nacional de estándares y tecnología (NIST) y la segunda, correspondiente al libro “Cloud Computing: Concepts, Technology & Architecture”.

1. *“Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models, and four deployment models”.* [11]
2. *“Cloud computing is a specialized form of distributed computing that introduces utilization models for remotely provisioning scalable and measured resources”.* [5]

Como factor común se entrevé en las definiciones que el paradigma Cloud Computing brinda una forma de gestionar recursos computacionales de manera remota y efectiva, reduciendo el esfuerzo administrativo y la interacción con el proveedor de dicho servicio. Por otra parte, advierte que es factible realizar lo descrito mediante la incorporación de ciertos “modelos” orientados a la escalabilidad. Por lo tanto, es importante entender que este paradigma no es una tecnología nueva per se, sino más bien un modelo de prestación de servicios el cual tiene ciertos principios tecnológicos. A modo de ejemplo, se considera necesaria la existencia de una fuerte **capa de virtualización, elasticidad** a la hora de escalar recursos TI, capacidad de **aprovisionar dichos recursos** (agrupar estos para servir a múltiples consumidores). También, se evidencia la necesidad de **orientar la arquitectura a servicios (SOA)** la cual promoverá la escalabilidad deseada y por último, **la estandarización de cada servicio** con el fin de facilitar la arquitectura orquestada.

En la siguiente sección se especificará sobre los modelos aludidos en las definiciones brindadas.

2.2. Modelos para una arquitectura “Cloud”.

En relación a los modelos especificados en las definiciones precedentes se manifiesta la existencia de dos conjuntos de modelos que propician la conformación de una arquitectura de este tipo. Una primera lista de modelos denominados “Cloud Delivery Models” que representan, a grandes rasgos, una combinación pre-ensada de recursos TI ofrecidas por un proveedor “Cloud”; y una segunda lista denominada “Cloud Deployment Models” los cuales describen un tipo específico de distribución en la nube que se distingue por el tamaño, el acceso y la propiedad. A continuación se describen detalladamente los entornos de cada modelo tipo.

2.2.1. Modelos de servicios.

Los servicios “Cloud” se distribuyen entre todas las capas arquitectónicas de un sistema de informático tradicional (sistema monolítico). Esto es, desde la capa de hardware hasta la capa de software. Desde acá entonces, se exhiben tres entornos que pueden estar interrelacionados en el sentido de que el alcance de uno puede abarcar el otro:

- **IaaS (Infraestructura como servicio):** Este entorno está compuesto por recursos TI centrados en la infraestructura los cuales pueden ser accedido a través de interfaces y herramientas basadas en servicios en la nube. Por consiguiente, los elementos que puede incluir este entorno abarcan no solo el hardware como tal, sino también la conectividad, los sistemas operativos y otros recursos “crudos”. Por lo tanto, el propósito general de esta categoría será brindar a los consumidores “Cloud” la posibilidad de adquirir recursos hardware (servidores, sistemas de almacenamiento, computadores, routers, etc.) como si se trataran meramente servicios externalizados. Por último, es importante resaltar que estos recursos TI difieren de un entorno tradicional ya que son virtualizados y empaquetados con el fin de personalizar la infraestructura en tiempo real.

- **PaaS (Plataforma como servicio):** Este modelo exhibe un conjunto de productos y/o herramientas desplegadas y preconfiguradas que permiten a los consumidores la creación de nuevas aplicaciones informáticas. En otras palabras se brindan acá entornos de desarrollo e interfaces de programación con objeto de conseguir escalabilidad y reducir costos asociados a la configuración y mantenimiento de recursos locales. Desde acá es importante evidenciar que estos recursos a su vez deben ser sustentados sobre una infraestructura subyacente, la cual es proporcionada por un entorno IaaS. La figura 2.1 exhibe una imagen extraída de [5] la cual expresa lo descrito.

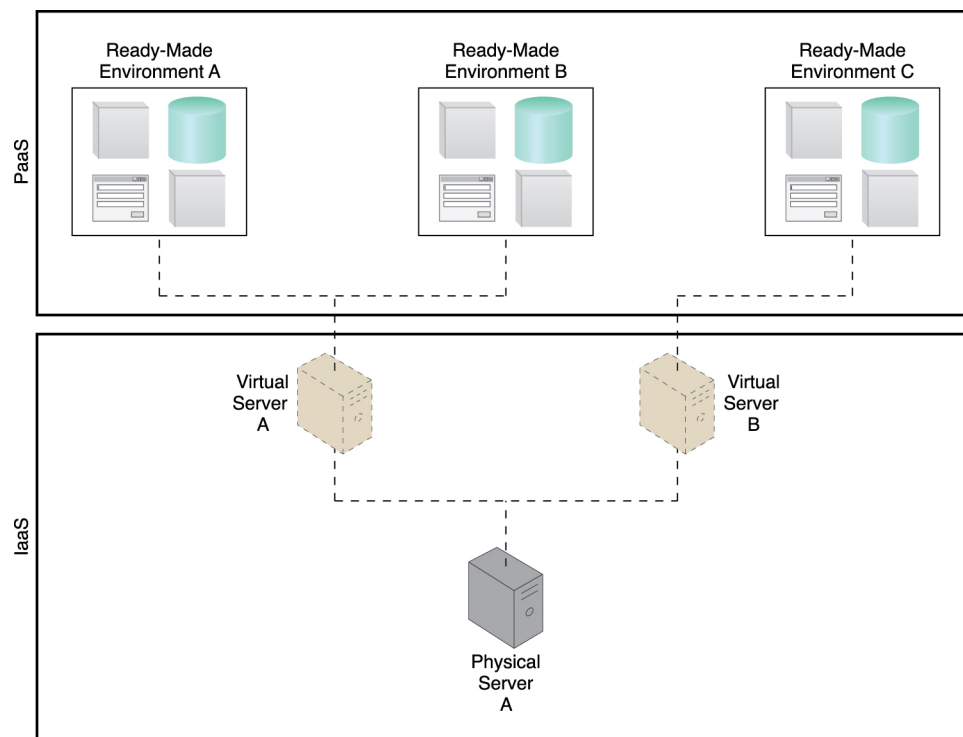


Figura 2.1: Entorno PaaS basado en los recursos TI proporcionados por un entorno IaaS subyacente.

- **SaaS (Software como servicio):** Este entorno brinda la posibilidad de posicionar una aplicación (software) como un servicio compartido en la nube. Por lo tanto, dicho servicio es reutilizable y alcanza una alta disponibilidad para los consumidores “Cloud”. Por otra parte, es importante manifestar que el ambiente PaaS puede crear un servicio en la nube que decida implementar un entorno del mismo, pero para un ambiente SaaS. Por lo tanto, se

distingue con esto que los tres modelos comprenden una jerarquía de aprovisionamiento natural la cual permite combinarlos para la conformación de sistemas más complejos. La figura 2.2 exhibe una imagen extraída de [5] que expresa lo descrito.

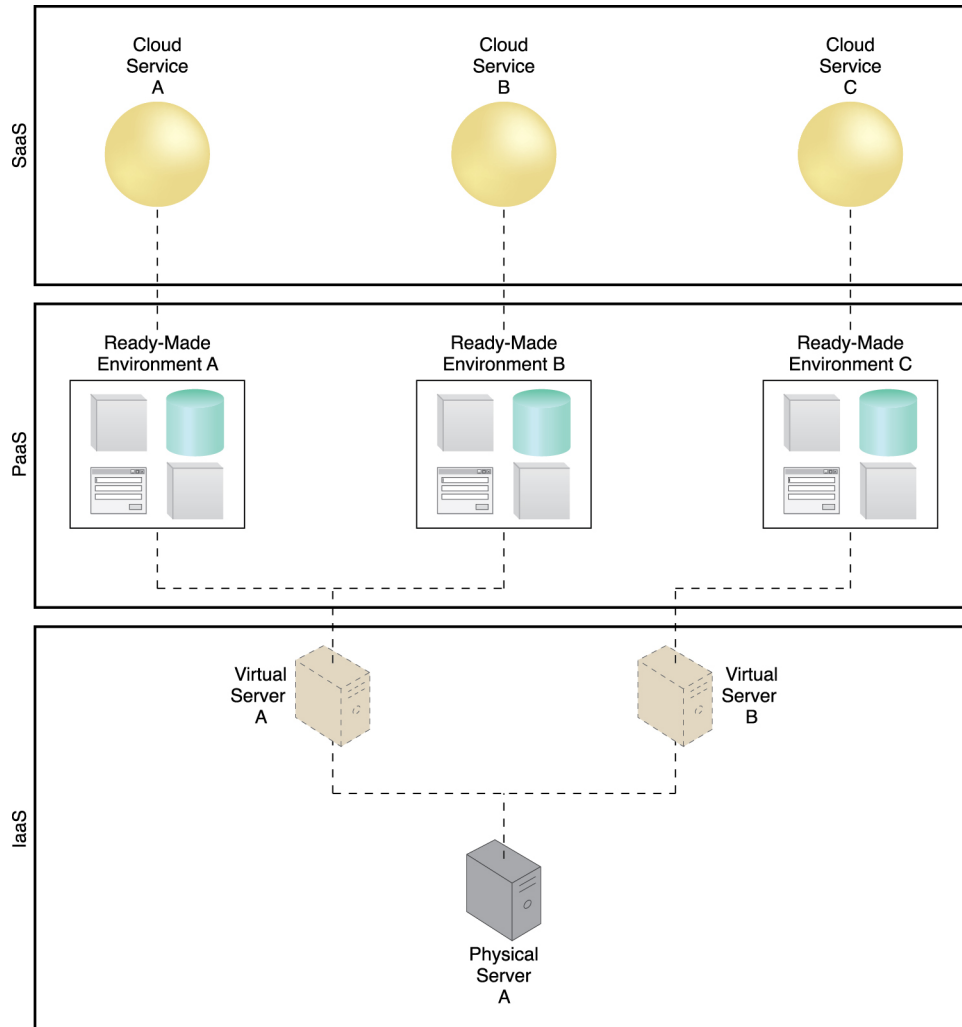


Figura 2.2: Una simple vista compuesta de entornos IaaS y PaaS que albergan tres implementaciones de servicios del entorno SaaS.

2.2.2. Modelos de despliegue

También denominados modelos de implantación. Describen esencialmente la manera en que se va a desplegar un servicio “Cloud” en términos de acceso, tamaño y propiedad. En otras palabras, manifiesta quien va a poder acceder a

dicho servicio y quien va a gestionar la infraestructura. Existen cuatro modalidades de implantación (tipos de nubes) de acuerdo a los criterios establecidos para el sistema.

- **Nube pública:** En este modelo el proveedor de servicios da acceso público a la nube, esto es, brinda la posibilidad a cualquier usuario de internet de acceder a sus servicios de infraestructura y/o software. Todo esto, mediante un costo de servicio o de manera gratuita. Este modelo tipo se utiliza cuando a la organización que ofrece el servicio no le importa que otros usuarios utilicen sus recursos virtualizados o cuando el despliegue de los servicios es provisorio.
- **Nube comunitaria:** Manifiesta lo mismo que una nube pública, sin embargo, el acceso se limita a una comunidad específica de consumidores.
- **Nube privada:** En este modelo se proporciona acceso privado a la organización, promoviendo la utilización de tecnología “Cloud Computing” y junto a ello centralizando el acceso a los recursos TI para una sección específica. Por ejemplo, para departamentos de la organización.
- **Nube híbrida:** Manifiesta un modelo de implantación compuesta por cualquiera de las alternativas señaladas anteriormente. Desde acá una organización (consumidor) puede optar para ciertos casos, por temas de seguridad, emplear una nube privada para datos confidenciales y por otra parte utilizar un modelo de nube pública para datos menos restrictivos.

2.3. Arquitectura orientada a servicios (SOA)

Tal como se describió en la sección anterior, en el proceso de conformación de una estructura “Cloud Computing” resulta fundamental la orientación a modelos de servicios. En términos prácticos, esto determinará la escalabilidad deseada de cualquier sistema informático que se desee desarrollar dado que permiten la reutilización de funcionalidades, encapsulamiento y nuevas oportunidades tanto para proveedores como consumidores.

2.3.1. ¿Qué es una arquitectura orientada a servicios?

Se define una arquitectura orientada a servicios como *“la representación de una estructura abierta, flexible y federada basada en composición, que promueve la orientación de servicios interoperables e independientes de los proveedores, los cuales pueden ser identificados en catálogos con un gran potencial de reutilización e implementados como servicios Web. SOA puede establecer una abstracción de la lógica de negocios y la tecnología, resultando un bajo acoplamiento entre dominios.”*.

Efectivamente, es factible entender SOA como un patrón de diseño que estandariza los aspectos de creación y uso de servicios durante su correspondiente ciclo de vida. Por lo tanto, el planteamiento de este tipo de arquitectura permite la construcción de aplicaciones distribuidas y en donde cada servicio implementado es distribuido en componentes que proveen interfaces definidas, que para el caso de los servicios web procesan y distribuyen mensajes basados en XML u otros formatos de intercambio de datos.

2.3.2. Relación Cloud Computing - SOA

En la figura 2.3 se puede apreciar entonces que SOA es uno de los elementos fundamentales de una arquitectura “Cloud Computing” se destaca esto dado que en las secciones venideras se exhibirá una lista de servicios encapsulados que en su conjunto conforman la plataforma ChiVO. Por lo demás, la imagen expuesta en la figura 2.3 representa un resumen de este capítulo en donde se utilizaron, en su mayoría, los conceptos que aparecen para su útil definición.

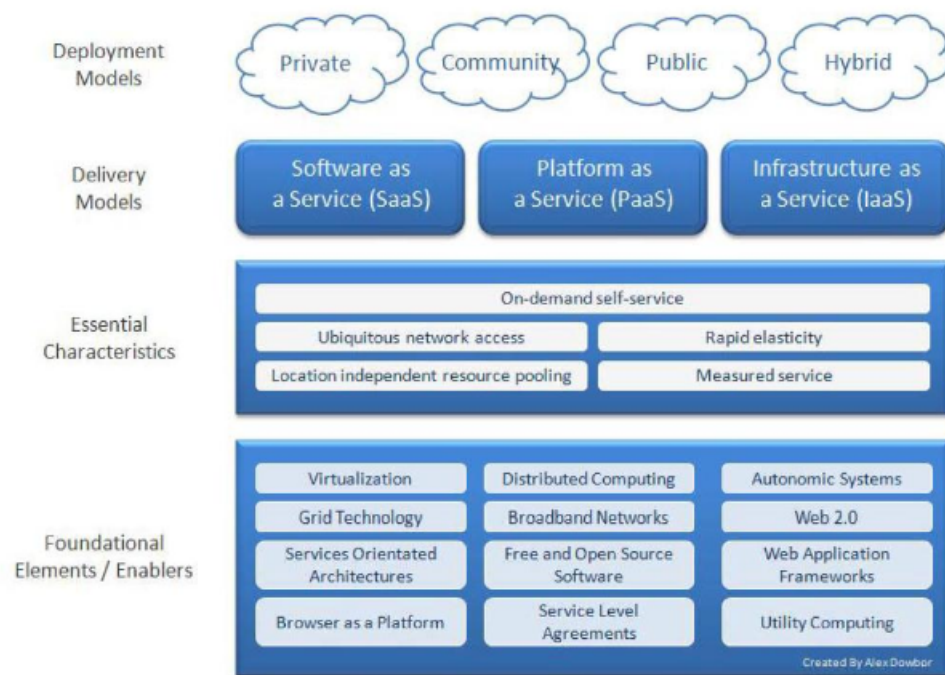


Figura 2.3: Elementos esenciales arquitectura “Cloud Computing”.

Capítulo 3

Arquitectura genérica de un observatorio virtual

En esta sección se detalla información asociada al proyecto con el fin de establecer la base conceptual para erigir la nueva arquitectura de ChiVO (observatorio virtual Chileno). De manera más específica, se brindarán fundamentos que permiten encauzar el proceso de diseño arquitectónico del observatorio virtual Chileno desde el más alto nivel, descendiendo hacia la exposición de los servicios y por último evocando los protocolos y estándares especializados para la conformación de cada uno.

3.1. Definición Observatorio Virtual

Para comenzar entonces es necesario contextualizar al lector detallando que es un observatorio virtual (VO) según la alianza internacional de observatorios virtuales (IVOA).

“The VO is not a concrete thing, like a data warehouse. Rather, it is more like an ecosystem of mutually compatible datasets, resources, services, and software tools which use a common set of technologies and a common set of standards. The idea is to make all these things inter-operable - i.e. to make them work nicely together. However, the VO is more than just a set of rules for everybody to follow; it also requires some specialised middleware to glue things together - for example Registry services, distributed storage, sign-on services and so on”. [1]

En términos prácticos, esta definición manifiesta que un observatorio virtual no es simplemente un almacén de datos, sino más bien, una estructura que permite orientar los requerimientos de la comunidad para poder trabajar con bases de datos extremadamente grandes y diversas a través de variados servicios y herramientas de software mutuamente compatibles que utilizan un conjunto de estándares y protocolos. Por lo tanto, para el caso de ChiVO (miembro de IVOA) lo descrito implica que el sistema debe adaptarse entonces a los estándares promovidos por la organización con el fin de conseguir interoperabilidad en sus servicios y herramientas de software. Por lo tanto, es menester precisar a continuación cuáles son dichos protocolos de acuerdo a los objetivos que persigue la plataforma astronómica Chilena.

3.2. Estándares de interoperabilidad de IVOA

Para comprender cabalmente los conceptos expuestos por IVOA, es preciso resaltar que la organización recomienda que el desarrollo de un observatorio virtual debe ceñirse bajo una arquitectura descrita en tres niveles lógicos con el objetivo de vislumbrar que debiese considerarse para la implementación de una plataforma VO. Ahora bien, dentro de estos tres niveles, también se identifican tres capas cada una con objetivos claramente definidos. Estas son:

- **La capa de recursos (“Resource Layer”):** Proporciona a la comunidad científica los servicios computacionales y el compilado de datos proveniente de las diversas fuentes competentes (observatorios, modelos teóricos, etc.).
- **La capa de usuarios (“User Layer”):** Representa a los consumidores de dichos datos y los servicios brindados.
- **La capa intermedia (“Middle Layer”):** Persigue el objetivo de conectar las dos capas predecesoras de manera **transparente**.

A continuación se profundizará sobre el concepto de transparencia en la capa intermedia, específicamente exhibiendo la figura 3.1 (extraída de [13]), la cual muestra el nivel dos de la arquitectura de IVOA y detalla los protocolos y estándares agrupados en subsecciones.

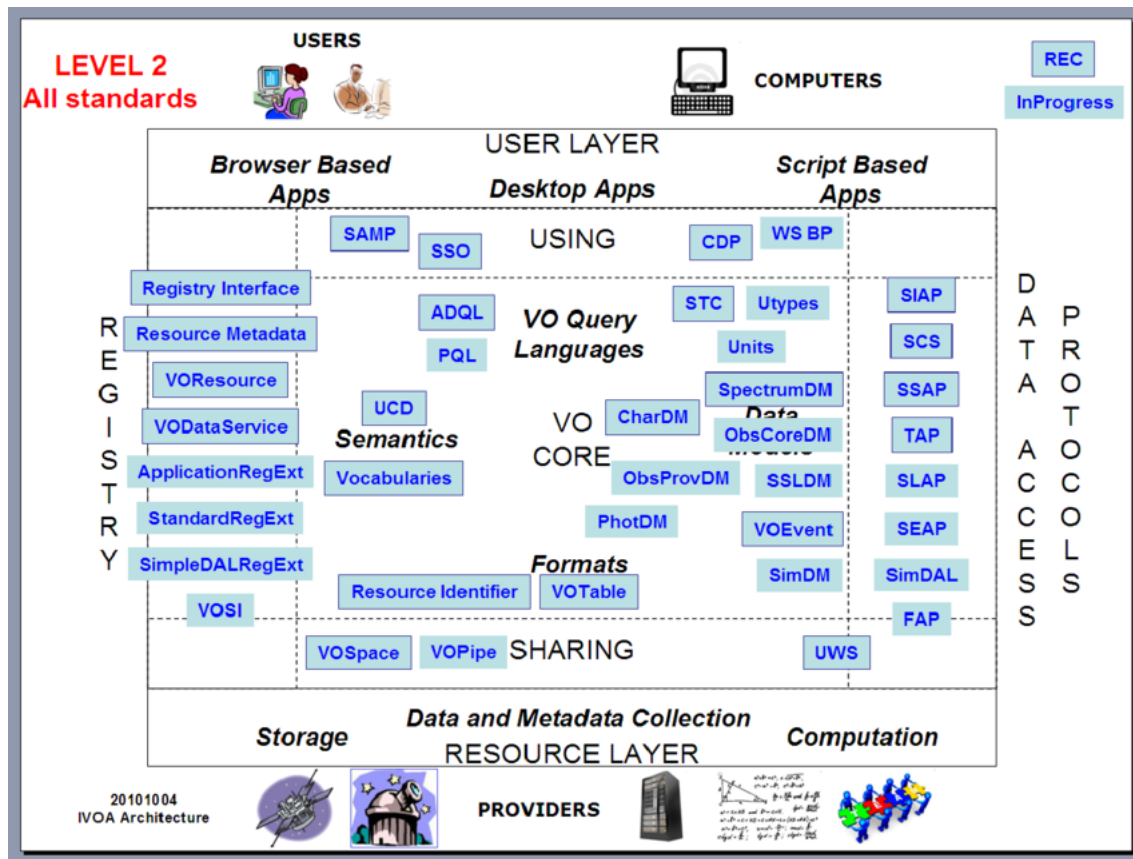


Figura 3.1: Arquitectura nivel 2 de IVOA.

3.2.1. Capa intermedia: Subsección REGISTRY.

En el panel izquierdo de la capa intermedia, se encuentra la subsección “REGISTRY” la cual contiene las descripciones de los recursos VO. Se entiende por un recurso VO al término general que alude a un elemento (VO); dicho elemento es descrito en términos prácticos con un nombre, un identificador único y quien lo “curará” (mantendrá). Junto a esto, es importante destacar que este concepto puede ser “cualquier cosa”, por ejemplo la configuración de un instrumento, una idea abstracta, una organización, una colección de datos, una aplicación, un servicio de acceso a los datos, entre otros. Por lo tanto, la finalidad de estos estándares será permitir a los agentes participantes en la capa de usuario descubrir los servicios y las colecciones de datos (y metadatos) de la capa de recursos a través de mecanismos de identificación. En otras palabras, un registro es un “servicio” de consulta para el cual la respuesta es una descripción estructurada de los datos.

- **VOResource:** Representa un estándar de codificación XML el cual describe un recurso utilizando los mismos conceptos expuestos en la tabla anterior. Por lo tanto, uno de los principales usos para este estándar será proporcionar un formato de intercambio XML para el uso de los elementos VO. Por otra parte, cabe destacar que VOResource también puede ser utilizado para pasar descripciones de editores a registros (repositorio de descripciones de los recursos) y luego dichos registros a aplicaciones.
- **Registry Interface:** Proporciona un modelo general para la descripción de los registros como recursos VO utilizando una extensión del esquema de metadatos de VOResource. Por lo tanto, este estándar define las interfaces para la comunicación e interacción de los registros con las aplicaciones y/o entre los propios registros concibiendo estos como recursos. Adicionalmente, se destaca que este es un modelo distribuido y está compuesto por dos interfaces denominadas: “Registros de búsqueda” y “Registros de publicación”. La primera interfaz (para buscar), permite a los agentes de la capa de usuario hallar registros de recursos utilizando ciertos criterios de selección; la segunda interfaz (para publicar), predispone un mecanismo de exposición de descripciones de recursos para el entorno VO con el fin de facilitar el proceso de recolección de los mismos. Por último, el lector podrá encontrar más información de este estándar en documento [14]

- **Resource Metadata:** Término genérico de alto nivel que promueve la independencia para un elemento VO específico mediante la descripción detallada del mismo. En efecto, considerando que los recursos VO siempre son descritos a través de metadatos, este estándar constituye una especie de “páginas amarillas” de la información astronómica. En la tabla 3.1 se indica la forma en la que se describe un elemento VO de acuerdo a este estándar [13].

Tabla 3.1: Tabla descriptiva para estándar “Resource Metadata”.

Identity Metadata	Recurso adquiere un nombre y un identificador (campos son obligatorios). Se tiene la posibilidad de agregar como atributo una abreviación para el recurso.
Curation Metadata	Describe quien brinda soporte al recurso y su disponibilidad. En este tópico existen atributos tales como el publicador (responsable por el recurso disponible), la id del publicador, el creador, la versión, el contacto, entre otros.
General Content Metadata	Exhibe el tipo de información de disponible para el recurso, de manera práctica. Puede contener los tipos de datos, la cobertura espectral o del cielo, entre otros. Dentro de los atributos considerados se exhibe una lista de palabras claves, una descripción del recurso, una definición, páginas de referencia, entre otros.
Collection and service content metadata	Este campo detalla una profundización del contenido general. Atributos de referencia pueden ser, “facility” (qué observatorio o instalación está facilitando los recursos), con qué instrumentos, qué zona cubren los datos, el formato, entre otros.
Data quality metadata	Exhibe a grandes rasgos la calidad de los datos, de los recursos y las interfaces (cuando es aplicable) mediante una calificación; dicha ponderación se obtiene mediante ciertos aspectos de calidad tales como calibración, documentación, entre otros. Desde acá atributos referenciales pueden ser, nivel de validación de los datos, comentarios, URLs con la (s) organización (es) que validan dichos datos, entre otros.
Service metadata	Cuando se desea describir un servicio, se utiliza este campo extra que tiene un patrón de exhibición según el tipo de servicio planteado. Desde acá atributos referenciales son como se puede acceder al servicio (interface metadata), descripción, limitaciones del mismo, entre otros.

3.2.2. Capa intermedia: Subsección DATA ACCESS PROTOCOL STANDARDS.

En el panel derecho de la capa intermedia, se encuentra la subsección “DATA ACCESS PROTOCOL” la cual provee un conjunto de servicios de consulta y acceso a colecciones de metadatos y datos distribuidos. En específico, cada uno de estos estándares se utilizan para un tipo particular de datos los cuales podrían ser, a modo de ejemplo, una imagen, un espectro, entre otros. Por lo tanto, a diferencia con la sección REGISTRY que exhibía simplemente las descripciones de los recursos VO contenidos, esta interfaz proporciona ahora (a los agentes de la capa de usuario) la posibilidad de acceder al conjunto de datos almacenados en distintas ubicaciones.

- **SIAP - Simple image Access Protocol:** Este protocolo proporciona los mecanismos para el descubrimiento y la recuperación de datos de imágenes astronómicas de cualquier dimensión. La importancia de este estándar es que las imágenes descritas contienen en sí, cubos de datos espectrales e hipercubos de dimensiones superiores, además de la imagen espacial bidimensional. Así, realizar una consulta que define una región rectangular en el cielo permite obtener una lista para consultar las imágenes candidatas formateadas como VOTable. Cada imagen candidata entonces posee una URL de referencia de acceso para la recuperación de la información.
- **SCS - Simple Cone Search:** Este protocolo precisa la posibilidad de recuperar registros de un catálogo de fuentes astronómicas a través de una “consulta cónica simple” (Simple cone Search). Por esta razón, dicha consulta deberá describir la posición del cielo y una distancia angular, con el objetivo de definir un cono en el cielo. Como resultado, el estándar entrega una lista de fuentes astronómicas la cuales están inscritas en el cono conformado. La respuesta a una consulta de este tipo se encuentra en formato VOTable para propiciar el intercambio de metadatos; además, se consideran otros protocolos: STC y UCD para la descripción de los mismos. Es importante recordar que los tres protocolos forman parte del panel central de la arquitectura nivel dos de IVOA, por lo tanto serán descritos con mayor detalle en la próxima sección.

- **TAP - Table Access Protocol:** Es un protocolo de servicio Web que permite el acceso a colecciones de datos tabulares (datos expresados a través de tablas). En otras palabras, define una interfaz para la consulta de catálogos astronómicos o tablas de bases de datos generales.
- **SSA - Service Spectral Protocol:** Define una interfaz uniforme para descubrir y acceder de forma remota a un espectro dimensional procedente de colecciones de datos y metadatos. Es capaz de describir la mayoría de los datos tabulares espectrofotométricos, incluyendo series temporales y distribuciones de energía espectral (SED), así como espectros en una dimensión.
- **SLAP - Simple Line Access Protocol:** Define una interfaz uniforme para descubrir y acceder a transiciones de líneas espectrales procedentes de colecciones de datos y metadatos.

3.2.3. Capa intermedia: Subsección VO CORE.

En el panel central de la capa intermedia, se encuentra la subsección “VO CORE” la cual proporciona un conjunto de estándares categorizados (en cuatro grupos), y dirigidos especialmente para desarrolladores y proveedores de servicios. En términos prácticos, estas utilidades clasificadas persiguen distintos objetivos tales como: determinar los tipos de modelo de datos permitidos, la semántica que se debe seguir, formatos especiales y cómo elaborar una consulta de manera estandarizada. A continuación, se detalla brevemente cada uno de estos con los protocolos de particular interés.

VO CORE: DATA MODELS STANDARDS

Tal como fue descrito al comienzo de este capítulo, un observatorio virtual provee un almacén de datos y ciertas herramientas para el análisis, procesamiento y visualización de datos astronómicos mediante la instauración de varias estructuras modeladas (estandarizadas) que se adecuan a los distintos objetivos planteados. Por lo tanto, dichos modelos gestan el marco ideal para describir las relaciones entre las colecciones de datos astronómicos y sus metadatos respectivos con el fin de alcanzar interoperabilidad y transparencia. A continuación, se describen los modelos de datos VO de relevancia para la plataforma astroinformática chilena en su etapa actual:

- **STC - Space Time Coordinate Metadata:** Este estándar proporciona la estructura para describir los metadatos espaciales y temporales. En la sección subsecuente el lector podrá vislumbrar la presencia de éste en la mayoría de los servicios y estándares de IVOA orientados a la búsqueda espacial y/o temporal. Por lo que sigue, dentro de la estructura se define metodologías de acceso a los datos y lenguajes de consulta VO.
- **Units:** Esta estructura evidencia la composición de una unidad de medición ligada a una determinada cuantía (valor) para un concepto concreto. De manera específica, representa una cantidad estandarizada de una determinada magnitud física, definida y adoptada por convención. Ejemplo de esto: metros por segundo, gramos, Joules, Kelvin, Ampere, etc. Por lo tanto, para el caso de un observatorio virtual las “Unidades VO DM” tiene como objetivo definir prácticas comunes en la manipulación de unidades en metadatos astronómicos.
- **Utypes:** Este modelo manifiesta unicidad o especificidad para un atributo (parámetro) explícitamente diseñado en algún modelo de datos. La noción central de esto, es poder dictaminar que ciertos campos tendrán, de manera universal, el mismo significado. Adicionalmente, se destaca que UTypes puede ser utilizado para definir recursos, tablas y grupos.
- **ObscoreDM - Observation Data Model Core Components:** Este modelo define los componentes básicos que deben tener los metadatos para la búsqueda oportuna de datos distribuidos. En profundidad, bajo la implementación de esta estructura se provee mayor facilidad de uso para la comunidad en la búsqueda de datos a través de distintos VO's asociados a IVOA; todo esto dado que incorpora los diversos protocolos recomendados, de gran utilidad para el descubrimiento global de observaciones (p. ej. OBS-TAP que es una interfaz TAP para ObscoreDM). Junto a esto, es importante destacar que en este marco se hace uso de de STC, Utypes y Units (descritos anteriormente) por lo que es factible aseverar la globalidad en el proceso de rastreo.

Ahora bien, dado que resulta imprescindible para el proceso de inserción de datos para la plataforma ChiVO y también para conseguir interoperabilidad

con otras fuentes de datos astronómicas se exhibe en la tabla 3.2 los campos obligatorios que considera este modelo con nombre, unidades recomendadas, tipo de datos y descripciones (La información expuesta fue extraída del documento [12]).

Tabla 3.2: Campos modelo ObscoreDM. Fuente [12]

Column Name	Unit	Type	Description
dataprodukt_type	unitless	String	Logical data product type.
calib_level	unitless	enum integer	Calibration level {0, 1, 2, 3, 4}
obs_collection	unitless	String	Name of the data collection.
obs_id	unitless	String	Observation ID
obs_publisher_did	unitless	String	Dataset identifier given by the publisher
access_url	unitless	String	used to access (download) dataset
access_format	unitless	String	File content format (see in App. B.5.2.)
access_estsize	kbyte	integer	Estimated size of dataset in kilo bytes
target_name	unitless	String	Astronomical object observed, if any
s_ra	deg	double	Central right ascension, ICRS
s_dec	deg	double	Central declination, ICRS.
s_fov	deg	double	Diameter (bounds) of the covered region
s_region	unitless	String	Sky region covered by the data product (expressed in ICRS frame)
s_xel1	unitless	integer	Number of elements along the first spatial axis
s_xel2	unitless	integer	Number of elements along the second spatial axis

Continúa en la siguiente página.

Column Name	Unit	Type	Description
s_resolution	arcsec	double	Spatial resolution of data as FWHM
t_min	d	double	Start time in MJD
t_max	d	double	Stop time in MJD
t_exptime	s	double	Total exposure time
t_resolution	s	double	Temporal resolution FWHM
t_xel	unitless	integer	Number of elements along the time axis
em_min	m	double	Start in spectral coordinates
em_max	m	double	Stop in spectral coordinates
em_res_power	unitless	double	Spectral resolving power
em_xel	unitless	integer	Number of elements along the spectral axis
o_ucd	unitless	String	UCD of observable (e.g. phot.flux.density, phot.count, etc.)
pol_states	unitless	String	List of polarization states or NULL if not applicable
pol_xel	unitless	integer	Number of polarization samples
facility_name	unitless	String	Name of the facility used for this observation
instrument_name	unitless	String	Name of the instrument used for this observation

VO CORE: QUERY LANGUAGE STANDARDS

“Structured Query Language” (más conocido como “SQL”) representa un estándar para emitir consultas a bases de datos. Este lenguaje actualmente posee ciertas variantes de acuerdo a las necesidades de un área específica. En términos prácticos, acá se presentan lenguajes adoptados en IVOA para las consultas a datos en el campo astronómico (p. ej. la búsqueda dentro de una región circular).

A continuación, se exhibe sólo uno de estos en función a los objetivos planteados en ChiVO.

- **ADQL - Astronomical Data Query Language:** Este estándar representa uno de los lenguajes ideados por IVOA para representar las consultas de astronomía enviadas a los servicios habilitados en un VO. De manera específica, un observatorio virtual puede poseer una serie de datos tabulares almacenados en una o más bases de datos relacionales. Por lo tanto, se justifica la incorporación de un lenguaje de consultas ad hoc para el área de la astronomía como un medio de acceso conveniente a dichos datos. ADQL, basado en SQL, cumple cabalmente el objetivo planteado y proporciona todos los mecanismos necesarios para suplir las insuficiencias de su predesor en este campo. Ejemplificando, se brindan a continuación algunas consultas “tipo” que dictaminan los alcances de este lenguaje:

- **SELECT TOP 150 * FROM “CHIVO_FitsTable”;**

Esta función entrega los primeros 150 registros de la tabla “ChiVO_fitstable”.

- **SELECT “RA(ICRS)”, “DE(ICRS)”, Btmag, VTmag, BTmag - VTmag AS BV FROM “CHIVO_FitsTable” WHERE Btmag - VTmag > 0 AND BTmag - VTmag < 0,2;**

Esta función devuelve los registros explicitados filtrados por el cálculo de magnitudes de color del catálogo “ChiVO_fitstable” que cumplen con las restricciones expuestas. Se puede apreciar que entre los registros seleccionados se encuentra la obtención de los parámetros posicionales en formato ICRS (Sistema de Referencia Celestial Internacional).

- **SELECT TOP 100 HIP, RA(ICRS)”, ”DE(ICRS)”, pmRA, pmDE, SQRT(POWER(pmRA,2)+POWER(pmDE,2)) AS pm FROM “ChiVO_fitstable”;**

Esta función retorna las posiciones y movimientos adecuados de la tabla “ChiVO_fitstable”, seleccionando los primeros 100 registros y calculando el módulo vectorial con los parámetros de distancia.

VO CORE: SEMANTICS

En esta sección se especifican formatos estandarizados para vocabularios basados en el Marco de descripción de recursos (RDF) del W3C y el Sistema de organización de conocimientos simple (SKOS). El uso de estos estándares propicia que las aplicaciones VO sean capaces de aprovechar los recursos combinando o gestando nuevos vocabularios en esta creciente red semántica.

- **UCD - Unified Content Descriptor:** Este estándar presenta el vocabulario IVOA controlado para describir las cantidades de datos astronómicos. En definitiva, un UCD es una cadena que contiene fichas textuales llamadas “words”, separadas por punto y coma (;). Dichas palabras a su vez están compuestas de “atoms” (abreviaciones), separados por puntos (.) y posiblemente ciertas palabras compuestas.

VO CORE: FORMATS STANDARDS

En esta sección se detalla uno de los formatos estándar de IVOA para la administración de datos astronómicos. De manera específica, es sabido que los datos presentan una estructura que muestra información descriptiva de los mismos (metadatos). Por lo tanto, dicha estructura fue estandarizada para poder adquirir, procesar, intercambiar y archivar cierto tipo de datos específicos (“crudos”) del área.

- **VOTable:** Es un formato XML flexible para el almacenamiento e intercambio de datos tabulares; en otras palabras, para la administración de tablas astronómicas. Es utilizado en diversos contextos VO dada la amplia gama de herramientas existentes en el área que permiten la exportación de datos a este formato. Por otra parte permite que tablas FITS puedan ser representadas en este estándar, resultando esto de gran utilidad dado el alto uso actual por la comunidad astronómica. Ciertamente, el uso de normas XML potencia la interoperabilidad entre plataformas virtuales y permite el uso de métodos efectivos para la validación de documentos de entrada. VOTable cuenta con funciones integradas para el manejo de grandes volúmenes de datos en entornos de computación distribuida. A su vez, posee un almacenamiento flexible y eficiente, en términos de representación; esto es, puede

ser estructurado en cuatro formatos diferentes (TABLEDATA, FITS, BINARY y BINARY2) de acuerdo a las pretenciones de un VO. Por último, es menester señalar que para representar metadatos, VOTable se basa en los otros estándares IVOA “UCD”, “Utype”, “Units” y “STC”.

3.2.4. Capa intermedia: Subsección SHARING.

En esta última sección a describir, se exhiben estándares para que los usuarios compartan sus datos y servicios.

- **VOSpace:** Es factible considerar a este estándar como un punto de acceso para una red distribuida de almacenamiento de datos, en el cual los usuarios VO y aplicaciones podrían administrar estos, manipular metadatos, obtener y compartir URIs.
- **UWS:** Como servicio “SHARING”, este estándar provee mecanismos para la distribución de datos en contexto de flujos de trabajo. Todo esto, mediante la ejecución asíncrona de trabajos (Se posibilita la ejecución de labores de manera diferida en el tiempo).

Capítulo 4

Arquitectura Cloud Computing para ChiVO

Esta sección tiene por objetivo definir la arquitectura Cloud para la plataforma astronómica ChiVO de acuerdo a los conceptos entregados en las secciones anteriores. Por consiguiente, se brindarán a continuación los diversos modelos funcionales para este sistema VO los cuales no sólo exhiben los servicios proyectados sino también la interacción y su comportamiento con los componentes especializados para este tipo de tecnologías en la nube.

Recordando la definición brindada por el instituto nacional de estándares y tecnología (NIST), el paradigma tecnológico Cloud Computing es una estructura que facilita el acceso a la red a un conjunto compartido de recursos computacionales configurables (redes, servidores, discos de almacenamiento, aplicaciones y/o servicios), los cuales pueden ser agrupados o liberados mediante diversos mecanismos de gestión rápidamente y con un mínimo esfuerzo. Así pues, la estructura descrita está conformada a su vez por modelos (de servicios y de implantación) que permiten, en definitiva, el despliegue y la construcción de aplicaciones distribuidas. Se vislumbra como primer paso para la conformación de la arquitectura de ChiVO la identificación y posterior posicionamiento de los recursos TI en los modelos descritos [5].

4.1. Definición modelos de servicios para ChiVO.

Tal como fue descrito en el capítulo dos, los modelos de servicios (“Cloud delivery models”) representan una combinación específica de recursos TI pre-ensados y ofrecidos por un proveedor Cloud. Estos modelos que manifiestan la estructura que puede adoptar la nube, se dividen en tres categorías de servicios (Infraestructura, Plataforma y servicio) las cuales se interrelacionan cuando el alcance de una abarca a la otra. A continuación, en la figura 4.1 se predispone un diagrama de alto nivel que manifiesta simplemente los conceptos a considerar en cada categoría, de acuerdo a los requerimientos actuales de ChiVO, todo esto con propósito de desarrollar luego cada abstracción con profundidad.

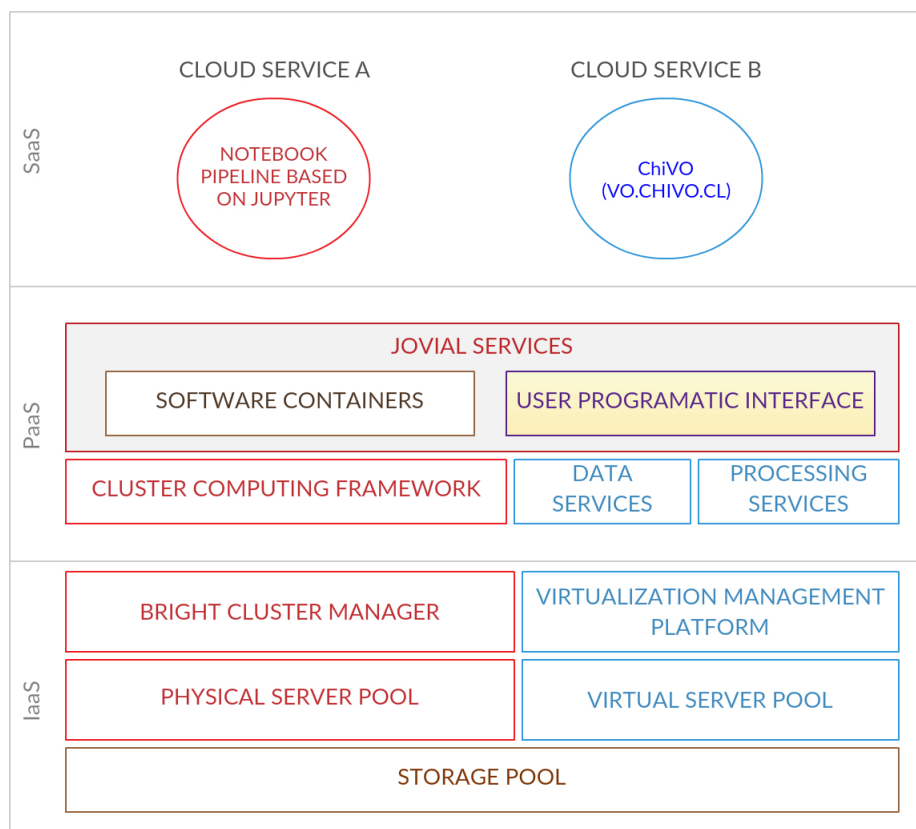


Figura 4.1: Diagrama alto nivel arquitectura ChiVO.

4.1.1. Infraestructura como servicio ChiVO

El modelo IaaS expuesto en la figura 4.2 expone los recursos TI centrados en las infraestructura los cuales serán accedido posteriormente a través de interfaces y herramientas basadas en servicios en la nube. Por lo tanto, resulta necesario describir cada concepto enmarcado en el diagrama para dilucidar la estructura de estos y junto a ello la relación entre cada componentes que tendrán en el plataforma astronómica Chilena.

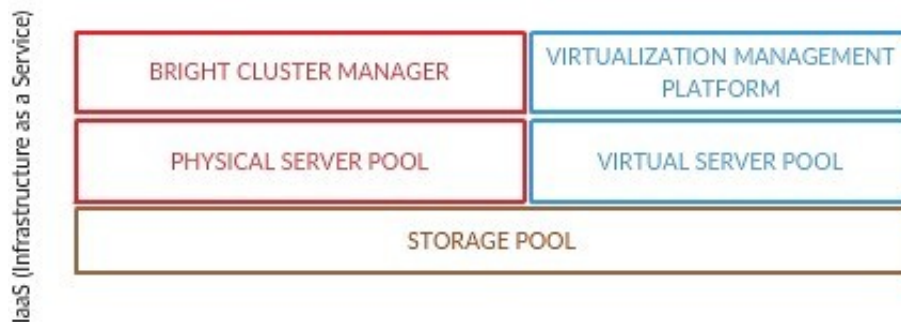


Figura 4.2: IaaS ChiVO Alto Nivel.

STORAGE POOL

Este término hace alusión a un conjunto de dispositivos de almacenamiento los cuales constan de estructuras basadas en bloques o archivos que poseen la capacidad de guardar recursos en entornos de almacenamiento compartido. De tal forma, estos pueden ser configurados en diversos tamaños permitiendo mejoras de rendimiento, administración y protección de datos.

CHIVO STORAGE POOL En el Data Center de ChiVO se tiene un **1 PB** de capacidad de almacenamiento distribuido en varios dispositivos HDD (Hard Disk Drive) y en dónde, se tiene como objetivo cubrir los dos grandes desafíos que posee este sistema: Proveer alta disponibilidad de los servicios proyectados y proporcionar computación de alto rendimiento (HPC). En consecuencia de lo descrito es que surgen dos grandes bloques que permiten cumplir cabalmente con la finalidad deseada. El primero de estos, que fue denominado “Lustre Storage” brindará soporte para el desarrollo de simulaciones computacionales, la detección

de patrones y el procesamiento de datos en problemas complejos (es decir, HPC). El segundo grupo que albergará el volumen de datos recibidos de los observatorios astronómicos (“Bulk Data Storage”) y administrará el espacio “privado” para los usuarios de la plataforma (“UserSpace”). Lo descrito se aprecia gráficamente en la figura 4.3.

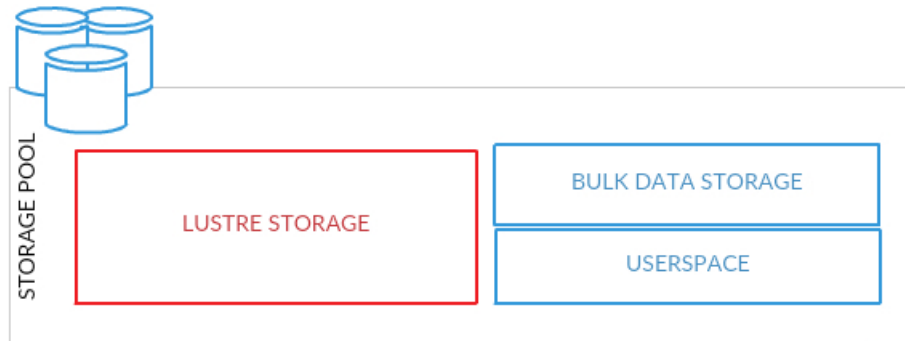


Figura 4.3: Storage Pool CHIVO.

PHYSICAL SERVER POOL

Es un conjunto de servidores físicos que poseen un set de aplicaciones instaladas necesarias para la conformación y posterior administración de un sistema de archivos distribuidos. Una de las más importantes es “Lustre”, que responde al cometido mediante la creación de una estructura estándar que permite maximizar la utilización eficiente de los recursos de almacenamiento (storage) incrementando el rendimiento y junto a ello la fiabilidad colectiva [3]. Se exhibe en la figura 4.4 un modelo estándar para la elaboración de la estructura de Lustre.

En relación a la figura 4.4 se percibe la existencia (de arriba hacia abajo) de un conjunto de discos de almacenamiento (“Lustre Storage”) seguido de un set de servidores (“Lustre Servers”) cada uno con funcionalidades específicas y también, mediante una conexión red, un conjunto de “Lustre Clients”. Es imperativo definir ahora las siglas presentes en el esquema ya que representan parte esencial para comprender la arquitectura del observatorio virtual Chileno. Estas se exhiben con detalle en la tabla 4.1:

Building blocks for Lustre scalable storage

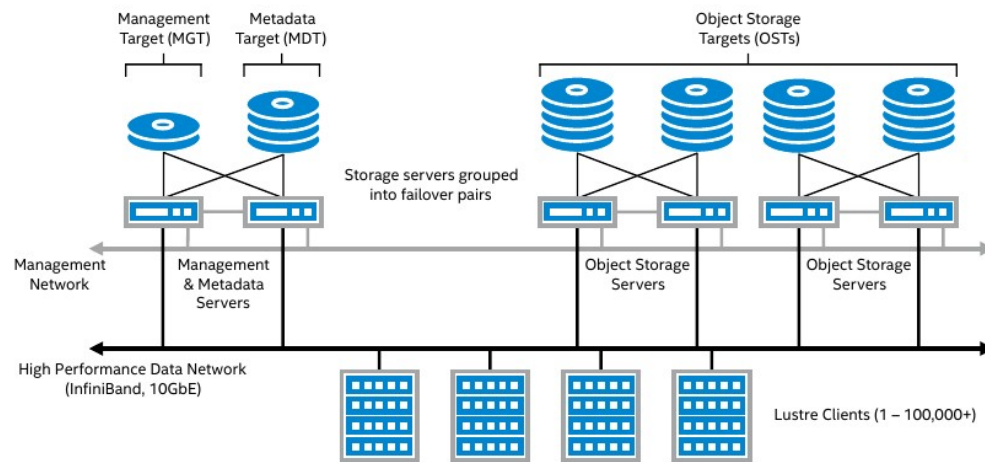


Figura 4.4: Estructura Lustre estándar para almacenamiento escalable.

Tabla 4.1: Detalles “Filesystem Lustre”

Metadata Server (MDS)	El servidor MDS brinda los metadatos almacenados en el (los) disco(s) lógico(s) a los “Lustre clients”. Cada MDS administra el espacio de nombres del sistema de archivos Lustre (“Lustre file system namespace”) y proporciona el manejo de solicitudes de red para uno o más MDT locales.
Metadata Target (MDT)	El MDT almacena los metadatos (Nombres de archivos, directorios, permisos, disposición de estos, etc.) en el almacenamiento adjunto a un Servidor de metadatos (MDS). Cada sistema de archivos tiene un MDT. En términos prácticos, un MDT es un destino de almacenamiento compartido que está disponible para varios MDS (aunque solo uno puede acceder a él a la vez). Por último si un MDS activo falla, un segundo servidor de metadatos puede servir al MDT y ponerlo a disposición de los clientes.

Continúa en la siguiente página.

Object Storage Target (OST)	En cada OST es posible almacenar uno o más objetos separados en un sistema de archivos Lustre. Es necesario entonces configurar el número de objetos por archivo para su posterior sincronización con el fin de optimizar el rendimiento de una carga de trabajo determinada.
Object Storage Servers (OSS)	Los servidores OSS proveen el servicio de entrada y salida de archivos y el manejo de solicitudes de red para uno o más OST locales. Normalmente, un OSS sirve entre dos y ocho OSTs, de hasta 16 TB cada uno.
Lustre clients	Los clientes lustre, en términos simples, son nodos computacionales destinados a la visualización o a la transferencia de datos que ejecutan un software lustre especializado para el cliente. Por lo tanto, esta aplicación provee una interfaz entre el sistema de archivos virtual de linux y los servidores lustre.

CHIVO PHYSICAL SERVER POOL En este tópico, ChiVO brinda una secuencia similar al esquema expuesto en la figura 4.4. Por este motivo es que presenta en su plataforma cuatro servidores “Lustre” que se encuentran divididos a su vez en dos servidores MDS y dos servidores OSS. Por lo tanto, se manifiesta la posibilidad de almacenar los metadatos en discos lógicos MDT y junto a ello los objetos (entrada y salida de archivos de estos) en los discos de almacenamiento OST cada uno proporcionando el manejo de solicitudes a los discos respectivos. Por lo que sigue, es interesante señalar que el proceso de inserción de datos, al menos en el caso de los metadatos, se basa en la utilización de los protocolos de registro de IVOA (Para más detalles de los estándares que proveen este servicio, ver capítulo tres sección “estándares de interoperabilidad de IVOA”). Por último, en la imagen 4.5 es posible visualizar una representación de lo descrito.

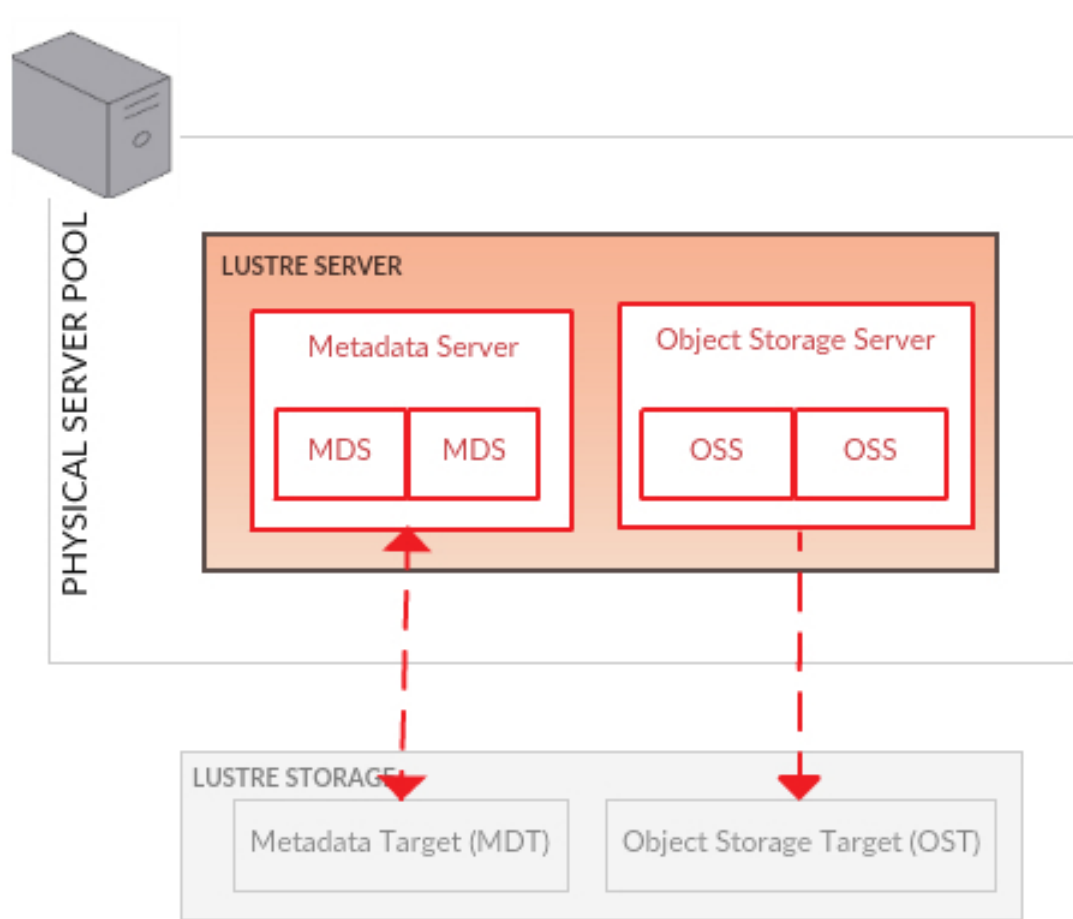


Figura 4.5: Conjunto de servidores físicos de ChiVO.

BRIGHT CLUSTER MANAGER

Dentro del esquema expuesto en la figura 4.4 se manifiesta la necesidad de implementar “Lustre Clients” los cuales se definieron anteriormente como nodos computacionales destinados a la visualización y/o a la transferencia de datos. Se destaca entonces la importancia de estos nodos dado que son los encargados de realizar el trabajo “duro” para la resolución de problemas complejos (procesamiento de datos). No obstante, es debido mencionar que algunos nodos tienen labores más simples por ejemplo, para el inicio de sesión o la ya mencionada transferencia de archivos.

En su forma básica, un clúster (conjunto de nodos computacionales) que ejecuta “Bright Cluster Manager[®]” (BCM) contiene esencialmente lo siguiente:

- Una máquina designada como nodo principal “Head Node”.
- Muchas máquinas designadas como nodos secundarios “Compute nodes”.
- Uno o más switches ethernet.
- Y opcionalmente, uno o más unidades de poder (“Power Distribution Units”).

A continuación se exhibe la imagen 4.6 que manifiesta la implementación estándar de un BCM (Imagen extraída de [2]).

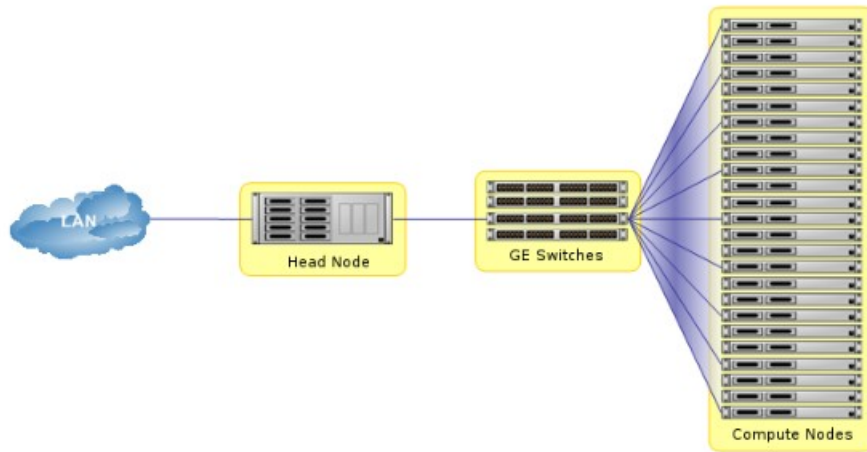


Figura 4.6: Brigh Cluster Manager[®].

Con los conceptos entregados se define entonces el BCM para la plataforma astronómica Chilena.

CHIVO BRIGHT CLUSTER MANAGER En la arquitectura de ChiVO se definió un modelo de clúster estándar de acuerdo a la lista entregada previamente. Esto es, se definieron ocho nodos computacionales que permitirán el desarrollo computacional de alto rendimiento “High Performance Computing”. Un nodo principal que no es más que un sistema que está configurado para actuar como un punto medio entre el clúster real y la red externa (y por ello es el primer agente de los trabajos que se realizan en el clúster). Un switch ethernet que posibilita la conexión “infiniband” privada e interna entre los nodos principal y de cómputo; y por último, unidades de poder para brindar energía a los nodos computacionales y conectar estos a la red privada. A continuación se exhibe en la imagen 4.7 la implementación estándar del Bright Cluster Manager de ChiVO.

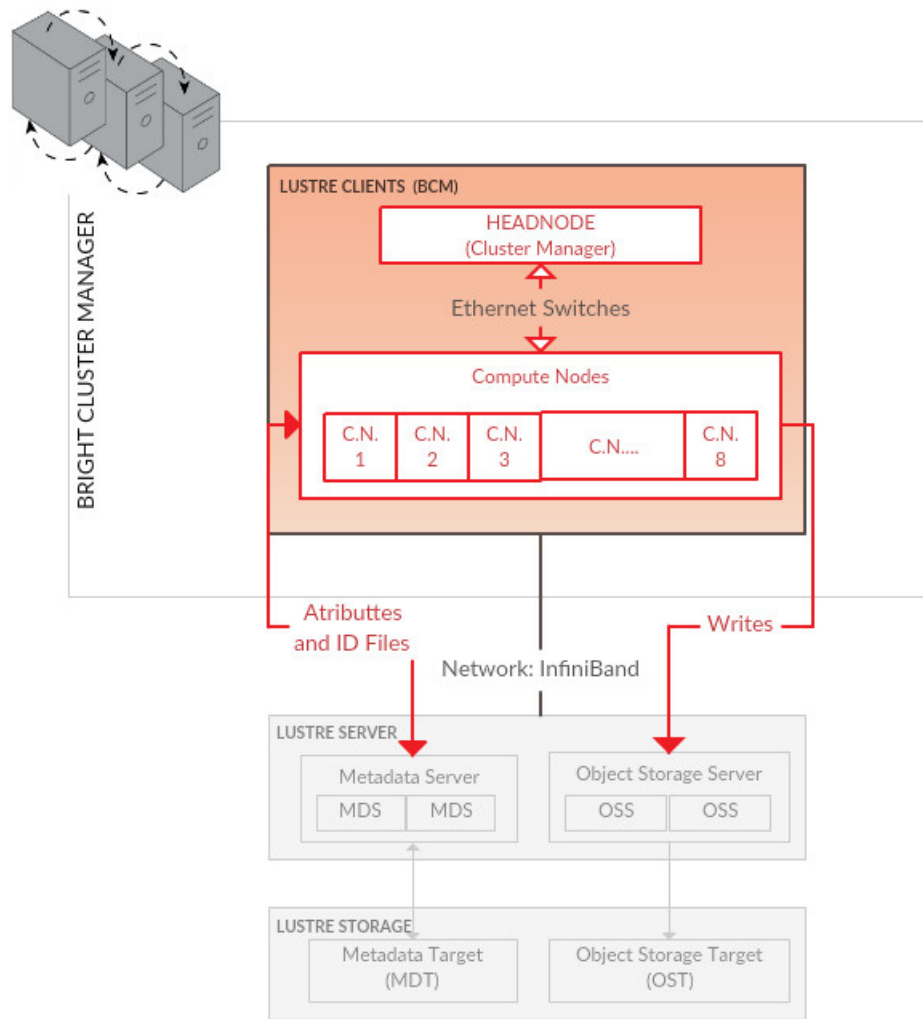


Figura 4.7: CHIVO Bright Cluster Manager

En este punto se han abordado todos los conceptos utilizados para el campo de procesamiento de alto rendimiento de datos de ChiVO. Es preciso ahora describir los componentes de virtualización para la disponibilidad de los servicios del observatorio virtual Chileno.

VIRTUAL SERVER POOL

Un conjunto de máquinas virtuales permite la utilización de una o varias plantillas las cuales pueden utilizarse a petición de cualquier usuario de un grupo determinado. Desde el lado del proveedor, este grupo de máquinas pueden configurarse rápidamente para los usuarios de la nube. Por lo tanto estos acceden a

un grupo de máquinas virtuales la cual es extraída de la lista disponible y tendrá la misma especificación de requerimientos de la plantilla creada. En relación a lo descrito resulta importante señalar dos aspectos técnicos: en primer lugar, que estas máquinas poseen información persistente predeterminada antes en la sesión; y en segundo lugar, que estas se encuentran sustentadas/soportadas por un clúster (conjunto de nodos computacionales).

En términos prácticos, se describe el proceso de **virtualización** para el cual existen distintas alternativas. Una de estas es oVIRT una plataforma de administración para la virtualización de sistemas que consiste la instauración de una estructura que abarca nodos computacionales (“oVirt Nodes”) y una plataforma para el proveedor “Cloud” pueda administrar estos sistemas virtualizados (“oVirt Engine”) sobre los nodos. A continuación se exhibe en la figura 4.8, la estructura de la plataforma de virtualización “oVirt” (extraída de [6]). Posteriormente, se entrega la tabla 4.2 que describe los conceptos presentes en el diagrama.

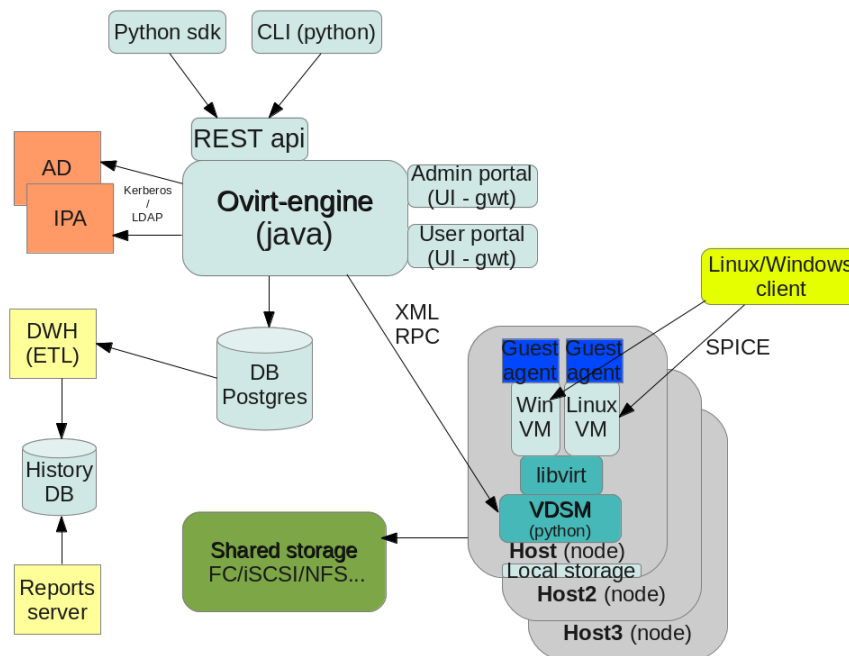


Figura 4.8: Arquitectura estandar de oVirt.

Tabla 4.2: Descripción de elementos arquitectura oVIRT

Engine (ovirt-engine)	Administra los host oVirt. Esto es, permite la creación y despliegue de sistemas virtualizados anfitriones.
Admin Portal	Aplicación web para la administración de tareas relacionadas.
REST API	Una API que permite realizar diversas tareas de virtualización se puede usar desde la línea de comandos o por medio del SDK de python.
CLI/SDK	La interface CLI y SDK proporcionan un método de comunicación hacia el engine por medio de scripts.
Database	La base de datos Postgres es usada por el engine para poder proveer persistencia para cada configuración que ovirt realiza en cada componente.
VDSM (Host Agent)	oVirt Engine se comunica con el host “agente” para la realización de acciones sobre las máquinas virtuales (VMs).
Guest Agent	“Agente” interno de cada VM que provee información sobre los recursos al oVirt Engine.
AD/IPA	Utilizados para determinar mecanismos de permiso hacia los usuarios y grupos.
DWH	Componente data warehouse que realiza tareas de extracción, transformación y carga datos extraídos. Posterior a ello, cada proceso se agrega en el historial de la base de datos postgres.
Reports Engine	Generación de reportes basado en la información histórica de la DB.
SPICE Client	Para el de acceso de los usuarios a las VMs.

CHIVO VIRTUAL SERVER POOL Entendiendo la estructura de una arquitectura de virtualización con oVIRT, se requiere de manera específica dos tipos de nodos: los nodos anfitriones capaces de ejecutar las máquinas virtuales y los nodos de almacenamiento en donde están las imágenes ISO correspondientes al encapsulamiento de los servicios de ChiVO. Estos nodos son distribuciones linux (con VDSM y libvirt preinstalados) que poseen ciertos paquetes para habilitar la virtualización de networking y otros servicios del sistema. ChiVO implementó dos nodos de servicio, cada uno con 126 GB de memoria RAM y 27 cores para el procesamiento efectivo del flujo de datos, los cuales persiguen el fin descrito conectándose mediante la red a los discos de almacenamientos lógico que reciben los

datos de los observatorios astronómicos (“Bulk Data Storage”). A continuación se exhibe la figura 4.9 que manifiesta la implementación de nodos de virtualización (diagrama de alto nivel).

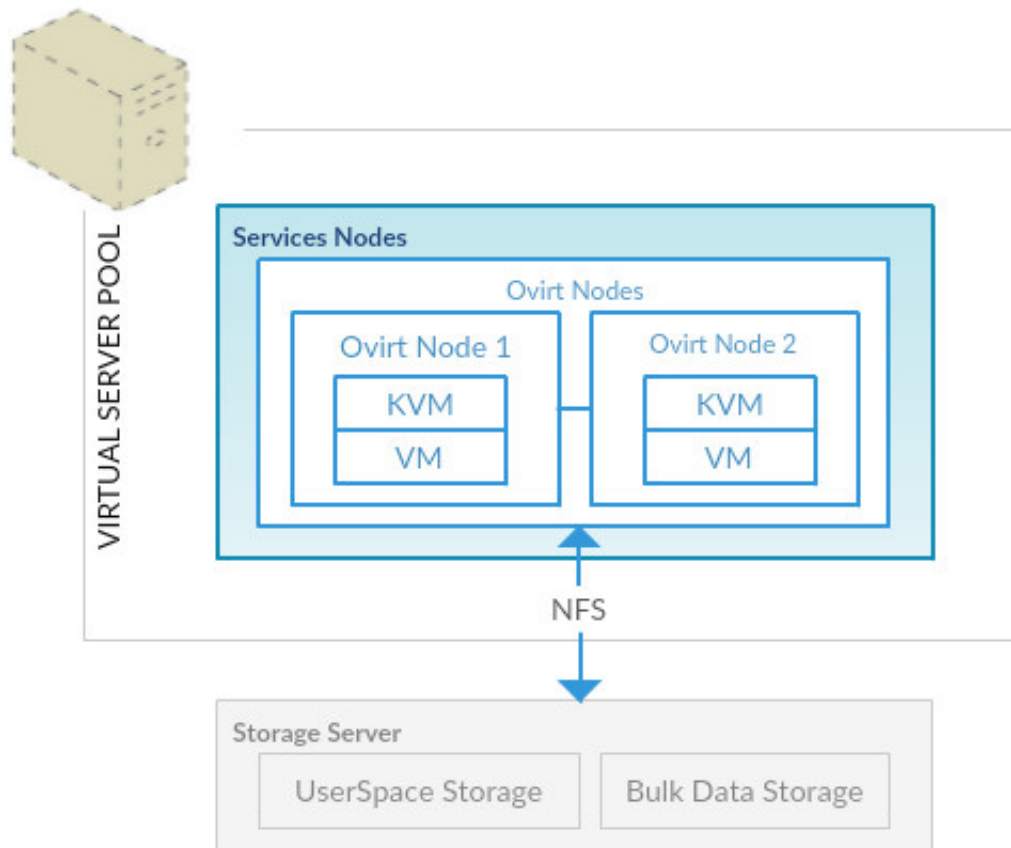


Figura 4.9: Nodos de virtualización de ChiVO

CHIVO VIRTUALIZATION MANAGEMENT PLATFORM A través de la plataforma oVirt Engine se puede administrar los nodos de servicios en términos de creación, modificación, eliminación de los sistemas virtualizados. De manera específica, todo esto a través de una aplicación web y/o una API que permite realizar las distintas tareas competentes al proceso virtualización. Se exhibe en la figura 4.10 un bosquejo de lo manifestado:

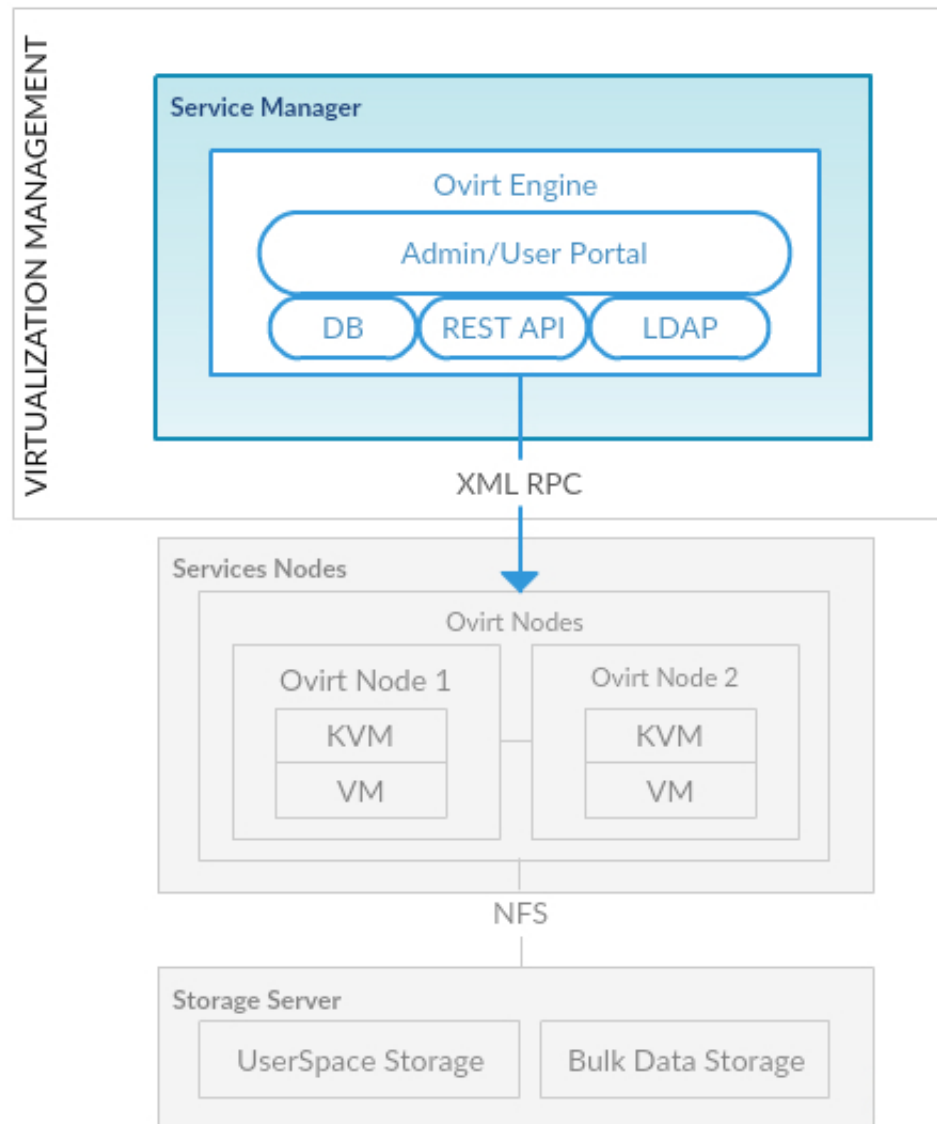


Figura 4.10: Plataforma de administración de virtualización de ChiVO

4.1.2. Plataforma como servicio ChiVO

El modelo PaaS expuesto en la figura 4.11 expone el conjunto de herramientas desplegadas y preconfiguradas que permitirán a los usuarios de la plataforma ChiVO trabajar con el conjunto de datos que se obtienen desde los observatorios astronómicos de forma on-line. En otras palabras, se proporciona un entorno de desarrollo para que la comunidad científica pueda no sólo visualizar dichos datos sino también realizar simulaciones, cruzamiento de información y en general

realizar análisis de datos a gran escala mediante la incorporación de las nuevas herramientas de la plataforma que proveen el procesamiento de datos. Cada proceso determinado acá se sustenta bajo los recursos TI proporcionados por el entorno IaaS de ChiVO, lo cual motivó el desarrollo de nuevas interfaces utilizando computación de alto rendimiento (HPC) para favorecer el procesamiento de Big Data. Por último, se manifiesta que los servicios detallados a continuación tendrán una etiqueta que indica si están ya totalmente implementados (Disponible) o están en proceso de desarrollo.

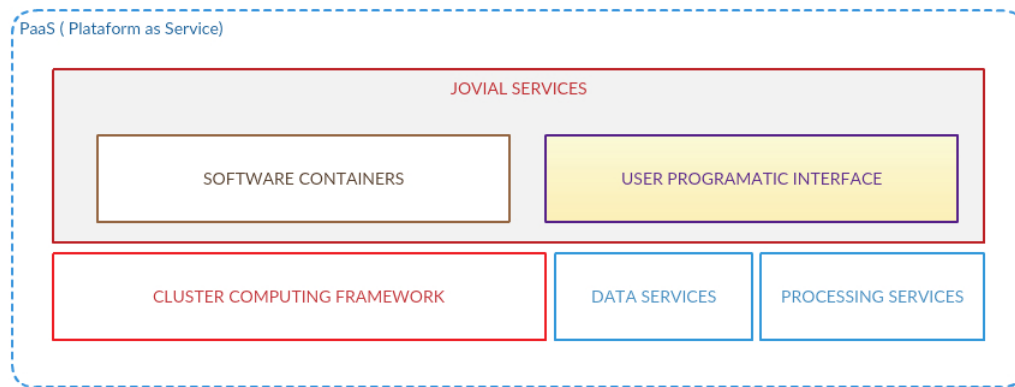


Figura 4.11: Diagrama PaaS (Alto nivel)

Data Services

En este campo se agrupan los servicios de consulta y acceso a las colecciones de metadatos y datos distribuidos de ChiVO. Todo esto, bajo la consideración de los estándares que recomienda IVOA para la interoperabilidad entre plataformas computacionales astronómicas. A continuación se exhibe en la figura 4.12 los servicios que serán abordados en esta sección. Para mayor información sobre cada servicio se recomienda leer el capítulo tres, sección 3.12 “Capa intermedia: DATA ACCESS PROTOCOL STANDARDS”.

SCS (Disponible): Este servicio brinda la posibilidad de realizar consultas por coordenadas en el espacio por medio de la definición de parámetros característicos: RA/DEC y radio angular. Por lo que sigue, esta utilidad representa una de las herramientas típicas de un observatorio virtual en donde se retorna como respuesta a la búsqueda una tabla en formato VOTable, la cual exhibe la lista de fuentes

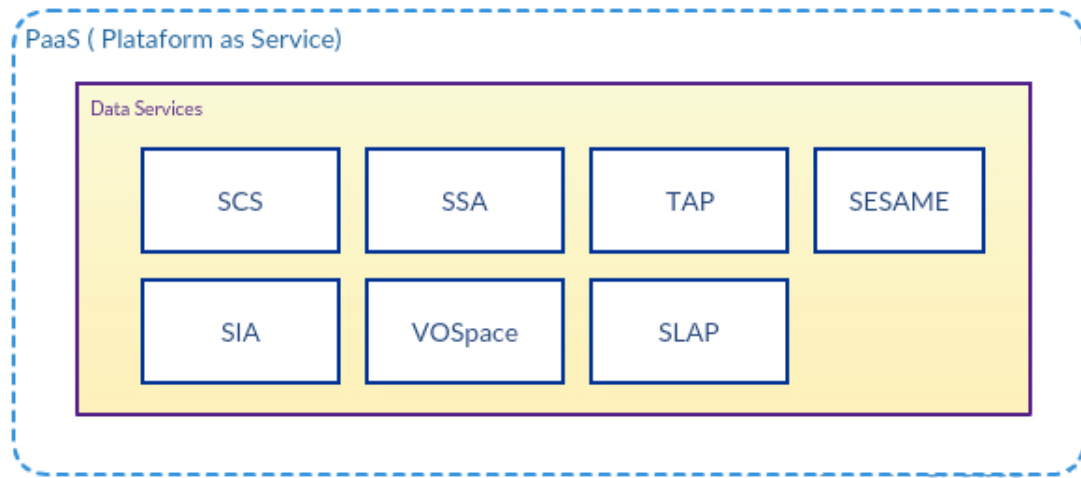


Figura 4.12: Diagrama Data Services ChiVO.

astronómicas de los catálogos asociados a CHIVO, cuyos resultados se encuentran dentro del cono especificado.

SIA (Disponible): Este servicio brinda la posibilidad de realizar consultas al catálogo astronómico provisto por ChiVO mediante la definición de una región rectangular en el cielo. Como resultado el sistema devolverá la lista de imágenes candidatas pertenecientes a la región inscrita, en formato VOTable; cada una con los vínculos de referencia (URL) para su acceso.

SSA (En desarrollo): Este servicio propicia la búsqueda por metadatos espectrales mediante la posibilidad de definir diversos parámetros, tales como: rango de frecuencia o banda, resolución espectral, línea espectral y redshift. Como resultado se retorna una lista con datos candidatos en formato VOTable.

TAP (Disponible): Este servicio expone diversas utilidades tales como, proveer la búsqueda por metadatos espectrales (al igual que SSA), la búsqueda por metadatos temporales, la búsqueda por metadatos espaciales (al igual que SIA y SCS), entre otras. Desde acá, su vasta extensión de funcionalidades se debe a que el servicio brinda la opción de realizar dichas búsquedas por medio de consultas SQL, con la extensión de lenguaje ADQL.

SESAME (Disponible): El servicio permite la búsqueda por nombre o tipo de objeto con datos almacenados en distintos catálogos astronómicos (no necesariamente los almacenados en la plataforma CHIVO) mediante la implementación de un resolovedor de nombres (cadenas de texto) llamado SESAME. Por lo que sigue, el resultado de la búsqueda retorna una lista con los elementos agrupados de acuerdo al criterio de entrada. En relación a SESAME cada consulta realizada a su plataforma es dirigida a tres dispositivos de almacenamiento: Simbad, NED y Vizier.

VOSPACE (En Desarrollo): Este servicio proporcionará un medio para que los agentes participantes de la plataforma (proveedores y usuarios) puedan compartir sus datos y servicios. De manera específica, se entregará un servicio Web que crea un ambiente para compartir elementos VO, en dónde un usuario puede acceder a su espacio personal en un disco lógico especializado para ello (“UserSpace”) y tendrá la posibilidad tanto de agregar o eliminar objetos de datos, obtener, manipular metadatos para dichos objetos, obtener URIs para acceder al contenido almacenado y por supuesto compartir información. Esta herramienta se encuentra en proceso de implementación dado que “UserSpace” no cumple por ahora con los estándares de IVOA.

SLAP (En Desarrollo): Este servicio en desarrollo, exhibirá una interfaz uniforme para el acceso a transiciones de líneas espectrales procedentes de colecciones de datos y metadatos. Desde acá se visualiza en el futuro, formalizar este servicio mediante la implementación del motor de búsqueda distribuido **ElasticSearch** que posibilita respuestas en tiempo real y soporta multi-tenencia. Esto es, se podrá albergar el servicio en micro contenedores, cada uno con los recursos necesarios para ser ejecutados, y en donde una premisa fundamental es que dichos microcontenedores estarán aislados unos de otros y se conectarán a un dispositivo de almacenamiento lógico preferencial para este tipo de trabajos (“Spectral Lines Storage”).

Processing Services

En esta sección se añade uno de los servicios de procesamiento y análisis de datos recomendados por IVOA. En donde, se dictamina como objetivo principal brindar a los investigadores un entorno que administra la ejecución asíncrona de trabajos. La figura 4.13 manifiesta lo descrito.

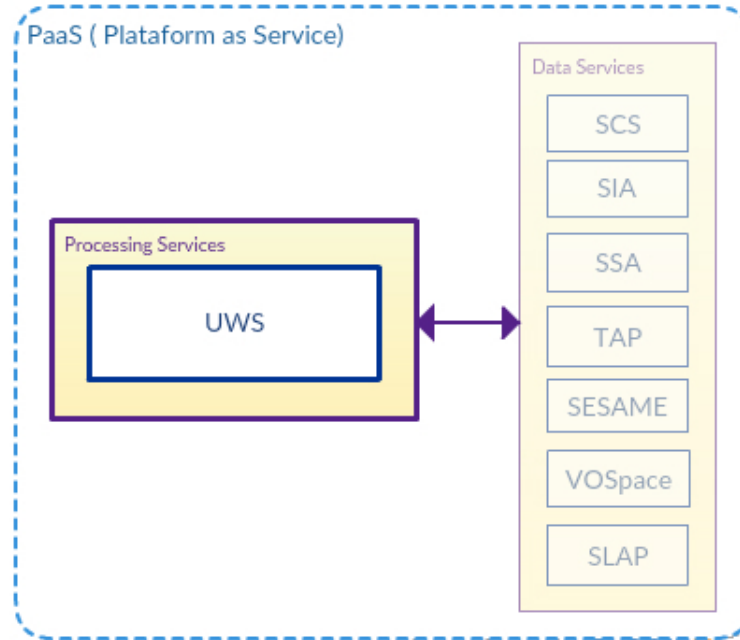


Figura 4.13: Diagrama Processing Services ChiVO.

UWS (En Desarrollo): Este servicio en desarrollo representa un claro ejemplo de un recurso TI perteneciente a PaaS (Platform as Service) dado que promueve la ejecución asíncrona de trabajos de cualquiera de los servicios señalados previamente (Data Services). Esto implica que el usuario no debe esperar que se cumpla cada solicitud, y por ende, si este llegase a desconectarse el flujo de trabajo realizado no se perderá. Por lo que sigue, el patrón UWS resulta efectivo cuando se realizan consultas complejas en donde el tiempo de respuesta aumenta exponencialmente y no se desea perder los trabajos realizados hasta el momento. Por lo tanto, la plataforma VO requiere de esta interfaz asíncrona para actividades de larga duración.

JOVIAL SERVICES

A continuación se profundiza en las posibilidades que brinda el servicio JOVIAL (“Jupyter Override For Astronomical Libraries”) como plataforma web que integra distintas librerías astronómicas para el procesamiento y posterior análisis de datos. Este servicio se encuentra basado en los notebooks de jupyter por lo que es debido, de acuerdo a la figura 4.14, definir y describir qué implica esto.

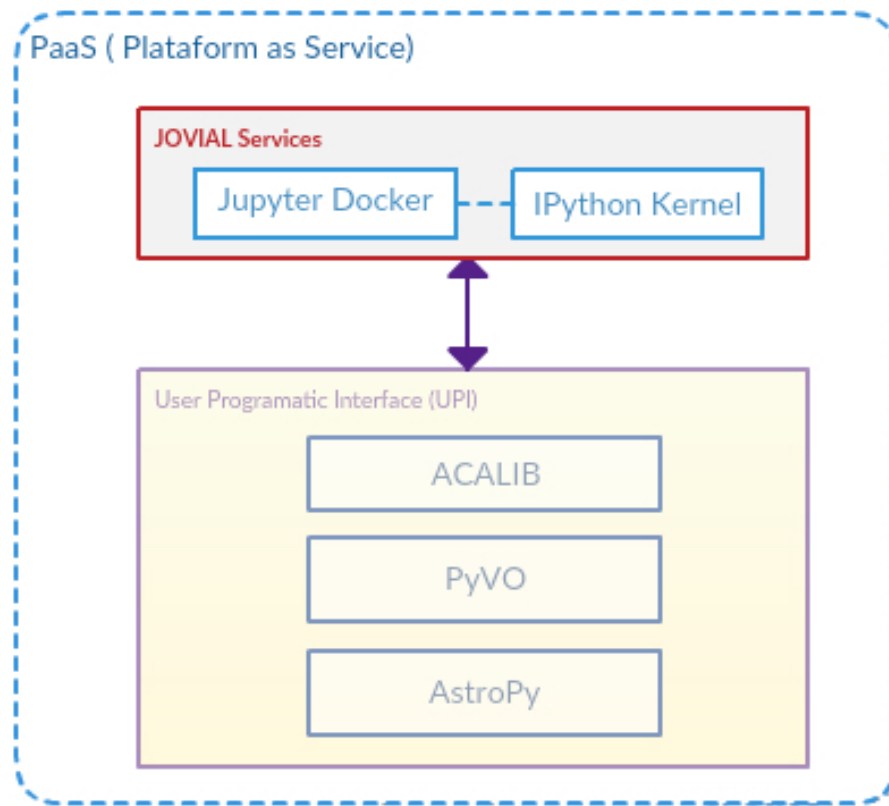


Figura 4.14: Diagrama Jovial Services ChiVO

JUPYTER NOTEBOOK Es un entorno web que extiende el enfoque basado en consola a la informática interactiva y en el cual se puede capturar todo el flujo de procesamiento en términos de documentación, desarrollo y ejecución de código, así como en la comunicación de resultados (los que pueden ser entregado en gráficas que promueven el análisis y la explicación de los datos de trabajo).

IPYTHON KERNEL Para la plataforma astronómica, un usuario puede abrir un Notebook de Jupyter y tener una sesión interactiva conectada a un **kernel de python** (“**IPython Kernel**”), encargado de recepcionar instrucciones de ejecución de los usuarios y comunicar resultados. Lo descrito presenta muchas ventajas para la comunidad científica, y se extiende exponencialmente al situar este servicio en un Data Center, dado que esta arquitectura es capaz de soportar diferentes tipos de paralelismo desde el paso de mensajes hasta la técnica “Task Farming” la cual permite a los usuarios lograr un alto rendimiento de tareas en serie (un solo procesador) en grandes computadoras paralelas. En resumen, este Kernel está diseñado para soportar computación paralela y colaborativa.

CHIVO JUPYTER DOCKER Dentro de los servicios de JOVIAL se provee una imagen de Docker que proporciona un Notebook de Jupyter y una terminal de acceso. En otras palabras, una imagen de docker hace alusión a una plataforma que permite utilizar contenedores los cuales incluyen en su interior distintas aplicaciones de interés para desarrolladores. Dicha imagen entonces posee una terminal de acceso en conjunto con Jupyter Notebook y también como extensión, una instalación completa del entorno Anaconda que proporciona un método eficiente para la administración de paquetes a través de la terminal. Por lo que sigue, mediante la unión de los componentes descritos se puede acceder a una interfaz usuaria de programación interactiva que contiene tres librerías astronómicas de gran utilidad para el análisis de datos. Se recuerda acá, que tanto estos servicios como las librerías que se especificarán ahora forman parte de los servicios de plataforma de ChiVO (PaaS).

USER PROGRAMMATIC INTERFACE (UPI)

Se exhibe a continuación en la figura 4.15, un set de librerías aptas para ser utilizadas con los Notebooks de Jupyter que brinda “JOVIAL”. Desde acá cada una manifiesta un entorno y/o herramientas especializadas para la programación interactiva.

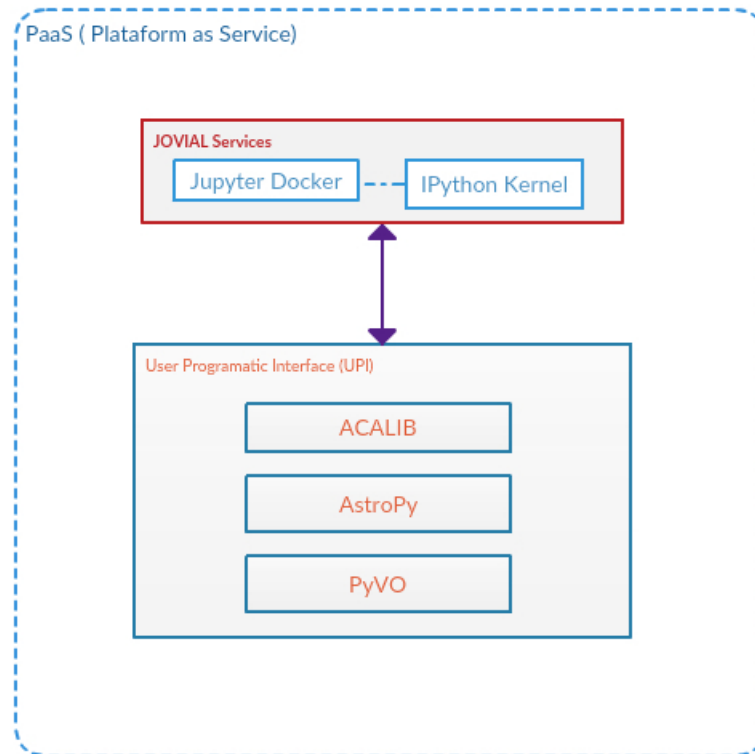


Figura 4.15: Diagrama Jovial Services extensión UPI ChiVO

ACALIB (“Advanced Computing for Astronomy Library ”) (En Desarrollo): Es una librería que dispone una interfaz web simple (compatible con Jupyter Notebook) la cual contiene un ensamblado de algoritmos de última generación para astrónomos y ciertas funcionalidades que propician el análisis de datos. A modo de ejemplo, se tienen algoritmos para la detección de regiones y posterior extracción de su forma; ejecución de un proceso de indexación para un cubo de datos; crear una imagen sobrepuesta por una plantilla y un conjunto de imágenes del mismo objeto; entre otros. Y funciones tales como realizar recortes de una lista de cubos de datos; eliminación de ruido en cubo de datos, etc. En

relación al desarrollo de esta librería se encuentra en un estado beta, pero persigue el principio de generar una interfaz usuaria de programación interactiva. Por último, esta librería propicia la interrelación de todos los servicios de acceso de datos ya descritos, por lo tanto se genera un nexo interesante que mezclará en definitiva los nodos de servicio con los de cómputo (HPC).

ASTROPY (Disponible): Tiene por objetivo incluir en un entorno único los paquetes básicos para el desarrollo efectivo de tareas en el campo de la astronomía. Por ejemplo, manejo de imágenes FITS, tablas, y coordenadas, lectura y escritura de archivos VOTable, cálculo estándar de distancias astronómicas, representaciones, entre otros. Gracias a esto, es posible gestar proyectos que busquen resolver problemas complejos en ámbito científico.

PYVO (En Desarrollo): Tiene como objetivo ser un agente especializado en el acceso a catálogos, datos, servicios de información y herramientas de análisis mediante el empleo de estándares de interoperabilidad VO. A su vez puede tener comunicación directa con AstroPy por lo que propicia mecanismos de migración para otras capacidades VO a esta última.

APACHE SPARK

Es un framework que permite el procesamiento de grandes cantidades de datos mediante la utilización de recursos de un clúster. En profundidad, Apache Spark alberga una aplicación a ejecutar en los nodos computacionales del clúster y determina los recursos a compartir para cada uno de estos, realizando un oportuno seguimiento del servicio. A continuación se exhibe en la figura 4.16 un esquema simple que manifiesta lo descrito.

Es debido señalar la importancia que tiene la incorporación de esta plataforma en un Data center. En primer lugar, el framework posee la característica “Resilient Distributed Datasets (RDD)” la cual propicia la realización de la operación descrita esto es, el trabajo sobre grandes cantidades de datos en clúster de manera rápida y tolerante a fallos. En segundo lugar, posee interesantes librerías que serán de gran utilidad para el proyecto ChiVO. [9]

- Spark SQL que permite el procesamiento de datos estructurados y puede

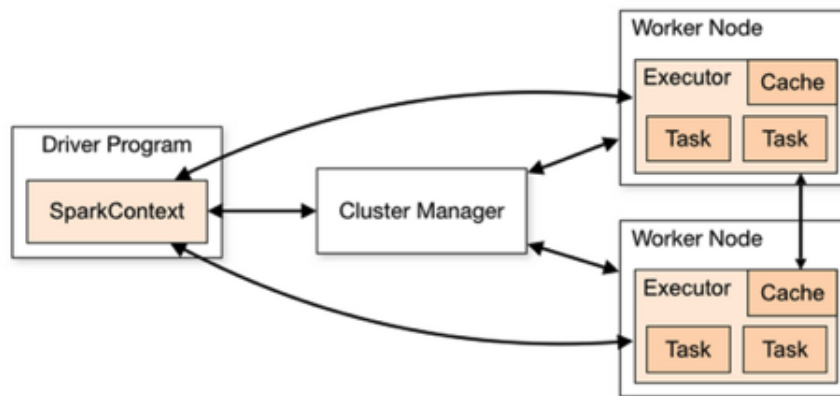


Figura 4.16: Diagrama Apache Spark ChiVO

trabajar con DataFrames (abstracción de programación). En adición, también puede actuar como motor de consultas SQL distribuido.

- Spark Streaming permite realizar un procesamiento de datos en tiempo real, escalable, de alto rendimiento, y con tolerancia a fallos.
- El objetivo de MLlib es hacer práctico, escalable y fácil el machine learning. Consiste en un conjunto de algoritmos y utilidades comunes, como clasificación, regresión, clustering, filtrado colaborativo y reducción de dimensionalidad. GraphX es el componente de Spark para visualizaciones y cálculo gráfico en paralelo.

CHIVO APACHE SPARK (En Desarrollo) El objetivo que se plantea en la instauración del framework en ChiVO es implementar los algoritmos descritos en la librerías UPI, de alto costo en términos computacionales, y ejecutar las tareas correspondientes de manera distribuida y HPC. Junto con fundamentar su desarrollo mediante las extensiones del framework Spark, en especial énfasis, se destaca MLIB que propicia la disciplina “Machine Learning”. Por lo que sigue, es debido manifestar cómo se conectaría esta plataforma con los nodos computacionales subyacentes en la capa IaaS. Ver figura 4.17.

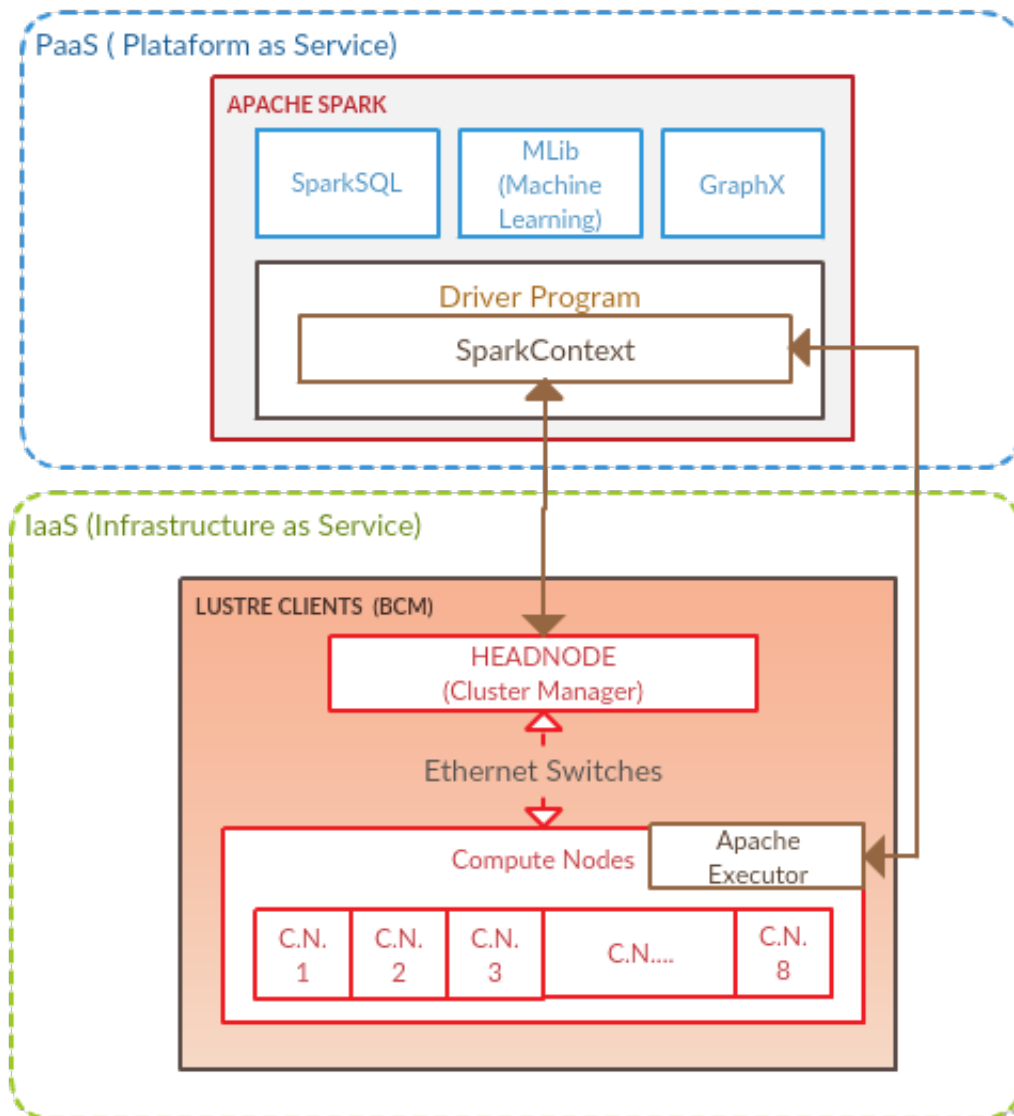


Figura 4.17: Arquitectura estándar de plataforma Apache Spark

4.1.3. Software como servicio ChiVO

El modelo SaaS brinda la posibilidad de posicionar las aplicaciones como un servicio compartido en la nube, esto implica que cada servicio será reutilizable y tendrá una alta disponibilidad para los consumidores “Cloud”. Esta capa representa la cúspide de un proceso que nace desde una infraestructura que permite HPC y una alta disponibilidad de servicios, sustentada por distintas plataformas (PaaS) que brindan los servicios esenciales de este VO (como por ejemplo las interfaces de programación interactiva, o frameworks para la distribución de carga) hasta llegar a estas aplicaciones (SaaS) que aprovisionan los recursos TI para brindar una alta disponibilidad y gestando, en definitiva, un servicio reutilizable para los usuarios “Cloud”. A continuación se señalan los dos software como servicio que tendrá ChiVO.

Notebook-based pipelines based on Jupyter (Disponible): Se predispone un **entorno web** para el uso de Notebooks basados en “Jupyter”. Anteriormente, en la sección ChiVO PaaS, ya fueron definidos los componentes subyacentes que conforman este servicio; a continuación, en SaaS, se considera simplemente la ejecución de estos, de forma interactiva, incluyendo visualización e inserción de código dentro de una plataforma web (Ver figura 4.18 - extraída de [7]).

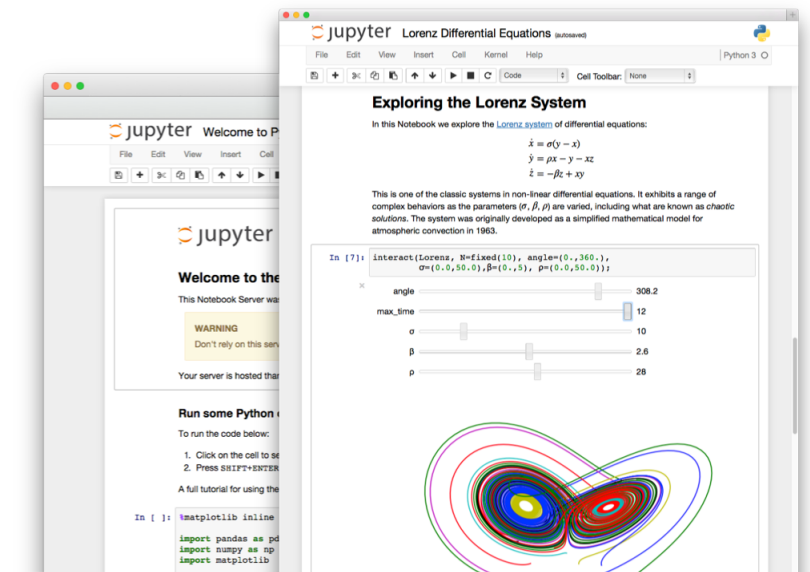
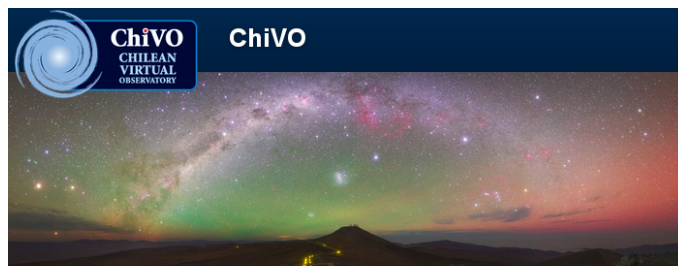


Figura 4.18: Muestra de Jupyter Notebook en SaaS

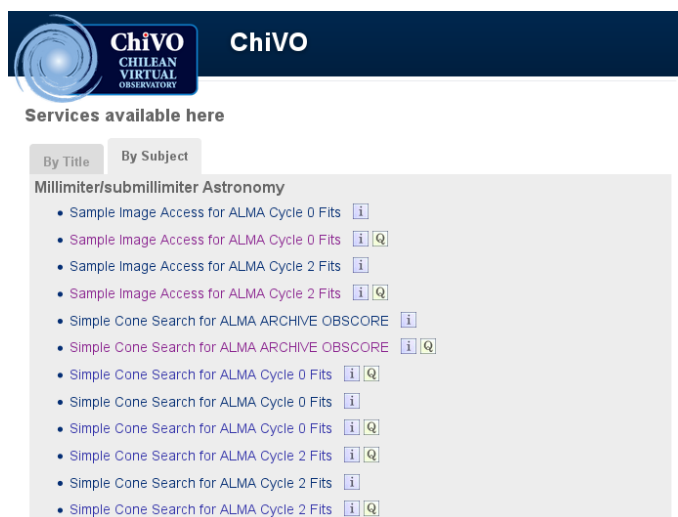
ChiVO (Disponible): Por último se exhibe en la figura 4.19 el portal web de ChiVO (vo.chivo.cl) el cual aprovisiona la lista de servicios disponibles para el trabajo con los dispositivos lógicos “Bulk Data Storage” (situado en la capa IaaS) encargados de almacenar los datos obtenidos por los observatorios astronómicos. Por lo tanto, esta plataforma web es el resultado de la virtualización de sistemas linux los cuales poseen conexión mediante NFS (protocolo que permite acceso remoto a un sistema de archivos a través de la red) con las bases de datos “Bulk”.



The Chilean Virtual Observatory (CHIVO) is a VO developed in Chile and it is one of many VO projects currently underway in the world. It was born out of the need to archive data that require large storage capacities and the need to develop new tools for the analyzing large volumes of data and better algorithms for intelligent processing of astronomical data, this due to the volumes of large scale data that will generate the astronomical observatories in Chile, mainly the ALMA project that will generate over 1 TB of data per day, and in this form be able to store the data in the local territory to optimize the observation time of the Chilean astronomers. Chivo is an initiative of the five most prestigious universities in the country, supported by the private institutions ALMA and REUNA. In addition of the use of the VO by scientists of astronomy area, the VO will be useful too for the students of schools, especially of schools of rural areas.

In addition to the services listed below, on this site you probably can access [numerous tables](#) using [TAP](#) or [form-based ADQL](#).

(a) Portal ChiVO



(b) Servicios disponibles

Figura 4.19: vo.chivo.cl

4.2. Arquitectura de servicios para ChiVO

Finalmente se exhibe en la figura 4.20, el esquema general de la arquitectura ChiVO. Tras esto, se describen ciertos recursos y consideraciones técnicas imposibles de manifestar mediante la abstracción de cada capa.

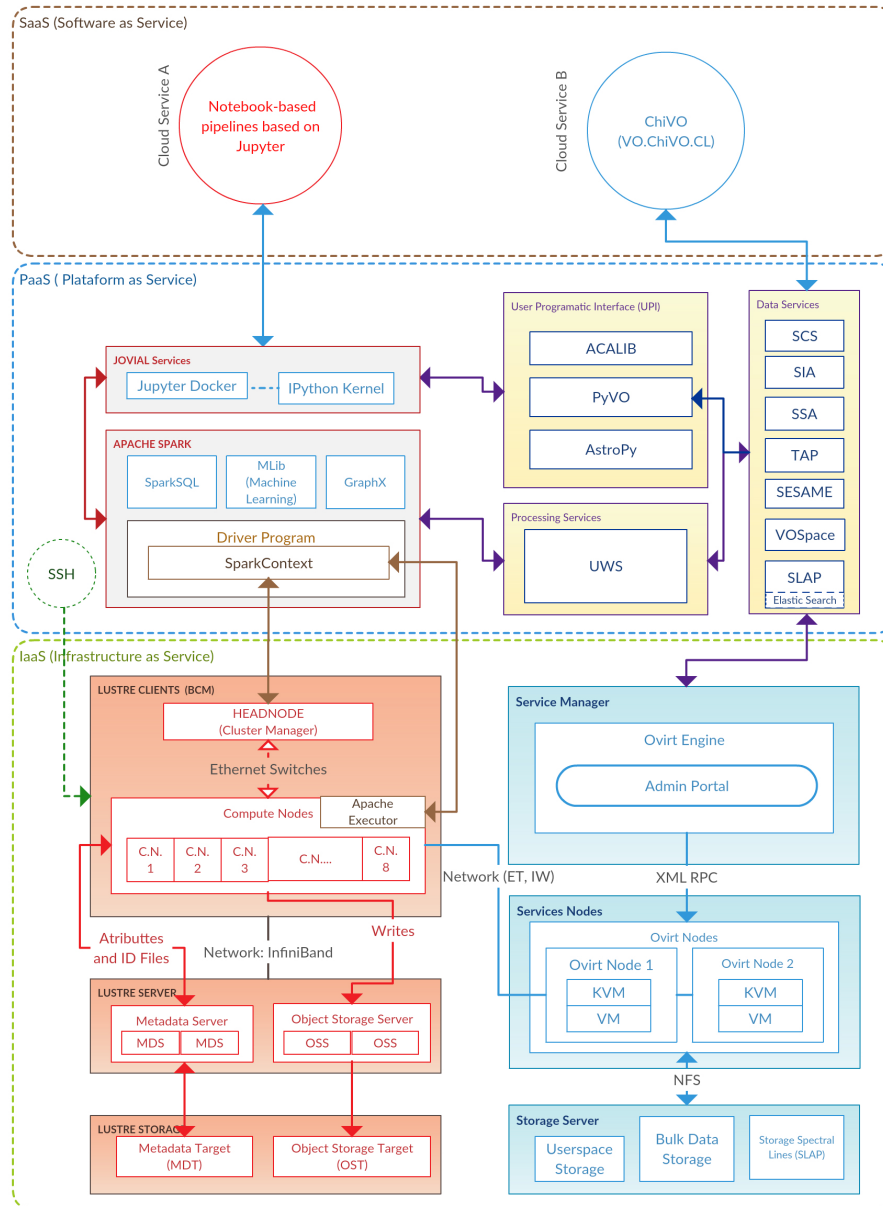


Figura 4.20: Arquitectura de servicios de ChiVO

4.2.1. Consideraciones técnicas generales

Es importante vislumbrar en la figura 4.20 los componentes, interacciones y detalles técnicos no contemplados en las secciones predecesoras. Los cuales se describen a continuación:

1. **SSH:** En la plataforma ChiVO existe la posibilidad de acceder via SSH a los nodos computacionales que provee la plataforma Chilena. Esto de manera paralela a las plataformas web implementadas y con el mismo fin de acceder a los servicios HPC y datos del VO.
2. **Apache Spark:** Dentro de la arquitectura de este framework existe SparkContext encargado de conectarse al clúster mediante el nodo principal y posteriormente a los nodos de cómputo. Por lo que sigue, Spark Executor es el módulo encargado de recepcionar y ejecutar las tareas emitidas por SparkContext.
3. **SLAP:** Se crea una base de datos dedicada al trabajo con líneas espectrales.
4. **Interacción de servicios:** Es factible visualizar la interacción de servicios mediante ciertas librerías y estándares de interoperabilidad que dispone IVOA. Por ejemplo, la librería PyVO permite acceder al conjunto “DataServices” para generar un “puente” en la obtención de los datos; UWS proporciona la ejecución asíncrona de trabajos mediante los datos presentes en “Bulk Data Storage”. Dichos datos pueden ser procesados en los nodos especializados a la computación de alto rendimiento (HPC).
5. **Interacción de nodos:** Nodos de servicio y de cómputo interactúan por medio de la red (InfiniBand y ethernet) con el fin de sustentar procesamiento y acceso a los datos.

Con esto se da término al proceso de definición de una arquitectura de servicios de ChiVO y se prosigue con vislumbrar la forma que adquirirá ChiVO en la “nube”. Es decir, se determinarán modelos de despliegues para la plataforma astronómica Chilena.

4.3. Modelo de implantación para ChiVO

En esta sección se presentan modelos funcionales de despliegue para los servicios Cloud que provee la plataforma ChiVO. De manera práctica, se busca exhibir qué forma adoptará la nube en términos de acceso (quién accede a los servicios y quien va gestionar la infraestructura), seguridad (propiedad) y diversidad. En el marco teórico expuesto en el capítulo dos se detalló la existencia de distintos tipos de nubes (pública, privada, comunitaria e híbrida); por lo tanto es necesario determinar cuál modelo se asemeja más a lo contemplado para el observatorio virtual.

ChiVO actualmente exhibe en su estructura una nube de tipo híbrida, una combinación de nubes públicas (comunitarias) y privadas que proporcionan a la plataforma chilena obtener el beneficio de ambas. Esto es, incorpora ciertas aplicaciones que manejan información “confidencial” en nubes privadas, y propicia el ambiente para aplicaciones que utilizan servicios de nubes tanto públicas como privadas. De este punto, esta decisión se fundamenta al analizar variables tales como:

1. **Interoperabilidad:** ChiVO busca administrar sus servicios y recursos internos descritos previamente y a la vez aprovechar las herramientas y procesos públicos existentes en otras plataformas astroinformáticas reduciendo así la complejidad y los gastos operativos. Lo descrito, se realizaría por medio de la inclusión de los servicios SESAME y PyVO.
2. **Segregación física:** Los recursos se agrupan físicamente y se provisionan como recursos virtuales.
3. **Control operacional:** Si bien ChiVO administra mayoritariamente los consumos de servicios, los usuarios pueden administrar cierto patrones de los servicios mediante la utilización de los entornos de desarrollo de la plataforma chilena.

De lo descrito se vislumbra en la figura 4.21, un modelo (extraído de [5]) que reúne las características mencionadas (nube híbrida). Este modelo permite visualizar la necesidad de definir los alcances de las nubes privada y pública (comunidad) para la plataforma ChiVO.

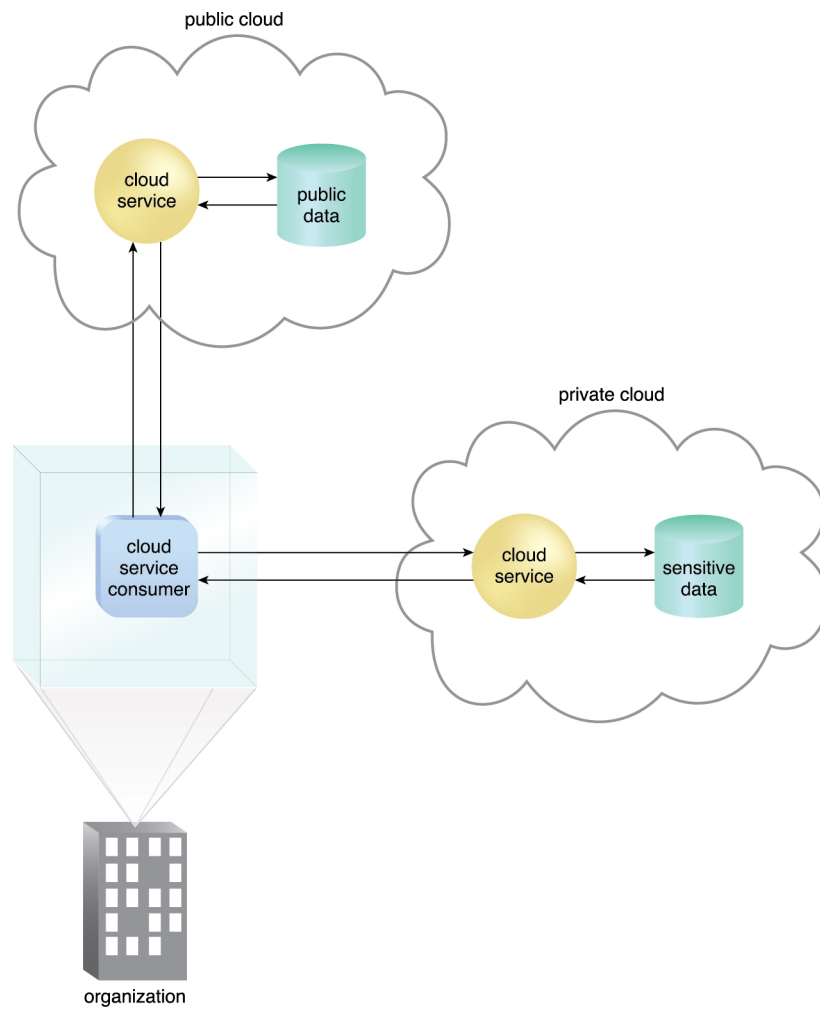


Figura 4.21: Diagrama de una organización usando una nube híbrida que utiliza nubes tanto privadas como públicas [5].

4.3.1. ChiVO PRIVATE CLOUD

En la figura 4.22 se exhibe una nube virtual privada que contiene todos los servicios de procesamiento que ofrece ChiVO (con sus respectivas librerías) y una gran mayoría de los servicios de acceso a datos a través del portal de ChiVO, con excepción de aquellos que proporcionan acceso a otros catálogos astronómicos fuera de la plataforma, pertenecientes a otros observatorios virtuales, aplicaciones y/o servicios con bases de datos externas. Descrito esto, se realiza entonces una representación de la nube privada de ChiVO.

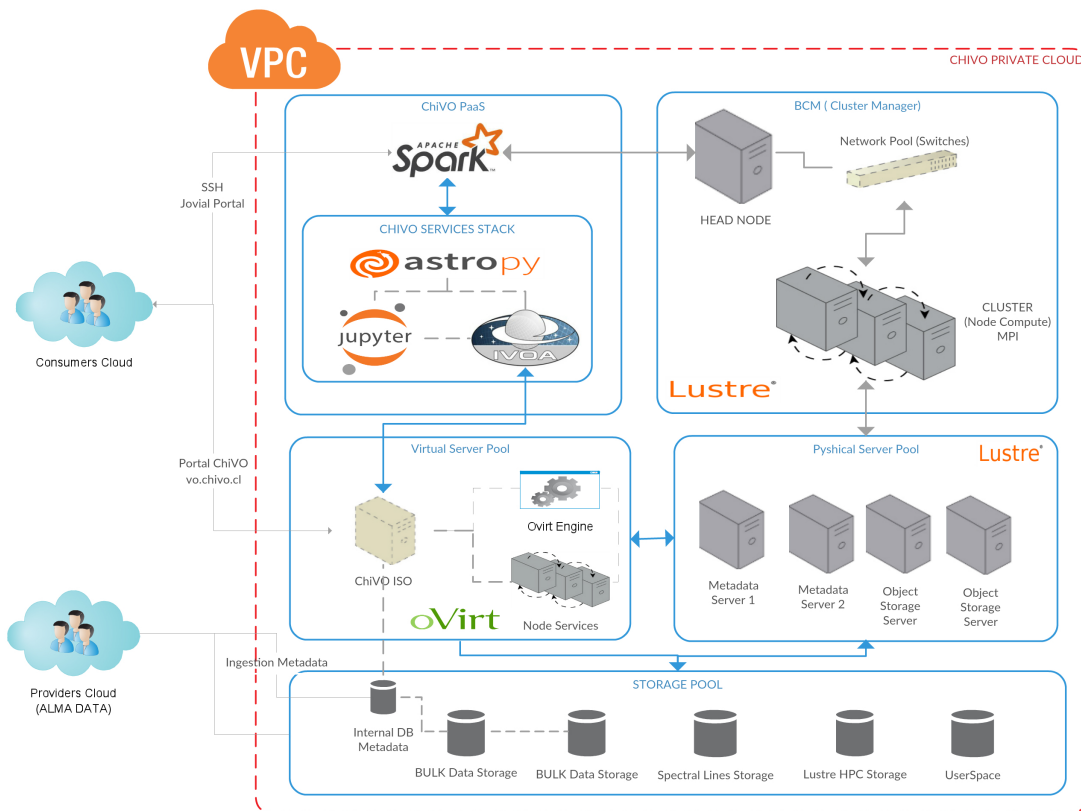


Figura 4.22: ChiVO Virtual Private Cloud

Es importante vislumbrar que el diagrama exhibe la totalidad de los componentes propuestos en los modelos de servicios. La diferencia se percibe solamente al profundizar sobre los servicios de acceso a datos; en detalle, SESAME y PyVO (IVOA). A continuación, se exhibe en la figura 4.23, la nube comunitaria que en conjunto con la privada recién expuesta manifiestan el modelo de despliegue de la plataforma virtual chilena.

4.3.2. ChiVO COMMUNITY CLOUD

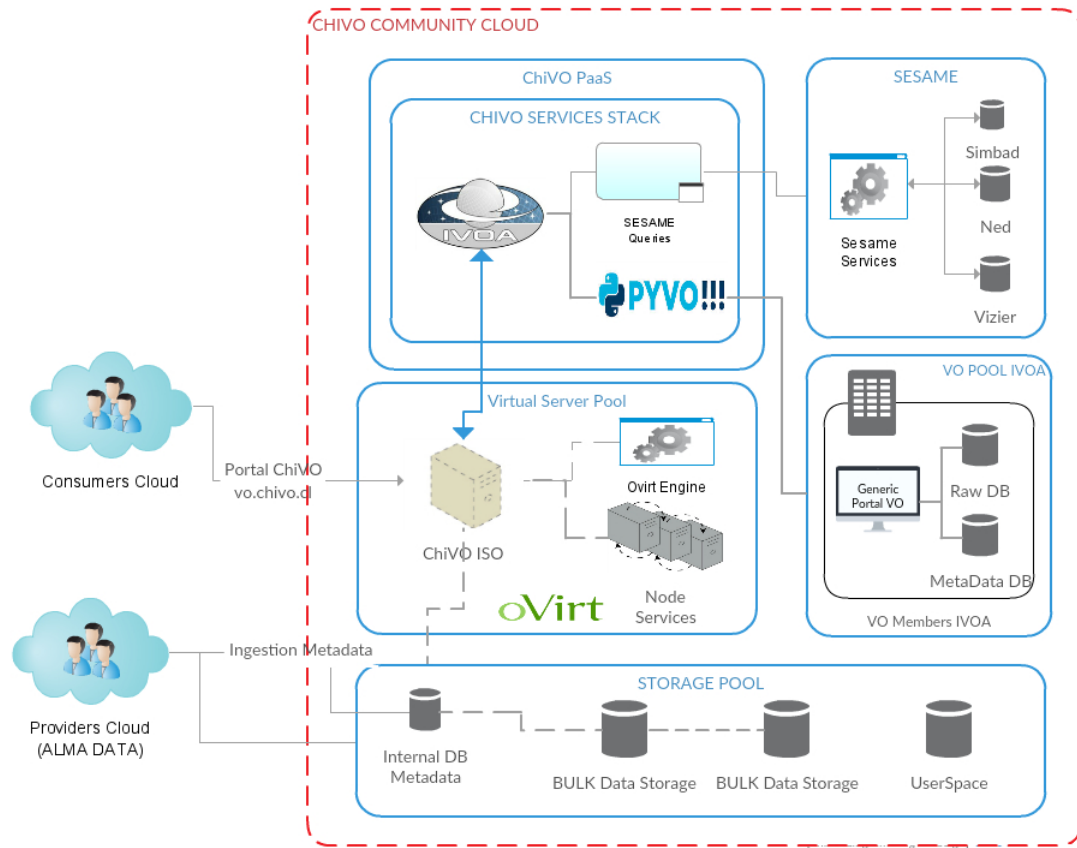


Figura 4.23: ChiVO Virtual Community Cloud

Por lo tanto, se han conformado los diagramas de despliegue que sustentan ChiVO en conformidad con las definiciones expuestas de un modelo Cloud Computing y un observatorio virtual con estándares de interoperabilidad. Considerar para la plataforma ChiVO una nube híbrida permite resaltar ciertas ventajas de interés para la comunidad astronómica: La principal, es la ampliación de la prestación servicios existentes de este VO por medio de la conexión con otras fuentes de datos astronómicos (nube pública - comunitaria); sin duda alguna, esto permitirá que el conjunto de plataformas astroinformáticas puedan sopesar las necesidades de crecimiento rápido y también liberar espacio de almacenamiento. Otra ventaja destacable, es satisfacer los requisitos de seguridad resguardando los datos y trabajos confidenciales de los usuarios de ChiVO (nube privada).



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

Esta página ha sido intencionalmente dejada en blanco.

Conclusiones

Acerca de la arquitectura de ChiVO.

El desarrollo del modelado arquitectónico para ChiVO, permitió detectar cada uno de los componentes que se deben implementar y configurar para el funcionamiento efectivo de la plataforma. Ciertamente, a la hora de considerar la inclusión de los servicios que se encuentran tanto disponibles como en desarrollo, se debió listar previamente las librerías y protocolos esenciales, así como el soporte físico que los sustentará. En definitiva, modelos de servicios y de despliegue que componen una plataforma “Cloud” y la consideración de una arquitectura orientada a servicios, alma de este paradigma, entrega ese mecanismo (soporte) deseado, flexible y abierto para alcanzar escalabilidad e interoperabilidad entre los servicios de este observatorio virtual y las diferentes plataformas externas existentes. Se determina entonces que la incorporación de tecnologías para el manejo de grandes volúmenes de datos permitirá acelerar los procedimientos de inclusión de servicios y junto a ello el procesamiento de los datos con los que se trabaja.

Es importante vislumbrar también que la prestación de servicios en la nube mediante una API (Application Programming Interface) que oculta la distribución de los datos, distribuye tareas en distintos nodos de cómputo y en definitiva, que dictamina el funcionamiento lógico de ChiVO, permitirá facilitar el proceso indagatorio de los astrónomos, mientras se progresa en nuevas soluciones para el problema de la “avalancha de datos”.

Ahora bien, en relación a las herramientas informáticas planteadas fue siempre importante identificar y analizar sistemáticamente los paradigmas tecnológicos en los que se encontraban envueltas. Por lo tanto, patrones de conceptos percibidos

en el proceso de investigación para el desarrollo de este documento (especial énfasis en las definiciones de problemáticas actuales del campo astronómico) tales como: “uso intensivo de datos”, “variedad de tipos de datos” o “problemas con técnicas tradicionales de gestión” acercaban inmediatamente a la noción de “**Big Data**” y los innumerables estudios, soluciones y aplicaciones asociadas, que generaban avances en las distintas disciplinas del quehacer humano; Adicionalmente, la gestación de frameworks para el análisis intensivo de los datos (p. ej. MapReduce), redirigían los esfuerzos por mejorar las tasas de respuesta y actividad y se incluían en “**sistemas distribuidos**”, o en otras palabras, centros de datos con altos niveles de almacenamiento y procesamiento (paralelo) de volúmenes de datos mediante un conjunto de clústeres. Por otra parte, la necesidad de brindar servicios a diversos usuarios sin la necesidad de obligarles a necesitar uno de estos super computadores y donde sólo mediante una capa de presentación especializada se acceda a una plataforma (desconocida) “virtualizada” que posee la oportunidad de realizar “minería de datos” sobre estos grandes volúmenes, todo esto bajo la implantación de una arquitectura “**Cloud Computing**”. Por último, la lista de normas y estándares de la comunidad IVOA, originados en base al estudio cooperativo de científicos de distinta índole e ingenieros simplificaron (en parte) el arduo ejercicio metodológico de fundamentar una arquitectura para ChiVO en cumplimiento con los requerimientos funcionales realizados por el equipo de astrónomos y no funcionales dictaminados por los ingenieros informáticos del proyecto. Es importante en el trabajo futuro, visualizar constantemente los avances en los paradigmas descritos e ir actualizando los servicios en pos de estos, sumarse a organizaciones certificadas, acercar estas implementaciones a otras disciplinas científicas, dado que solo así se podrá abordar cabalmente los retos en el área astroinformática o de cualquier ciencia de datos intensivo.

Acerca de la importancia de ChiVO para el desarrollo de la astronomía en Chile.

La riqueza patrimonial “del cielo” de la segunda región de Antofagasta, específicamente en el desierto de Atacama, y la instalación de los radio observatorios más grandes del mundo en Chile, convierten a esta región en un centro astronómico mundial. Esta zona que posee características excepcionales tales como la escasa humedad atmosférica o un promedio de 330 noches estrelladas al año generan una

condición única que permite realizar observaciones astronómicas privilegiadas [4]. En el caso del observatorio ALMA, enfocado en la astronomía milimétrica, realiza observaciones a galaxias, planetas y estrellas formadas en regiones frías difícilmente detectables en otras longitudes de onda. Dichas observaciones, luego de ser procesadas, resultan ser imágenes de alta definición que poseen una notable reducción en la distorsión producto de las condiciones ambientales descritas. ChiVO como receptor de los datos de ALMA, expande la posibilidades de entender la formación de las primeras galaxias que aparecieron en el universo, de las estrellas y de los planetas.

Por lo tanto, las pretenciones de ChiVO potencian el desarrollo científico a nivel nacional (e internacional) por lo que es debido seguir con la gradual evolución de este en la incorporación de los servicios restantes, en el desarrollo de nuevas herramientas y en la inclusión de otras fuentes de datos procedentes de otros observatorios astronómicos tales como GEMINI, VLT, E-ELT, LSST, entre otros. Sin duda alguna, esto permitirá situar al país como un actor principal en el fomento de descubrimientos globales en el campo de la astronomía, mejorando las falencias actuales del sistema chileno, en el campo de la astroingeniería. Por último, se listan a continuación algunas de estas falencia:

1. Problemas de colaboración entre grupos de astroningeniería y astrónomos.
2. Inexistencia de un ecosistema tecnológico que apoye líneas de investigación en astroingeniería.
3. Carencias en metodologías y certificación para la gestión de proyectos complejos en astroingeniería.
4. Fragmentación entre el desarrollo científico y la articulación de centros de estudio.
5. En comunión con el punto anterior mejoras en este, fomentaría un aumento en la demanda actual de los observatorios.
6. Bajo posicionamiento en decisiones del área a nivel global.
7. Baja promoción de políticas públicas en el campo astronómico.

Bibliografía

- [1] IVOA Introduction to VO Concepts. http://www.ivoa.net/deployers/intro_to_vo_concepts.html. 2017-03-15.
- [2] Bright Computing. *Bright Cluster Manager 6.1 - Administrator manual*, 2016.
- [3] Research Computing University Of Colorado. *Parallel IO on Janus Lustre*, 2016.
- [4] Valeria Medina D. *Centro turístico astronómico en Toconao*. Universidad de Chile, 2009.
- [5] Thomas Erl. *Cloud computing. Concepts, Technology & Architecture*. Prentice Hall, 2013.
- [6] Itamar Heim. *Ovirt Architecture*. 2016.
- [7] Mauricio Araya Mauricio Solar. *Cloud Services on an astronomy data center*. 2016.
- [8] Faviola Molina. La nueva era de datos en astronomía. *Revista Bits De Ciencia* n^o 11, págs. 63–84, 2014.
- [9] Allen Malony Nicholas Chaimov. *Scaling Spark on HPC Systems*. 2016.
- [10] Marcella Marconi Maurizio Paolillo Enrichetta Iodice Nicola R. Napolitano, Giuseppe Longo. *The Universe of Digital Sky Surveys: A Meeting to Honour the 70th Birthday of Massimo Capaccioli*. Springer, 2016.
- [11] Timothy Grance Peter Mell. The NIST definition of Cloud Computing. págs. 6–7, 2011.

-
- [12] IVOA Recommendations. *Observation Data Model Core Components and its Implementation in the Table Access Protocol*, 2017.
 - [13] IVOA Working Groups. *IVOA Architecture*, 2010.
 - [14] IVOA Working Groups. *Registry Interfaces*, 2017.



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

Esta página ha sido intencionalmente dejada en blanco.