



1	OBJETO.....	2
2	DESCRIPCIÓN DE LOS SERVICIOS	2
2.1	DESCRIPCIÓN COMÚN DE LOS SERVICIOS INCLUIDOS	2
2.2	SERVICIOS	2
2.3	TAREAS COMUNES A LOS SERVICIOS	4
3	ENTORNO TECNOLÓGICO	5
3.1	Plataforma física.....	5
3.1.1	Plataforma física de la capa de servidor.....	5
3.1.2	Plataforma física de la capa de cliente.....	5
3.2	Plataforma lógica	5
3.2.1	Características de aplicaciones analíticas	8
3.2.2	Arquitectura de servicios analíticos.....	11
3.2.3	Descripción de infraestructura de Sistemas, BD y Cargas.....	15
3.3	Entorno tecnológico Proyecto NIDEL	27
3.4	Entorno tecnológico Proyecto RIFA.....	27
4	EQUIPO DE TRABAJO	27
5	EJECUCIÓN DEL SERVICIO.....	29
6	TRANSFERENCIA DEL CONOCIMIENTO	30
7	DIRECCIÓN Y CONTROL. REUNIONES DE SEGUIMIENTO.....	31
8	COMPOSICIÓN DEL EQUIPO DE TRABAJO	31
9	GESTIÓN DEL EQUIPO DE TRABAJO REQUERIDO POR LA AEAT.	32



1 OBJETO

El objeto de la licitación es la contratación de servicios de gestión y sistemas, mantenimiento y desarrollo de aplicaciones para el Departamento de Informática Tributaria en los ámbitos del Sistema de Análisis de la Información, Big Data, Minería de Datos y Tecnologías del Lenguaje, responsabilidad de la Subdirección de Tecnologías de Análisis de la Información e Investigación del Fraude (en adelante, TAIIF)

2 DESCRIPCIÓN DE LOS SERVICIOS

2.1 DESCRIPCIÓN COMÚN DE LOS SERVICIOS INCLUIDOS

Los servicios objeto del contrato son los correspondientes a gestión y sistemas, desarrollo, mantenimiento (evolutivo, correctivo, preventivo), migración, adaptación y optimización de los sistemas de información relacionados a continuación, mediante un sistema basado en horas y categorías.

Todos los proyectos se realizarán bajo el compromiso del adjudicatario en el cumplimiento de los aspectos de planificación, gestión, seguimiento y calidad que la AEAT le demande.

2.2 SERVICIOS

Los sistemas y subsistemas de Información que corresponden a este expediente se enumeran a continuación:

Sistema de información Zújar – Cliente

Subsistemas:

- Parametrización de consultas
- Expresiones
- Colectivos

Sistema de información Prometeo/Pandata – Cliente

Subsistema:

- Modelo de seguridad

Sistema de información Teseo-Cliente

Subsistema:

- Colectivos

Sistema de información Zújar - Motor

Subsistemas:

- Operaciones tipo unión
- Modelo de seguridad



- Operaciones diferidas
- Alta disponibilidad

Sistema de información Prometeo/Pandata – Motor

Subsistema:

- Modelo de seguridad

Sistema de información Dédalo – Motor

Subsistema:

- Motor de búsquedas Solr

Sistema de información Genio – Motor

Subsistema:

- Nuevo Visor
- Operaciones diferidas

Sistema de información para servicios de Extracción, Transformación, Envío y Carga de la Información

Subsistemas:

- Zújares de relaciones entre contribuyentes
- Zújar de Recaudación
- Zújares de Declaraciones informativas y modelos tributarios
- Zújares de Inmuebles
- Zújares Explotación del Sistema Z/OS
- Zújares de Recursos Humanos
- Zújares del SETE
- Migración de Zújares a entornos abiertos
- Zújares para Comunidades Autónomas

Infraestructura de sistemas para servicios de Sistemas de Análisis.

Subsistemas:

- Sistema operativo
- Entorno de virtualización
- Herramientas de infraestructura
- Herramientas de apoyo al desarrollo
- Alta disponibilidad
- Servidores web y de aplicaciones,
- BigData
- Sistemas de análisis avanzado SAS

Infraestructura de Bases de Datos para servicios de Sistemas de Análisis.

Subsistemas:

- BBDD analíticas SAP IQ
- BBDD de infraestructura MySQL para entorno analítico
- BBDD distribuidas IMPALA

Infraestructura para servicios de Extracción, Transformación, Envío y Carga de la Información

Subsistemas:

- Sistemas ETL IBM DataStage y desarrollo
- Librerías ETL transversales sobre IBM DataStage

Sistema de información NIDEL-Big Data

Subsistemas:

- Optimización de descargas de datos
- Elaboración de riesgos NIDEL para Recaudación

Sistema de información RIFA (Recuperación de Información en Fuentes Abiertas)

Subsistema:

- Módulos integrados con aplicaciones externas de visualización de información geográfica



2.3 TAREAS COMUNES A LOS SERVICIOS

TAREAS DE ANÁLISIS:
Análisis de requisitos tanto funcionales como técnicos de las aplicaciones de los sistemas analíticos. Análisis de requisitos tanto funcionales como técnicos de los procesos de extracción, transformación y cargas de información.
TAREAS DE DISEÑO:
Diseño de aplicaciones de los sistemas analíticos. Diseño de procesos batch para los servicios de extracción, transformación y cargas de información.
TAREAS DE IMPLEMENTACIÓN:
Desarrollo de aplicaciones de los sistemas analíticos. Desarrollo de procesos batch para los servicios de extracción, transformación y cargas de información.
Construcción de juegos de pruebas unitarias, de integración, de aplicación, de carga y pruebas automáticas. Planificación de procesos
TAREAS DE IMPLANTACIÓN:
Despliegue e implantación de los sistemas y objetos en los entornos de desarrollo, preproducción y producción.
TAREAS DE MIGRACIÓN E INTEGRACIÓN:
Migración e integración de aplicaciones y sistemas.
TAREAS DE SISTEMAS /BASES DE DATOS
Instalación, despliegue, configuración, operación, monitorización, integración y mantenimiento de sistemas
Consultoría, análisis y diseño de arquitectura de sistemas para proyectos de sistemas nuevos o evolución de los ya existentes.
Desarrollo y mantenimiento de programas y procedimientos encaminados a facilitar el uso de los sistemas por parte de todas las piezas del entorno analítico.
Instalación, configuración, operación, optimización y monitorización de bases de datos
Instalación, despliegue, configuración, monitorización, integración y mantenimiento del sistema de extracción, transformación y carga (ETL)
Análisis, diseño, implementación, pruebas y mantenimiento de sistemas de cruces masivos de datos
Desarrollo de librerías y contenedores para integración en sistema ETL
TAREAS DE MANTENIMIENTO:
Mantenimiento preventivo, evolutivo, correctivo y adaptativo de aplicaciones y sistemas.
Optimización y automatización de sistemas y procedimientos.
Adaptación de módulos, aplicaciones y sistemas a la evolución tecnológica y a nuevas necesidades, así como a la normativa prevista por la A.E.A.T. para garantizar los servicios comprometidos.
TAREAS DE DOCUMENTACIÓN:
Elaboración de documentación de usuario, técnica y de procedimientos.
TAREAS DE GESTIÓN DE INCIDENCIAS:
Soporte y resolución de incidencias.
TAREAS DE USABILIDAD:
Mejoras de usabilidad.



3 ENTORNO TECNOLÓGICO

3.1 Plataforma física

3.1.1 Plataforma física de la capa de servidor

La plataforma central de proceso y almacenamiento actualmente consta de los siguientes elementos:

- Cabinas de discos HITACHI modelo VSP-G800
- Cabinas de discos 3PARdata modelo 3PAR_8440
- Cabinas de discos HITACHI modelo VSP-G1500
- Blades HP Proliant BL460c Gen9
- Blades HP Proliant BL460c Gen10
- Servidores HP Proliant DL380 Gen10
- Servidores HP Proliant DL580 Gen9

3.1.2 Plataforma física de la capa de cliente

Puestos de trabajo (PCs con Windows 10) de los usuarios de la Agencia

3.2 Plataforma lógica

- Sistemas operativos: SuSe Linux Enterprise Server 11,12 y 15 y Z-OS
- Entorno de virtualización: Xen Server
- SAP SyBase IQ 15.4 y 16.0.1 y MySQL 5.7 y 8
- Apache HTTP Server y Apache Tomcat
- Servicios de infraestructura de sistemas: OpenLDAP, Openfire, Bind, FTP, SMT, Pacemaker
- Servicios de infraestructura para desarrollo: redmine, eventum, NX, OBS.
- Servicios de monitorización: Zabbix
- Servicios de infraestructura de indexado: solr/lucene
- Servicios de MPP: zookeeper, Parallel Python, Cloudera 5.14 y 6.2
- Lenguajes: C, C++, Java, Perl, Python, Bash Shell Script; .NET(C#); Natural, Cobol y JCL



- Bases de datos: SAP SyBase IQ, MySQL, DB2 y ADABAS
- Sistema de ficheros VSAM.
- ERWIN
- Infraestructura de desarrollo: Visual Studio 2010, .NET
- Herramientas de desarrollo: gcc, gdb, Autotools, Eclipse
- Sistemas de control de versiones: CVS, TFS, Subversion, SourceSafe y Team Foundation Server 2010
- APIS y Frameworks: Developer Express 9.3, yFiles
- Tecnologías de virtualización
- ETL IBM Infosphere (DataStage) 11.7

En la plataforma descrita se ejecutan, por un lado, las necesarias instancias del almacén de datos SAP SyBase IQ, y, por otro, las distintas aplicaciones desarrolladas en la Subdirección General de Tecnologías de Análisis de la Información e Investigación del Fraude, que, a su vez, hacen uso de los servicios de SAP SyBase IQ.

Entre éstas cabe distinguir entre las que se corresponden con servicios comunes u horizontales (e.g: Motor Zújar) y las diferentes aplicaciones verticales, que hacen uso de las anteriores, y proporcionan servicios de análisis de la información y de minería de datos.

En esta infraestructura también se ejecutan trabajos de ETL correspondientes a toda la organización.

La plataforma lógica, como se ha indicado, se basa en la base de datos orientada a columnas SAP SyBase IQ en la que se manejan más de 16.000 tablas, algunas con más de 18.000 columnas, 9.000 millones de registros y 1,8 TB de tamaño.

La ocupación actual del repositorio es de 160 TB.

A título informativo puede indicarse que los sistemas de análisis de la información de la AEAT manejan datos de 10 ejercicios, con más de 200 conceptos, 30.000 dimensiones, 50 millones de contribuyentes y 300 millones de relaciones, más de 5.000 millones de registros y 2 billones de celdas.

Por otra parte el sistema se recarga periódicamente de dos maneras diferentes:



- Las cargas online son 7x24, actualmente hay un volumen de carga de 570 operaciones diarias, algunas de más de 50 millones de registros. Estas operaciones no solo suponen carga de los datos enviados, sino que realizan contrastes posteriores entre la información enviada y la existente en el sistema.
- Las cargas batch de almacenes de datos se realizan entre las 18:30 y las 6:00 de la mañana y en los fines de semana en 7x24, la media mensual de este tipo de cargas es de 29000, que agregadas dan una media mensual de 200000 millones de registros modificados y/o añadidos, en cuanto al número de columnas modificadas la media mensual es de 2'5 millones. Por último en cuanto a tamaño de datos la media es de 254 TB mensuales.

En cuanto a las consultas, el sistema tiene una volumetría con las siguientes características:

1. Número medio diario de consultas en horario online : 576.000
2. Número máximo de consultas diarias en horario online: 1.452.000
3. Número medio de tiempo de respuesta por consulta (por tiempo de respuesta se entiende desde que el cliente solicita la petición en su PC hasta que obtiene la respuesta): 2,3 sg
4. Número medio de informes diarios generados: 3.640
5. Número máximo de informes generados diariamente: 30.910
6. Tiempo medio diario de generación de informes: 15 sg
7. Número medio de grafos de Teseo en horario online : 127
8. Número medio de consultas por grafo: 2000
9. Número medio de usuarios distintos del sistema al día: 6.154
10. Número máximo de usuarios distintos al día: 8.500

Por **consulta** se entiende cualquier petición al entorno analítico solicitando datos, estas pueden ser multiregistros, desgloses, totales, etc...

Por **informe** se entiende una colección de consultas que se solicitan secuencialmente al motor según la lógica con que se haya definido aquel y cuyo resultado, nuevamente, se maqueta y presenta de manera directa o condicional.

En cuanto al sistema de indexado de información no estructurada (Buscón) la volumetría registrada es la siguiente:



1. Número medio diario de páginas procesadas con OCR: 84.650
2. Número máximo diario de páginas procesadas con OCR: 531.000
3. Número medio de documentos indexados diariamente: 18.477
4. Número máximo de documentos indexados en un día: 123.000
5. Número medio de páginas por documento: 7
6. Número máximo de páginas por documento: 65.000

3.2.1 Características de aplicaciones analíticas

En la AEAT existe un portafolio de aplicaciones analíticas, la mayoría de las cuales se sustentan en una infraestructura lógica y física común.

Todas ellas son desarrollos hechos a medida para la AEAT.

Aunque internamente la mayoría de estas aplicaciones se conocen con el nombre genérico de Zújar (o en ocasiones productos de la factoría Zújar) en realidad tiene nombres diferentes, aunque comparten muchos de los componentes, a continuación, se muestra una lista con todas ellas:

1. Zújar:

Es la aplicación de análisis multidimensional interactiva que permite trabajar contra almacenes de datos. Estos almacenes pueden ser solicitados, incluso creados por cualquier usuario de la organización. Permite realizar cruces de datos y filtrar mediante álgebra booleana por cualquiera de los atributos que están presentes en las entidades contenidas en dichos almacenes de datos.

La aplicación Zújar es de carácter horizontal, por tanto no está orientada a un negocio concreto ni ligada a un contenido concreto. Es utilizada por todas las áreas de la organización (gestión, inspección, recaudación, adunas, rrhh, ...).

2. Prometeo:

Prometeo es una especialización de Zújar para la carga de datos específicos de procedimientos de inspección y gestión. Tiene las mismas características que Zújar pero orientado a los usuarios de estas áreas.



Los almacenes de datos ligados a Prometeo están relacionados con la información de la actividad de los contribuyentes: contabilidad, facturación y movimientos de cuentas corrientes fundamentalmente. La información se asocia a expedientes, típicamente (aunque no necesariamente) de Inspección.

3. TESEO:

Teseo es una aplicación que permite el análisis gráfico de vínculos entre elementos de cualquier modelo de negocio de la organización (por ejemplo contribuyentes), y que posibilita al usuario la capacidad de generar de forma automática e interactiva, los grafos con las relaciones seleccionadas, hasta un nivel de profundidad dado y aplicando filtros sobre los atributos asociados a las entidades relacionadas.

Al igual que Zújar, es una aplicación de tipo horizontal que no está ligada a un negocio ni unos datos concretos. Permite analizar relaciones entre entidades de diversa naturaleza como contribuyentes, cuantas bancarias, operadores, países, ...

4. Genio (Infonor):

Se trata de una aplicación de gestión de informes que permite a los usuarios confeccionar y ejecutar sus propios informes alimentados con consultas creadas con la aplicación Zújar. Estos informes no solo contienen datos formateados y contenido estático, sino que se pueden añadir combinaciones sobre los datos recuperados de Zújar, expresiones condicionales en base al resultado de consultas e incluso código javascript para tratamiento avanzado de la información.

Dado que esta aplicación se nutre de las consultas Zújar, su contenido tampoco está ligado a un área de negocio concreta. Permite generar informes para todo el ámbito de la aeat, desde informes con información de la distribución de los trabajadores por área geográfica o por ámbito funcional, hasta informes sobre la evolución de la recaudación por determinados impuestos.

5. DÉDALO:

DÉDALO permite identificar contribuyentes cuando no se conoce con precisión los datos identificativos (NIF y Nombre), o solo se conocen



parcialmente, junto con otros indicios asociados al contribuyente como, por ejemplo, información del domicilio. Estos indicios pueden ser del propio contribuyente o de personas relacionadas con él.

El contenido, por tanto es información de identificación de contribuyentes junto con datos relacionados con él (domicilios, cuentas bancarias, vehículos, barcos o aeronaves).

6. BUSCÓN:

Es un entorno de indexado y búsqueda de términos sobre información no estructurada (documentos de imagen, binarios y de texto) ligado al gestor documental de la organización, pero abierto a aceptar datos de cualquier otra fuente. Este entorno tiene dos partes bien diferenciadas:

a/ una parte online con un interfaz tipo google en el que los usuarios pueden buscar los términos deseados y recuperar (y seleccionar por) información de metadata asociada, tanto cargada directamente (expedientes, procedimientos, ..) como descubierta en tiempo de proceso (entidades (NIF's, Matrículas, etc..)). La búsqueda tiene funcionalidades avanzadas como highlight o búsqueda basada en patrones.

b/ una parte batch que tiene una arquitectura grid y que permite procesar todos los documentos de entrada para indexar su contenido. Este proceso incluye filtros, proceso de mejora gráfica, OCR, reconocimiento de entidades, procesos de diccionario, etc...; la información se almacena en una infraestructura lucene que se gestiona con una capa basada en solr.

7. INTER:

La aplicación INTER gestiona los requerimientos internacionales de información de IVA y de Impuestos directos. Su gestión se puede separar en dos ámbitos. Por un lado, en el ámbito internacional, permite la gestión de dicho requerimiento por parte del equipo central de información de la AEAT, quien es el responsable de comunicarse con los demás países, enviando y recibiendo las peticiones de información y sus respuestas. Por otro lado, en el ámbito nacional, coordina el traslado de dicha información a los equipos de trabajo.

8. RIFA:



Sistema de Recuperación de Información de Fuentes Abiertas. Se trata de un framework Python que de una forma estandarizada permite desarrollar scripts para obtener información estructurada de sitios concretos en Internet. Todos estos datos descargados se pueden cruzar con información de cualquier otro almacén del sistema analítico.

9. NIDEL:

Es una herramienta que haciendo uso de la infraestructura de Big Data del DIT y de librerías especializadas en el análisis de grafos está orientada a encontrar estructuras y relaciones indirectas en redes de millones de nodos.

10. HERMES:

Elaboración y obtención de los riesgos de los contribuyentes mediante la infraestructura de Zújar / Genio. Los riesgos serán utilizados en los procesos de selección y de tramitación por parte de las aplicaciones gestoras

3.2.2 Arquitectura de servicios analíticos

Las aplicaciones que comparten plataforma de diseño con Zujar, es decir: Zujar/Prometeo/TESEO/DEDALO enumeradas más arriba se basan en una arquitectura de tres capas:

1. Cliente pesado: Implementado en .NET y desplegado en prácticamente todos los ordenadores personales de la AEAT, se encarga del interfaz de presentación de los datos, así como de la gestión de todos los eventos generados por el usuario dirigidos al sistema, básicamente la captura de peticiones/respuestas de datos y metadatos, la exportación de datos, preparación y carga de datos, gestión de consultas almacenadas, peticiones de gestión de metadata de colectivos, enlace con otras aplicaciones/tramitadores, enlaces entre almacenes de datos, etc..... Gestiona un tercer nivel de cache personal para cada instalación.
2. Capa middleware (motor Zújar): se comparte por todas las aplicaciones de diseño Zújar (Zujar/Prometeo/TESEO/DÉDALO), así como por otras como Genio o incluso clientes genéricos de Web Services. Es un componente especializado desarrollado en C++ y en Java, se encarga de gestionar las peticiones/respuestas de todos los clientes hacia los almacenes de datos de manera eficiente, para



ello es capaz de aplicar optimizaciones y transformaciones de dichas peticiones (que los clientes describen en un dialecto llamado zSQL) para convertirlas en sentencias SQL interpretables por el SGBD o en operaciones específicas del propio motor (e.g: creación de colectivos, ...), este diseño permite manejar conceptos que no soporta la BD pero que son muy importantes para el negocio, como por ejemplo los datos vectoriales. Esta pieza se encarga de gestionar el segundo nivel de cache compartido por todos los usuarios. Por último, se encarga de todas las labores de seguridad, tanto a nivel de perfilado, como seguridad por ámbito o por filtros específicos y del registro de los accesos.

3. Capa de gestión de datos: Se implementa sobre el mencionado SGBD orientado a columnas SAP SyBase IQ. Se encarga de recibir peticiones tanto de consulta de datos como de peticiones de carga de los mismos. La arquitectura es de varios nodos que tienen el almacenamiento de datos compartido y los catálogos sincronizados, esta arquitectura se conoce como multiplex y permite resolver consultas y cargas simultáneamente, así como escalar y compartimentar el acceso para gestionar los cambios de volumen de peticiones de manera eficiente. Este sistema recibe peticiones de todas las aplicaciones analíticas (excepto Buscón), todas las mencionadas en el punto 2 desde el Motor Zujar. Aquí se implementa el primer nivel de cache de datos. También existen almacenes de datos basado en Cloudera sobre una infraestructura HW MPP sobre los blades de HP, sobre este entorno y usando servicios adicionales (impala, Sqoop, Hue, etc...) se realizarán labores analíticas sobre objetos grandes usando la mismas capas middleware y cliente.

Además de las capas mencionadas existen una serie de piezas transversales a las aplicaciones de la suite Zujar que son imprescindibles para su correcto funcionamiento:

- A. Diccionario de Metadata: gestiona toda la metainformación de los almacenes de datos, tanto la de carga, como la de consulta y la de cliente. Esta implementado en MySQL.
- B. SGBD de infraestructura: siete instancias de MYSQL que sirven para almacenar toda la información accesoria del sistema: consultas almacenadas, datapedia, indexado de metainformación



- de almacenes, configuraciones de clientes, registro de accesos, etc....
- C. Servicio de árboles: gestiona la metadata de cliente de los almacenes de datos (carpetas, campos,...) en XML para que se puedan construir las peticiones. Implementado en PERL.
 - D. Herramienta de gestión de metadatos: se encarga de permitir a algunos usuarios especializados la gestión de la metadata de los almacenes de datos mencionada en A. Implementado en .NET
 - E. Servicios de menú: el acceso a todas las herramientas zújar se realiza desde el navegador web, hay una primera capa que gestiona estos accesos (incluso solicitando información necesaria a los usuarios) y si todo está correcto permite el arranque parametrizado de los clientes pesados (incluso la instalación desasistida de dichos clientes en caso de que el PC no disponga de ellos). Implementado en JAVA.
 - F. Sistema de Cargas automáticas: Sistema que permite gestionar las peticiones de cargas de almacenes de datos, de manera eficiente, consumiendo la metadata de carga, alimentando la metadata de consulta y posibilitando la sincronización de la metadata de cliente. En las cargas de datos se realizan determinadas transformaciones encaminadas a la optimización de las consultas (creación de tiles, tramos y desnormalización de literales), implementada en C y PERL.
 - G. Sistema de Cargas Prometeo: capa que rodea el sistema anterior y que permite una carga especializada del sistema Prometeo que como se ha comentado más arriba permite cargas online desde los clientes. Estas cargas tienen preprocesos y postprocesos específicos. Implementada en Shell Script, PERL y C.
 - H. Sistema de ETL Infosphere information Server (DataStage): Sistema ETL fabricado por IBM, sobre él se realizan parte de los procesos de extracción de datos del sistema operacional, transformaciones del negocio encaminadas a construir los almacenes de datos.
 - I. Procesos ETL implementados en JCL: un gran número de los procesos de ETL actuales que realizan las mismas labores enumeradas en el punto H, se realizan en el z/OS usando JCL's.

La aplicación Genio tiene cuatro capas:



- A. Cliente, bien el diseñador (implementado en Java) que permite confeccionar y gestionar informes, bien el navegador con un interfaz simple que permite invocar su ejecución.
- B. Genio servidor, es la parte servidora que se encarga de coordinar la ejecución de los informes contra el motor Zujar, así como maquetar las respuestas y gestionar los aspectos lógicos del informe (condiciones, combinaciones, etc...) (se puede considerar también parte del propio motor con lo que la arquitectura sería de tres capas).
- C. Motor Zujar, traduce y optimiza las consultas ZSQL de los informes.
- D. SGBD resuelve las peticiones SQL de datos

La aplicación Buscón tiene una arquitectura un poco más compleja, no la parte de consulta que es simple:

- A. Navegador como cliente de acceso, que presenta la interfaz de búsqueda, el manejo de los resultados y el salto al gestor documental.
- B. Aplicación Buscón servidor, hecha en Java, permite traducir las peticiones de búsqueda en llamadas a solr.
- C. Gestor de índices, es el propio sistema solr (hecho en java) que permite las búsquedas de datos sobre los índices lucene.

sino la parte batch que consta de las siguientes capas:

- A. Capa de recepción de órdenes, implementa web services para recibir peticiones de indexado, modificación de metadata o directamente peticiones adhoc de OCR.
- B. Demonio de órdenes, valida el XML de las ordenes, descarga los documentos y encola los trabajos. Implementado en Python.
- C. Demonio de procesado de órdenes, categoriza las ordenes, decide los pasos por los que tienen que transitar para ser procesadas (transformaciones, extracción de metadata, separación de páginas, OCR, extracción de entidades, diccionario, mejoras gráficas, etc...), gestiona el paralelismo de todo el proceso, gestiona las colas de proceso, gestiona la



ordenación para garantizar la consistencia de la información, gestiona la distribución de órdenes a los nodos de proceso, así como la finalización del trabajo. La parte adhoc está implementada en Python y Parallel Python, y se usan multitud de herramientas de SW libre: aspell, tesseract, tika, imagemagick,

- D. Demonio de indexado, se encarga de gestionar la petición de escritura de la información en el sistema de indexado solr. Implementado en Python.
- E. Infraestructura de índices, es el mencionado almacén de índices lucene gestionado con librerías solr.

En Buscón además de las capas mencionadas, existen herramientas accesorias necesarias para el funcionamiento del sistema:

- SGBD MySQL para almacenar estado de las ordenes y registro de su proceso
- Web de Soporte de Buscón, para poder ver el estado del sistema de procesado de órdenes. Hecho en Java.

3.2.3 Descripción de infraestructura de Sistemas, BD y Cargas

Para que todos los servicios anteriores funcionen correctamente sobre la infraestructura física y lógica en los apartados 2.1 y 2.2. existe una capa de infraestructura de Sistemas, Bases de Datos y Cargas, que se detalla a continuación.

3.2.3.1 Infraestructura de Sistemas

Entornos

Actualmente existen 5 entornos de sistemas:

- A. Entorno de test: Es el ámbito donde se realizan las pruebas de nuevas versiones de productos, o de nuevos productos. También hace las veces de entorno de desarrollo/preproducción del entorno de infraestructura de sistemas.



- B. Entorno de Infraestructura: Es el ámbito donde funcionan de manera productiva todos los servicios necesarios para la administración de sistemas, y los servicios transversales necesarios para que todos los entornos funcionen, e.g: DNS, LDAP, Zabbix, etc...
- C. Entorno de Desarrollo: Es el ámbito donde los equipos de desarrollo construyen nuevas aplicaciones, evolucionan o corrigen las ya existentes. En este entorno también se hacen labores de integración entre servicios.
- D. Entorno de Preproducción: Es el ámbito donde se realizan los pases de las aplicaciones/servicios exactamente a como se harán en producción. En este entorno es donde el grupo de verificación hace las pruebas de validación y de regresión sobre los nuevos productos o las nuevas versiones de los existentes.
- E. Entorno de Producción: Es el entorno donde se realiza la producción de los servicios analíticos.

A continuación, se detalla cada uno de los entornos, indicando la infraestructura física y lógica de cada uno de ellos.

La mayoría de los servicios/plataformas enumeradas están dotadas de alta disponibilidad (en adelante HA) que garantizan que tras la caída de uno de los nodos el otro se haga cargo de ofrecer el servicio. Esta HA normalmente es activa/pasiva y se implementa de dos maneras:

- Con mecanismos propios del producto: e.g infraestructura Apache-Tomcat
- Con mecanismos del sistema operativo: Pacemaker

Todos los entornos marcados como virtualizados con Xen Server se ejecutan en un pool de servidores formado por:

- 40 servidores x86 HP Blade BL460c G9 con 512 o 256 Gb memoria y 24 cores cada uno

Entorno de Test

Actualmente consta de la siguiente infraestructura (por servidor se entiende un sistema con Suse Linux Enterprise Edition 12 con los productos necesarios instalados sobre él):



- Dos servidores de balanceo. Virtualizados con Xen Server.
- Dos servidores web apache, clusterizados con Pacemaker
- Dos servidores del sistema de monitorización zabbix, clusterizados con Pacemaker y virtualizados con Xen Server.
- Dos servidores del middleware de Zujar, clusterizados con Pacemaker y virtualizados con Xen Server.
- Un servidor de control de versiones CVS, virtualizado con Xen Server.
- Un servidor de control de versiones SVN, virtualizado con Xen Server.
- Un servidor de gestor de tareas Redmine, virtualizado con Xen Server.
- Un servidor de compilación OBS, virtualizado con Xen Server.
- Dos servidores de mensajería instantánea openfire, clusterizados con Pacemaker y virtualizados con Xen Server.
- Dos servidores DNS, virtualizados con Xen Server.
- Dos servidores OpenLDAP, virtualizados con Xen Server.
- Dos servidores del servicio de árboles, virtualizados con Xen Server.
- Cuatro servidores Tomcat, virtualizados con Xen Server.
- Un servidor de compilación maven, virtualizado con Xen Server.
- Cuatro servidores de proceso batch del entorno de indexación Buscón, virtualizados con Xen Server.
- Dos servidores base de datos MySQL, clusterizados con Pacemaker y virtualizados con Xen Server.
- Dos servidores de correo, clusterizados con Pacemaker y virtualizados con Xen Server.
- Cuatro servidores de índices solr, virtualizados con Xen Server
- Un servidor del gestor de paquetes de Suse SMT y virtualizado con Xen Server
- Tres servidores para SAS virtualizados con Xen Server
- Diez servidores para Cloudera virtualizados con Xen Server
- Seis servidores de Sybase IQ en una máquina física x86 de 12 cores y 96 GB de RAM
- Dos servidores del sistema de intercambio de mensajes Rabbit-MQ virtualizados con Xen Server
- Dos servidores del sistema de gestión de datos compartidos Redis virtualizados con Xen Server



Entorno de Infraestructura

Actualmente consta de la siguiente infraestructura (por servidor se entiende un sistema con Suse Linux Enterprise Edition 12 con los productos necesarios instalados sobre él):

- Dos servidores de balanceo de carga. Virtualizados con Xen Server.
- Dos servidores web apache, clusterizados con Pacemaker
- Dos servidores del sistema de monitorización zabbix, clusterizados con Pacemaker y virtualizados con Xen Server.
- Un servidor de control de versiones CVS, virtualizado con Xen Server.
- Un servidor de control de versiones SVN, virtualizado con Xen Server.
- Un servidor de gestor de tareas Redmine, virtualizado con Xen Server.
- Un servidor de compilación OBS, virtualizado con Xen Server.
- Dos servidores de mensajería instantánea openfire, clusterizados con Pacemaker y virtualizados con Xen Server.
- Dos servidores DNS, virtualizados con Xen Server.
- Dos servidores OpenLDAP, virtualizados con Xen Server.
- Dos servidores de compilación maven, virtualizados con Xen Server.
- Dos servidores de correo, clusterizados con Pacemaker y virtualizados con Xen Server.
- Un servidor del gestor de paquetes de Suse SMT y virtualizado con Xen Server
- Un servidor del gestor de configuración Spacewalk y virtualizados con Xen Server
- Dos servidores base de datos MySQL, clusterizados con Pacemaker y virtualizados con Xen Server.

Entorno de desarrollo

Actualmente consta de la siguiente infraestructura (por servidor se entiende un sistema con Suse Linux Enterprise Edition 12 con los productos necesarios instalados sobre él):

- Dos servidores web apache, clusterizados con Pacemaker
- Dos servidores del middleware de Zujar, clusterizados con Pacemaker y virtualizados con Xen Server.



- Un servidor del servicio de árboles, virtualizado con Xen Server.
- Cuatro servidores Tomcat, virtualizados con Xen Server.
- Dos servidores de proceso batch del entorno de indexación Buscón, virtualizados con Xen Server.
- Cuatro servidores base de datos MySQL, clusterizados con Pacemaker y virtualizados con Xen Server.
- Cuatro servidores de índices solr, virtualizados con Xen Server
- Dos servidores de DataStage sobre servidor físico x86
- Seis servidores de Sybase sobre servidor x86 y en los que reside el sistema de carga de datos analíticos.
- Cuatro servidores para Cloudera virtualizados sobre servidores físicos x86
- Cuatro servidores para SAS virtualizados con Xen Server
- Un servidor para crawler virtualizados con Xen Server
- Dos servidores para los servicios de Buscon, virtualizados con Xen Server.
- Dos servidores del sistema de intercambio de mensajes Rabbit-MQ virtualizados con Xen Server
- Dos servidores del sistema de gestión de datos compartidos Redis virtualizados con Xen Server
- Dos servidores de balanceo de carga virtualizados con Xen Server.

Entorno de preproducción

Actualmente consta de la siguiente infraestructura (por servidor se entiende un sistema con Suse Linux Enterprise Edition 12 con los productos necesarios instalados sobre él):

- Dos servidores web apache, clusterizados con Pacemaker
- Dos servidores del middleware de Zujar, clusterizados con Pacemaker y virtualizados con Xen Server.
- Dos servidores del servicio de árboles, virtualizado con Xen Server.
- Cuatro servidores Tomcat, virtualizados con Xen Server.
- Dos servidores de proceso batch del entorno de indexación Buscón, virtualizados con Xen Server.
- Dos servidores para los servicios de Buscon, virtualizados con Xen Server.
- Cuatro servidores base de datos MySQL, clusterizados con Pacemaker y virtualizados con Xen Server.



- Seis servidores de Sybase IQ sobre servidores x86. En este servidor reside el sistema de carga de datos analíticos.
- Cuatro servidores para SAS virtualizados con Xen Server
- Cuatro servidores de DataStage en servidores físico Blade HP
- Diez servidores para Cloudera virtualizados sobre servidores físicos x86
- Dos servidores del sistema de intercambio de mensajes Rabbit-MQ virtualizados con Xen Server
- Dos servidores del sistema de gestión de datos compartidos Redis virtualizados con Xen Server
- Dos servidores de balanceo de carga virtualizados con Xen Server

Entorno de producción

Actualmente consta de la siguiente infraestructura (por servidor se entiende un sistema con Suse Linux Enterprise Edition 12 con los productos necesarios instalados sobre él):

- Dos servidores web apache, clusterizados con Pacemaker
- Dos servidores del middleware de Zujar, clusterizados con Pacemaker sobre servidores x86.
- Dos servidores del servicio de árboles, virtualizado con Xen Server.
- Cuatro servidores Tomcat, virtualizados con Xen Server.
- 16 servidores de proceso batch del entorno de indexación Buscón, virtualizados con Xen Server.
- Cuatro servidores base de datos MySQL, clusterizados con Pacemaker y virtualizados con Xen Server.
- Cuatro servidores de DataStage, sobre un servidores físicos x86 potenciados.
- 16 servidores de Sybase IQ en sobre un servidores físicos x86 potenciados.
- 28 servidores físicos HP Blade BL460c G9 para Cloudera
- Cuatro servidores físicos x86 potenciados para DataStage
- Cuatro servidores para SAS virtualizados con Xen Server
- Dos servidores físicos x86 para SAS
- Dos servidores del sistema de intercambio de mensajes Rabbit-MQ virtualizados con Xen Server
- Dos servidores del sistema de gestión de datos compartidos Redis virtualizados con Xen Server



- Dos servidores de balanceo de carga virtualizados con Xen Server

Descripción de productos y herramientas

Las herramientas y productos sobre los que se implementan los entornos descritos más arriba se describen a continuación.

- Servidores Web Apache: Se encargan de atender las peticiones de contenido estático, como por ejemplo los binarios del cliente Zujar para que clickonce lo pueda autoinstalar y canalizar las llamadas a los servidores de contenido dinámico. Se usan como canal de entrada para multitud de servicios.
- Servidores de aplicación Tomcat: Se encargan de servir contenido dinámico que generan determinadas aplicaciones. Es el servidor de aplicaciones corporativo de la parte de análisis y sobre el se ejecutan muchas de las aplicaciones transversales desarrolladas internamente como por ejemplo los servicios de menús, frontales del Motor Zujar, servicios online y de recepción de ordenes de Buscón, o InfoCargas, otras que son productos de software libre como solr o CAS.
- Servidores de Middleware Motor Zujar: Son los servidores donde se ejecutan los procesos binarios del middleware de todo el sistema Zujar. Estos procesos no se ejecutan al amparo de ningún servidor de aplicaciones, son procesos que funcionan directamente sobre el sistema operativo
- Servidores zookeeper: Son los servidores encargados de soportar el mencionado producto de Apache que sirve para poder gestionar configuraciones de entornos cloud (en nuestro caso el solrcloud).
- Servidores solr: Son los servidores donde se almacenan los índices de términos de Buscón y las piezas necesarias para poder gestionarlos (altas, baja, modificaciones, réplicas, administración, optimizaciones de índices, etc...)
- Servidores de MySQL: Son los SGBD en los que se gestiona la información correspondiente a la metadata de análisis, así como cualquier otro dato necesario para el correcto funcionamiento de las aplicaciones.
- Servidores Buscón: En estos servidores se ejecutan todos los procesos necesarios para poder procesar órdenes de indexación, de modificación o borrado de metadata de índices y puras de OCR.



Estos procesos se apoyan en multitud de software libre: aspell, tika, tesseract, imagemagick, ..., así como en mencionados desarrollos con parallel python.

- Servidores de balanceo de carga: Son los equipos desde los que se repartirán las peticiones a los servidores web que vayan dirigidas a las aplicaciones de análisis.
- Servidores de ETL: Son los servidores donde se diseñan, prueban y ejecutan todos los procesos de carga de datos implementados con el producto de ETL IBM Infosphere Information Server (DataStage).
- Servidores del servicio de árboles: Este servicio, como se ha comentado más arriba, se encarga de generar la metadata de los almacenes de análisis que es consumida por los clientes. Este servicio es un desarrollo propio que está basado en PERL y que se ejecuta con módulos específicos de Apache, está separado del resto por facilitar la gestión y minimizar las interferencias entre servicios.
- Servidores SyBaseIQ y cargas: Son los equipos que más recursos consumen de toda la instalación y están instalados en la tecnología más potente de la que dispone la parte de análisis de la organización. En parte de estos servidores también se ejecutan los procesos de carga automática y carga Prometeo que posibilitan la carga de datos de los almacenes analíticos.
- Servidores Zabbix: Zabbix es el producto de SW libre que se usa para monitorizar toda la infraestructura de análisis, desde la parte más baja (hardware vía SNMP) hasta la parte más alta de servicios de análisis.
- Servidores Redmine: Se usan para gestionar tareas de los proyectos de desarrollo del grupo de análisis, para gestionar los pases, para documentar toda la instalación de manera colaborativa, de acceso al repositorio de versiones SVN, etc...
- Servidores SVN y CVS: Son los gestores de repositorios de software, no solo de los proyectos de análisis, sino en el caso de CVS de todos los desarrollos de la infraestructura de la AEAT.
- Servidores de compilación: En ellos se implementa el entorno OBS que es un producto de SW libre que permite gestionar la compilación de los productos para varias plataformas (SLES sobre Power, SLES sobre x86, etc...)
- Servidores de gestión de paquetes Suse: Sirven para hacer de proxy de descarga de paquetes de SLES hacia todos los servidores,



así como para gestionar las suscripciones de los servidores en el soporte de Suse.

- Servidores de LDAP: En estos se implementan la infraestructura de autenticación y perfilado para los usuarios que acceden tanto a los sistemas de análisis y a las herramientas internas del grupo de informática de análisis como a los entornos de desarrollo. Los usuarios finales corporativos (todo el personal de la AEAT) de los servicios de análisis usan la autenticación corporativa (RACF) para validarse y se perfilan un mecanismo propietario de la organización, ninguno de estos dos productos residen en la infraestructura de análisis pero hay enlace a ellos para que su uso sea transparente.
- Servidores de correo: Solo se usan para canalizar los correos de aviso del sistema de monitorización, del gestor de tareas y algunos otros al sistema corporativo de correo.
- Servidores de mensajería instantánea: Sirven para lo mismo que los anteriores pero en este caso a través del canal de mensajería instantánea (jabber).
- Servidores de nombres: Se encargan de gestionar los nombres de los servidores del entorno analítico en aquellos dominios internos que tenemos delegados.
- Servidores maven: Son necesarios para realizar compilaciones que tienen dependencias externas a un servidor concreto.
- Servidores de SAS: Se encargan de dar soporte a determinadas labores de análisis avanzado
- Servidores de crawler: Soportan las labores de selección de datos externos para alimentar los servicios de análisis
- Servidores de Cloudera: Dan soporte al entorno MPP sobre el que se hace análisis sobre grandes volúmenes de información.
- Servidores de mensajería Rabbit-MQ, se usan para comunicar procesos que residen en máquinas diferentes, por ejemplo para señalar procesos distribuidos que deben coordinarse.
- Servidores de almacenamiento de datos Redis, se usan para almacenar datos no persistentes con alto rendimiento y consultables de manera distribuida, por ejemplo, en cachés distribuidas.

3.2.3.2 Infraestructura de Base de Datos



En esta área existen dos partes bien diferenciadas:

- Gestores de datos orientados a columnas (SAP Sybase IQ) que se usan básicamente para gestionar los datos correspondientes al entorno de análisis.
- Gestores de datos operacionales (Oracle MySQL), se usan para multitud de fines, entre otros:
 - Gestionar los metadatos del entorno analítico
 - Gestionar los datos de la herramienta de gestión de almacen Dolibarr
 - Gestionar los datos de informes de Genio
 - Gestionar los logs de los clientes de los entornos de análisis, todos ellos guardan remotamente la información de registro en este entorno central.
 - Gestionar los datos de registro y de estado del sistema de cargas automáticas y Prometeo
 - Gestionar los datos de zabbix, redmine y openfire
 - Gestionar todos los datos de infraestructura que necesita para poder funcionar la parte batch de Buscón, como por ejemplo el estado de las órdenes. También se usa como registro de todas las operaciones, incluidas las de consulta.
 - Gestionar toda los datos de infraestructura que necesita la factoría Zujar: Consultas almacenadas, configuraciones, datos de búsqueda en metadata (buscazujar), metadatos colaborativos del sistema de análisis (datapedia), etc...

En el caso de SAP SyBase IQ, como ya se ha comentado se dispone de un multiplex de varios nodos en cada uno de los entornos de desarrollo, preproducción y producción. Este multiplex permite garantizar tanto la alta disponibilidad, como la distribución de la carga (de consultas y de carga de datos).

IQ se usa como SGBD de alto rendimiento orientado a columnas, lo que posibilita que las consultas típicas de entornos analíticos, como son los agrupados, sean muy rápidas. Otra característica muy importante de este gestor es que dispone de multitud de índices, algunos muy especializados, otros más generales, que posibilitan un rendimiento muy bueno en diferentes perfiles de uso, esto es fundamental dado que el entorno analítico de la AEAT tiene un corte de Inteligencia de Negocio Operacional, es decir



los usuarios tienen libertad para solicitar cualquier consulta hacia su almacén de datos.

Prueba de que el SGBD se ajusta muy bien a las necesidades de la AEAT, es que el modelo de datos no ha parado de crecer y se ha pasado de un modelo puro de almacenes totalmente desnormalizados en una sola tabla, a modelos en estrella y el futuro es permitir la combinación de datos entre almacenes de manera relativamente libre para el usuario, lo que permitirá el ideal de que cualquier usuario pueda cruzar datos para cualquier objeto tributario, sin necesidad de personal informático que lo asista.

IQ es la pieza fundamental de todo el sistema de análisis y su rendimiento tiene que ser óptimo en todo momento, porque cualquier degradación en el mismo impacta inmediatamente en los servicios de análisis.

Sobre IQ existen multitud de desarrollos de procesos para mejorar la administración personalizada para el entorno de la AEAT, como por ejemplo los procesos de backup y recuperación, monitorización, exportación, chequeos, etc...

En cuanto al entorno MySQL, aunque con menor músculo que IQ, no es menos importante dado que sobre él se implementa la gestión de la metadata del entorno de análisis, así como datos de casi todas las piezas tanto del entorno de análisis como de infraestructura de sistemas. Si MySQL deja de funcionar automáticamente todas las aplicaciones del entorno paran o se degradan.

Actualmente se está reorganizando las instancias y se les está dotando de alta disponibilidad usando PaceMaker y replicación propia de MySQL.

Sobre estos gestores también existen tanto procesos realizados a medida para intercambiar datos entre instancias y con otros repositorios (incluido el propio IQ), como herramientas propias de Oracle para gestionar el MySQL, e.g MYSQL Monitor.

3.2.3.3 Infraestructura de Cargas

Por último, se detalla el sistema de carga de datos, que no debe ser confundido con el entorno sobre el que se realizan las labores del grupo de Extracción, Transformación y Carga y que se verán más adelante, en este último entorno se realizan procesos de transformación de datos de dos tipos: en el entorno Z/OS usando JCL's como lenguaje de programación, orientados al negocio y en entornos SuSe Linux con la herramienta ETL IBM Infosphere Information Server (DataStage).



Sin embargo, en el sistema de Cargas (que para distinguirlo le llamamos Cargas Automáticas y Prometeo) lo que se realizan son:

- transformaciones de optimización en el caso de las cargas automáticas para datos cuyo origen es la salida de los procesos JCL mencionados antes y cuyo origen son los datos operacionales residentes en el DB2 o en Oracle.
- y transformaciones de negocio y de optimización (estas cargas usan para este fin las cargas automáticas) en el caso de las Cargas Prometeo para datos que los usuarios de inspección suben directamente desde sus clientes, así como datos de control de facturas (no SII, o Lineas PNI) del departamento de Gestión o de la DCGC.

El sistema de cargas automáticas es una máquina de estados que extiende las labores del SGBD IQ hasta cubrir la consistencia con la metadata del sistema analítico, es decir, permite cargar datos en IQ de manera muy eficiente y garantizando que la metadata residente en MySQL queda consistente con lo cargado, esta pieza es la que garantiza que todos los datos cargados serán usables por todas las piezas.

El sistema de Cargas Prometeo es una capa que recubre a la carga automática y que permite cargas especializadas de datos de Inspección, esta especialización se muestra en funciones como la alimentación de metadata específica, realización de contrastes, creación de almacenes al vuelo, etc...

Todos estos sistemas son desarrollos específicos de la AEAT, y disponen de una serie de herramientas relacionadas, también desarrollos propios, que permiten gestionar dichas cargas, e.g InfoCargas.

El producto IBM Infosphere Information Server (DataStage) se usa para implementar sin necesidad de desarrollar procesos eficientes de extracción, transformación y carga. El objetivo es hacer reingeniería de todos los procesos existentes de ETL, tanto los JCL's, como procesos adhoc de movimiento de datos, e incluso partes de las cargas mencionadas anteriormente, hacia el nuevo sistema.

La gestión de toda la infraestructura del producto en todos sus entornos, así como la consultoría, optimización de procesos e incluso desarrollo de trabajos, será responsabilidad del grupo de análisis.



3.3 Entorno tecnológico Proyecto NIDEL

El entorno tecnológico es el siguiente:

- Cluster Hadoop de Cloudera
- Librerías de Spark: SparkSQL y Spark GraphFrame
- Lenguaje de programación: Scala, Java y script de Linux.

3.4 Entorno tecnológico Proyecto RIFA

Este framework cuenta con los siguientes elementos:

- a. Sistema operativo Linux (Suse)
- b. Bases de datos MySQL / Oracle
- c. Lenguaje de programación Python 2.7
- d. Desarrollo de scripts haciendo uso de la librería Scrapy 0.24 y Scrapy 1.0. Scrapy es una librería open source desarrollada en Python que permite acceder a posiciones concretas dentro de una página HTML haciendo uso de rutas XPath. También gestiona el flujo de peticiones, reconexiones, etc.
- e. Aplicación Selenium (aplicación de libre distribución para automatización de pruebas)

4 EQUIPO DE TRABAJO

Dado el carácter esencial del elemento personal para el desarrollo y mantenimiento de los sistemas y aplicaciones que se llevan a cabo en el ámbito de la AEAT y la importancia de un conocimiento profundo de los entornos tecnológicos similares al especificado en este Pliego de Prescripciones Técnicas para la correcta ejecución del contrato, los licitadores deberán ofertar los perfiles técnicos con los requisitos, certificaciones, capacidad técnica y los meses de experiencia exigida para cada uno de los perfiles en el Anexo II, para el desempeño de las actividades descritas en este Pliego y deberán acreditarlo de manera obligatoria de acuerdo con lo exigido en el PCAP.

En determinados perfiles se exige experiencia específica en entornos similares al descrito en el apartado 3. Dado que exigir experiencia específica en un entorno objetivamente similar al descrito en dicho apartado, si bien sería lo ideal de cara a la prestación del servicio, pudiera resultar restrictivo en cuanto a la presentación de ofertas, en el Anexo II se describe lo que a estos efectos se entiende como instalación similar, tratándose de una instalación cuyas características se han reducido al máximo posible para



facilitar la presentación de ofertas, sin que ello ponga en riesgo la prestación del servicio por falta de cualificación técnica.

Los perfiles requeridos agrupados por categorías son:

Consultor	4
Consultor BigData	1
Analista-1	7
Analista-2	4
Analista Programador-1	6
Técnico de Sistemas-2	3
Analista Programador-2	5
Programador-2	1

En total, 31 perfiles, que corresponden al siguiente detalle:

Categoría genérica	Perfil específico	Recursos
Consultor	C01	1
Consultor	C02	1
Consultor	C03	2
Consultor BigData	CB01	1
Analista-1	A101	4
Analista-1	A102	2
Analista-1	A103	1
Analista-2	A201	2
Analista-2	A202	2
Analista Programador-1	AP101	2
Analista Programador-1	AP102	2
Analista Programador-1	AP103	2
Técnico de Sistemas-2	TS201	3
Analista Programador-2	AP201	2
Analista Programador-2	AP202	1
Analista Programador-2	AP203	2
Programador-2	P201	1

El número de horas por categoría es de **1.700 horas/año**, excepto en el caso del Consultor BigData, en que son 400 horas/año.

Debido a la criticidad de los sistemas objeto del contrato, es necesario que los perfiles de Sistemas (Consultor C03, Analista-1 A103 y Técnico de Sistemas-2 TS201) tengan la posibilidad de realizar guardias si así se les requiere. El importe del presupuesto para atender el pago del "servicio de guardia" que sea solicitado por la AEAT, se ha estimado, atendiendo a la



experiencia de contratos anteriores, en un 0,5% del presupuesto base de licitación. Dicho importe se establece **como un máximo**, de manera que la AEAT no estará obligada al pago del citado servicio, salvo que previamente haya solicitado su prestación, todo ello sin perjuicio de las modificaciones del contrato que, en su caso, puedan tramitarse de conformidad con lo prevenido en los Pliegos y en la LCSP.

5 EJECUCIÓN DEL SERVICIO

Los trabajos serán llevados a cabo en las instalaciones del contratante.

En circunstancias ordinarias, la jornada ordinaria no podrá exceder de las 40 horas semanales por cada recurso asignado (35 en los meses de verano en el caso de que dicha reducción sea adoptada para la organización en su conjunto).

La franja horaria de cobertura obligatoria de los servicios reflejados en el presente pliego, a los efectos de su consideración como horario del centro de trabajo destino al que deberá atenerse la jornada diaria del trabajador (Ex. Art. 20.6 XVII convenio colectivo estatal de empresas de consultoría y estudios de mercado y de la opinión pública), salvo excepciones expresamente autorizadas, será de 9:00h a 14:30h de lunes a viernes y dos tardes de lunes a jueves, de 16:00h a 17:00h.

En los meses de verano, del 16 de junio al 15 de septiembre, la jornada diaria del trabajador se reducirá a 7 horas diarias, con carácter general estas horas se prestarán durante la mañana y sin interrupción para la comida.

El resto del horario se realizará de forma flexible de lunes a viernes en la franja horaria de 7.00h a 20:00h.

Sin perjuicio de las necesidades del servicio anteriormente descritas y su necesaria adaptación al horario del cliente, el horario concreto a realizar, será determinado por la empresa.

Por necesidades del servicio, el responsable del área correspondiente podrá autorizar excepciones a este horario.

Los servicios referidos contemplan la disponibilidad para actuaciones, fuera del horario citado, programadas para cambios fuera del horario normal o requeridas por la AEAT por ocurrir incidencias graves.



La prestación del servicio de guardia incluirá durante una semana al mes la atención y resolución de incidencias críticas del servicio en cualquier horario, 24x7, mediante conexión remota por ordenador proporcionado por la AEAT y, en caso de requerirse, desplazamiento a cualquiera de los dos CPDs del DIT (en Madrid capital) en menos de 1 hora. Podrá ser llevada a cabo, indistintamente, por cualquiera de los tres perfiles anteriormente indicados (Consultor C03, Analista-1 A103 y Técnico de Sistemas-2 TS201).

La prestación del servicio de guardia semanal será de lunes de 8 de la mañana hasta el lunes siguiente hasta las 8 de la mañana. La AEAT deberá solicitar la prestación del servicio de guardia con una antelación mínima de 24 horas

De manera excepcional, algunas tareas que puedan impactar en el correcto funcionamiento del servicio deberán ser realizadas fuera de la franja horaria indicada (horas nocturnas, fines de semana, etc.). En estos supuestos, el responsable de la AEAT informará con la debida antelación al responsable de la empresa, para obtener la oportuna autorización.

No obstante, el adjudicatario deberá comprometerse a una plena disponibilidad, para la correcta prestación del servicio en presencia de circunstancias excepcionales o imprevistas, sin que la realización del trabajo en estos casos, fuera del horario y franja habitual, tenga una consideración especial a efectos de cómputo de horas o tarifa aplicables a las mismas, sin perjuicio de la adecuación legal que requieran dichos aspectos en el estricto ámbito laboral de la relación del trabajador con su empresa.

Sólo se facturarán las horas efectivamente realizadas y las reducciones de horario no serán recuperables en meses posteriores.

6 TRANSFERENCIA DEL CONOCIMIENTO

El adjudicatario saliente deberá realizar el proceso de transición de salida, asegurando que se mantienen los servicios a la AEAT durante el traspaso del control de servicios y deberá colaborar activamente con el futuro suministrador y con la AEAT durante este proceso, para facilitar la transición sin causar perjuicios.

Durante un mes tras la ejecución del objeto del contrato, y de manera gratuita, el adjudicatario saliente deberá poner a disposición del entrante los técnicos que han prestado el servicio, con el objetivo de transferir dicho conocimiento.

Los objetivos de esta fase de transferencia son los siguientes:



- Garantizar la continuidad y estabilidad de los servicios en la fase de transición.
- Transferir el conocimiento en detalle de los servicios a la AEAT y a la nueva empresa adjudicataria.
- Generar la documentación y entregables que la AEAT le exija.

7 DIRECCIÓN Y CONTROL. REUNIONES DE SEGUIMIENTO

La dirección, organización y control del equipo de trabajo es responsabilidad de la empresa adjudicataria, como encargada de la prestación integral del servicio.

El adjudicatario nombrará a un responsable que actuará como interlocutor único a todos los efectos frente al responsable del seguimiento y control de la ejecución del contrato nombrado por la administración contratante, canalizando las comunicaciones y responsabilizándose de la gestión de los trabajos, y estando facultado para llevar adelante las acciones que se acuerden como consecuencia de las reuniones de seguimiento.

Las reuniones de seguimiento tendrán lugar, con carácter ordinario, al menos una vez al mes, y con carácter extraordinario siempre que las necesidades del servicio lo aconsejen, bien a propuesta de la empresa adjudicataria o de la organización. En dichas reuniones de seguimiento podrán participar, si la complejidad o diversidad del servicio así lo requiere, otros responsables designados al efecto tanto por la empresa adjudicataria como por la administración contratante.

El adjudicatario se compromete a facilitar en dichas reuniones la información y documentación necesaria para que por parte del DIT se obtenga un adecuado conocimiento de las circunstancias en que se desarrollan los trabajos, así como de los eventuales problemas que puedan plantearse y de las tecnologías, métodos y herramientas de resolución.

8 COMPOSICIÓN DEL EQUIPO DE TRABAJO

El equipo humano que se incorporará para la ejecución de los trabajos, tras la formalización del contrato, deberá estar formado por las personas relacionadas en la oferta adjudicataria y consecuentemente valoradas o en su defecto, en casos excepcionales en que no pueda mantenerse en su totalidad la propuesta inicial valorada, personas que reúnan las mismas características o superiores, en cuyo caso se requerirá la conformidad expresa del DIT.



9 GESTIÓN DEL EQUIPO DE TRABAJO REQUERIDO POR LA AEAT.

Una vez constituido el equipo inicial, la incorporación de un nuevo profesional contará con un periodo de prueba de 10 días hábiles. Dicho periodo se plantea para facilitar la adaptación al entorno y a los servicios a desarrollar del profesional propuesto, así como para comprobar el cumplimiento efectivo con el perfil solicitado. Si transcurrido el periodo de prueba la AEAT no expresa rechazo de manera expresa (mediante una solicitud de cese o sustitución del candidato), se entenderá por aceptado el profesional incorporado. En caso de que, durante ese periodo de prueba, el candidato haya sido rechazado, NO se facturará por él durante dicho periodo.

Los responsables de la gestión del presente contrato podrán solicitar la incorporación de un determinado número de profesionales cuando así lo requiera el proyecto, mediante notificación al adjudicatario.

Asimismo, podrán solicitar los cambios necesarios en el equipo, en caso de ser necesaria la variación del perfil de alguno de los componentes del mismo (estos cambios no se considerarán imputables a la empresa).

El adjudicatario facilitará la incorporación de los profesionales requeridos según lo expresado en párrafos anteriores, en un plazo máximo de un mes desde la solicitud de la Subdirección General de Planificación y Coordinación Informática de la AEAT. Con carácter general, una vez que haya sido aceptado un profesional, su permanencia en el equipo de trabajo será como mínimo de 3 meses.

Si el adjudicatario propusiera el cambio de una de las personas del equipo de trabajo por motivos ajenos al DIT, imputables o no a una baja en la empresa del personal afectado, deberá solicitarlo por escrito con 10 días hábiles de antelación, exponiendo las razones que obligan a ello y deberá facilitar la incorporación del sustituto, en caso de que se requiera, en un plazo máximo de un mes a contar desde la solicitud de la Subdirección General de Planificación y Coordinación Informática de la AEAT. Todo cambio, sustitución o incorporación debe ser aprobado por los responsables afectados y su permanencia en el equipo de trabajo igualmente será como mínimo de 3 meses.

Si se superara el plazo máximo de un mes para la incorporación del sustituto o la permanencia del mismo fuera inferior a 3 meses, se aplicarán las penalidades descritas en el Pliego de Cláusulas administrativas particulares.

Se establece un factor de rotación máximo de 10% anual del equipo medio del servicio, redondeando al alza, como número de cambios en los recursos sobre el equipo inicial, realizados durante 1 año por razones imputables al adjudicatario.



Se considera razón imputable al adjudicatario toda aquella situación ajena a la voluntad del Departamento de Informática Tributaria, que origine una o varias de las siguientes situaciones: que se sobrepase el periodo máximo de 1 mes establecido para los cambios/sustituciones o incorporaciones (salvo retrasos imputables a la AEAT), que la persona no permanezca en los servicios asignados un periodo mínimo de 3 meses, que la persona no se adecúe al perfil demandado por la AEAT (siempre y cuando la AEAT no haya demandado variación de perfil) o cualquier otra causa no imputable a una propuesta que parta del Departamento de Informática Tributaria.

En todos los casos anteriores se produce una afección en la continuidad del servicio, que podrá derivar en las penalizaciones correspondientes señaladas en el PCAP.

Documento firmado electrónicamente según detalle del encabezado.

Subdirector General de Tecnologías de Análisis de la Información e Investigación del Fraude.