

satec_

Plan TL
Plan de Impulso de las
Tecnologías del Lenguaje



Explicación de los desarrollos realizados en la solución de la herramienta Corpus Viewer

InfoDay sobre INTELIGENCIA ARTIFICIAL EN POLÍTICAS PÚBLICAS DE I+D+i – 26/06/2019

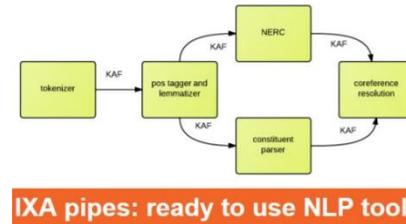
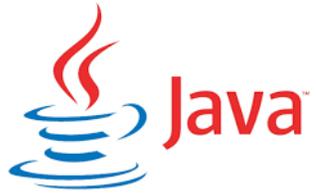
Breve descripción de Corpus Viewer

- Es una herramienta basada en técnicas de procesamiento de lenguaje natural (PLN), machine learning (ML) y traducción automática (MT) que permite el análisis conjunto de datos textuales no estructurados y metadatos estructurados en grandes corpus documentales
- Se basa, entre otras técnicas, en el modelado de tópicos y el análisis de grafos
- Corpus Viewer es una **plataforma genérica que puede ser explotada con prácticamente cualquier colección de documentos de texto**, el despliegue actual de la plataforma aloja principalmente corpus relacionados con el I+D
- Para más detalle, pueden consultar la web del Plan de Tecnologías del Lenguaje: <https://www.plantl.gob.es/tecnologias-lenguaje/actividades/plataformas/Paginas/corpus-viewer.aspx>

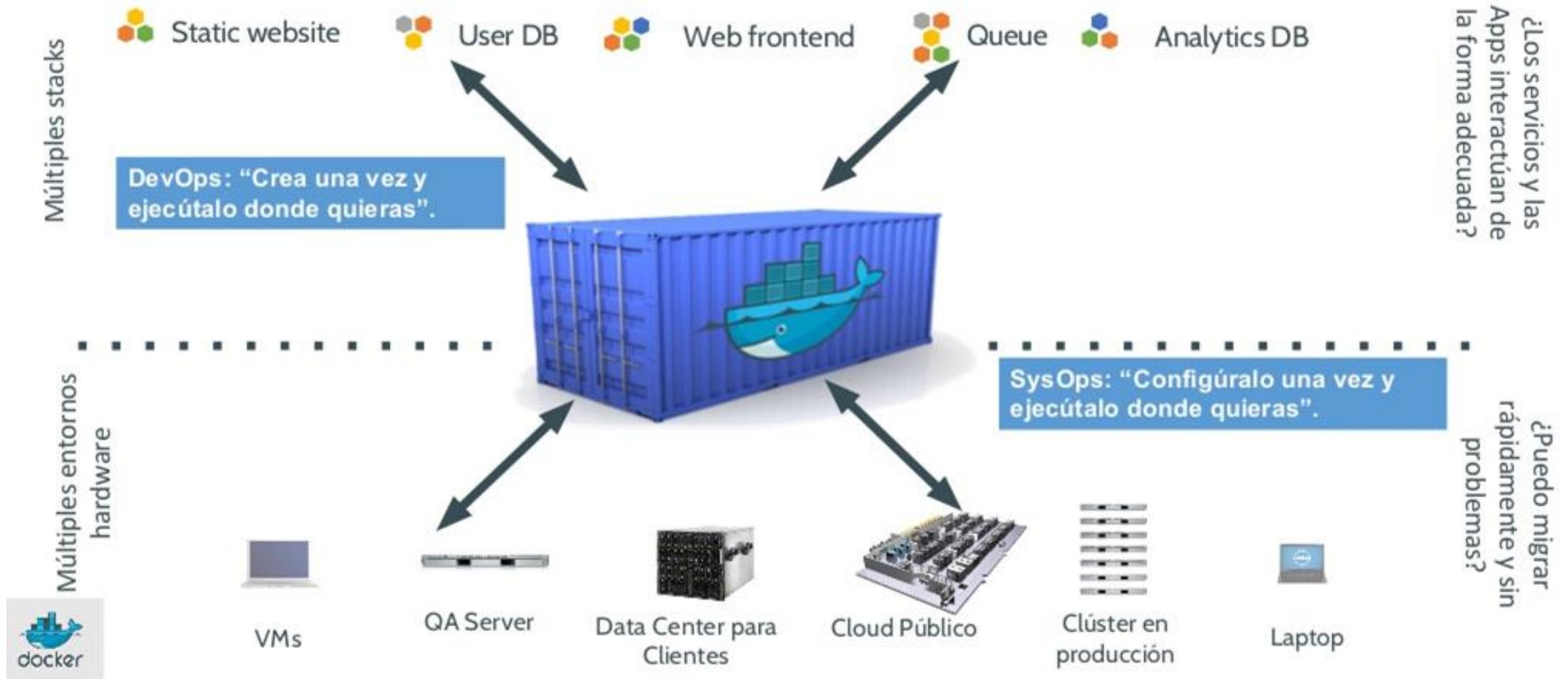
Corpus Viewer - funcionalidades

- Comparar las áreas de conocimiento y financiación de I+D+i en diferentes regiones geográficas
- Identificar ventajas competitivas entre países, regiones, organizaciones
- Identificar áreas de conocimiento de I+D+i, así como su emergencia, evolución e incluso hibridación con otras áreas de conocimiento (proporciona también agregación de metadatos y visualización de tablero de instrumentos de tipo BI)
- Perfilado de agentes de I+D (organizaciones, investigadores y empresas)
- Ayuda en la evaluación del impacto de las políticas públicas mediante el seguimiento de los resultados de las subvenciones, resultados a corto y largo plazo

Componentes y herramientas opensource

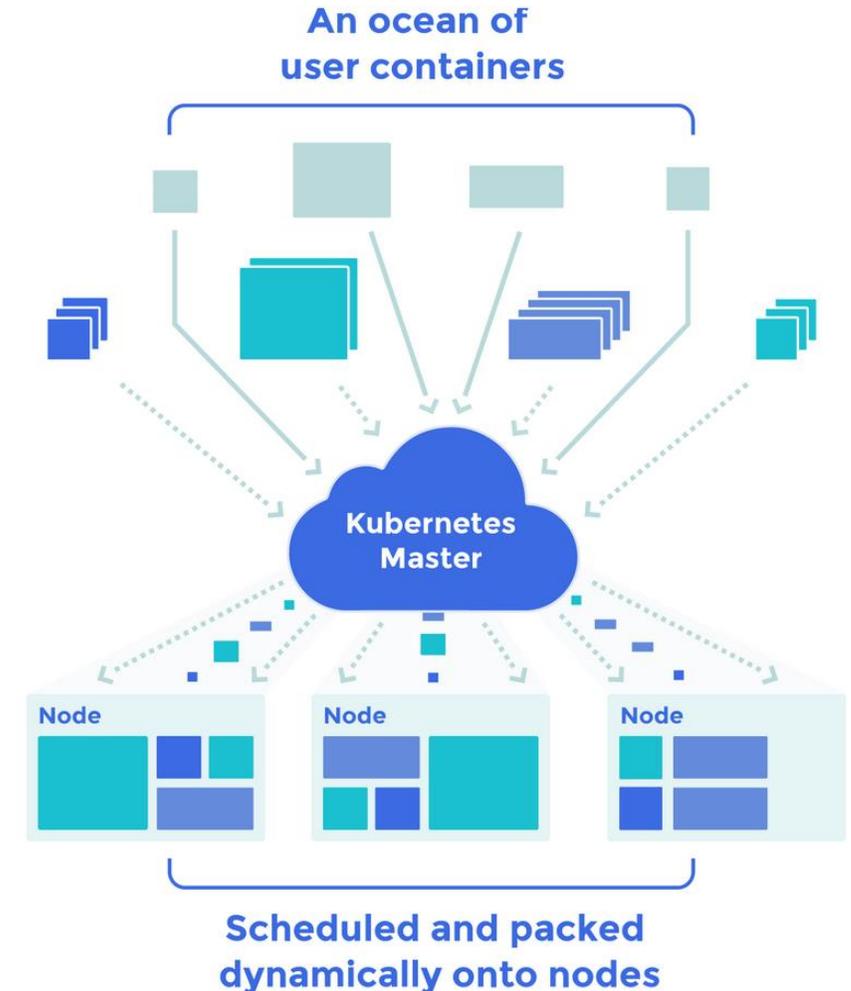


Despliegue basado en docker

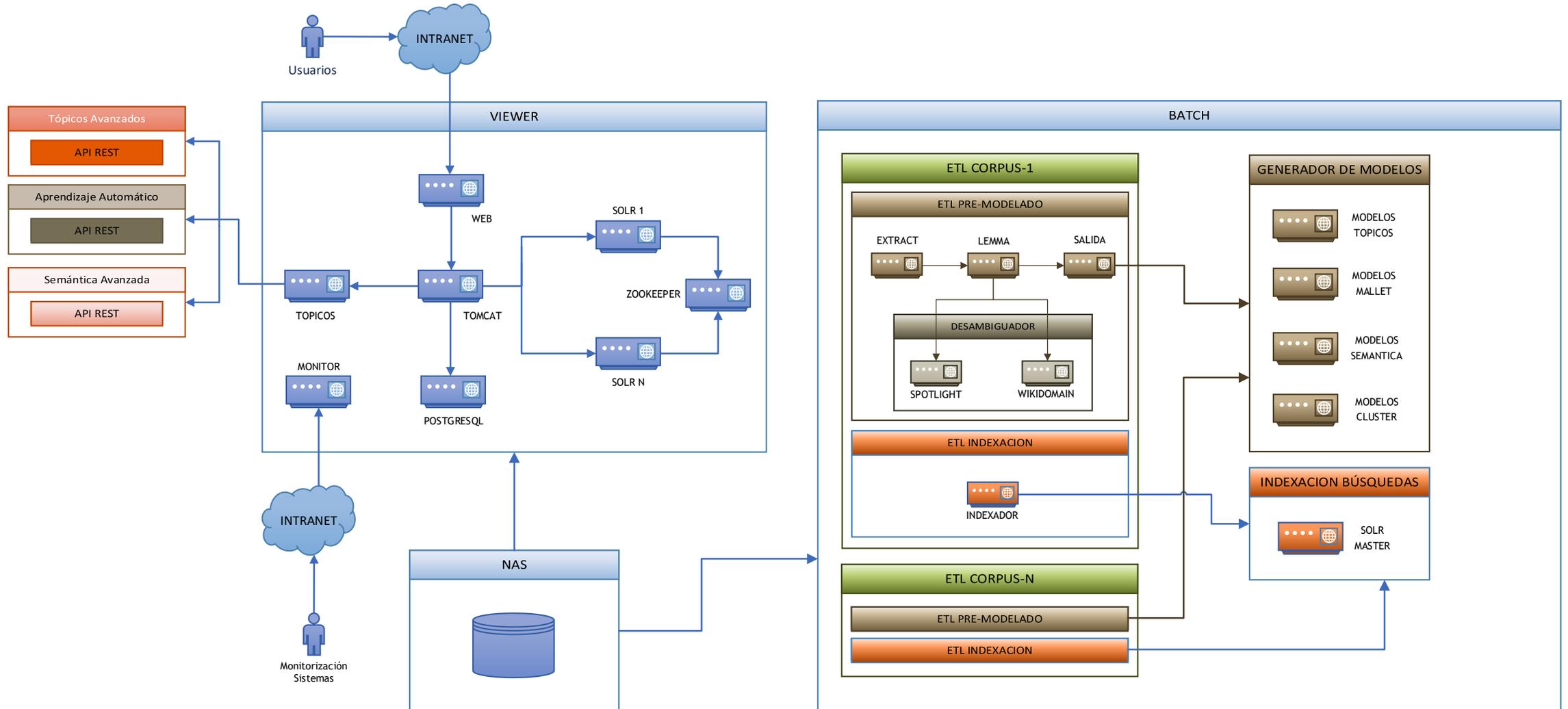


Orquestación mediante Kubernetes

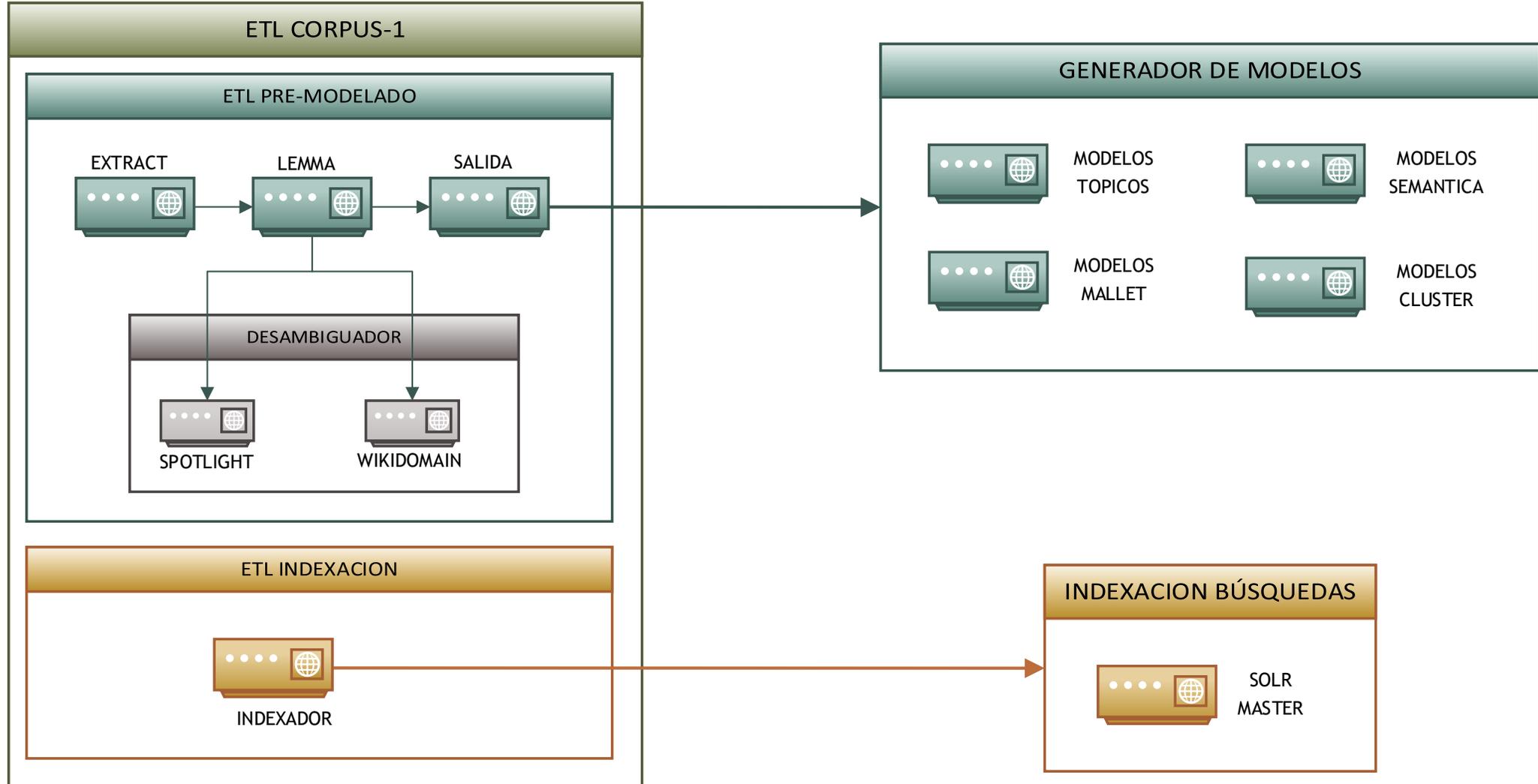
- Sistema de código libre para la automatización del despliegue, ajuste de escala y manejo de aplicaciones en contenedores
- Originalmente diseñado por Google y donado a la Cloud Native Computing Foundation
- Soporta diferentes ambientes para la ejecución de contenedores, incluido Docker



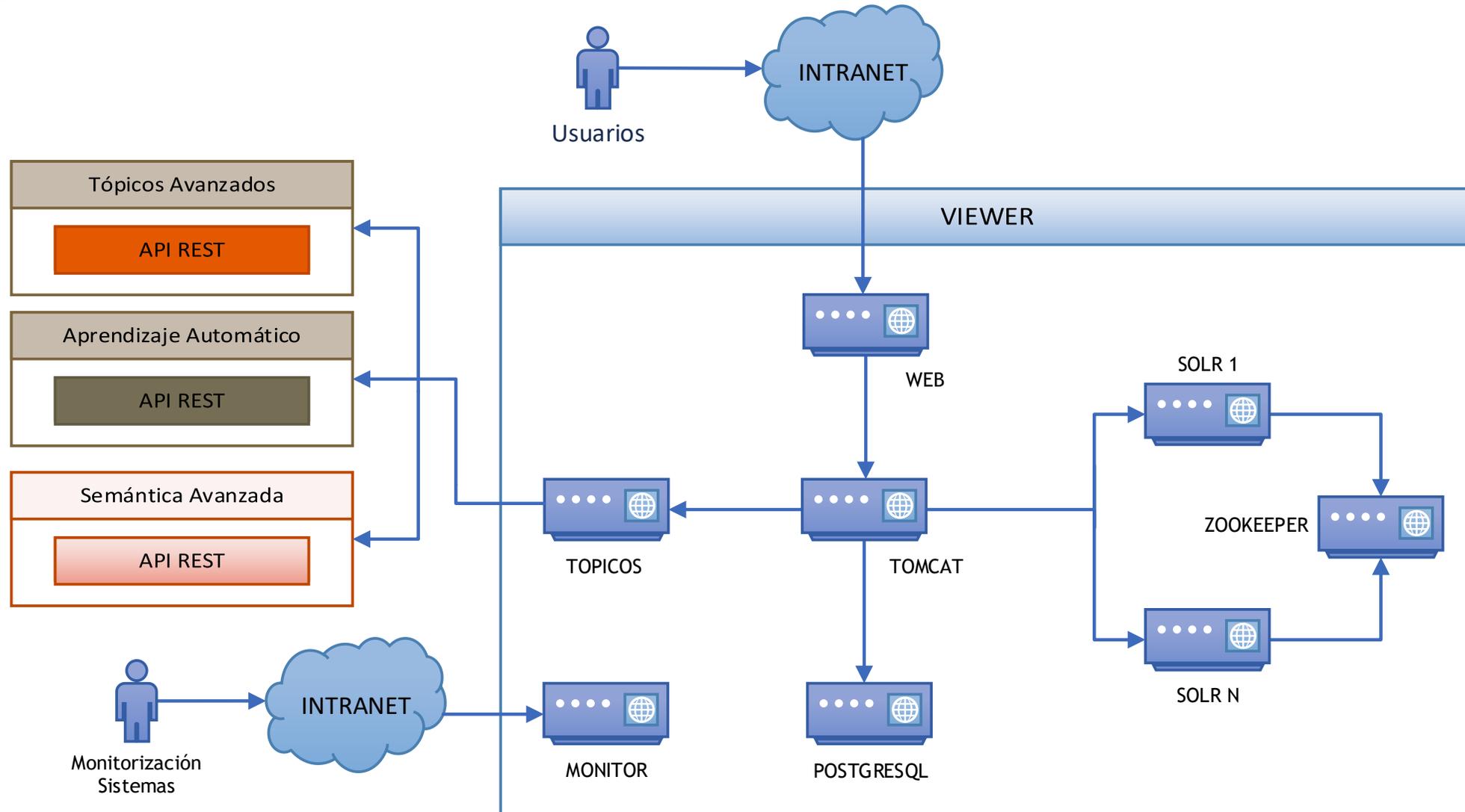
Arquitectura de componentes



BATCH



VIEW



Reingeniería de la solución

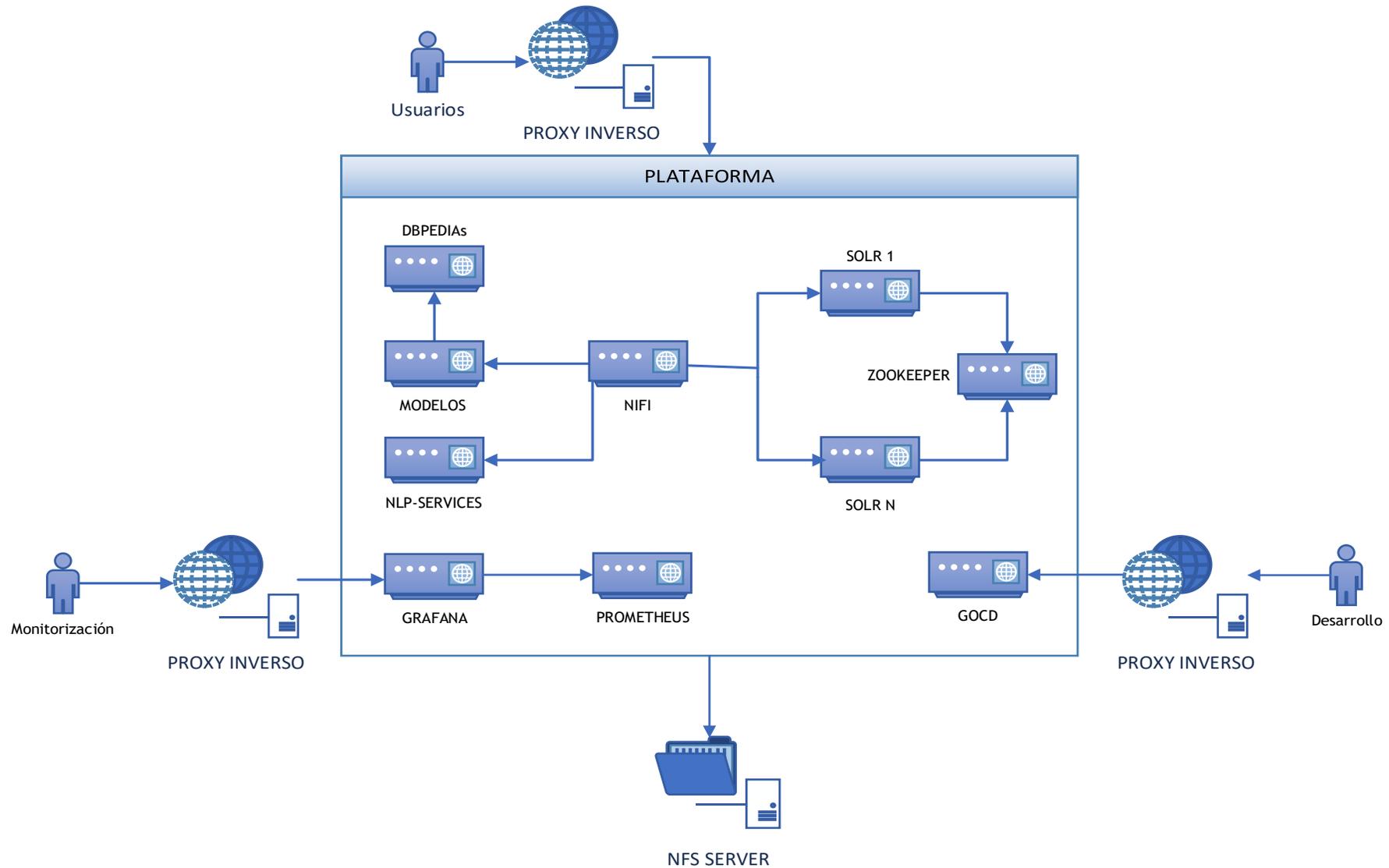
Arquitectura orientada a microservicios que permite:

- La escalabilidad de los componentes del sistema
- Definir de forma explícita la lógica que debe exponer cada componente por medio de un API
- Disponer de una base sólida para la distribución geográfica
- Sentar las bases para la implementación de una arquitectura federada

Componentes de alto nivel:

- Carga: incorporación de nuevos recursos a partir de fuentes de información
- Procesamiento: transformación y anotación de recursos en un formato común estructurado
- Análisis: creación de los modelos correspondientes
- Publicación: consulta de los resultados obtenidos a partir del análisis realizado
- Monitorización del sistema
- Integración y despliegue continuo

Arquitectura servicios





MUCHAS GRACIAS

david.nieto@satec.es

satec_

www.satecgroup.com