

Necesidades identificadas

La compañía Vocento requiere de una plataforma que garantice el **Gobierno y la Calidad de los Datos** y que les permita conocer el origen del dato para tener una trazabilidad end-to-end.

Para ello, se precisa de una hoja de ruta que incluye la implementación gradual de casos de uso, por lo que la plataforma debe ser flexible, escalable y configurable.

Las necesidades identificadas tienen que ver con los siguientes aspectos:

- **Funciones administrativas:** todas las tareas relacionadas con la gestión de información.
 - Glosario de negocio
 - Catálogo de datos
 - Unidades organizativas, roles, accesos y flujos de trabajo
- **Funciones operativas:** todas las tareas relacionadas con la operación diaria de la información.
 - Definición y desarrollo de seguridad/privacidad
 - Definición y desarrollo de reglas de negocio
 - Definición y desarrollo de reglas de calidad
 - Trazabilidad e identificación de datos
 - Generación de workflows para procesos de solicitud de cambios
- **Funciones ejecutivas:** todos los aspectos relacionados con la visualización y presentación de la información.
 - Integración de la información
 - Monitores de calidad
 - Remediación de calidad

Caso de uso

Para el primer caso de uso que cubra con las necesidades identificadas, se requiere gobernar un **cuadro de mando de operativa diaria** en Vocento (registro de usuarios, suscriptores, altas/bajas, newsletters, etc.) y un **cuadro de mando de “cierre de portadas”**, con la información en tiempo real de las noticias en portada de cada uno de los periódicos. Además de los cuadros de mando, se gobernarán todos los activos de datos (objetos) relacionados, con el objetivo de testear las funcionalidades y capacidades de Anjana Data y Omma Data (motor de calidad de datos integrado en Anjana), su flexibilidad y adecuación a las necesidades de Vocento. De esta forma, se consigue además una traza completa del ciclo de vida de los datos, desde su generación en las fuentes origen, hasta su consumo en los cuadros de mando y otras herramientas útiles para el negocio, pasando por los diversos procesos de transformación de datos en los diferentes sistemas.

Es importante destacar, que para poder implementar este caso de uso en Anjana Data, ha sido necesario definir y configurar aspectos globales y transversales como las diferentes entidades, jerarquías organizativas, roles, permisos y flujos de aprobación.

En cuanto a la implementación de calidad, se ha definido una estructura para la ejecución de históricos, datos diarios y comprobaciones entre cargas de información. Los resultados de estas ejecuciones se integran en Anjana Data de forma automática para su monitorización.

Las tablas provenientes de Hive (sistema de almacenamiento) son tablas desarrolladas por el equipo de Data Science, por lo que se han añadido los orígenes de datos e incluido además la documentación del desarrollo.

También se han creado procesos de solicitud de cambios, para que los usuarios de negocio (consumidores de información) puedan solicitar modificaciones sobre los cuadros de mando (estructura, dimensiones, indicadores, etc.).

Por tanto, adicionalmente a las entidades estándar como Términos de Negocio, Datasets, Soluciones, Instancias, Procesos, Reglas de Calidad y Data Sharing Agreements, se han definido e implementado en Anjana Data una serie de entidades no nativas específicas que cubren todos los aspectos necesarios para la implementación del caso de uso:

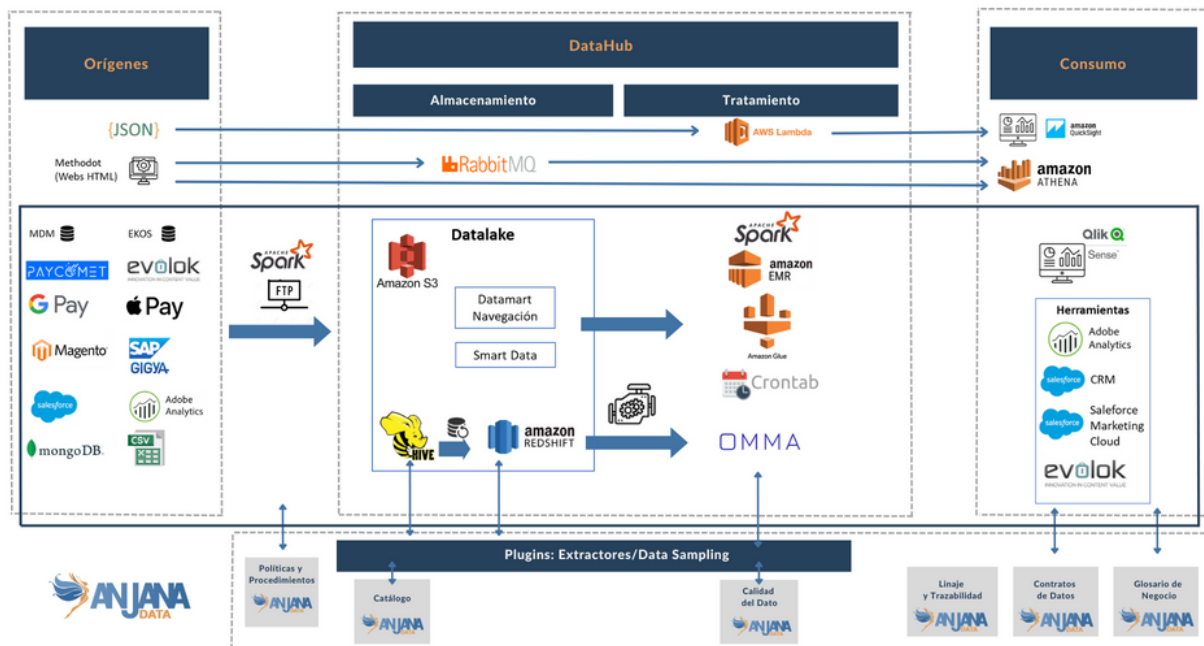
- Solicitud Modificación
- Documento de Apoyo
- Origen Fuente
- Dato Analítico de Visualización
- Proceso Analítico de Visualización
- Medida y Métrica (KPI)
- Incidencia
- Visualización de Datos

Objetivos planteados

El caso de uso seleccionado y desarrollado en Anjana Data tiene como objetivo poner a disposición de Vocento una plataforma que permita:

- Implantar las políticas de gobierno y normas de calidad necesarias.
- Identificar los orígenes de datos utilizados en los procesos de transformación.
- Identificar el movimiento de estos datos hacia el DataHub (centralizar de datos) de procesamiento y consumo.
- Garantizar la democratización y federación de los datos en Vocento.
- Definir roles y responsabilidades dotándolos de las herramientas y funcionalidades necesarias para llevar a cabo las acciones que les permitan cumplir con dichas responsabilidades.
- Trazar el ciclo de vida de datos gobernado con la plataforma Anjana Data, permitiendo obtener métricas e indicadores para las siguientes iteraciones y casos de uso sobre los modelos de gobierno y metamodelo planteados.
- Dotar al sistema de medidas de calidad del dato y planes de remediación con la ayuda de Omma Data y su integración en Anjana Data.

Arquitectura funcional y técnica



Beneficios obtenidos

Una vez conseguidos los objetivos planteados en la plataforma Anjana Data, se obtienen beneficios tales como:

- Centralizar la definición y ejecución ágil de políticas y procedimientos de acceso a datos.
- Personalizar totalmente el modelo de gobierno donde se encuentran catalogados los activos técnicos de datos, los activos de negocio y la relación existente entre ellos, pudiendo obtener el linaje desde producción hasta consumo.
- Disponer de una visión única de toda la información de forma ordenada para su tratamiento y consumo.
- Comunicar los distintos sistemas de Vocento e intercambiar información de una manera gobernada gracias a la centralización de su gestión. Se consigue acercar los datos de Vocento a todo tipo de perfiles, tanto técnicos como de negocio, y permite que los usuarios puedan conocer la información de una forma mucho más cercana y menos técnica.
- Poner a disposición de los roles distintas capacidades de consulta, modificación y acceso a la información.
- Disponer de un gobierno proactivo y preventivo mediante funcionalidad que permite monitorizar con grafos las relaciones de todas las entidades disponibles para realizar una gestión de impactos ante cambios de estructura, de forma que sea posible ver y analizar cómo fluyen y se comportan los datos entre los diferentes tipos de entidades y las transformaciones que experimentan.
- Tener controlados los accesos a datos a través de un Data Marketplace. Mediante la firma de contratos de datos es posible adherirse a entidades y ejecutar flujos de acceso a la información subyacente como si fuera un “carrito de la compra”.
- Disponer de un motor de ejecución de reglas de calidad de datos integrado en Anjana Data y en el ecosistema de Vocento, de modo que se monitorice y se garantice la calidad de la información que se está compartiendo.

Retos superados

Se han superado retos, tanto tecnológicos como funcionales. Algunos de los más importantes son:

- Crear un entorno común para un despliegue coordinado de Anjana Data y Omma Data sobre una arquitectura ya existente con gran complejidad tecnológica.
- Cubrir completamente las necesidades planteadas en el caso de uso. Algunas de ellas han supuesto desarrollos personalizados.
- Establecer las normas básicas y la manera de funcionar para empezar a gobernar la información.
- Identificar y conocer la calidad de los datos en todo momento.
- Poner a disposición de los usuarios de negocio un procedimiento y mecanismo que centralice todas las solicitudes de cambios sobre los cuadros de mando.
- Generar cultura del dato: Anjana Data ha servido para ayudar a traccionar sobre Vocento las iniciativas de gobierno definidas. Se ha conseguido que las unidades de negocio empiecen a trabajar con la plataforma, considerándola una pieza clave y amigable para entender mejor la información que utiliza la organización, ya que se han alineado y relacionado conceptos técnicos y de negocio, buscando un lenguaje común para todos.