

HERA, HERRAMIENTA DE EVALUACIÓN RÁPIDA DEL AGUA

PROTOCOLO DE INFORMACIÓN

DICIEMBRE 29, 2023

1. VISIÓN GENERAL

1.1. Contexto

La actual condición de sequía ha relevado la importancia de la producción, almacenamiento y despliegue de información relacionada con el ciclo del agua, para la gestión efectiva de los recursos hídricos. En este contexto, diferentes grupos de I+D+i se han involucrado en desarrollar plataformas de producción, gestión y visualización de información relacionada con los recursos naturales y cuyo trabajo ha abordado diferentes aspectos del ciclo del dato – captura, almacenamiento, procesamiento, despliegue, uso y a diferentes grupos objetivos – planificadores, agricultores, reguladores, etc.

Los resultados de distintos diagnósticos coinciden en la falta de coordinación de los productores de datos, la dispersión o inexistencia de formatos y protocolos de manejo de datos y la falta de catalogación de procesos de toma de decisión. Así, el problema es la coordinación y consideraciones de gobernanza de datos sobre comprar más y más poderosos computadores. Más aún, la penetración de conceptos como la inteligencia artificial, la analítica de datos o el análisis de *big data* han presionado a levantar estos tipos de acciones de manejo de datos. Sin embargo, los conceptos antes mencionados dependen de la calidad, volumen y confiabilidad de los datos, además de un proceso de catalogación de procesos de toma de decisiones. Todo, al final, se reduce a programar esos procesos. Por lo que si no hay procesos, aunque existan datos, no hay producción de información.

1.2. Descripción

HERA es una plataforma de información diseñada para ser colaborativa, dinámica y confiable, presentando los datos desde un nivel puramente descriptivo a un nivel analítico, propiciando la toma de decisiones a nivel sectorial.

HERA tiene por propósito ser un concentrador (hub) de información desde fuentes públicas, en el que diferentes proveedores de datos se encuentren mediante la co-creación con información fidedigna relacionada con el agua en la Región Metropolitana. Y, a la vez, esta herramienta ayudará a los tomadores de decisiones y al público general, permitiendo acceder a la información con un alto estándar técnico, como los principios FAIR (findability, accessibility, interoperability, and reusability; Wilkinson et al. 2016), considerar las variables hídricas esenciales declaradas por la organización meteorológica mundial (<https://gcos.wmo.int/en/essential-climate-variables>) y visualizar y generar reportes relacionados con la oferta, demanda, calidad y uso de agua en la Región Metropolitana.

1.3. Objetivo

Conformar un entorno abierto para compartir y visualizar información técnica de uso público, de disponibilidad y demanda de agua a nivel espacial y temporal a través de estaciones de monitoreo disponibles pertenecientes a instituciones públicas, incluyendo además información de contexto comunal aportada por los 52 municipios de la Región Metropolitana.

2. BASE DE DATOS

2.1 Consideraciones preliminares

La plataforma HERA, se basa en datos públicos recopilados/descargados desde la web de la Dirección General de Aguas (DGA), Dirección Meteorológica de Chile (DMC), Superintendencia de Medio Ambiente (SMA), y Superintendencia de Servicios Sanitarios (SISS). Base de datos que fue descargada, depurada, y preparada para su posterior visualización e interacción con los (las) usuarios (as).

En ningún caso, la base de datos es manipulada para un posible relleno y/o reemplazo de información, pues el objetivo de la plataforma es concentrar los datos relacionados al agua y por sobre todo entregar datos confiables, manteniendo la información real generada por cada una de las instituciones o entidades mencionadas anteriormente.

Datos de Oferta

Los datos de oferta son estructurados de manera que permitan entregar un análisis a nivel de cuenca y subcuenca, generando una visualización espacio - temporal a través de paneles interactivos amigables e intuitivos para el (la) usuario (a).

Datos de Demanda

Los datos de demanda, reflejados a través de la información de consumo, son estructurados de manera que permitan entregar un análisis a nivel de comuna, generando una visualización espacio - temporal a través de paneles interactivos amigables e intuitivos para el (la) usuario (a).

2.2 Fuentes de información

Las instituciones que proporcionan información son: Dirección General de Aguas (DGA), Dirección Meteorológica de Chile (DMC), Superintendencia de Medio Ambiente (SMA) y Superintendencia de Servicios Sanitarios (SISS).

Para el lanzamiento y puesta en marcha de la plataforma, las instituciones entregaron datos a través de descargas libres desde la web de cada una de ellas; con excepción de la SISS, quien colaboró con información solicitada por medio de Ley de Transparencia.

Para el funcionamiento regular de la plataforma, el Fondo de Agua Santiago-Maipo ha celebrado diversos convenios de colaboración, por los que tanto la DGA como la SISS se han comprometido a poner a disposición del FDA-SM, de manera permanente y continua, y siempre actualizada, los datos respectivos.

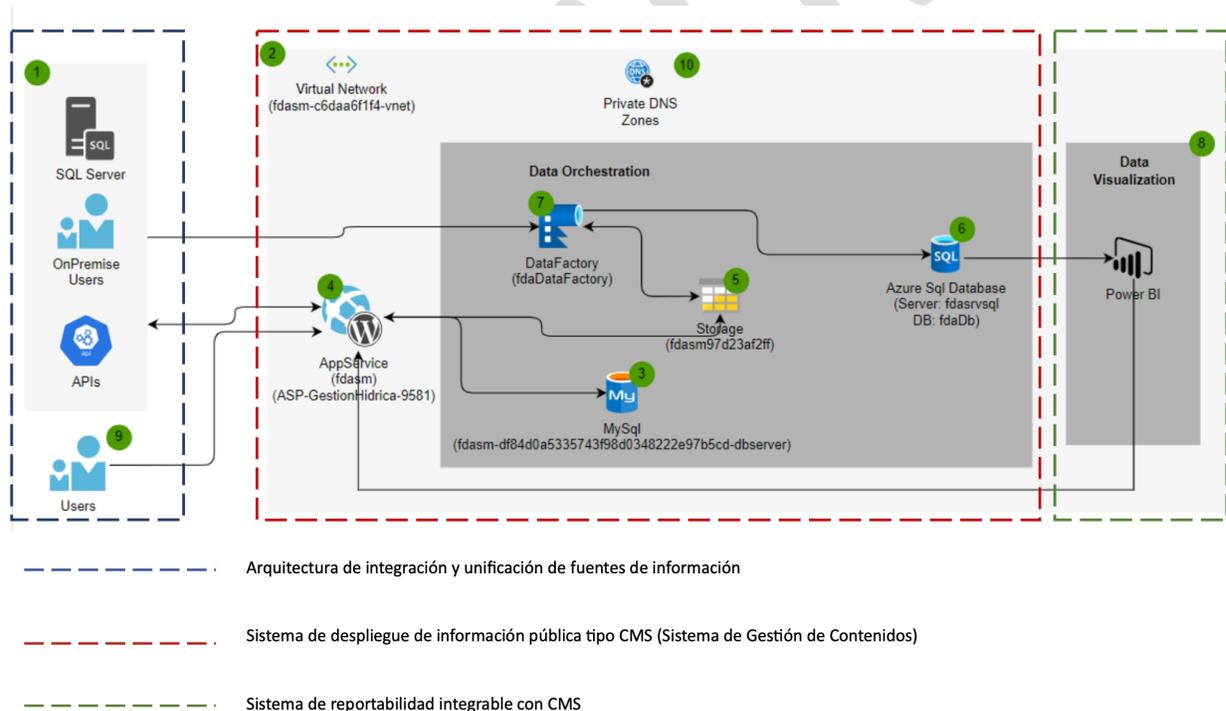
2.3 Estructura

HERA tiene la capacidad de unificar diversas fuentes de datos, ya sea que se trate de fuentes estructuradas o no estructuradas, lo que permitirá analizar, depurar, transformar y poner a disposición esa información tanto para el uso público, mediante información estática en su capa de representación, como para uso por parte de equipos especialistas en desarrollo de análisis descriptivo / prescriptivo, para la generación de reportes dinámicos y dashboards que puedan ser generados por el equipo de desarrollo y colaboradores de FDA-SM, quienes además podrán publicar dichos resultados en la capa de representación, Factor IT (2023).

Su arquitectura TO-BE, alojada en un ecosistema Cloud Azure, permitirá que la plataforma sea flexible para potenciales necesidades futuras; teniendo siempre como pilares fundamentales: la Reusabilidad, Interoperabilidad y Portabilidad, Disponibilidad y Privacidad.

La estructura cuenta con 3 capas, de acuerdo a la siguiente imagen:

Imagen No 1. Diagrama Estructura Plataforma HERA



Fuente: Factor IT, 2023

El WordPress de la Plataforma HERA, está contenida en Azure App Service para el despliegue de sus vistas. Se alimenta directamente desde el servidor de base de datos MySQL con el fin de poder

consumir la información almacenada. Por otro lado, existe un repositorio que se alimenta de las imágenes y logos generados en cada pestaña del sitio Web, Factor IT (2023).

2.4 Preparación de base de datos

Para una primera etapa, la plataforma HERA cuenta con 5 archivos de origen, distribuidos en 2 campos: oferta y consumo:

Modelo	Archivo	Variable	Detalle
Oferta	<u>pp_mm_day_actualizado.xlsx</u> <i>Periodo_data:</i> Inicio 01-01-1950 Fin 31-03-2020	Precipitación	Datos de precipitación acumulada diaria para 62 estaciones pluviométricas ubicadas en la región Metropolitana. Las estaciones son de las instituciones Dirección General de Aguas (DGA) y Dirección Meteorológica de Chile (DMC)
	<u>q_m3/s_day_actualizado.xlsx</u> <i>Periodo_data:</i> Inicio 01-01-1950 Fin 31-03-2020	Caudal	Datos de caudal promedio diario para 39 estaciones fluviométricas ubicadas en la región Metropolitana. Las estaciones son de la institución Dirección General de Aguas (DGA).
	<u>Cobertura_nival_Maipo.csv</u> <i>Periodo_data:</i> Inicio 25-02-2000 Fin 21-01-2023	Nieve	Datos diarios de superficie de cobertura nival en la cuenca del Maipo.
	<u>Embalses_RM.xlsx</u> <i>Periodo_data:</i> Inicio 01-01-2022 Fin 31-12-2022	Embalse	Datos diarios de volumen promedio de agua almacenados en embalses de la Región Metropolitana que son monitoreados por DGA y cuyos datos pueden ser obtenidos mediante la página web de la institución. Dado esto, sólo se tienen datos del embalse El Yeso.

Consumo	<u>SIFAC - CONSUMOS Y CIENTES MENSUALES 2012-2021.xlsx</u> <i>Fecha inicio:</i> 01-2012 <i>Fecha final:</i> 12-2021	Consumo	Hoja 1: Clientes 2012-2020 Datos de los clientes: región, rut, razón social, tipo de contrato, fecha de contrato, tipo de servicio Hoja 2: Consumo 2012-2020 Consumo de agua potable y Descarga de agua servida Hoja 3: Sistema-Localidad
---------	---	---------	--

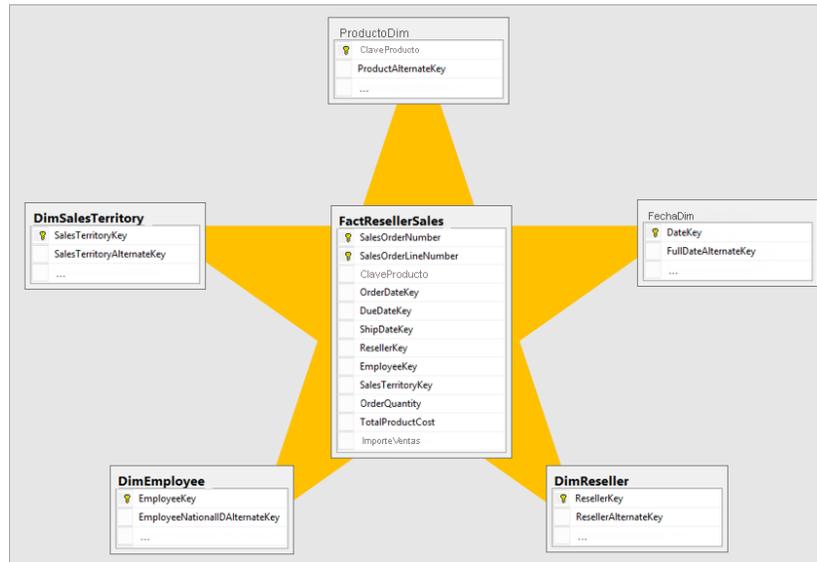
A partir de estos, se crearon en base al modelo de estrella, 2 modelos que se detallan a continuación:



- a) FACT_OFERTA
- b) FACT_CONSUMO

El modelo estrella es un enfoque utilizado en la creación de esquemas de bases de datos para sistemas de Business Intelligence (BI) y Datawarehousing, ya que proporciona un rendimiento rápido y eficiente para consultas. Asimismo, se aplica este modelo porque es escalable, por lo que, si en el futuro se amplía la cantidad de datos, se pueden agregar fácilmente sin dificultad, Factor IT (2023).

Imagen No 2. Modelo ESTRELLA

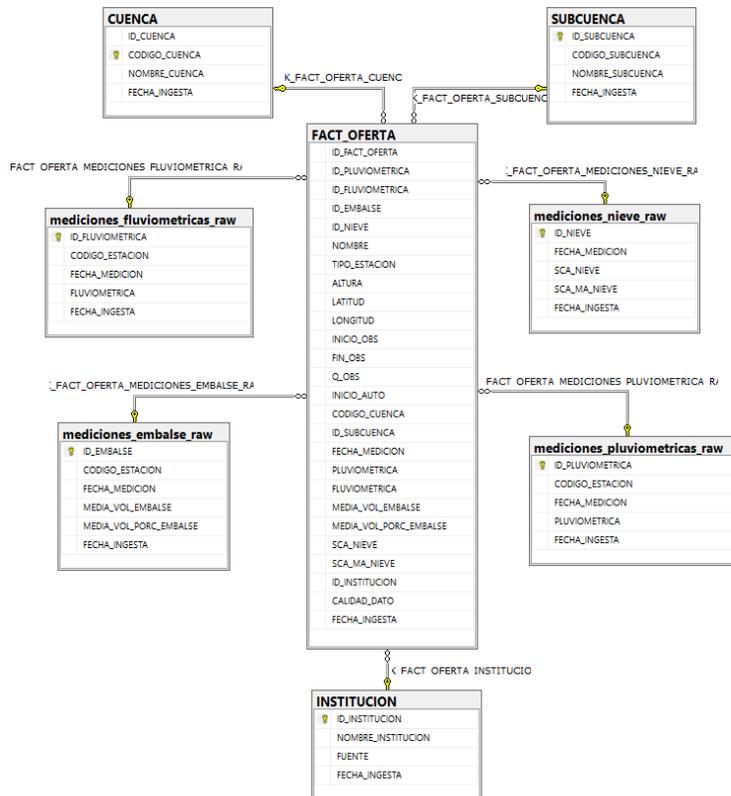


Fuente: Factor IT, 2023

Para el modelo de OFERTA, se ha considerado sólo mediciones reales, es decir, mediciones a partir de la fecha de inicio de observaciones de cada estación.

Tablas	Nombre
Tablas Raw	Información_Estaciones_Raw
	Mediciones_Pluviometricas_Raw
	Mediciones_Fluviometricas_Raw
	Mediciones_Embalse_Raw
	Mediciones_Nieve_Raw
Tablas Nuevas	Cuenca
	Subcuenca
	Institucion
Tabla General	FACT_OFERTA

Imagen No 3. Diagrama Relacional Modelo FACT_OFERTA



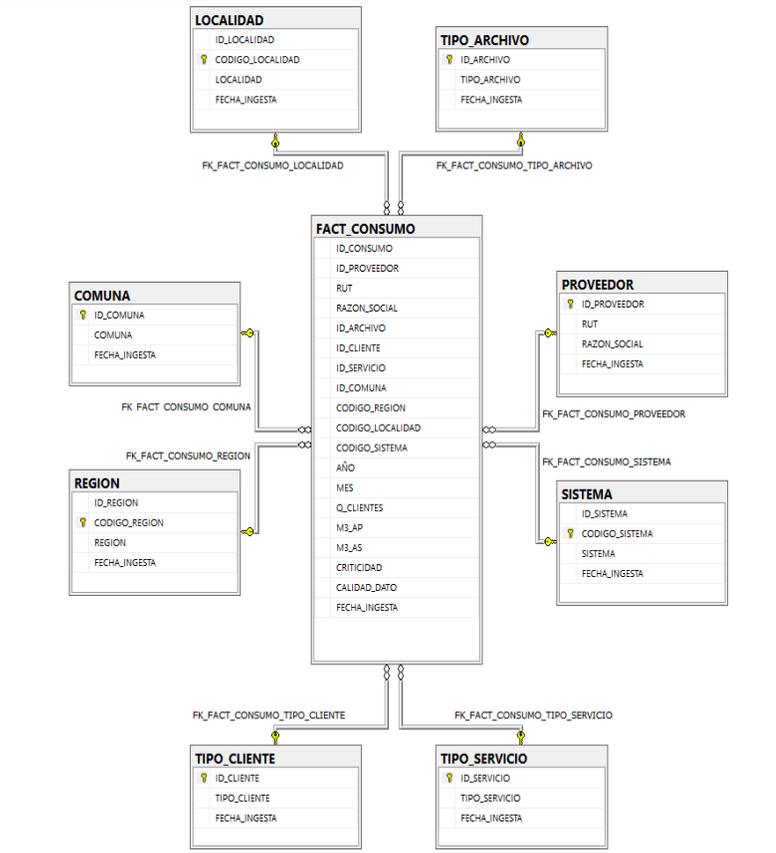
101

Fuente: Factor IT, 2023

Para el modelo de CONSUMO, se ha considerado los datos correspondientes a la Región Metropolitana

Tablas	Nombre
Tablas Raw	Informacion_Cliente_Consumo_Raw
	Sistema_Localidad_Raw
Tablas Nuevas	Tipo_Servicio
	Tipo_Cliente
	Tipo_Archivo
	Proveedor
	Region
	Localidad
	Comuna
Tabla General	FACT_CONSUMO

Imagen No 4. Diagrama Relacional Modelo FACT_CONSUMO

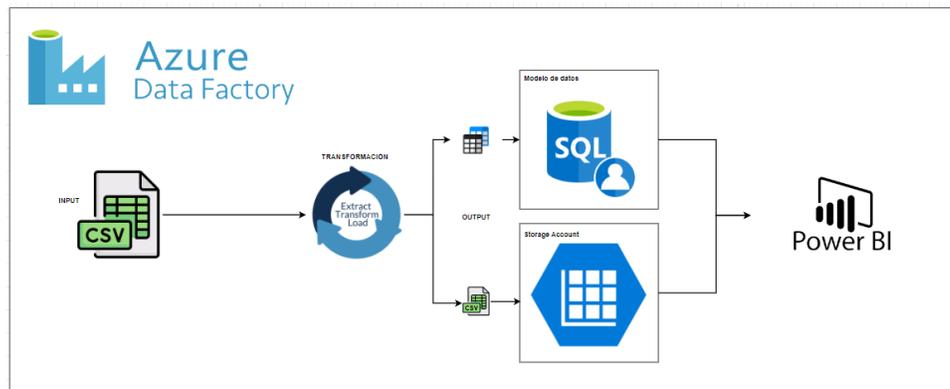


Fuente: Factor IT, 2023

Ahora bien, la ingesta de datos se realiza a partir de 5 archivos crudos previamente cargados en el contenedor “raw data” en el Storage Account de Azure, siendo consumidos en los pipelines del flujo de Data factory. Lo anterior, con el objetivo de procesar la información, pivotarla desde origen y obtener una salida en el formato correcto.

En esta etapa, la plataforma HERA presenta una estructura donde se inserta la información al Azure Database SQL, para cargar el modelo de datos generado para lo que es consumo y oferta.

Imagen No 5. Diagrama flujo de datos



Fuente: Factor IT, 2023

2.5 Reportes

HERA, cuenta con una sección técnica/científica que permite profundizar a través de un análisis del comportamiento histórico de los datos hídricos, tanto de Oferta como de Demanda (consumo). De esta manera, dentro del desarrollo se incluyen paneles interactivos utilizando Power Bi¹, mostrando de forma amigable e intuitiva diferentes parámetros estadísticos y tendencias de cada uno de los indicadores incluidos en esta etapa.

2.5.1 Reportes de OFERTA

Se detalla a continuación la distribución de paneles generados:

Oferta de agua	Caudales_2
	Datos_secos
	Precipitaciones_Maipo_1
	Acumulada_anual_estaciones

- Datos_Secos

Este reporte indica la cantidad de días secos por estación agrupados por año.

La cantidad de registros del reporte generado es 511 871.

Se incluyen 62 estaciones pluviométricas con un período de data desde 01-01-1980 hasta 31-03-2020.

Contiene 16 campos, los cuales se detallan en la siguiente tabla:

Columna	Detalle
Date	Fecha (YYYY-MM-DD) en la que se realiza la medición

¹ Power BI es un servicio de análisis de datos de Microsoft orientado a proporcionar visualizaciones interactivas y capacidades de inteligencia empresarial con una interfaz lo suficientemente simple como para que los usuarios finales puedan crear por sí mismos sus propios informes y paneles.

Estacion	Código único que identifica a la estación pluviométrica
Value	Valor de la medición de precipitación
Nombre	Código de la estación pluviométrica
Altura	Valor de altura geográfica (en metros), con respecto al nivel del mar, de cada estación pluviométrica
Latitud	Latitud geográfica de cada estación pluviométrica.
Longitud	Longitud geográfica de cada estación pluviométrica.
Codigo_Cuenca	Código identificador de la cuenca en la que se ubica cada estación.
Nombre_Cuenca	Nombre de la cuenca en la que se ubica la estación.
Codigo_Sub_Cuenca	Código identificador de la subcuenca en la que se ubica cada estación.
Nombre_Sub_Cuenca	Nombre de la subcuenca en la que se ubica la estación.
Inicio_Observaciones	Fecha de inicio del registro de observaciones de cada estación.
Fin_Observaciones	Fecha de término del registro de observaciones de cada estación.
Cantidad_Observaciones	Cantidad de mediciones registradas hasta la actualidad
Year	Año en el que se realiza la medición,
Dray_Days	Cantidad de días secos acumulado al año.

- Caudales_2

Este reporte indica las mediciones fluviométricas en conjunto con información de la estación de medición.

La cantidad de registros del reporte generado es 307 918.

Contiene 12 campos, los cuales se detallan en la siguiente tabla:

Columna	Descripción
Fecha_Medicion	Fecha de medición en cada estación. Los datos corresponden a valores diarios de caudal, el campo de fecha representa una serie de tiempo diaria desde 1980-01-01 hasta 2020-03-31.
Year	Año en el que se realiza la medición en la estación
Month	Mes en el que se realiza la medición en la estación
Day	Día en el que se realiza la medición en la estación
Value	Valor de la medición de caudal
Gauge_Id	Código de la estación fluviométrica
Gauge_Name	Nombre de la estación fluviométrica
Gauge_Lat	Latitud geográfica de cada estación fluviométrica.
Gauge_Lon	Longitud geográfica de cada estación fluviométrica
Record_Period_Start	Fecha en la que se inició el registro de observaciones en cada estación.

Record_Period_End	Fecha en la que terminó el registro de observaciones en cada estación, limitado según la fecha de recopilación de la base de datos.
N_Obs	Cantidad de mediciones registradas hasta el momento en cada estación.

- Precipitaciones_Maipo_1

Este reporte indica todas las mediciones pluviométricas en conjunto con información de la estación de medición. La cantidad de registros del reporte generado es 511 871.

Contiene 14 campos, los cuales se detallan en la siguiente tabla:

Columna	Detalle
Date	Fecha de la medición en la estación pluviométrica
Estacion	Código único que identifica a la estación pluviométrica
Value	Valor de la medición de precipitación
Nombre	Nombre de la estación pluviométrica
Altura	Valor de altura geográfica de la estación pluviométrica
Latitud	Latitud geográfica de cada estación pluviométrica.
Longitud	Longitud geográfica de la estación pluviométrica
Codigo_Cuenca	Código identificador de la cuenca en la que se ubica cada estación.
Nombre_Cuenca	Nombre de la cuenca en la que se ubica la estación.
Codigo_Sub_Cuenca	Código identificador de la subcuenca en la que se ubica cada estación.
Nombre_Sub_Cuenca	Nombre de la subcuenca en la que se ubica la estación.
Inicio_Observaciones	Fecha de inicio del registro de observaciones de cada estación.
FIN_OBSERVACIONES	Fecha de término del registro de observaciones de cada estación.
CANTIDAD_OBSERVACIONES	Cantidad de mediciones registradas hasta la actualidad.

- Acumulada_anual_estaciones

Este reporte indica todas las mediciones pluviométricas en conjunto con información de la estación de medición. La cantidad de registros del reporte generado es 511 871.

Contiene 14 campos, los cuales se detallan en la siguiente tabla:

Columna	Detalle
Date	Fecha de la medición en la estación pluviométrica
Estacion	Código único que identifica a la estación pluviométrica
Value	Valor de la medición de precipitación

Nombre	Nombre de la estación pluviométrica
Altura	Valor de altura geográfica de la estación pluviométrica
Latitud	Latitud geográfica de cada estación pluviométrica
Longitud	Longitud geográfica de cada estación pluviométrica.
Codigo_Cuenca	Código identificador de la cuenca en la que se ubica la estación.
Nombre_Cuenca	Nombre de la cuenca en la que se ubica la estación.
Codigo_Sub_Cuenca	Código identificador de la subcuenca en la que se ubica la estación.
Nombre_Sub_Cuenca	Nombre de la subcuenca en la que se ubica la estación
Inicio_Observaciones	Fecha de inicio del registro de observaciones de cada estación.
Fin_Observaciones	Fecha de término del registro de observaciones de cada estación.
Cantidad_Observaciones	Cantidad de mediciones registradas hasta la actualidad.

2.5.2 Reportes de Consumo

Para el grupo de Consumo, se entregan 3 reportes.

Consumo de agua	clientes
	clientes_consumo_completo
	consumo_agua_comunas_RM

- Clientes

Este reporte indica la cantidad de clientes por empresa proveedora.

La cantidad de registros del reporte generado es 138 106.

Contiene 8 campos, los cuales se detallan en la siguiente tabla.

Columnas	Descripción
Año	Año del contrato
Mes	Mes del contrato
Razon_Social	Nombre legal de la empresa a la que está asociado el contrato.
Localidad	Localidad en la cual se ubica el cliente asociado al contrato.
Comuna	Comuna en la cual se ubica el cliente asociado al contrato.
Tipo_Cliente	Clasificación del cliente según tipo de cliente.
Tipo_Servicio	Clasificación del(los) cliente(s) según el tipo de servicio que tiene asociado el contrato.
Q_Clientes	Cantidad de clientes asociados al contrato. Si es 0 (NULL en archivo de origen) significa que no tiene clientes.

- Clientes_consumo_completo

Este reporte indica la cantidad de clientes por empresa proveedora.

La cantidad de registros del reporte generado es 138 106.

Contiene 12 campos, los cuales se detallan en la siguiente tabla.

Columna	Detalle
Tipo_Archivo	Clasificación del contrato según las características que tiene.
Año	Año del inicio del contrato
Mes	Mes del inicio del contrato
Rut	RUT de la empresa a la que está asociado el contrato.
Razon_Social	Nombre legal de la empresa a la que está asociado el contrato
Localidad	Indica el número de la región a la que corresponde el contrato según la codificación de regiones.
Comuna	Comuna en la cual se ubica el cliente asociado al contrato.
Tipo_Cliente	Clasificación del cliente según tipo de cliente.
Tipo_Servicio	Clasificación del cliente según el tipo de servicio
M3_Ap	Consumo de agua potable (m3) según contrato y cliente.
Fecha	Registra todos los días 1 con las columnas año y mes (01-mes-año)
Q_Clientes	Cantidad de clientes asociados al contrato.

- Consumo_agua_comunas_RM

Este reporte indica la cantidad de consumo de agua por comuna.

La cantidad de registros del reporte generado es 138 106.

Contiene 11 campos, los cuales se detallan en la siguiente tabla.

Columnas	Descripción
Región	Indica el nombre de la región a la que corresponde el contrato.
Año	Año del inicio del contrato.
Mes	Mes del inicio del contrato.
Razon_Social	Nombre legal de la empresa a la que está asociado el contrato.
Tipo_Cliente	Clasificación del cliente según tipo de cliente.
Tipo_Servicio	Clasificación del(los) cliente(s) según el tipo de servicio que tiene asociado el contrato.
M3_Ap	Descarga de agua potable (m3) según contrato y cliente.
Fecha	Registra todos los días 1 con las columnas año y mes (01-mes-año)
Población	No tenemos la información de la columna.
Localidad	Indica el número de la región a la que corresponde el contrato según la codificación de regiones.

Comuna	Comuna en la cual se ubica el cliente asociado al contrato.
--------	---

3. GLOSARIO

Palabra	Descripción
Data Factory	Servicio de Azure para la orquestación de datos en la nube.
Storage Account	Contenedor de datos en Azure para almacenar y acceder a objetos de datos.
Power BI	Plataforma de análisis de datos y visualización de informes interactivos.
Etl	Proceso de Extracción, Transformación y Carga de datos en sistemas de información.
Claves Foráneas	Restricción que asegura la integridad referencial de datos en una base de datos relacional.
Datawarehousing	Almacenamiento y gestión de grandes cantidades de datos en un entorno empresarial.
Business Intelligence	Proceso de análisis de datos para mejorar la toma de decisiones empresariales.
Azure SQL Database	Servicio de base de datos relacional en la nube de Azure.
API	Interfaz de programación de aplicaciones, permite a las aplicaciones comunicarse entre sí.
Dataset	Conjunto de datos lógicos, representando un conjunto específico de datos tabulares o jerárquicos.
Dataflow	Flujo de datos de extracción, transformación y carga (ETL) que se ejecuta en la nube de Azure para preparar los datos para su análisis.
Pipeline	Flujo de trabajo que organiza y ejecuta actividades de procesamiento de datos de forma programática en Azure Data Factory, incluyendo los dataflows, datasets y otras tareas.

4. BIBLIOGRAFÍA

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1), 1-9. <https://doi.org/10.1038/sdata.2016.18>.

5. WEBGRAFÍA

- USG. Science for a changing world (Julio 2022). California drought, visualized with open data. https://labs.waterdata.usgs.gov/visualizations/ca_drought/index.html.
- Agromet (Julio 2022). Banco de Información de Recursos Hídricos de la Región de O'Higgins. <http://www.birh.cl/mapas/agromet/>.



— FONDO DE AGUA —
SANTIAGO - MAIPO



- SANKEY DIAGRAMS (Julio 2022). <https://www.sankey-diagrams.com/tag/water/>.
- Sistema Nacional de Información Ambiental (Julio 2022), Ministerio Medio Ambiente. <https://sinia.mma.gob.cl>.
- Water Data Visualization for Stakeholder Engagement and Improved Decision Making (Julio 2022). https://colloque.inrae.fr/anaee-conference/content/download/3632/37416/version/1/file/Sébastien_Soleille_AnaEE_March_2016.pdf.
- Storage Capacity and Moving Water North and South of Lake Okeechobee 2021-2022 (Julio 2022). <https://sfwmd.maps.arcgis.com/apps/MapTour/index.html?appid=cca90b996fbc45ea9b9c39539738f9a2&webmap=96cc4a7b87f74d35a6a9dcd5c012ef5f>.
- USG. Science for a changing world (Julio 2022). U.S. Water Use from 1950-2015 (Julio 2022). <https://labs.waterdata.usgs.gov/visualizations/water-use/index.html>.
- USG. Science for a changing world (Julio 2022). <https://www.usgs.gov/media/images/may-2022-streamflow>.
- USG. Science for a changing world (Julio 2022). From Snow to Flow. https://labs.waterdata.usgs.gov/visualizations/snow-to-flow/index.html#.
- USG. Science for a changing world (Julio 2022). National Water Information System: Web Interface. https://waterdata.usgs.gov/nwis/water_use?wu_year=2015&wu_category=l&submitted_form=introduction&wu_category_nms=Irrigation%2C+in+Bgal%2Fd