

## Proyecto 1: Detección de nodos con degradación y agrupación de estos elementos según tiempo y área

El objetivo de este proyecto es detectar automáticamente y en tiempo real (o lo más cercano posible) los recursos con incidencias de red y correlacionar estos elementos en tiempo y espacio para administrar estas incidencias de manera eficiente. Actualmente, se proyecta un retraso de dos horas como lo más "real time" que es posible con los datos disponibles

### Consideraciones a priori sobre las incidencias

Antes de abordar la detección de incidencias es importante definir que es considerada un incidencia y como se le asigna alguna ponderación de gravedad

### Sobre la base que guiará este desarrollo

#### Enunciados

*La idea se planteará con enunciados*

Al contar con enunciados claros y aceptados, se reduce la posibilidad de malinterpretaciones o confusiones en el proceso de detección y se asegura que los criterios de evaluación sean uniformes y estén basados en un acuerdo previo. Por lo tanto, es importante que cualquier proceso de detección de incidencias cuente con un enunciado claramente definido y aceptado por todos los involucrados.

Entendiendo esto se hace evidente la necesidad que exista esta declaración sobre lo que es una incidencia y que sea aceptado por todos

Al explorar el concepto de incidencias deberemos descubrir que cosas necesitamos investigar de ella y descubriremos las variables que la explican

En pocas palabras, para tener éxito en una investigación o en la detección de incidencias es fundamental contar con enunciados claros y aceptados por todos los involucrados. Estos enunciados permiten definir de manera precisa el objeto de estudio, las variables a considerar y establecer criterios de evaluación uniformes, reduciendo así la posibilidad de malinterpretaciones o confusiones. Es

importante tener en cuenta que todos deben estar de acuerdo con estos enunciados para evitar desacuerdos o problemas futuros. Así que, se manifiesta la importancia de establecer estos y que sean claros y aceptados para garantizar el éxito en tus investigaciones o detección de incidencias.

## Enunciado de Incidencias Nodos

La incidencia de Nodo se refiere al desvío en el comportamiento habitual de algún indicador de cada nodo de la red (Degradación de Nodo) o que el nodo pase a un estado Fuera de Servicio (desaparición del Nodo)

Sobre la valoración de la incidencia de NODO, esta puede ser solamente positiva o negativa respecto de la experiencia de red de nuestros usuarios, en cuanto a su capacidad de Navegar (Navega Lento –No Navega)

Para la valoración de si la incidencia de NODO fue positiva o negativa es necesario diagnosticar en base al conjunto de las variables del Nodo se hayan detectado desviado de su comportamiento

Además de la evaluación de cada nodo en particular se debe considerar que la red móvil es un sistema que la afectación de uno de sus componentes altera el comportamiento de su entorno

Esto exige incluir en la evaluación de las incidencias la su dimensión zonal. Esta será política (Comuna-Región –País) y, aunque no se desarrollará en este proyecto también puede ser lógica (LAC,TAC,RA)

Con lo anterior es importante entender que las incidencias de nodos es importantísima para el análisis del clasificación del tipo de incidencia que se observa en cada nodo (Fuera de servicio -Pérdida cobertura – Pérdida de transporte)

Pero también es importante el análisis de como este nodo afecta su entorno y descubrir los recursos que fueron afectados de manera indirecta

Entonces es fácil entender que los nodos realmente involucrados en una incidencia tener dos dimensiones principales: Nodo en incidencia y Nodo afectado por incidencia (Afectación Directa – Afectación Indirecta)

Cuando un Nodo esta en incidencia puede afectar tanto a nodos WOM o ROAM. Esta clasificación también proporciona información útil al negocio

## RESPECTO DE LAS VARIABLES

### De la Identidad de los Nodos

Los nodos se presentaran en esta solución respecto de su nombre y su área lógico o política (comuna, región, país)

Esta información es útil para hacer los análisis individuales de los nodos y los análisis de entorno de los nodos

### Respecto del comportamiento de los Nodos

**Numero de celdas disponibles:** esta variable es importante porque afecta directamente la capacidad de la red para manejar la carga de los usuarios. Si el número de celdas disponibles es bajo, puede provocar una congestión en la red, lo que a su vez puede resultar en una disminución en la calidad del servicio o una interrupción total.

**Número de usuarios únicos atendidos:** esta variable permite monitorear la cantidad de usuarios que están siendo atendidos por cada nodo en particular. Si el número de usuarios atendidos es alto, es posible que el nodo esté experimentando una sobrecarga, lo que puede afectar la calidad del servicio.

**Volumen de tráfico del nodo:** esta variable se refiere a la cantidad de datos que se transmiten a través del nodo en un período de tiempo determinado. Si el volumen de tráfico es alto, puede indicar que el nodo está experimentando una gran demanda de datos, lo que puede afectar la velocidad y calidad del servicio.

La elección de estas variables está orientada a detectar si estamos ante una incidencia provocada por los usuarios, por la cobertura o por el backhaul

De los cómo se detectará estas categorías de incidencias, esto se atenderá más adelante en detalle, pero es importante hacer una pausa en este punto y atender lo siguiente

De la detección de problemas de backhaul usando este método, solo se espera descubrir si el nodo tiene un problema relacionado a su capacidad de transporte

No se resuelve la identificación de las dependencias de los nodos en el ámbito de transporte. Para esto se desarrollará otro proyecto, cuando este sea implementado y que será complementario donde se atenderá el análisis de dependencia en la red de transporte de los nodos para detectar la fuente de origen del problema o si la incidencia no guarda relación de dependencia de transporte

Esta idea se presenta para poder enmarcar los límites de conocimiento que este proyecto puede entregar y poner de manifiesto de la importancia de incluir en un futuro sprint esta consideración que ayudará que los resultados sean de mayor impacto en cuanto a su utilidad

## DE LA ADMINISTRACION DE LAS INCIDENCIAS

Una Vez que hemos detectado los nodos en incidencias, sabemos si fueron afectados directamente o indirectamente, cuantificamos su pérdida y diagnosticamos si es un problema de fuera de servicio, cobertura o transporte, corresponde atender lo relativo de administración de la incidencia

Aquí corresponde encontrar correlaciones temporales y asociaciones que nos indique que cierto grupo de nodos están comprometidos en la misma incidencia y poder entregar a este grupo una identidad de incidencia de red formalmente

Además esta parte del desarrollo es la encarga de la valoración de gravedad de la incidencia para que a la hora de la asignación de los tickets de control estos reflejen de buena manera incidencia

### Principios de la administración de incidencias

Una incidencia es la detección de 2 o más nodos que hayan sido diagnosticados por algún tipo de afectación en particular (fuera de servicio-cobertura-transporte) y que sus horas de inicio de incidencia sea el mismo.

Dependiendo el tipo de afectación los nodos podrán recuperar su comportamiento habitual (problemas de transporte) de manera simultánea o por bloques de simultaneidad, mientras otras veces la recuperación ocurre de manera particular para cada Nodo (problemas de cobertura)

Cuando los nodos en la incidencia están la misma comuna su administración es fácil. Todos los nodos compartirán el ID de la incidencia, pero cuando el número de nodos en la incidencia es muy grande, genera un clúster lógico que contiene todos los nodos en incidencia. Estos clúster lógicos pueden contener varias comunas o varias regiones

¿Cuál es el tamaño de la incidencia en términos geográfico para asignar el ID?

Podemos asumir que el tamaño del clúster es el tamaño verdadero de la incidencia y que para su administración, todas las comunas y por ende los nodos están en su interior deben estar asociado al mismo ID

El objetivo de la Valoración de la incidencia en cuanto al nivel de gravedad es útil poder comunicar de manera eficiente como nuestros usuarios están siendo afectados por el comportamiento anormal de la red

Las valoraciones actuales P1, P2, P3 son valoraciones que principalmente evalúa la incidencia en cuanto a la disponibilidad de los nodos y otras consideraciones más. Esta valoración iguala la importancia de los nodos en un área y no refleja el peso de los nodos en cuanto a su relevancia

Por ejemplo, es diferente informar para la valoración lo siguiente:

- Existen 3 nodos que están fuera de servicio pero existen nodos roam que soportan el servicio degradado en la zona

A informar

- Existen 3 nodos fuera de servicio, no se observa pérdida de usuarios conectados y el tráfico perdido es de 20G correspondiente al 30% del tráfico habitual. Se diagnostica problema raíz es de transporte

En el ámbito zonal, el procedimiento actual de las incidencias adquieren una asignación comunal, pero se ha detectado que cuando se revisa los elementos de la comuna declarada en los tickets, están incluyen nodos de otras comunas.

Independiente de si hay una lógica detrás que explique por qué ocurre esto, es evidente que es confuso y para quien no este interiorizado de la lógica subyacente que explica esto, lo induce a error o simplemente no lo entiende

El desafío que se asume en este aspecto es hacer un sistema de asignación de ID de incidencia que tenga una lógica que pueda ser entendida por cualquiera sin necesidad de una explicación

Volvamos al tema del clúster que se genera cuando se hace la detección de nodos con incidencia simultáneas

El ID de la incidencia es en relación al clúster que generen todos los nodos involucrados en la incidencia, por lo tanto todas las comunas y por consecuencia todos los nodos de todas esas comunas tendrán el mismo valor de ID

Pero ¿Qué pasa con la gravedad?

Para la observación de la gravedad he resuelto el problema de la siguiente manera

La gravedad debe ser referenciada al respecto del tráfico fundamentalmente, desde que es la dimensión que refleja y que da peso a la relación de Usuarios – red

- Suponga que se pierden 5 Nodos en una incidencia y se pierden 7% usuarios (No Navega) y se tiene una caída de 10%
- Ahora tenemos una incidencia donde no se perdió usuarios ni nodos, pero el tráfico de la comuna cayó 20%
- Otro caso es uno igual al anterior pero con pérdidas de usuarios y Nodos o celdas

Anteriormente dije que el tráfico es fundamental, pero los ejemplos expuestos muestran que también debemos considerar las otras variables de este proyecto

En los casos mostrados vemos que el menos grave, mientras el segundo y el tercero solo podríamos decir que son iguales de graves, solo cambian su motivos (el segundo aparenta ser un problema de transporte mientras en el tercero aparenta ser por cobertura). También existe la posibilidad que los motivos de la incidencia sea una mezcla entre transporte y cobertura

Este tema se abordará con mayor cuidado cuando sea atendida su lógica en detalle, estas apreciaciones solo están orientadas a que se entienda la lógica que planteo al respecto de este punto

## Variables de entidad

Sus valores son estáticos durante todo el día

- Celdas
- Nodos

## Variable de comportamiento

Sus valores son varían horariamente durante todo el día

- Volumen de tráfico
- Usuarios

Con esto propongo que la gravedad sea evaluada considerando las posibles combinaciones de estas categorías.

Las combinaciones de entidades están determinadas por los siguientes comportamientos

- Con/Sin pérdida de usuarios
- Con/Sin pérdida de celdas
- Con/Sin pérdidas de Nodos
- Con/Sin pérdidas de Tráfico

Así por ejemplo, una incidencia tiene perdida de nodos o celdas, pero todos los clientes que vieron desaparecer a su mejor servidor migraron a otros nodos. El ticket debería reflejar eso.

Perdida de Nodos- Celdas en la comuna Sin pérdida de Usuarios

También se puede dar el mismo caso de pérdida de entidades de red pero que también se pierdan usuarios. Aquí el ticket debería reflejar

Perdida de Nodos- Celdas en la comuna Sin pérdida de Usuarios



En este punto cabe preguntarse. ¿Cuál de los casos anteriores es el más grave?

Aquí la observación del tráfico se vuelve esencial. Es la única variable que mide realmente la magnitud del problema al darle peso a las entidades involucradas en la incidencia ya que es la única variable de entre las utilizadas que habla de la relación usuarios - nodos

Ahora ya podemos saber por ejemplo que una incidencia tiene perdida de entidades de red (cobertura) pero una pérdida de tráfico del 10%. También puede existir una incidencia donde no se detecten pérdidas de entidades de red pero la pérdida de tráfico sea de un 25% (transporte). ¿Qué incidencia es peor suponiendo que ambas incidencias ocurren en la misma comuna en diferentes días?

Se debe tener en cuenta que en la ponderación de variables siempre hay subjetividad por lo que esta debe ser cuidadosamente acordada entre los que toman decisiones, para que la ponderación sea coherente y justa en términos de la importancia que cada variable aporta a la gravedad de la incidencia

Hasta aquí propongo las consideraciones que son necesarias atender para valorizar la gravedad de las incidencias, pero el determinar cómo se valorarán debe ser discutido con los encargados de mantener la red para ajustar la valoración de las incidencias según sus intereses

Aunque no es un asunto que se atenderá en esta etapa del desarrollo, es importante notar que se hace evidente que para una gestión aún más eficiente de las incidencias, se deber agregar la información de las dependencias de los nodos en el BackHaul.

Con esta información no solo sabremos la magnitud de la incidencia sino que sabremos donde está el nodo raíz de la problemática para incidencias de transporte y así poder enfocar los esfuerzos donde corresponde de forma temprana

Actualmente no se dispone de esta información y es difícil proponer una lógica, pero se estima que poseyendo esa información, con una exploración de cómo está construida esta tabla se puede, sin inconvenientes crear un procedimiento de sea fácil de implementar y de acoplar a esta solución, lo que permitirá la concentración de esfuerzos de manera eficiente por parte de WOM en la normalización de sus servicio con la información que provea este sistema

## Sobre el desarrollo de “detección de incidencias en nodos de red” basado en análisis automático y nativo en SQL

La incidencia corresponde al conjunto de nodos que se ven afectados por la misma. Esto se refiere a que existe una correlación temporal en el inicio de las incidencias particulares de cada nodo, y alguna correlación espacial comuna, comunas adyacentes, regiones, etc.

Si solo se detecta un nodo directamente afectado por incidencia, independiente de si genera impacto en otros nodos, no se considera incidencia y se considerará fallo aislado

Para la detección de incidencias de nodos (la base de este desarrollo) se espera que en cada actualización de los datos, todos los elementos de red encontrados en los ellos serán sometidos a análisis, por lo que la detección de nuevo nodos será automática

La solución se desarrolló nativamente en SQL como una forma de reducir la complejidad del sistema y aumentar su portabilidad. Utilizar SQL también es una buena opción para realizar el análisis de datos en grandes cantidades, como en el caso de la información de la red

## **Sobres las tablas que se crean en base a esta idea**

### **TABLA PRS\_DATA**

La tabla PRS\_DATA es la tabla que alimenta todo el sistema de este desarrollo ya que contiene los datos de red provenientes desde el PRS y se actualiza cada hora, será tabla se llamará " PRS\_DATA " y contendrá información como la fecha y hora, nombre de los nodos, comuna del nodo, región del nodo, número de celdas en el nodo, número de usuarios atendidos y tráfico en GB

Los Datos provenientes desde el PRS presentan un desafío que debe ser resuelto antes de continuar con los procesos de análisis

Este desafío es respecto de los nodos que dejan de emitir datos al OSS (recursos que desaparecen) por lo tanto en el análisis de su línea temporal existen vacíos

de tiempo que afectan los procesos posteriores. Para corregir esto y hacer que el nodo aparezca incluso en los momentos en donde pierde comunicación con el OSS, se creará una tabla que será la base para los procesos posteriores. Esta tabla se llamará DATA\_PRS\_NORMALIZADA y contendrá los nodos en el tiempo, si vacíos temporales. En caso que el nodo no reciba información de OSS en esta tabla para las variables de número de usuarios atendidos y volumen de tráfico se valorizará como 0. Esta tabla contendrá los mismos campos de DATA\_PRS pero además un campo que indicara el estado del nodo. Si el tráfico y el número de abondos es cero, el nodo se marca Fuera de Servicio, en caso contrario en servicio

Con esto se simplifica la detección de nodos fuera de servicio y se puede pasar directamente a la etapa de predicción de variables de la tabla PRS-DATA

## **TABLA PRS\_DATA\_NORMALIZADA**

El objetivo principal de este conjunto de queries es mantener actualizada la tabla PRS\_DATA\_NORMALIZADA con los datos más recientes de la tabla PRS\_DATA, incluso cuando haya nodos que dejen de emitir datos desde el OSS.

La tabla PRS\_DATA\_NORMALIZADA\_NODO contendrá información de fecha y hora, nombre de nodos, número de celdas, número de usuarios atendidos y tráfico en GB. Además, se resolverá el problema de vacíos temporales que pueden ocurrir en PRS\_DATA cuando los nodos dejan de emitir datos, mediante la valorización en 0 de los campos de número de usuarios atendidos y volumen de tráfico en las fechas y horas faltantes.

De la tabla PRS\_DATA\_NORMALIZADA\_NODO, que no olvidemos siempre se actualizara a cada hora, se alimentaran otras tablas convenientemente creadas pata este desarrollo

- PRS\_DATA\_NORMALIZADA\_COMUNALES
- PRS\_DATA\_NORMALIZADA\_REGIONALES
- PRS\_DATA\_NORMALIZADA\_NACIONAL

Para estas tablas se harán agrupaciones de los nodos por su comuna, region y país para cada hora de registro. La agrupación será la suma de las variables de tráfico y usuarios. Además estas tablas incluirán el conteo de los nodos y celdas vistas en los últimos registros

En resumen, todas las tablas PRS\_DATA acumularan el comportamiento histórico de la red y serán actualizadas con los datos más actualizado que se dispongan. La retención de esta data al ser horaria será evaluada en el modo de prueba de la maqueta.

Estas tablas son consideradas las tablas que contienen el comportamiento real de las entidades observadas

## QUERY TABLA PRS\_DATA\_NORMALIZADA\_NODO

La primera query detecta si existen registros nuevos en la tabla PRS\_DATA que aún no han sido cargados en la tabla PRS\_DATA\_NORMALIZADA, y en caso de

```
INSERT INTO PRS_DATA_NORMALIZADA (fecha_hora, nombre_nodo, n_celdas, n_usuarios_atendidos,
volumen_trafico_GB, estado_nodo)
SELECT a.fecha_hora, b.nombre_nodo, b.n_celdas, COALESCE(a.n_usuarios_atendidos, 0) AS n_usuarios_atendidos,
      COALESCE(a.volumen_trafico_GB, 0) AS volumen_trafico_GB,
      CASE WHEN a.n_usuarios_atendidos IS NULL AND a.volumen_trafico_GB IS NULL THEN 'Fuera de Servicio' ELSE 'En
Servicio' END AS estado_nodo
FROM (
  SELECT fecha_hora, nombre_nodo, n_celdas
  FROM (
    SELECT DISTINCT fecha_hora
    FROM PRS_DATA
  ) a
  CROSS JOIN (
    SELECT DISTINCT nombre_nodo, n_celdas
    FROM PRS_DATA
  ) b
  LEFT JOIN PRS_DATA_NORMALIZADA c ON a.fecha_hora = c.fecha_hora AND b.nombre_nodo = c.nombre_nodo
AND b.n_celdas = c.n_celdas
  WHERE c.fecha_hora IS NULL AND c.nombre_nodo IS NULL AND c.n_celdas IS NULL
) c
LEFT JOIN PRS_DATA a ON c.fecha_hora = a.fecha_hora AND c.nombre_nodo = a.nombre_nodo AND c.n_celdas =
a.n_celdas;
```

existir, procede a insertarlos y normalizarlos.

La segunda query detecta si existen nodos que no tienen registros en la tabla PRS\_DATA para ciertas horas y fechas, y en caso de ser así, procede a insertar registros con valor 0 para las variables de número de usuarios atendidos y volumen de tráfico en esas fechas y horas.

```
INSERT INTO PRS_DATA_NORMALIZADA (fecha_hora, nombre_nodo, n_celdas, n_usuarios_atendidos,
volumen_trafico_GB, estado_nodo)
SELECT a.fecha_hora, b.nombre_nodo, b.n_celdas, 0 AS n_usuarios_atendidos, 0 AS volumen_trafico_GB, 'Fuera de
Servicio' AS estado_nodo
FROM (
    SELECT DISTINCT fecha_hora
    FROM PRS_DATA
) a
CROSS JOIN (
    SELECT DISTINCT nombre_nodo, n_celdas
    FROM PRS_DATA
) b
LEFT JOIN PRS_DATA_NORMALIZADA c
```

Estos dos procesos tienen como objetivo asegurar que la tabla PRS\_DATA\_NORMALIZADA contenga todos los nodos en todas las horas, independientemente de si existen o no registros en la tabla PRS\_DATA.

En conjunto, estos procesos garantizan la disponibilidad y calidad de las informaciones necesarias para realizar los siguientes procesos de análisis y predicción.

Esta idea es de gran utilidad ya que simplificará la tarea de la detección de nodos fuera de servicio sin tener que crear tablas adicionales y cruce entre tablas en etapas posteriores de este desarrollo

Una vez finalizada estos procedimientos los datos están disponibles para el siguiente stage

QUERY TABLA PRS\_DATA\_NORMALIZADA\_GEO

Siguiendo la misma lógica anterior aquí se presenta la query para la consolidación de las tablas de predicción a nivel comunal y regional

```
-- Insertar datos en tabla PRS_DATA_COMUNALES
INSERT INTO PRS_DATA_COMUNALES (fecha_hora, comuna, numero_celdas, numero_usuarios, volumen_tráfico, numero_nodos,
numero_celdas_vistas)
SELECT
fecha_hora,
comuna,
SUM(numero_celdas) as numero_celdas,
SUM(numero_usuarios) as numero_usuarios,
SUM(volumen_tráfico) as volumen_tráfico,
COUNT(DISTINCT nodo) as numero_nodos,
FROM PRS_DATA
GROUP BY fecha_hora, comuna;

-- Insertar datos en tabla PRS_DATA_REGIONALES
INSERT INTO PRS_DATA_REGIONALES (fecha_hora, region, numero_celdas, numero_usuarios, volumen_tráfico, numero_nodos,
numero_celdas_vistas)
SELECT
fecha_hora,
region,
SUM(numero_celdas) as numero_celdas,
SUM(numero_usuarios) as numero_usuarios,
SUM(volumen_tráfico) as volumen_tráfico,
COUNT(DISTINCT nodo) as numero_nodos,
FROM PRS_DATA
GROUP BY fecha_hora, region;

-- Insertar datos en tabla PRS_DATA_NACIONAL
INSERT INTO PRS_DATA_NACIONAL (fecha_hora, numero_celdas, numero_usuarios, volumen_tráfico, numero_nodos,
numero_celdas_vistas)
SELECT
fecha_hora,
SUM(numero_celdas) as numero_celdas,
SUM(numero_usuarios) as numero_usuarios,
SUM(volumen_tráfico) as volumen_tráfico,
COUNT(DISTINCT nodo) as numero_nodos,
FROM PRS_DATA
GROUP BY fecha_hora;
Approach
```



## **STAGE 2**

### **TABLAS REDICCION\_DATA\_NORMALIZADAS**

#### **TABLA PREDICCION\_DATA\_NORMALIZADA\_NODO**

La tabla PREDICCION\_DATA\_NORMALIZADA\_NODO se utilizará para almacenar las predicciones de las variables de tráfico y usuarios de cada nodo en la red en el tiempo.

Las predicciones de las variables de comportamiento se harán a las 23:30 de cada día para todas las horas del siguiente día solamente

Esto tendrá utilidad a la hora de la visualización de los datos. Imagine una gráfica donde a cada comienzo de día ya se pueda visualizar el comportamiento esperado de toda la hora de ese día. Luego a medida que pasa el día en la misma gráfica irá apareciendo el comportamiento real de nodo ¿no sería una gráfica interesante?

La predicción se basará en el análisis de series temporales, utilizando datos de los últimos 30 días menos 1 día

Si la predicción se realiza para un día de semana, entonces solo se consideraran los días de semana dentro de la ventana de tiempo de los datos para la predicción. Si el día para el que se realiza la predicción es fin de semana de la misma ventana solo se consideraran los días fin de semana.

Los datos para el cálculo de la predicción provienen de la tabla PRS\_DATA\_NORMALIZADA

Los resultados que se obtengan deberán ser almacenados en la tabla PREDICCION\_DATA\_NORMALIZADA\_NODO de manera horaria y continua en el tiempo

## PREDICCION\_DATA\_NORMALIZADA\_NODO

Esta tabla contendrá los siguientes campos:

- fecha\_hora

--- Datos de la entidad de red

- Nombre del nodo
- Comuna
- Región

-- - Administración de incidencia

- Estado de servicio (en servicio –Fuera de servicio)
- Diagnostico falla (cobertura – transporte)
- Tipo de afectación (directa – indirecta)
- **Tipo Nodo en Transporte (Maestro - Esclavo) – No considerado ahora**
- **Tipo de Nodo Maestro (Nodo origen – en cascada) – No considerado ahora**

--- Valores de predicción

- Max capacidad Celdas Vistas en Nodo
- Prediccion\_número de usuarios
- Prediccion\_volumen de tráfico
- MAX\_predicción\_número de usuarios
- MIN\_predicción\_número de usuarios
- MAX\_predicción\_volumen de tráfico
- MIN\_predicción\_volumen de tráfico

--Valores de comportamiento Actual

- tráfico actual
- usuarios actuales
- Numero de celdas actual

--- Valores de Variación

- volumen de usuarios perdido
- volumen de tráfico perdido
- volumen celdas perdidas
- % volumen de usuarios perdido
- % volumen de tráfico perdido

--- Ranking de nodo

- Ranking de tráfico en comuna (ejemplo 13 de 40)

- Aporte % tráfico en comuna (ejemplo 5% )

Cada “—” marca un ítem que se refiere al grupo de campos de la tabla según su naturaleza de uso tienen la misma clasificación o “ITEM”

La actualización de los datos en esta tabla no se realizará en una sola vez,

Este enfoque de etapas consecutivas permite manejar mejor la complejidad del proceso y asegurarse de que cada etapa se complete de manera adecuada antes de pasar a la siguiente.

También esto hace que el proceso sea más escalable y adaptable a medida que cambian las necesidades del proyecto.

## ANALISIS DE SERIES TEMPORALES DE TRÁFICO Y USUARIOS

La primera etapa será el cálculo de predicciones de Tráfico y usuarios.

Para que cada nuevo día cuente con su predicción para cada hora de su día antes que ese día llegue, a las 23:30 del día presente se realizara la predicción solamente para el día siguiente.

Esta estrategia de predicción en tiempo real permitirá que se puedan visualizar los datos de manera oportuna y tomar decisiones basadas en la información más actualizada.

Esta query realiza un cálculo de predicción de tráfico y usuarios para cada hora del siguiente día, utilizando análisis de series temporales en una tabla de datos normalizados de nodos de red. Se obtienen las muestras de los últimos 30 días (menos el día anterior) y se calcula la predicción para el siguiente día de la ejecución de la query. Además, se considera si el día siguiente es un día de semana o fin de semana para seleccionar las muestras correspondientes. El resultado es un promedio de las predicciones de tráfico y usuarios para cada hora de cada nodo de la red. La query es apta para ser ejecutada en BigQuery.

## CALCULO DE CURVAS MAX Y MIN DE TRÁFICO Y USUARIOS

Lo que hace esta consulta es actualizar la tabla PREDICCION\_DATA\_NORMALIZADA\_NODO para agregar los valores máximos y mínimos de las predicciones de tráfico y usuarios para cada nodo en particular. Para ello, se utiliza una subconsulta que calcula las distancias máximas entre las predicciones de tráfico y usuarios para cada nodo. Luego, se une esta subconsulta con la tabla PREDICCION\_DATA\_NORMALIZADA\_NODO a través del nombre del nodo, y se actualizan los campos MAX\_prediccion\_numero\_de\_usuarios, MIN\_prediccion\_numero\_de\_usuarios, MAX\_prediccion\_volumen\_de\_trafico y MIN\_prediccion\_volumen\_de\_trafico para cada registro en la tabla.

```
UPDATE PREDICCION_DATA_NORMALIZADA_NODO AS A
LEFT JOIN (
SELECT
  MAX(Prediccion_numero_usuarios) - (MAX(Prediccion_numero_usuarios)/100*90) AS distancia_usuarios,
  MAX(Prediccion_volumen_trafico) - (MAX(Prediccion_volumen_trafico)/100*90) AS distancia_trafico,
  Nombre_nodo
FROM PREDICCION_DATA_NORMALIZADA_NODO
WHERE fecha_hora >= DATE_TRUNC('day', CURRENT_DATE()) + INTERVAL 1 DAY
  AND fecha_hora < DATE_TRUNC('day', CURRENT_DATE()) + INTERVAL 2 DAY
GROUP BY Nombre_nodo
) AS B ON A.Nombre_nodo = B.Nombre_nodo
SET A.MAX_prediccion_numero_de_usuarios = A.Prediccion_numero_usuarios + b.distancia_usuarios,
  A.MIN_prediccion_numero_de_usuarios = A.Prediccion_numero_usuarios - b.distancia_usuarios,
  A.MAX_prediccion_volumen_de_trafico = A.prediccion_volumen_de_trafico + b.distancia_usuarios,
  A.MIN_prediccion_volumen_de_trafico = A.prediccion_volumen_de_trafico - b.distancia_usuarios
WHERE A.fecha_hora >= DATE_TRUNC('day', CURRENT_DATE()) + INTERVAL 1 DAY
  AND A.fecha_hora < DATE_TRUNC('day', CURRENT_DATE()) + INTERVAL 2 DAY;
```

## Detección Max capacidad Celdas Vistas en Nodo

Como se mencionó ya en este documento, la tabla max\_capacidad\_celdas\_por\_nodo se actualizará periódicamente para buscar si existen nuevos máximos en la cantidad de celdas vistas por cada nodo. Esta tabla es esencial en el proceso de predicción de nodos, por lo que se debe asegurar que siempre tenga la información más actualizada.

La actualización se realiza mediante la query UPDATE que busca el nuevo máximo en la tabla PRS\_NORMALIZADA\_NODO para cada nodo y periodo de predicción. Si se encuentra un nuevo máximo, se actualiza el campo num\_celdas en la tabla max\_capacidad\_celdas\_por\_nodo. La query se ejecuta antes de pasar los datos a la tabla de predicción de nodos para asegurar que

```
SELECT
DATETIME(DATE_ADD(DATE_TRUNC(DATE_ADD(CURRENT_DATE(), INTERVAL 1 DAY), DAY), INTERVAL t.n HOUR)) AS fecha_hora,
prn.nombre_nodo AS Nombre_nodo,
prn.comuna AS Comuna,
prn.region AS Región,
AVG(prn.num_usuarios) AS Prediccion_num_usuarios,
AVG(prn.volumen_trafico) AS Prediccion_volumen_trafico
FROM
PRS_DATA_NORMALIZADA_NODO prn
CROSS JOIN
UNNEST(GENERATE_ARRAY(0, 23)) AS t
WHERE
prn.fecha_hora >= DATETIME(DATE_ADD(CURRENT_DATE(), INTERVAL -31 DAY))
AND prn.fecha_hora < DATETIME(DATE_ADD(CURRENT_DATE(), INTERVAL -1 DAY))
AND EXTRACT(DAYOFWEEK FROM DATETIME(DATE_ADD(CURRENT_DATE(), INTERVAL 1 DAY))) = 7
AND EXTRACT(DAYOFWEEK FROM prn.fecha_hora) IN (2, 3, 4, 5, 6)
GROUP BY
prn.nombre_nodo, prn.comuna, prn.region, DATETIME(DATE_ADD(DATE_TRUNC(DATE_ADD(CURRENT_DATE(), INTERVAL 1 DAY),
DAY), INTERVAL t.n HOUR))
ORDER BY
fecha_hora ASC;
```

se esté utilizando la información más reciente

```
UPDATE MAX_CAPACIDAD_CELDAS_POR_NODO AS mc
SET mc.num_celdas = (
  SELECT MAX(num_celdas_vistas)
  FROM PRS_NORMALIZADA_NODO
  WHERE nombre_nodo = mc.nombre_nodo
    AND fecha_hora >= (SELECT MAX(fecha_hora) FROM PRS_NORMALIZADA_NODO) - INTERVAL 1 DAY
)
WHERE EXISTS (
  SELECT 1
  FROM PRS_NORMALIZADA_NODO
  WHERE nombre_nodo = mc.nombre_nodo
    AND fecha_hora >= (SELECT MAX(fecha_hora) FROM PRS_NORMALIZADA_NODO) - INTERVAL 1 DAY
    AND num_celdas_vistas > mc.num_celdas
);
```

Una vez actualizada esta tabla podemos incluir el valor de la prediccion de celdas por nodo en la tabla PREDICCION\_DATA\_NORMALIZADA\_NODO

## Query para actualización en tabla PREDICCION\_DATA\_NORMALIZADA\_NODO

```
UPDATE PREDICCION_DATA_NORMALIZADA_NODO AS pdn
SET pdn.Max_capacidad_celdas_vistas = mcn.num_celdas
FROM (
  SELECT nombre_nodo, MAX(num_celdas) as num_celdas
  FROM MAX_CAPACIDAD_CELDAS_POR_NODO
  GROUP BY nombre_nodo
) AS mcn
WHERE pdn.nombre_nodo = mcn.nombre_nodo
  AND pdn.fecha_hora >= TIMESTAMP(DATE_ADD(DATE_TRUNC(DATE_ADD(CURRENT_DATE(), INTERVAL 1 DAY), DAY), INTERVAL
0 HOUR))
  AND pdn.fecha_hora < TIMESTAMP(DATE_ADD(DATE_TRUNC(DATE_ADD(CURRENT_DATE(), INTERVAL 2 DAY), DAY), INTERVAL 0
HOUR));
```

Esta consulta actualiza la columna Max\_capacidad\_celdas\_vistas de la tabla PREDICCION\_DATA\_NORMALIZADA\_NODO con los valores correspondientes a la tabla MAX\_CAPACIDAD\_CELDAS\_POR\_NODO, para los registros de la predicción del día siguiente, es decir, para las horas del siguiente día. Se utiliza una subconsulta para obtener el valor máximo de la capacidad de celdas para cada nodo en la tabla MAX\_CAPACIDAD\_CELDAS\_POR\_NODO, y se hace un JOIN con la tabla de predicción para actualizar los registros correspondientes.

Hasta este punto, hemos creado una tabla de predicción y actualizado sus campos según el ítem "Valores de predicción". Ahora es necesario considerar la llegada de datos reales y utilizarlos para actualizar la tabla de predicción de nodos. Para ello, crearemos consultas de actualización horaria de los datos reales en la tabla de predicción de nodos.

## UPDATE DE LOS CAMPOS

tráfico actual - usuarios actuales - Numero de celdas actual

```
UPDATE PREDICCION_DATA_NORMALIZADA_NODO AS pdn
SET
  pdn.trafico_actual = IFNULL(prn.volumen_trafico, pdn.trafico_actual),
  pdn.usuarios_actuales = IFNULL(prn.num_usuarios, pdn.usuarios_actuales),
  pdn.num_celdas_actual = IFNULL(mcc.num_celdas, pdn.num_celdas_actual)
FROM (
  SELECT nombre_nodo, fecha_hora, num_usuarios, volumen_trafico, num_celdas
  FROM PRS_NORMALIZADA_NODO
  WHERE fecha_hora = (
    SELECT MAX(fecha_hora)
    FROM PRS_NORMALIZADA_NODO
    WHERE nombre_nodo = pdn.nombre_nodo AND fecha_hora <= pdn.fecha_hora
  )
) AS prn
LEFT JOIN MAX_CAPACIDAD_CELDAS_POR_NODO AS mcc
ON prn.nombre_nodo = mcc.nombre_nodo
WHERE pdn.fecha_hora = (
  SELECT MAX(fecha_hora)
  FROM PREDICCION_DATA_NORMALIZADA_NODO
  WHERE nombre_nodo = pdn.nombre_nodo AND fecha_hora <= pdn.fecha_hora
) AND NOT EXISTS (
  SELECT 1
  FROM PREDICCION_DATA_NORMALIZADA_NODO
  WHERE nombre_nodo = pdn.nombre_nodo AND fecha_hora = pdn.fecha_hora
); FROM PREDICCION_DATA_NORMALIZADA_NODO
WHERE nombre_nodo = pdn.nombre_nodo
) AND pdn.nombre_nodo = prn.nombre_nodo; FROM PRS_NORMALIZADA_NODO
WHERE fecha_hora <= pdn.fecha_hora
) AND pdn.nombre_nodo = prn.nombre_nodo;
```

Esta query actualiza los campos de tráfico actual, usuarios actuales y número de celdas actual en la tabla PREDICCION\_DATA\_NORMALIZADA\_NODO con los valores correspondientes de la tabla PRS\_NORMALIZADA\_NODO y la tabla MAX\_CAPACIDAD\_CELDAS\_POR\_NODO, respectivamente.

La subconsulta prn busca los últimos valores de tráfico, usuarios y número de celdas en la tabla PRS\_NORMALIZADA\_NODO para cada nodo y hora. La



subconsulta mcc busca el número de celdas máximo para cada nodo en la tabla MAX\_CAPACIDAD\_CELDAS\_POR\_NODO.

La cláusula LEFT JOIN une las subconsultas prn y mcc a la tabla PREDICCION\_DATA\_NORMALIZADA\_NODO. La cláusula WHERE filtra los registros de la tabla PREDICCION\_DATA\_NORMALIZADA\_NODO para actualizar solamente los valores correspondientes a la hora más reciente para la cual existen datos en la tabla PRS\_NORMALIZADA\_NODO

## EVALUACION DE SI NODO ESTA EN INCIDENCIA

En este punto cuando se insertaron los valores más recientes de la red podemos atender los otros ítems de la tabla de predicción de nodos

Comenzaremos con la evaluación de los registros más actuales de la tabla

¿Será que los valores más recientes están fuera de los rangos Mac y Min aceptados?

Estas preguntas pertenecen al ítem “Valores de Variación” y posee 6 campos donde se guardan las respuestas a esta consulta

Esta query actualiza la tabla de predicción de nodos, específicamente los campos relacionados a la cantidad de tráfico, usuarios y celdas perdidas, y sus porcentajes correspondientes. Para esto, se compara la cantidad de tráfico y usuarios actuales con las predicciones de volumen de tráfico y número de usuarios, respectivamente. Si la cantidad actual es mayor que la predicción, entonces se calcula la cantidad perdida y se actualiza el campo correspondiente; en caso contrario, se establece el valor en cero. Para el volumen de celdas perdidas, se utiliza la diferencia entre la capacidad máxima de celdas y el número de celdas actuales. Finalmente, se calculan los porcentajes correspondientes y se actualizan los campos en la tabla de

```
UPDATE `PREDICCION_DATA_NORMALIZADA_NODO` AS pdn
SET
  pdn.volumen_usuarios_perdido = IF(pdn.prediccion_numero_de_usuarios > pdn.usuarios_actuales,
pdn.prediccion_numero_de_usuarios - pdn.usuarios_actuales, 0),
  pdn.volumen_trafico_perdido = IF(pdn.prediccion_volumen_de_trafico > pdn.trafico_actual,
pdn.prediccion_volumen_de_trafico - pdn.trafico_actual, 0),
  pdn.volumen_celdas_perdido = IFNULL(mcc.num_celdas, 0) - IFNULL(pdn.num_celdas_actual, 0),
  pdn.porcentaje_volumen_usuarios_perdido = IF(pdn.prediccion_numero_de_usuarios > 0, pdn.volumen_usuarios_perdido /
pdn.prediccion_numero_de_usuarios, 0),
  pdn.porcentaje_volumen_trafico_perdido = IF(pdn.prediccion_volumen_de_trafico > 0, pdn.volumen_trafico_perdido /
pdn.prediccion_volumen_de_trafico, 0),
  pdn.porcentaje_volumen_celdas_perdido = IF(mcc.num_celdas > 0, pdn.volumen_celdas_perdido / mcc.num_celdas, 0)
FROM (
  SELECT nombre_nodo, fecha_hora, num_usuarios, volumen_trafico, num_celdas
  FROM `PRS_NORMALIZADA_NODO`
  WHERE fecha_hora = (
    SELECT MAX(fecha_hora)
    FROM `PRS_NORMALIZADA_NODO`
    WHERE nombre_nodo = pdn.nombre_nodo AND fecha_hora <= pdn.fecha_hora
  )
) AS prn
LEFT JOIN `MAX_CAPACIDAD_CELDAS_POR_NODO` AS mcc
ON prn.nombre_nodo = mcc.nombre_nodo
WHERE pdn.fecha_hora = (
  SELECT MAX(fecha_hora)
  FROM `PRS_NORMALIZADA_NODO`
  WHERE fecha_hora <= pdn.fecha_hora
) AND pdn.nombre_nodo = prn.nombre_nodo;
```

predicción de nodos.

AHORA VAMOS A EVALUAR LO RELATIVO A LA ADMINISTRACION DE LA INCIDENCIA

Recordemos que la administración en la tabla de predicción está relacionada con los siguientes campos

-- - Administración de incidencia

- Estado de servicio (en servicio –Fuera de servicio)
- Diagnostico falla (cobertura – transporte)
- Tipo de afectación (directa – indirecta)
- **Tipo Nodo en Transporte (Maestro - Esclavo) – No considerado ahora**
- **Tipo de Nodo Maestro (Nodo origen – en cascada) – No considerado ahora**

Es preciso recordar que



La tabla PRS\_DATA\_NORMALIZADA\_NODO se utilizará para hacer las predicciones y para extraer los valores reales más actuales de cada variable. Al inicio de cada día se insertará en la tabla PREDICCION\_DATA\_NORMALIZADA\_NODO los resultados de la predicción para ese día con base en los datos vistos en PRS\_DATA\_NORMALIZADA\_NODO. A medida que pase el día, para cada hora con valores predichos, se insertará en otras columnas previamente creadas (valor real de tráfico y real de usuarios) los valores más actualizados de la tabla PRS\_DATA\_NORMALIZADA\_NODO para las variables correspondientes.

La tabla PREDICCION\_DATA\_NORMALIZADA\_NODO permitirá visualizar el comportamiento predicho de los nodos para todo el día y en tiempo real se irá actualizando con los valores reales, con un desfase de dos horas debido a la temporalidad de los datos. Si los valores reales que ingresan son 0 en ambas variables (tráfico, usuarios), se escribirá "Fuera de Servicio" en la columna estado de servicio, y en "Servicio" si al menos una variable real es mayor a 0.

Una vez descrita la tabla PREDICCION\_DATA\_NORMALIZADA\_NODO y explicado las consideraciones para el cálculo de las series temporales e este proyecto, se procede a explicar en detalle los pasos requeridos para su correcto poblamiento

- Cálculo de predicciones para todas las variables en todas las horas venideras del nuevo día en cuanto a tráfico y número de usuarios (variable de comportamiento)
- Cálculo de Max y MIN para todas las variables en todas las horas venideras del nuevo día en cuanto los resultados de las predicciones del paso anterior
- Insertar valores más actuales (reales) de tráfico y usuarios
- Detección del estado de servicio de los registros más recientes
- Detección de variación de celdas (variable entidad)

En este momento la tabla estará disponible para su explotación

SOBRE LA QUERIES PARA EL PROCESO DE UPDATE DE LA TABLA A CADA HORA

PUNTO 1

Calculo de predicciones para todas las variables en todas las horas venideras del nuevo día en cuanto a tráfico y número de usuarios (enunciadas como variables de comportamiento)