LABORATORIO

Ecosistema Hadoop: Fundamentos

Análisis de datos de movilidad

Claudio Rodríguez Meza
Mayo-2019

Exploración de datos

Se realizó análisis de los dos archivos entregados (excel y csv), para responder las preguntas.

Para responder las preguntas como "expediciones realizadas", "tiempo detenido fuera del terminal", "cantidad de kms recorridos", (basándonos en la data entregada) es necesario realizar cálculos y volver a generar otra BD.

La sección de preguntas de georeferenciacion como "calles", "comunas" es necesario cargar una nueva BD que NO FUE PROPORCIONADA. Para ello se estudió la solución de determinar la ubicación mediante latitud y longitud, utilizando la API de Google Maps. El problema de esto, es que hay costos asociados por cada "pulso" gps enviado a la API, pero enviar 9Millones de pulsos es inviable. Pero cabe destacar, que es la solución más exacta. (Se explica con mayor detalle en el punto más abajo)

Por lo explicado anteriormente, se cargó una serie de archivos excel, descargado desde la web http://www.dtpm.cl/descargas/poss18/09.%20PO%20vigente%20(07Jul2018)_v2.rar
Cada excel posee una "hoja" distinta con el recorrido. Se tuvo que trabajar en la unificación de cada archivo-hoja, para crear una nueva BD

Generación de nueva BD con nuevos campos

Se calculó distancias GPS entre un registro y otro, ordenado por patente y fecha/hora, además de diferencia de tiempos, y sentidos de dirección (ida, retorno)

Nuevos Campos

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Patente	Latitud	Longitud	Tiempogps	Ruta	Velocidad	Ignicion	FECHA2	geoloc	correl	metros distancia	segundos distancia	cant sentidos
2	FLXB-68	-33.3756332	-70.6476517	20170401011521	B02 C2 00I	0	NO	2017-04-01 01:15:21	0xE6100000010C5BA5AABF14B	155016	0.00000	300	2
3	FLXB-68	-33.3756332	-70.6476517	20170401012021	B02 C2 00I	0	NO	2017-04-01 01:20:21	0xE6100000010C5BA5AABF14B	155017	0.00000	300	2
4	FLXB-68	-33.3756332	-70.6476517	20170401012521	B02 C2 00I	0	NO	2017-04-01 01:25:21	0xE6100000010C5BA5AABF14B	155018	0.00000	300	2
5	FLXB-68	-33.3756332	-70.6476517	20170401012645	B02 C2 00I	0	SI	2017-04-01 01:26:45	0xE6100000010C5BA5AABF14B	155019	0.00000	84	2
6	FLXB-68	-33.3756332	-70.6476517	20170401012715	B02 C2 00I	0	SI	2017-04-01 01:27:15	0xE6100000010C5BA5AABF14B	155020	0.00000	30	2
7	FLXB-68	-33.3756332	-70.6476517	20170401012745	B02 C2 00I	0	SI	2017-04-01 01:27:45	0xE6100000010C5BA5AABF14B	155021	0.00000	30	2
8	FLXB-68	-33.3756332	-70.6476517	20170401012815	B02 C2 00I	0	SI	2017-04-01 01:28:15	0xE6100000010C5BA5AABF14B	155022	0.00000	30	2
9	FLXB-68	-33.3756332	-70.6476517	20170401012845	B02 C2 00I	0	SI	2017-04-01 01:28:45	0xE6100000010C5BA5AABF14B	155023	0.00000	30	2
10	FLXB-68	-33.3756332	-70.6476517	20170401012915	B02 C2 00I	0	SI	2017-04-01 01:29:15	0xE6100000010C5BA5AABF14B	155024	0.00000	30	2
11	FLXB-68	-33.3756332	-70.6476517	20170401012945	B02 C2 00I	0	SI	2017-04-01 01:29:45	0xE6100000010C5BA5AABF14B	155025	0.00000	30	2
12	FLXB-68	-33.3756218	-70.6476669	20170401013015	B02 C2 00I	3.89	SI	2017-04-01 01:30:15	0xE6100000010C135509601 4B0	155026	1.89719	30	2
13	FLXB-68	-33.3756142	-70.6476746	20170401013045	B02 C2 00I	0	SI	2017-04-01 01:30:45	0xE6100000010CE474482014B0	155027	1.10631	30	2
14	FLXB-68	-33.3755341	-70.6478500	20170401013115	B02 C2 00I	16.85	SI	2017-04-01 01:31:15	0xE6100000010C33045B8011B0	155028	18.58304	30	2
15	FLXB-68	-33.3748322	-70.6503906	20170401013145	B02 C2 00I	36.86	SI	2017-04-01 01:31:45	0xE6100000010C6E3E6480FAAF	155029	248.90464	30	2
16	FLXB-68	-33.3729820	-70.6500244	20170401013215	B02 C2 00I	15.93	SI	2017-04-01 01:32:15	0xE6100000010C93FFC9DFBDA	155030	208.01804	30	2
17	FLXB-68	-33.3751488	-70.6468735	20170401013245	B02 C2 00I	40.93	SI	2017-04-01 01:32:45	0xE6100000010C1B9139E004B0	155031	379.11553	30	2
18	FLXB-68	-33.3745766	-70.6459274	20170401013315	B02 C2 00I	44.82	SI	2017-04-01 01:33:15	0xE6100000010C666C4320F2AF	155032	108.52969	30	2
19	FLXB-68	-33.3734207	-70.6453934	20170401013345	B02 C2 00I	17.96	SI	2017-04-01 01:33:45	0xE6100000010C2013DF3FCCAF	155033	137.49594	30	2
20	FLXB-68	-33.3722610	-70.6437607	20170401013415	B02 C2 00I	36.86	SI	2017-04-01 01:34:15	0xE6100000010CC3499A3FA6AF	155034	199.06935	30	2
21	FLXB-68	-33.3724976	-70.6419144	20170401013445	B02 C2 00I	6.85	SI	2017-04-01 01:34:45	0xE6100000010C54EB5800AEAF	155035	173.80663	30	2
22	FLXB-68	-33.3728752	-70.6391144	20170401013515	B02 C2 00I	33.89	SI	2017-04-01 01:35:15	0xE6100000010CA7BEE25FBAAF	155036	263.90756	30	2
23	FLXB-68	-33.3743286	-70.6380844	20170401013545	B02 C2 00I	14.82	SI	2017-04-01 01:35:45	0xE6100000010C8E7AE3FFE9AF	155037	187.54177	30	2
24	FLXB-68	-33.3745880	-70.6370239	20170401013615	B02 C2 00I	12.96	SI	2017-04-01 01:36:15	0xE6100000010CAEBCE47FF2AF	155038	102.79461	30	2
	.....												
80	FLXB-68	-33.4389267	-70.6348343	20170401020415	B02 C2 00I	0	SI	2017-04-01 02:04:15	0xE6100000010CADEB06C02EB	155094	128.55650	30	2
81	FLXB-68	-33.4378967	-70.6351013	20170401020445	B02 C2 00I	26.85	SI	2017-04-01 02:04:45	0xE6100000010C5FC3C2FF0CB8	155095	116.90646	30	2
82	FLXB-68	-33.4377708	-70.6351166	20170401020515	B02 C2 00R	0	SI	2017-04-01 02:05:15	0xE6100000010C6792A2DF08B8	155759	14.03615	30	3
83	FLXB-68	-33.4377708	-70.6351166	20170401020545	B02 C2 00R	0	SI	2017-04-01 02:05:45	0xE6100000010C6792A2DF08B8	155760	0.00000	30	3
84	FLXB-68	-33.4376831	-70.6351471	20170401020615	B02 C2 00R	0	SI	2017-04-01 02:06:15	0xE6100000010C8641F4FF05B8	155761	10.13205	30	3
85	FLXB-68	-33.4373283	-70.6354446	20170401020645	B02 C2 00R	14.82	SI	2017-04-01 02:06:45	0xE6100000010CC10EAD5FFAB	155762	48.10268	30	3
86	FLXB-68	-33.4375916	-70.6365585	20170401020715	B02 C2 00R	20.93	SI	2017-04-01 02:07:15	0xE6100000010C8D80650003B8	155763	107.61880	30	3
87	FLXB-68	-33.4397278	-70.6399689	20170401020745	B02 C2 00R	54.82	SI	2017-04-01 02:07:45	0xE6100000010C2E12240049B8	155764	395.86027	30	3
88	FLXB-68	-33.4421501	-70.6442032	20170401020815	B02 C2 00R	30.93	SI	2017-04-01 02:08:15	0xE6100000010C29B6DD5F98B	155765	476.65735	30	3
	.....												
150	FLXB-68	-33.3755722	-70.6473541	20170401023915	B02 C2 00R	13.89	SI	2017-04-01 02:39:15	0xE6100000010CB524F6BF12B	155827	315.94690	30	3
151	FLXB-68	-33.3756294	-70.6470490	20170401023945	B02 C2 00I	0	SI	2017-04-01 02:39:45	0xE6100000010C4335CA9F14B	155096	29.09133	30	4
152	FLXB-68	-33.3756294	-70.6470490	20170401024015	B02 C2 00I	0	SI	2017-04-01 02:40:15	0xE6100000010C4335CA9F14B	155097	0.00000	30	4
153	FLXB-68	-33.3756294	-70.6470490	20170401024515	B02 C2 00I	0	NO	2017-04-01 02:45:15	0xE6100000010C4335CA9F14B	155098	0.00000	300	4
154	FLXB-68	-33.3756294	-70.6470490	20170401025015	B02 C2 00I	0	NO	2017-04-01 02:50:15	0xE6100000010C4335CA9F14B	155099	0.00000	300	4
155	FLXB-68	-33.3756294	-70.6470490	20170401025515	B02 C2 00I	0	NO	2017-04-01 02:55:15	0xE6100000010C4335CA9F14B	155100	0.00000	300	4
156	FLXB-68	-33.3756294	-70.6470490	20170401025726	B02 C2 00I	0	SI	2017-04-01 02:57:26	0xE6100000010C4335CA9F14B	155101	0.00000	131	4
157	FLXB-68	-33.3756294	-70.6470490	20170401025756	B02 C2 00I	0	SI	2017-04-01 02:57:56	0xE6100000010C4335CA9F14B	155102	0.00000	30	4
158	FLXB-68	-33.3756294	-70.6470490	20170401025826	B02 C2 00I	0	SI	2017-04-01 02:58:26	0xE6100000010C4335CA9F14B	155103	0.00000	30	4
159	FLXB-68	-33.3756294	-70.6470490	20170401025856	B02 C2 00I	0	SI	2017-04-01 02:58:56	0xE6100000010C4335CA9F14B	155104	0.00000	30	4
160	FLXB-68	-33.3756218	-70.6470795	20170401025926	B02 C2 00I	3.89	SI	2017-04-01 02:59:26	0xE6100000010C135509601 4B0	155105	2.96071	30	4
161	FLXB-68	-33.3755264	-70.6473160	20170401025956	B02 C2 00I	0	SI	2017-04-01 02:59:56	0xE6100000010C6E64C33F11B0	155106	24.41903	30	4
162	FLXB-68	-33.3755913	-70.6473083	20170401030026	B02 C2 00I	0	SI	2017-04-01 03:00:26	0xE6100000010CC1142F6013B0	155107	7.23371	30	4

Nuevos campos:

- Fecha2 : Se pobló con el tiempogps original, pero en formato fecha – hora
- Geoloc : Punto gps en formato espacial
- Correl : Correlativo único para cada registro
- Metros_distancia : Metros de distancia del campo "geoloc" entre un registro y otro
- Segundos_distancia : Segundos de distancia entre un registro y otro
- Cant_sentidos : Es un contador por cada patente - sentido (ida – retorno) ordenado por patente y fecha. Para obtener el sentido se unieron los 2 archivos entregados (JOIN) por la llave "RUTA", por ello desaparecen sentidos ya que existe casos con rutas "huérfanas".

Finalmente se generó una nueva interfaz CSV, y se inyectó al entorno hadoop para responder las preguntas

Hallazgos

- Cuando el bus está detenido (ignicion=NO), envía pulsos de señales cada 300 secs
- Cuando el bus está encendido, envía pulsos de señales cada 30 secs
- Existen registros que NO cruzan entre excel y csv entregado con la llave ruta
- Existen recorridos que NO cruzan entre el csv y la nueva base de excels descargadas del sitio DTPM

SOLUCIÓN

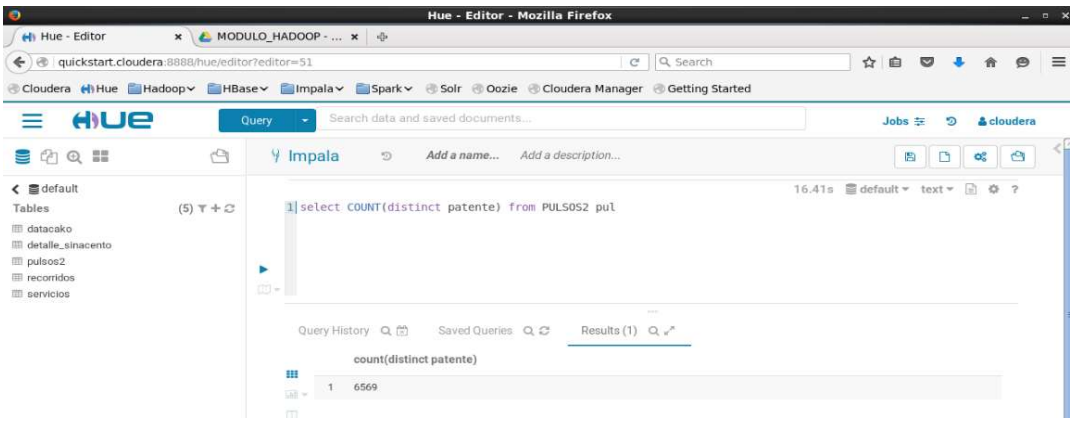
Se utilizó MSSQL e IMPALA para responder las preguntas, y cotejar sus diferencias. NO existió ninguna. Se muestra solo ejercicio en IMPALA

Preguntas simples

1. **Indique la cantidad de buses distintos que forman la flota. Especifique el lapso en que realizó el cálculo.**

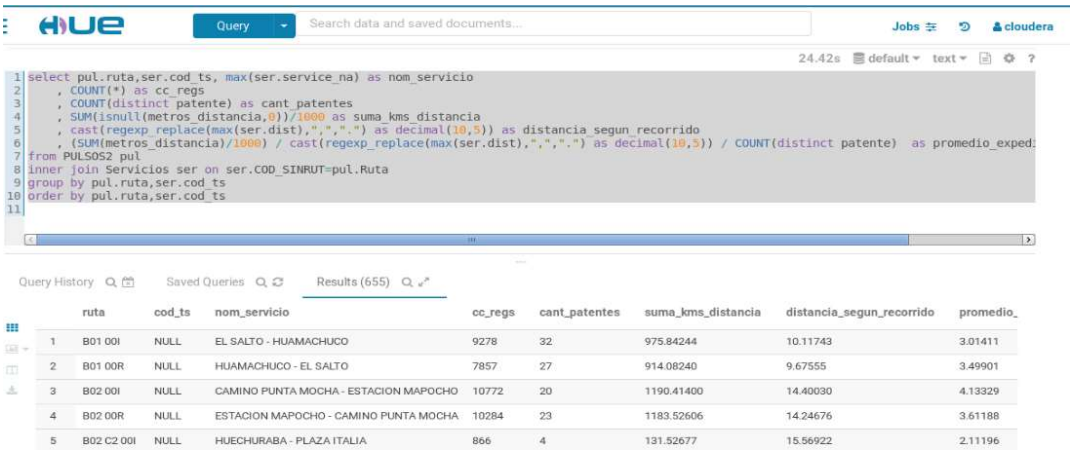
```
select COUNT(distinct patente) from PULSOS2 pul
```

R: 6569
solución en 16.41Secs



2. **Indique para cada recorrido, la cantidad de buses que lo realizan, y la cantidad promedio de expediciones realizadas.**

```
Select pul.ruta,ser.cod_ts, max(ser.service_na) as nom_servicio
       , COUNT(*) as cc_regs
       , COUNT(distinct patente) as cant_patentes
       , SUM(isnull(metros_distancia,0))/1000 as suma_kms_distancia
       , cast(regexp_replace(max(ser.dist),",",".") as decimal(10,5)) as
distancia_segun_recorrido
       , (SUM(metros_distancia)/1000) /
cast(regexp_replace(max(ser.dist),",",".") as decimal(10,5)) /
COUNT(distinct patente)  as promedio_expediciones_por_ppu
from PULSOS2 pul
inner join Servicios ser on ser.COD_SINRUT=pul.Ruta
group by pul.ruta,ser.cod_ts
order by pul.ruta,ser.cod_ts
```



Query History | Saved Queries | Results (655)

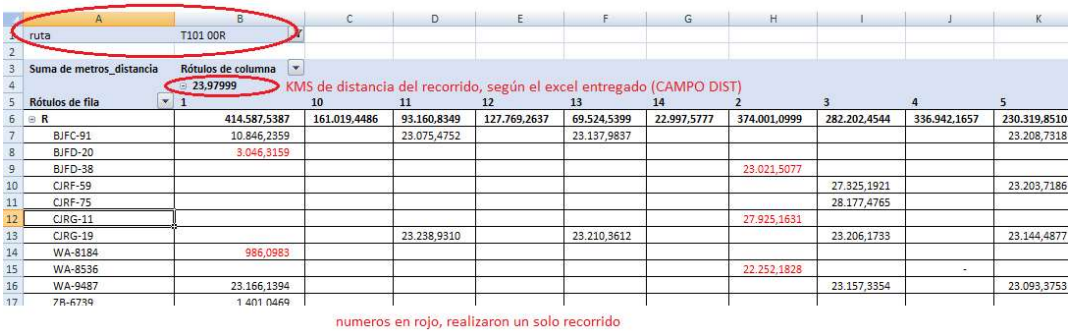
	ruta	cod_ts	nom_servicio	cc_regs	cant_patentes	suma_kms_distancia	distancia_segun_recorrido	promedio_
1	B01 00I	NULL	EL SALTO - HUAMACHUCO	9278	32	975.84244	10.11743	3.01411
2	B01 00R	NULL	HUAMACHUCO - EL SALTO	7857	27	914.08240	9.67555	3.49901
3	B02 00I	NULL	CAMINO PUNTA MOCHA - ESTACION MAPOCHO	10772	20	1190.41400	14.40030	4.13329
4	B02 00R	NULL	ESTACION MAPOCHO - CAMINO PUNTA MOCHA	10284	23	1183.52606	14.24676	3.61188
5	B02 C2 00I	NULL	HUECHURABA - PLAZA ITALIA	866	4	131.52677	15.56922	2.11196

Para verificar, se tomó un registro aleatorio y se verifico en excel

	ruta	cod_ts	nom_servicio	cc_regs	cant_patentes	suma_kms_distancia	distancia_segun_recorrido	promedio_expediciones_por_ppu
1	T101 00R	101	CERRILLOS - RECOLETA	21514	44	2979.36128	23.97999	2.82371

Por ejemplo la ruta "T101 00R", hay patentes que realizaron solo 1 recorrido, o bien, venían finalizando (del día anterior) el recorrido.

En el excel de servicios, viene la distancia de cada recorrido, se dividió la SUMA_DISTANCIA_METROS_RECORRIDO contra la distancia



ruta	T101 00R									
Suma de metros_distancia	Rótulos de columna									
	23,97999	KMS de distancia del recorrido, según el excel entregado (CAMPO DIST)								
Rótulos de fila	1	10	11	12	13	14	2	3	4	5
R	414.587,5387	161.019,4486	93.160,8349	127.769,2637	69.524,5399	22.997,5777	374.001,0999	282.202,4544	336.942,1657	230.319,8510
BJFC-91	10.846,2359		23.075,4752		23.137,9837					23.208,7318
BJFD-20	3.046,3159									
BJFD-38							23.021,5077			
CJRF-59								27.325,1921		23.203,7186
CJRF-75								28.177,4765		
CJRG-11							27.925,1631			
CJRG-19			23.238,9310		23.210,3612			23.206,1733		23.144,4877
WA-8184	986,0983									
WA-8536							22.252,1828		-	
WA-9487	23.166,1394							23.157,3354		23.093,3753
7B-6739	1.401,0469									

numeros en rojo, realizaron un solo recorrido

Para resolver esto, se recomienda utilizar la "moda" por cada ruta. Por ejemplo la moda de esta ruta, es 6 recorridos por patente.

3. **Indique para cada recorrido, el bus que más tiempo estuvo detenido fuera del terminal.**

Para esto se generó un primer grupo, de rutas, patentes, y cantidad de veces que los "metros distancia recorridos" sea igual a cero, y su "ignicion" sea SI

Se asume ignicion='SI' cuando el bus esta fuera del terminal.

```
select pul.ruta, pul.Patente, COUNT(*) as cont
from PULSOS2 pul
inner join Servicios ser on ser.COD_SINRUT=pul.Ruta
where pul.Ignicion='SI' and pul.metros_distancia=0
group by pul.ruta, pul.Patente
ORDER BY 1, 3 DESC
```

14.38s default

Query History | Saved Queries | Results (1,024+)

	ruta	patente	cont
1	B01 00I	CJRH-14	418
2	B01 00I	CJRC-60	321
3	B01 00I	FLXB-22	276
4	B01 00I	SW-6460	260
5	B01 00I	CJRH-12	229
6	B01 00I	CJRY-51	159
7	B01 00I	CJRD-80	140

Se creó VISTAS para resolver estas preguntas, ya que IMPALA no acepta subquerys.

```
create view view_rutas_mas_detenidas
as
select pul.ruta, pul.Patente, COUNT(*) as cont
from PULSOS2 pul
inner join Serviciosser on ser.COD_SINRUT=pul.Ruta
where pul.Ignicion='SI' and pul.metros_distancia=0
group by pul.ruta, pul.Patente
order by ruta, 3 desc;
```

```
create view view_rutas_mas_detenidas_max_cont
as
select ruta,max(cont) as cont
from view_rutas_mas_detenidas
group by ruta
order by ruta;
```

Luego se cruzan ambas vistas, para obtener el resultado

```
select pul.*
from view_rutas_mas_detenidas_max_cont grp
inner join view_rutas_mas_detenidas pul on pul.ruta=grp.ruta and
pul.cont=grp.cont
order by 1,2
```

23.89s default text

```
select pul.*
from view_rutas_mas_detenidas_max_cont grp
inner join view_rutas_mas_detenidas pul on pul.ruta=grp.ruta and  pul.cont=grp.cont
order by 1,2
```

Query History | Saved Queries | Results (660)

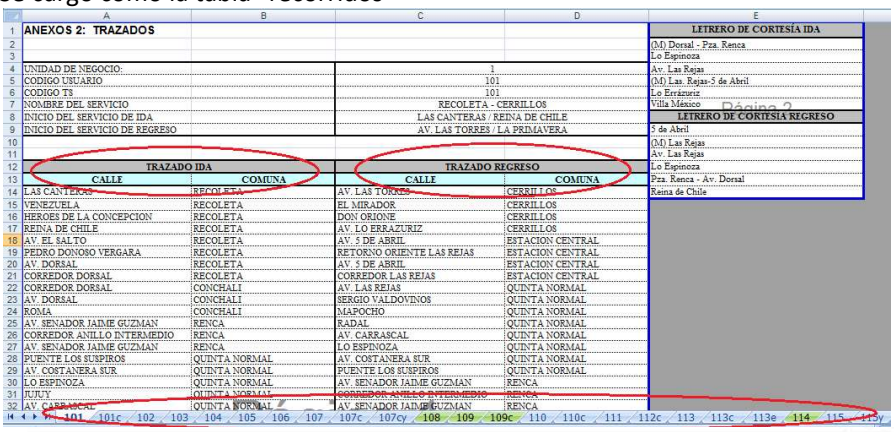
	ruta	patente	cont
1	B01 00I	CJRH-14	418
2	B01 00R	CJRH-12	275
3	B02 00I	CJRF-34	515
4	B02 00R	FLXB-59	415
5	B02 C2 00I	FLXB-72	119
6	B02 C2 00R	FLXB-72	8
7	B02 C2 00R	FLXB-93	8

Preguntas medias

1. **Genere una base de datos, donde se tenga para cada recorrido, las comunas que este visita.**

2. **Genere una base de datos, donde se tenga para cada recorrido, las calles que este recorre.**

Esta información se bajó de la URL de DTPM, y se unificaron todos los Excel-HOJA
http://www.dtpm.cl/descargas/poss18/09.%20PO%20vigente%20(07Jul2018)_v2.rar
Se cargo como la tabla "recorridos"



ANEXOS 2: TRAZADOS				
UNIDAD DE NEGOCIO:		1		
CODIGO USUARIO		101		
CODIGO TS		101		
NOMBRE DEL SERVICIO		RECOLETA - CERRILLOS		
INICIO DEL SERVICIO DE IDA		LAS CANTERAS / REINA DE CHILE		
INICIO DEL SERVICIO DE REGRESO		AV. LAS TORRES / LA PRIMAVERA		
TRAZADO IDA		TRAZADO REGRESO		
CALLE	COMUNA	CALLE	COMUNA	
LAS CANTERAS	RECOLETA	AV. LAS TORRES	CERRILLOS	
VENEZUELA	RECOLETA	EL MIRADOR	CERRILLOS	
HEROES DE LA CONCEPCION	RECOLETA	DON ORIONE	CERRILLOS	
REINA DE CHILE	RECOLETA	AV. LO ERRAZURIZ	CERRILLOS	
AV. EL SALTO	RECOLETA	AV. 5 DE ABRIL	ESTACION CENTRAL	
PEDRO DONOSO VERGARA	RECOLETA	RETORNO ORIENTE LAS REJAS	ESTACION CENTRAL	
AV. DORSAL	RECOLETA	AV. 5 DE ABRIL	ESTACION CENTRAL	
CORREDOR DORSAL	RECOLETA	CORREDOR LAS REJAS	ESTACION CENTRAL	
CORREDOR DORSAL	CONCHALI	AV. LAS REJAS	QUINTA NORMAL	
AV. DORSAL	CONCHALI	SERGIO VALDOVINOS	QUINTA NORMAL	
ROMA	CONCHALI	MAPOCHO	QUINTA NORMAL	
AV. SENADOR JAIME GUZMAN	RENCA	RADAL	QUINTA NORMAL	
CORREDOR ANILLO INTERMEDIO	RENCA	AV. CARRASCAL	QUINTA NORMAL	
AV. SENADOR JAIME GUZMAN	RENCA	LO ESPINOZA	QUINTA NORMAL	
PUENTE LOS SUSPIROS	QUINTA NORMAL	AV. COSTANERA SUR	QUINTA NORMAL	
AV. COSTANERA SUR	QUINTA NORMAL	PUENTE LOS SUSPIROS	QUINTA NORMAL	
LO ESPINOZA	QUINTA NORMAL	AV. SENADOR JAIME GUZMAN	RENCA	
JUJUY	QUINTA NORMAL	CORREDOR ANILLO INTERMEDIO	RENCA	
AV. CARRASCAL	QUINTA NORMAL	AV. SENADOR JAIME GUZMAN	RENCA	

LETRERO DE CORTESÍA IDA
(M) Dorsal - Pza. Renca
Lo Espinoza
Av. Las Rejas
(M) Las. Rejas-5 de Abril
Lo Errázuriz
Villa México

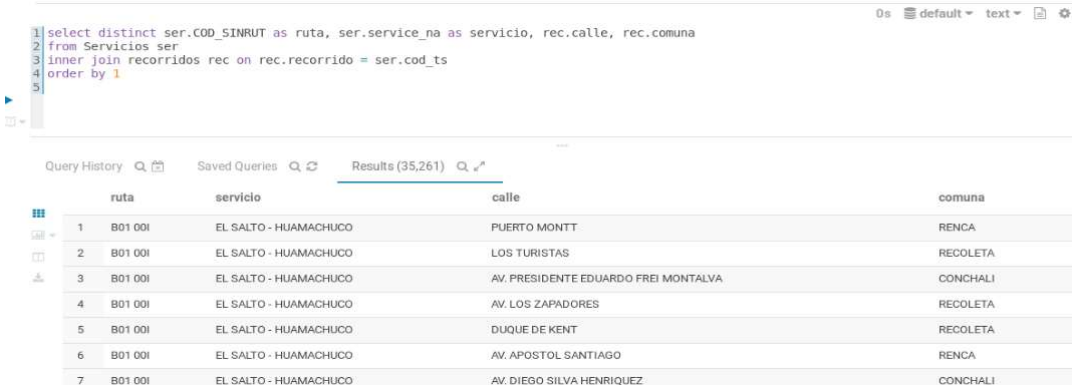
LETRERO DE CORTESÍA REGRESO
5 de Abril
(M) Las Rejas
Av. Las Rejas
Lo Espinoza
Pza. Renca - Av. Dorsal
Reina de Chile

```
select * from recorridos
```

	recorrido	tipo	calle	comuna
1	101	I	LAS CANTERAS	RECOLETA
2	101	I	VENEZUELA	RECOLETA
3	101	I	HEROES DE LA CONCEPCION	RECOLETA
4	101	I	REINA DE CHILE	RECOLETA
5	101	I	AV. EL SALTO	RECOLETA
6	101	I	PEDRO DONOSO VERGARA	RECOLETA
7	101	I	AV. DORSAL	RECOLETA
8	101	I	CORREDOR DORSAL	RECOLETA

Con esta consulta se responden las 2 preguntas

```
select distinct ser.COD_SINRUT as ruta, ser.service_na as servicio,
rec.calle, rec.comuna
from Servicios ser
inner join recorridos rec on rec.recorrido = ser.cod_ts
order by 1
```



```
select distinct ser.COD_SINRUT as ruta, ser.service_na as servicio, rec.calle, rec.comuna
from Servicios ser
inner join recorridos rec on rec.recorrido = ser.cod_ts
order by 1
```

	ruta	servicio	calle	comuna
1	B01 001	EL SALTO - HUAMACHUCO	PUERTO MONTT	RENCA
2	B01 001	EL SALTO - HUAMACHUCO	LOS TURISTAS	RECOLETA
3	B01 001	EL SALTO - HUAMACHUCO	AV. PRESIDENTE EDUARDO FREI MONTALVA	CONCHALI
4	B01 001	EL SALTO - HUAMACHUCO	AV. LOS ZAPADORES	RECOLETA
5	B01 001	EL SALTO - HUAMACHUCO	DUQUE DE KENT	RECOLETA
6	B01 001	EL SALTO - HUAMACHUCO	AV. APOSTOL SANTIAGO	RENCA
7	B01 001	EL SALTO - HUAMACHUCO	AV. DIEGO SILVA HENRIQUEZ	CONCHALI

Preguntas complejas

1. **Busque por cada recorrido, el bus que realizó la expedición completa más rápidamente.**

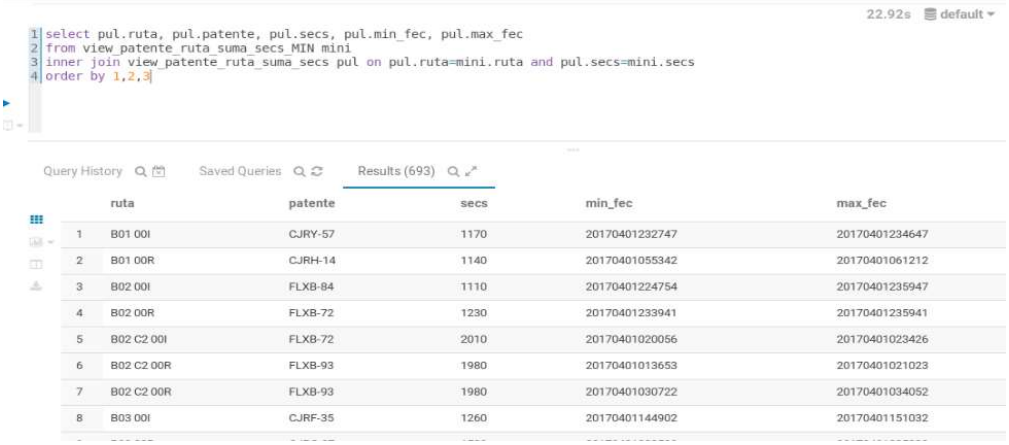
Se crean 2 VISTAS

```
Create view view_patente_ruta_suma_secs
as
select Ruta,Patente,cant_sentidos,SUM(segundos_distancia) as secs,
MIN(tiempogps) as min_fec, MAX(tiempogps) as max_fec
from pulsos2
where Ignicion='SI' and cant_sentidos>0 and metros_distancia>0
group by Ruta,Patente,cant_sentidos
having SUM(segundos_distancia)>1000
order by ruta,patente, secs ;
```

```
create view view_patente_ruta_suma_secs_MIN
as
select ruta, min(secs) as secs
from view_patente_ruta_suma_secs
group by ruta;
```

Finalmente respondemos la pregunta

```
Select pul.ruta, pul.patente, pul.secs, pul.min_fec, pul.max_fec
From view_patente_ruta_suma_secs_MIN mini
Inner join view_patente_ruta_suma_secs pul on pul.ruta=mini.ruta and
pul.secs=mini.secs
order by 1,2,3
```



```
select pul.ruta, pul.patente, pul.secs, pul.min_fec, pul.max_fec
from view_patente_ruta_suma_secs_MIN mini
inner join view_patente_ruta_suma_secs pul on pul.ruta=mini.ruta and pul.secs=mini.secs
order by 1,2,3
```

22.92s default

Query History | Saved Queries | Results (693)

	ruta	patente	secs	min_fec	max_fec
1	B01 00I	CJRY-57	1170	20170401232747	20170401234647
2	B01 00R	CJRH-14	1140	20170401055342	20170401061212
3	B02 00I	FLXB-84	1110	20170401224754	20170401235947
4	B02 00R	FLXB-72	1230	20170401233941	20170401235941
5	B02 C2 00I	FLXB-72	2010	20170401020056	20170401023426
6	B02 C2 00R	FLXB-93	1980	20170401013653	20170401021023
7	B02 C2 00R	FLXB-93	1980	20170401030722	20170401034052
8	B03 00I	CJRF-35	1260	20170401144902	20170401151032

2. Para una comuna en particular, encuentre la calle por la que más recorridos pasan.

Nuevamente se crean 2 vistas

```
create view view_comuna_calle_cant_rec
as
select comuna, calle,COUNT(distinct recorrido) as cant_rec
from recorridos
group by calle,comuna;
```

```
create view view_comuna_calle_cant_rec_MAX
as
select comuna,max(cant_rec) as cant_rec
from view_comuna_calle_cant_rec
group by comuna;
```

Finalmente respondemos la pregunta

```
select rec.*
from view_comuna_calle_cant_rec_MAX maxi
inner join view_comuna_calle_cant_rec rec on rec.comuna=maxi.comuna and
rec.cant_rec=maxi.cant_rec
order by 3 desc
```

```
select rec.*
from view_comuna_calle_cant_rec_MAX maxi
inner join view_comuna_calle_cant_rec rec on rec.comuna=maxi.comuna and rec.cant_rec=maxi.cant_rec
order by 3 desc
```

Query History | Saved Queries | Results (39)

	comuna	calle	cant_rec
1	SANTIAGO	AV. LIBERTADOR BERNARDO O'HIGGINS	81
2	ESTACION CENTRAL	AV. LIBERTADOR BERNARDO O'HIGGINS	44
3	PUENTE ALTO	AV. CONCHA Y TORO	38
4	PROVIDENCIA	AV. PROVIDENCIA	34
5	MAIPU	AV. LOS PAJARITOS	33
6	LA CISTERNA	GRAN AVENIDA JOSE MIGUEL CARRERA	33
7	QUILICURA	AV. AMERICO VESPUCIO	27
8	LA FLORIDA	AV. AMERICO VESPUCIO	25

3. **Para una comuna en particular, encuentre la calle con más kilómetros de viaje.**

Para esto se estudió la solución de determinar la ubicación mediante latitud y longitud, utilizando la API de Google Maps. El problema de esto, es que hay costos asociados por cada "pulso" gps enviado a la API, pero enviar 9Millones de pulsos es inviable.

La solución propuesta es tomar alguna patente aleatoria, y por cada N metros recorridos (por ejemplo cada 1000) buscar la señal de GPS, y se almacena el resultado.

Para ello activamos una "key" GPS como ejemplo, y enviamos una coordenada mediante PHP y logramos capturar el JSON devuelto por la API

Tomamos el primer registro que pertenece a la patente **"FLXB-68"** ruta **"B02 C2 00I"** latitud y longitud **-33.3756332 -70.6476517**

```
https://maps.googleapis.com/maps/api/geocode/json?latlng=-33.3756332,-70.6476517&key=AIzaSyB1mrKoPUeLgs7xmVb5GVHNphbqTSvO30o
```


```
"plus_code" : {
   "compound_code" : "J9F2+PW Santiago, Chile",
   "global_code" : "47RFJ9F2+PW"
},
"results" : [
   {
      "address_components" : [
         {
            "long_name" : "1315",
            "short_name" : "1315",
            "types" : [ "street_number" ]
         },
         {
            "long_name" : "José Joaquín Aguirre Luco",
            "short_name" : "José Joaquín Aguirre Luco",
            "types" : [ "route" ]
         },
         {
            "long_name" : "Huechuraba",
            "short_name" : "Huechuraba",
            "types" : [ "locality", "political" ]
         },
         {
            "long_name" : "Huechuraba",
            "short_name" : "Huechuraba",
            "types" : [ "administrative_area_level_3", "political" ]
         },
         {
            "long_name" : "Santiago",
            "short_name" : "Santiago",
            "types" : [ "administrative_area_level_2", "political" ]
         },
         {
            "long_name" : "Región Metropolitana",
            "short_name" : "Región Metropolitana",
            "types" : [ "administrative_area_level_1", "political" ]
         },
         {
            "long_name" : "Chile",
```


Desde el JSON de la API se puede rescatar, rutas cercanas, ciudad, comuna, calle, y otros datos.

Se deberia almacenar la información para poder responder la pregunta, mediante la fabricación de algun "BOT" programado en PHP, Python u algún otro lenguaje

Por un asunto de tiempo, no fue implementada, solo se expone en éste documento.

ENLACES DESCARGA BD

- Recorridos -> recorridos https://drive.google.com/file/d/1n84KJzUVbJY9mq0PrJqoBiGFj8fQX2uj/view?usp=sharing

- Pulsos GPS version2 -> pulso2
  https://drive.google.com/file/d/19x2-YzBSaVMeKcHrD7p0sKazXrOXDU7Q/view?usp=sharing

- Diccionario de servicios PO (excel entregado transformado a csv) -> servicios
  https://drive.google.com/file/d/10w9lpoqthQKmyEZas1gpR33fFxYkxVp9/view?usp=sharing

ESTRUCTURA DE LAS TABLAS EN IMPALA

pulsos2
- patente (string)
- latitud (string)
- longitud (string)
- tiempogps (bigint)
- ruta (string)
- ignicion (string)
- metros_distancia (decimal(10,5))
- segundos_distancia (bigint)
- cant_sentidos (bigint)

recorridos
- recorrido (string)
- tipo (string)
- calle (string)
- comuna (string)

servicios
- route_id (string)
- route_name (string)
- service_na (string)
- un (bigint)
- op_noc (string)
- dist (string)
- po_mod (string)
- sentido (string)
- cod_usuari (string)
- cod_ts (string)
- cod_sinser (string)
- cod_sinrut (string)
- cod_ususen (string)
- tipo_serv (bigint)
- frec_pm (bigint)
- frec_pt (bigint)
- plazas_pm (bigint)
- plazas_pt (bigint)
- sel_pm (bigint)
- sel_pt (bigint)