QUESTION 1

Using the dataset below, please find the following:

1. **The time (hour, minute and second) of all the times a parcel has been "READY"**
2. **Last time each parcel changed state**
3. **Last state of each parcel**

For simplicity, you can limit the timeframe to events that happened during the current month.

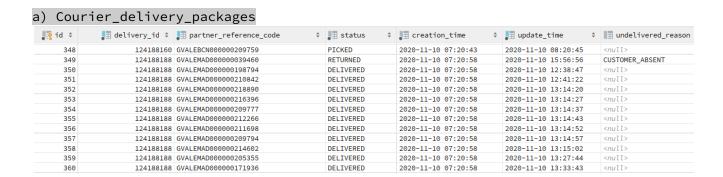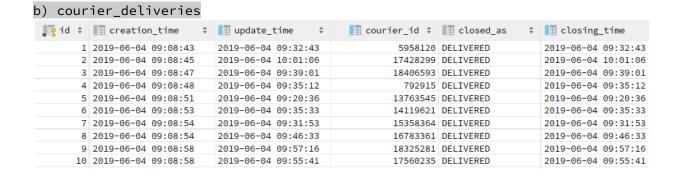| id	changed_at	state	parcel_id	reason	action
1	2020-11-06 15:14:34	CREATED	1	<null>	CREATE_PARCEL
37890	2020-11-10 10:55:46	CREATED	10000	<null>	CREATE_PARCEL
5542492	2021-01-31 09:14:08	CREATED	454357	<null>	UPDATE_PARCEL
11619959	2021-05-29 02:13:34	CREATED	1000000	<null>	CREATE_PARCEL
4930762	2021-01-24 09:13:27	CREATED	435987	<null>	CREATE_PARCEL
97611	2020-11-11 11:16:44	DELIVERED	10000	<null>	MARK_DELIVERED
5241489	2021-01-27 11:27:55	DELIVERED	435987	<null>	MARK_DELIVERED
11672758	2021-05-31 15:05:35	DELIVERED	1000000	<null>	MARK_DELIVERED
5693180	2021-02-01 13:35:48	DELIVERED	454357	<null>	MARK_DELIVERED
11671612	2021-05-31 12:54:29	DISPATCHED	1000000	<null>	MARK_DISPATCHED
76	2020-11-06 16:06:43	DISPATCHED	1	<null>	MARK_DISPATCHED
5211002	2021-01-27 08:30:12	DISPATCHED	435987	<null>	MARK_DISPATCHED
5683340	2021-02-01 11:58:28	DISPATCHED	454357	<null>	MARK_DISPATCHED
83886	2020-11-11 09:10:22	DISPATCHED	10000	<null>	MARK_DISPATCHED
5584484	2021-01-31 13:34:13	IN_RETURNED_BIN	454357	<null>	PICKER
5164186	2021-01-26 14:05:05	IN_RETURNED_BIN	435987	<null>	PICKER
5208508	2021-01-27 08:10:55	IN_RETURNED_BIN	454357	<null>	PICKER
88880	2020-11-11 09:15:43	PICKED	10000	<null>	MARK_PICKED
11671650	2021-05-31 12:56:09	PICKED	1000000	<null>	MARK_PICKED
5213574	2021-01-27 08:55:33	PICKED	435987	<null>	MARK_PICKED
5683651	2021-02-01 12:02:04	PICKED	454357	<null>	MARK_PICKED
11	2020-11-06 15:14:36	READY	1	<null>	PREPARE_FOR_BATCHING
5554825	2021-01-31 09:26:49	READY	454357	<null>	PREPARE_FOR_BATCHING
11622805	2021-05-29 02:20:19	READY	1000000	<null>	PREPARE_FOR_BATCHING
5598659	2021-01-31 17:17:52	READY	454357	<null>	RETURN_FLOW
45774	2020-11-10 11:29:19	READY	10000	<null>	PREPARE_FOR_BATCHING
4945109	2021-01-24 09:21:56	READY	435987	<null>	PREPARE_FOR_BATCHING
5171423	2021-01-26 18:17:41	READY	435987	<null>	RETURN_FLOW
5192221	2021-01-27 00:32:58	ROUTED_IN_BIN	435987	<null>	PICKER
31	2020-11-06 15:18:02	ROUTED_IN_BIN	1	<null>	ADMISSION_HANDSHAKE
5626975	2021-02-01 00:46:27	ROUTED_IN_BIN	454357	<null>	PICKER
11648876	2021-05-30 20:49:39	ROUTED_IN_BIN	1000000	<null>	PICKER
71857	2020-11-11 06:13:13	ROUTED_IN_BIN	10000	<null>	ADMISSION_HANDSHAKE
21	2020-11-06 15:16:19	ROUTED_NOT_CONFIRMED	1	<null>	BATCHING_EXECUTION
5608219	2021-01-31 18:19:36	ROUTED_NOT_CONFIRMED	454357	<null>	BATCHING_EXECUTION
5187658	2021-01-26 22:23:51	ROUTED_NOT_CONFIRMED	435987	<null>	BATCHING_EXECUTION
64335	2020-11-10 21:15:23	ROUTED_NOT_CONFIRMED	10000	<null>	BATCHING_EXECUTION
11647612	2021-05-30 18:23:42	ROUTED_NOT_CONFIRMED	1000000	<null>	BATCHING_EXECUTION
46	2020-11-06 15:18:59	SCHEDULED	1	<null>	BATCHING_SCHEDULING
11659576	2021-05-31 08:13:58	SCHEDULED	1000000	<null>	BATCHING_SCHEDULING
5640046	2021-02-01 06:12:30	SCHEDULED	454357	<null>	BATCHING_SCHEDULING
5198190	2021-01-27 06:12:17	SCHEDULED	435987	<null>	BATCHING_SCHEDULING
77286	2020-11-11 06:25:11	SCHEDULED	10000	<null>	BATCHING_SCHEDULING

Make it scalable.

QUESTION 2

Dataset a) contains information regarding the last status of each parcel for each time it was attempted to deliver. Dataset b) contains information regarding the status of all orders (remember 1 order is composed of multiple parcels). The primary key in this table matches with the "delivery_id" field of the table above.

Using the datasets below, and for parcels attempted last week, please find:

1. **Which courier did the 1st attempt of each day?**
2. **What's the main reason why parcels are not delivered? What % does it represent of the total unsuccessful attempts?**
3. **Is there a time of day you'd recommend to stop delivering? Why?**

a) Courier_delivery_packages

id	delivery_id	partner_reference_code	status	creation_time	update_time	undelivered_reason
348	124188160	GVALEBCN000000209759	PICKED	2020-11-10 07:20:43	2020-11-10 08:20:45	<null>
349	124188188	GVALEMAD000000039460	RETURNED	2020-11-10 07:20:58	2020-11-10 15:56:56	CUSTOMER_ABSENT
350	124188188	GVALEMAD000000198794	DELIVERED	2020-11-10 07:20:58	2020-11-10 12:38:47	<null>
351	124188188	GVALEMAD000000210842	DELIVERED	2020-11-10 07:20:58	2020-11-10 12:41:22	<null>
352	124188188	GVALEMAD000000218890	DELIVERED	2020-11-10 07:20:58	2020-11-10 13:14:20	<null>
353	124188188	GVALEMAD000000216396	DELIVERED	2020-11-10 07:20:58	2020-11-10 13:14:27	<null>
354	124188188	GVALEMAD000000209777	DELIVERED	2020-11-10 07:20:58	2020-11-10 13:14:37	<null>
355	124188188	GVALEMAD000000212266	DELIVERED	2020-11-10 07:20:58	2020-11-10 13:14:43	<null>
356	124188188	GVALEMAD000000211698	DELIVERED	2020-11-10 07:20:58	2020-11-10 13:14:52	<null>
357	124188188	GVALEMAD000000209794	DELIVERED	2020-11-10 07:20:58	2020-11-10 13:14:57	<null>
358	124188188	GVALEMAD000000214602	DELIVERED	2020-11-10 07:20:58	2020-11-10 13:15:02	<null>
359	124188188	GVALEMAD000000205355	DELIVERED	2020-11-10 07:20:58	2020-11-10 13:27:44	<null>
360	124188188	GVALEMAD000000171936	DELIVERED	2020-11-10 07:20:58	2020-11-10 13:33:43	<null>

b) courier_deliveries

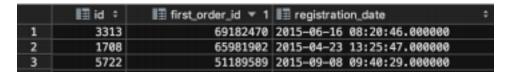| id	creation_time	update_time	courier_id	closed_as	closing_time
1	2019-06-04 09:08:43	2019-06-04 09:32:43	5958120	DELIVERED	2019-06-04 09:32:43
2	2019-06-04 09:08:45	2019-06-04 10:01:06	17428299	DELIVERED	2019-06-04 10:01:06
3	2019-06-04 09:08:47	2019-06-04 09:39:01	18406593	DELIVERED	2019-06-04 09:39:01
4	2019-06-04 09:08:48	2019-06-04 09:35:12	792915	DELIVERED	2019-06-04 09:35:12
5	2019-06-04 09:08:51	2019-06-04 09:20:36	13763545	DELIVERED	2019-06-04 09:20:36
6	2019-06-04 09:08:53	2019-06-04 09:35:33	14119621	DELIVERED	2019-06-04 09:35:33
7	2019-06-04 09:08:54	2019-06-04 09:31:53	15358364	DELIVERED	2019-06-04 09:31:53
8	2019-06-04 09:08:54	2019-06-04 09:46:33	16783361	DELIVERED	2019-06-04 09:46:33
9	2019-06-04 09:08:58	2019-06-04 09:57:16	18325281	DELIVERED	2019-06-04 09:57:16
10	2019-06-04 09:08:58	2019-06-04 09:55:41	17560235	DELIVERED	2019-06-04 09:55:41

Make it scalable.

QUESTION 3

Build one SQL query to create a cohort of Signup to First Order and show the result. The objective of this cohort is to see, out of the users that signed up in Week N, how many did their first order in Week N+1, N+2, N+3...

The **users** table has 1M+ rows; here's the first three rows:

	id	first_order_id ▼ 1	registration_date
1	3313	69182470	2015-06-16 08:20:46.000000
2	1708	65981902	2015-04-23 13:25:47.000000
3	5722	51189589	2015-09-08 09:40:29.000000

The **orders** table has 1M+ rows; here's the first row:

	id	customer_id	activation_time_local	first_order
1	985	596	2015-03-15 22:10:34.000000	☐

The output does not require to be in a cohort format. The end user could potentially use the pivot function from Excel or Google sheets to do so.

Make it scalable