ORGANIZATIONAL REPORTING

Consider a Stream **ORG_HIER_STREAM** that is receiving data representing the employee-supervisor reporting in the Organization. The message from the stream would be in JSON format as below.


```
{
 "emp": "E3000",
 "grade": "G8",
 "effective_dt": "10-Jan-2019",
 "suprvsr": "E1000",
 "status":"SUPERVISOR_CHANGE",
 "posted_dt": "10-Jan-2019"
}
```


The above sample indicates employee **E3000** with grade **G8** reporting to Supervisor **E1000** starting from **10-Jan-2019** because of change in Supervisor.

- status can have any of these 4 values: "SUPERVISOR_CHANGE", "NEW", "ROLE_CHANGE", "RESIGNED"

- effective_dt can be today or past day. It can't be future date

- Employee E1 is the CEO of the Organization and is at top of the hierarchy chain. So ultimately every employee in the organization would directly/indirectly reports to employee E1

- grade ranges from G8 to G18, G18 being the grade of CEO

- Employees with same grades cannot report to each other

Requirements:

Write a Consumer Process that will refresh table **ORG_HIER_TBL** with new records coming from the input **ORG_HIER_STREAM**. The table should have **RPT_END_DT** properly calculated. Also, calculate the reporting hierarchy (till the top-level) and populate **SUPRVSR_HIER** column accordingly Overall this will be the structure of **ORG_HIER_TBL** table

EMP	GRADE	RPT_ST_DT	RPT_END_DT	SUPRVSR	SUPRVSR_HIER
E3000	G8	10-Jan-2019	14-Mar-2020	E1000	E1000-E500-E100-E1
E3000	G9	15-Mar-2020	31-Jul-2020	E200	E200-E50-E20-E1
E3000	G9	1-Aug-2020	31-Dec-9999	E200	E200-E10-E1

ORG_HIER_TBL	
Column	**Description**
EMP	Employee
GRADE	Employee Grade
RPT_ST_DT	Start date
RPT_END_DT	End date
SUPRVSR	Current Supervisor during this period/date range
SUPRVSR_HIER	List of all Supervisors (separated with **hyphen**) falling in the hierarchy till the top-level during this period/date range

*Populate **RPT_END_DT** with **31-Dec-9999** when the chain cannot be broken further to indicate that this is the current hierarchy for this period

Format of the Date can be **DD-MM-YYYY (21-01-2020) or **DD-MMM-YYYY (21-Jan-2020)** or **any other format** whichever suits you

Write jobs which will generate below outputs which are processed by a **BI reporting tool** to generate reports every fortnight giving insights into reporting structure with-in the Organization.

Report A: Rank each Supervisor (of same Grades) based on number of reportees reporting. If the number of reportess is same then rank them based on employee id giving preference to employee who has joined organization earlier. Sample report below.

EMP	Grade	Num. of Reportees	Rank
E2000	G10	10	1
E2001	G10	9	2
E2002	G10	9	3
E1000	G12	4	1
E1001	G12	3	2

Report B: Find average number of **direct** reportees and **total** (direct+indirect) reportees for each grade. Sample report below.

Grade	Avg Dir Reportees	Avg Total Reportees
G10	5	5

G11	6.5	6
G13	8	14
G14	9	16

Report C: List of Eligible Supervisors (G10 and above only) who were not assigned any reportees under them. Order the list based on Grade starting from Grade 10 and so on. Sample report below.

EMP	Grade
E3124	G10
E3000	G10
E2000	G11
E2996	G11
E5	G12

Report D: Generate a report which can give insight on number of employees that have resigned and joined at different grades in the past 30 days. Sample report below.

Grade	resigned	joined
G8	10	5
G9	7	1
G10	0	2
G11	3	1

Key points:

- Choose data types as per your convenience
- Use logical and valid assumptions whenever deemed fit. Should be able to explain during discussion
- Assume corrupted or invalid data in input. Please use valid filter as necessary
- Please do consider late arrival of messages in the input stream. Hence handle accordingly
- Please treat the use case as Big Data problem with millions of records/messages coming through and hence provide solution with best performance analysis and considerations