# SUTD 2022 Term 7 50.046 Homework 4

Hand-out: 14 April

Hand-in: 20 April (Wednesday 23:59)

Name: James Raphael Tiovalen Student ID: 1004555

**Question 1. [10pt]** Consider a streaming data processing in a distributed sensor network. The ID of the sensor nodes are labelled as A, B, C, …, and H. Each sensor collects data, i.e., the number in the circle, independently and all the nodes send data to a single sink node where the summation of all the sensing data is calculated by a streaming processing engine. Due to the time delay in the data forwarding, the processing time might be later than the event time. The figure below illustrates the distribution of the sensing data in a coordinate system of event time and processing time. The format of time is *hh:mm*. Please answer the following questions:

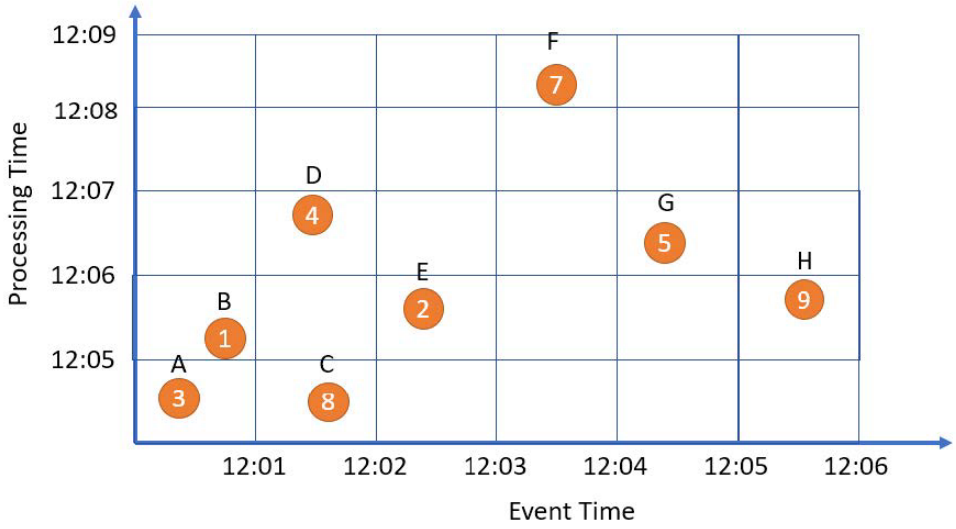
a) What are the time skews for Node A, F, and H, respectively in this example? Assume that all timestamps are *rounded down* to the closest minute and the temporal resolution is one minute. **[3pt]**

b) Recall that we use the 'watermark' to mark the input completeness with respect to event times. Please plot the *perfect watermark* line and the *heuristic watermark* line in the figure. Assume that the event time window is one minute, and the tolerable latency is also one minute. **[2pt]**

c) In task (b) please indicate the data summation so far on top of each watermark within each processing time window. **[2pt]**

d) To provide timely update of the data summation, a 'trigger' is applied in the streaming processing engine. Assume that the trigger is fired every one minute to update the summation of data so far in each event time window. Please mark in the diagram the reported data (i.e., the summation) when the trigger is fired. For each report, please indicate whether it is an *ontime/late/early* report. (You can duplicate the diagram as needed.) **[3pt]**

Hint: You can refer to https://www.oreilly.com/radar/the-world-beyond-batch-streaming-102/



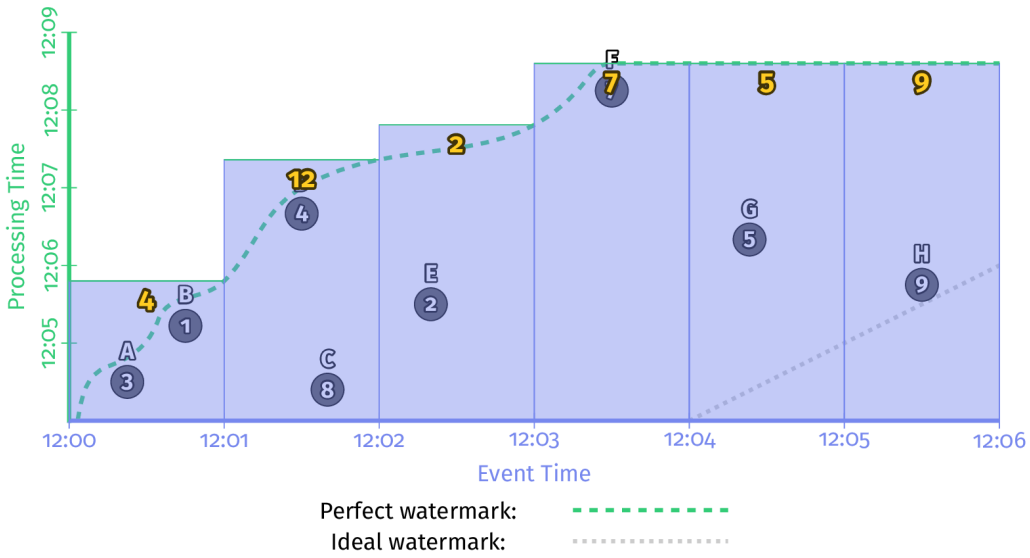

**Answer:**

a) The time skews for:

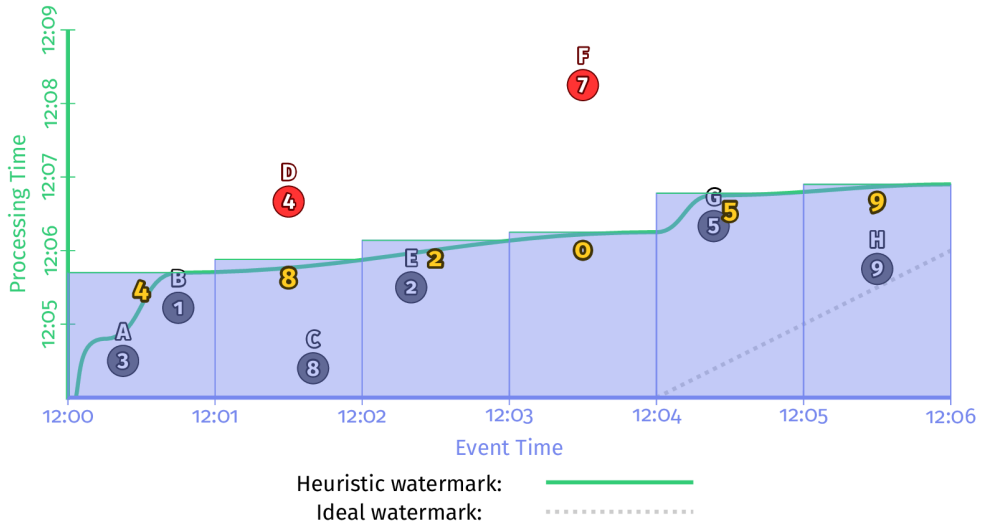
- Node A: 12:04 – 12:00 = **4 minutes**
- Node F: 12:08 – 12:03 = **5 minutes**
- Node H: 12:05 – 12:05 = **0 minutes**

b) Assuming that windowing is fixed, windows are materialized as the watermark passes the end of the window, and that low watermarks are used, then the requested plots for:

- The perfect watermark line:



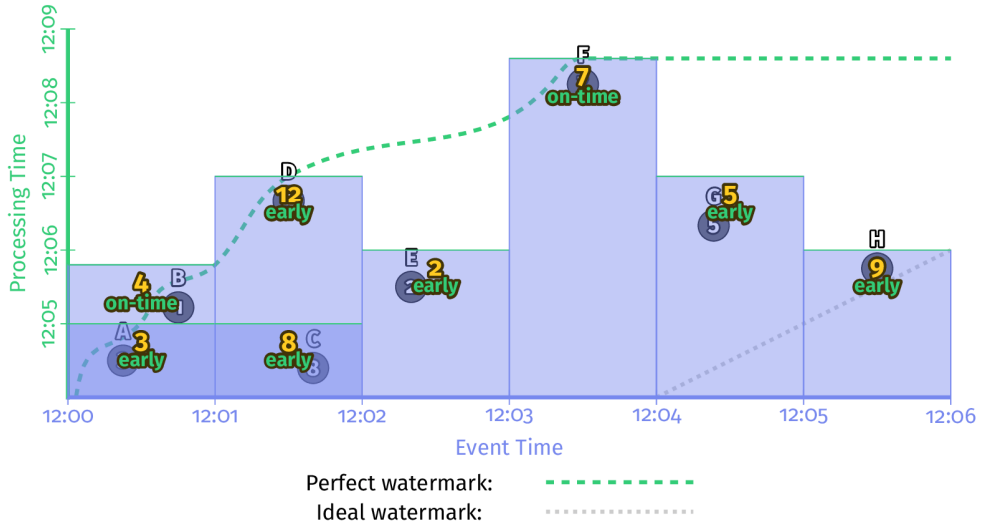
- One possible heuristic watermark line:



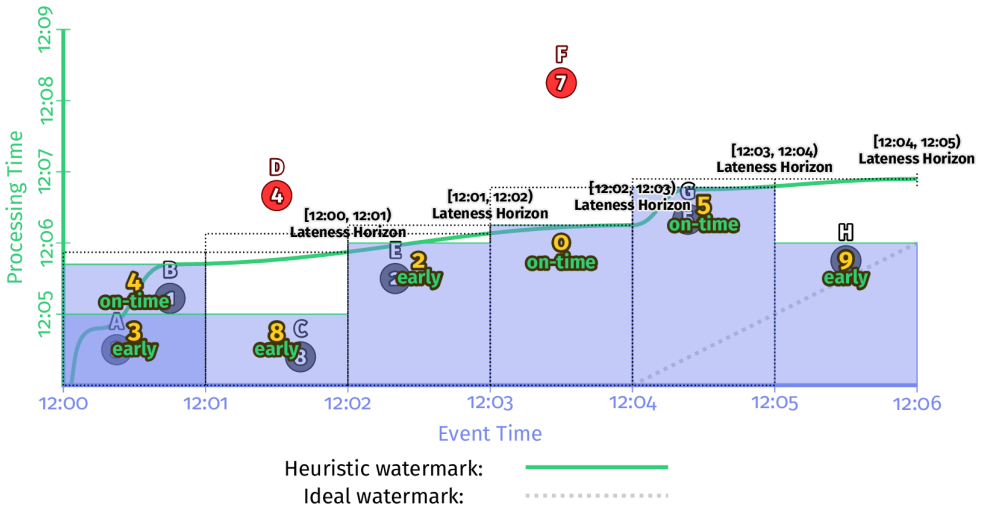

c) Solution for this part is included in the diagrams in part (b) above.

d) The requested plots for:

- The perfect watermark line:




- The aforementioned heuristic watermark line:



**Question 2. [10pt]** You are tasked to design a VPC to host components of a web service. In this case, the VPC is assigned IP addresses 10.0.0.0/16 (in CIDR notation). These IP addresses will be *allocated evenly* to two public subnets residing in two availability zones AZ1 and AZ2. The web service is comprised of two web servers S1 and S2 residing in AZ1 and AZ2 respectively. The two web servers share an Internet Gateway (IGW) to connect to the Internet.

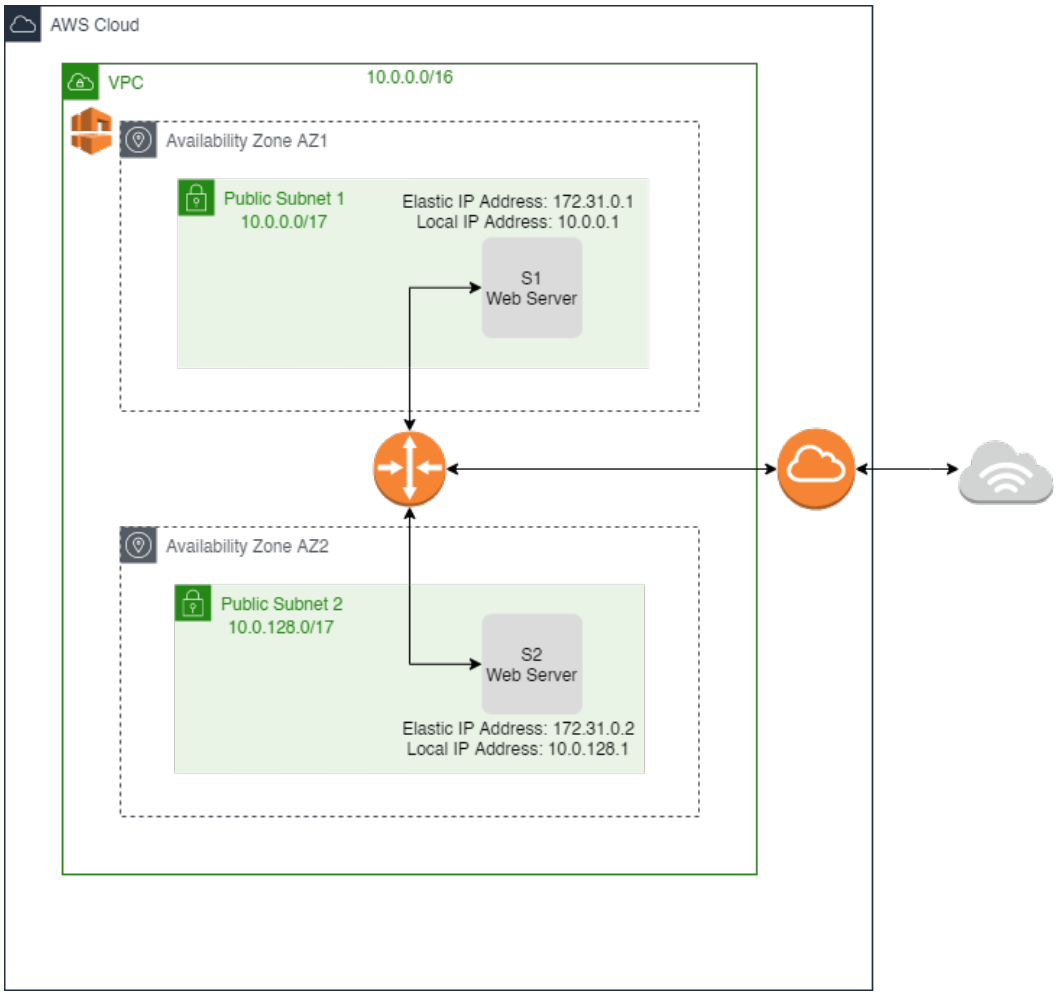
a) Please use standard AWS icons to plot the architecture of the VPC. You can add network components as needed to enable the above application scenario. In your diagram, please specify the IP addresses (in CIDR notation) associated with each subnet. You can choose random internal IP address from 10.0.0.0/16 and elastic IP address from 172.31.0.0/24 as needed. **[4pt]**

b) Assume that there is only one route table associated with the two subnets. How to design the route table to enable the above service? Please specify the 'Destination' and 'Target' fields of the route table. Briefly elaborate the behavior of the route table you designed. **[2pt]**

c) Assume that the subnet in AZ2 is set to be a *private* subnet. What changes need to be made to the architecture and the route table to keep server S2's connection to the Internet? **[4pt]**

**Answer:**

a) The VPC architecture is as follows:

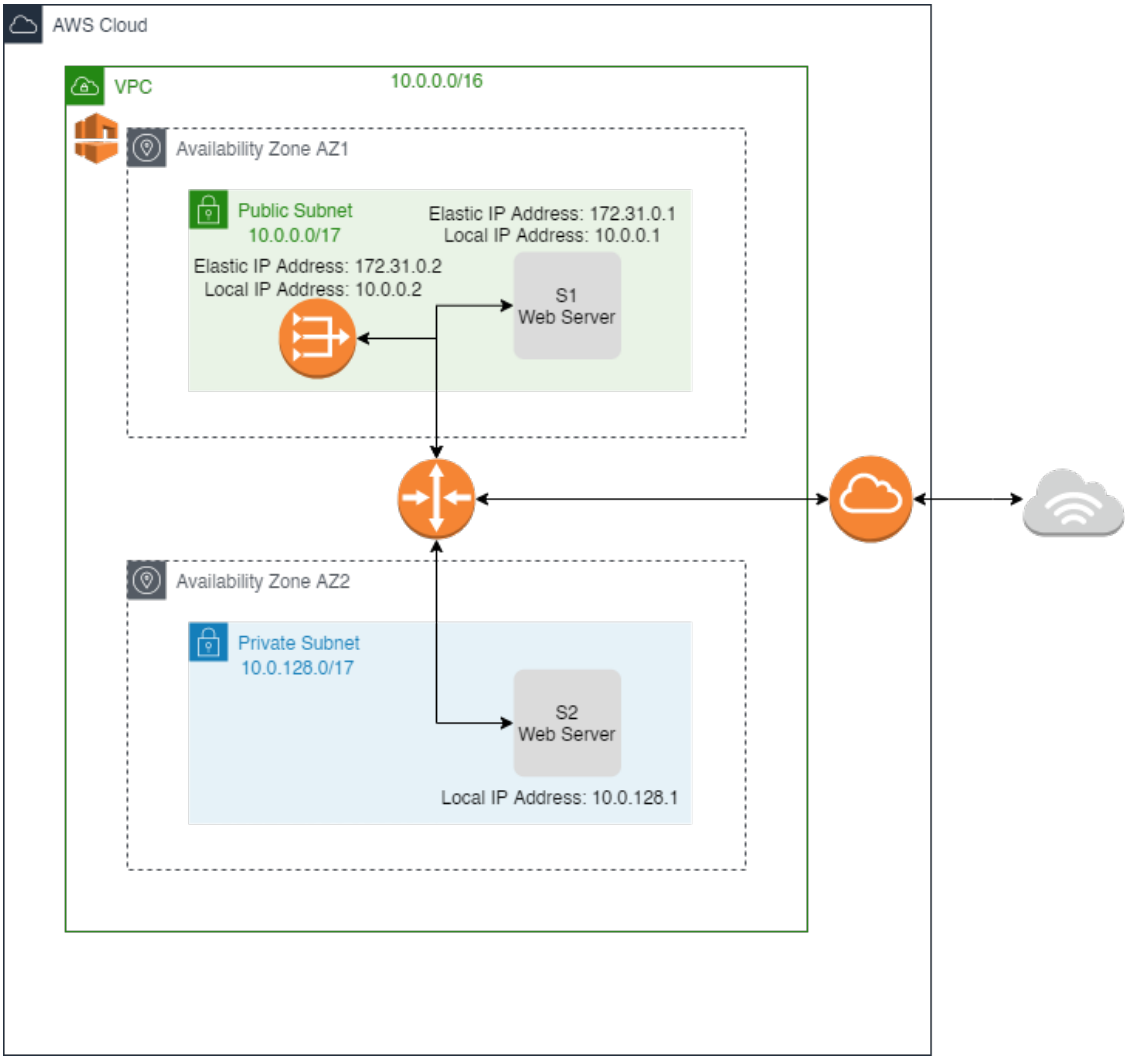


b) The route table should contain the following information:

| Destination | Target |
|---|---|
| 10.0.0.0/16 | Local |
| 0.0.0.0/0 | igw-id |

The first entry is to allow instances in the VPC to communicate with each other, while the second entry is to allow all other traffic to go through the IGW (Internet Gateway).

c) A NAT Gateway needs to be added to the Public Subnet in AZ1, and an Elastic IP Address needs to be allocated to the NAT Gateway as such:




For the Public Subnet, it should still have the same route table:

| Destination | Target |
|---|---|
| 10.0.0.0/16 | Local |
| 0.0.0.0/0 | igw-id |

Meanwhile, for the Private Subnet, it should have this route table instead:

| Destination | Target |
|---|---|
| 10.0.0.0/16 | Local |
| 0.0.0.0/0 | nat-gateway-id |

**-- END --**