**Summary of my understanding for implementation**

* There are 2 types of event logs one which has one complete timestamp and the other which has 2 times stamps namely start and end times stamps.
* We have the following weakness to be identified in these event logs.

|  |
| --- |
| 1. Unwanted Activity |
| 1. Back-loop |
| 1. Redundant Activity |
| 1. Interface |
| 1. Switch of Media |
| 1. Idle Time |
| 1. Variance of Process Time |
| 1. Bottle Neck |
| 1. Parallelizable tasks |

So having a closer look at event log we can see the components which consists of event log.

Application, table, Excel

Description automatically generated

Our main consideration are the first 5 columns where we can classify the data into certain levels:

1. **Case level**: All the rows with same CaseID
2. **Activity level**: All the rows with same activity name
3. **Event Level**: Individual rows(events are the instances of the activity)
4. **Log level**: Whole event log has some characteristics which are captured at this level.
5. **Trace level**: All the rows with same sequence of activities between start and end activities.
6. Process level?

Let’s consider the hierarchy of these levels:

Traces(Unique)

Log

Collection of Instances

Collection of Instances

Collection of Events

Collection of Cases

Activities(Unique)

Events

Cases

We can now associate the 9 weaknesses mentioned to the different levels discussed above based on at what level we can abstract the weakness.

1. Unwanted Activity:

By definition it is the unwanted activities. Since we can point out the occurrence of the activity which is an event and all the details of this event can be obtained like timestamp of occurrence , we can classify this as **Event level**

1. Back-loop :

Sequence of several activities in a loop. This loop can be best found out at **case level**. As anything below it will be just event and here we have sequence of events. This can also be a trace level but we have more details at case level.

1. Redundant Activity:

Repeating sequence of several activities. These sequences can also be best found out at **case level**. As anything below it will be just event and here we have sequence of events. This can also be a trace level but we have more details at case level.

1. Interface:

Resource change within process flow. We can identify the event at which the resource changes. So this can be considered as an **event level** weakness.

1. Switch of Media:

Media change within process flow. We can identify the event at which the media changes. So this can be considered as an **event level** weakness.

1. Idle Time:

Idle time is the time between an end of an event and start of next event. Since this time can be calculated for the start of every event we can consider this as **event level** weakness.

1. Variance of Process Time:

(Min, Max, Average, Variance) for each activity. Since this can be calculated for every activity it must be an **Activity level** weakness.

1. Bottle Neck:

Activity with longest cycle time. Since we have to compare the time duration of all events and give out the max time duration we must consider the entire log. So we can consider this as **Log level** weakness.

1. Parallelizable tasks:

We can find out parallelisable task for each case and hence classify this as a **case level** weakness. This can also be a trace level but we have more details at case level.

**What are the additional potential weakness we can find out in different abstraction level:**

1. Event level:
2. Activity Level:
3. Case level:
4. Trace Level:
5. Log Level:

**Additional Weakness:**

**Time related:**

1. We can find out on which day all the trace instances(Cases) take a longer time or shorter time than usual and hence try to figure out the underlying conditions which lead to this situation.
2. We can compare the event durations from different resources and hence find out the weakest or strongest resource. The same with media.
3. We can take the daily/shift wise list of completed events and check how many are pending for the next day/shift and check subtract the waiting time from duration of that event.
4. ~~We can collect the data from the experts or production in-charge of what is reason in case of delays and unlikely changes in the environment. This can be used to predict the reason in case if similar delays occurs.~~
5. ~~Find out the resource or machine idle times and schedule better.~~
6. Grouping the resulting data frame according to levels.