Unit-2

Date: / /
Page:

Five fundamental process of data science :

① Start with a What-if question -
- Decide what you want to know.
- If you use subset of data lake, which is good start.

② Create a hypothesis :-
- If you want to explore your data in a pattern, use you experience or insights/Intellegence.

③ Gather observation and use in hypothesis -
- Hypothesis is used as a starting point on the basis of limited evidence.

④ Verify the hypothesis (by using real-world evidence) -
- Now, verify your hypothesis by comparing with real-world evidence.

⑤ Promptly (तुरंत) & regularly collaboration -
- promptly & regularly collaborate with customer & subject matter expert.

Six Functions of Operational Managment layer -

① Management of data processing stream -
- manage the correct execution order a/q to workflow designed by data scientist.

② Processing parameters are stored -
- All the parameters are stored in a single location & called them when needed.

③ Overall Process scheduling -
- Plan the complete scheduling for entire system.

Unit-2

Date : / /
Page:

Five fundamental process of data science :

① Start with a What-if question -
- Decide what you want to know.
- If you use subset of data lake, which is good start.

② Create a hypothesis :-
- If you want to explore your data in a pattern, use you experience or insights/Intellegence.

③ Gather observation and use in hypothesis -
- Hypothesis is used as a starting point on the basis of limited evidence.

④ Verify the hypothesis (by using real-world evidence) -
- Now, verify your hypothesis by comparing with real-world evidence.

⑤ Promptly (जल्दी) & regularly collaboration -
- promptly & regularly collaborate with customer & subject matter expert.

Six Functions of Operational Mangment layer -

① Management of data processing stream -
- manage the correct execution order a/q to workflow designed by data scientist.

② Processing parameters are stored -
- All the parameters are stored in a single location & called them when needed.

③ Overall Proces scheduling -
- Plan the complete scheduling for entire system.

④ Overall Process monitoring -
- There is a single unified view of the complete system

⑤ Overall Communication -
- Any activities that are happening are communicated with the system.

⑥ Overall alerting -
- Inform the correct status of the complete system, at correct time, to the correct person.

⇒ #One of the Process Scheduling method is-

Drum-Buffer-Rope method :
- It Identify the slowest process among entire system.
- Then control the processing speed.
- Bind all the process for fast processing.

Drum :- Drum is placed at the slow part of the pipeline.
Rope :- Rope attached/Bind all the process together form beginning to end of the pipeline.

This method ensure that all the process complete more efficiently.

These are 5 Watchers for each logging locations/
or, Good Indicators for Audit purpose :
or, Different ways of implementing the built-in logging ;

(1) Debug Watcher - [Capture every relevent log entry in one place]
- If any debug logs discovered in ecosystem, it should raise an Alarm.

(2) Information Watcher -
- Watch logs information ; and notify - that is fanificial to the running & management of system.

(3) Warning Watcher -
- Used for exceptions handling
- And other important log events.

(4) Error Watcher -
- When error found, ecosystem handle the issue
- And take the necessary action for recovery.

(5) Fatal Watcher -
- Identify the event cause immediately.

Different ways of implementing the Basic-logging in the Audit phase.

[Allows to log everything that occurs into a central file]

(1) Process Tracking -
- Monitors system hardware like -
  → voltage, fan speed, sensors temperature clock speed.

(2) Data Provenance -
- System track every data entry, & generate a record.
- (1) Reproduced the data (2)
  (2) supply a detailed history of data.

(3) Data Lineage -
- Keeping record of every changes in data value in data lake.