

Q6.

Another processing service is DataPrep

Differences

A few differences include that the automation of clusters are proposed by DataPrep and Dataflow, whereas on the other hand manual setups are heavily required for DataProc.

Advantages

A few advantages include improved data visualization, higher security and protection, as well as utilized machine learning for reductions of unimportant data. This helps keeping the size minimal.

Disadvantage:

It is slightly more complex to apply/use. It also is not completely integrated with Apache Hadoop as the processing service.

Limitations:

There is a limit of 1000 datasets that are able to be stored within the user interface. However, there is only a minimal amount of memory size that can be used and is shorthanded for storing these datasets as well. Another extremely important thing is that it is not UTF-32 compatible.

Q7.

The Application:

- An application of this could be used in perhaps an ATM machine, as stream processing would be the act of withdrawing or depositing, whereas the batch processing would be finding the user of the atm through the database of the entire bank franchise. We can implement the LIDAR dataset could be applied here as when more than one user is detected in the camera of the ATM, the ATM machine locks out and returns the card cancelling any transaction. Or if money was deposited, it will be deposited and then the card will instantly be returned, or perhaps kept for safety measures.

Its Impact:

- It will allow for a more fluent integrated system. A team already is dedicated to the ATMs therefore it would be still using a limited amount of resources (in this

case; workers). Real-time operating systems are great however they are extremely difficult to implement and therefore difficult to troubleshoot with newer employees or typically just people who are new to the organization of the system.

In terms of the object detection dataset that we are implementing, it would prove useful as it would allow for victims to be safe from robberies. It would try to detect any malicious intent and instantaneously react accordingly - by blocking out any withdrawals and perhaps even withholding the card until a set time frame has passed i.e.. 15 minutes.

Datasets:

- The dataset used is the LIDAR dataset called NCLT dataset, along side the long-term vision. The latter aids in the object detection area, which can be configured to notice more than one person within a camera shot at once.

Tools:

- Tools for this would include the ATM health monitoring system, which is currently a tool monitoring ATM health status, cash, consumable levels, alerts, warnings, and information for the management of ATMS.
- HDFS also could be used to keep wanted data used by the application, such as bank balances and withdrawals limits, etc.
- Starforce solutions also provides ATM security as well as ATM source code protection, against decompilation, disassembly, and reverse engineering.
- TPSecure also helps in the protection of ATMS and payment terminals that run against unauthorized programs as well as connecting with external devices.