**Talend Demo (6th ETL Demo)**

June 08, 2017, 3:00-4:00 pm EST

Discussion Meeting:

June 08, 2017, 5:00 – 5:15 pm EST

PCORTF Team Comments:

**Advantages:**

* Platform for data agility (adopt to changes in data quickly)
* One unified platform
* Native tool generation code (sometimes changing the code is one click away).
* Open source platform
* Ties to Apache community (e.g. Spark)
* Predicable environment (priced is per developers only), no cost for CPU or run time.
* Web UI (monitor job execution)
* Has ESB capability by integrating to another Talend application
* You use Talend Studio, which has GUI and it generates codes on the back end. The language is standard in java and can be run anywhere.
* 900 Connectors (for databases, etc.) if not on the list, use JDBC connector.
* Matching capability to find duplicates in the data
* Data preparation capability

**Disadvantages:**

* Not able to import existing mappings into Talend. (Maybe none of the ETL tools have this capability).
* Complicated to use.

**Discussion:**

* Eclipse-based Software Development Kit (SDK) environment. They compile an .exe to run, which increases the efficiency and the exe file can be provided behind the firewall to the data partners of the 4 networks.
  + There is a step to describe the mapping and that is saved as a file, and the second step is to run the input file to generate the output file. It does not require Talend Studio to run it in data partners’ site. The .exe can be run in any environment that uses any of the 4 CDMs.

Q: Are we expecting researchers to use this tool?

A: Did not come away from Talend, it might be complicated.

* Ken mentioned that if you take the Gardner magic quadrant, both these 2 tools (Pentaho and Talend) made it to the magic quadrant.

**Questions for Talend ETL Vendor:**

* Mapping between fields were not demoed today. Can it be demoed to us?
* Does your tool provide mapping across terminologies and value sets?
* Can one query be transformed into 4 different queries for 4 CDMs?