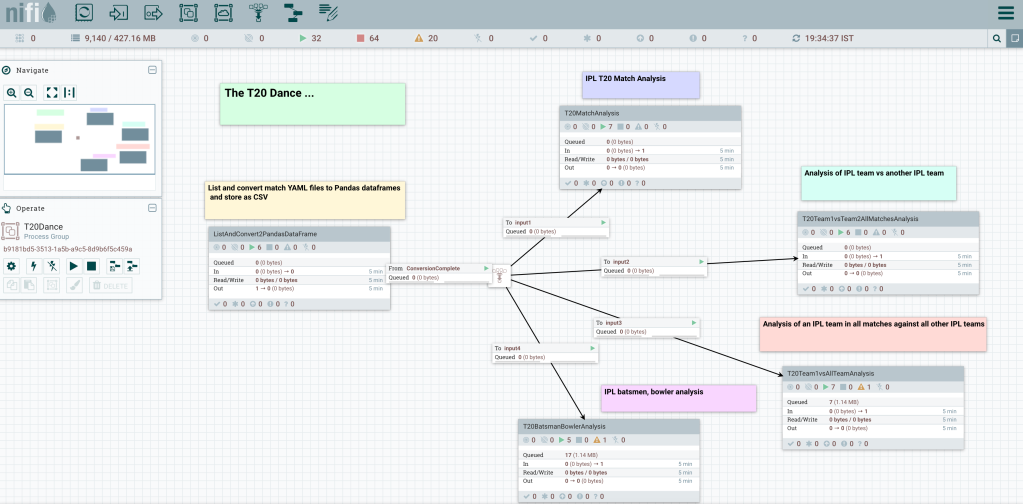
All the data are taken from the R packages yorkpy and Cricsheet. In that post I had mentioned that we could create a similar pipeline to create a real time dashboard of IPL Analytics. I could have have done this but I needed to know how to create a Web UI. After digging and poking around, I have been able to create a simple Web UI running off Apache Web server. This UI uses basic JQuery and CSS to display a real time IPL T20 dashboard. As in my previous post, this is an end-2-end Big Data pipeline which can handle large data sets at scheduled times, process them and generate real time dashboards.

We could imagine an inter-galactic T20 championship league where T20 data comes in every hour or sooner and we need to perform analytics to see if us earthlings are any better than people with pointy heads  or little green men. The NiFi pipeline could be used as-is, however the yorkpy package would have to be rewritten in Pyspark. That is in another eon, though.

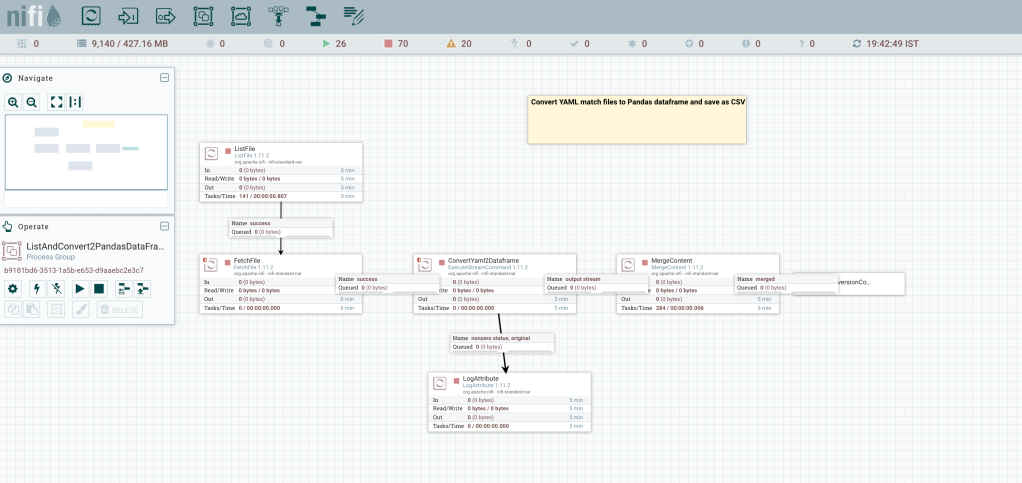
**1. T20 Dance – Overall NiFi Pipeline**



There are 5 process groups

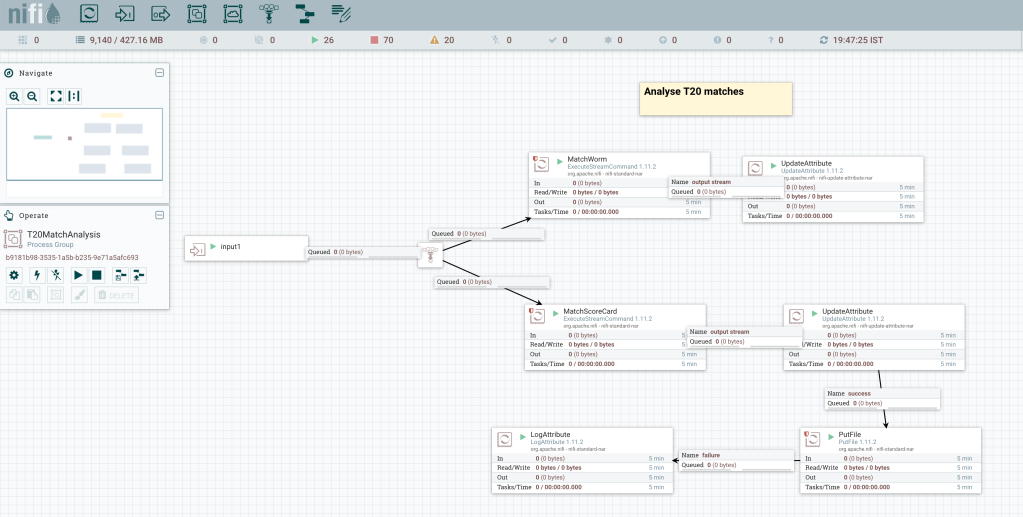
**2. ListAndConvertYaml2DataFrames**

This post starts with having the YAML files downloaded and unpacked from Cricsheet.  The individual YAML files are converted into Pandas dataframes and saved as CSV. A concurrency of 12 is used to increase performance and process YAML files in parallel. The processor MergeContent creates a merged content to signal the completion of conversion and triggers the other Process Groups through a funnel.



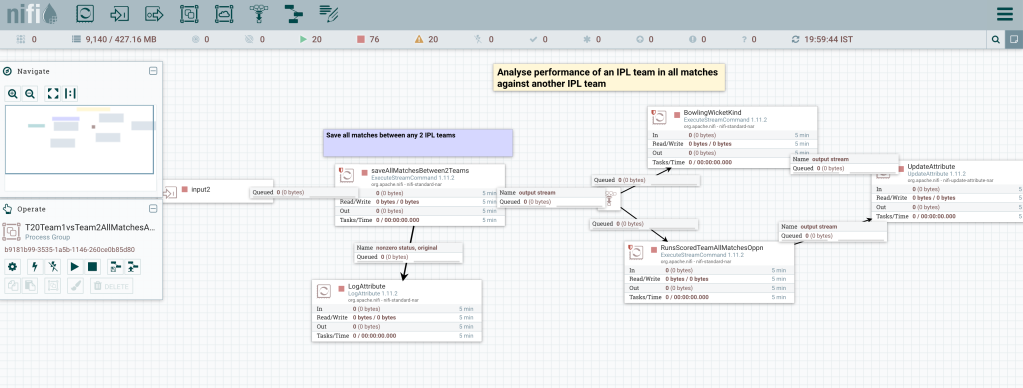
**3. Analyse individual IPL T20 matches**

This Process Group ‘**Analyse T20 matches’**  used the yorkpy’s Class 1 functions which can perform analysis of individual IPL T20 matches. The matchWorm() and matchScorecard() functions are used, through any other function could have been used. The Process Group is shown below



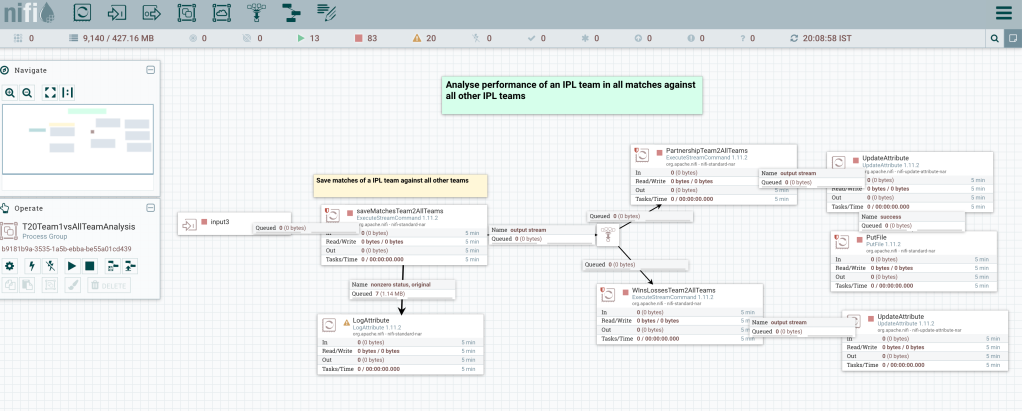
**4. Analyse performance of an IPL team in all matches against another IPL team**

This Process Group ‘**Analyse performance of IPL team in all matched against another IPL team**‘ does analysis in all matches between any 2 IPL teams (Class 2) as shown below



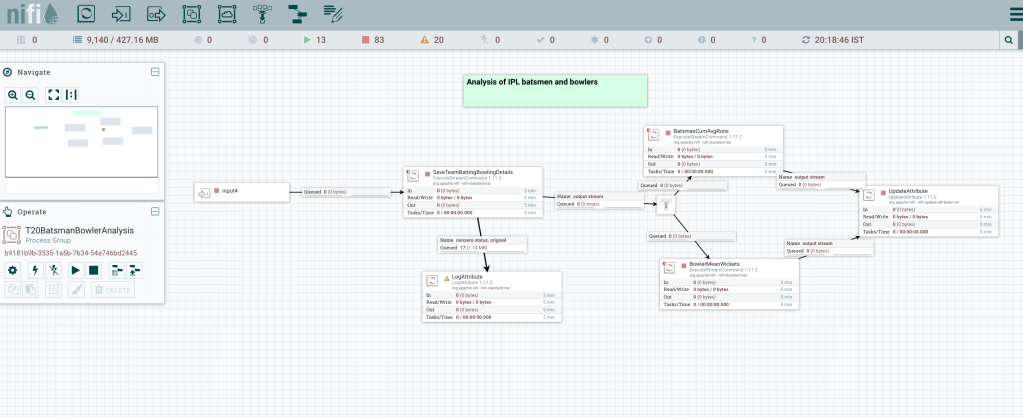
**5. Analyse performance of IPL team in all matches against all other IPL teams**

This uses Class 3 functions. Individual data sets for each IPL team versus all other IPL teams is created before Class 3 yorkpy functions are invoked. This is included below



**6. Analyse performances of IPL batsmen and bowlers**

This Process Group uses Class 4 yorkpy functions. The match CSV files are processed to get batting and bowling details before calling the individual functions as shown below



**7. IPL T20 Dashboard**

The IPL T20 Dashboard is shown

