*I don’t count my sit-ups. I only start counting once it starts hurting. ”*

Muhammad Ali

*“Hard work beats talent when talent doesn’t work hard.”*

Tim Notke

In my previous post Big Data 5: kNiFI-ing through cricket data with Apache NiFi and [yorkpy](https://pypi.org/project/yorkpy/), I created a Big Data Pipeline that takes raw data in YAML format from a [Cricsheet](https://cricsheet.org/) to processing and ranking IPL T20 players. 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.

My package yorkpy has around ~45+ functions which fall in the following main categories

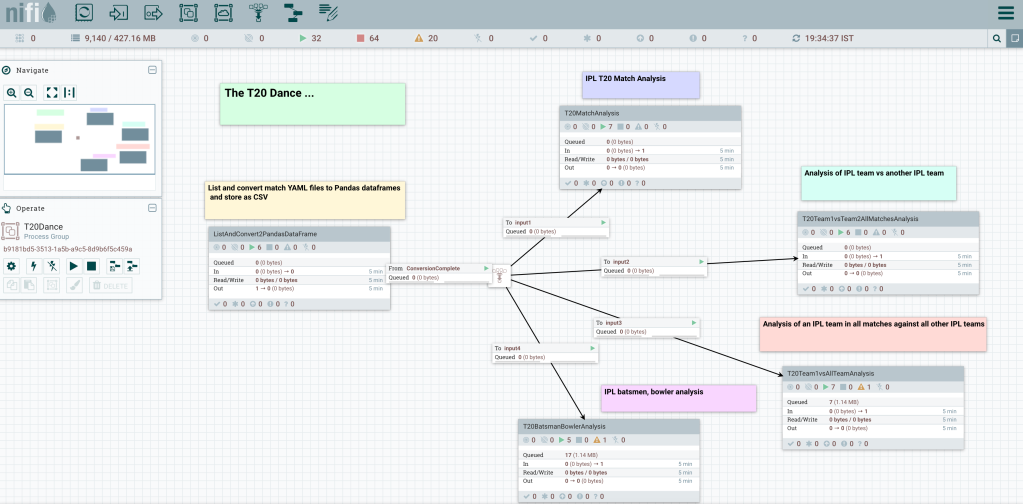
1. [Pitching yorkpy . short of good length to IPL – Part 1](https://gigadom.in/2018/12/28/pitching-yorkpy-short-of-good-length-to-ipl-part-1/) :**Class 1**: This includes functions that convert the yaml data of IPL matches into Pandas dataframe which are then saved as CSV. This part can perform analysis of individual IPL matches.  
2. [Pitching yorkpy.on the middle and outside off-stump to IPL – Part 2](https://gigadom.in/2019/01/27/pitching-yorkpyon-the-middle-and-outside-off-stump-to-ipl-part-2/) :**Class 2**:This part includes functions to create a large data frame for head-to-head confrontation between any 2IPL teams says CSK-MI, DD-KKR etc, which can be saved as CSV. Analysis is then performed on these team-2-team confrontations.  
3. [Pitching yorkpy.swinging away from the leg stump to IPL – Part 3](https://gigadom.in/2019/02/03/pitching-yorkpyswinging-away-from-the-leg-stump-to-ipl-part-3/) **Class 3**:The 3rd part includes the performance of any IPL team against all other IPL teams. The data can also be saved as CSV.  
4. [Pitching yorkpy … in the block hole – Part 4](https://gigadom.in/2019/02/26/pitching-yorkpy-in-the-block-hole-part-4/) :**Class 4**: This part performs analysis of individual IPL batsmen and bowlers

Watch the live demo of the end-2-end NiFi pipeline at ‘[The T20 Dance](https://www.youtube.com/watch?v=s_jnqIyOn-8)‘

You can download the NiFi template and associated code from Github at  [T20 Dance](https://github.com/tvganesh/T20Dance)

The Apache NiFi Pipeline is shown below

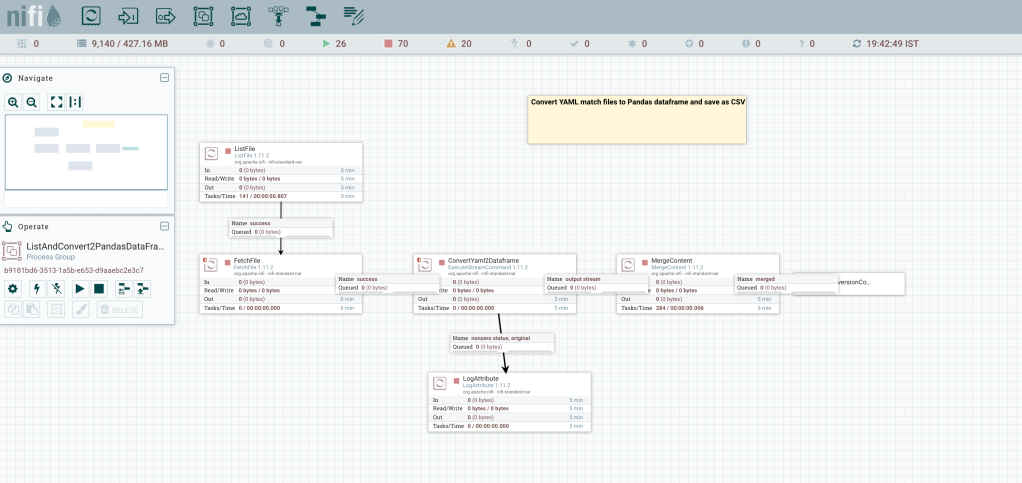
**1. T20 Dance – Overall NiFi Pipeline**



There are 5 process groups

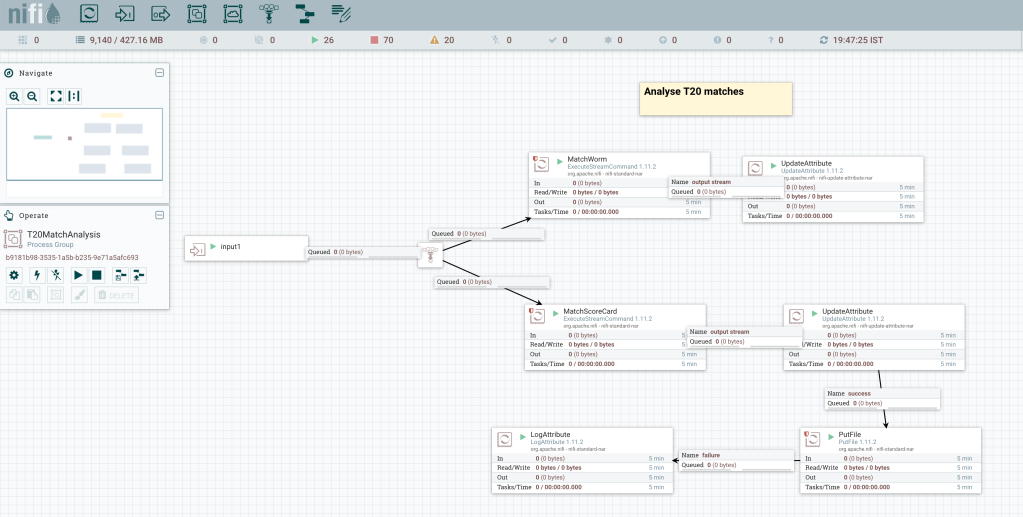
**2. ListAndConvertYaml2DataFrames**

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.



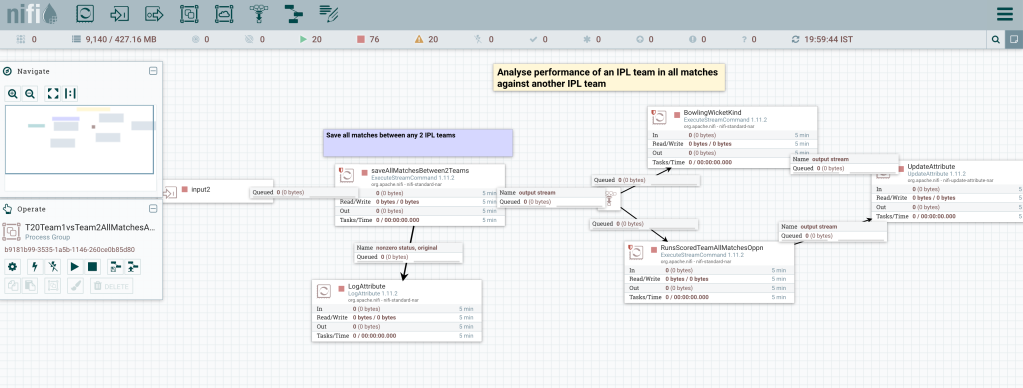
**3. Analyse individual IPL T20 matches**

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



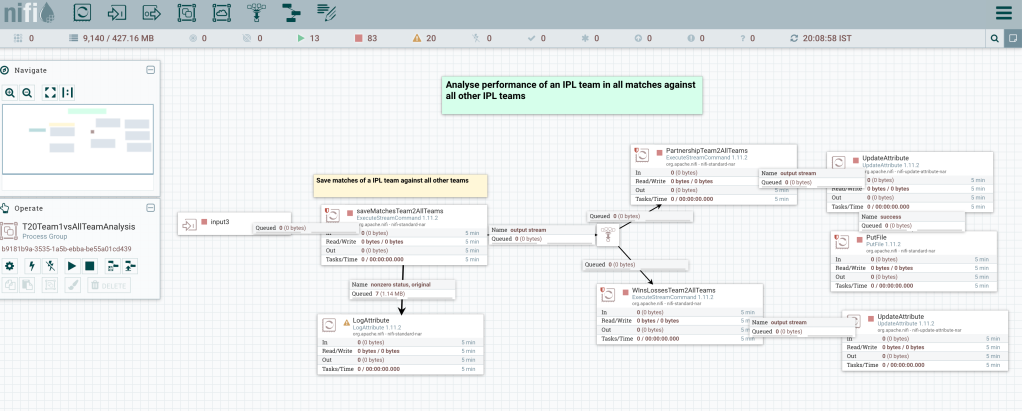
**4. Analyse performance of an IPL team in all matches against another IPL team**

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



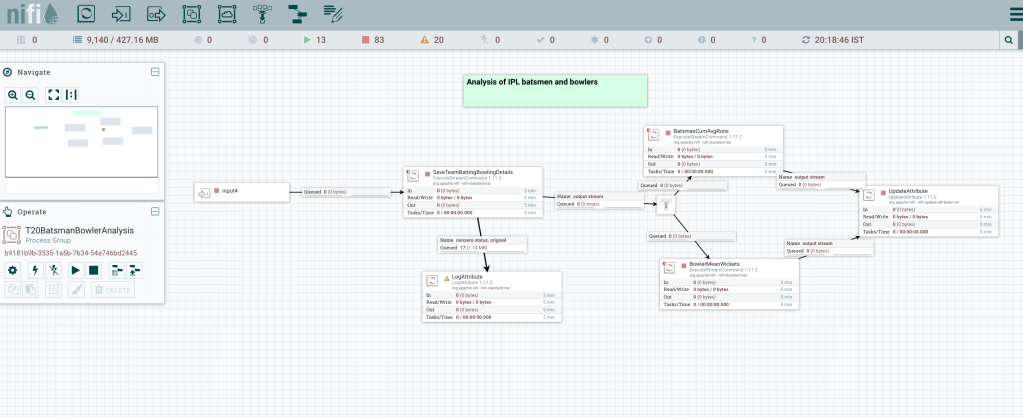
**5. Analyse performance of IPL team in all matches against all other IPL teams**

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

