Application developer – Risk and Mid office

Python, pandas, numpy, pl/sql, elastic search (ELK), redis, kafka, tableau, finance, Java

ETL, real-time, distributed processing systems

Browers based, FRM

Enso is our brand name for a web-based application, we help hedge fund clients to perform ETL and data analytics, provide centralized view of trade position on our web portal or reports.

Working mainly as a backend Python developer here at Enso, Some of the projects I am working on are Data Integration (Extra and Load data from data source to our DB)

Web Analytics, build calculations in Pandas, will serves as back-end for web-portal or reports.

I spend around 90% time coding in python and 10% time in JS and Java. Main skill sets I use daily are Python, SQL, Pandas and Linux.

Question:

What kind of data feed do you have? What is structure / size of data, how do you load it into your system. Sounds like the dataset is big and you are using a distributed system, is it HDFS?

Also looks like you guys are doing some real time processing, I guess you guys are using Hadoop and Spark for batched loading and real time loading, all these frameworks are written in JVM languages, are you guys also using Java or scala? or you just use the python interface such as pyspark? Have you migrate to the cloud (AWS)

Can you introduce the flow control system? Is it a crontab based scheduling system or it is built on top of frameworks such as airflow or luigi?

Do you use a lot of open-source software?