Python Assignment

3. PaaS, SaaS and IaaS

SaaS (Software as a Service) is a cloud application service for users to accessed application through website (no need to download or installation), and users could pay the service as monthly subscription. SaaS is a quick option for some small projects, which is good for startups or small business. Google Doc is a good example!

PaaS (Platform as a Service) is a cloud platform service for developers to build cloud components for their software, and multiple developers could manage the application components through the framework, which is provided by the third-party vendor. AWS Elastic Beanstalk is a good example.

IaaS (Infrastructure as a Service) is Cloud-based Pay-as-you-go substructure service (business could decide how much service they want and only pay for the service they use). IaaS provide highly scalable and automated compute resources. Computing, networking, storage, and other services can be accessed and monitored through the SaaS model. Amazon EC2 is the most popular IaaS.

4. ETL and ELT

Current data teams are processing to migrate from ETL to ELT. The mainly difference between those two is the way raw data is managed that distinguishes these two approaches. ETL moves data from the data source through staging to the warehouse before performing all transformations. ELT is a modern alternative to ETL where analysts load their data into a warehouse before it is transformed, allowing them to work more agilely and efficiently.

	Pros	Cons
ETL	• good for bulk data movements with complex rules and transformations • make maintenance and traceability much easier than hand-coding • good for data warehouse environment	• you must be a data-oriented developer or database analyst to use • not ideal for near real-time or on-demand data access, where fast response is required • takes months to put into place • difficult to keep up with changing requirements
ELT	• Faster loading from the original, raw data format • You can transform data at runtime, without starting a new process or project, which reduces administration time and collaboration time • Can be easier to automate • Uses the popular data lake structure • Eliminates the main data staging part of the process	• fairly new, there is not as much documentation • overall process could be more expensive, which might not be suitable for all companies