**Cloud Analytics Notes**

**Cloud Foundation Track**

**301 In the Cloud: ELT vs. ETL:**

*Exam:*

* A viable approach to data manipulation is using an intermediate database and SQL to transform raw data before it is appended to a read-only warehouse
* Idempotence and uniqueness constraints are easier to achiever with ETL than with ELT
* Backfilling aggregate data can be easier in an ELT data process if…
  + Base data is preserved inside the warehouse with due retention
  + Base data is semi-structured
* ELT approach…
  + Makes retention policies more transparent
  + Retention policies more flexible
* Compared to ETL, ELT warehouses often require…
  + More awareness and contextual knowledge among the user base
  + More database views
* ELT improves turnaround time for getting analysts connected to new data sources
* ELT is a way to bring simplicity, expressiveness, and transparency to your data warehouse
* ETL is a vital part of most mature data processing systems

*Video:*

* ETL
  + Extract, Transform, Load
  + Requires staging area, means it needs extra time to load data
  + As data volume increases, so does transformation time
  + Choose which data to load and transform upfront, and then do it again to do backfill
  + Requires less space and a cleaner warehouse
  + Multiple processes to create transformations along the way
  + Retention and data limits are determined before processes
  + Geared towards on-premise and relational structured data
  + Not cost-effective for small and medium sized businesses
* ELT
  + Extract, Load, and Transform
  + Single system, data loaded immediately only once
  + Staging and production occurs in the same system
    - Speed is not directly related to data size
  + Backfill is readily available as it already has the raw data
  + Requires in-depth knowledge of tools and expert design of the depository
  + Low-cost for building and maintaining while analysts have access to the raw and transformed data
  + Retention and data limits are determined by hardware and policies, these policies can be changed at any time at the warehouse level
  + Scalable cloud infrastructure to handle structured, semi-structured, and unstructured data sources