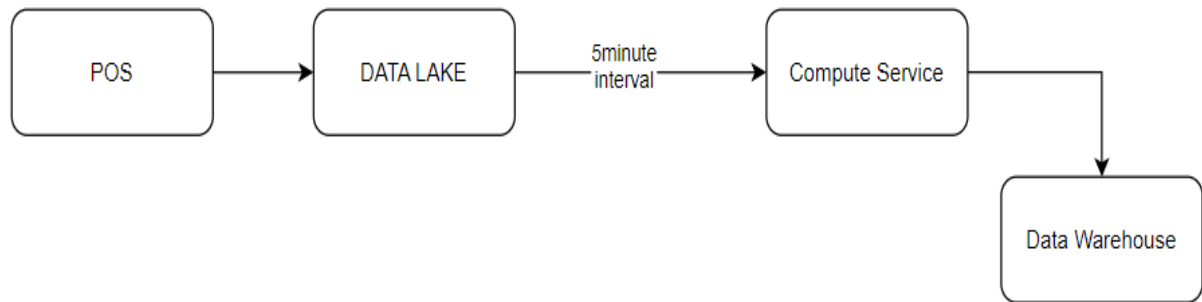Current System Architecture

As mentioned; data flow from POS to the data lake which gets loaded into the data warehouse every 5 minutes via a scheduled job (compute service). Realistically this service does some transformation/logic which loads staging tables/reporting tables and then updates information in the warehouse.

Concerns:
- Ideally the warehouse should avoid as many write transactions as possible as OLAP is different from OLTP databases.
- 5 minutes interval to trigger jobs can be overload with multiples runs could have timeout issues as mentioned
- Low latency is important here as order requests within 5 minutes could have wrong information i.e. strong consistency vs eventual consistency

New Approach
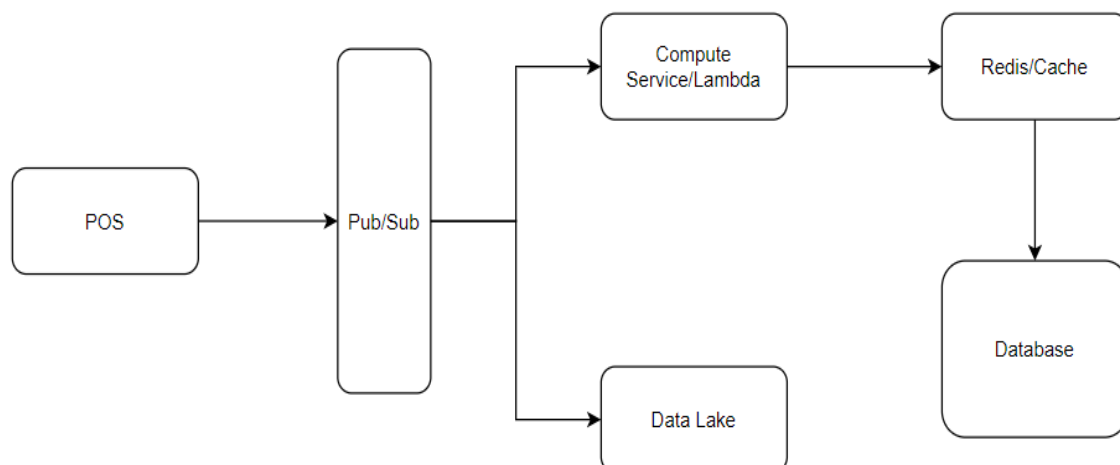
Benefits
- Real time event driven approach to getting data across to the systems quicker

- Using a event bus or pub/sub systems means there can be multiple consumers such as the data lake and the compute without including time to store the data and get the job to perform logic on the data
- Having Redis/Cache can reduce workload on the database regarding write updates from the compute service and can also improve consistency of inventory data.
- Using a database rather than a traditional warehouse can improve OLTP access as they are usually optimized for it.

Cons
- Adds an extra layer of complexity, working with distributed systems can be hard.
- Possibility of duplicates in messages, however there are ways to combat that using keys and partitions within topics, using redis etc.
- Adding new services can bring extra cost to the overall spend of the platform.