Kushagra Dixit

The problem presented by the Bond challenge question presented the architectural design challenge of a pizza shop with 2000 locations. Some of the key elements that were taking into consideration are listed below:

- Transactions that are online and POS
- ANY Online transaction is streaming data
- POS are 15 min synced Batch transactions
- Logistical/inventory control
- Running analytics to perform analysis like time-series, projections and other Business analysis.
- Varying data loads within hundreds of gigs/ month that may increase during special sale days

To create an architecture, I chose Google Cloud Platform (GCP) for this problem as I have done cloud architect certificates within this cloud platform. For a quick reference, the diagram below showcases a very high-level skeletal view of the architectural implement that can be used to solve this problem.
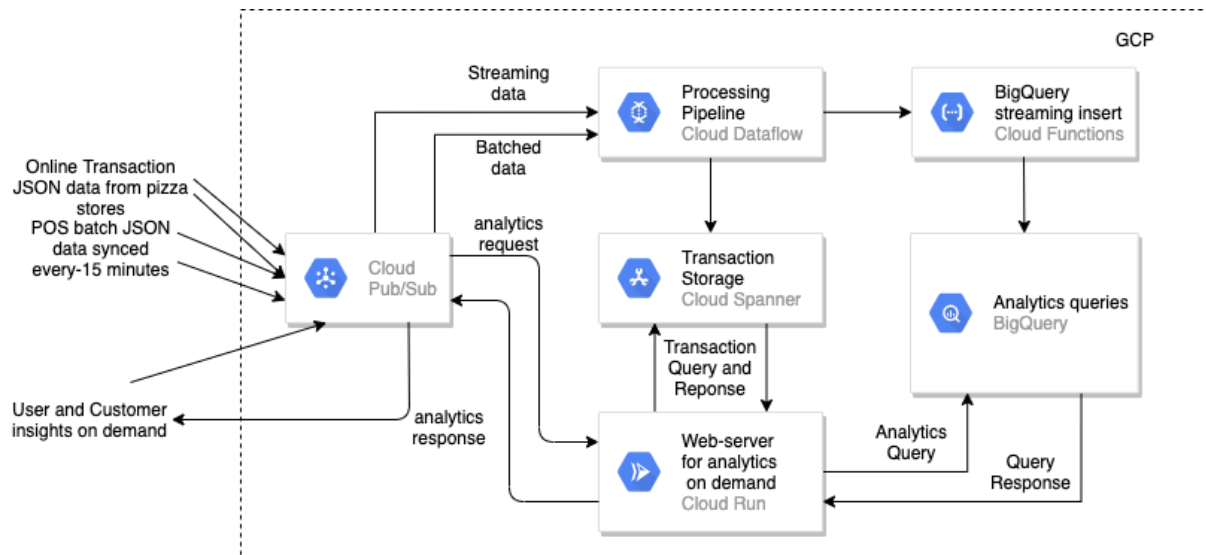
The flow will be explained as a sequence of the following steps:

**Step-1-Pub/Sub(Big-Data reception and redistribution)**

Pizza stores use an in-store app that behaves as an API that sends POS transactions as **batched JSONs** every 15 mins. This is coupled with online transactions that stream the transaction information as a JSON to a Pub/Sub topic. Pub/Sub enables creation of systems of event producers and consumers, called publishers and subscribers. Pub/Sub uses 2 separate topics 1 for streaming data and batched data that publish these JSON messages to subscribers. Pub/Sub allows messages to be received asynchronously and is
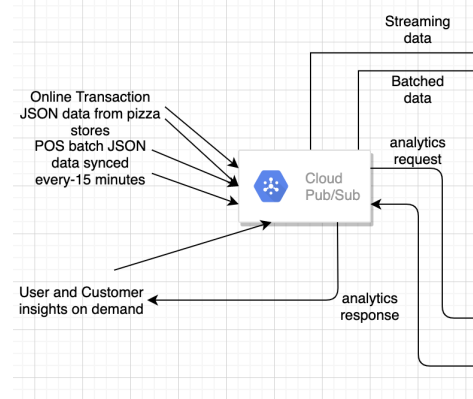
commonly used for streaming and data integration pipelines and can also be used as message oriented middleware between various services.

Step-2 Dataflow(Pre-processing pipeline for data storage and warehousing)

The subscribers in this case is a cloud dataflow service account that receives the published JSON messages. Dataflow allows for unified stream and batch processing that's serverless, fast and cost-effective. It has automated provisioning and management of processing resources and allows for horizontal autoscaling of worker resources to maximize resource utilization. A Dataflow job allows for data pre-processing to ensure that data around transaction processing is sent to cloud Spanner and the data for business/customer related analytics is transferred for appending on to BigQuery.
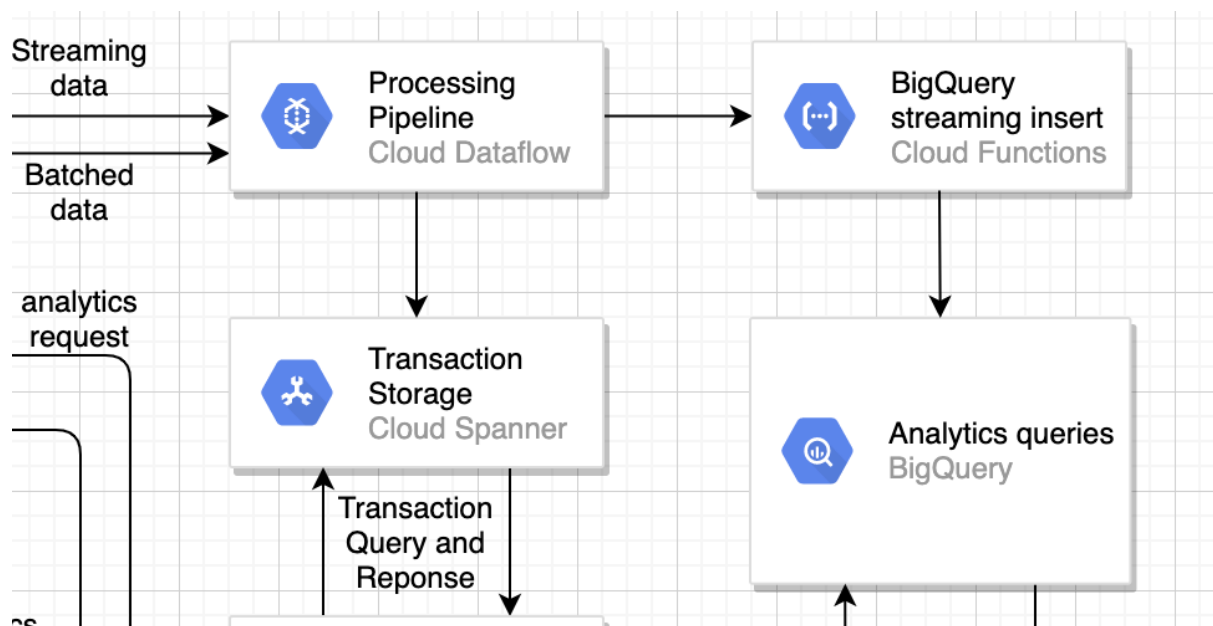
**Step-3a-Cloud Spanner (transaction processing RDBMS)**

Cloud Spanner is a fully managed RDBMS with unlimited scale and has 99.999% availability resulting in very little to almost no downtime. It is used primarily for architecture that needs to be robust enough to work with **varying data loads** as it is highly scalable both horizontally and vertically. It is known for high performance data transactions with strong consistency across regions **allowing for billions of write events per day.** This allows for real time transaction and inventory processing and record keeping. The high availability further allows for on-demand customer analytics and real time consistent updates for pizza shop owners and customers across multiple regions for online transaction processing.

Step-3b- BigQuery (Data warehousing and analytics)

BigQuery is a data warehousing solution designed for businesses. It is highly scalable and has a lot of machine learning, time series and other business analytics based key-features like BigQuery-ML, BigQuery BI engine and BigQuery Omni that allow for rapid

analytics processing of **petabytes of historical data with 0 operational overhead**. This also allows reprocessing of historical data in case of bugs in the pre-processing logic.

The cloud functions streaming insert allows for a faster stream-based integration of data into BigQuery as mass loading of data into BigQuery can result in batch-related delays in BI analytics processing for business owners and customers.
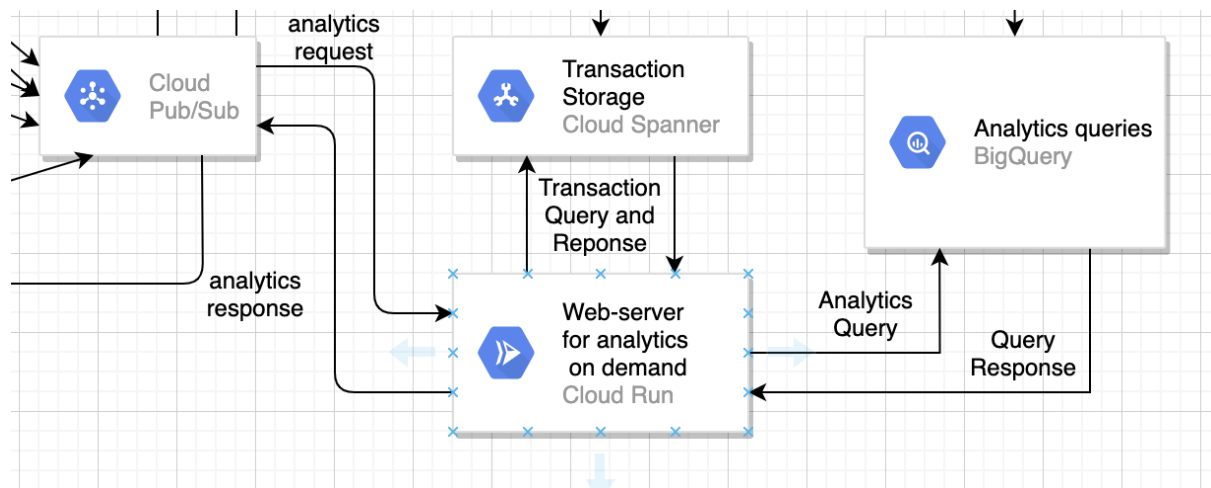
Step-4- Customer/Business on demand analytics

A user/business can request analytics on-demand and hundreds of thousands of these requests can be accepted as JSON by the Pub/Sub topic dedicated to customer analytics. This topic will be subscribed by the Cloud-Run based web-server that can **handle up to a 1 million concurrent requests** on a single worker model. Cloud run is a fully managed platform that allows deployment of containerized applications which can take up to 1000 requests per container and can deploy upto 1000 containers per instance. This application can create a custom query and read the resulting analytics from BigQuery and send a message back to the Pub/Sub topic that will push a response back to the customer using the Application/api to request business analytics. This allows millions of business-owners and customers to request analytics that will result in minimum response latencies.

The above solution is one of many that can be utilized for the specific problem as mentioned by the challenge. There are always ways to improve and optimize the design solution to this problem. This is only a very skeletal high level solution of the problem while taking into consideration the demands and requirements of this infrastructure. This also incorporates default limitations of the GCP products used here. These quotas and limits can be adjusted for businesses by requesting increases for enterprise solutions.

Feel free to get in touch with me and ask me any questions.