Google Cloud BigQuery ML

Demand Forecasting

Aman Rajvardhan, Dev Rao, Jessica Cheng, Ishita Chaudhary, Xiaoxue Liu

BigQuery ML Introduction

- BigQuery ML lets you create and run machine learning (ML) models by using Google SQL queries.

- BigQuery ML also lets you access LLMs and Cloud AI APIs to perform artificial intelligence (AI) tasks. The following models are built in to BigQuery ML:
  - Linear regression
  - Logistic regression
  - K-Means Clustering
  - Principal component analysis (PCA)
  - Matrix factorization
  - Time Series

BigQuery ML Advantages

Top 5

Widely Used and Secure

Ranked in the top 5 for database management warehouses and trusted as a secure platform.

47%

Unified, Faster Performance

47% of BigQuery queries finish in less than 10 seconds compared to 20% for alternative cloud services.
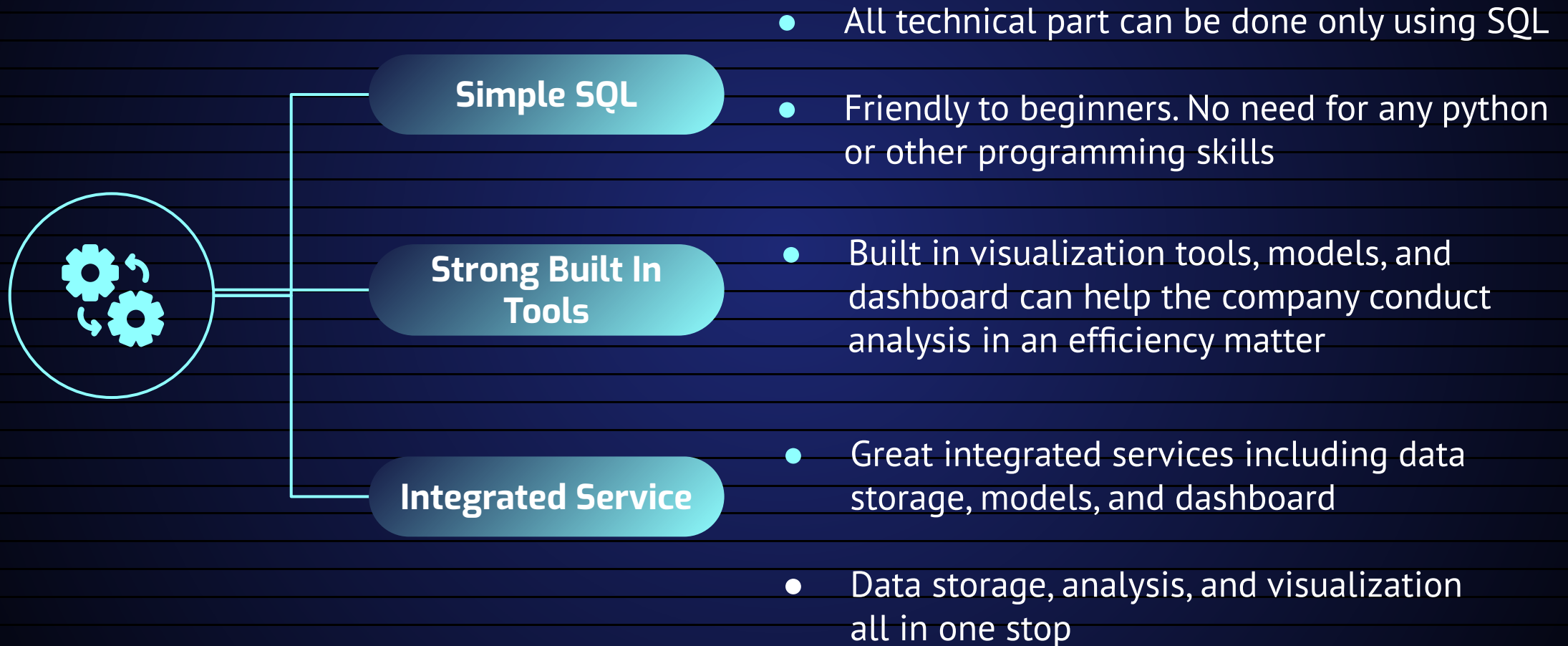
>26%

Reduced Costs

Google Cloud BigQuery's pay-as-you-go structure is 26-34% less costly than competitors.
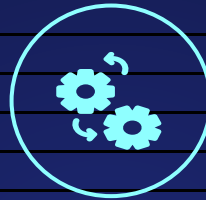
BigQuery ML Business Values

Simple SQL

- All technical part can be done only using SQL

- Friendly to beginners. No need for any python or other programming skills

Strong Built In Tools

- Built in visualization tools, models, and dashboard can help the company conduct analysis in an efficiency matter

Integrated Service

- Great integrated services including data storage, models, and dashboard

- Data storage, analysis, and visualization all in one stop

BigQuery ML (Built in Models) Use Cases

Forecast Demand

Create Product Recommendation System

Identify Potential Customers

Time Series

Matrix Factorization

K-means clustering

Predict product sales to help with inventory management - OUR FOCUS

Using historical customer behavior, transactions, and product rating information to make recommendations

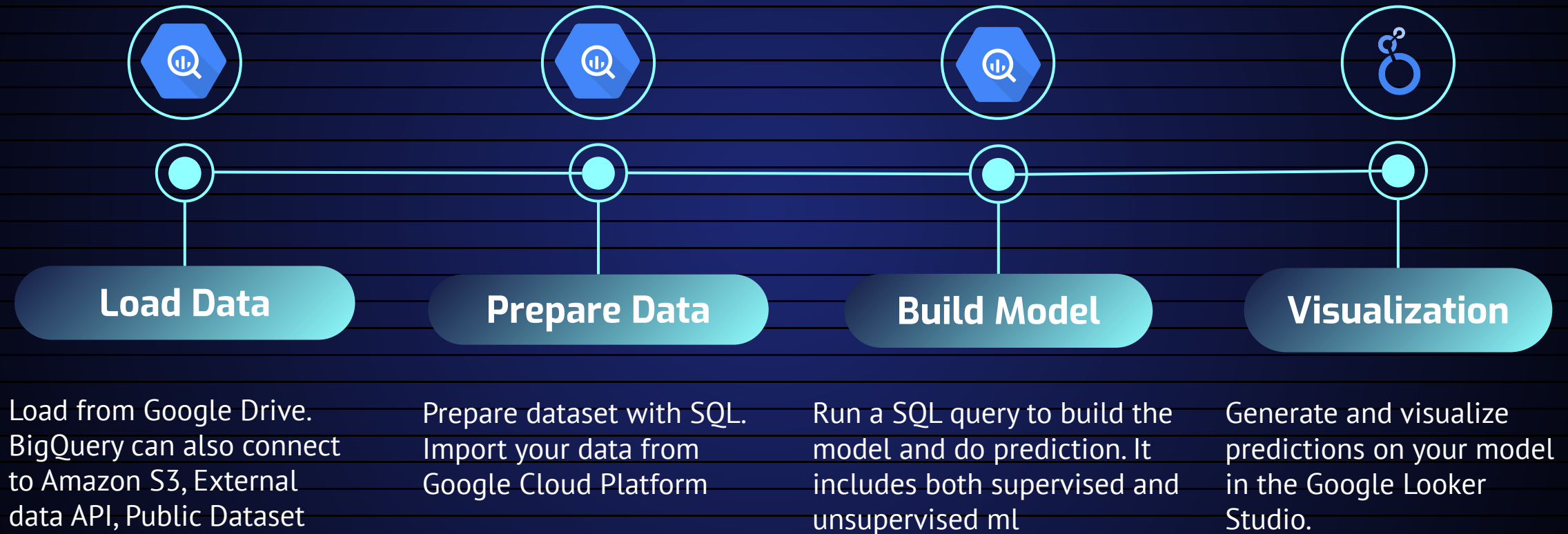Segment customers to identify their characteristics and target potential customers
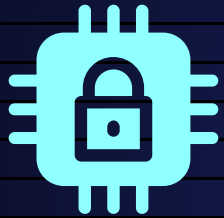
BigQuery ML Workflow

Load Data

Prepare Data

Build Model

Visualization

Load from Google Drive. BigQuery can also connect to Amazon S3, External data API, Public Dataset

Prepare dataset with SQL. Import your data from Google Cloud Platform

Run a SQL query to build the model and do prediction. It includes both supervised and unsupervised ml

Generate and visualize predictions on your model in the Google Looker Studio.

BigQuery Demo: Demand Forecast

Data Source

- The data source is from Kaggle at https://www.kaggle.com/competitions/demand-forecasting-kernels-only/data?select=train.csv

- The data is 5 years of store-item sales data. We utilize the data to predict the product demand.

BigQuery Demo: Demand Forecast

Easy Model Building Using SQL

BigQuery allows us to use its built- in models to train the model of the entire data depending on the problem context

```sql
CREATE OR REPLACE MODEL Store_item.arima_model

OPTIONS(
  MODEL_TYPE='ARIMA',
  TIME_SERIES_TIMESTAMP_COL='date',
  TIME_SERIES_DATA_COL='total_amount_sold',
  TIME_SERIES_ID_COL='store_item_id',
  HOLIDAY_REGION='US'
) AS

WITH store_item_data AS(
  SELECT
    store_no,
    item_name,
    DATE(date) AS date,
    total_amount_sold,
    CONCAT(CAST(store_no AS STRING), "_", item_name) AS store_item_id
  FROM
    `Store_item.train_data`
)

SELECT
    date,
    store_item_id,
    total_amount_sold
FROM
  store_item_data
```

BigQuery Demo: Demand Forecast

Quick Model Prediction

Model Prediction is simple in three lines of codes. Can adjust hyperparameters if needed

```
SELECT *
FROM ML.FORECAST(MODEL Store_item.arima_model,
                 STRUCT(30 AS horizon, 0.9 AS confidence_level))
```

BigQuery Demo: Demand Forecast

MAPE: 43

Efficient Model Evaluation

Model can be evaluated efficiently with simple line of codes

```sql
WITH predictions AS (
  SELECT
    CAST(forecast_timestamp AS DATE) AS date,
    store_item_id,
    forecast_value AS predicted_total_amount_sold
  FROM
    ML.FORECAST(MODEL Store_item.arima_model, STRUCT(30 AS horizon, 0.9 AS confidence_level))
)

SELECT
  AVG(ABS((p.predicted_total_amount_sold - a.total_amount_sold) / a.total_amount_sold)) * 100 AS mape
FROM
  predictions p
JOIN
  `Store_item.true_data` a
ON
  p.store_item_id = a.store_item_id AND p.date = a.date;
```

BigQuery Demo: Demand Forecast

Simple and Straightforward Visualization

Looker Studio allows simple visualization and has the ability to compare both the historical and forecasted data together
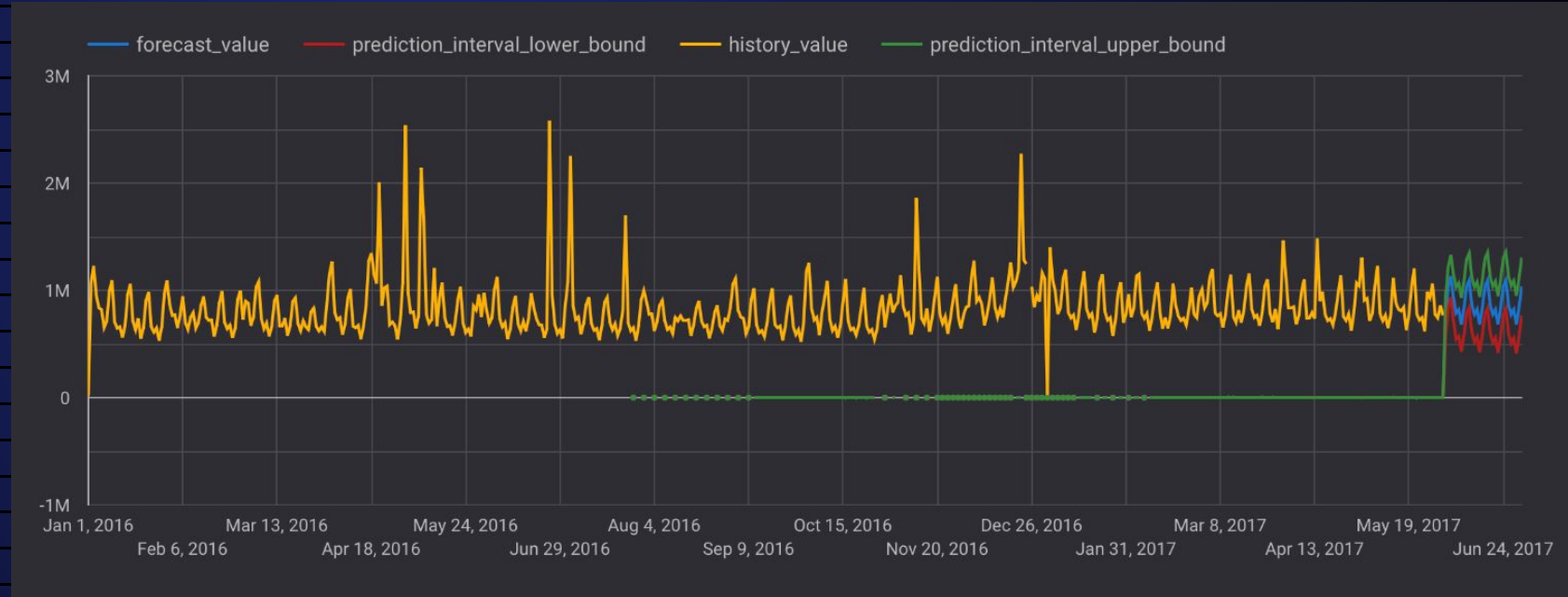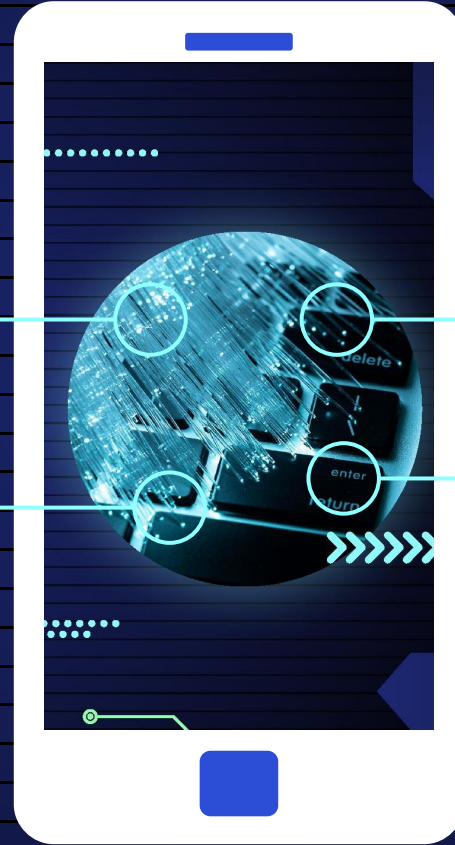

click here to see more detail

Concluding Remarks

Powerful BigQuery ML Tool

BigQuery provides SQL based modeling

Precious Learning Experience

Exploring a powerful, and advanced tool thrilled us on our way of analysis

Great Business Values

Built in models to forecast demand, do segmentation, and create product recommendation systems

Smooth With Big Data

Platforms such as Amazon Forecast cannot accommodate big data

Thank you