

salesforce

# Bring your own AI in Data Cloud

**Sharda Rao**

Distinguished, Technical Architect



# Thank You

# Introduction



**Sharda Rao**

*Distinguished, Technical Architect  
Marketing Cloud Specialists Team*

- 5 Years @Salesforce specializing in CRM Analytics and Tableau
- 15+ experience designing and executing successful enterprise rollouts of analytics & Salesforce solutions in Fins
- MS Mathematics & Statistics, MS Computer Science & Engineering, MS in Data Science ( 12/23)
- Close partnership with product to build real life demos of BYOM Sagemaker, VertexAI & Databricks

# Google Brings Powerful New Capabilities to Salesforce Data Cloud

NEW!

Open Data Access

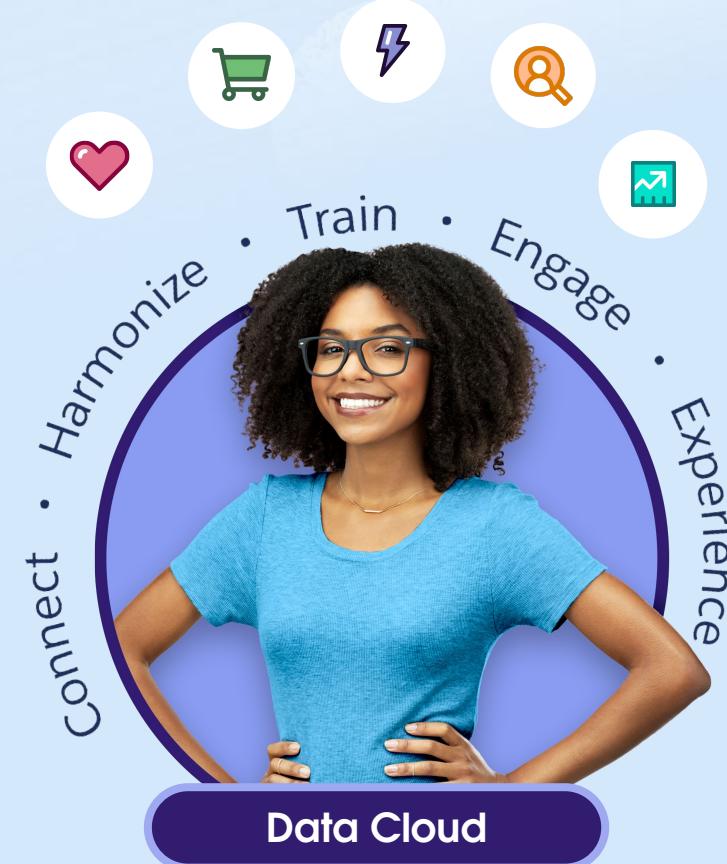
BigQuery



NEW!

Bring your own AI

VertexAI



NEW!

First Party Advertising



Google Audience Insights

NEW!

Mobile & Web Analytics



Google Analytics 4

# Essential Terms

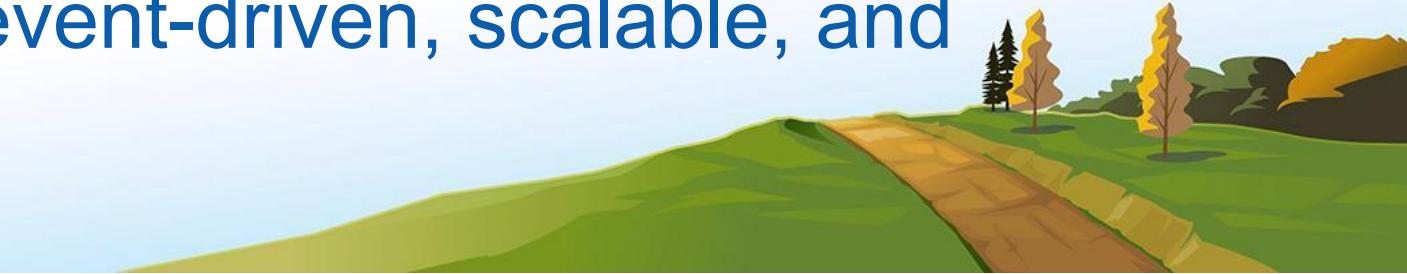


**BYOM** - Bring your own model

**Vertex AI** - Machine learning platform powered by Google to build, deploy, and manage machine learning models at scale

**Sagemaker** - Machine learning platform powered by AWS to build, deploy, and manage machine learning models at scale

**Serverless Framework** - Simplifies infrastructure management and enables event-driven, scalable, and cost-effective applications.



# Bring Your Own AI to Drive Hyper-Personalization



Salesforce Data Cloud & Google Vertex AI

**Supercharge** the Customer 360 with your own AI models

Simplify access to data for AI model training

Realize value of your custom AI investments **faster**

**NEW**

Register an External Model

Inference Endpoint: https://

Security Token:

Endpoint Type: Online

Optional:

Uploaded DataSchema: data\_schema\_filename.ext

Created with AWS Sagemaker

Sagemaker Registry Name:

Previous Register Model

Search Salesforce

AI Studio Home

Overview Templates Monitoring Models Learning

Build with Code Use our coding experience and build a custom AI process

View Monitoring

Average Latency 12.1ms Last Hour 12% Benchmark 13.1ms

GA | Oct '23

# Bring Your Own AI to Drive Hyper-Personalization



Salesforce Genie & Amazon SageMaker

**Supercharge** the Customer 360 with your own AI models

**Simplify** access to data for AI model training

Realize value of your custom AI investments **faster**

**NEW**

Pilot | Jan'23+

# End to End Solution



**Harmonize &  
Unify**



**Ingest, build ML  
model and create  
endpoint**



**Deploy and  
predict**



**Personalize &  
Act**



# End to End Solution



**Harmonize &  
Unify**

**Ingest, build ML  
model and create  
endpoint**

**Deploy and  
predict**

**Personalize &  
Act**



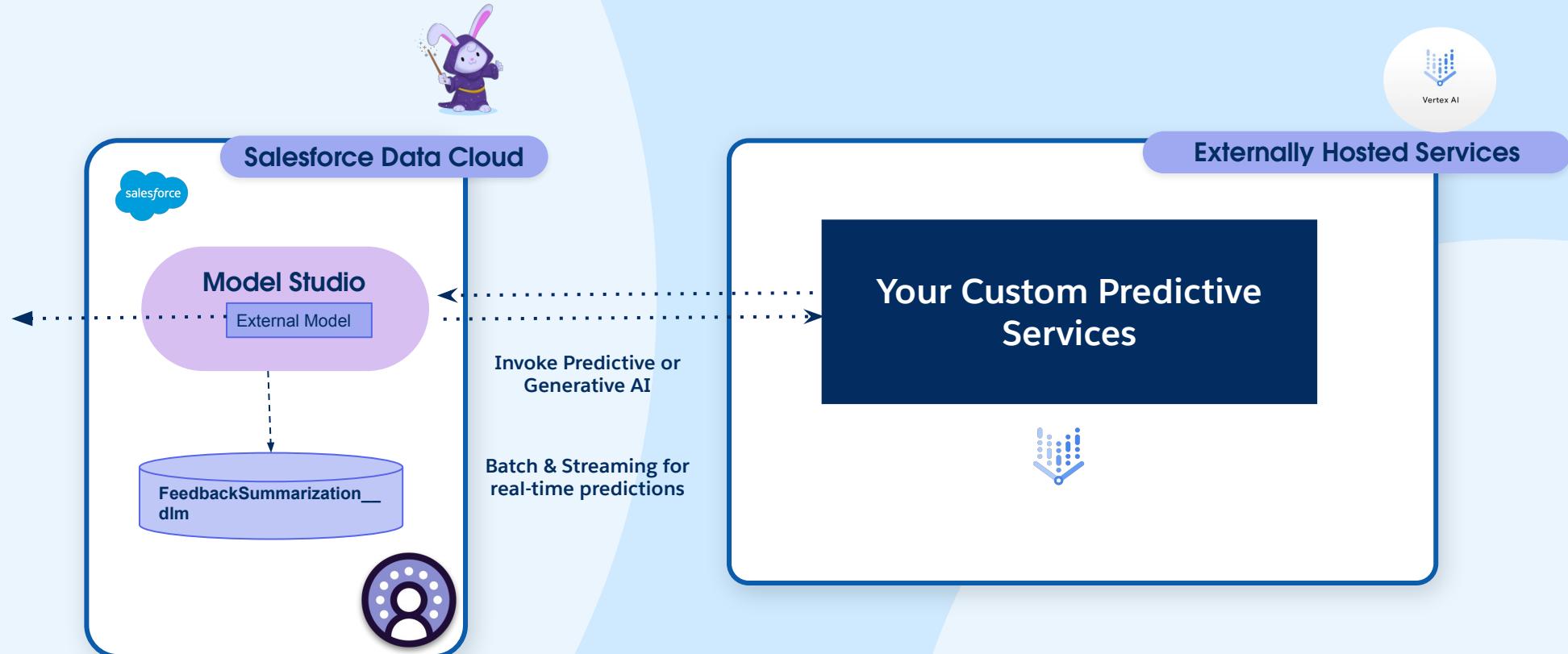
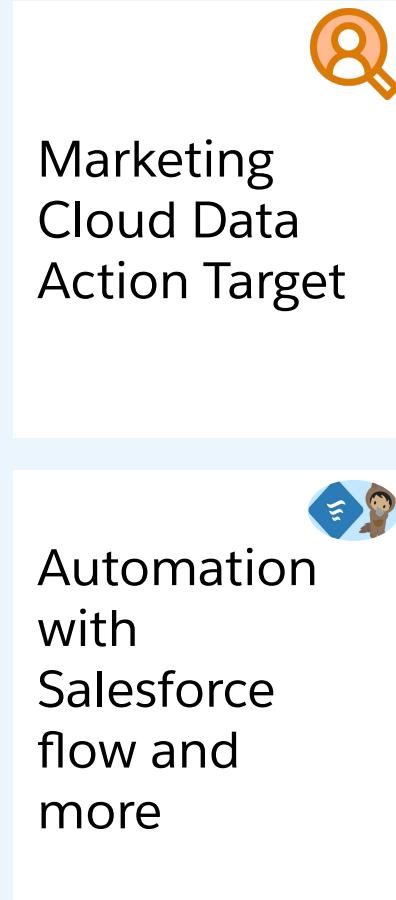
  
**databricks**



  
**salesforce**



# Consume Predictions and Generations in Salesforce



# Use Case



## About the Company

- ❑ An online retail store specializing in clothing & winter gear
- ❑ Utilizing various Salesforce clouds i.e., Sales, Service & Marketing for their CRM strategy
- ❑ Use Salesforce data cloud to ingest data from the multiple salesforce clouds and external sources to create a true customer 360

## Goal

- ❑ Tailor marketing promotions based on **predicted** customer behaviors and interests
- ❑ Increase overall customer satisfaction by providing personalized touch points based on **predictive models** especially for high-value customers

## Challenges

- ❑ Siloed Systems for developing machine learning models
- ❑ Inability to get predictive insights in the flow of work for timely action

# Why Data Cloud?

## Data Systems

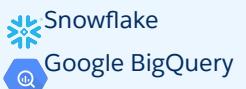
### CRM



### Cloud Storage



### Data Lakes & Warehouses



### Mobile & Web

### APIs & SDKs

### Legacy Systems

## Connect at Hyperscale



Out-of-the-Box  
Connectors



MuleSoft Anypoint  
Platform



Bring Your Own  
Lake



Data Bundles



Streaming & Batch  
Data Ingestion



Streaming & Batch  
Data Transforms

## Harmonize



Data Spaces



Data Models



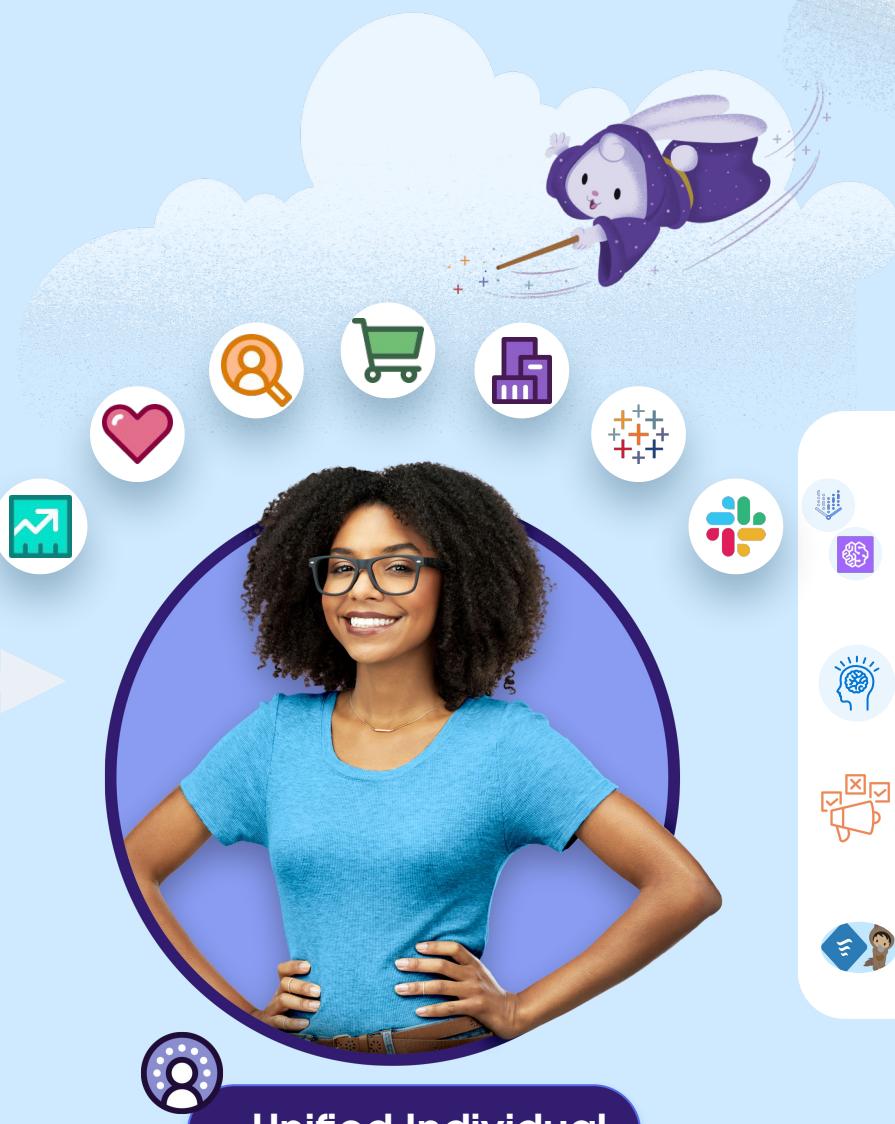
Data Mapping



Identity Resolution



Data Graphs



## Predict & Act



Einstein Copilot Studio  
Google Vertex AI  
Amazon SageMaker



Intelligent Segment  
Generation



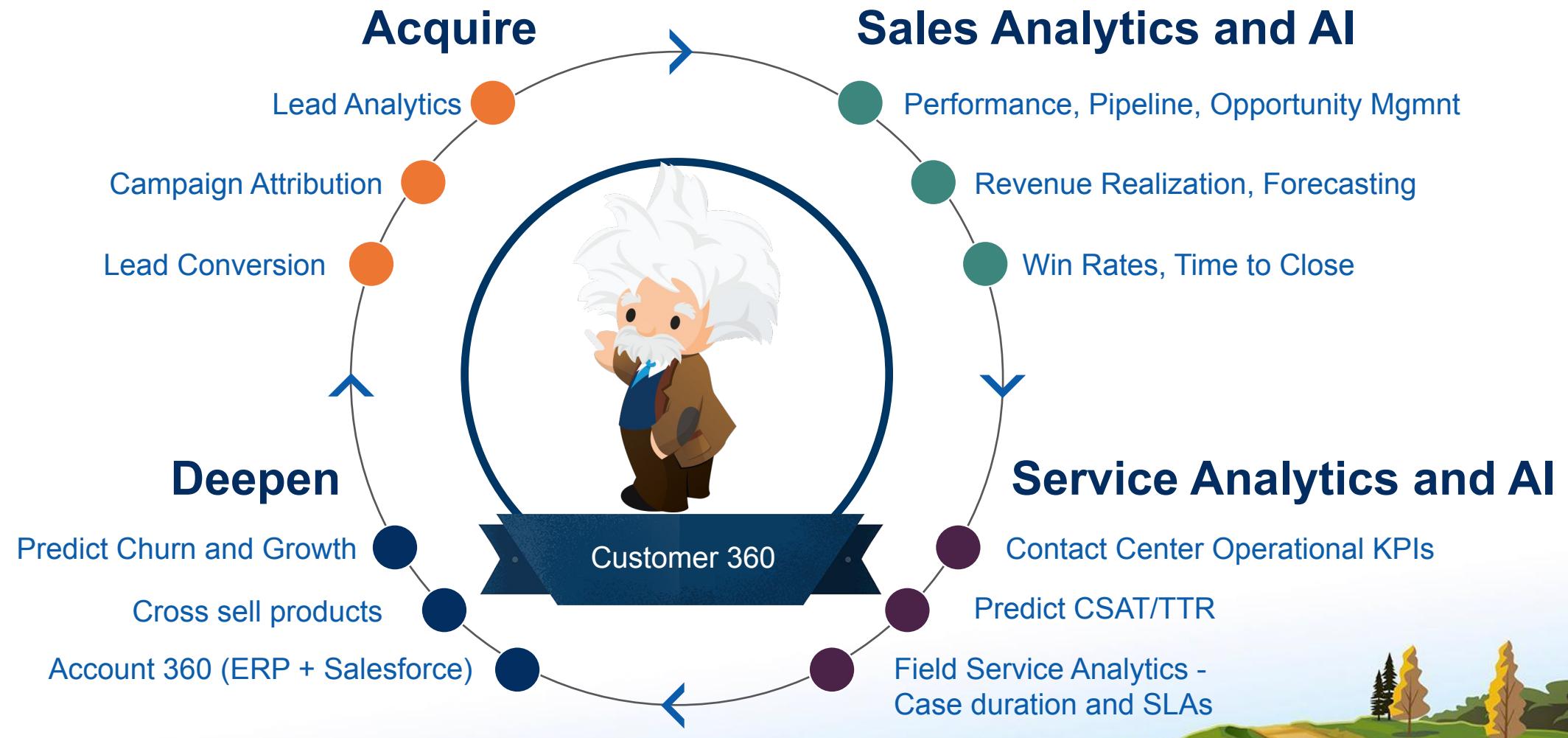
Targeted Marketing  
Campaigns



Automate in core  
salesforce platform

Unified Individual

# Typical Use Cases Across the Customer Life Cycle



# How to use predictive & generative in Salesforce?



*Output from traditional ML models can be served as an input to generative AI and vice versa to provide a more refined and enhanced personalized Salesforce experience*

	Generative AI		Predictive Machine Learning
<b>End to end call center analytics</b>	Case Classification Sentiment Entity Extraction Summarization	→ → → ↔	Propensity to meet SLA* Predicted Days to Close. Predicted CSAT^
<b>Customer 360</b>	Hyper personalization using timely summarization of customer queries and trends, search and content generation	↔ ↔ ↔	Propensity to attrit* Propensity to Deepen* Product Recommendation^^ Next best Campaign^^ Customer Segmentation*

\*Binary Classification, ^Numerical Outcome, ^^ Multi Classification, \*Unsupervised Learning

# Training Data to predict product interest



Historical data that includes the following factors:

- Customer demographic details, such as location, gender, age range, CSAT / NPS score, and loyalty status.
- Case information related to previous purchases, including the number of cases raised and whether any cases were escalated.
- Purchase history, encompassing details about the products bought and the dates of purchase.
- Website and engagement data, including the number of website visits, clicks, and engagement score



# Sales use cases



## Example Predictors

Predict Opportunity Win Rate	Predict at Risk Accounts	Predict Upsell Opportunities
<ul style="list-style-type: none"><li>❖ Historical won / lost opportunity data</li><li>❖ Opportunity demographics e.g., price bucket, product line bundles, days to close etc</li><li>❖ Account demographics e.g., region, company size, revenue, Industry etc</li><li>❖ Activities around opportunities / accounts</li></ul>	<ul style="list-style-type: none"><li>❖ Historical attrited / not attrited accounts</li><li>❖ Account demographics e.g., region, company size, revenue, Industry etc</li><li>❖ CSAT, Sentiment, Cases</li><li>❖ Account product penetration, pricing, tenure</li><li>❖ YOY balances, revenue growth</li><li>❖ Current opportunities stats</li></ul>	<ul style="list-style-type: none"><li>❖ Historical Products purchased at account level</li><li>❖ Account demographics e.g., region, company size, revenue, Industry etc</li><li>❖ CSAT, Sentiment, Cases</li><li>❖ Account product penetration, pricing, tenure with bank</li><li>❖ Historical YOY balances, revenue growth</li></ul>

# End-to-End Machine Learning life cycle



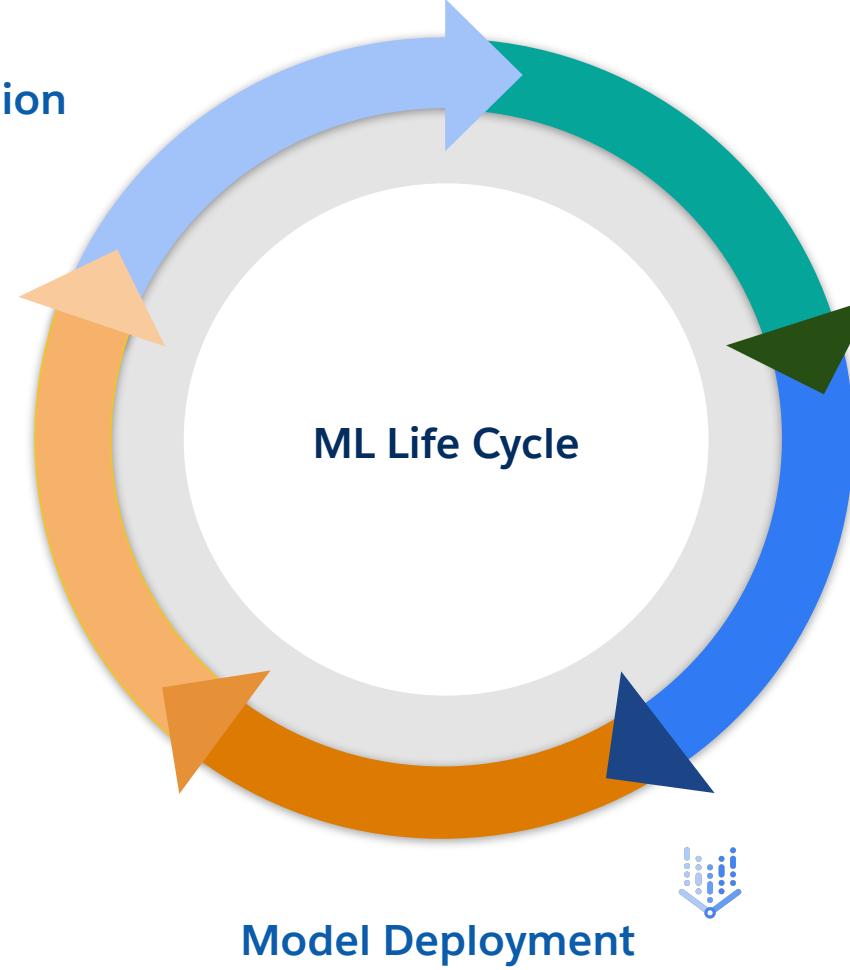
Data ingestion and Unification



Data Exploration and Analysis



Model Consumption and Actions



Model Deployment

Model Development



# Prepare the data



# Ingest & Unify

- Ingest data from multiple sources
- Unify, prepare and cleanse
- Create training dataset to build ML model on

The image displays three screenshots of a data pipeline platform:

- Data Streams:** A list of 7 data streams. The columns are: Data Stream Name, Data Connector Type, Last Run Status, Data Stream ID, Last Processed..., Total Records, and Last Refreshed.
- New Data Transform:** A dialog for creating a new data transform. It shows a query editor with the following SQL-like code:

```
SELECT
<SourceDataLakeObject.SourceField1> as TargetField1,
<SourceDataLakeObject.SourceField2> as TargetField2
FROM <SourceDataLakeObject>
WHERE <predicate on rows> [Optional]
```

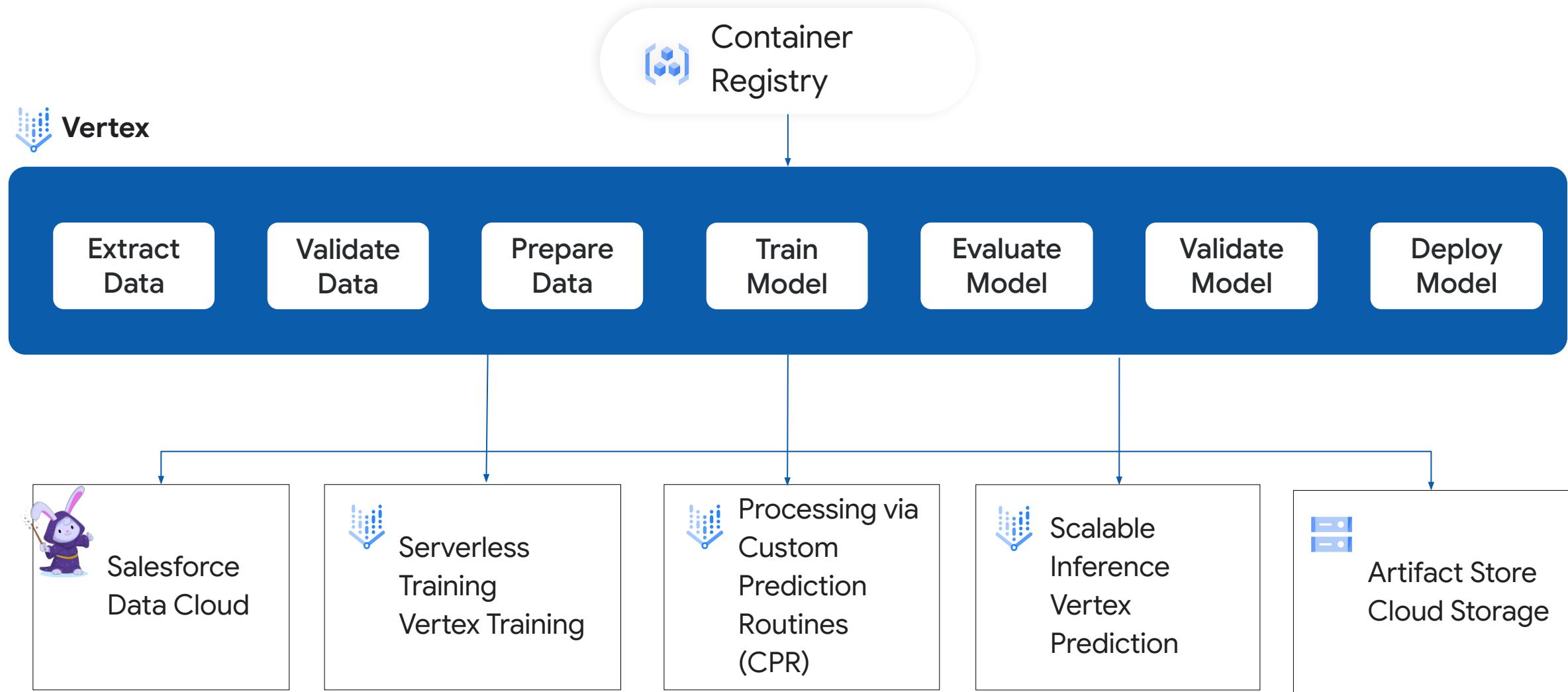
A dropdown menu on the left lists various source tables: Campaign\_NonProfit\_TabAccelerators, CarreraEvenFunnier-controller\_data, CarreraEvenFunnier-event\_data, CarreraFun-controller\_data, CarreraFun-event\_data, CarreraFun-event\_type, CarreraFun-lap\_data, CarreraMostFun-controller\_data, CarreraMostFun-event\_data, CarreraMostFun-event\_type, and Case\_0000900000050Ue9.
- Batch Data Transform:** A visual workflow editor showing a sequence of nodes: Input, Join 0, Transform 0, Join 1, Filter 0, Transform 1, and Output 0. The input nodes are AJ\_Account\_00D2 w00000PhYwk, AJ\_Case\_00D2w00 000PhYwk, and AJ\_Contact\_00D2 w00000PhYwk. The output node is Output 0.



# Explore, Build, Train and Deploy Model in Vertex AI



# ML Ops with Vertex AI



# Query



- Query data from data cloud into Vertex

Data Model Object  
**XGBoost Training Data Stream**

Type: Custom, Object Status: Ready, Mapped data streams: 1, Mapped data lake objects: 1

Fields (17)

Field Label	Field API Name	Data Type	Is Mapped	Enable Value Suggestion	Key Qualifier
1 Campaign	campaign__c	Text	✓		
2 Case Count	case_count__c	Number	✓		
3 Case Type Return	case_type_return__c	Number	✓		
4 Case Type Shipment Damaged	case_type_shipment_damaged__c	Number	✓		
5 Clicks	clicks__c	Number	✓		
6 Club Member	club_member__c	Number	✓		
7 Data Source	DataSource__c	Text	✓		
8 Data Source Object	DataSourceObject__c	Text	✓		
9 Engagement Score	engagement_score__c	Number	✓		
10 Id	id__c	Text	✓		
11 Internal Organization	InternalOrganization__c	Text	✓		
12 Key Qualifier Id	KQ_id__c	Text	✓	"Id" Key Qualifier	

Search...



vertex.ai

```
<salesforcecdpconnector.connection.SalesforceCDPConnection object at 0x7f1ba640bca0>
```

Query training data from CDP into dataframe

```
query = "SELECT product_purchased__c, club_member__c, campaign__c, state__c, month__c, case_count__c, case_type_return__c, clicks__c, engagement_score__c, InternalOrganization__c, id__c, KQ_id__c, InternalOrganization__c, club_member__c, campaign__c, state__c, month__c, case_count__c, case_type_return__c, clicks__c, engagement_score__c, InternalOrganization__c, id__c, KQ_id__c"
```

```
df = conn.get_pandas_dataframe(query)
print("Fetched")
```

Fetched

# Transform



- Data Exploration
- Transformation
- Create Preprocessing pipeline

Colab Enterprise

NOTEBOOKS RUNTIMES RUNTIME TEMPLATES

Product Interest Prediction.ipynb +

Applying out Custom Transformations here

```
# Apply custom transforms here

def transform(X):
    # Convert predictor types
    X['club_member_c'] = X['club_member_c'].astype(object)
    X['month_c'] = X['month_c'].astype(object)
    X['case_type_return_c'] = X['case_type_return_c'].astype(object)
    X['case_type_shipment_damaged_c'] = X['case_type_shipment_damaged_c'].astype(object)
    # Implement your custom formula with if statement
    # For example, if you want to create a new column based on a condition:
    X['transformed_state_c'] = X['state_c'].apply(lambda x: 'Other' if x not in States_List else x)
    X['transformed_cases_c'] = X['case_count_c'].apply(lambda x: 'No Cases' if x == 0 else '1 to 2 Cases' if x <= 2 else 'Greater than 2 Cases')
    X['transformed_tenure_c'] = X['tenure_c'].apply(lambda x: 'Less than 1' if x < 1 else '1 to 2 Years' if x == 1 else '1 to 2 Years' if x == 2 else '2 to 3 Years' if x == 3 else 'Greater than 3 Years')
    X = X.drop(['state_c', 'case_count_c', 'tenure_c'], axis=1)
    X = X.rename(columns={'transformed_state_c': 'state_c', 'transformed_cases_c': 'case_count_c', 'transformed_tenure_c': 'tenure_c'})
    return X

X_train = transform(X_train)
X_test = transform(X_test)

[43] numeric_features = ['engagement_score_c', 'clicks_c', 'pages_visited_c']
categorical_features = ['club_member_c',
                        'campaign_c',
                        'state_c',
                        'month_c',
                        'case_count_c',
                        'case_type_return_c',
                        'case_type_shipment_damaged_c', 'tenure_c']

[44] preprocessor = ColumnTransformer(
        transformers=[
            ('numeric', StandardScaler(), numeric_features),
            ('categorical', OneHotEncoder(), categorical_features)
        ]
)
```



# Train



- Build, train and test ML model using Vertex

Colab Enterprise

NOTEBOOKS RUNTIMES RUNTIME TEMPLATES

+ Code + Text Commands All changes saved

Product Interest Prediction.ipynb

Build out the XGBoost decision tree components

```
# Define the KNN classifier
xgb_classifier = xgb.XGBClassifier(objective='multi:softmax')

# Define the hyperparameters to tune
param_grid = {
    'classifier__n_estimators': np.arange(100, 1001, 100),
    'classifier__max_depth': np.arange(3, 9, 4),
    'classifier__learning_rate': np.logspace(-2, 0, 3),
    'classifier__subsample': [0.8, 1.0],
    'classifier__colsample_bytree': [0.8, 1.0],
    'classifier__gamma': np.linspace(0, 0.2, 3), # Range from 0 to 0.2 (inclusive) with 3 values
    'classifier__reg_alpha': np.linspace(0, 1.0, 3), # Range from 0 to 1 (inclusive) with 3 values
    'classifier__reg_lambda': np.linspace(0, 1.0, 3),
}

[47] # Create the ML pipeline
pipeline = Pipeline(steps=[
    ('transformer', preprocessor),
    ('classifier', xgb_classifier)
])

[48] # Perform grid search for hyperparameter tuning
random_search = RandomizedSearchCV(pipeline, param_grid, cv=5)
random_search.fit(X_train, y_train)
```

RandomizedSearchCV  
estimator: Pipeline  
transformer: ColumnTransformer  
numeric categorical  
StandardScaler OneHotEncoder  
XGBClassifier



# Create Endpoint



- Register model
- Create End point

Model Registry    [+ CREATE](#)    [IMPORT](#)

Models are built from your datasets or unmanaged data sources. There are many different types of machine learning models available on Vertex AI, depending on your use case and level of experience with machine learning. [Learn more](#)

Region  ▼ ?

Filter Enter a property name

	Description	Default vers
—	—	1

Online prediction

[ENDPOINTS](#) [DEPLOYMENT RESOURCE POOLS](#)

Endpoints are machine learning models made available for online prediction requests. Endpoints are useful for timely predictions from many users (for example, in response to an application request). You can also request batch predictions if you don't need immediate results.

To create an endpoint, you need at least one machine learning model. [Learn more](#)

Region  ▼ ?

Endpoints    [+ CREATE](#)

Filter Enter a property name

<input type="checkbox"/>	Name	ID	Status	Models	Deployment resource pool	Region	Monitoring
<input type="checkbox"/>	<a href="#">spotlight-ml-demo</a>	5719732055458185216	<span>Active</span>	1	—	us-central1	Disabled
<input type="checkbox"/>	<a href="#">e2e-text-classification-model-unique_endpoint</a>	355522686794858496	<span>Ready</span>	0	—	us-central1	Disabled



# Bring Inferences into Datacloud



# Deploy Model in Data Cloud



- No code deployment of Google Vertex model
- Score unified individuals

The screenshot displays the Salesforce Data Cloud interface, specifically the ML Model and Data Model Object sections.

**ML Model - Product Interest Prediction Inference:**

- Details:** ML Model API Name: Product\_Interest\_Prediction\_Inference, Status: Active, Last Run Status: SUCCESS, Last Refresh Date: 6/16/2023, 11:08 AM, Last Processed Records: 262.
- Endpoint:** Endpoint Name: BYOMSagemaker, Endpoint API Name: BYOMSagemaker, Inference Endpoint URL: https://drf23262j.execute-api.us-east-1.amazonaws.com, Request Format: JSON, Response Format: JSON.
- Input Features:** ML Model API Name: Product\_Interest\_Prediction\_Inference, Status: Active, Last Run Status: SUCCESS, Last Refresh Date: 6/16/2023, 11:08 AM, Last Processed Records: 262.

**Data Model Object - Product Prediction Inference:**

- Details:** Type: ML\_Prediction, Object Status: Ready, Mapped data streams: 0, Mapped data lake objects: 0.
- Relationships:** A table showing relationships between the Product Prediction Inference object and other entities like Account Contact, Campaign, State, Month, Case Count, Case Type Return, Case Type Shipment, Pages Visited, and Engagement Score.



# Act of the inferences



# Create automated tasks for Service Agents



- Use data actions to update platform events when records are scored
- Kick start lightning flows automatically when a platform event is created
- Curated call list for high value customers created for service reps to act on the insights

The image displays four components illustrating automated tasks:

- Data Cloud Product Interest Inference:** A screenshot of the Data Cloud interface showing the configuration of a "Product Interest Inference" data action. It includes sections for Action Targets, Event Rules, Conditions, and Attributes.
- New Flow:** A screenshot of the Lightning Flow builder interface showing a list of available flow types under the "Core" category, including Session Flow, Record Triggered Flow, Schedule-Triggered Flow, AutoLaunched Flow (No Trigger), Record Triggered Orchestrator, and Record Triggered Orchestration.
- My Cases:** A screenshot of the Salesforce My Cases dashboard. It shows a summary of 94 cases vs prev 30 days, a "Cases Breakdown" chart with categories: Closed (48), New (24), Working (12), Escalated (8), and Waiting on Customer (4). To the right is a "My Call List" section showing 262 entries with columns for Full Name, Email, and Product Interest, listing contacts like Judith Lopes, Alidis Lorenzin, Sharl Danforth, Yolane Lattka, Rube Whitney, Francesco Beneyto, and Marcia Capenor.
- Autlaunched Flow Diagram:** A detailed view of an Autlaunched Flow. The flow starts with "Start Autlaunched Flow", followed by "GetContacts Get Records". This leads to a "Loop Loop" which contains a "For Each" loop. Inside the loop, there is an "ML Prediction Action" (represented by a lightning bolt icon) and a "Check Contact Type Decision" (represented by a diamond icon). The flow then branches into two paths based on the decision: one path leads to "Create Task Create Records" and the other to "End". Finally, the main loop ends with "End" and the entire flow concludes with "End".



# Create marketing segmentation and Journey



- Create segmentation and activation targets based on the predictions
- Personalized promotions automatically sent to customers

The screenshot shows the Salesforce Data Cloud interface with a segment named "Marketing Promotions". The segment details include:

- Segment On: Account Contact
- Segment Status: Active
- Last Publish Completed: Not specified
- Publish Status: Published
- Publish Schedule: Don't refresh
- Next Publish Date Time: Not specified

Segment Population: 0 (Segment Status: Active)

Activations (0) [New]

**Details** [Related]

Owner Name: Sharda RAO	Last Publish Completed:
Segment On: Account Contact	Publish Schedule: Don't refresh
Segment Name: Marketing Promotions	Next Publish Date Time:
Segment Description:	Publish Schedule Start Date and Time:
Segment Status: Active	Expiration Date:
Published Status:	Segment Type: UI
Created By: Sharda RAO, 6/16/2023, 4:10 PM	Publish Type: Standard
Created Date: 6/16/2023, 4:10 PM	

**northern trail** outfitters

Thank you for being a customer,  
Delaria!

Based on your history with NTO, we've found these products we think you'll enjoy. We've even included a special offer to help you with your decision...

**Latest Recommendations**

**Paramount Peak Convertibles**  
Lightweight, convertible pants for hiking in variable summer conditions.  
\$88

**Motivation High-rise Crop**  
Spring training sessions just got real with these crops made with 4-way stretch fabric.  
\$65

**On the Go Mid-rise Pants**  
We noticed you needed the perfect pants for warming up before your summer hiking.  
\$90

**10% OFF + FREE SHIPPING**  
Valid today only with code NEW10

VISIT MY CART



# Follow Up Materials



Request your own Vertex AI or AWS Environment -

<https://embark.sfdcvt.net/?tab=0>

Vertex AI Demo Script -

<https://salesforce.quip.com/kTIfAXgAqChC>

Build your own demo with Vertex AI -

<https://salesforce.quip.com/Sap1AbLFhMa2>

Sagemaker Demo Script -

<https://salesforce.quip.com/CnbrA6sIu2PR>

Predictors by different use cases and industries -

[https://docs.google.com/presentation/d/1FIXHRFgKD7TKHt822P1sIQxmwJ2gMEo33SHBT1j8aZc/edit#slide=id.gadfbfbcc8d6\\_1\\_3048](https://docs.google.com/presentation/d/1FIXHRFgKD7TKHt822P1sIQxmwJ2gMEo33SHBT1j8aZc/edit#slide=id.gadfbfbcc8d6_1_3048)





# Thank You