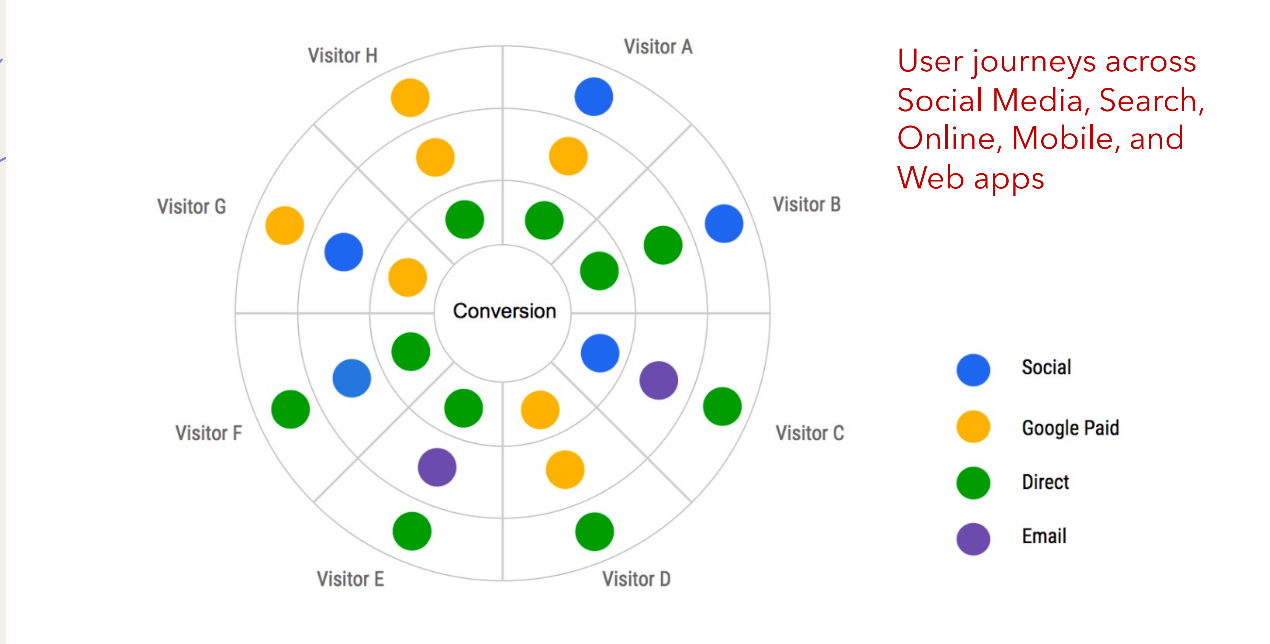
Optimizing Advertising Spend with **Machine Learning & Databricks**

Demian Esnaurrizar

Shift the bricks **X-Challenge**

+The original accelerator just does Attribution



Source: https://www.adalyz.com/multi-channel-attribution/

Ex: *How much Google or Social Media contributes to overall sales of your product.*

.

Attribution Model
Assigns credit to various touchpoints on customer journey

Attribution model answers the business question: *How much money is produced or can be attributed to each touch point?*

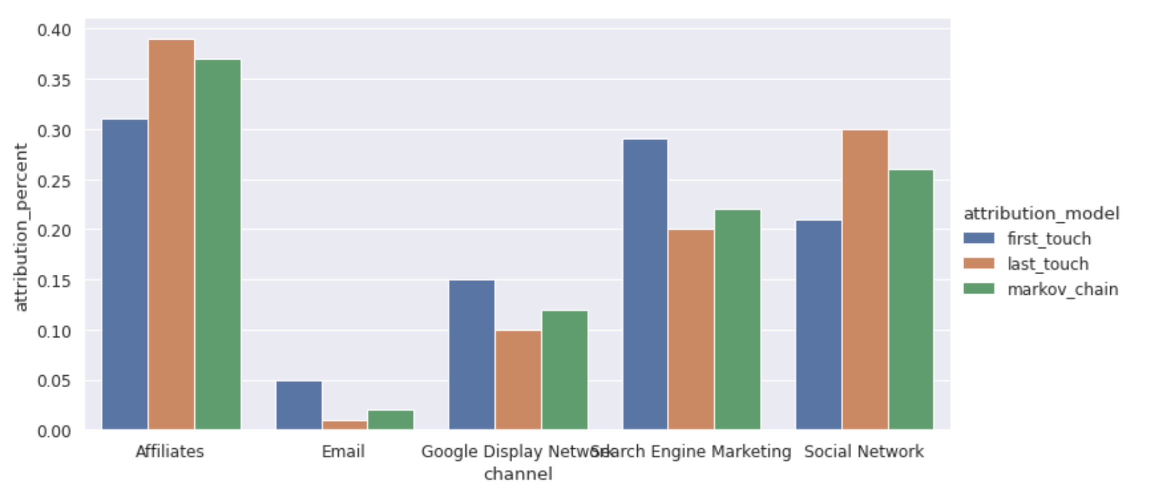
.

Attribution Model

Assigns credit to various touchpoints on customer journey

+Using 3 methods:

+ First Touch
+ Last Touch
+ Markov Chain



For example, in chart above from the original Databricks Accelerator Notebook. Email channel is very ***inefficient****.*

What's the Purpose?
With these attribution insights, you could then adjust your spend allocation accordingly and yield a higher return on ad spend **(ROAS).**

Why Stop There?

Multi-Channel Attribution

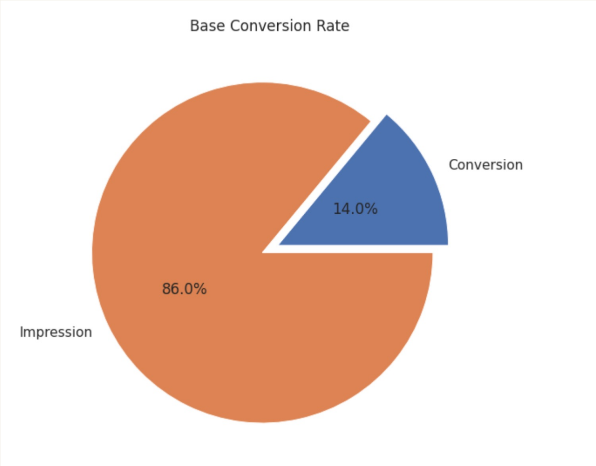
- User Journey
- Target selection of items relevant for ad placement
  - We call this set the “bid unit”

- Business Problem:
  - Spending money on Social Network or Google Ads Networks can have different returns


The right Ad at the right time…
Understanding the Customer Journey

+The original accelerator maps the entire user journey

*to compute the **overall conversion rate** at a very high level*



*And to allocate the percentage of ad spend per **marketing campaign***

	channel	pct_spend	dollar_spend
1	Social Network	0.2	2000
2	Search Engine Marketing	0.2	2000
3	Google Display Network	0.2	2000
4	Affiliates	0.2	2000
5	Email	0.2	2000

These are actual snapshots from original accelerator, just for illustrative purposes

Why stop there?

Perhaps this is enough to start a conversation with potential customers…

Going back to Differentiation

We want to have actionable insights

Original Accelerator gives only Information

We want to do something that improves the process

+In this day and age, who does not need advertising?

- Almost any company use advertising in some way
- Blueprint uses advertising
- Almost any company that wants to use their ad dollars more effectively
- Millions of dollars are spent every day in branded ad placement

How can we combine detailed insights to Optimize Ad spend?

+Leverage Machine Learning + Databricks

Data Preparation

Spark Streaming

Spark

DELTA LAKE

Raw Data
(Bronze Table)

DELTA LAKE

Clean Data
(Silver Table)

DELTA LAKE

Refined Data
(Gold Table)

Machine Learning & Optimization

Spark MLlib

COMPUTATIONAL INFRASTRUCTURE FOR OPERATIONS RESEARCH

COIN|OR

SCIPY PYOMO PULP CVXOPT

Feature
Engineering

Train Machine
Learning Models

Optimize Ad
Spend

explore

We want to **predict** return on ad spend on any ad **before** we buy it so we can take decision not to

+Had to create synthetic data

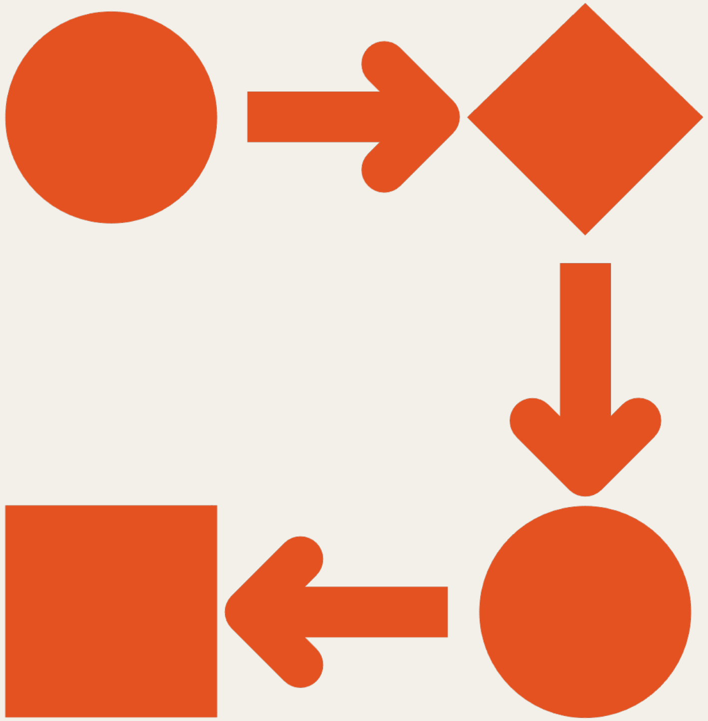
Original Accelerator	**New**
Steps in Customer Journey	Detailed clicks, spend, and revenue data
Per Marketing Campaign	Per Bid Unit

Steps in Customer Journey → Detailed Clicks, Spend and Revenue → Predictive Modeling & Data Engineering → Ad Spend Optimizer

For solving this X-Challenge we are using standard python optimization engine Coin-Or CBC solver PuLP
https://realpython.com/linear-programming-python/
https://coin-or.github.io/pulp/

Business Friendly Configurable Objective Function

- **Goal**:
  - No black-box or mysterious AI/Deep Learning
  - Simply:
    - Select your targets
    - Plug-in your data
    - Maximize Revenue and Profitable **Growth**
    - Create and explore new sources of income
    - Across multiple channels
- **Take into consideration easily**: Operational constraints, Spend, and ROI constraints



Epic user experience:
Win at each step of the customer journey

+Bid units

"Utility is the happiness a person gets by consuming goods and services"

Source: [Brigham Young University - Idaho](https://courses.byui.edu/econ_150/econ_150_old_site/lesson_05.htm)
https://courses.byui.edu/econ_150/econ_150_old_site/lesson_05.htm

This is the set or bundle we optimize for

Customer	Product Stars	Channel

Bidding Unit

This is what gives us the score to optimize

Utility Function	**Spend**	**Revenue**

Machine Learning Scores

All is driven by detailed cost and click share data

- Impressions
- Clicks
- Transactions, etc.

DELTA LAKE

Spark MLlib

Once a customer provides its own data we can simply point and click and append extra features

Utility is important and any decision we make, and we want to maximize utility, but utility is not profit

databricks

+ Detailed Flow (Original)

Steps in Journey Raw Data

Silver User Journey Table

Gold User Journey Table

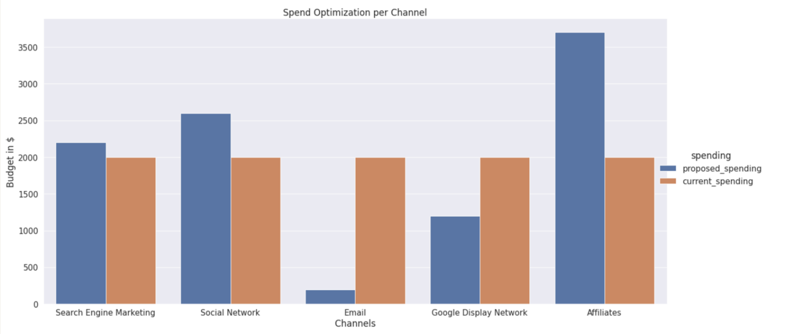
Attribution Model

First Touch Last Touch

Markov Chain

Spend Optimization per Channel

Proposed vs current channel spending



Per Marketing Campaign

Gold Ad Spend

Per Marketing Campaign

Budget Allocation

Per Channel

Spend Dashboard

+ Detailed Flow (New)

Raw Click Share Data

Silver Spend and Revenue Table

Gold Feature Engineering

Split Data

Train

Test

Evaluate

New - Ad Spend Optimizer

MLlib

Machine Learning Models

Spend Prediction

Markov Chain 2

Utility and Risk Aversion Estimator

Per Bid Unit

Ad Spend Optimizer

Demo & Code Samples

+ Detailed Marketing Contribution per Channel

+ Example

Detailed Marketing Contribution per Channel

Python

```python
pnl_per_channel = (detailed_costs3.groupby("channel").sum(*["impression", "click", "conversion", "cpi", "cpc", "gpt"])
                   .withColumn("marketing_cost", col("sum(cpi)") + col('sum(cpc)'))
                   .withColumn("marketing_contribution", col("sum(gpt)") - col("marketing_cost"))
                   .withColumn("ROAS", col("sum(gpt)") / col("marketing_cost"))
                  )
display(pnl_per_channel)
```

(2) Spark Jobs

pnl_per_channel: pyspark.sql.dataframe.DataFrame = [channel: string, sum(impression): long ... 8 more fields]

Table Data Profile

	channel	sum(impression)	sum(click)	sum(conversion)	sum(cpi)	sum(cpc)	sum(gpt)	marketing_cost	marketing_contribution	ROAS
1	Search Engine Marketing	729649	32670	13977	1807.011180001734	1948300.1199977694	1393286.6840000194	1950107.1311777711	-556820.4471777517	0.7144667396598571
2	Google Display Network	364481	16535	7117	1159.3857599992753	234106.74000008628	709840.0902500082	235266.12576008556	474573.9644899226	3.0171793238686346
3	Email	120676	3774	716	1820.8799999993455	110466.71999998602	68701.84624999996	112287.59999998537	-43585.75374998541	0.6118382283529874
4	Affiliates	727482	45761	27247	15094.57999998888	664161.5199999834	2728374.5582500203	679256.0999999723	2049118.458250048	4.016709689099784
5	Social Network	486233	33413	20959	4057.535999965367	466616.6399993911	2099586.268250019	470674.17599938763	1628912.0922506312	4.46080616977146

Showing all 5 rows.

Command took 0.62 seconds -- by desnaurrizar@bpcs.com at 3/22/2022, 7:48:34 PM on coursework

Key Decision Variables

- Key Decision variables:
    - User Behavior
        - Ex. Risk Aversion
    - Marketing Campaign Targets
        - Ex. boost in relevant target audience, clicks, or sales conversions.
    - Marketing Contribution per Channel
        - Sales Gross Margin
        - Spend
            - Cost-per-Mille-Impressions
            - Cost-per-Click
    - Share of Voice
        - Sales Transactions
        - Revenue
        - User Clicks
    - Inventory
        - Ad Impressions
        - Product Stars or Reviews
        - Bid units
    - Return on Advertising Spend
        - ROAS = Revenue / Spend

Levers the user can Tweak

- Exploration Vs Exploitation
- Marketing Contribution and ROAS Efficiency
- New markets exploration
- Customer Behavioral profiling sales booster (Ex. “Premium Risk Seeker” vs “Risk Averse”)

Disclaimer: All these features are experimental for now

Return on Ad Spend

Additional Slides

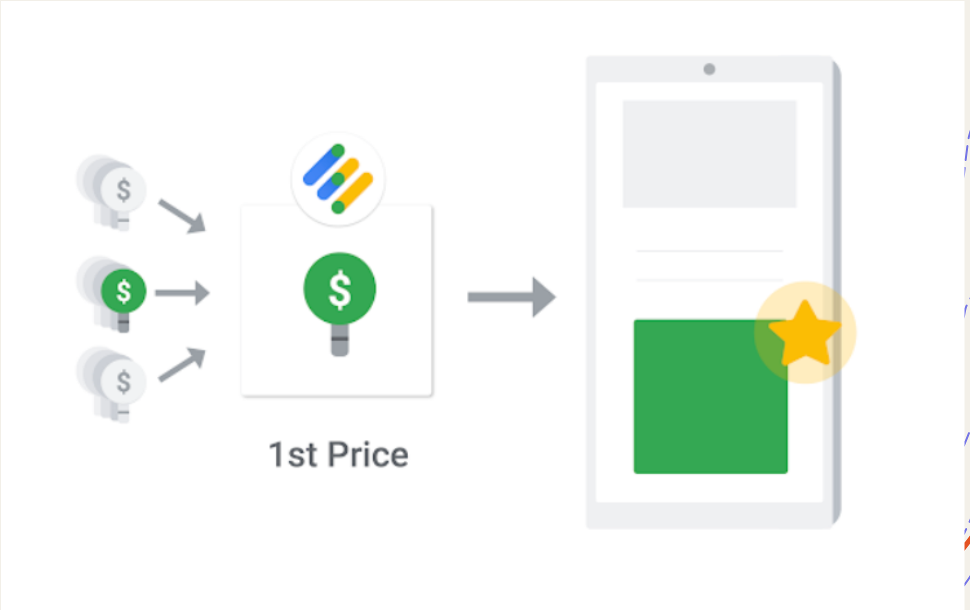
Key Concepts & Terminology

Just as an example we'll mention Google, similar applies to other platforms.

+How does ad-bidding works?

- When you seen an online add, many times in the backend there is a bidding war that is live.
  - + Begins with a user visiting a website
  - + Advertisers submit bids or pre-allocate bid budget
  - + The Ad goes to the highest bidder

Google Ads bidding example



Source: https://www.blog.google/products/admanager/simplifying-programmatic-first-price-auctions-google-ad-manager/

Mathematical Optimization

- A basic **optimization process** consist of:
  - an objection function f(x)
  - a vector x of decision variables
  - a set of constraints Ci

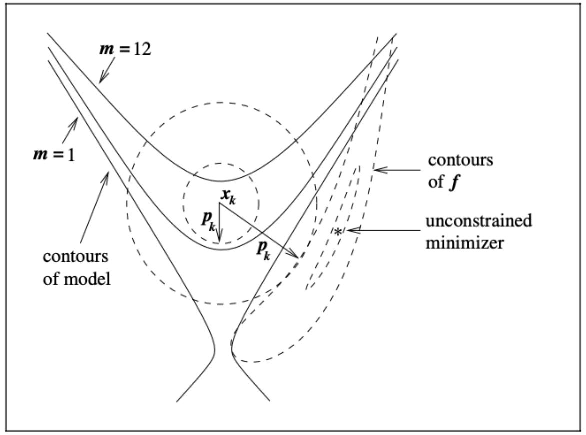


Figure 2.4 Two possible trust regions (circles) and their corresponding steps p_k. The solid lines are contours of the model function m_k.

Source: Numerical Optimization. Nocedal, Wright.2006

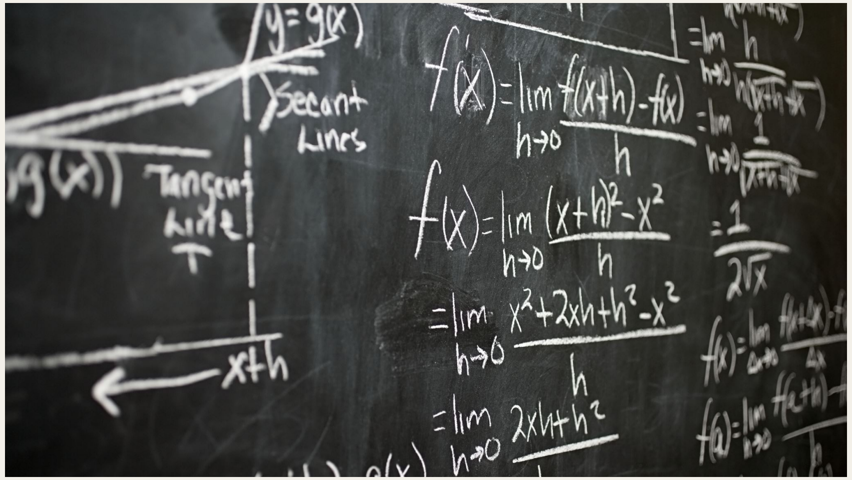
Mathematical Optimization

+ In classical operations research and optimization literature, objective function typically reflects the **total cost** or **total profit**

+ Constraints are logical or mathematical conditions to a solution, which an optimization problem must satisfy

+ Constraints usually reflect real-world limits on raw materials, stocks, quantities, or budget available, among others.

+ Mathematical optimization is both and area of innovation and research and a mature field, going back to the 1950s, traditionally considered a branch of applied mathematics that has applications in many different fields. From Manufacturing, to Inventory Control, and Engineering.



Key Concepts and Terminology

+Utility

Imagine you have a home improvement project, and you need a hammer

Key Concepts and Terminology

+Utility

In Retail…

"Utility is the happiness a person gets by consuming goods and services"

Different hammers same or different levels of utility

Light, easy to use

rubber

In any Marketplace you can find bunch of different options.

Soft sided

SledgeHammer

Key Concepts and Terminology

+Utility

In Financial and Economics…

"Utility is a function…

Assigns a score to a product set or bundle

+

=

Utility Score 2.0

+

=

Utility Score 0.6

Key Concepts and Terminology

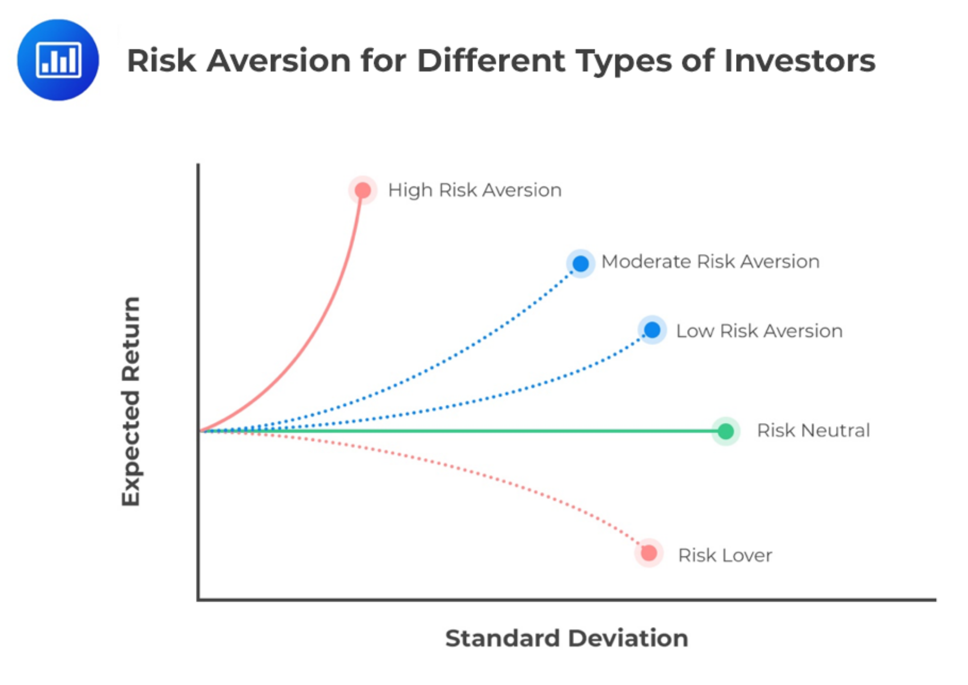
+Utility

In Stock Market and Investments Portfolio Management…

"Utility can be a tool …

Sometimes used to capture the satisfaction or level of happiness of an investor

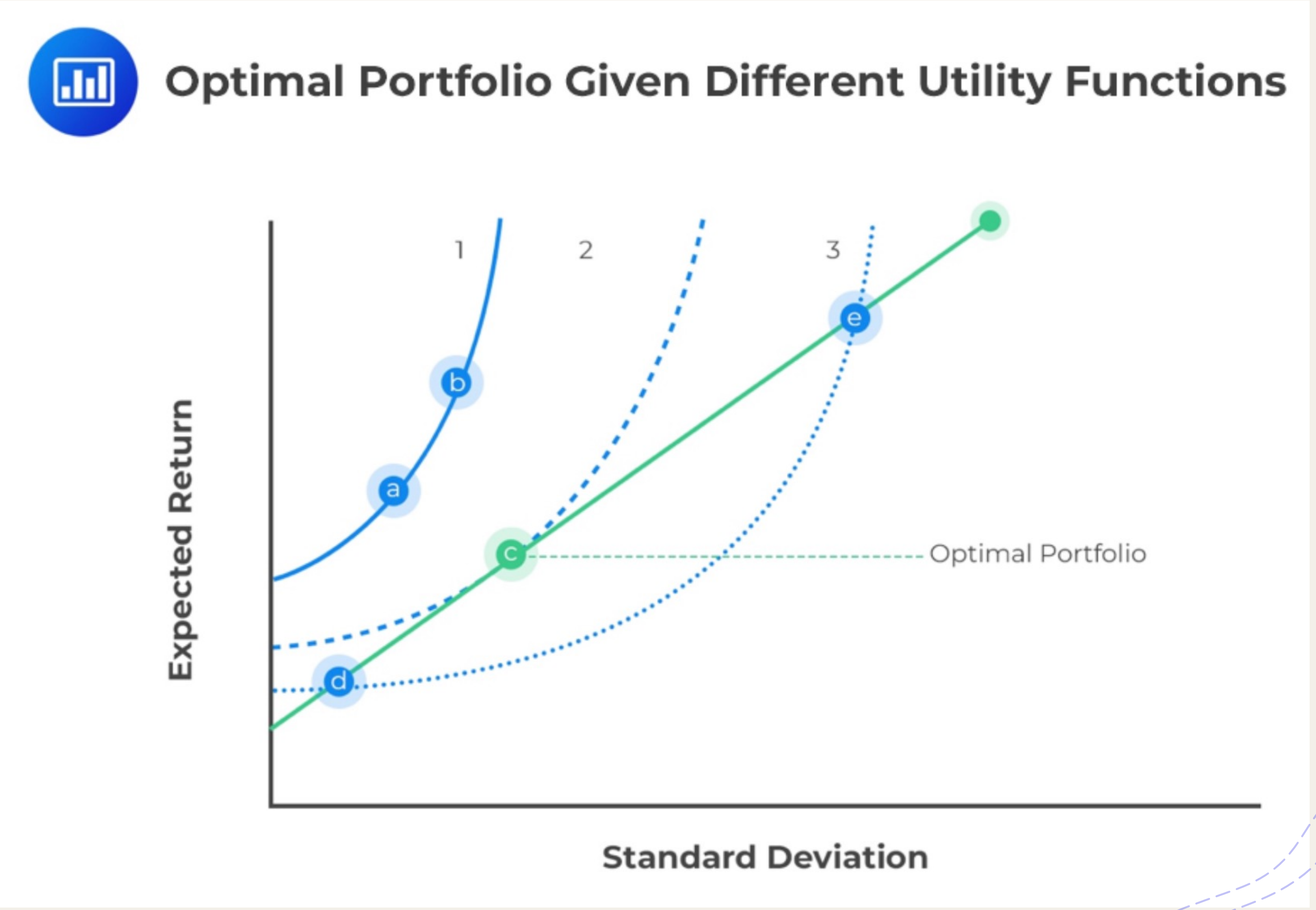
"Utility and Indifference Curves
Utility is a measure of relative satisfaction that an investor derives from different portfolios"



Source: https://analystprep.com/cfa-level-1-exam/portfolio-management/optimal-portfolios/

+ Key Concept:

Utility Function can be used both to score product bundles and to determine Optimal Portfolio and Assortment Planning



Source: https://analystprep.com/cfa-level-1-exam/portfolio-management/optimal-portfolios/

Key Concepts and Terminology

+Utility

In Financial, and Economics…

"Utility is a function…

Assigns a score to a product set or bundle

Classic Utility Functions

Example 3.2 Some utility functions.

$$
\begin{aligned}
u_1(w) &= w - \frac{b}{2}w^2, \qquad b > 0 \qquad \text{(Quadratic Utility)} \\
u_2(w) &= \frac{1}{1-\gamma}w^{1-\gamma} \qquad w > 0, \gamma > 1 \qquad \text{(Power Utility)} \\
u_3(w) &= -e^{-aw} \qquad a \geq 0 \qquad \text{(Negative Exponential Utility)}
\end{aligned}
$$

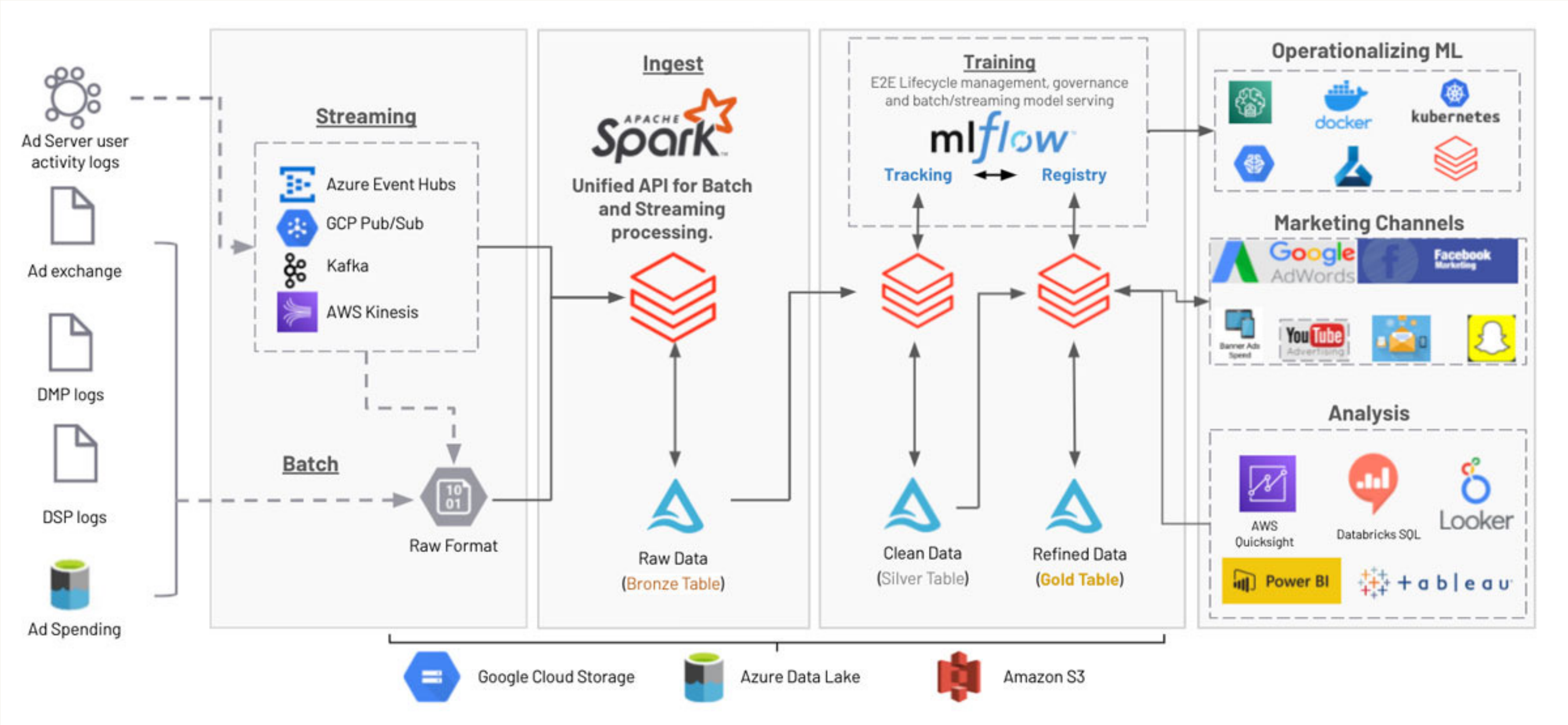
+Original accelerator architecture diagram

Lakehouse pattern:

- Bronze -> Silver -> Gold

Markov Chain :

- Transition Probability Matrix



These are actual snapshots from original accelerator

Source: https://databricks.com/blog/2021/08/23/solution-accelerator-multi-touch-attribution.html

Predictive Models

- We want to predict return on ad spend on any ad before we buy it so we can take decision not to
- Natural way to do this is with a **Machine Learning** model:
  - Random Forests, Neural Networks, Boosted Trees, Time Series Forecasting, are well known methods and APIs to do this.
  - We can leverage Spark ML and Databricks ML Runtime
- In addition to those ML models, we take the original accelerator inspiration of a Markov Chain upgrade one level up:
  - Use existing Markov Chain transition probability matrix to predict the likelihood of a user **"to explore"**
  - We define the likelihood of a user to explore in a similar way as portfolio managers define "Risk Seekers" or "Risk Aversion"
- With plug in all these feature engineering and ML predictions into a standard **Utility Function** and feed it as parameters to an **Optimization Engine**