Module 2: NewSQL Analytic Functions

Day on the life of a Data Scientist Workshop

Objectives

After completing this module, you will be able to:

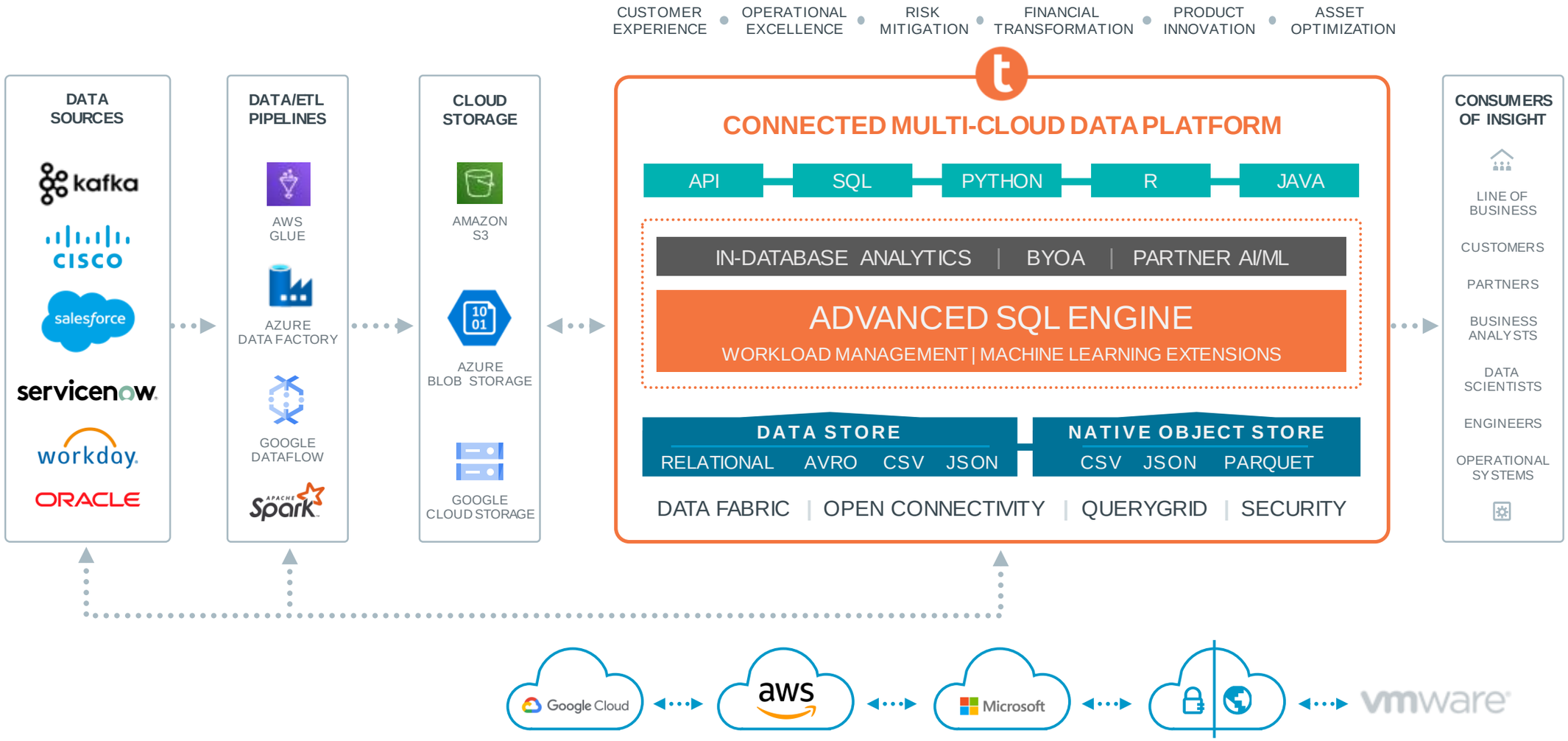
- Write queries in SQL, using the main SQLE analytic functions
- Understand the available capabilities in SQLE

Topics

- **Introduction**
- 4D Analytics
- SQLE Analytic Functions

Teradata Vantage

The Connected Multi-Cloud Data Platform for Enterprise Analytics



NewSQL Engine Functions

Category	Function	Description
Path & Pattern	nPath	Pattern matching in an ordered collection of rows.
Data Prep & Transformation	Sessionize	Maps each click in a clickstream to a unique session identifier.
Attribution	Attribution	Calculates attributions with a wide range of distribution models. Often used in webpage analysis.
Time Series	Aggregation	Allows aggregate data using grouping of time buckets.
Data Exploration	StringSimilarity	Calculates the similarity between two strings, using various metrics
Data preparation/ Transformation	Antiselect	Returns all columns except those specified in the EXCLUDE argument so that users can select all columns in a data set except a few.
	Pack	Combines multiple columns of raw data into one column for further processing.
	Unpack	Splits raw data into multiple columns using a delimiter or a regular expression for further processing.
	nGramSplitter	Tokenizes (splits) an input stream of text and outputs n-multigrams
Statistical Analysis	MovingAverage	Creates series of averages of different subsets of the full data set and associated weighting scheme. The following Moving Average functions are now available: cumulative, exponential, modified, simple, triangular, weighted

Topics

- Introduction
- **4D Analytics**
- SQLE Analytic Functions

4D Analytics in Action

Use Case: How Can We Engage Customers?

Business data
- High-value customers
- Sales/order history

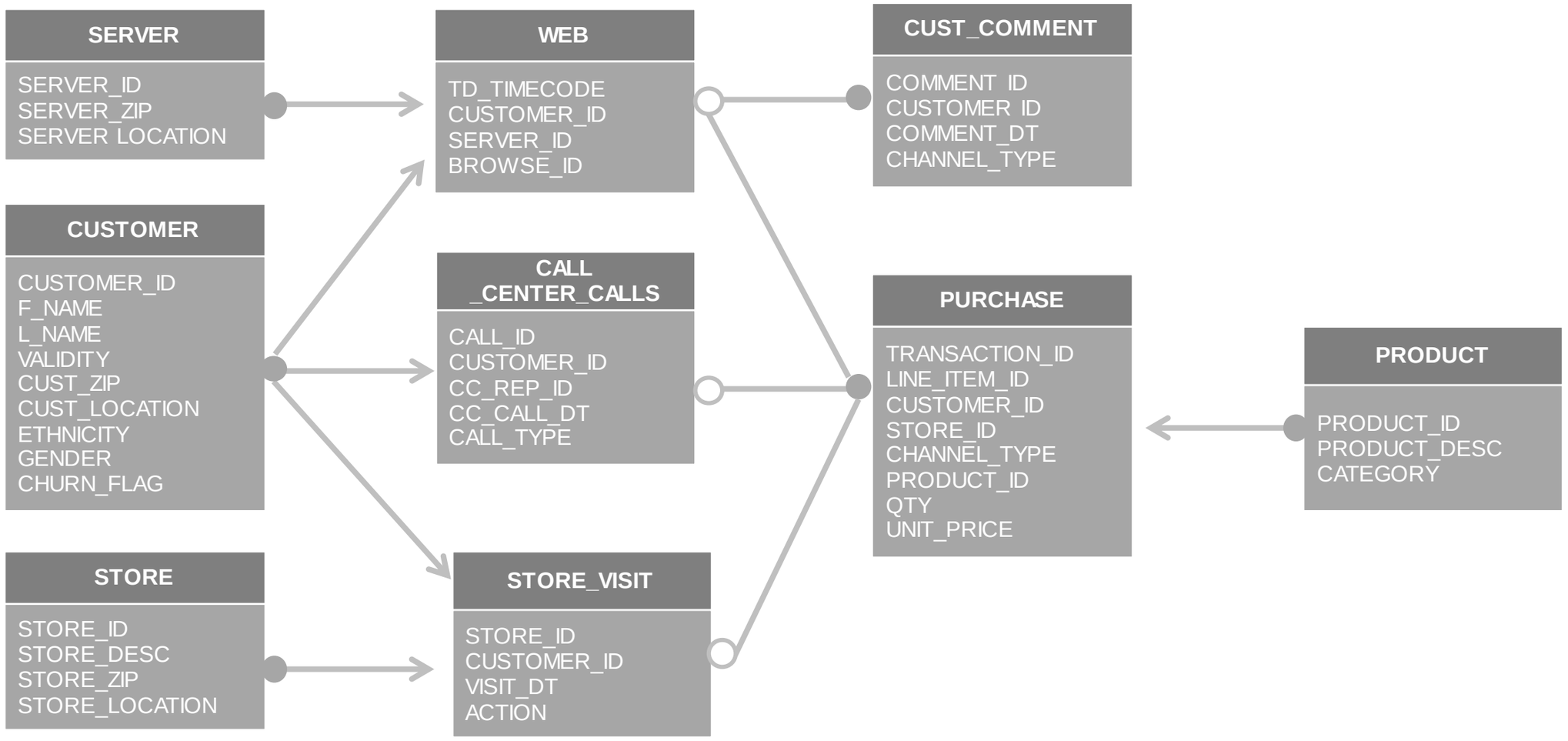
> Identify high-value customers whose purchase interests match our current product promotions and currently live near one of our stores. Send an e-mail with a special discount coupon.

Customer locations (Geospatial)
- Current customers' addresses and distance to our store(s) and competitors
- Beacon data

Customer interests and product availability (Time-based)
- Customers' recently and frequently searched/browsed products
- Current product sell rate and inventory at our store(s)

Valid promotion periods and seasonality (Temporal)
- Store hours
- Current promotion begin/end dates
- Holidays and special events

Data Model


SERVER
SERVER_ID
SERVER_ZIP
SERVER LOCATION
WEB
TD_TIMECODE
CUSTOMER_ID
SERVER_ID
BROWSE_ID
CUST_COMMENT
COMMENT_ID
CUSTOMER_ID
COMMENT_DT
CHANNEL_TYPE
CUSTOMER
CUSTOMER_ID
F_NAME
L_NAME
VALIDITY
CUST_ZIP
CUST_LOCATION
ETHNICITY
GENDER
CHURN_FLAG
CALL _CENTER_CALLS
CALL_ID
CUSTOMER_ID
CC_REP_ID
CC_CALL_DT
CALL_TYPE
PURCHASE
TRANSACTION_ID
LINE_ITEM_ID
CUSTOMER_ID
STORE_ID
CHANNEL_TYPE
PRODUCT_ID
QTY
UNIT_PRICE
PRODUCT
PRODUCT_ID
PRODUCT_DESC
CATEGORY
STORE
STORE_ID
STORE_DESC
STORE_ZIP
STORE_LOCATION
STORE_VISIT
STORE_ID
CUSTOMER_ID
VISIT_DT
ACTION

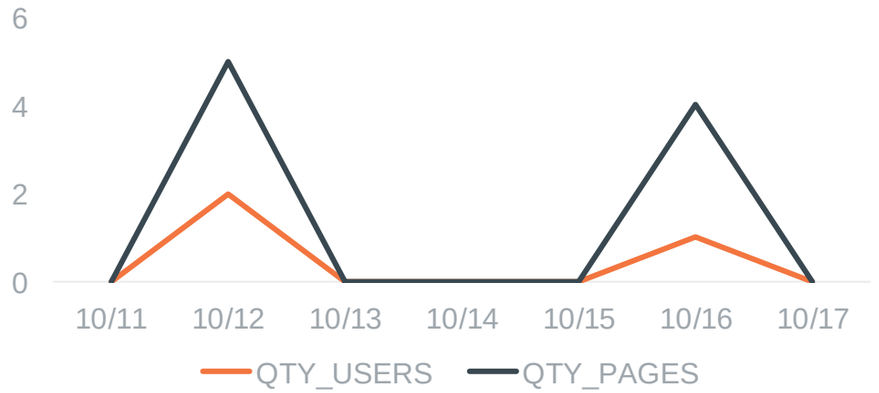
Time Series Example

Time Aware Aggregation

Pages Visited per User		
USER	CLICK TIME	WEB PAGE
333	2018-10-12 09:35:00	home
333	2018-10-12 09:51:00	product
333	2018-10-12 10:05:00	warranty
6740	2018-10-12 03:50:00	home
6740	2018-10-12 04:15:00	product
6740	2018-10-16 04:01:00	home
6740	2018-10-16 04:15:00	product
6740	2018-10-16 09:27:00	review
6740	2018-10-16 09:27:55	checkout

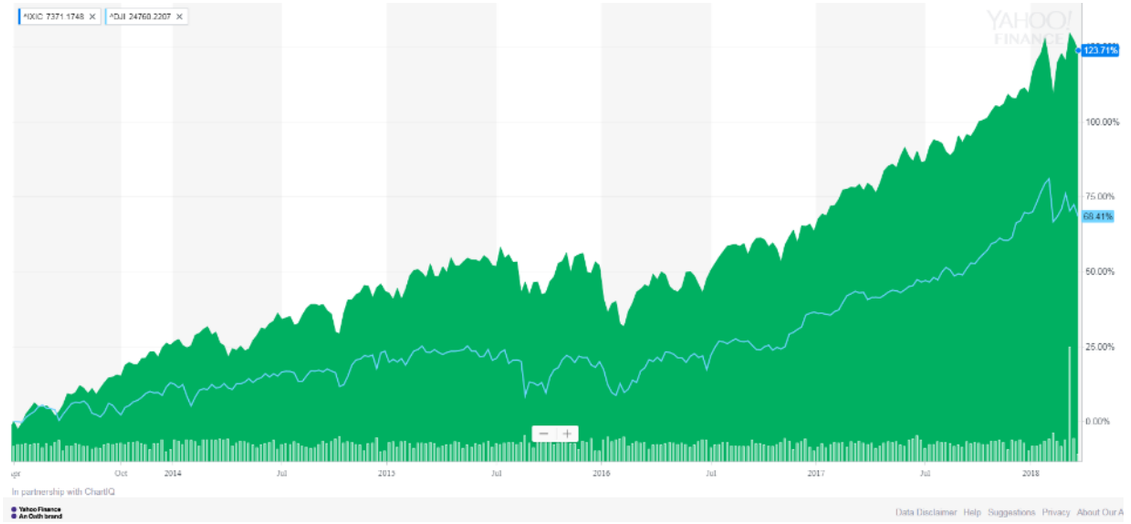
Time Series by Days

RANGETIME	QTY_USERS	QTY_PAGES
2018-10-12 00:00:00	2	5
2018-10-13 00:00:00	0	0
2018-10-14 00:00:00	0	0
2018-10-15 00:00:00	0	0
2018-10-16 00:00:00	1	4



Time-based Data and Analysis

- Understanding data's historical trend
  - Based on natural chronological ordering of data by time (e.g. hour, day, week, etc.)
  - May provide patterns in data based on time
    - Seasonal
    - Upward/downward trend over time
  - "What happened between time X through time Y?"
- May provide basis for forecasting
- Note: Not every time-based data is *time series* data



Source: Yahoo Finance

Teradata Time Series

Time Aware Aggregation Functions and PTI

High-Performance enabled by Primary Time Index (PTI)

- Enables a new way of storing and ordering time-based data (time series as well as date/timestamp data)
- Supports time sensitive decisions
- Fast access through:
  - Hash distribute by time bucket
  - AMP-local processing
  - Sequenced data

Agile Analysis enabled by Time-Aware Aggregate Functions

- Time period aware aggregations
- Work with ANY time component data - PTI or non-PTI
- Handle missing values
  - Ignore or fill with a value

Existing Aggregate Functions

Average	Count
Describe	Kurtosis
Maximum	Minimum
Percentile	Rank
Skew	Sum
Std. population deviation	Std. sample deviation
Population variance	Sample variance

New Aggregate Functions

Bottom	Delta_T
First	Last
Median	Mode
Top	Mean absolute deviation

Teradata Time-Series

Time Aware Aggregation

“For each sensor, show me the average temperature in a ½ hour increment, over 3 hours”

TIMECODE_RANGE	GROUP BY TIME(MINUTES(30))	SENSOR_ID	AVG(TEMPERATURE)
'2017-05-02 08:00:00', '2017-05-02 08:30:00'	1	22	63.5
'2017-05-02 08:30:00', '2017-05-02 09:00:00'	2	22	64.6
'2017-05-02 09:00:00', '2017-05-02 09:30:00'	3	22	65.0
'2017-05-02 09:30:00', '2017-05-02 10:00:00'	4	22	65.1
'2017-05-02 10:00:00', '2017-05-02 10:30:00'	5	22	64.7
'2017-05-02 10:30:00', '2017-05-02 11:00:00'	6	22	64.8
'2017-05-02 08:00:00', '2017-05-02 08:30:00'	1	23	66.4
'2017-05-02 08:30:00', '2017-05-02 09:00:00'	2	23	65.1
'2017-05-02 09:00:00', '2017-05-02 09:30:00'	3	23	64.9
'2017-05-02 09:30:00', '2017-05-02 10:00:00'	4	23	64.8
'2017-05-02 10:00:00', '2017-05-02 10:30:00'	5	23	64.9
'2017-05-02 10:30:00', '2017-05-02 11:00:00'	6	23	65.0

Resources:: Teradata DBS Time Series Technology Overview Gregory H. Milby May 31st, 2017

Temporal Table Example - DataType

Transformation to Build Small Tables

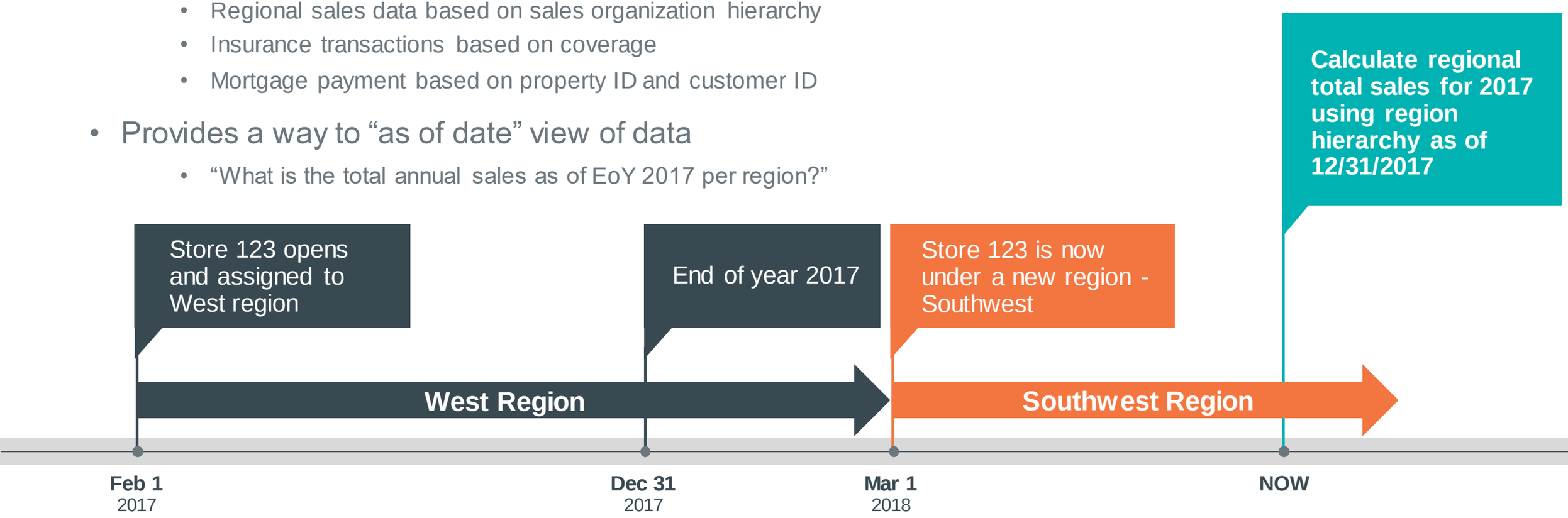
Fidelity Points		
USER	DATE	BALANCE
333	2018-10-12 00:00:00	400
333	2018-10-13 00:00:00	400
333	2018-10-14 00:00:00	400
333	2018-10-15 00:00:00	400
6740	2018-10-12 00:00:00	1000
6740	2018-10-13 00:00:00	1000
6740	2018-10-14 00:00:00	1500
6740	2018-10-15 00:00:00	1500
6740	2018-10-16 00:00:00	1500

Temporal Table

USER	BALANCE	VALIDITY
333	400	(2018-10-12, 2018-10-15)
6740	1000	(2018-10-12, 2018-10-13)
6740	1500	(2018-10-14, 2018-10-16)

Temporal Data and Analysis

- Track history and understand data changes over time
- Analysis of time-referenced data
  - Regional sales data based on sales organization hierarchy
  - Insurance transactions based on coverage
  - Mortgage payment based on property ID and customer ID
- Provides a way to “as of date” view of data
  - “What is the total annual sales as of EoY 2017 per region?”



Temporal Data and Analysis

Automatically tracking data versioning

```
SEQUENCED VALIDTIME
select * from subsrptn_equip_hist order by 1;
```

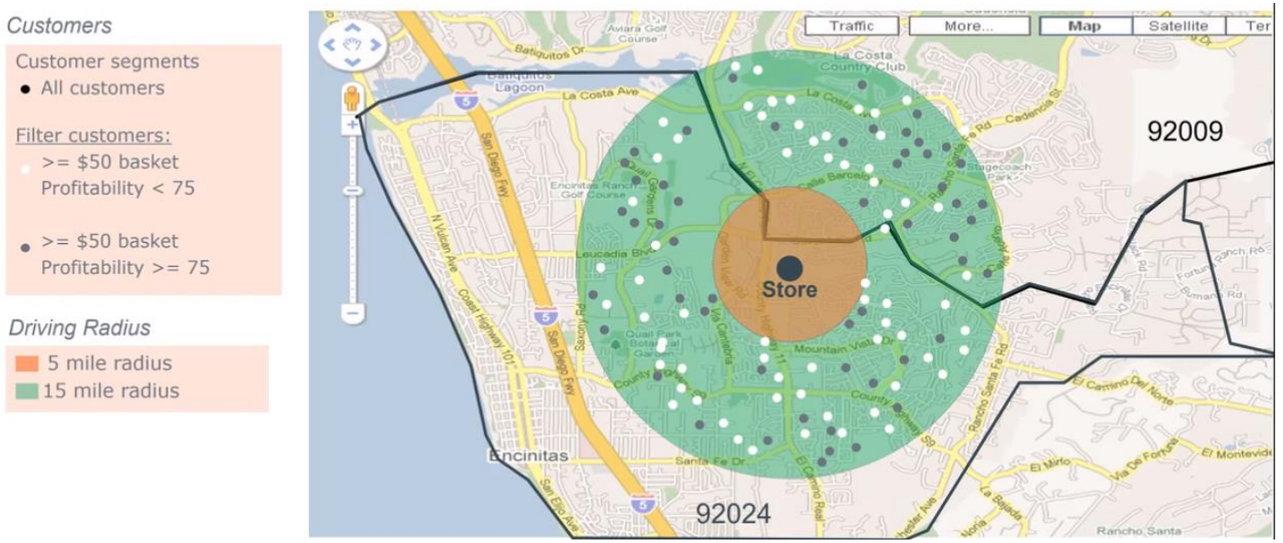
srv_accs_id	equip_nm	VALIDTIME
1234	Razr	('2010-05-10', '9999-12-31')
1235	Blackjack	('2011-01-10', '9999-12-31')
1236	*iPhone*	*('2009-12-10', '2011-05-01')*
1236	*iPhone 4*	*('2011-05-01', '9999-12-31')*
1237	iPhone 3S	('2010-02-28', '9999-12-31')

```
VALIDTIME AS OF date '2011-04-01'
select * from subsrptn_equip_hist order by 1;
```

srv_accs_id	equip_nm
1234	Razr
1235	Blackjack
1236	*iPhone*
1237	iPhone 3S

Geospatial Analysis

- Understanding data in association with a geographical or spatial aspect
  - Typically requires visualization of the data (e.g. map and graphs) to make it meaningful
- Analysis of data in relation to its location
  - How many customers with value scores in the top quartile **live within five miles** of a store?
  - How is promotion response related to home **distance from store**?
  - How many target customers are **within** my sales territory?



Geospatial Capabilities

- Teradata-native data type ST_Geometry supporting most standard geospatial types such as:
  - Point (x y (z)): Single location points such as GPS location
  - Line or curve (xy, xy, xy): Lines and curves to represent roads, tracks, or rivers
  - Polygon (xy, xy, xy, xy..): Represents area objects (e.g. sales regions, neighborhoods, etc.)
  - … and more

Available Geospatial Functions

Attributes

ST_AsBinary

ST_AsText

ST_CoordDim

ST_Dimension

ST_GeometryType

ST_IsEmpty

ST_IsSimple

ST_IsClosed

ST_NumPoints

ST_SRID

…

Spatial Operators

ST_Buffer

ST_Intersection

ST_Boundary

ST_Difference

ST_Envelope

ST_ExteriorRing

ST_GeometryN

ST_InteriorRingN

ST_Transform

SimplifyPreserveTopology

Spatial Relationships

ST_Intersects

ST_Overlaps

ST_Touches

ST_Within

ST_Contains

ST_Crosses

ST_Equals

…

Spatial Table Operators

AggGeom

PolygonSplit

GeometryToRows

Measurements

ST_Area

ST_Distance

ST_SphericalDistance

ST_SpheroidalDistance

ST_Perimeter

ST_Length

Topics

- Introduction
- 4D Analytics
- **SQLE Analytic Functions**

TIME SERIES DATA PREP: Sessionization

What Is It:

- **Sessionization** function stitches together all activities within a time window and designates an event as having occurred within that window (session)
- Sessionization is required to prepare the input for nPath

Sample Use Cases:

- All path analysis use cases:
  - Fraud detection
  - Service usage attrition
  - Customer churn
  - Network security threat determination

Example:

- Session 1: Visitor comes to a website and spends the first 30 minutes (time window) browsing the products page
- Session 2: Visitor spends the next 30 minutes reading product reviews
- The two sessions and others are used to determine the sequential sessions (path) taken to a product purchase

Visual: List of Sessions Outputted

ID	CLICK TIME	USER ID	PRODUCT	PAGE TYPE	REFERRER	PRICE	USER SESSION ID
1	1110000	333		home	www.yahoo.com		0
1	1112000	333	Ipod	checkout	www.yahoo.com	200.2	0
1	1160000	333	bose	checkout		340	0
1	1200000	333		home			0
1	1203000	67403		home	www.google.com		0
1	1300000	67403		home	www.google.com		1
1	1301000	67403		home			1
1	1302000	67403		Home			1

Sessionization Example

Starting at 0, create session ID for each day, grouping of events per customer

Pages Visited per User			
USER	CLICK TIME	WEB PAGE	**SESSION ID**
333	2018-10-12 09:35:00	home	**0**
333	2018-10-12 09:51:00	product	**0**
333	2018-10-12 10:05:00	warranty	**0**
6740	2018-10-12 03:50:00	home	**0**
6740	2018-10-12 04:15:00	product	**0**
6740	2018-10-16 04:01:00	home	**1**
6740	2018-10-16 04:15:00	product	**1**
6740	2018-10-16 09:27:00	review	**1**
6740	2018-10-16 09:27:55	checkout	**1**

Sessions by day

User 333 Session 0

User 6740 Session 0

User 6740 Session 1

PATTERN MATCHING: nPath

What Is It:

- **nPath** is a behavioral analytic technique designed for time-series sequence analysis of data to link an outcome with a preceding set of events

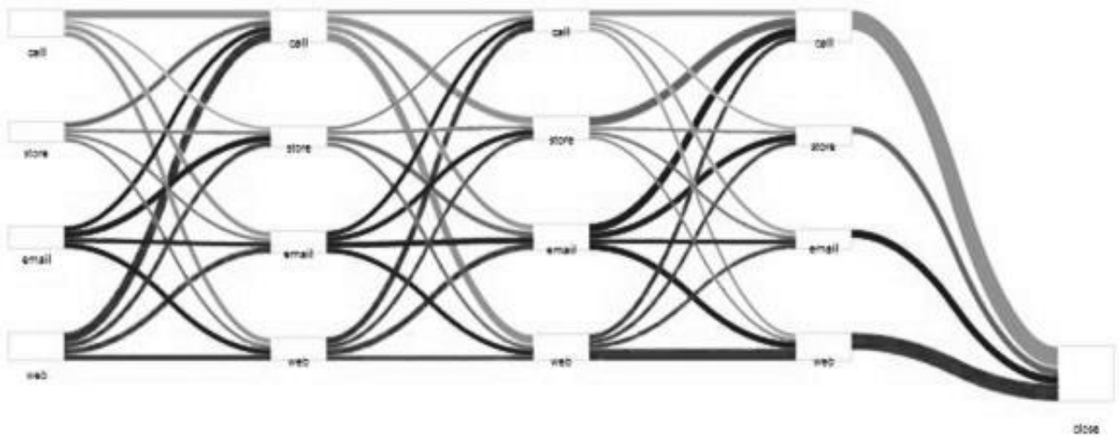
Sample Use Cases:

- Channel activities that lead to customer churn
- Web clickstream activities that lead to a product sale
- Fraud pattern identification in financial transactions

Example:

- Customers interact through stores, online portal, and call centers
- Some customers are cancelling their service contracts
- Identify all incident combinations over time (paths) that lead to service cancellation

Visual: Output is a Sankey Diagram



nPath Example

Transpose events into sequences

Pages Visited per User			
USER	CLICK TIME	WEB PAGE	**SESSION ID**
333	2018-10-12 09:35:00	home	**0**
333	2018-10-12 09:51:00	product	**0**
333	2018-10-12 10:05:00	warranty	**0**
6740	2018-10-12 03:50:00	home	**0**
6740	2018-10-12 04:15:00	product	**0**
6740	2018-10-16 04:01:00	home	**1**
6740	2018-10-16 04:15:00	product	**1**
6740	2018-10-16 09:27:00	review	**1**
6740	2018-10-16 09:27:55	checkout	**1**

Sessions by day

User 333 Session 0

User 6740 Session 0

User 6740 Session 1

Paths by User & Session

[home, product, warranty]

[home, product]

[home, product, review, checkout]

nPath Example

Accumulates rows and 'flattens' into single CSV row

Often used with SUM to view most popular paths across all partitions

```
SELECT * FROM NPATH
(ON RETAIL.DAILY_SESSIONS_NOCHURN
PARTITION BY CUSTOMER_ID
    ORDER BY DATESTAMP
    USING
    MODE(NONOVERLAPPING)
    PATTERN('E*.C')
    SYMBOLS  ( EVENT = 'Purchase'  AS C
                     , EVENT <> 'Purchase' AS E )
    RESULT    (FIRST(CUSTOMER_ID OF ANY(E,C))          AS CUSTOMER_ID
               ,FIRST(DATESTAMP OF ANY(E,C))           AS DS_START
               ,LAST(DATESTAMP OF ANY(E,C))            AS DS_END
               ,COUNT(* OF E)                          AS EVENT_CNT
               ,ACCUMULATE(EVENT OF ANY(E,C))          AS PATH )
) AS DT;
```

	CUSTOMER_ID	DATESTAMP	EVENT	CHURN_FLAG
1	1895	2018-04-17 19:19:00.000000	Product Browsing	N
2	1895	2018-04-17 19:20:00.000000	Store Visit	N
3	1895	2018-04-17 19:23:00.000000	Service Inquiry	N
4	1895	2018-04-17 19:36:00.000000	Purchase	N

Result Set - 2.1 - nPath - Solution.sql (1)

	customer_id	ds_start	ds_end	event_cnt	path
1	1895	2018-04-17 19:19:00.000000	2018-04-17 19:36:00.000000	3	[Product Browsing, Store Visit, Service Inquiry, Purchase]

ATTRIBUTION: Attribution

What Is It:

- **Attribution** function calculates the contribution (weight) of an event to a specific outcome
- Weights are determined based on the attribution model type:
  - Uniform (equal weights to all events)
  - Exponential (successive events given more weight than previous ones)
  - First in (all weight to the first event)
  - Last in (all weight to the last event)

Sample Use Cases:

- Identifying the strength of key contributors to specific outcomes based on which operational decisions (e.g., which channel to advertise on) are taken
- Determining inefficient processes that contribute to outcomes

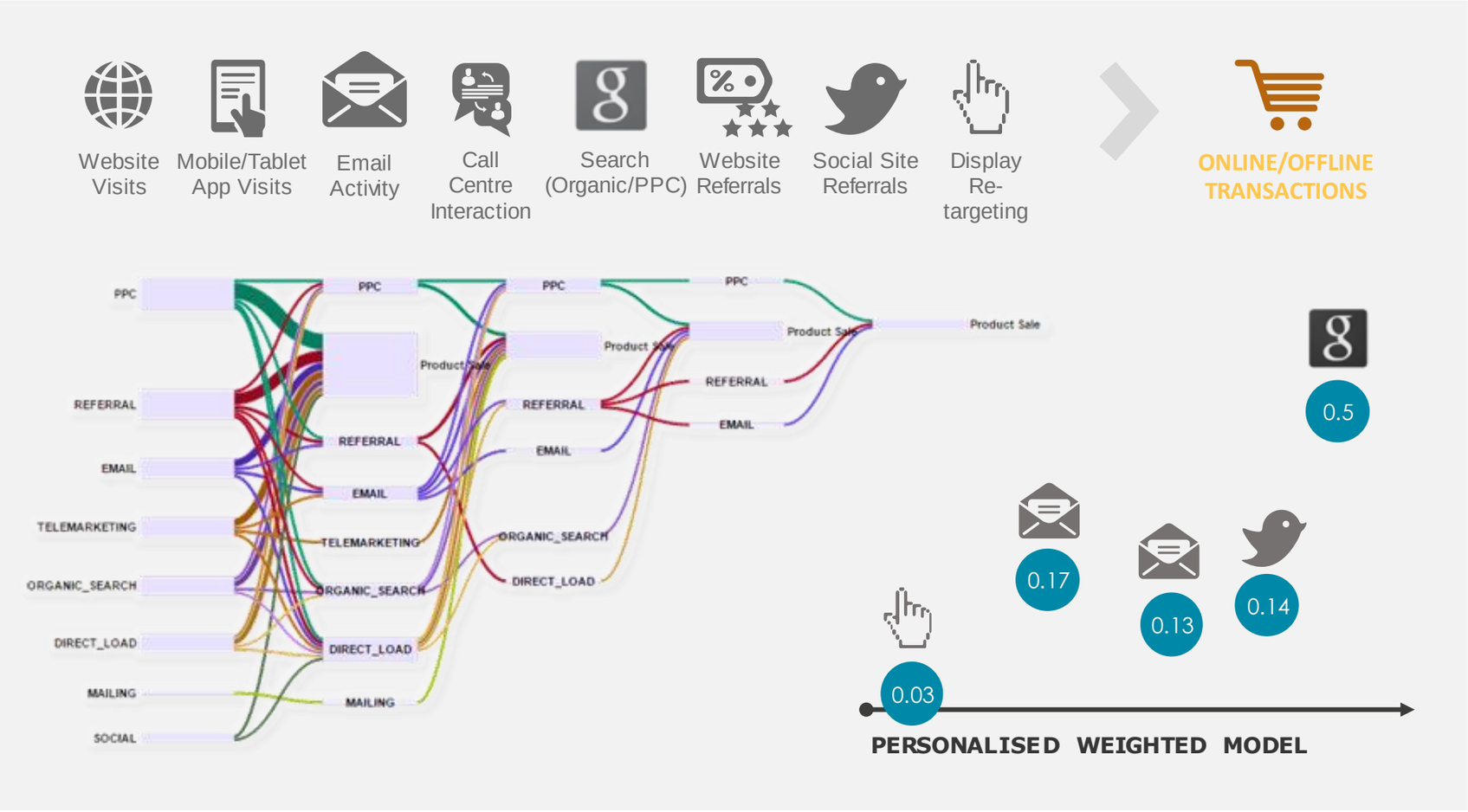
Example:

- Customer clicked on a banner ad, signed up for promo email, received emails, visited the website, saw ad in newspaper, heard radio ad, liked Facebook page, then made a purchase
- Attribution scores pinpoint which channel contributed the most to product purchase

Visual: Weighted Chart of Actions

ID	EVENT	TIMESTAMP	ATTRIBUTION	TIME TO CONVERSION
1	Store Visit	09/27/2017 23:00	0.09	-19
1	Paper Ad	09/29/2017 23:00	0.09	-17
1	Online Visit	10/01/2017 23:00	0.09	-15
1	Online Ad Click	10/03/2017 23:00	0.09	-13
1	Browse Offers	10/05/2017 23:00	0.09	-11
1	Store Visit	10/07/2017 23:00	0.09	-9
1	Call Center Question	10/09/2017 23:00	0.09	-7
1	Email Click	10/11/2017 23:00	0.19	-5
1	Visit Online	10/15/2017 23:00	0.18	-1
1	Purchase Product	10/16/2017 23:00		

Attribution: Path to Acquisition



RESULT (Retailer):
10% Saving in Annual Digital Marketing Budget

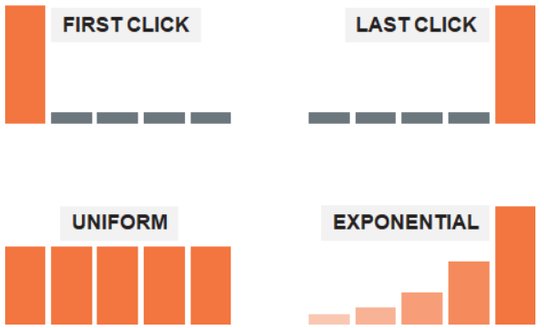
Sample Attribution Table 'Weight' Types

Weight types determines weight percentage the qualifying row receives

	CUSTOMER_ID	DATESTAMP	EVENT	FIRST_CLICK_AT...	LAST_CLICK_AT...	UNIFORM_ATTR...	EXPONENTIAL_...
1	66592	2018-05-03 13:30...	Complaint Call	1	0	0.2	0.032
2	66592	2018-05-04 19:33...	Web Chat	0	0	0.2	0.065
3	66592	2018-05-04 19:37...	Product Browsing	0	0	0.2	0.129
4	66592	2018-05-05 23:25...	Online Feedback	0	0	0.2	0.258
5	66592	2018-05-05 23:26...	Return Policy In...	0	1	0.2	0.516
6	66592	2018-05-07 09:12...	Product Return	null	null	null	null

Weight types

- **First Click** – Score only first click (Complaint Call)
- **Last Click** – Score only last click (Return Policy Inquiry)
- **Uniform** – Each click gets equal weight (.2 for all 5 clicks)
- **Exponential** – Starting at last click, decay by defined %



! Total Weight must equal 1

DATA PREP / TRANSFORMATION: Antiselect

What Is It:

- **Antiselect** function enables the user to specify the set of columns that are NOT needed when selecting data from a table with a large number of columns.

Sample Use Cases:

- Needing to analyze a data set with a large number of columns in its entirety with the exception of a small number of columns.
- Applicable for implementation across all use cases in any vertical.

Example

Input Dataset:

ID	SRCE	AGE	GENDER	RACE	NUM BUYS	NUM SELLS
1	Site A	62	Male	White	30	44
2	Site B	29	Female	Asian	33	23

Antiselect Output: An anti-select on SRCE, AGE, RACE will give us:

ID	GENDER	NUM BUYS	NUM SELLS
1	Male	30	44
2	Female	33	23

Antiselect Example

created_at	id	full_text	retwee	fav	user_id	user_nam	user_screen_na	user_description	user_location	user_created_at
2019-12-05 05:54:00	1	Australia's PM Launches	0	0	8.16E+17	greeen	greeenorg	followback	[NULL]	2017-03-01 01:58:00
2020-01-05 21:24:00	1	Australian election 2019	0	0	17664300	jeisea?	jeisea	Living in the Shire of Byron and	Byron Bay, Australia	2010-03-11 00:29:00
2019-11-05 07:28:00	1	Australian election 2019	1	0	114878082	Ken	QldProgressive	An Australian Progressives Que	City of Gold Coast (	2011-04-02 22:18:00
2020-01-05 11:35:00	1	@CliveFPalmer If your c	0	0	1.05E+18	Steve	Istvan04922424	[NULL]	[NULL]	2019-05-10 23:34:00
2019-12-05 05:54:00	1	It's time to take your lea	0	0	1.07E+18	rintinhtinh?	rintinhtinh	[NULL]	Australia	2020-05-11 01:13:00
2020-02-05 11:18:00	1	@opa1420 Liberal party	3	2	1715727252	mira mcnair	McnairMira	[NULL]	[NULL]	2015-07-08 13:56:00
2019-11-05 07:28:00	1	@collias_bill Labors Mini	9	14	1.04E+18	ALLRight	EternalSaint	R.N Midwife Lady :Loves God,h	KEEP LaborOUT put	2019-04-09 09:56:00

```
SELECT * FROM Antiselect (
ON aus_election_twitter
USING
Exclude ('id', 'user_created_at')
) AS dt
```

created_at	full_text	retweet_c	favorit	user_id	user_name	user_screen_name	user_description	user_location
2020-04-05 09:27:00	@knarfnamduh But ask Morrison or Barnaby Joy	1	1	3341241538	Mel	mel_giancarlo	An Aussie en Suisse, I twe	Switzerland
2020-02-05 11:18:00	@opa1420 Liberal party breached electoral laws	3	2	1715727252	mira mcnair	McnairMira	[NULL]	[NULL]
2020-06-05 07:36:00	As an immigrant who made Australia home 8 yea	0	0	8.50E+17	Yak Suri	freakflyers	Full time nomad.... part tir	Northern Territory
2020-01-05 11:35:00	@CliveFPalmer If your coming to Western Austra	0	0	1.05E+18	Steve	Istvan04922424	[NULL]	[NULL]
2019-11-05 18:26:00	Top story: Australia Votes: Election words - Learı	0	0	8.78E+17	Art Icon	IdiomsWorld	Give a new meanings for y	[NULL]
2020-01-05 21:24:00	Australian election 2019: how to avoid voting for	0	0	17664300	jeisea?	jeisea	Living in the Shire of Byro	Byron Bay, Australia
2020-07-05 15:51:00	'We have lost Australia for now,' warns climate sc	0	0	50645038	Frank Langenfeld	FELDart	The focus of humanistic tl	planet earth
2019-11-05 07:28:00	Australian election 2019: how to avoid voting for	1	0	114878082	Ken	QldProgressive	An Australian Progressive	City of Gold Coast Qu

DATA PREP / TRANSFORMATION: Pack / Unpack

What Is It:

- **Pack** transforms the input table columns and merges all columns into a row. Packing columns frees up disk space, and speeds query results retrieval.

- **Unpack** function expands data from a single packed column into multiple columns.

Sample Use Cases:

- In healthcare diagnostic data is available as strings and unpack is used to deconstruct into columns for analysis
- Speedy information retrieval from data sources where packed data is stored for efficiency

Examples

Pack Input:

ID	SOURCE	AGE	GENDER	RACE	NUM BUYS	NUM SELLS
1	site_a	62	male	white	30	44

Pack Output:

ID	PACKED DATA
1	src:site_a,age:62,gender:male,race:white,numBuys:30,numSells:44

Unpack Input:

ID	SOURCE	PACKED DATA
1	Site_a	62,male,white,30,44

Unpack Output:

AGE	GENDER	RACE	NUM BUYS	NUM SELLS	ID	SOURCE
62	male	white	30	44	1	Site_a

Unpack Example

created_at	id	full_text	retweet_count	favorite_count	user_id
2020-04-05 09:27:00	1	@knarfnamduh But ask Morrison or Barnaby Joy	1	1	3341241538
2020-02-05 11:18:00	1	@opa1420 Liberal party breached electoral laws	3	2	1715727252
2020-06-05 07:36:00	1	As an immigrant who made Australia home 8 yea	0	0	8.50E+17
2020-01-05 11:35:00	1	@CliveFPalmer If your coming to Western Austra	0	0	1.05E+18
2019-11-05 18:26:00	1	Top story: Australia Votes: Election words - Learr	0	0	8.78E+17
2020-01-05 21:24:00	1	Australian election 2019: how to avoid voting for	0	0	17664300
2020-07-05 15:51:00	1	'We have lost Australia for now,' warns climate sc	0	0	50645038
2019-11-05 07:28:00	1	Australian election 2019: how to avoid voting for	1	0	114878082

```
SELECT * FROM Unpack (
        ON (select full_text as fl
                        ,aus_election_twitter.*
                from USER1.aus_election_twitter
            )
        USING
                TargetColumn ('fl')
                OutputColumns ('first mention')
                OutputDataTypes ('VARCHAR')
                Regex('(@[\S]+)')
                IgnoreInvalid ('true')
) AS dt;
```

first mention	created_at	id	full_text	retweet_count
@knarfnamduh	2020-04-05 09:27:00	1	@knarfnamduh But ask Morrison or Barnaby Joy	1
@opa1420	2020-02-05 11:18:00	1	@opa1420 Liberal party breached electoral laws	3
@andrewbrodie331	2020-06-05 07:36:00	1	@andrewbrodie331 #ausvotes here in Australia t	0
@CliveFPalmer	2020-01-05 11:35:00	1	@CliveFPalmer If your coming to Western Austra	0
@BillHareClimate	2020-04-05 03:13:00	1	How Australia?s election will decide its role in #c	4
@BBCBreaking	2020-06-05 17:25:00	1	@BBCBreaking Not unexpected, not shock result	3
@wine_australia	2020-03-05 01:12:00	1	A different angle on election coverage from @wi	0
@collias_bill	2019-11-05 07:28:00	1	@collias_bill Labors Mining tax fail¶Pink bats fail'	9
@donaldh66287394	2020-07-05 07:19:00	1	@donaldh66287394 @JeffreyMeursing @Odder	0
@mirandadevine	2020-05-05 07:53:00	1	@mirandadevine This issue is nearly as divisive a	0
@annabelcrabb	2020-06-05 07:36:00	1	@annabelcrabb Australia was the second countr	0

Pack Example

created_at	id	full_text	retweet_count	favorite_count	user_id
2020-04-05 09:27:00	1	@knarfnamduh But ask Morrison or Barnaby Joy	1	1	3341241538
2020-02-05 11:18:00	1	@opa1420 Liberal party breached electoral laws	3	2	1715727252
2020-06-05 07:36:00	1	As an immigrant who made Australia home 8 yea	0	0	8.50E+17
2020-01-05 11:35:00	1	@CliveFPalmer If your coming to Western Austra	0	0	1.05E+18
2019-11-05 18:26:00	1	Top story: Australia Votes: Election words - Learı	0	0	8.78E+17
2020-01-05 21:24:00	1	Australian election 2019: how to avoid voting for	0	0	17664300
2020-07-05 15:51:00	1	'We have lost Australia for now,' warns climate sc	0	0	50645038
2019-11-05 07:28:00	1	Australian election 2019: how to avoid voting for	1	0	114878082

```
SELECT * FROM Pack (
        ON USER1.aus_election_twitter
        USING
                TargetColumns ('retweet_count'
                               ,'favorite_count')
                Delimiter ('-')
                IncludeColumnName ('true')
                OutputColumn ('packed_data')
) AS tb
```

packed_data	created_at	id	full_text	user_id
retweet_count:0-favorite_count:1	2019-10-05 17:08:00	1	With #Australia facing a knife-edge federal poll c	4853065283
retweet_count:0-favorite_count:1	2019-10-05 17:10:00	1	THE GREAT FLUSHING OUT!¶Toxic Speech Floo	64001425
retweet_count:0-favorite_count:4	2019-10-05 17:17:00	1	Proud to see Waleed Aly on prime time CNN in T	28261687
retweet_count:0-favorite_count:0	2019-10-05 17:26:00	1	Dutton and France in Dickson divided by politics	88890364
retweet_count:0-favorite_count:0	2019-10-05 17:27:00	1	"Here is what you need to know about the partie	55545669
retweet_count:0-favorite_count:0	2019-10-05 17:32:00	1	Election Day Is Coming. Australia Says: 'Meh.' - T	1267062757
retweet_count:1-favorite_count:0	2019-10-05 17:35:00	1	Two weeks out from the general election, #Austr	80876514
retweet_count:0-favorite_count:1	2019-10-05 17:44:00	1	@VicGarciaNou @robertoantoniow Tambi?n en A	9.60E+17
retweet_count:0-favorite_count:2	2019-10-05 17:48:00	1	I cant wait until tuesday but at the same time im	2828559474
retweet_count:0-favorite_count:0	2019-10-05 18:02:00	1	Being Australia's biggest electorate puts its vote	9.38E+17
retweet_count:0-favorite_count:0	2019-10-05 18:19:00	1	@JoshFrydenberg @billshortenmp @Bowenchris	625172280

STATISTICAL ANALYSIS: Simple Moving Average

What Is It:

- **Simple Moving Average (SMA)** is calculated by adding current data to and dropping the oldest data from the series and then dividing the total by the number of time periods.

Sample Use Cases:

- The simplest of all moving average calculations, it is used to determine basic data patterns such as outliers as well as comparing trends over time.
- Ability to filter out data noise and t data volatility into more discernible trends.

SMA Calculation:

$$SMA = \frac{A_1 + A_2 + \ldots + An}{n}$$

SMA Trend

STATISTICAL ANALYSIS: Cumulative Moving Average

What Is It:

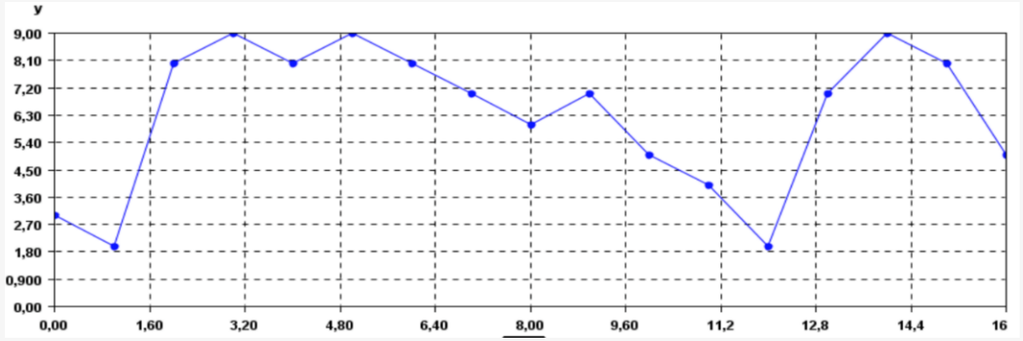
- **Cumulative Moving Average (CMA)** computes an average on data that arrive in an ordered stream. When more data arrive, the average is recalculated with the new data point. Unlike SMA the oldest data point is NOT removed when the CMA is computed.

Sample Use Cases:

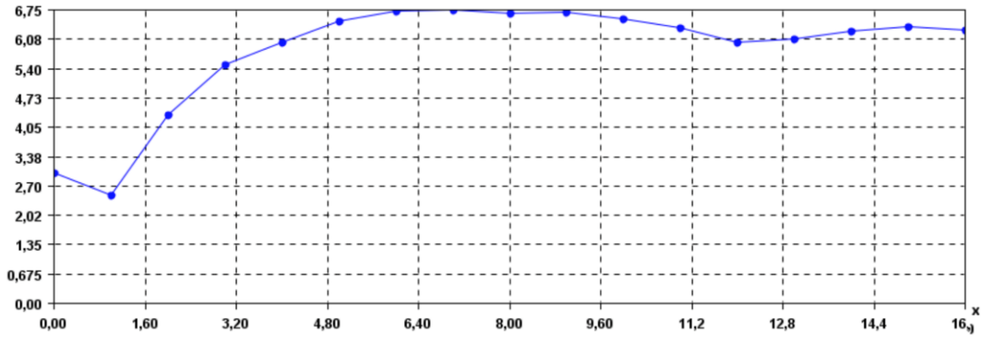
- CMA is commonly used with time series data to smooth out short-term fluctuations and highlight longer-term trends or cycles.
- For example, it is often used in technical analysis of financial data, like stock prices, returns or trading volumes.

Example:

Original time series



CMA time series: Notice how the average changes as more observations are added.



STATISTICAL ANALYSIS: Exponential Moving Average

What Is It:

- **Exponential Moving Average (EMA)** computes the average over a number of points in a time series while applying an exponentially decaying or damping (weighting) factor so that more recent values are given a heavier weight.

Sample Use Cases:

- Stock price movements are analyzed with different moving average indicators based on which trends can be prognosticated.

EMA Calculation:

- Initial simple moving average: 1-period total / 1
- 2nd simple moving average: 2-period total / 2
- Multiplier for 2nd period: (2 / (Time periods + 1)) = (2 / (1 + 1)) = 1 (100%)
- 2nd time period EMA: {2nd time period's value - EMA(1st period*)} x Multiplier + EMA(1st period*)

…

- Multiplier 12th period: (2 / (Time periods + 1)) = (2 / (11 + 1)) = 0.1666 (17%)
- 12th time period EMA: {12th time period's value - EMA(11th period)} x Multiplier + EMA(11th period)

* Note 1st Period's EMA is the SMA of the 1st period which is also the 1st period value

Comparing EMA to SMA

STATISTICAL ANALYSIS: Triangular Moving Average

What Is It:

- **Triangular Moving Average (TMA)** is a double smoothed (double averaged) metric that averages the averages of the SMAs of each period.

Sample Use Cases:

- Due to the way it is calculated (double average) there is a greater emphasis made on the middle of the series. This enables to analyze the rate at which change in the data occur. This is a big metric used in financial calculations.

TMA Calculation:

- First, calculate the simple moving average (SMA) for, say, 4 time periods:
  SMA = (P1 + P2 + P3 + P4) / 4
- Then, take the average of all the SMA values to get TMA values.
  TMA = (SMA1 + SMA2 + SMA3 + SMA4)/4
- The TMA can also be expressed as: TMA = SUM (SMA values) / N

Comparing TMA to SMA

STATISTICAL ANALYSIS: Weighted Moving Average

What Is It:

- **Weighted Moving Average (WMA)** gives you a weighted average of the last **n** data points (e.g. prices), where the weighting *decreases* with each previous price.

- Similar to EMA but the calculation is done slightly differently and can be customized (unlike EMA)

Sample Use Cases:

- Stock price movements are analyzed with different moving average indicators based on which trends can be prognosticated.
- Also used to test out different trend predictions based on various weighting factors.

WMA Calculation:

- Weighted moving average calculation
  = (Current data * weighting factor)
  + (Previous period data * weighting factor-1)

Comparing WMA to SMA

MovingAverage Example

123 par	123 id	123 val
1	1	2
1	2	2
1	3	2
1	4	3
1	5	4

```
SELECT * FROM MovingAverage (
        ON   moving_average
        PARTITION BY par
        ORDER BY id asc
        USING
                MAvgType ('C')
                TargetColumns ('val')
) AS dt ORDER BY id;
```

123 par	123 id	123 val	123 val_cmavg
1	1	2	2
1	2	2	2
1	3	2	2
1	4	3	2.25
1	5	4	2.6

Type	Result
'C' (Default) Cumulative moving average.	C= 2.6
'E' Exponential moving average.	Alpha= 0.5, E = 3.25
'M' Modified moving average.	Window size = 5, M = 2.56
'S' Simple moving average.	Window size = 5, S = 2.6
'T' Triangular moving average.	Window size = 5, T = 2.44
'W' Weighted moving average.	Window size = 5, W = 2.93

Thank you.

teradata.