Busting eCommerce Scammers with Splunk

Measuring Transactional Risk

J.R. Murray | VP Technical Services, Gemini Data
Juan Morales | Industry Professional

August 2018 | Version 1.2

Forward-Looking Statements

During the course of this presentation, we may make forward-looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC.

The forward-looking statements made in this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward-looking statements we may make. In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

splunk> .conf18

Our Speakers

J.R. MURRAY

VP Technical Services, Gemini Data

@jrzmurray

JUAN MORALES

Industry Professional

@juanmmoralesv

Order Risk Problem Space

eCommerce Risk Mitigation

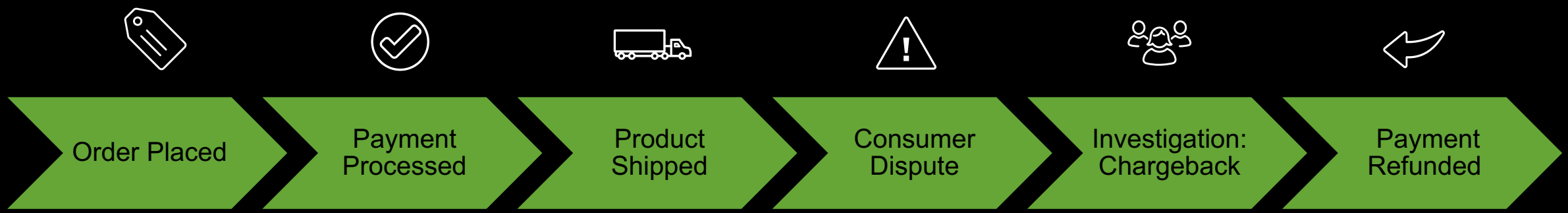
Mission

1. Prevent payment card chargebacks
2. Prevent shipment of goods
3. Prevent campaigns/repeat offenders

What is a Chargeback?

a frustrating threat to your livelihood

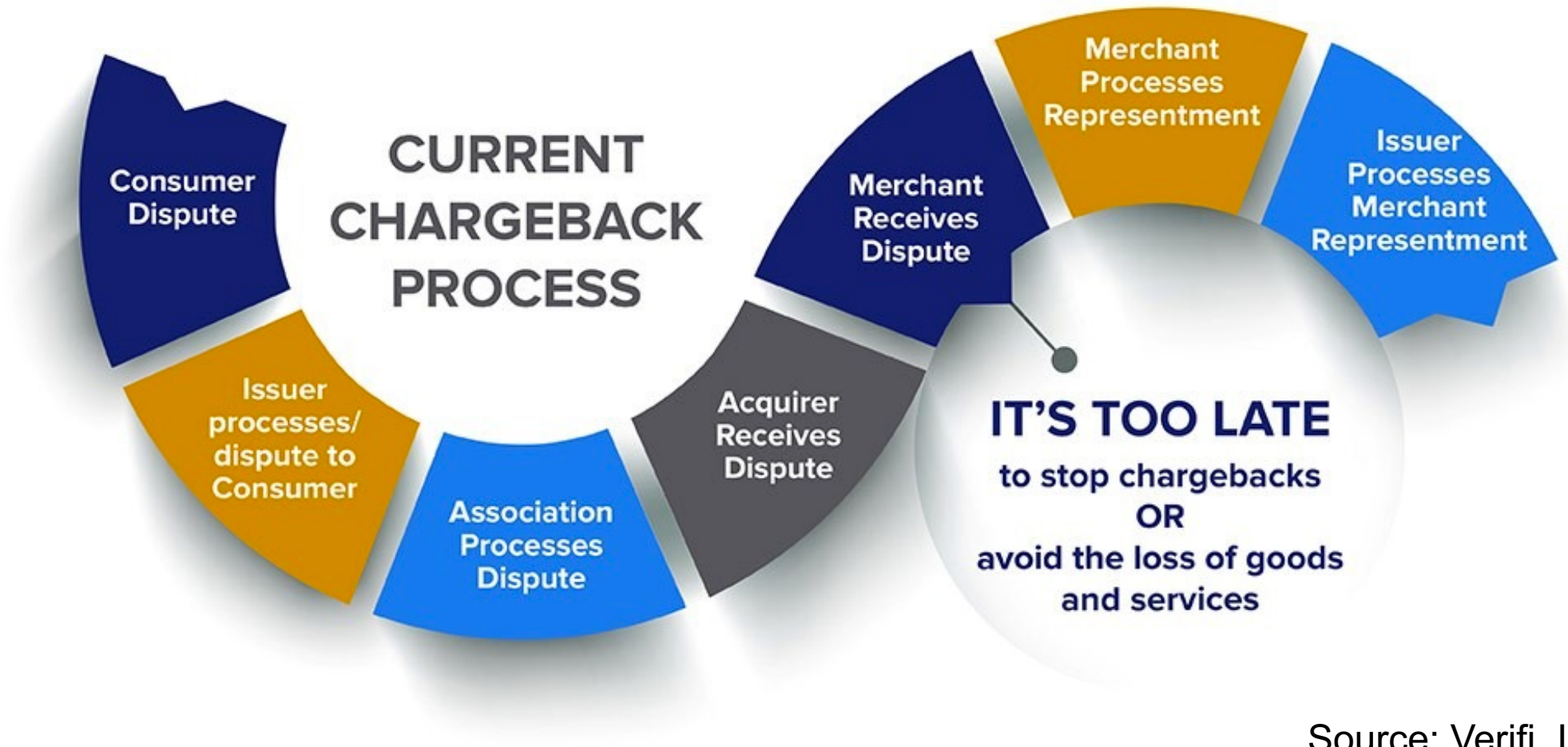
A Chargeback is a reversal of a payment originally from a consumer to a merchant that is forced on the merchant by a financial institution.



Chargebacks were developed and are utilized as a consumer protection control.

How Does the Process Work?

Your stuff becomes their stuff.



Source: Verifi, Inc.

The Problem & The Effects

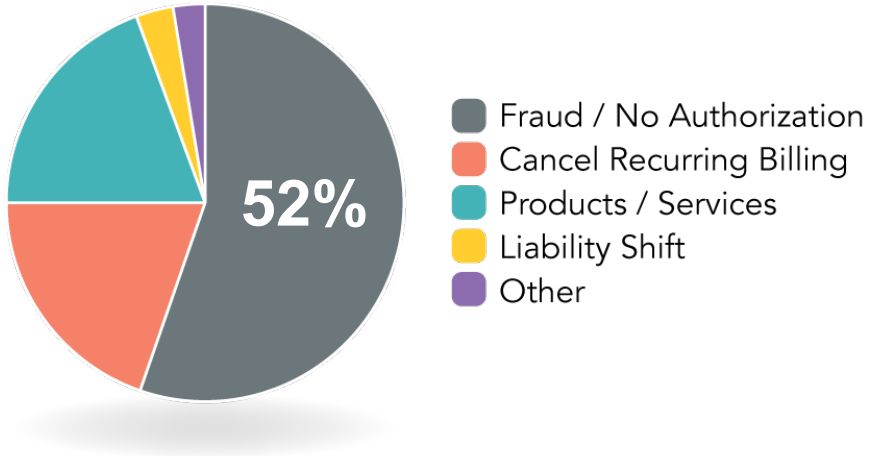
Risks for Payment Reversal

$19B

Merchants chargeback problem in 2017*

+$1.50

Additional cost for merchants and issuers per dollar spent



Overall, "Fraud/No Authorization" chargebacks account for 52% of all disputes**

*Sources: *Javelin Strategy & Research (May 2018) – Published by CNP (Card Not Present)*
**https://chargeback.com/5-chargeback-reason-code-categories/

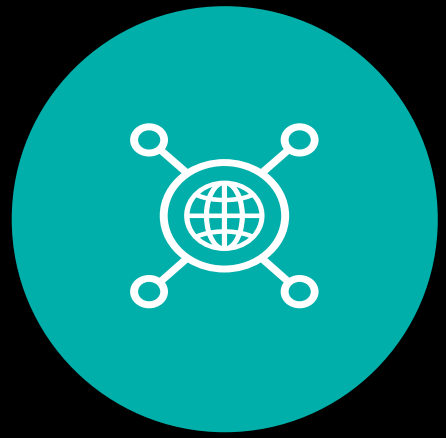
Not only is cybersecurity a risk for online merchants, but accepting high-risk payments can also affect when having to refund the charges.

splunk> .conf18

One Response (Among Others)

Using the Power and Extensibility of Splunk

Rule Engine with Workflow Management



Search-Time Data Enrichment with Risk Scoring

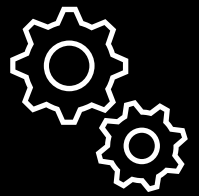
Advanced Searches and Visualizations

Rule Engine with Workflow Management

splunk> .conf18

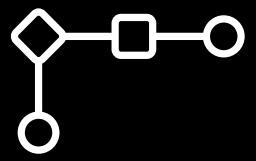
Workflow Management Goals

Pillars of Success



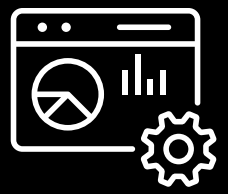
Rule Engine

- Scheduled Alerts
  - **Velocity**
  - **Blacklist matches**
  - **High Risk Score**
- Advanced Macros
  - **Automatic aggregation of overlapping alerts**
  - **Zero duplication**



Interactive Workflow

- Delete orders from cases
- Blacklist order attributes
- Expand table rows
- Add case comments
- Set order flags
  - **Bad Actor**
  - **Stopped shipping/billing**
  - **False positive (et al.)**

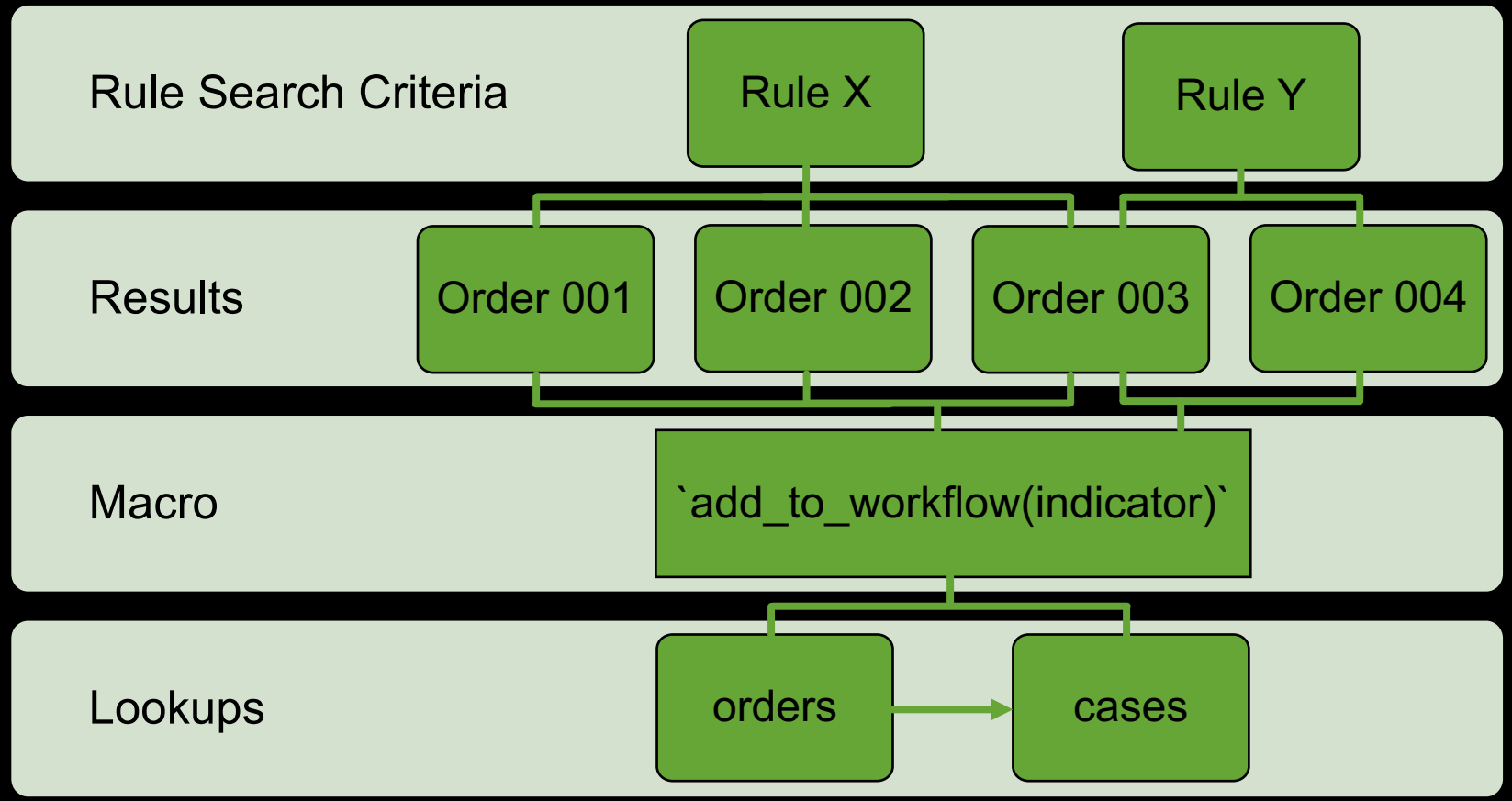


Organization

- Cases with order grouping
- Campaigns with case grouping
- Activity auditing
- Metrics and reports

Rule Engine - Workflow

Status = “It’s Complicated”



Rule Engine - Example

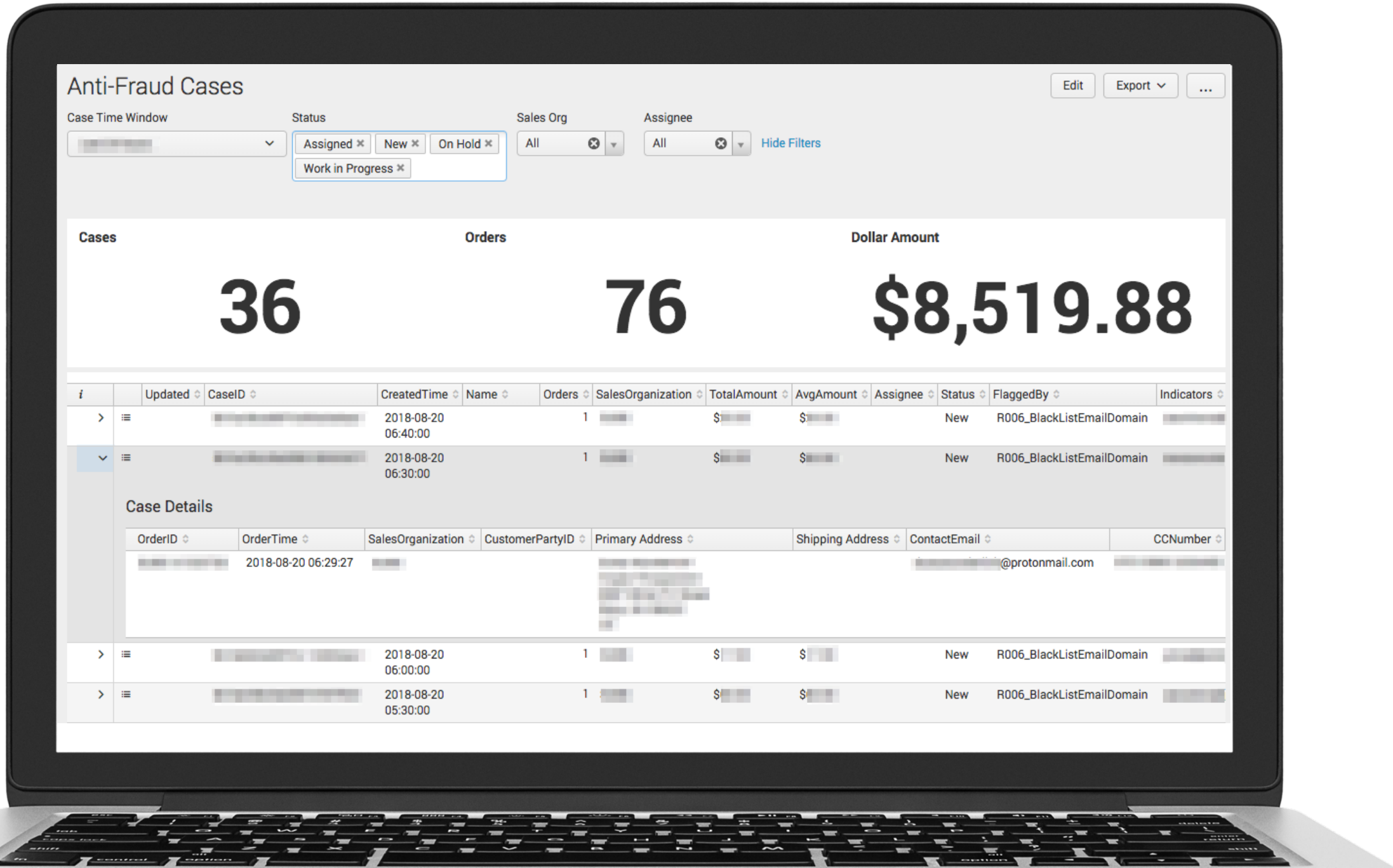
Writing to two lookups in one search

```
index=orders sourcetype=order | lookup orders _key AS OrderID
[ rule filter / logic ]
[ filtered lookup for *open* cases w/matching $indicator ] # output _key=NewCaseID
| eval CaseID=coalesce(CaseID, NewCaseID)
[ add the case fields / set the priority / set default values for CaseID==null() ]
[ append indicator field values (make multi-value if one is set) ]
| appendpipe
    [| stats first(CaseID) first(Priority), first(CampaignID), first(CreatedTime),
first(ResolvedTime), values(IndicatorType1), values(IndicatorType2), first(Assignee),
first(Status), first(CaseName) count by Indicator
    | fields - Indicator
    | rename values(*) AS *, first(*) AS *
    | rename CaseID AS _key
    | outputlookup append=t cases                    # Write to the Cases lookup
    | search NOT * ]                                 # Blank _key = new case
| lookup cases_open $indicator_field$ AS Indicator OUTPUT _key AS CaseIDNew
| eval CaseID=coalesce(CaseID, CaseIDNew)           # Grab the case ID (if newly created)
| fields - Indicator CaseIDNew
| rename OrderID AS _key
| outputlookup append=t orders                      # Write to the Orders lookup
```

Core Views/Dashboards

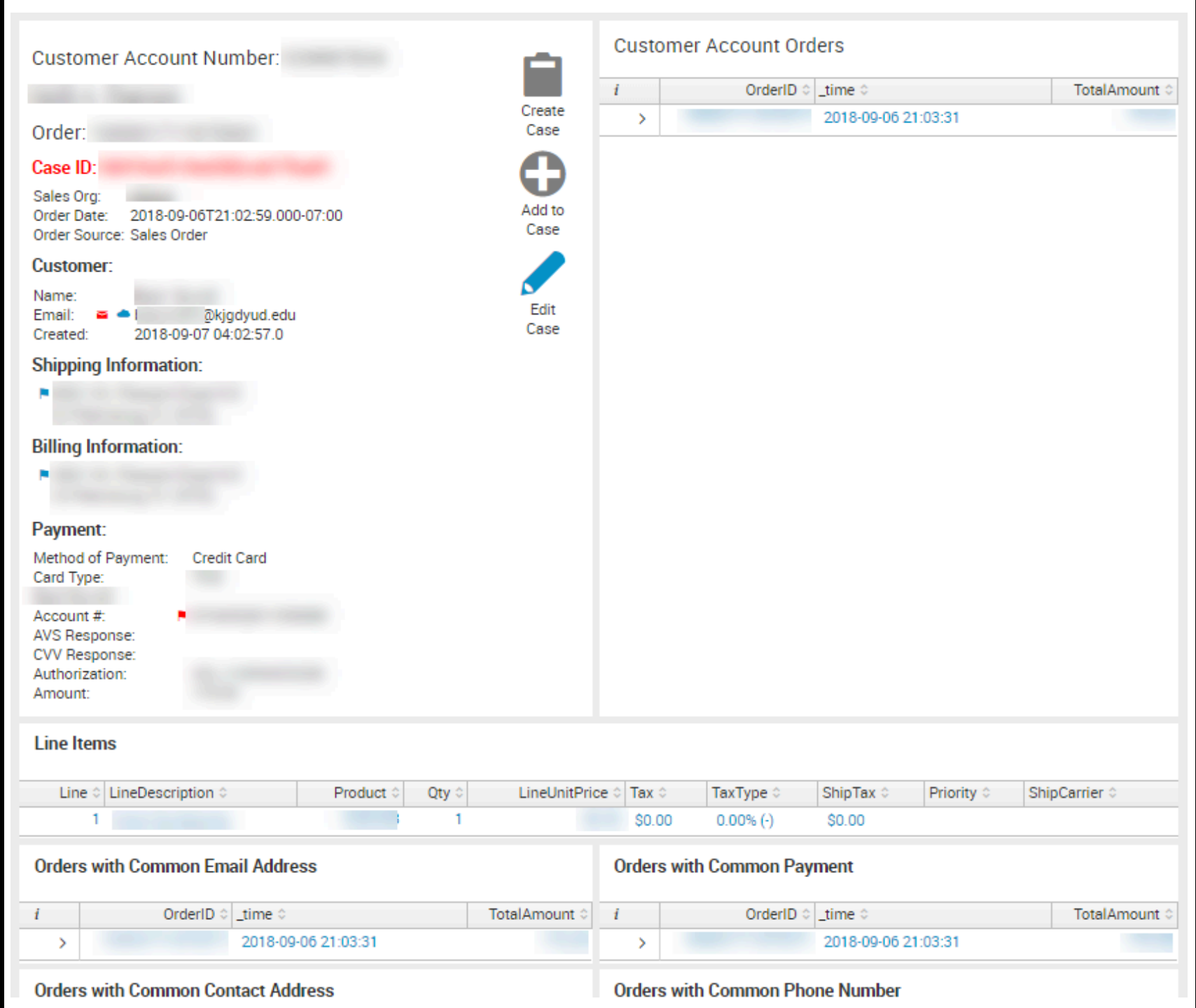
Bread and Butter

- Order Search
  - Filter by index/sourcetype/keywords
  - Includes Create Case button for matching orders
- Recursive (Multi-Level) Search
  - Digs for linked orders with up to 3 degrees of separation
- Velocity Search
  - Finds orders with a common field value X times over Y hours
- Order View
  - Custom HTML view of a single order
  - Case functions (add/edit/delete)
  - Related order searches
- Case View
  - Manage cases, included orders, related attributes, and see graph relationships



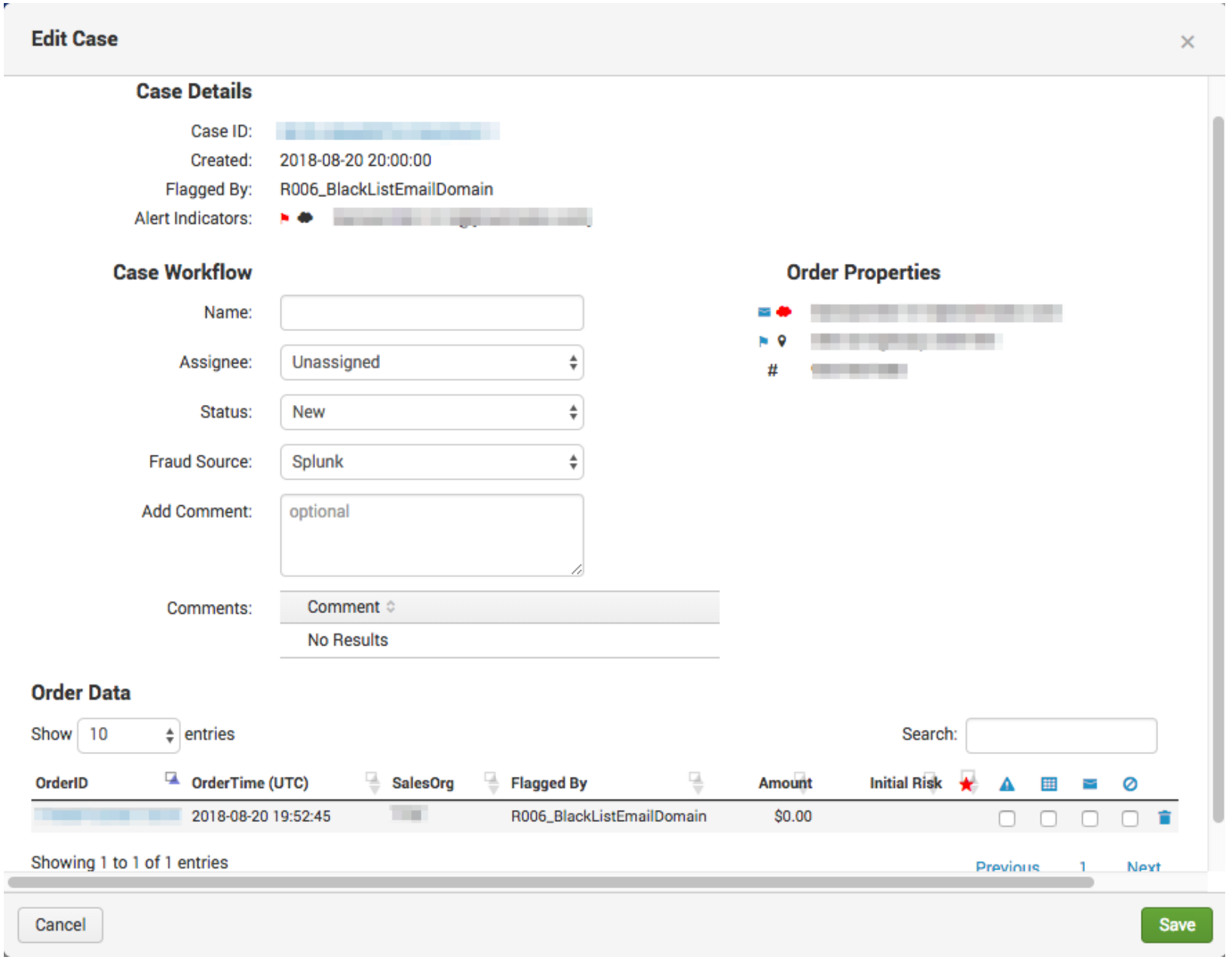
Order View

For when it's all about the details



Edit Cases

Set extended case attributes, comments, and order flags

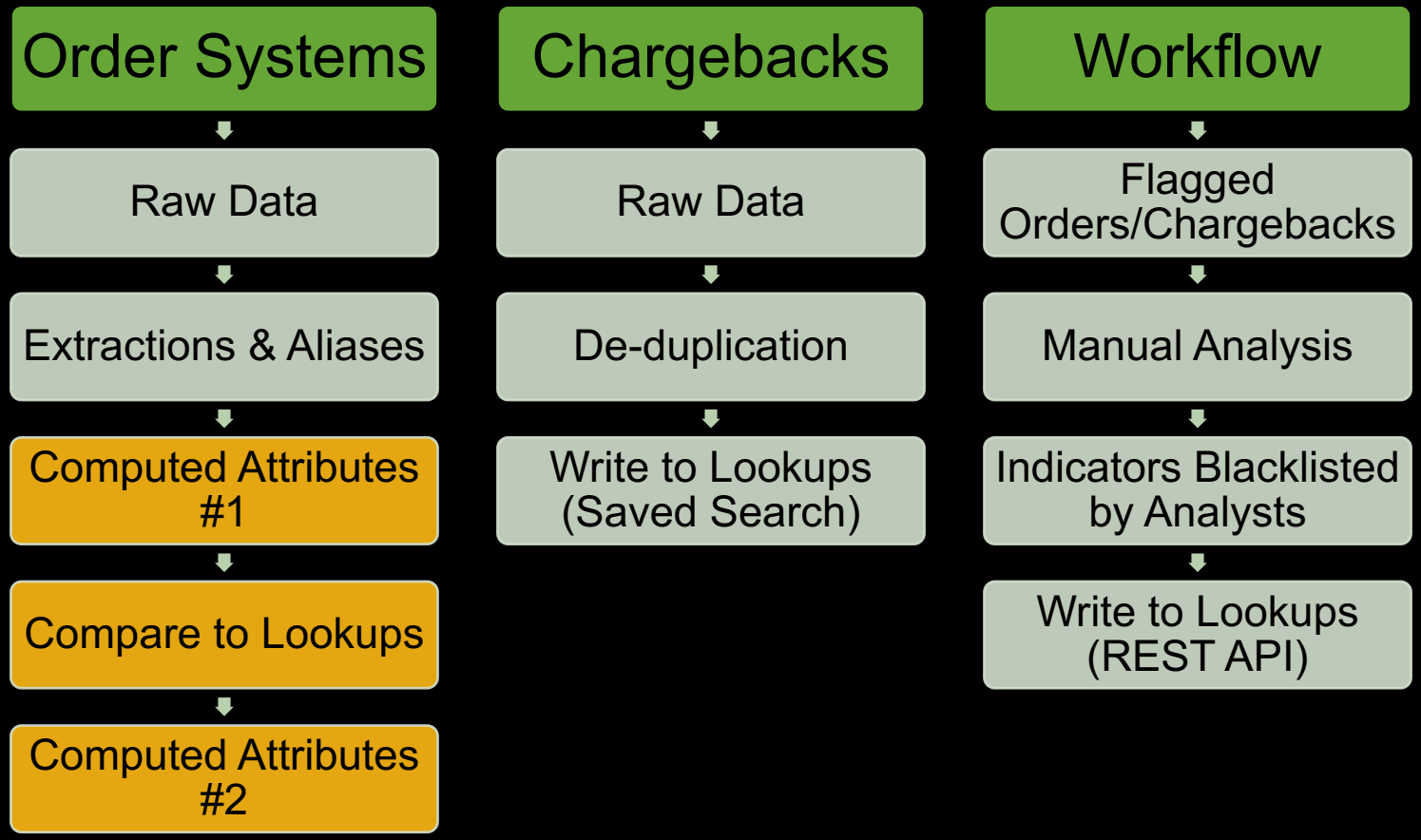


Search-Time Data Enrichment with Risk Scoring

splunk> .conf18

Search-Time Risk Attributes – Concept

Calculating Badness at Search-Time



Search-Time Risk Attributes – Examples

What is bad? Start with suspicious.

Intelligence

- Blacklist Attributes
- Attributes Matching Chargeback Orders
- DNS Domain Validation

Email

- Contains Contact Name
- Abnormal # of Digits in Address
- Foreign Email Domain TLDs
- Free Email Domains
- Invalid Email Domains
- Email Subdomains

Suspicious

- New Customer & Large Order
- Multiple Payment Cards per Customer
- Other/Secret Sauce

Shipping

- Billing State != Shipping State
- Rush Shipment
- Shipping Distance > # miles

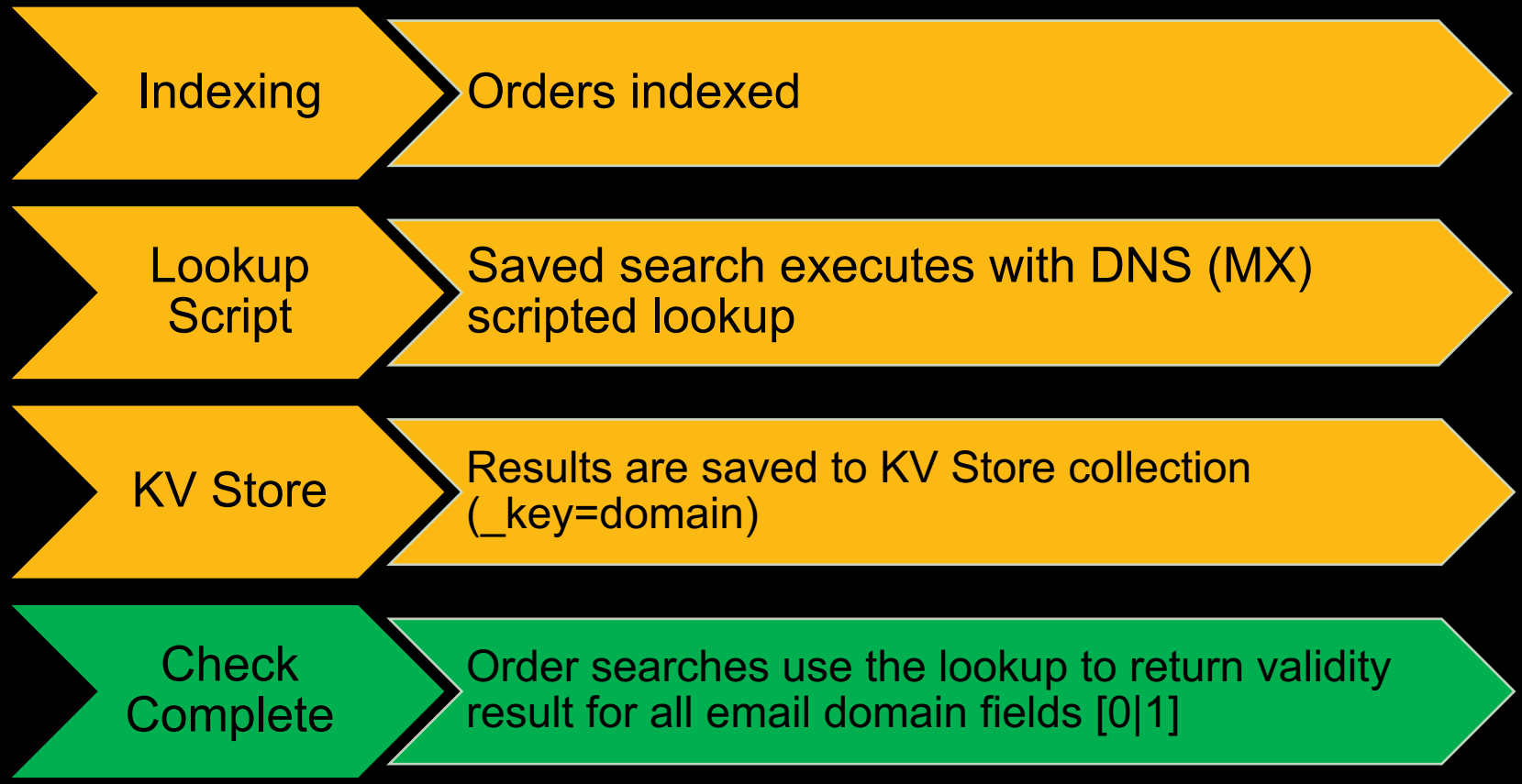
Search-Time Risk Attributes – Examples

Configuration Details

```
# Phone/Payment Method/Address/etc.
LOOKUP-[attribute] = af_[attribute] _key AS [attribute] OUTPUT BL_[attribute] FA_CB_[attribute]

# Multi-level (4) email domains / foreign domain
EVAL-FA_ContactEmailDomain_4Levels = if(match(ContactEmail, "@[^.]+\.[^.]+\.[^.]+\.[^.]+$"), 1, 0)
LOOKUP-email_foreign_tld = foreign_domain_suffix DomainSuffix AS ContactEmailTLD OUTPUT Foreign AS
    FA_Email_Foreign_TLD

# Email Address does not include the Contact Name
EVAL-FA_Email_NotIncludes_Name = if(like(lower(Email), "%"+lower(FirstName)+"%") OR like(lower(Email),
    "%"+lower(LastName)+"%"), 0, 1)

# Billing and shipment states are different
EVAL-FA_Bill_Ship_State_Diff = if(upper(PrimaryAddressState)!=upper(ShipAddressState) AND
    isnotnull(ShipAddressState) AND isnotnull(PrimaryAddressState), 1, 0)

# Rush shipment (not ground)
EVAL-FA_Rush_Shipping = if(like(ShipCarrier, "%Day%") OR like(ShipCarrier, "%Overnight%"), 1, 0)

# High dollar amount on the order
EVAL-FA_HDA1 = if(TotalAmount>#####, 1,0)
```

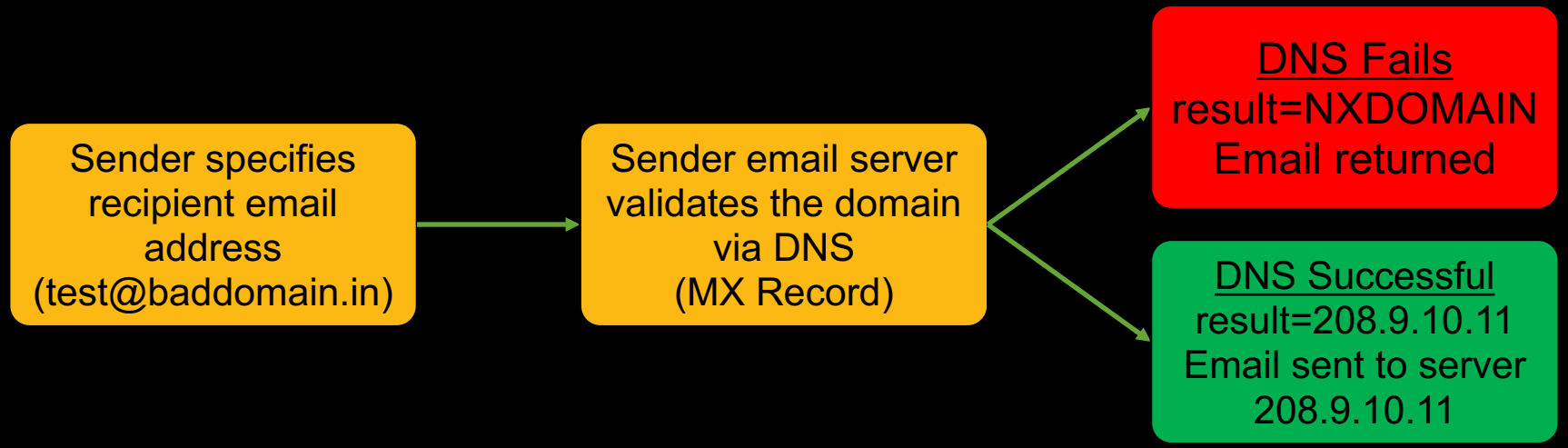
Email Domain Validation – Concept

Validation Process



How Email Works

Email Validation Process Review (Simplified)



- Scammers use bad email addresses to avoid being traced.
- We prove the bad email domains are invalid and store the result.
- We can reference all observed domains to validate orders.
- Only 1.5% of orders with invalid email domains result in a chargeback.

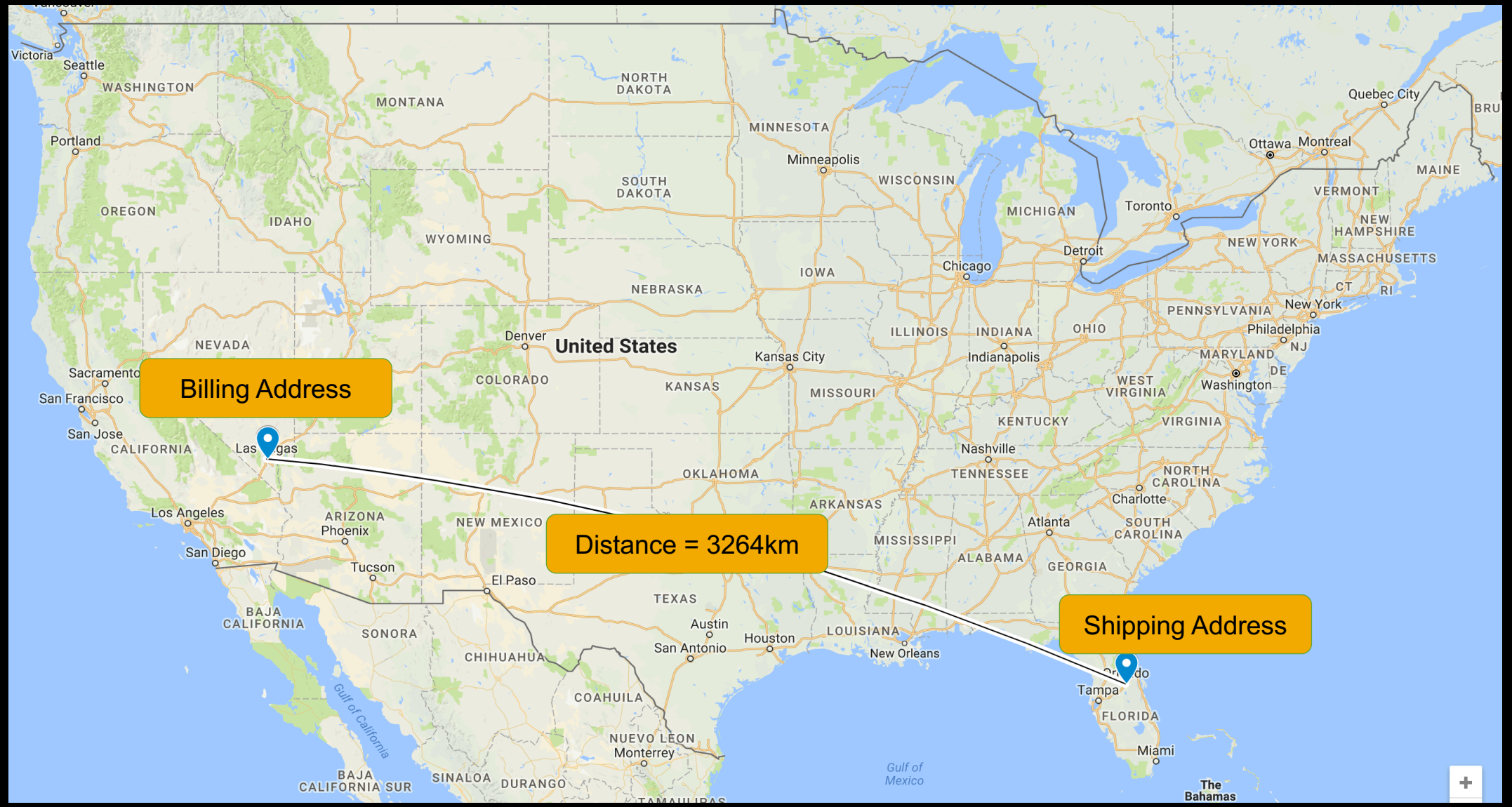
Email Domain Validation – Example

This Isn't the Email You're Looking For

```
LOOKUP-contact_email_domain = af_email_domain _key AS ContactEmailDomain OUTPUT
    BL_EmailDomain AS BL_ContactEmailDomain          # Blacklist flag
    FA_CB_EmailDomain AS FA_CB_ContactEmailDomain    # Previous chargeback
    WL_EmailDomain AS WL_ContactEmailDomain          # Whitelist flag
    FA_Email_Error AS FA_ContactEmail_Error          # DNS MX Error
    FA_Email_Invalid AS FA_ContactEmail_Invalid      # DNS MX Invalid
```

FA_ContactEmail_Error	1	
FA_ContactEmail_Invalid	1	

Geographic Distance – Billing vs. Shipping



Geographic Distance – Billing vs. Shipping

Configuration Details

props.conf:

```
[order]
# Normalize the postal codes to 5-digit numbers
EVAL-PrimaryAddressZip = replace(PrimaryAddressPostalCode, "^(\d{5}).*", "\1")
EVAL-ShipAddressZip = replace(ShipAddressPostalCode, "^(\d{5}).*", "\1")

# Perform a lookup against the zipcode table for latitutde/longitude
LOOKUP-primary_zip_geo = zipcode_geo Zipcode AS PrimaryAddressZip OUTPUTNEW Lat AS
PrimaryAddressZipLat Long AS PrimaryAddressZipLong
LOOKUP-ship_zip_geo = zipcode_geo Zipcode AS ShipAddressZip OUTPUTNEW Lat AS ShipAddressZipLat
Long AS ShipAddressZipLong
```

Search:

```
index=orders sourcetype=order
| `globedistance(AddressDistance, PrimaryAddressZipLat, PrimaryAddressZipLong,
ShipAddressZipLat, ShipAddressZipLong, "m", "0")`
| eval FA_ShipAddress_Far=if(AddressDistance>####, 1, 0)
```

Search-Time Risk Scoring

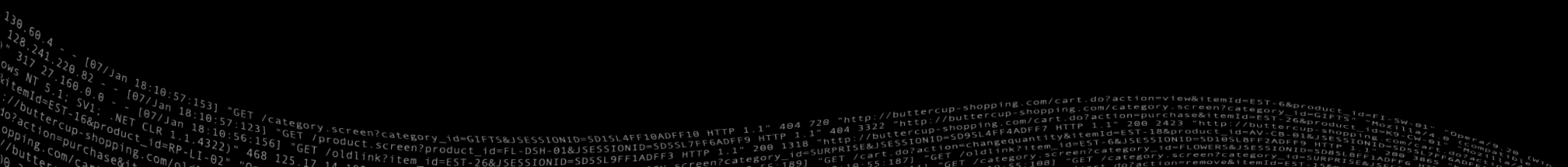
You can too, in 5 easy steps!

Create Attributes → Create Risk Scores Lookup → Import Scores into Search → Multiply Attributes (0|1) x Scores → Total All Scores

Create all Risk Attributes (calculated fields and/or lookups) to produce a Boolean value for each (0 or 1).

props.conf:

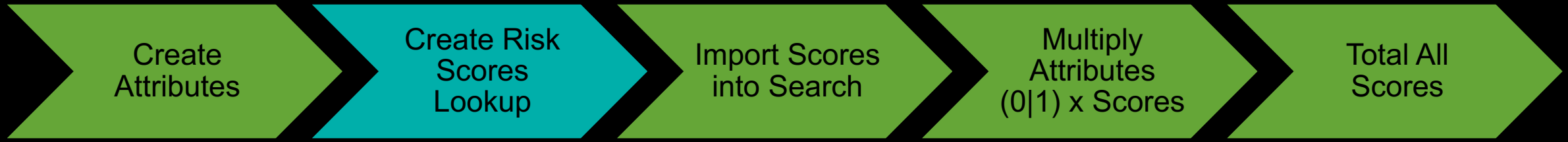
```
# Email Address does not include the Contact Name
EVAL-FA_Email_NotIncludes_Name = if(like(lower(Email), "%"+lower(FirstName)+"%") OR
    like(lower(Email), "%"+lower(LastName)+"%"), 0, 1)

# Credit card is blacklisted
LOOKUP-cc = cc _key AS CCNumber OUTPUT FA_BL_cc FA_CB_cc
```



splunk> .conf18

Search-Time Risk Scoring

You can too, in 5 easy steps!



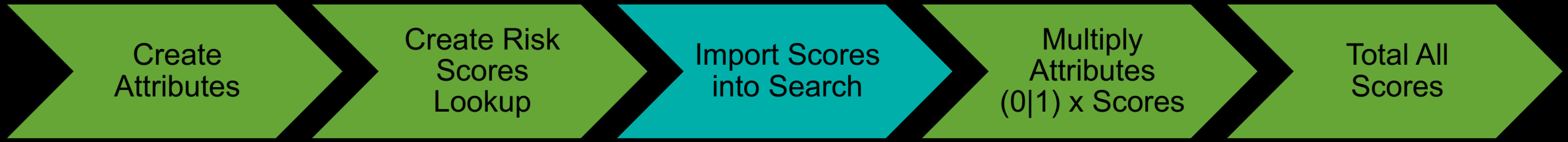
Each app has a different risk profile (per-attribute) for chargebacks. These scores are stored in a lookup.

Run a search to compute the probability for each attribute that it results in a chargeback. Split by application or business unit.

- **((#CB/#Non-CB)-1)*100**
- **-100 score = no (or negative) correlation. Normalize to 0**
- **Normalize to 0-100 range by adding the totals for each row and dividing each field by that total**
- **Output the results to a lookup**

Search-Time Risk Scoring

You can too, in 5 easy steps!



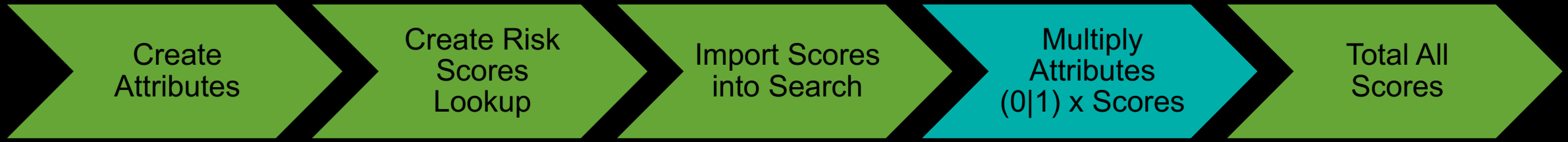
Run a search and reference the Risk Scores Lookup values for each attribute based on that business unit or application.

```
| lookup scores app AS app OUTPUT Score_FA_Email_NotIncludes_Name
Score_FA_CB_cc …
```

splunk> .conf18

Search-Time Risk Scoring

You can too, in 5 easy steps!

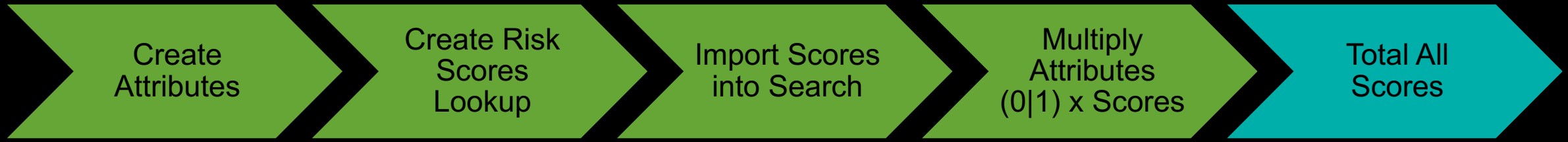


For each risk attribute, multiply the value of the attribute (0 or 1) by the Risk Score Lookup value.

```
| foreach Score_* [eval <<FIELD>>=FA_<<MATCHSTR>>*<<FIELD>>]
```

Search-Time Risk Scoring

You can too, in 5 easy steps!



Add all of the resulting values together for the Order Risk Score.

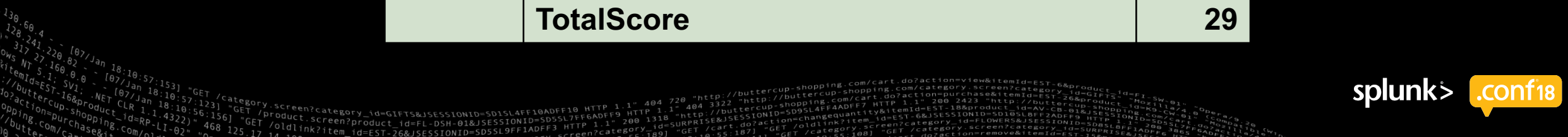
```
| eval TotalScore=0
| foreach Score_* [eval TotalScore=TotalScore+<<FIELD>>]
```

Search-Time Risk Scoring

Example Results

Order ID	App	FA_Example1	FA_Example2	Score_FA_Example1	Score_FA_Example2	Risk_Score
ABC1	ABC	0	1	20	30	30
ABC2	ABC	1	1	20	30	50
XYZ3	XYZ	1	0	5	60	5

Order ID	Field	Value
ABC1	Score_FA_Bill_Ship_State_Diff	10
	Score_FA_CB_cc	15
	Score_FA_ContactEmailDomain_4Levels	0
	Score_FA_Email_NotIncludes_Name	4
	Score_FA_Rush_Shipping	0
	Score_FA_ShipAddress_Far	0
	TotalScore	**29**



splunk> .conf18

Advanced Searches & Visualizations

splunk> .conf18

Recursive (Multi-Level) Search – Concept

How deep does the rabbit hole go?

Order IDs 1000-1002	
Customer ID	12345
Order Email	triples@graph.net
Payment Method	CC# XXXXXX…1554
Billing Address	22 Courtless Ln|99991
Shipping Address	5 Curve Cir|80808

Order ID 301	
Customer ID	12001
Order Email	cool@theschool.edu
Payment Method	CC# XXXXXX…9999
Billing Address	9 Bendy Way|19809
Shipping Address	5 Curve Cir|80808

Order ID 542	
Customer ID	8969
Order Email	cool@theschool.edu
Payment Method	CC# XXXXXX…3951
Billing Address	5720 Round Dr|80521
Shipping Address	6 Drive Pkwy|80808

Order IDs 806-807	
Customer ID	9111
Order Email	third@fourth.net
Payment Method	CC# XXXXXX…3951
Billing Address	5720 Round Dr|80521
Shipping Address	N/A

Recursive (Multi-Level) Search – Example

Configuration Details

Search:

```
| tstats summariesonly=t values(Orders.CCNumber) AS Orders.CCNumber … values([ all order fields]) AS
Orders.[fieldX] …
    from datamodel=Orders where $index$ Orders.SalesOrganization=$sales_org$
    [| tstats summariesonly=t values(Orders.CCNumber) AS Orders.CCNumber … values([ key fields only]) AS
Orders.[fieldX] from datamodel=Orders where $index$ Orders.SalesOrganization=$sales_org$
        [| tstats summariesonly=t values(Orders.CCNumber) AS Orders.CCNumber … values([ key fields only])
AS Orders.[fieldX] from datamodel=Orders where $index$ Orders.SalesOrganization=$sales_org$
            [ search $index$ eventtype=order SalesOrganization=$sales_org$ $base_search$
            | stats values(CCNumber) AS CCNumber  … values([ key fields only]) AS Orders.[fieldX]
            | fields CCNumber field1 field2 field3 field4
            | rename * AS Orders.*
            | format "(" "" "OR" "" "OR" ")"
            | fields search]
        | format "" "" "OR" "" "OR" ""]
    | format "" "" "OR" "" "OR" ""] by _time span=1s Orders.OrderID Orders.SalesOrganization
| `remove_dm_name(Orders)`
```

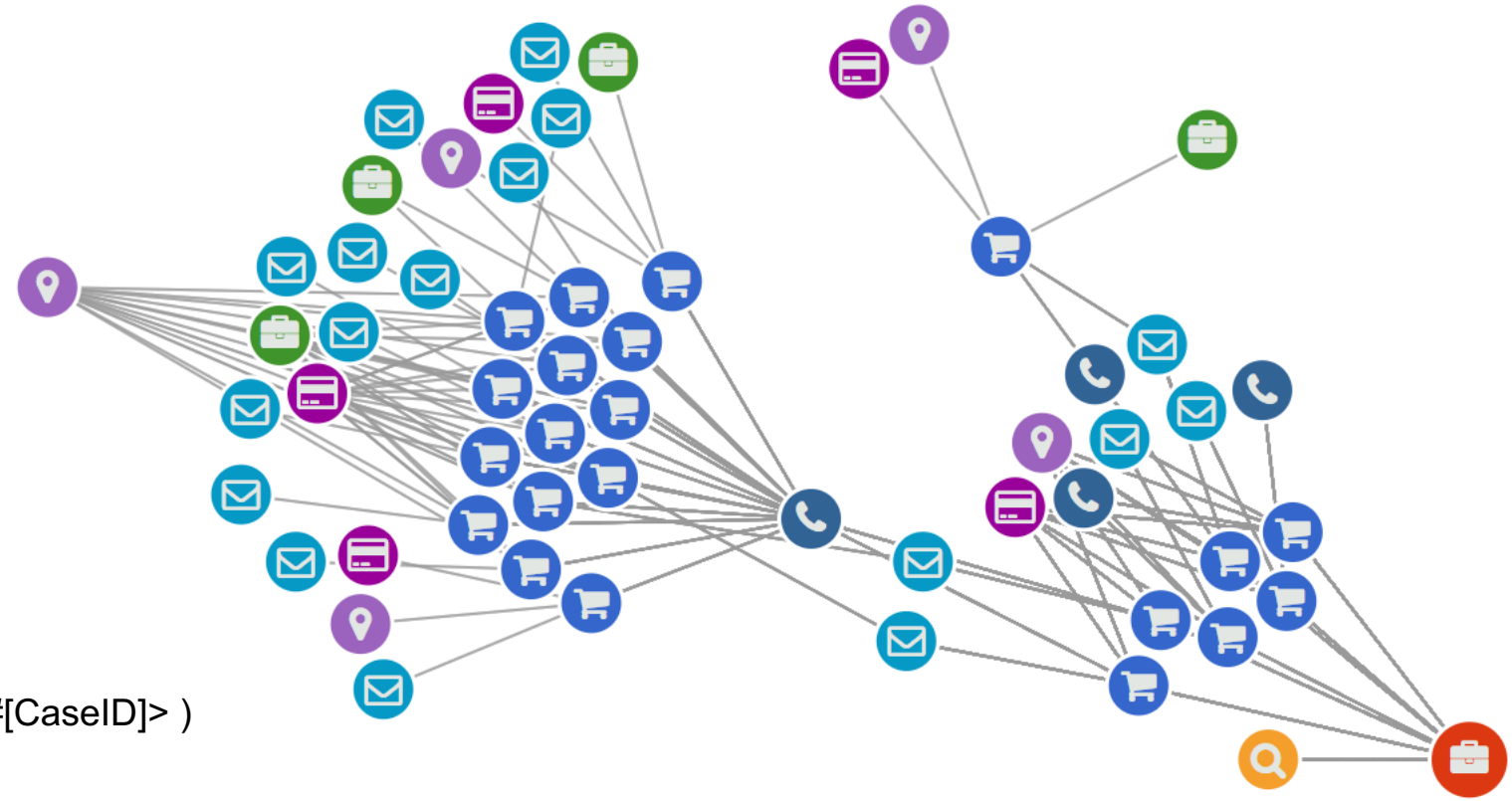
Level 1 Level 2 Level 3 Level 4 (Root Search)

Recursive (Multi-Level) Search – Graph Version

Splunk doesn't speak Graph (natively)

Search (Custom to RDF):

```
| triples query="SELECT * WHERE {
?s rdfs:label ?sLabel .
?s rdf:type ?sType .
?s ?p ?o .
?o rdfs:label ?oLabel .
?o rdf:type ?oType .
?o ?p2 ?o2 .
?o2 rdfs:label ?o2Label .
?o2 rdf:type ?o2Type .
?o2 ?p3 ?o3 .
?o3 rdfs:label ?o3Label .
?o3 rdf:type ?o3Type .
?o3 ?p4 ?o4 .
?o4 rdfs:label ?o4Label .
?o4 rdf:type ?o4Type .
FILTER (?s = <http://www.customer.com/ns/case#[CaseID]> )
}"
```



Predictive Analysis via Machine Learning

Using the Splunk Machine Learning Toolkit to Predict Chargeback Outcome

Prediction Results

BadOrder	predicted(BadOrder)	fs_SalesOrganization=	fs_SalesOrganization=	fs_SalesOrganization=	fs_SalesOrganization=	fs_SS_TotalAmount	fs_SalesOrganization=
Y	N	0.0	0.0	1.0	0.0	1.26011654042	
Y	N	0.0	0.0	1.0	0.0	0.99708302206	
Y	Y	0.0	0.0	1.0	0.0	3.73426686204	
Y	N	0.0	0.0	1.0	0.0	3.10915576648	
Y	N	0.0	0.0	1.0	0.0	1.09055236839	
Y	N	0.0	0.0	1.0	0.0	2.17186077063	
Y	N	0.0	0.0	1.0	0.0	0.297642426425	
Y	N	0.0	0.0	1.0	0.0	2.97554849011	
Y	N	0.0	0.0	1.0	0.0	0.987606010801	
Y	N	0.0	0.0	1.0	0.0	0.477984375963	

« prev 1 2 3 4 5 6 7 8 9 10 next »

Open in Search | Show SPL

Precision	Recall	Accuracy	F1
1.00	1.00	1.00	1.00

Open in Search | Show SPL

Classification Results (Confusion Matrix)

Predicted actual	Predicted N	Predicted Y
N	1337931 (99.8%)	3033 (0.2%)
Y	370 (92%)	32 (8%)

Open in Search | Show SPL

* Work-in-progress: Not yet operationalized

Machine Learning

Starter Guidelines for Implementation

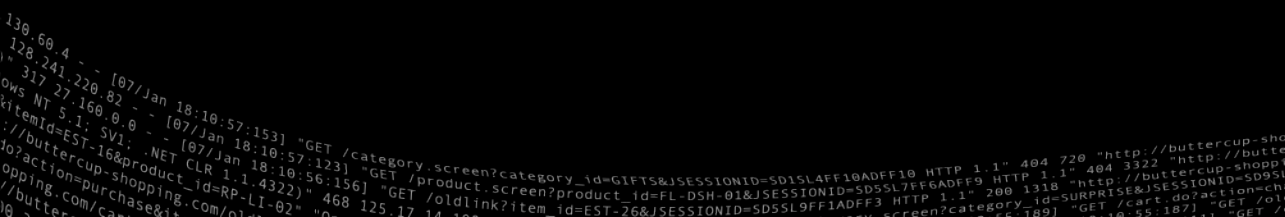
- Use Case = Predict Categorical Fields (i.e. Chargeback Y/N).
- Select or create Boolean or numeric metadata fields (like the risk attributes).
  - Starter apps for string field metadata ideas: URL Toolbox, DGA, Jellyfisher.
- Remove string fields with too many unique values (>150).
- Use fillnull for any fields that may have null values.
- Push numeric values through StandardScaler.
- Use FieldSelect to narrow down your fields.
- Start with tried & true, old-school LogisticRegression. Branch out from there to get more advanced.
- Increase limits in mlspl.conf.
  - Memory limits, categorical values, runtime, event count

Conclusion

Results

Did we make a dent?

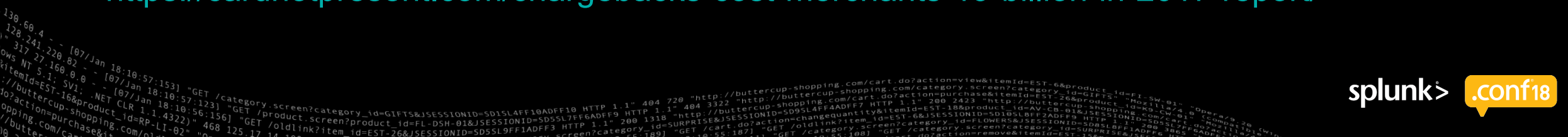
- Analyst time per incident was reduced by 50%.
  - More stopped shipments and payments
  - Reduced chargebacks by preventing charges to credit cards and banks
- Analyst team is able to identify/respond to new patterns within hours, instead of days or weeks
- Senior leadership reporting visibility for year-over-year performance is 100% improved
- Customer stakeholders have approved more investment in Splunk
- The work continues…



Additional Resources

- Gemini Data Web Site
  https://www.geminidata.com/
- Gemini Data on Twitter
  @geminidataco
- Gemini KV Store Tools App for Splunk
  https://splunkbase.splunk.com/app/3536/
- Gemini Data Quality App for Splunk
  https://splunkbase.splunk.com/app/3481/

- Chargebacks Cost Merchants $19 Billion in 2017: Report
  https://cardnotpresent.com/chargebacks-cost-merchants-19-billion-in-2017-report/



splunk> .conf18

Thank You

Don't forget to rate this session in the .conf18 mobile app