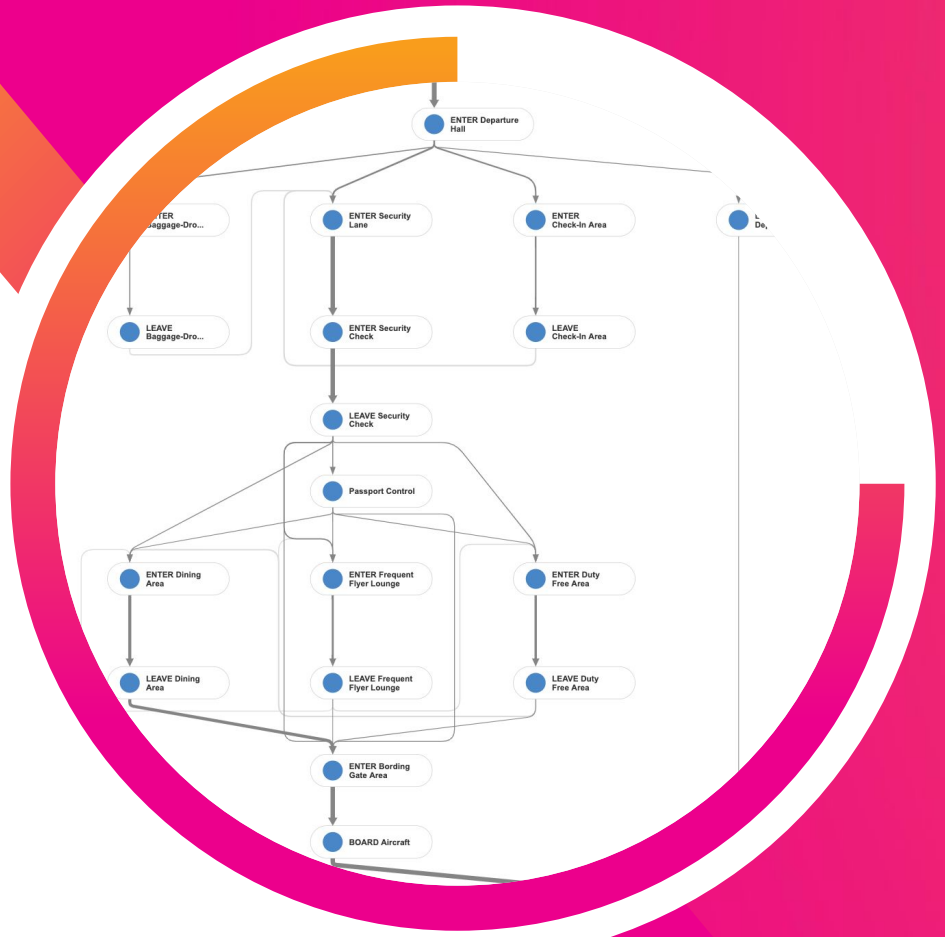
© 2019 SPLUNK INC.



BA1130

Demonstrating the Value of a Business Flow Use Case

Dirk Beerbohm
Senior Sales Engineer | Splunk Germany

splunk> .conf19

© 2019 SPLUNK INC.

Forward-Looking Statements

During the course of this presentation, we may make forward-looking statements regarding future events or plans 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 may differ materially. The forward-looking statements made in the this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, it may not contain current or accurate information. We do not assume any obligation to update any forward-looking statements made herein.

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 functionalities described or to include any such feature or functionality in a future release.

Splunk, Splunk>, Turn Data Into Doing, The Engine for Machine Data, Splunk Cloud, Splunk Light and SPL are trademarks and registered trademarks of Splunk Inc. in the United States and other countries. All other brand names, product names, or trademarks belong to their respective owners. © 2019 Splunk Inc. All rights reserved.

© 2019 SPLUNK INC.

Today's Agenda

Splunk Business Flow is a fast, flexible, and intuitive process mining solution for discovering and investigating issues within processes

In this session, I will share tips and tricks for Business Flow covering missing, wanted or future functionality

- Feedback from customers, presentations, workshops,etc.
- Do not link with promises, roadmaps or future functionality. Things may come or may not come.

© 2019 SPLUNK INC.

What is Splunk Business Flow?

splunk> .conf19

© 2019 SPLUNK INC.

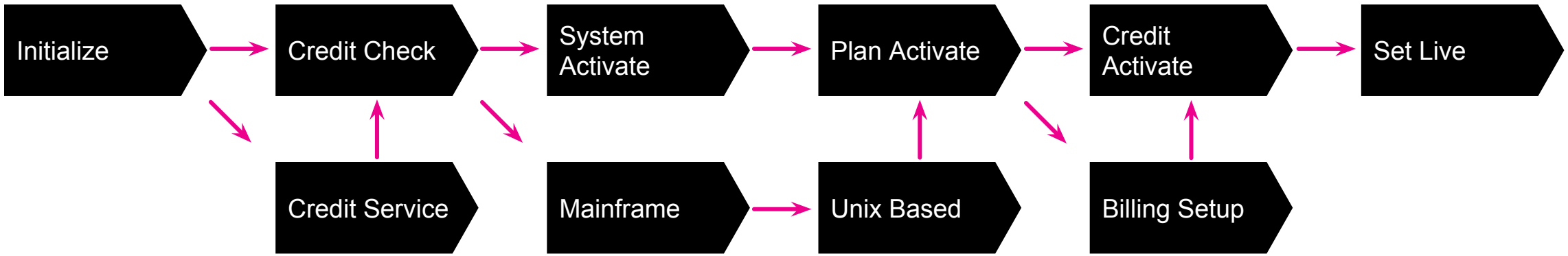
Process Mining Defined

Process Mining is a family of techniques in the field of process management that support the *analysis of business processes based on event logs.* During process mining, specialized data mining algorithms are applied to event log data in order *to identify trends, patterns and details* contained in event logs recorded by an information system.

Source = Wikipedia

© 2019 SPLUNK INC.

Example End-To-End Process: Telco SIM Card Activation



© 2019 SPLUNK INC.

Splunk Business Flow

End-to-end process discovery through event stitching

Investigate drill-down with exploration interface

Side-by-side A/B comparison of process flows

Conformance checking and deviation notifications

Splunk Business Flow

splunk>enterprise splunk>cloud

© 2019 SPLUNK INC.

Splunk's Event-Based Approach is Optimized for Process Mining

Legacy	Splunk
Schema at Write	Schema at Read
SQL	Search
ETL	No ETL Delays - Universal Indexing

Not only do legacy data architectures lack the speed for fast-paced business, but it also lacks the flexibility for agile, continuous testing of changes

Splunk's Event-Based Approach Eliminates Data Integration and Analytics Complexity

© 2019 SPLUNK INC.

Unlike legacy approaches, Splunk's architecture only requires time-stamped events to construct a model of the process

- Very easy to develop and modify process models
- Very easy to combine disparate sources of data
- All computation can be done at runtime to ensure freshness of insights

Minimal Required Fields In the Data:

- Timestamp
- Correlation Field(s)
- Step Name Field
- Attribute Field(s)

TIMESTAMP	CORRELATION IDS	STEP NAME	STEP VALUES	ATTRIBUTES
2018-11-10 11:15:00	OrderID CustomerID	ACTION	New Account Created Add To Cart Apply Coupon Submit Purchase Game	Code Country Product

© 2019 SPLUNK INC.

Splunk's Approach Enables Greater Time to Value, Agility, and Objectivity

Easy correlation of any data across silos **enables greater time to value** without relying on predefined process models

Easy iteration of process changes **enables greater agility & flexibility** without complexity and delays of data integration

Easy drill-down to raw event data **enables greater objectivity & transparency** without predefined references

© 2019 SPLUNK INC.

Three Tips & Tricks in Demonstrating Value With Business Flow

Prepare Your Data Intelligently!

Optimise your SPL!

Set Up Advanced Detection!

© 2019 SPLUNK INC.

Prepare Your Data Intelligently

splunk> .conf19

© 2019 SPLUNK INC.

Data Sampling

What happens if the dataset is too big?

- You need to minimize the number of events
- Standard Event-Sampling does not work <->
  It is not transaction aware
- Selecting different time periods may not help to solve you problem

What is “Too Big”?

- 1M events, 10M events, 100M events

The answer is to sample the data within the SPL base query using the Splunk MLTK

© 2019 SPLUNK INC.

Successful Data Sampling

Real World vs. Planned World

Splunk Business Flow is a premium application located on the SearchHead itself

As such, it can retrieve events from different Indexers, but can also be memory intensive

Stitching of events into transactions is done with a new “journey”

This event stitching may be executed as as Python script or natively, depending on the Splunk version used.

Specify a correct value for **Max Duration** when configuring the Event Sources

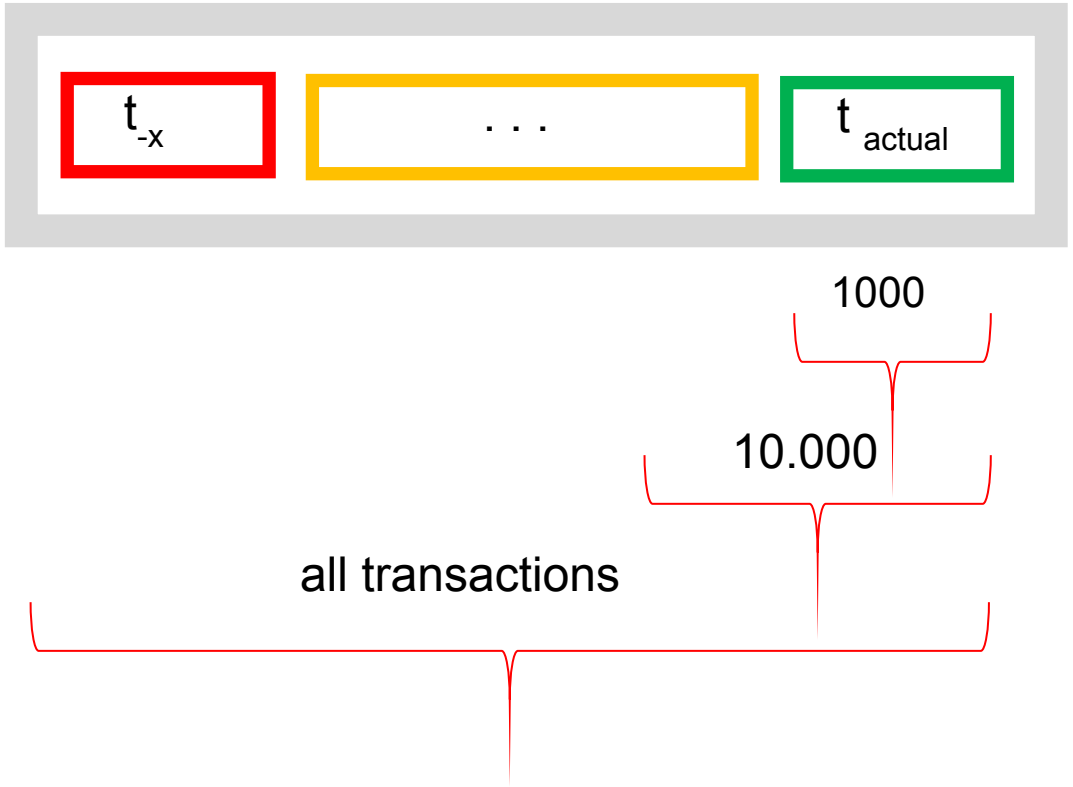
© 2019 SPLUNK INC.

Inside the Basics of Data Sampling

Splunk Business Flow allows to sample for first 1000, 10.000 or the entire dataset

However you may be interested in a custom strategy for partitioning – some reasons why:

- Use only a tiny portion of you entire time period.
- Provide insights for all times

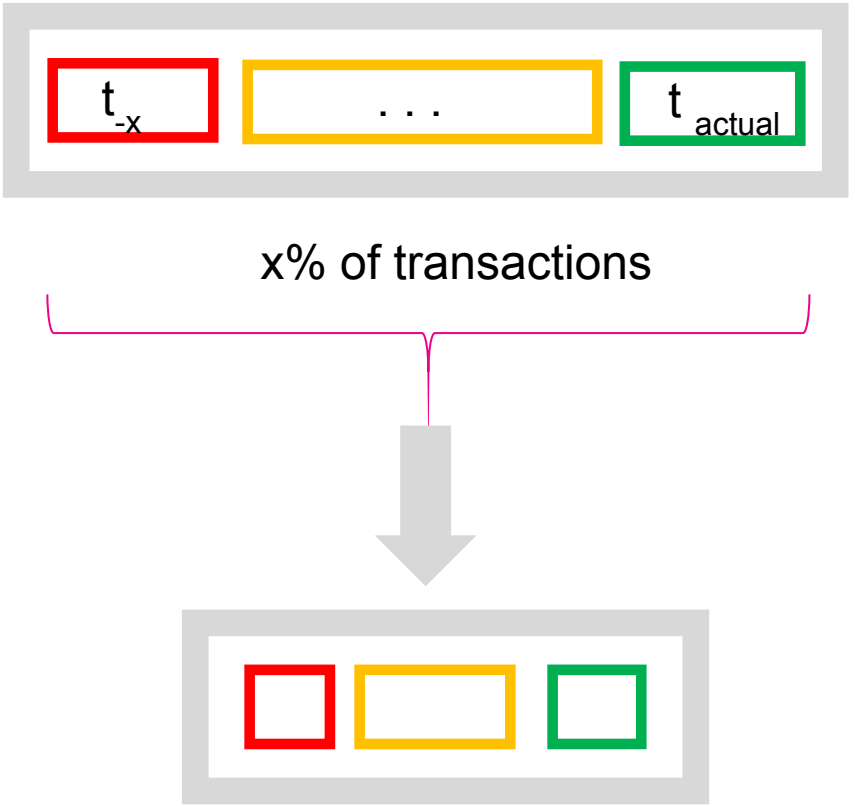


© 2019 SPLUNK INC.

Inside the Basics of Data Sampling

Suggested Method

- Sample the entire dataset during data preparation in SPL
- Reasonable size of data set for entire range of transactions (time)
- Results need to be cross checked / verified
  - Use “seed” functionality
  - Use several “seeds” and compare for plausibility
- Honors entire time range
  - Better insights in your overall process behavior



© 2019 SPLUNK INC.

Example Data Sampling

```
index=sbf_airport_sampling
    [| makeresults count=100
    | eval acc=1
    | accum acc
    | eval acc=acc-1
    | sample count=10 seed=4321
    | eval new_id="*".if(acc<10,"0".acc,acc)
    | table new_id
    ]
| sort 0 - _time
```

- **Works with numbers, only**
- **Example for a single Correlation ID**
- **This is a simple example and not production ready!**
- **Sampling individual per flow model**
- **Splunk MLTK required**

© 2019 SPLUNK INC.

Example Data Sampling

```
index=sbf_airport
| sort 0 + pax_id
| autoregress pax_id as pax_id_old
| eval change=0
| autoregress change as change_old
| eval change = if(pax_id=pax_id_old, change_old,change_old+1)
| streamstats count(eval(change==1)) as new_id
| table  _time new_id STEP LOCATION GATE
| sort 0 - _time
| collect index=sbf_airport_sampling
```

- **Create new arbitrary Correlation-ID's**
- **Example for a single Correlation ID**
- **Take the result as the input for the sampling**

© 2019 SPLUNK INC.

Some Things to Watch Out For

- The demo algorithm does not partition 100% precisely, for demo purposes is it sufficient ;-)
- Data set is fairly small (too small for real world example), prone to errors

- Complete data set: Avg. Journey Duration 23,5 days
- 50% of data set: Avg. Journey Duration 24,2 days
- 25% of data set: Avg. Journey Duration 21,4 days / 20,2 days (different seeds)

- Deviation in this example max. 15%
  - Results will become better for larger data sets
  - Do not rely on one simple partition or seed, only – be curious, verify with multiple different seeds

© 2019 SPLUNK INC.

Adding Numeric Values

Numeric values enhance the ability for filtering, but:

- Too many different values, e.g.
  - Product / Ticket prices in a retail shop
  - Durations of a desired action

Create discrete value ranges / bands of values / price groups

- | bin bins=10 price as PriceGroup
- bucket is an alias for bin

© 2019 SPLUNK INC.

Adding Numeric Values

Your SPL is NOT (only partial) THE SPL in Business Flow

```
index=sbf_retail*
| lookup lookup_products.csv product_id
| bin bins=4 price as price_group
| fields basket_id session_id shipping_id action country product_id stage slogan price_group, introduction_year
| sort 0 - _time
```

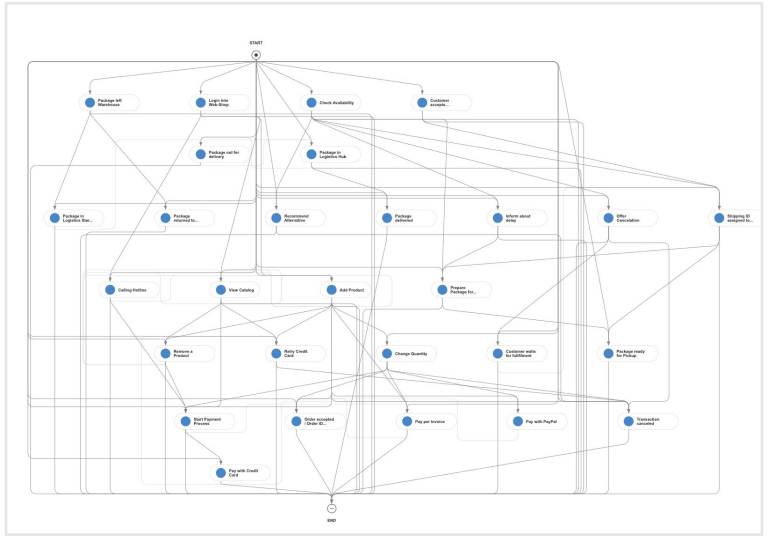
```
index=sbf_retail*
| lookup lookup_products.csv product_id
| bin bins=4 price as price_group
| fields basket_id session_id shipping_id action country product_id stage slogan price_group, introduction_year
| sort 0 - _time
| search (("action"="*" AND "action"!="") OR (basket_id="*" AND session_id="*") OR (basket_id="*" AND shipping_id="*") OR (session_id=
| eval _cja_step = 'action'
| eval _cja_stepTime = _time
| eval _cja_stepName = '_cja_step'
| head 100000000000
| jsontxn action="_cja_stepName" time_field="_cja_stepTime" maxspan="42d" correlation="basket_id" correlation="session_id" correlation
    attribute="price_group" attribute="introduction_year"
| spath output=_recompute_duration path=sequence{}._time
| eval total_duration = exact(tonumber(mvindex(_recompute_duration,-1)) - tonumber(mvindex(_recompute_duration,0)))
| eval "country" = 'country'
| eval "slogan" = 'slogan'
| eval "stage" = 'stage'
| eval "price_group" = 'price_group'
| eval "introduction_year" = 'introduction_year'
| fields - date_hour date_mday date_minute date_month date_second date_wday date_year date_zone eventtype host index linecount punct s
    timestartpos
| table *
```

© 2019 SPLUNK INC.

Adding Numeric Values

Your SPL is NOT (only partial) THE SPL in Business Flow

- The bin command destroys the sorting of the events -> It destroys everything…
- Add "| sort 0 - _time" as the last row in your SPL



© 2019 SPLUNK INC.

Assign Numerical Values to STEPS

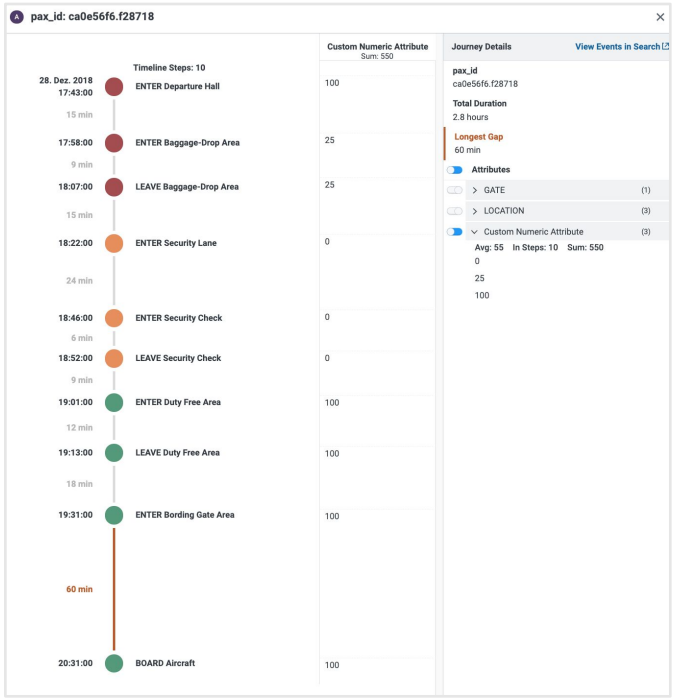
In Business Flow you can assign numbers to step within a saved “Flow”

- Treat it as a priority of steps, the higher the number the more important is this step for your “business”
- Business Flows uses statistical functions (avg, sum, count) on this individually assigned numbers
- Example the higher the average the closer you come to the “Happy Path”
- Example from the Airport Demo:
  - “Departure Hall -> Security -> Duty Free -> Boarding Gate” would be the ideal path
  - Dining Area and Frequent Flyer Lounge are seen as the “errors” in this process as they are not or only partial of interest for the airport operator with regards to additional revenue.
  - Security is of no interest, as all passenger have to pass it
  - Check-In and Baggage Drop are seen as less important.

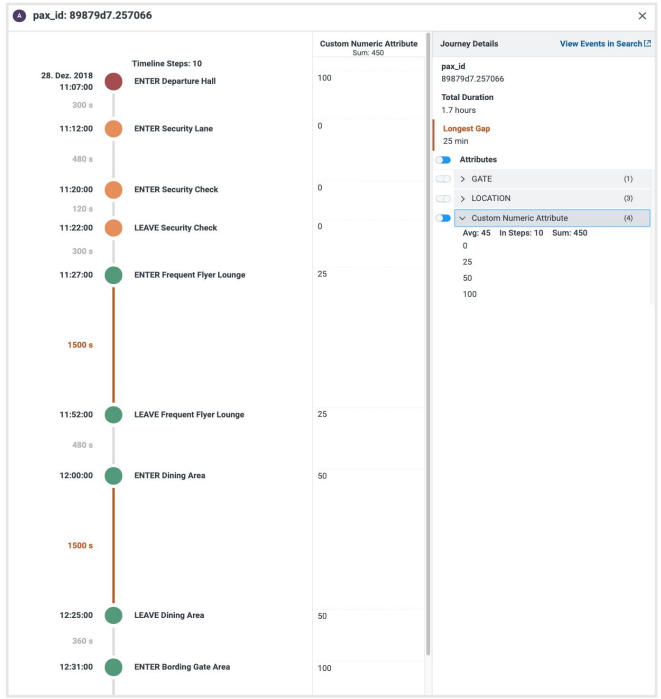
© 2019 SPLUNK INC.

Assign a Value to STEPS

Simulation of “Happy Path”



AVG: 55 Only Duty Free in Secure Area



AVG: 45 No Duty Free in Secure Area

splunk> .conf19

© 2019 SPLUNK INC.

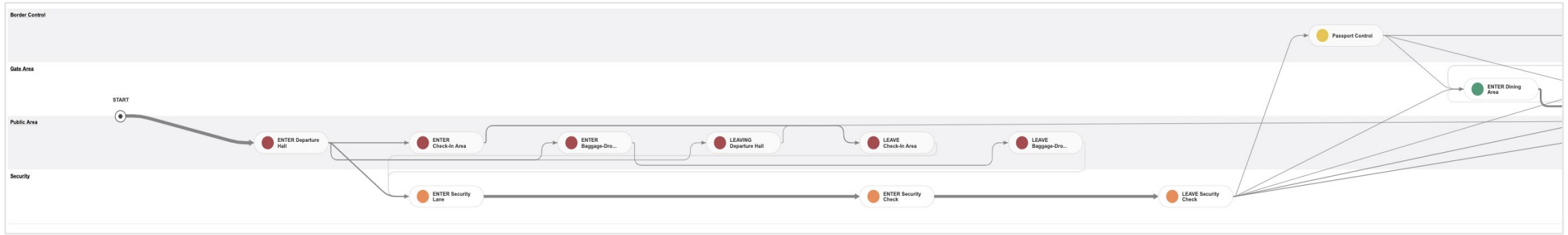
Assign Responsibilities to STEPS

Who controls which STEP

Within a ”Flow” assign categories / responsibilities to a STEP (“Lane”)

Shows clearly when, for example, a responsibility within a process changes

- Better than assign a value to an attribute for the purpose of visualization
- Use an attribute for filtering



© 2019 SPLUNK INC.

Optimising SPL Queries

splunk> .conf19

© 2019 SPLUNK INC.

SPL Query Optimization

Use Non-Streaming Commands as late as possible

- stats, sort, dedup, bin (streaming only with span argument)
- All (partial) results need to be returned before these commands can be run

Restrict the number of fields to be returned to the absolute minimum

- Only Correlation-ID‘s, STEP, Attributes used in flow model
- Faster transfer of result sets from Indexers to SearchHead
- Less memory consumption on SearchHead (Business Flow is a SH application)

© 2019 SPLUNK INC.

Set Up Advanced Detection With A Reference Model

splunk> .conf19

© 2019 SPLUNK INC.

Reference Models

Real World vs. Planned World

Business Flow does not offer the functionality of a reference model

- How to compare the exploration of the flow models against the intended behaviour?

Create a second flow model with the layout of the process flow you want to use as a reference

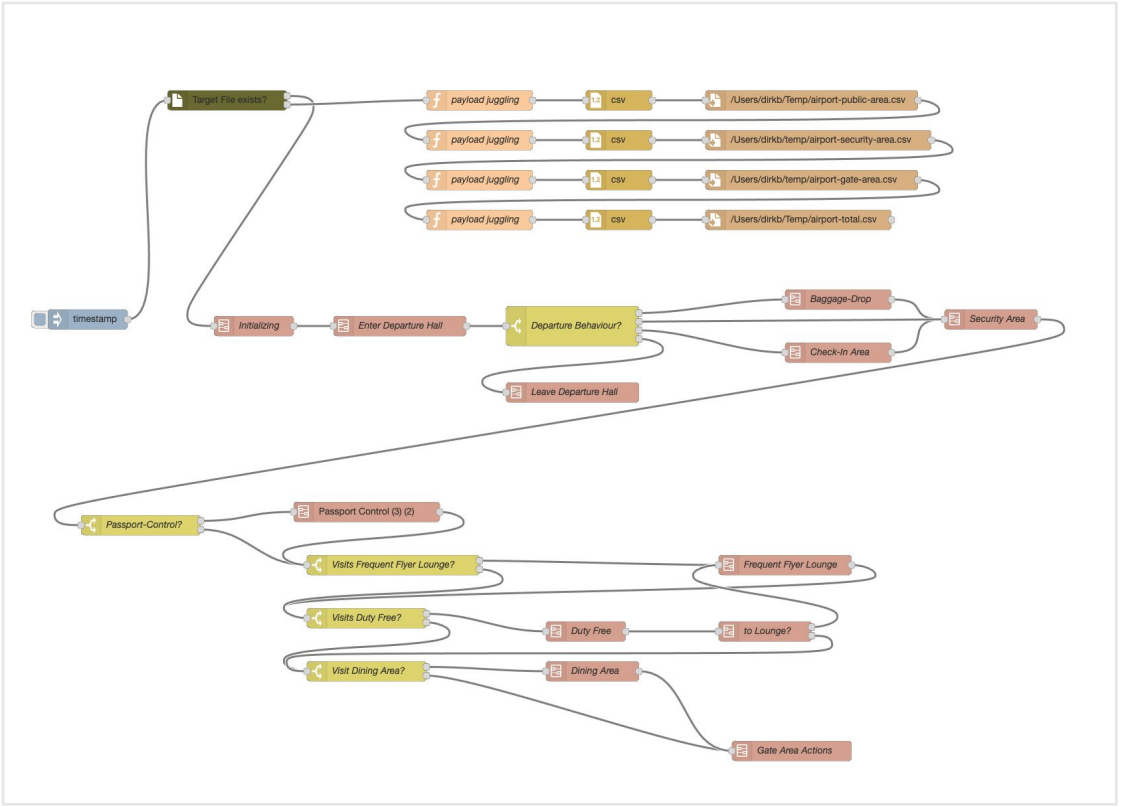
- Text-Edit
- Node-Red (Graphical layout of processes and generator for data sample sets)
- Outline the exact flow of the intended process flow.
- Distinguish reference model by index, sourcetype or special attribute
- Load both flow models in one SPL,e.g index=model **OR** index=reference_model
- Flag, if reference model: „| eval ref = if(index==reference_model,1,0)
- Hint: Save as a “Flow“ with reference attribute set to „0“!

© 2019 SPLUNK INC.

Reference Models

Example: Node-Red

- Node-Red gives you great flexibilty in designing exemplary processes, either reference models or sample data sets for mockup or testing
- It documents itsself
- Beware: Originally not designed for this task, but it does ist job.
- Used to build Busienss Flow demos on Splunk Oxygen Demo Server.



© 2019 SPLUNK INC.

Reference Models

Real World vs. Planned World

For „Mining“ create a Flow with reference attribute set to „0“

Use comparison „A/B“ and set reference attribute to „1“

This is not ideal, but it helps to make quick judgements and comparisons.

Use the tool of choice to create a reference flow

- One „journey“ per selected path

© 2019 SPLUNK INC.

Three Tips & Tricks In Demonstrating Value With Business Flow

Prepare Your Data Intelligently!

Optimise your SPL!

Set Up Advanced Detection!

© 2019 SPLUNK INC.

.conf19

splunk>

Thank You !

Go to the .conf19 mobile app to

RATE THIS SESSION