#splunkconf

.conf2013

YOUR DATA
NO LIMITS

Splunk as a BI Tool:
Business metrics to know everything about your product

Ekaterina Koptilina

Analyst, PLAY

#splunkconf

splunk>

About Me

- Turning machine data into business valuable information:
  - Data inputs preparing and integrating with Splunk Enterprise
  - Querying with spl language, SQL
  - Visualizing query results in dashboards
- Maintaining company's Splunk Enterprise infrastructure
- Exploring, utilizing Splunk apps and new features
- Teaching and consulting co-workers in Splunk Enteprise issues

About PLAY

- PLAY – leading video on demand service in Russia in premium segment
- www.playfamily.ru
- Presented on TV, WEB, MOB platforms though focused on smart TV platform

About PLAY

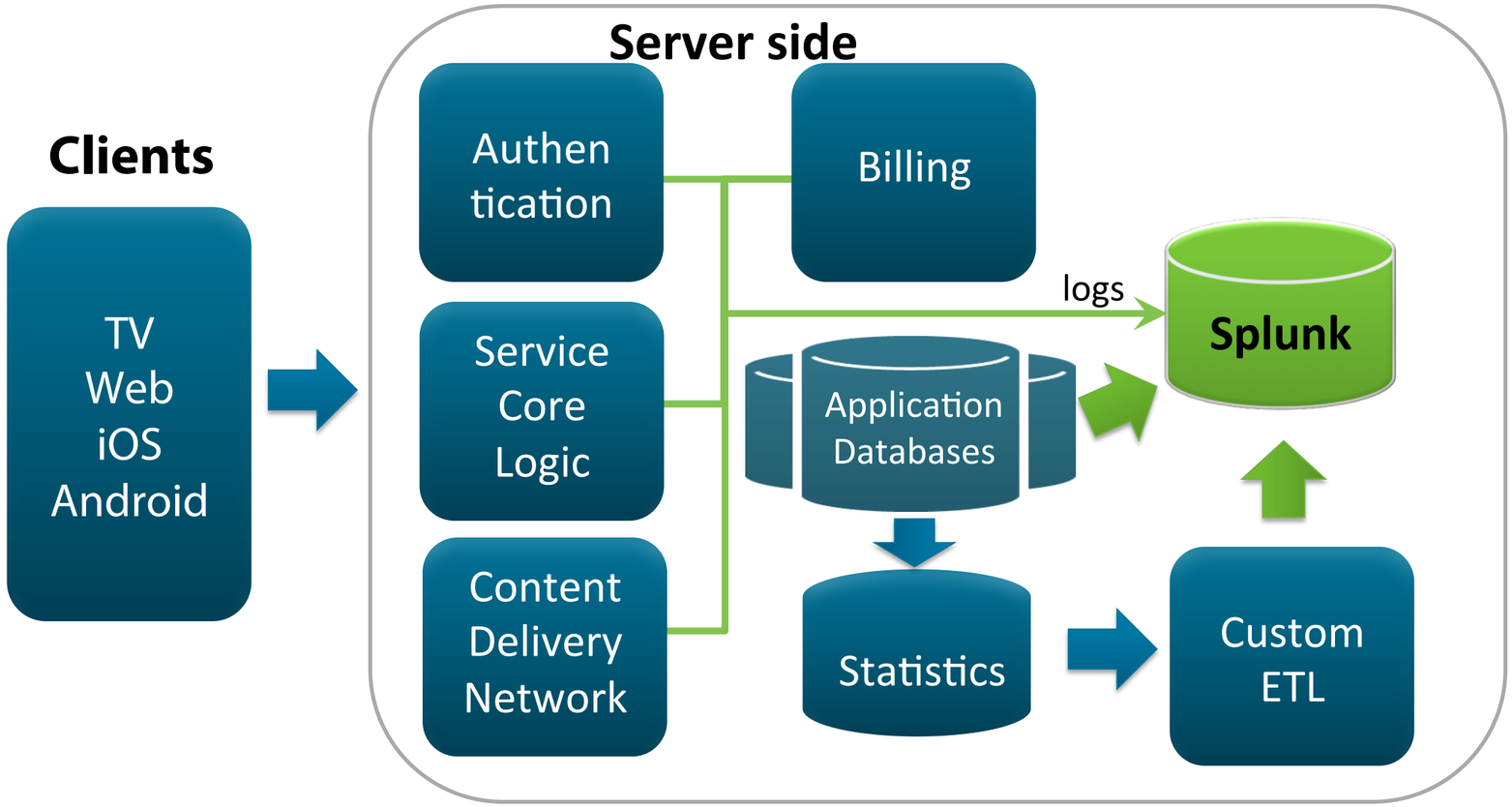
- Features: streaming, downloading, subscription, HD, 3D, no advertising, >4700 movies
- Partnership with >50 film studios including all major Hollywood ones
- Application is preinstalled on TV devices of main vendors

Agenda

- PLAY & Splunk
- What to measure?
- How should we estimate business success?
- Who are our customers?
- How do they use our service?
- What is interesting for them in our content?
- How to calculate and implement it all in Splunk?

PLAY & Splunk

- Log-files navigation
  - >70 servers
  - >20 applications
  - Up to 1Gb logs per day
- Business statistics



What to Measure?

- How should we estimate business success?
- Who are our customers?
- How do they use our service?
- What is interesting for them in our content?

- How to calculate it all?

Starting to Estimate Business Success

→ Define key user's action in your service:

Visit

Rate

Registration

Monetary purchase

Watch

Promo purchase

Bookmark

Basic Metrics to Start With

→ Elementary metrics to start with:

1. Revenue
2. Number of customers

3. ARPU= Revenue / Number of customers

ARPU (Average Revenue Per User)

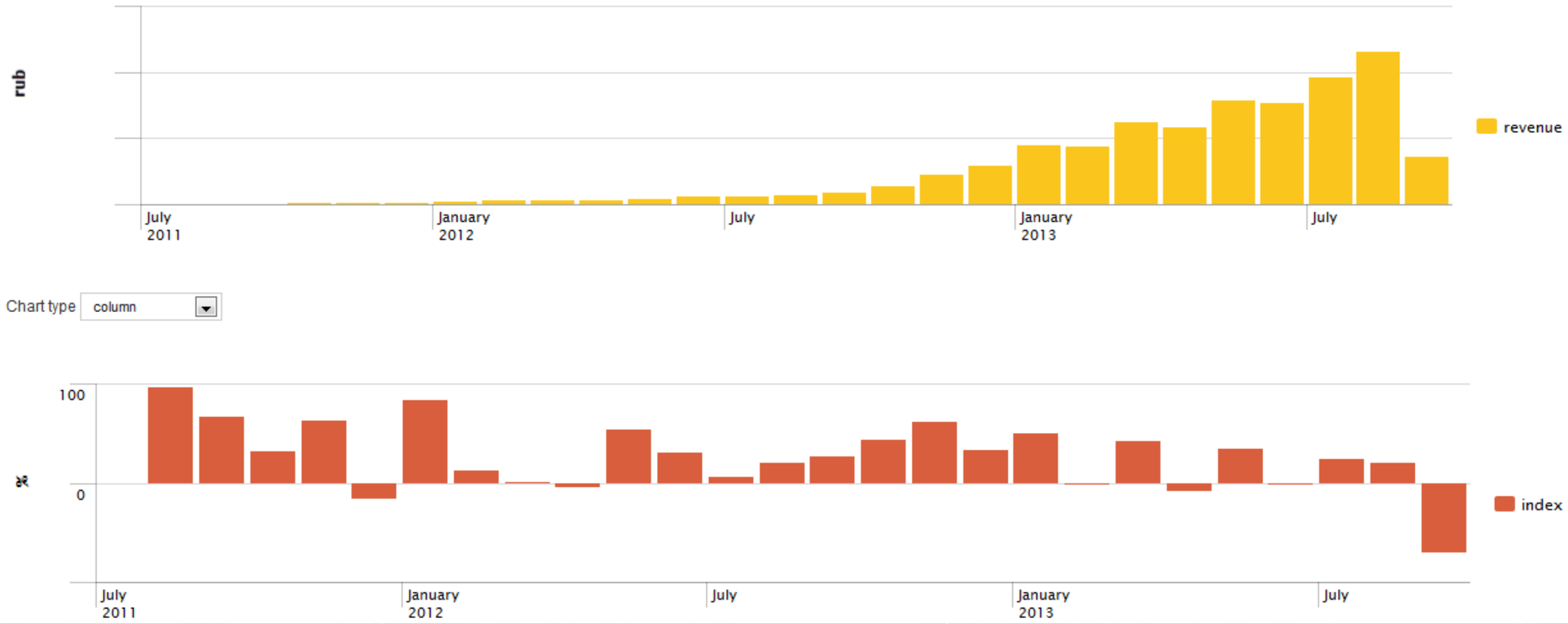
A measure of the revenue generated by one customer per time unit, typically per month or year

Index Metric

→ Index

Percent change of estimated value per time unit in comparison with value in previous time unit

```
Index = 100 * (currentValue - previousValue) / previousValue
```



Mean Revenues

→ Mean revenue per day

Mean revenue per day = Revenue for the period / Number of days

① Per weekend day

```
...
| where (like(date_wday, "saturday") OR like(date_wday, "sunday")
```

+ add implicitly official holidays of your country

② Per working day

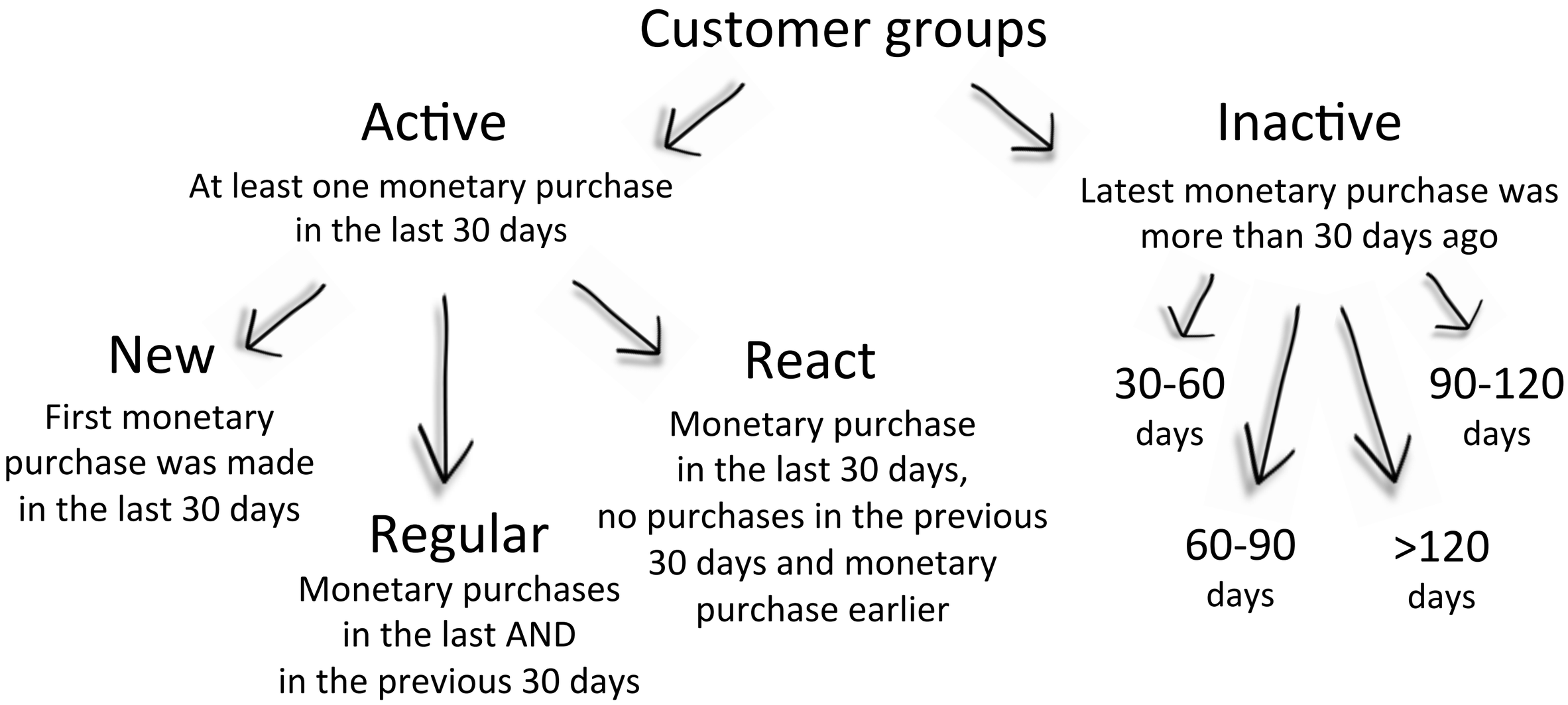
Exclude weekend days and official holidays

Who are Our Customers?

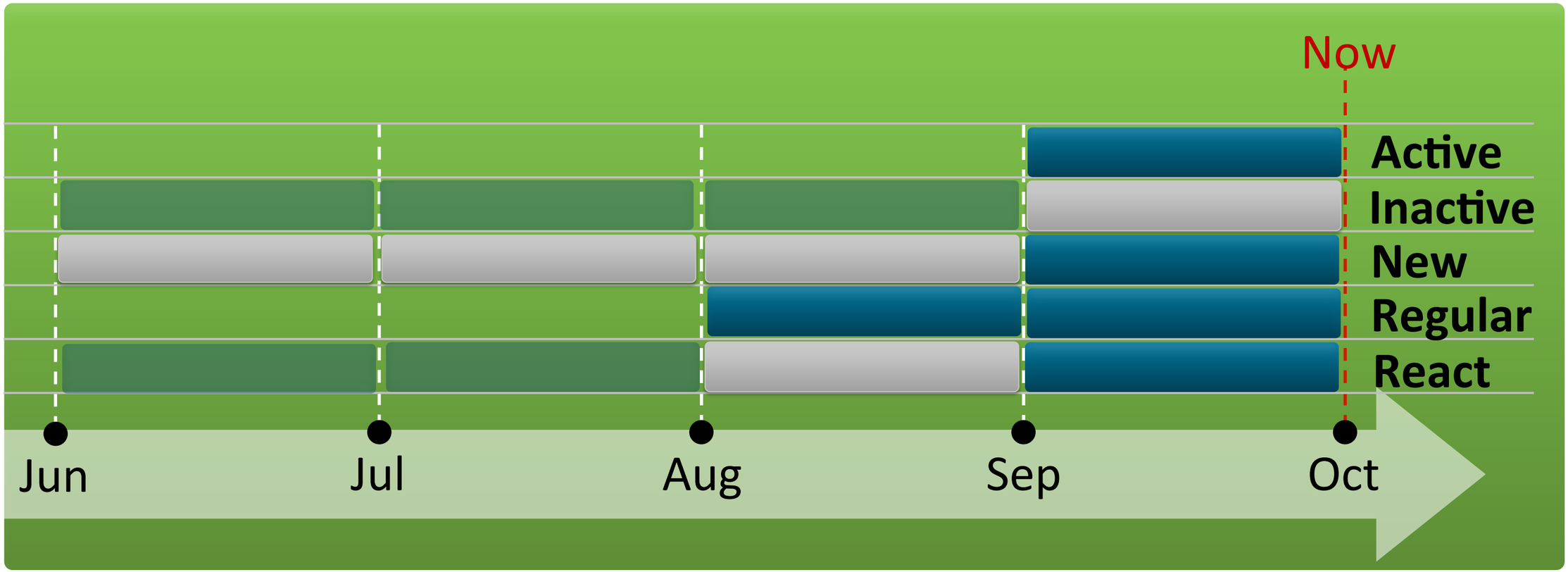
→ Users' segmentation

1. **Visitor**
   Visitor of the service without registration
2. **Prospect user**
   Registered user without any purchase
3. **Promo user**
   Registered user with only promo purchase
4. **Customer user**
   Registered user with monetary purchase (can have promo purchases as well)
5. **Buyer user**
   Registered user with any purchase (promo/monetary/both)

Customers' Segmentation



Customers' Segmentation on Timeline



Monetary purchase in each colored month

Monetary purchase at least in one of the months

No monetary purchases in any month

Active / Inactive Customers: How?

```
source="transactions_view" tx_amount>0
| addinfo
| eval limit_time=if(isnum(info_max_time), info_max_time, time())
| stats dc(eval(_time>= limit_time-2678400)) as active, dc(eval(_time<
limit_time-2678400)) as inactive by user_uid
| eval usage=case(active>0, "ACTIVE users", inactive>0 AND active=0, "INACTIVE
users")
| stats dc(user_uid) as count by usage
```

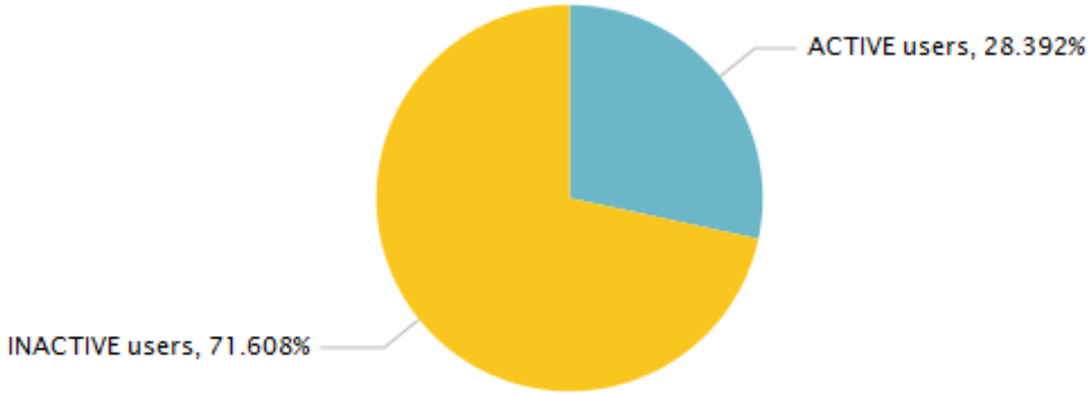
New / Regular / React Customers: How?

```
…
| addinfo
| eval limit_time=if(isnum(info_max_time), info_max_time, time())
| stats dc(eval(_time>= limit_time-2678400)) as this_month_purchase,
dc(eval(_time<limit_time-2678400 AND _time >= limit_time-2*2678400)) as
previous_month_purchase, dc(eval(_time < limit_time-2*2678400)) as preprevious_month_purchase
by user_uid
| eval group=case(preprevious_month_purchase>0 AND previous_month_purchase=0 AND
this_month_purchase>0, "REACT", previous_month_purchase>0 AND this_month_purchase>0,
"REGULAR", preprevious_month_purchase=0 AND previous_month_purchase=0 AND
this_month_purchase>0, "NEW")
```

Diagrams Active/Inactive, New/Regular/React

Active/Inactive CUSTOMER users

All time | Custom time



	usage	number
1	ACTIVE users	
2	INACTIVE users	

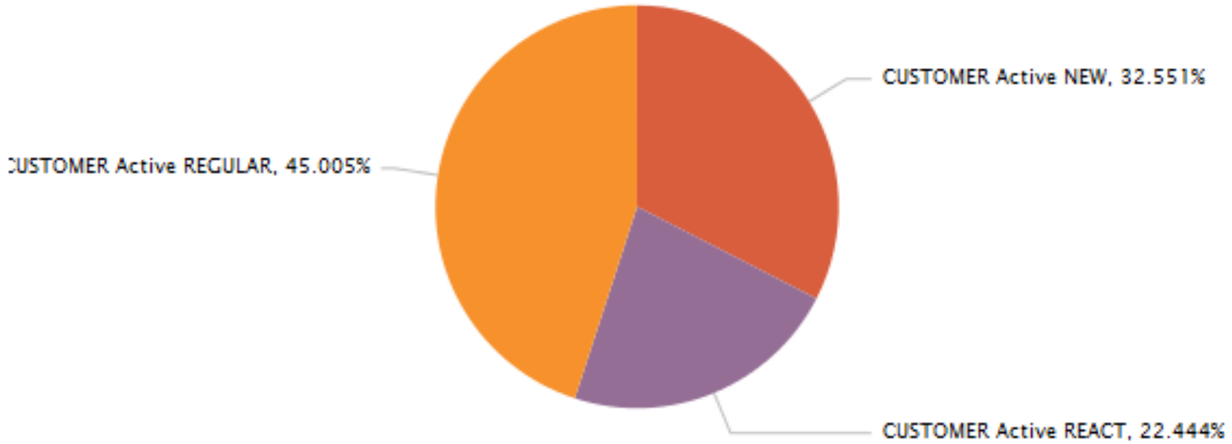
ACTIVE - users that had purchases during the latest 30 days

INACTIVE - users that had purchases earlier than the latest 30 days

Export

New, Regular, React CUSTOMER users

All time | Custom time

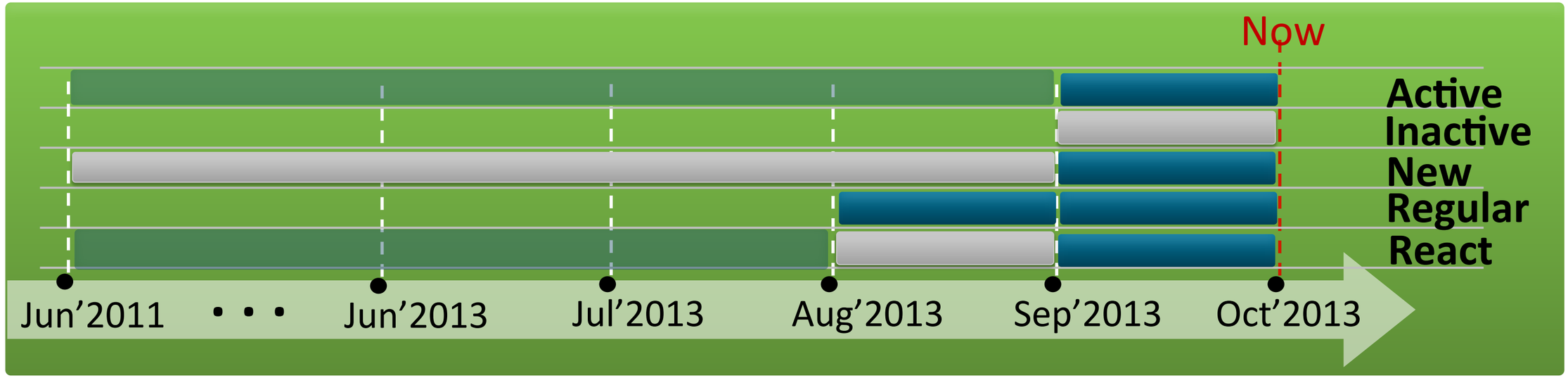


	group	number
1	CUSTOMER Active NEW	
2	CUSTOMER Active REACT	
3	CUSTOMER Active REGULAR	

Export

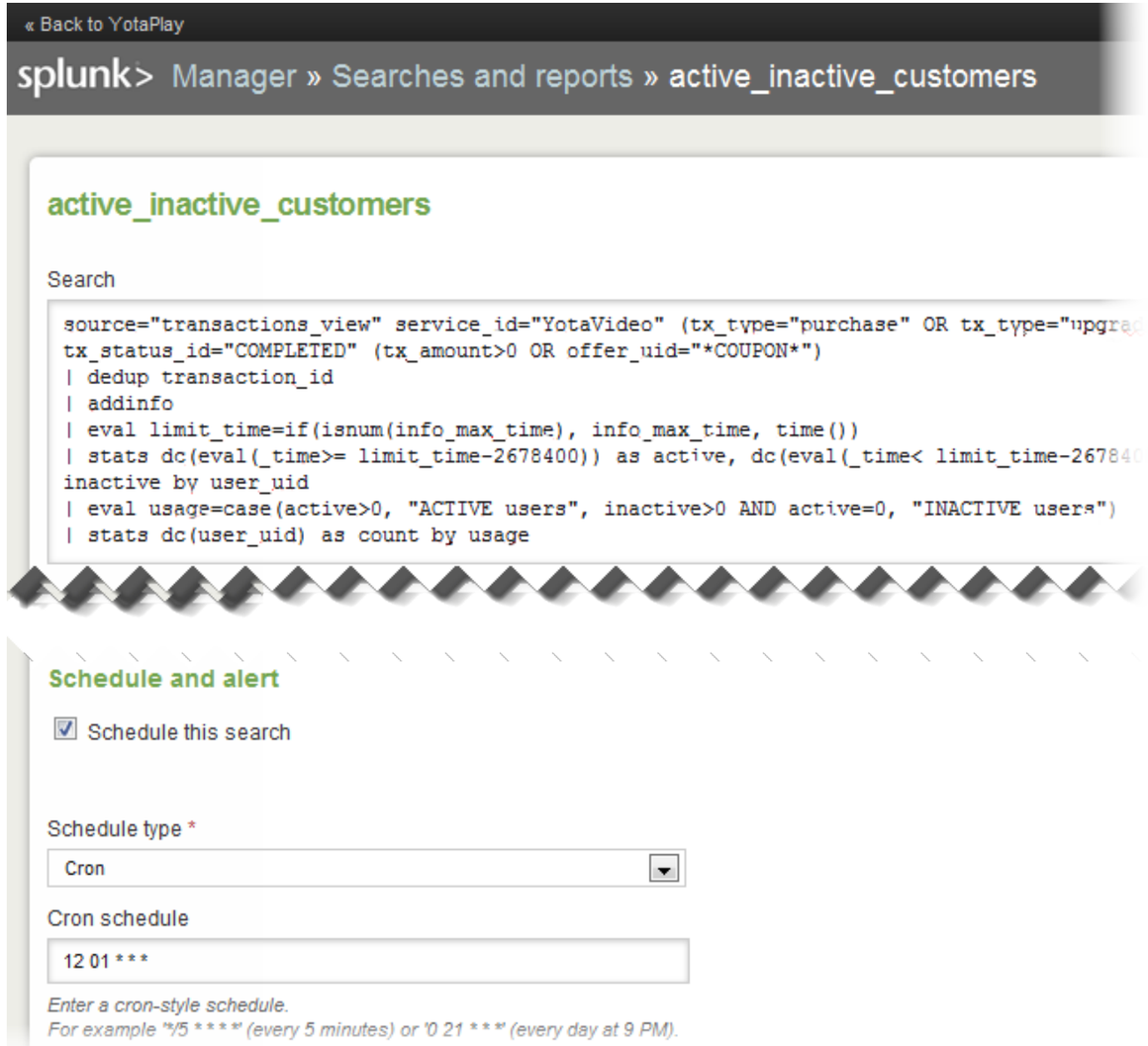
Calculating Historical Data

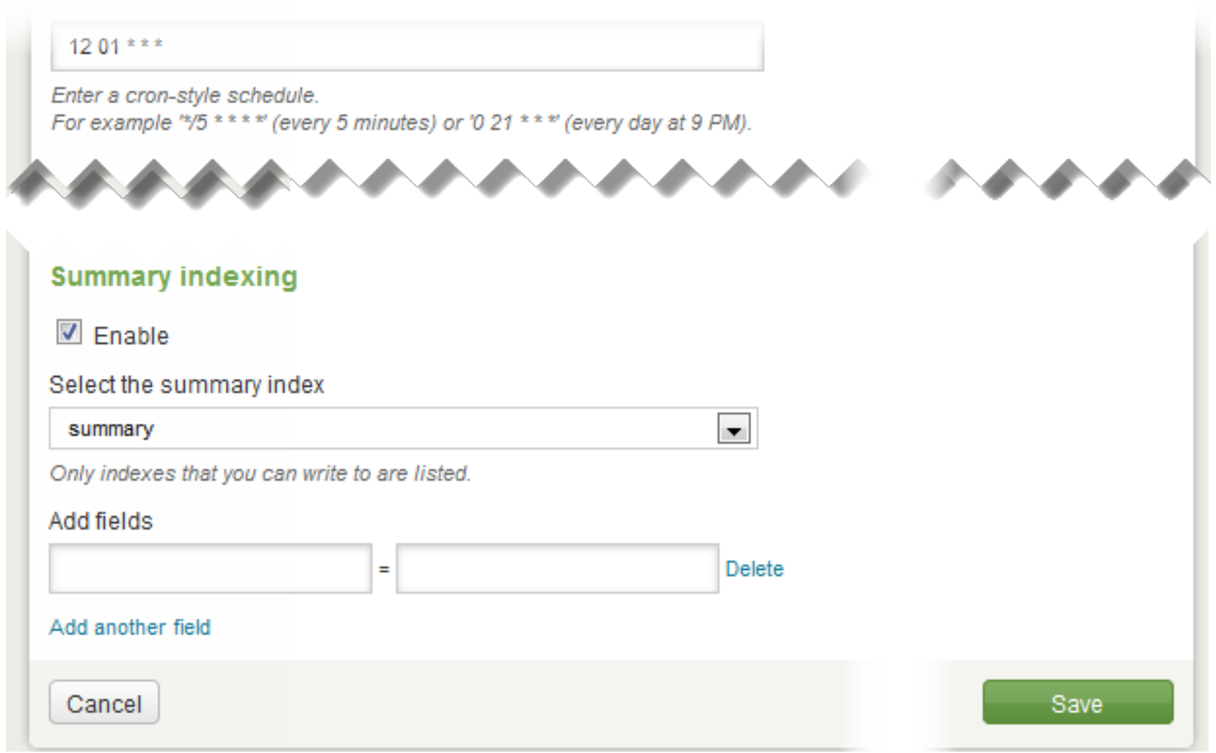
→ Active, Inactive, New, Regular, React –
Each time the need to count till the very beginning…



→ Solution: **summary indexing**

Summary Indexing





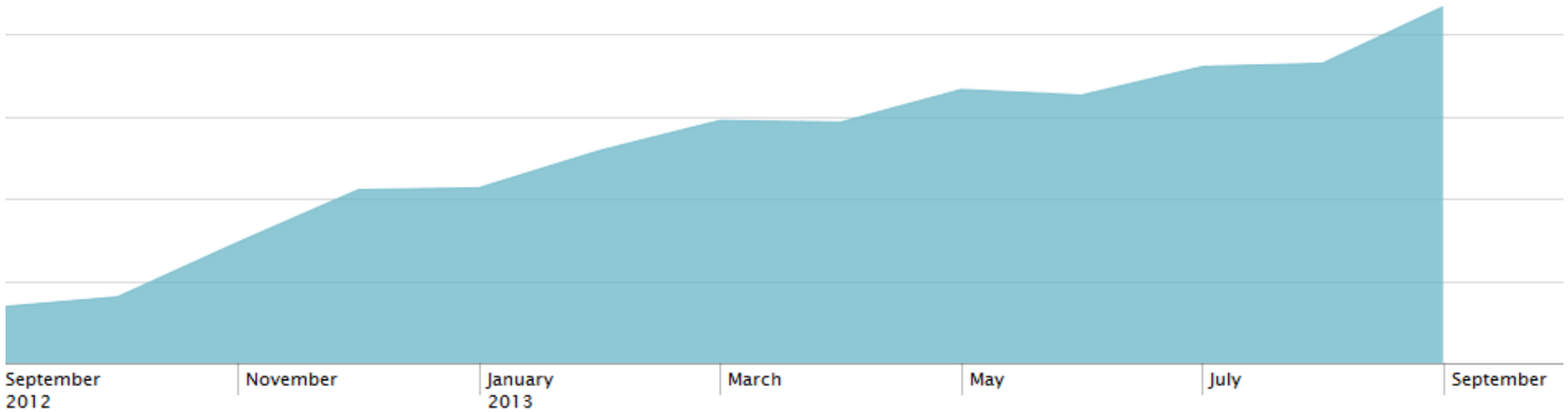
Backfill Script

- How to get historical data for our active/inactive customers and other time critical characteristics?
- Solution: **backfill script**

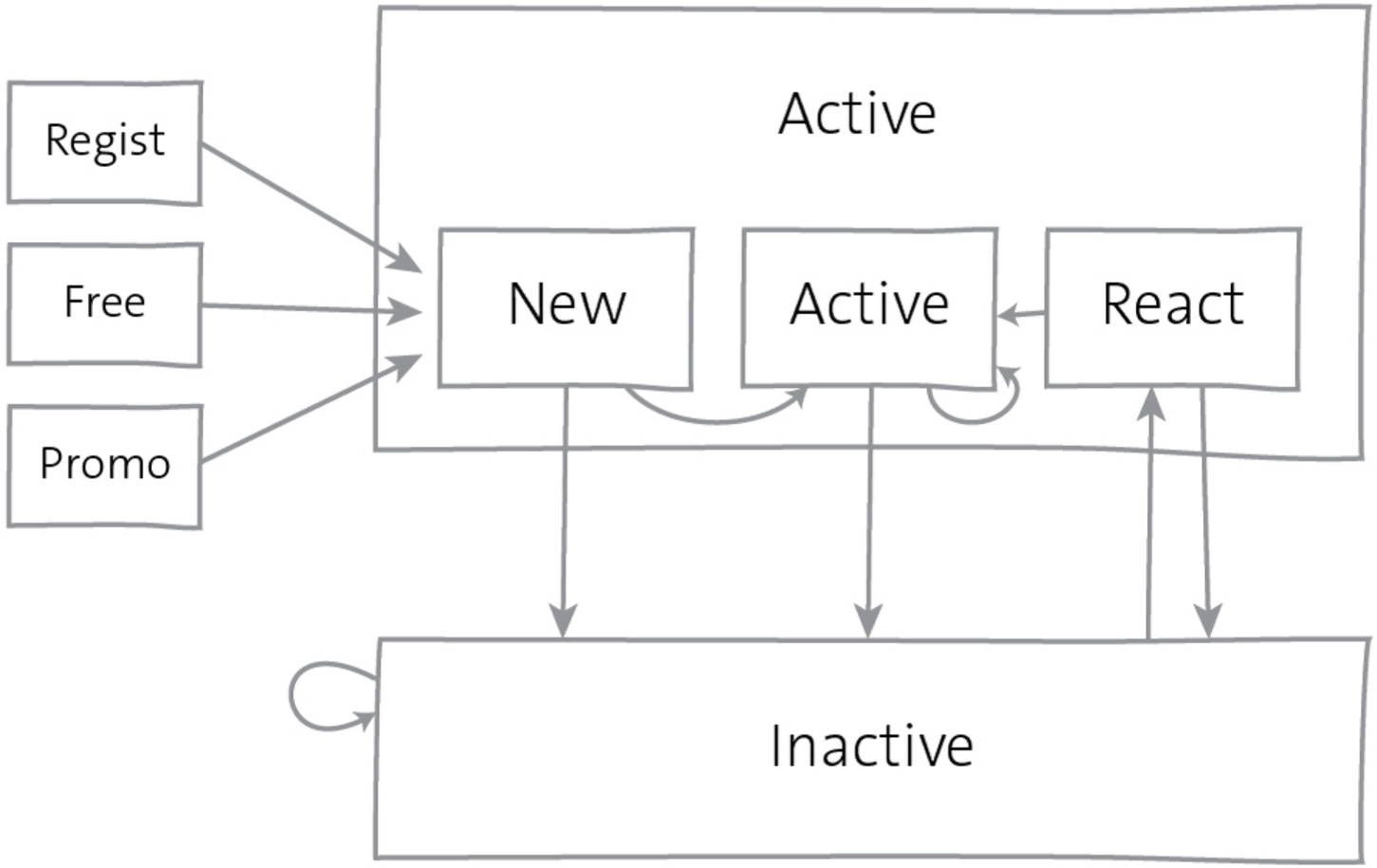
```
./splunk cmd python fill_summary_index.py -app yotaplay -name
active_inactive_customers -et “1310947200” -lt now
```

- Usage of summary indexing and backfill results :

```
index=summary source=active_inactive_customers usage="ACTIVE users"
| timechart span=1mon latest(count) as count
```

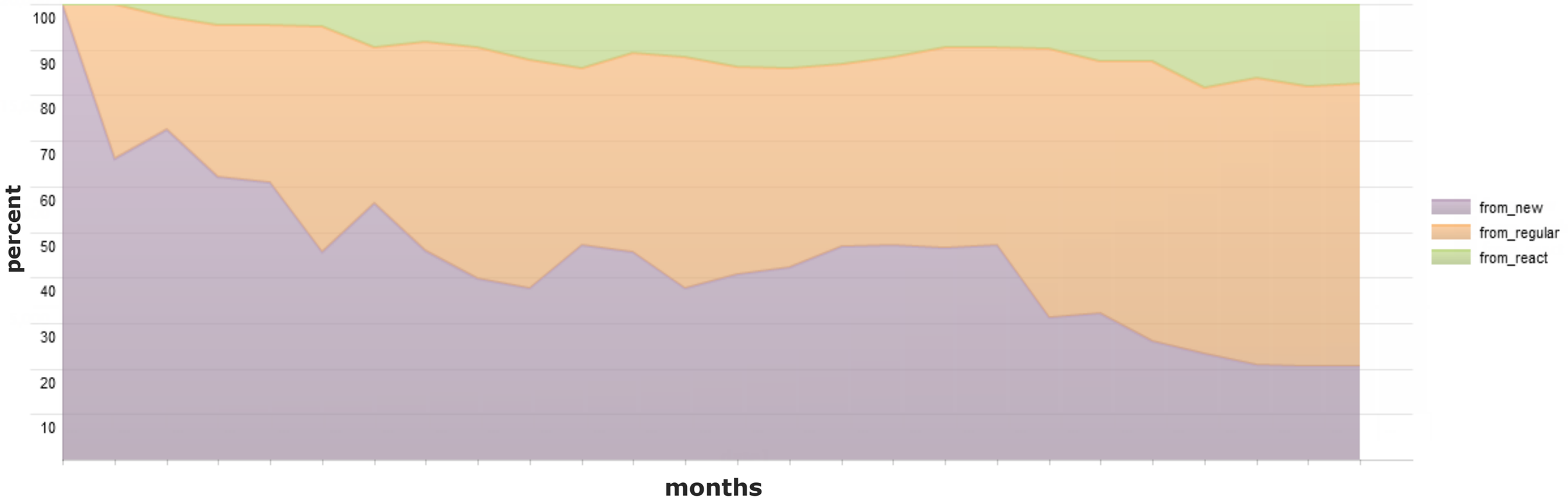


Status Map of Users' Pre-history

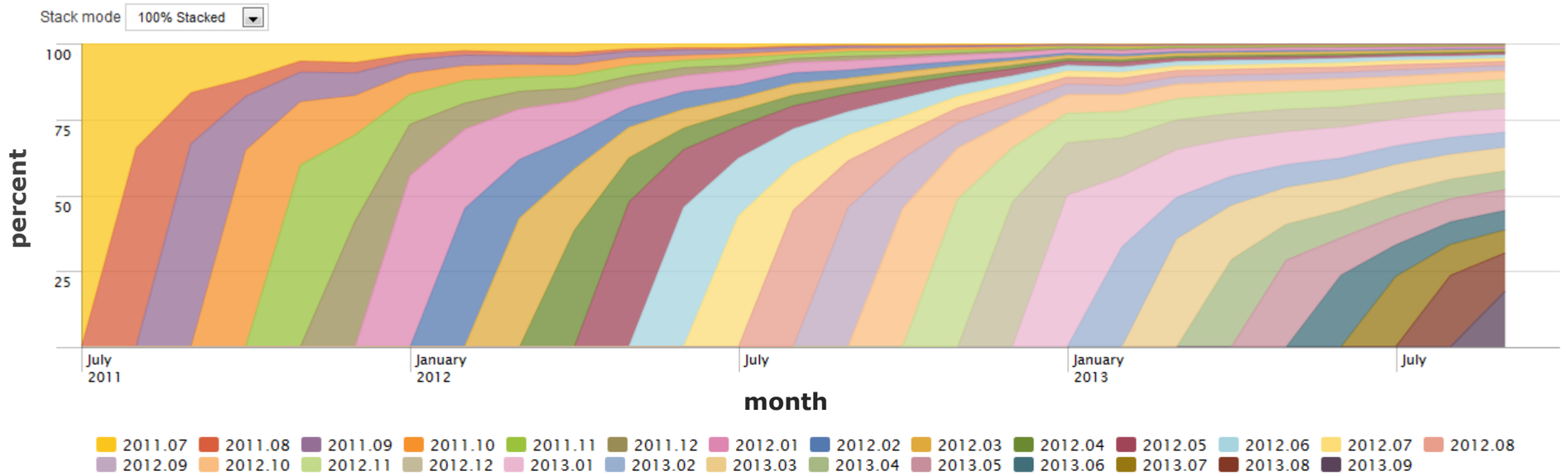


Map of Users' Pre-history

Origin and dynamics of New CUSTOMER users | **Origin and dynamics of Regular CUSTOMER users** | Origin and dynamics of Inactive CUSTOMER users



New Customers Cohorts Diagram



Percent of each cohort in amount of month revenue

« prev 1 2 3 next »

	Time	2011.07	2011.08	2011.09	2011.10	2011.11	2011.12	2012.01	2012.02	2012.03	2012.04	2012.05	2012.06	2012.07	2012.08	2012.09	2012.10	2012.11	2012.12	2013.01	20
1	2011.07	100.0 %																			
2	2011.08	67.6 %	100.0 %																		
3	2011.09	53.2 %	43.2 %	100.0 %																	
4	2011.10	50.2 %	19.9 %	34.9 %	100.0 %																
5	2011.11	40.7 %	20.6 %	31.2 %	52.5 %	100.0 %															
6	2011.12	37.1 %	16.1 %	20.0 %	27.6 %	40.0 %	100.0 %														
7	2012.01	39.1 %	15.4 %	22.2 %	27.1 %	25.6 %	75.3 %	100.0 %													
8	2012.02	28.6 %	14.8 %	19.3 %	21.6 %	21.7 %	43.0 %	52.1 %	100.0 %												
9	2012.03	36.1 %	10.9 %	16.3 %	18.9 %	13.9 %	29.3 %	33.7 %	43.1 %	100.0 %											
10	2012.04	35.6 %	12.6 %	15.3 %	14.8 %	12.2 %	20.8 %	22.3 %	23.7 %	45.9 %	100.0 %										

Lookup Tables (1)

→ "Dynamic fields lookup feature adds fields to your events with information from an external source, such as a static table (CSV file) or an external (Python) command." Creating lookup table with new customers id-s:

1. Create saved search that should be run daily by a schedule

```
…
| eval group="NEW"
| stats earliest(first_purchase_month) as first_purchase_month,
earliest(first_purchase_month_day) as first_purchase_month_day by
user_uid group
| outputlookup append=f group_new_customers.csv
```

2. In the file props.conf file, located in either $SPLUNK_HOME/etc/system/local/ or $SPLUNK_HOME/etc/apps/<app_name>/local/, put:

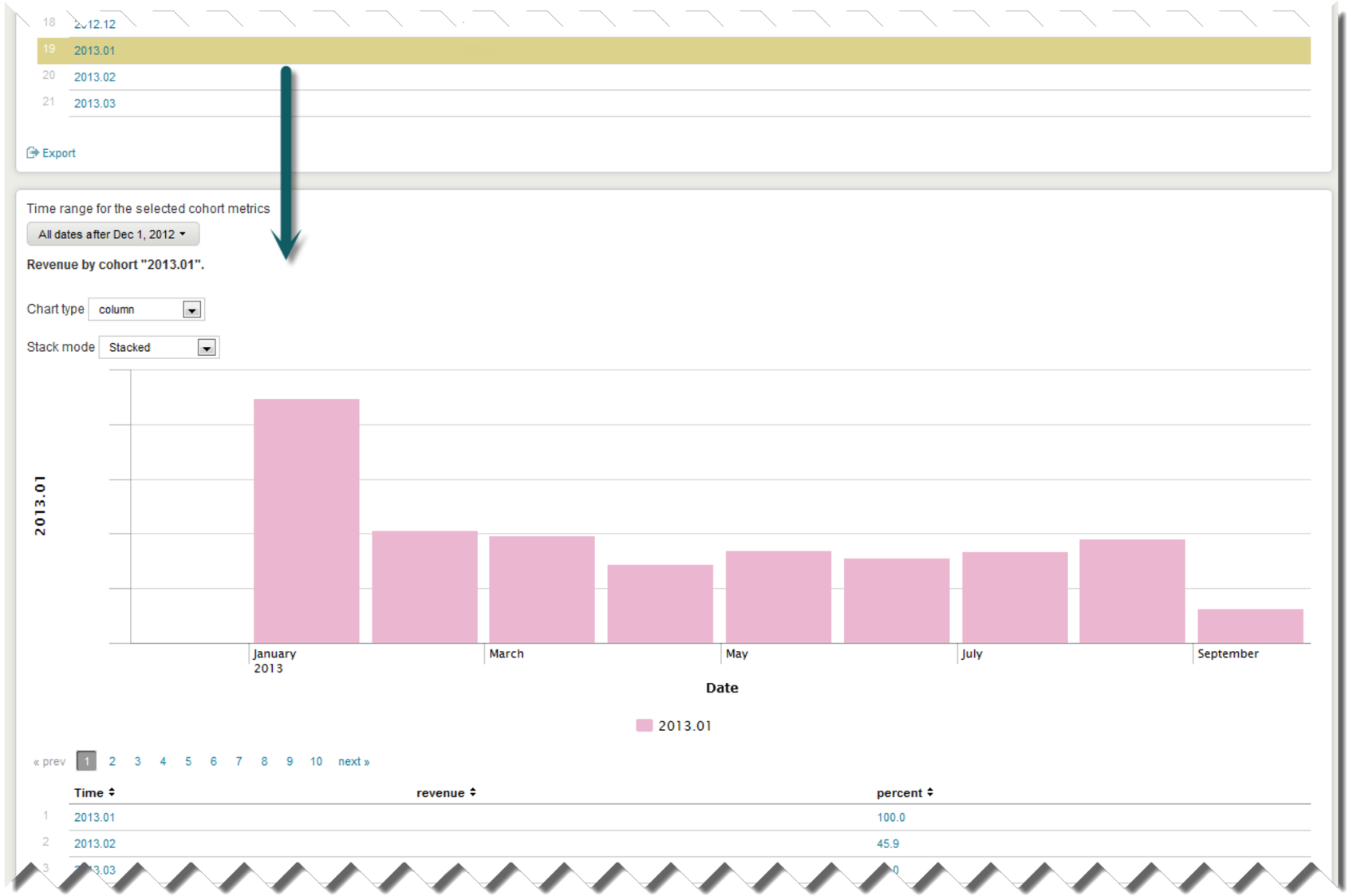
```
[source::transactions_view]
LOOKUP-group_new_customers = group_new_customers user_uid OUTPUTNEW group
first_purchase_month first_purchase_month_day
```

Lookup Tables (2)

3. Restart Splunk

Result: New fields are added to events

Drilldown into Details



Drilldown Into Details Without “Intentions”

→ Ease drilldown usage(and lot more) with SideviewUtils app

```
<?xml version='1.0' encoding='utf-8'?>
<view onunloadCancelJobs="False" autoCancelInterval="100" template="dashboard.html">
    <!-- autoCancelInterval is set here to 100 -->
    <label>Cohorts by new customers</label>
    <module name="AccountBar" layoutPanel="appHeader" />
    <module name="AppBar" layoutPanel="navigationHeader" />
    <module name="SideviewUtils" layoutPanel="appHeader" />
    <module name="Message" layoutPanel="messaging">
        <param name="filter">*</param>
        <param name="clearOnJobDispatch">False</param>
```

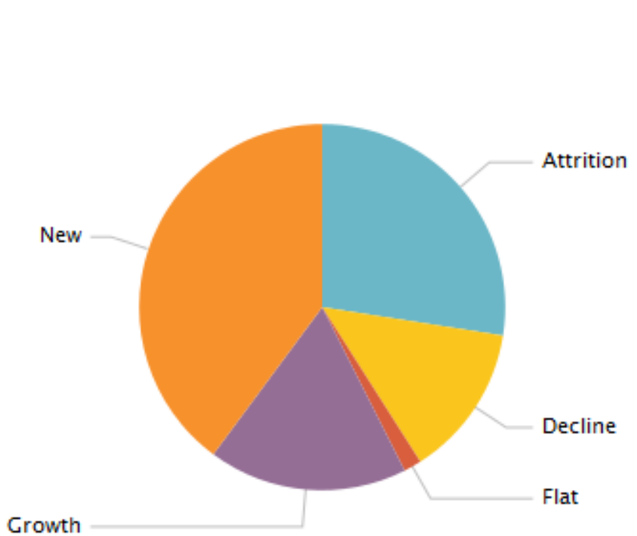
```
        <param name="count">7</param>
        <module name="SimpleResultsTable">
            <param name="entityName">results</param>
            <param name="drilldown">row</param>
            <param name="allowTransformedFieldSelect">True</param>

            <module name="TimeRangePicker" layoutPanel="panel_row3_col1" >
                <param name="selected">All time</param>
                <param name="searchWhenChanged">True</param>
                <param name="label">Time range for the selected cohort metrics</param>
                <module name="Search" autoRun="True" >
                    <param name="search">
                        source="transactions_view" first_purchase_time="$click.value$" group="NEW" tx_amount>0
                        | timechart span=1month sum(tx_amount) as revenue by first_purchase_time where revenue>0
                    </param>
                        <module name="SimpleResultsHeader">
                            <param name="entityName">results</param>
                            <param name="headerFormat">Revenue by cohort "$click.value$".</param>
                        </module>
                        <module name="HiddenChartFormatter">
```

Flux Analysis

Flux analysis – revenue alteration from the customer in the current month compared with previous month revenue

Calculated for each customer with monetary purchases:

1. Revenue for the previous month (Revenue1) and revenue for the current month (Revenue2)
2. Revenue difference: *MoM_Change = Revenue2 - Revenue1*
3. Revenue difference in percent:
   % Change = MoM_Change/Revenue 1
4. Customer is assigned a group according the value of % Change:
   - Attrition (Change = -100%)
   - Decline (Change < 0%)
   - Flat (Change = 0%)
   - Growth (Change > 0%)
   - New (previous period revenue is null, Change = 100%)



Group	Users	MoM_Change sum	Users' perce
Attrition (Change = -100%)			27.5 %
Decline (Change < 0%)			13.5 %
Flat (Change = 0%)			1.6 %
Growth (Change > 0%)			17.6 %
New (Change = 100%)			39.9 %

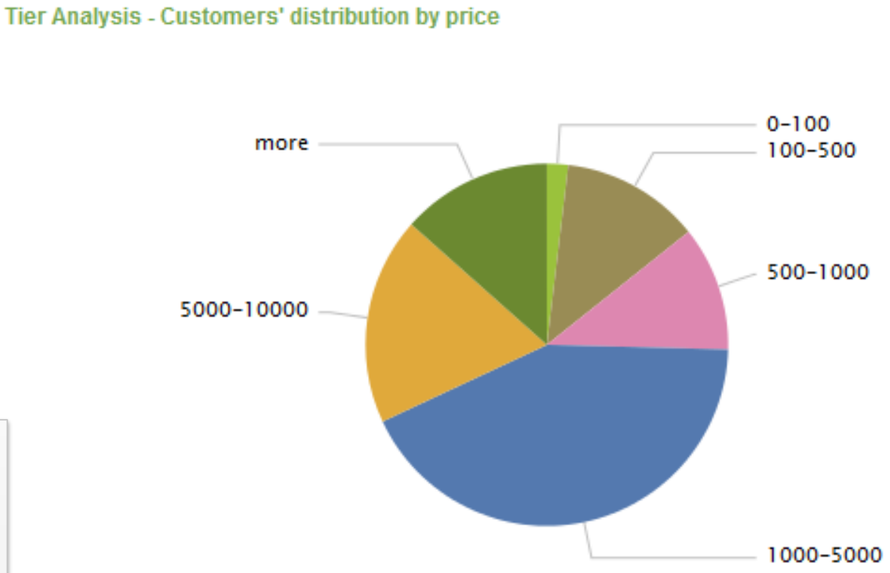
Tier Analysis

- Tier Analysis - customers' distribution by price for the period

Groups example:

<100 rub
> 100 < 500 rub
> 500 < 1000 rub
> 1000 < 5000 rub
> 5000 < 10000 rub
> 10000 rub

```
…
| chart sum(tx_amount) as revenue, count as
purchases by user_uid
| rangemap field=revenue 0-100=1-100
100-500=101-500 500-1000=501-1000
1000-5000=1001-5000 5000-10000=5001-10000
more=10001-10000000
| chart sum(revenue) as Revenue,
count(purchases) as "Number of Customers"
by range
```



	range	Revenue	Number of Customers
1	0-100		
2	100-500		
3	500-1000		
4	1000-5000		
5	5000-10000		
6	more		

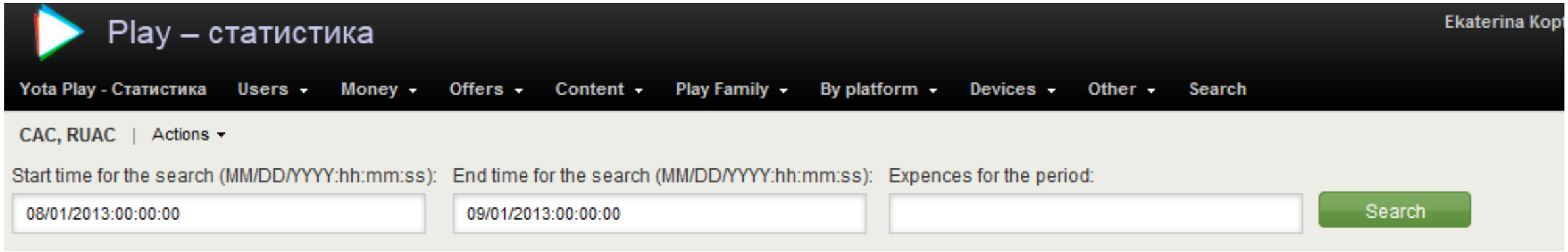
Acquisition Costs

- **CAC** – customer acquisition cost

  CAC = expenses / new_customers

- **RUAC** – registered user acquisition cost

  RUAC = expenses / new_users



Acquisition Costs: Form Search

"A form is a Splunk view similar to a dashboard, but provides an interface for users to supply values to one or more search terms, typically using text boxes, dropdown menus, or radio buttons."

http://docs.splunk.com/Documentation/Splunk/5.0.4/Viz/Buildandeditforms

```
<?xml version='1.0' encoding='utf-8'?>
<form>
    <label>CAC, RUAC</label>
    <fieldset>
        <input type="text" token="start">
            <label>Start time for the search (MM/DD/YYYY:hh:mm:ss):</label>
            <default>07/19/2011:00:00:00</default>
            <seed>07/19/2011:00:00:00</seed>
        </input>
        <input type="text" token="end">
            <label>End time for the search (MM/DD/YYYY:hh:mm:ss):</label>
            <default>09/19/2011:00:00:00</default>
            <seed>09/19/2011:00:00:00</seed>
        </input>
        <input type="text" token="expences">
            <label>Expences for the period:</label>
            <default>1000</default>
            <seed>1000</seed>
        </input>
    </fieldset>
    <row grouping="2">
        <single>
            <title>Registered User Acquisition Cost (RUAC)</title>
            <searchString>
                source="profiles" earliest=$start$ latest=$end$
                | stats dc(user_uid) as new_customers
                | eval Expences=$expences$
                | eval RUAC=round(Expences/new_customers,3)
                | stats values(RUACb) as RUAC
            </searchString>
            <option name="field">RUAC</option>
            <option name="beforeLabel">RUAC:</option>
        </single>
        <table>
            <se  Str
```

More Wonderful Metrics

CHURN - the value of loosing active customers

CHURN = B/A *100

A = active customers in previous 30 days

B = inactive customers in current month, that were active in previous month

Attrition - difference between this month and previous month inactive Customers

attrition=inactive_current - inactive_previous

Grow rate – the value of new Customers in the service within active Customers of previous month

Grow = new_current/active_previous *100

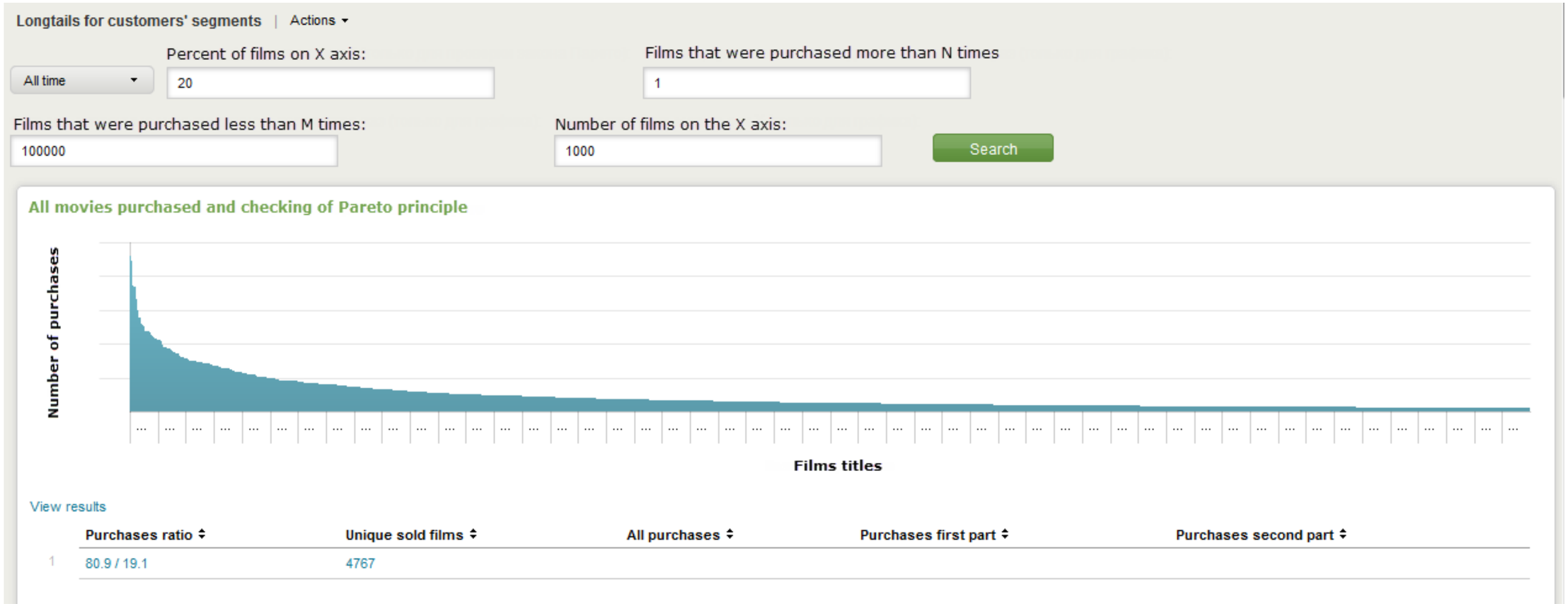
Retention - the value of active Customers percent in target month within all Customers in the previous month

Retention =100*active_current/total_previous

Checking of Pareto Principle in Your Business (80–20 Rule)

"For many events, roughly 80% of the effects come from 20% of the causes."

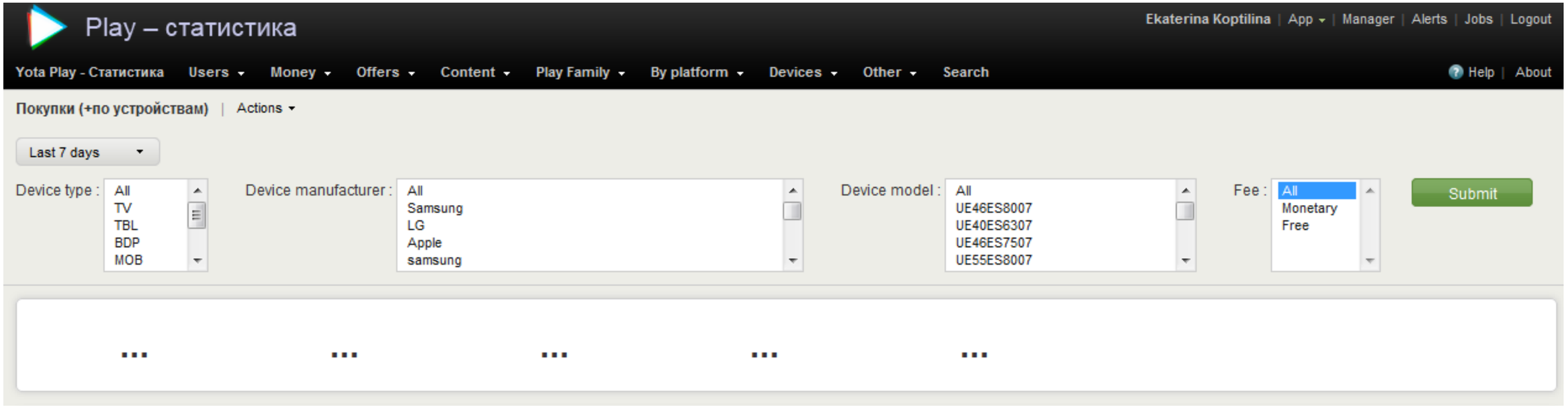
Joseph M. Juran
business-management consultant



Query for Checking Pareto Principle

```
…
| stats count as sold_films by ru_title
| eval counter=1
| accum counter
| sort 0 - sold_films
| eventstats max(counter) as all_films
| eval film_number_for_percent=all_films*$percent$/100
| stats sum(eval(if(counter<=film_number_for_percent, sold_films, null()))) as sold_films_first,
sum(eval(if(counter>film_number_for_percent, sold_films, null()))) as sold_films_second,
sum(sold_films) as all_purchases, max(all_films) as all_films
| eval purchases_ratio=round(sold_films_first*100/all_purchases, 1)." / ".round(sold_films_second*100/all_purchases, 1)
| table purchases_ratio  all_films  all_purchases  sold_films_first  sold_films_second
```

Specifying Series of Results in a Dashboard



Have seen such header first in S.o.S - Splunk on Splunk (Thanks!)

→ Pulldown module + layoutPanel="mainSearchControls"

Specifying Series of Results in a Dashboard (2)

```
</module>

<module name="Search" layoutPanel="mainSearchControls" autoRun="True">
    <param name="search">source="transactions_view" | table device_type device_manufacturer device_model tx_amount offer_uid
    <module name="TimeRangePicker">
        <param name="selected">Last 7 days</param>
        <param name="searchWhenChanged">True</param>
        <param name="label"> </param>

        <module name="JobProgressIndicator"></module>

        <module name="Pulldown">
            <param name="name">device_type</param>
            <param name="template">$name$="$value$"</param>
            <param name="postProcess"> stats count by $name$ | sort - count </param>
            <param name="label">Device type :</param>
            <param name="float">left</param>

            <param name="size">5</param>
            <param name="separator">+OR+</param>
            <param name="outerTemplate">( $value$ )</param>

            <param name="searchFieldsToDisplay">
                <list>
                    <param name="label">device_type</param>
                    <param name="value">device_type</param>
                </list>
            </param>

            <module name="Pulldown">
                <param name="name">device_manufacturer</param>
                <param name="template">$name$="$value$"</param>
                <param name="postProcess">search $device_type$ | stats count by $name$ | sort - count </param>
                <param name="label">Device manufacturer :</param>
                <param name="float">left</param>
```

```
                <param name="value">false</param>
            </list>
        </param>

        <module name="Button">
            <param name="allowSoftSubmit">False</param>
            <param name="allowAutoSubmit">False</param>
            <param name="label">Submit</param>

            <module name="Search" layoutPanel="panel_row1_col1" autoRun="True">
                <param name="search">
                    source="transactions_view" service_id="YotaVideo" $device_type$ $device_manufacturer$ $device_model$
                    (tx_type=purchase OR tx_type=upgrade) tx_status_id=COMPLETED
                    | `monetary`
                    | search $monetary$
                    | `coupon_price`
```

Service Health Indexes

Service health indexes | Actions

Registrations

	Registrations in latest 30 days	Registrations grow index
1		22 %

ARPU

	ARPU in latest 30 days, rub.	ARPU grow index
1		16 %

Revenue

	Revenue in latest 30 days, rub.	Revenue grow index
1		40 %

Metrics

	Metric	Index, %	Value in latest 30 days	date	Value in previous 30 days	date1
1	CHURN	-10	43 %	2013.09.09	48 %	2013.08.09
2	Grow	15	39 %	2013.09.09	34 %	2013.08.09
3	Retention	11	31 %	2013.09.09	28 %	2013.08.10
4	LT	10	69 days	2013.09.08	63 days	2013.08.08
5	LTV	14		2013.09.09		2013.08.09
6	New	22		2013.09.09		2013.08.09
7	Attrition	-31		2013.09.09		2013.08.10

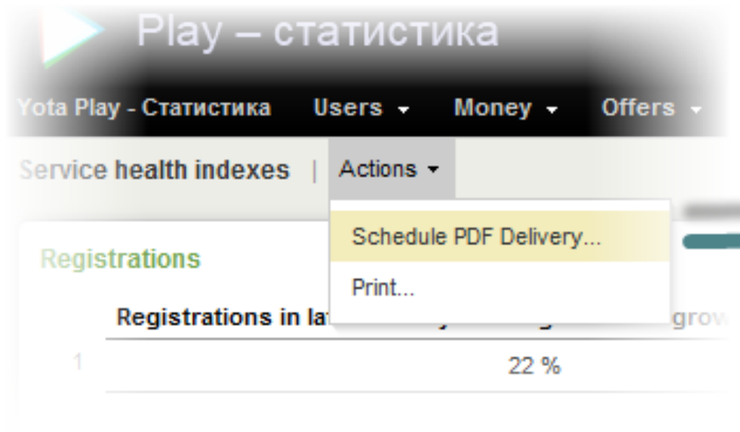
Customers

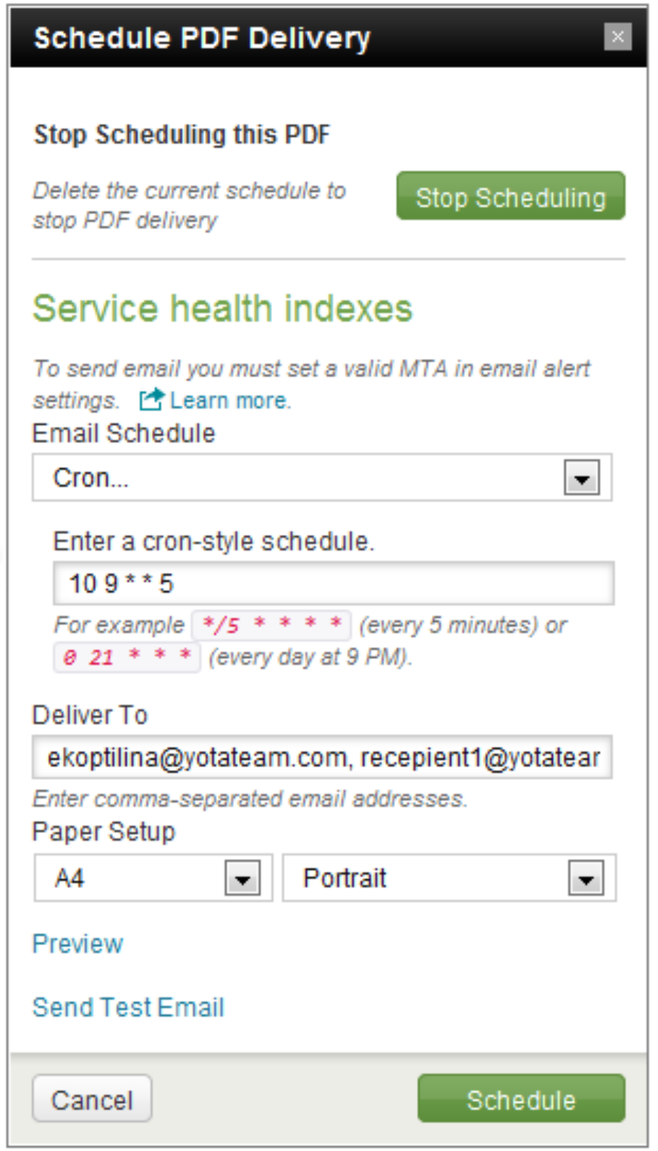
	Metric	Index, %	Value in latest 30 days	date	Value in previous 30 days	date1
1	Active Customers	20		2013.09.09		2013.08.09
2	New Customers	22		2013.09.09		2013.08.09
3	Regular Customers	15		2013.09.09		2013.08.09
4	React Customers	29		2013.09.09		2013.08.09
5	Inactive Customers	7		2013.09.09		2013.08.09

Offers

	Offer name	Purchases in latest month	Purchases in previous month	Purchases grow index	Conversion	Customer users	Promo users
1	FREE_PURCHASES_REGISTRATION			37 %			
2	F_PU[illegible]SES[illegible]ST			6.0			

Service Health Indexes: Schedule





Repeat at Home

Metrics:

- Revenue
- Number of customers
- ARPU
- Index metric
- Mean revenue per day, weekday, working day
- Visitor, Prospect, Promo, Customer, Buyer
- Active, Inactive
- New, Regular, React
- By absence periods

- Map of users' pre-history
- Cohorts diagram
- Flux analysis
- Tier analysis
- CAC
- RUAC
- CHURN
- Attrition
- Grow rate
- Retention
- Longtail and Pareto principle
- Service health indexes

Techniques:

- Summary indexing
- Backfill script
- Lookup tables
- Drilldown
- Form search
- Pulldown

One Splunk, One Love

What do we value most in Splunk:

- Acceleration of business and IT processes in the company
- Velocity of operating
- Browser-like search
- Search language and interface
- Neat dashboards design

Direct assistance in:

- Product management
- Content management
- Financial management
- Business strategy planning
- IT-operations management

"Splunk is a dashboard with almost endless number of leverages to drive your business right in the direction you need."

Vitaly Grigorash
Product director

Summary

- Splunk gives Play key business insights from machine data
  - **Churn** – The value of losing active customers
  - **Attrition** – The difference in inactive customers between this month & previous month
  - **Growth** – the value of new customers this month
  - **Retention** – the value of active existing customers

- Splunk gives us the ability to accelerate business and IT processes
- With Splunk we now have a higher operating velocity

Next Steps

1. **Download the .conf2013 Mobile App**
   If not iPhone, iPad or Android, use the Web App

2. Take the survey & **WIN A PASS FOR .CONF2014**… Or one of these bags!

3. **View all of the "Beyond IT" sessions**
   All sessions are available on the Mobile App
   Videos will be available shortly


.conf2013
YOUR DATA
NO LIMITS
Questions and Answers
splunk>


.conf2013
YOUR DATA
NO LIMITS
THANK YOU
splunk>