

Sankey Path Analysis of Web Button Clicks and Page Views

Nancy Abramson

Product Data Scientist

nabramson@ieee.org

720-468-1796

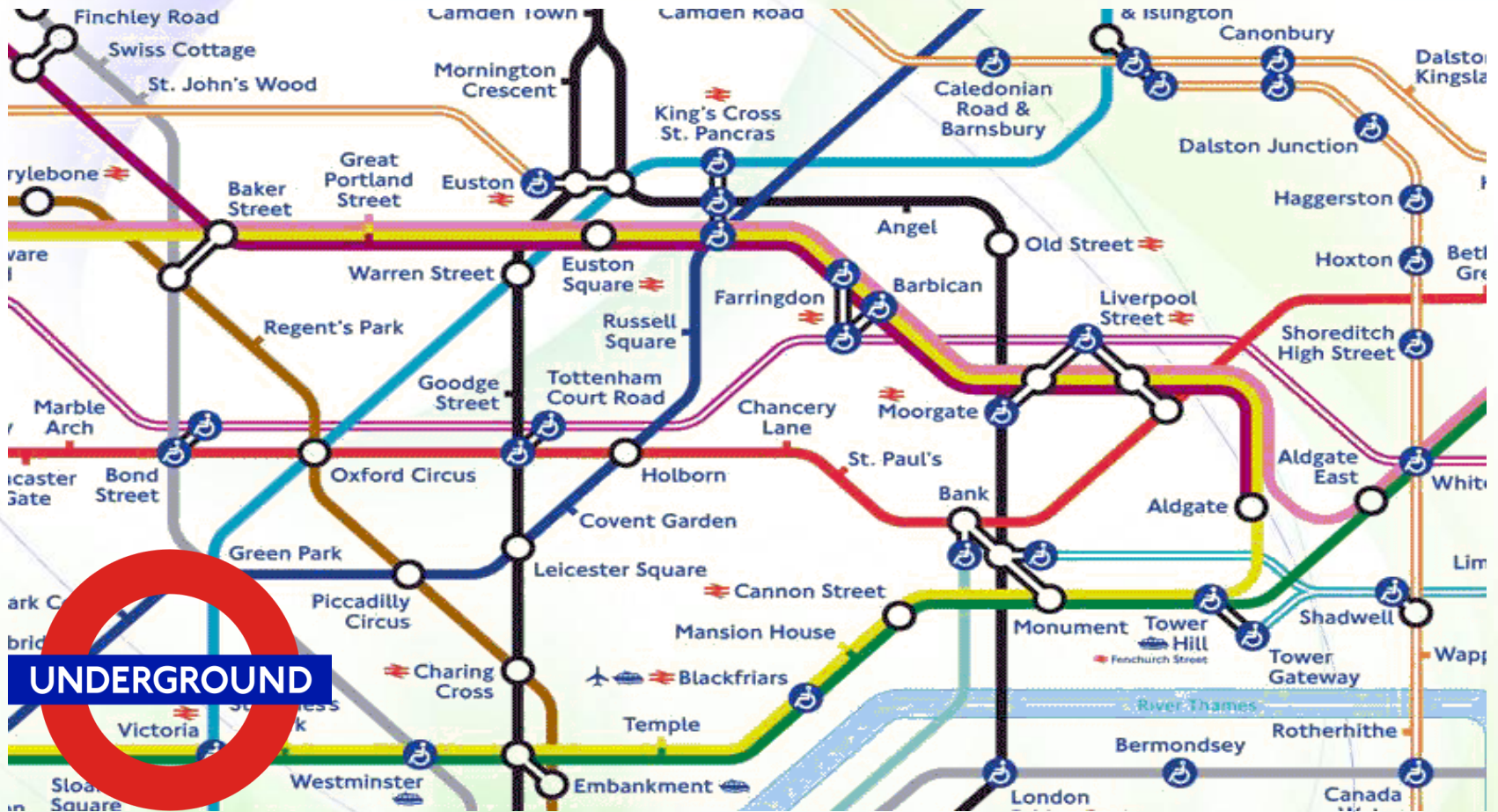
Customer Application Fails

Understand which customer actions create a poor experience

- Replicate for testing
- Focus for programming fix
- Redirect user away from an error to safe haven

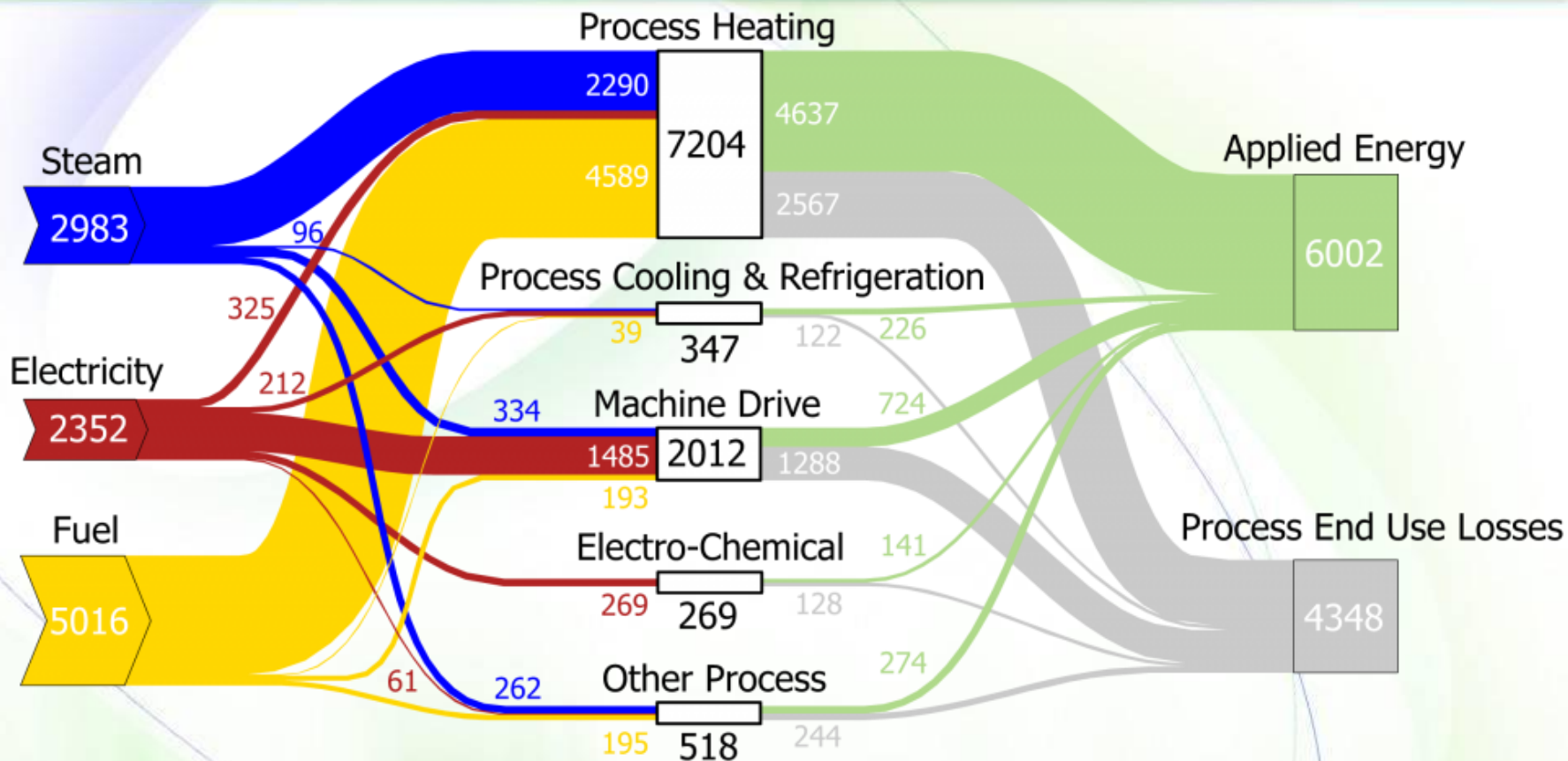


Mind the (Sankey) Gap



A Sankey diagram is a map showing how information flows through events

Sankey Diagram Power



- Flow diagram
- Width proportional to flow
- Dense Data Display

Source: <http://www.energy.gov/eere/amo/static-sankey-diagram-process-energy-us-manufacturing-sector>

Analysis Overview

- Combine Web and Server Data Sources
 - ▶ Best Practices to Blend Data Sources
 - ▶ Method for Disparate Session
 - ▶ SQL Sample to Fill in Sparse Data
 - ▶ Aggregate Values from Dual Sources
- Result: Sankey Visual Analysis
- Application: Automation and Targeted Debug

Blend Data Sources

- Adobe Omniture Web
 - ▶ Third party service
 - ▶ Collects website clicks made by visitors
 - ▶ Organized by events in a session
- Apache Tomcat Server
 - ▶ System health
 - ▶ System errors
 - ▶ Log report
 - ▶ Aggregate Values from Dual Sources



Union data together within a database table

Data Blending Difficulties

● Each data source often has it's own...

- ▶ Surrogate Key
- ▶ Session Definition
- ▶ Data Elements
- ▶ Server Location
- ▶ Reporting Speed
- ▶ Sparse Data



Benefit of using different data sets is seeing the data from multiple perspectives

Harmonize

- Impose structure to blend
 - ▶ Need common natural or passed key between data sets (e.g. account, device id, user id)
 - ▶ Store timestamps without time zone (UTC)
 - ▶ Synchronize servers with Network Time Protocol (NTP)
 - ▶ Combine data sources into conformed columns to preserve the richness of the user experience
- Insert metadata columns
 - ▶ Source of data for traceability
 - ▶ Insert date timestamp
 - ▶ Session using ***your own business*** definition

Business Defined Session and Data

- Partition Window Over Natural Key
 - ▶ Order by time
 - ▶ Use business rules for max/min session times
 - ▶ Require session events such as login/logout
 - ▶ Select values in window (min, max, first, last)
- Conform Sparse Events Across Columns

```
COALESCE(post_pagename, '') ||  
COALESCE(post_buttonclick, '') ||  
COALESCE(post_togglename, '') as behavior
```

Compress Events Within Column

- Replace NULL Row Values within Session

```
UPDATE event_table set user_id = v.user_id
FROM
  (SELECT max(user_id) as user_id, session_id
    FROM event_table
   WHERE session_id IN
      ( SELECT distinct session_id FROM
        event_table WHERE source = 'omniture' AND user_id
        is NULL )
     GROUP BY session_id ) v
WHERE v.session_id = event_table.session_id
AND event_table.user_id IS null AND source =
'omniture';
```

Simplify Events Across Columns

● Use Business for Multiple Column Compression

CASE

WHEN (event is NULL OR event LIKE 'APP') AND
channel is null THEN error_code

WHEN (event is NULL OR event LIKE 'APP') AND
(error_code is null OR error_code LIKE 'APP') THEN
channel

WHEN (error_code is NULL OR error_code LIKE 'APP')
AND channel is NULL THEN event

WHEN (event is NULL OR event LIKE 'APP') THEN
error_code

WHEN channel is NULL THEN error_code

WHEN (event is NULL OR event LIKE 'APP') THEN
channel

ELSE event

END AS click

Example Row Count Aggregation

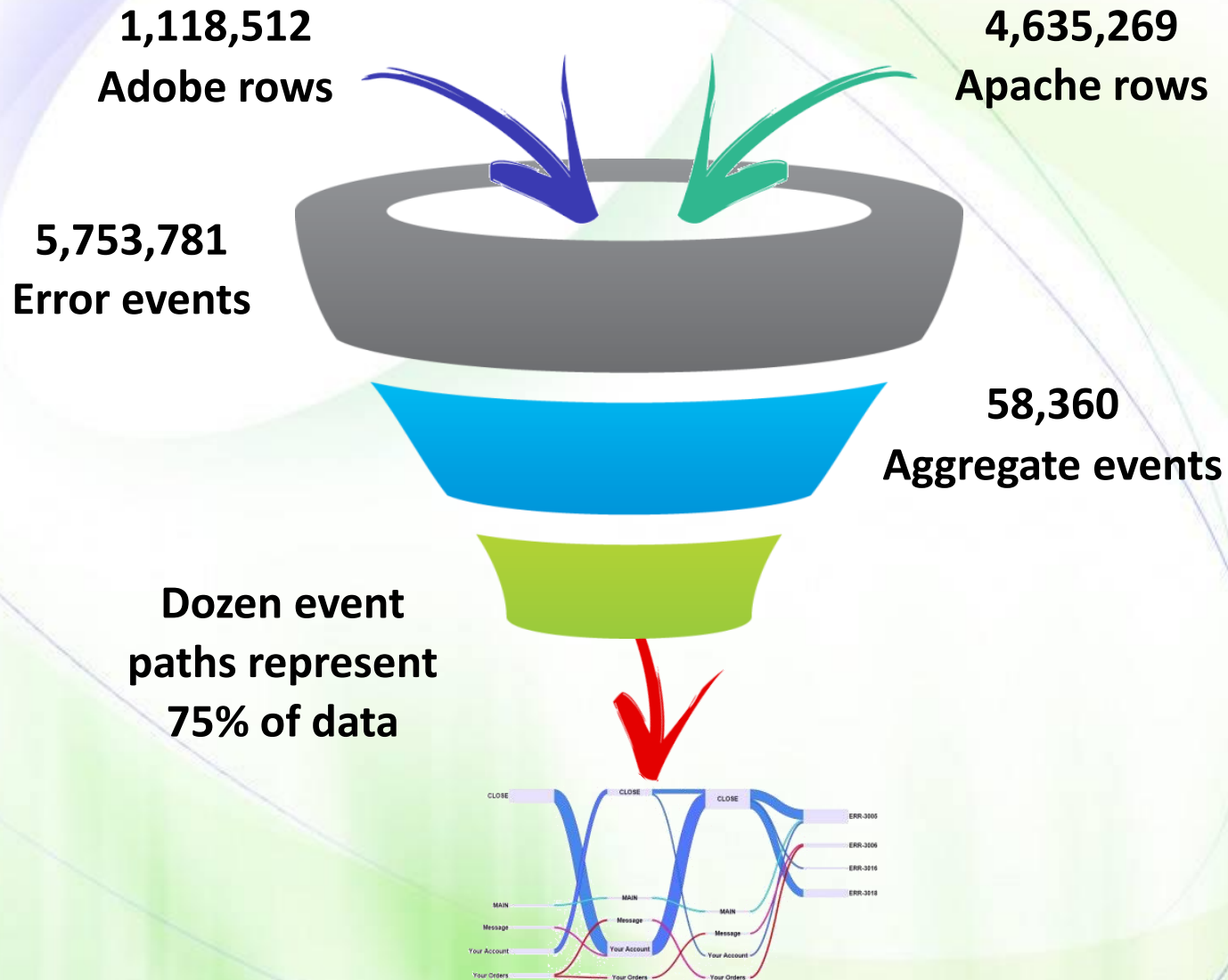


common_id	event	source	time
100ASD39Q	Your Account	<u>omniture</u>	2015-07-23 16:00:02
100ASD39Q	Message	<u>omniture</u>	2015-07-23 16:00:45
100ASD39Q	CLOSE	<u>omniture</u>	2015-07-23 16:01:54
100ASD39Q	ERR-3005	tomcat	2015-07-23 16:03:19
100ASDAFX	CLOSE	<u>omniture</u>	2015-07-16 22:35:09
100ASDAFX	Your Account	<u>omniture</u>	2015-07-16 22:41:15
100ASDAFX	CLOSE	<u>omniture</u>	2015-07-16 22:42:38
100ASDAFX	ERR-3005	tomcat	2015-07-16 22:50:26
111C320L20	CLOSE	<u>omniture</u>	2015-07-23 15:50:56
111C320L20	Your Account	<u>omniture</u>	2015-07-23 15:51:03
111C320L20	CLOSE	<u>omniture</u>	2015-07-23 15:55:23
111C320L20	ERR-3018	tomcat	2015-07-23 16:00:12

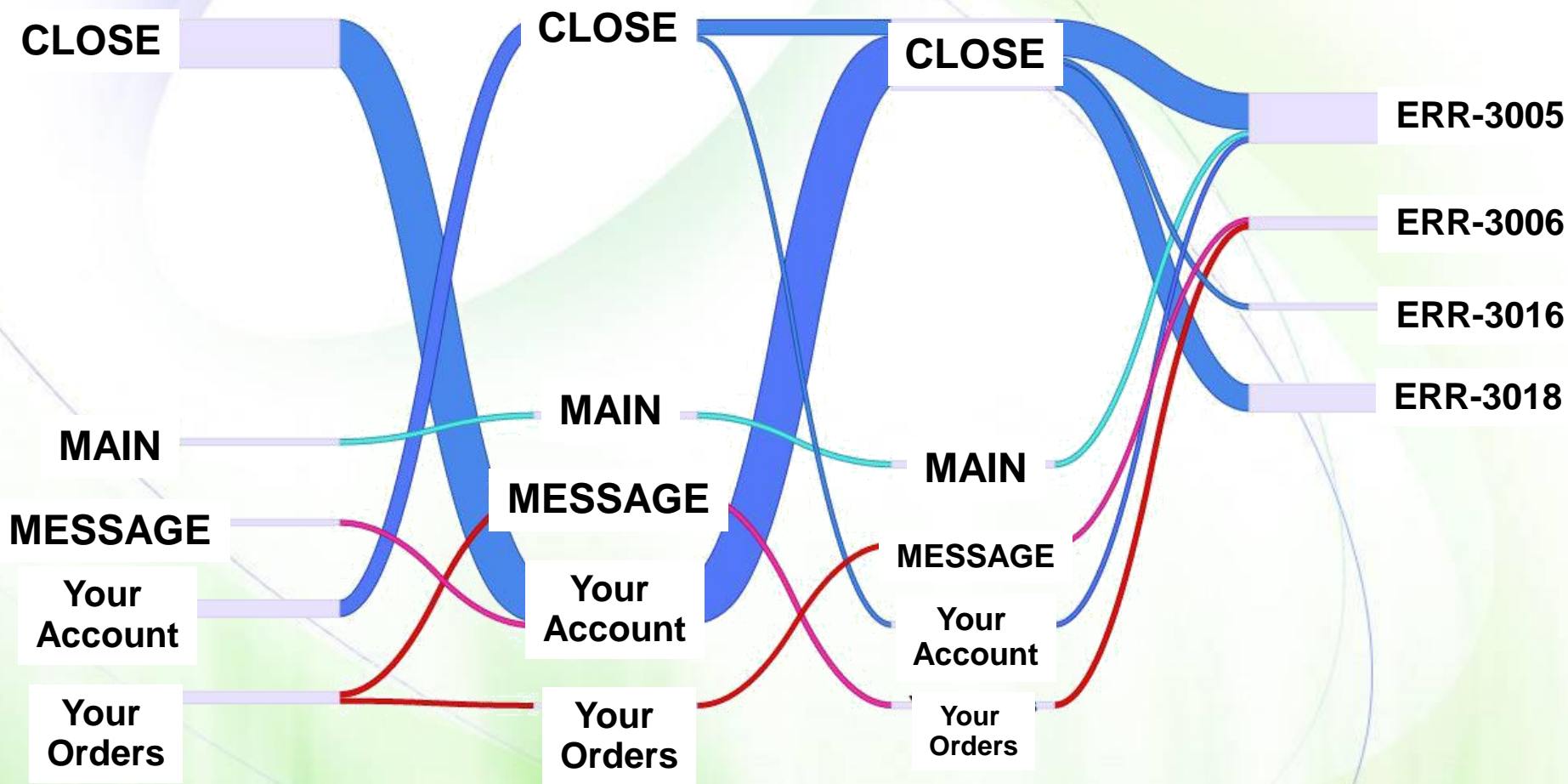


common_id	error_code	any_event	time_start	time_end
100ASD39Q	ERR-3005	[Your Account, Message, CLOSE, ERR-3005]	2015-07-23 16:00:02	2015-07-23 16:03:19
100ASDAFX	ERR-3005	[CLOSE, Your Account, CLOSE, ERR-3005]	2015-07-16 22:35:09	2015-07-16 22:50:26
111C320L20	ERR-3018	[CLOSE, Your Account, CLOSE, ERR-3018]	2015-07-23 15:50:56	2015-07-23 16:00:12

Row Event Compression



Resulting Sankey Diagram



Results

- Manual
 - ▶ Business Review
 - ▶ Targeted Development Debug
 - ▶ Focused Testing and Replication
- Automatic
 - ▶ Top error path re-routes to working page

More Information

- Free Classes: <https://www.coursera.org/>
- *The Visual Display of Quantitative Information*
Edward Tufte
- Teradata Aster Big Data Analytics