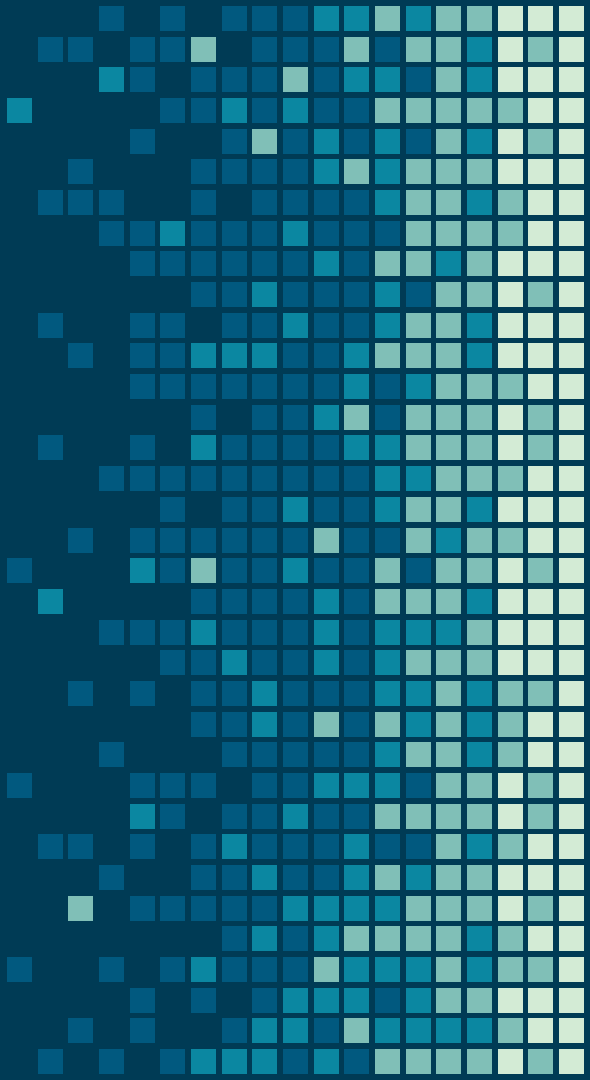
DASHBOARDS AND DATA LAKES WITH TABLEAU



BACKGROUND

Food Protection and Defense Institute (FPDI)

Research Institution located at the University of Minnesota. Started by Department of Homeland Security to look at threats of **"Intentional contamination"** of the Food Supply. FPDI is a part of the Academic Health Center (AHC).

Ted Steinmann

"IT manager" at FPDI and pursuing masters in the University of Minnesota's the Technology Leadership Institute **"Management of Technology"** program.

Why did we build a data lake? Why are we embedding dashboards in web applications?

To centralize numerous distributed datasets into a large-scale data repository and processing engine and to provide access to a diverse group of sponsors, stakeholders and consumers.

DEFINITIONS

FOOD SAFETY

Reducing exposure to **natural hazards**, errors, failures.

- Accidental
- Food poisoning
- Cleanliness

FOOD SECURITY

"When all people at all times have **access** to sufficient, safe, nutritious food to maintain a healthy and active life".

- World Health Organization

FOOD DEFENSE

Reducing the impact of **system attacks** by assuring the stability or continuity of the food supply.

FOOD ADULTERATION INCIDENTS REGISTRY

- When and where is it taking place?
- What foods are adulterated, and by what mechanism?
- How often does it occur?

Primary concerns and data collected are type of incident , motivation , adulteration type.

MELAMINE IN MILK

Used for making whiteboards

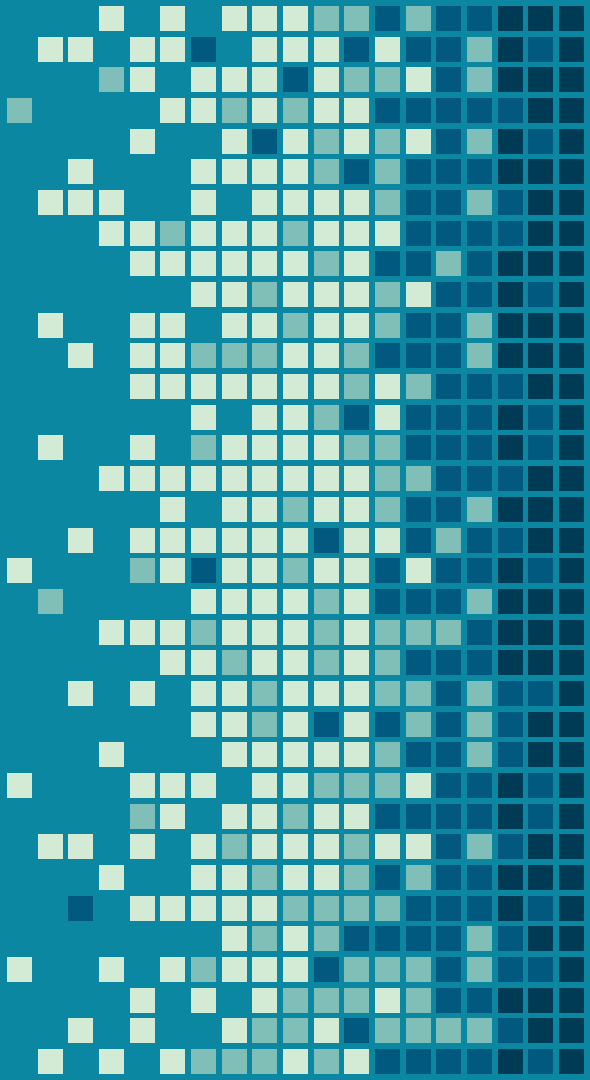
Added to milk to improve appearance

DASHBOARDS

Embedded in web applications

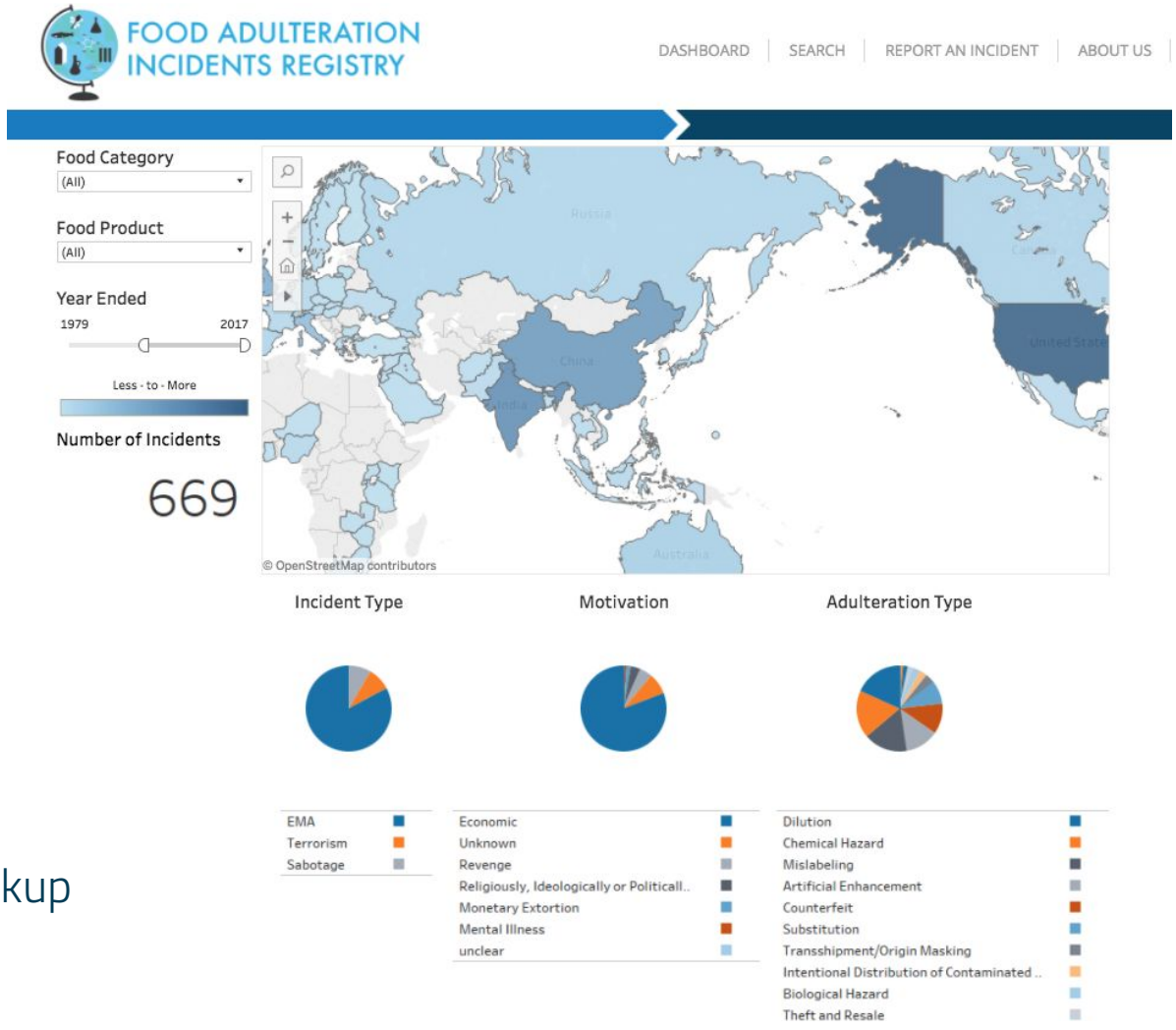
“Dashboards increase the analytical power of your visualization by showing multiple perspectives in the same location

- Tableau



DEMO

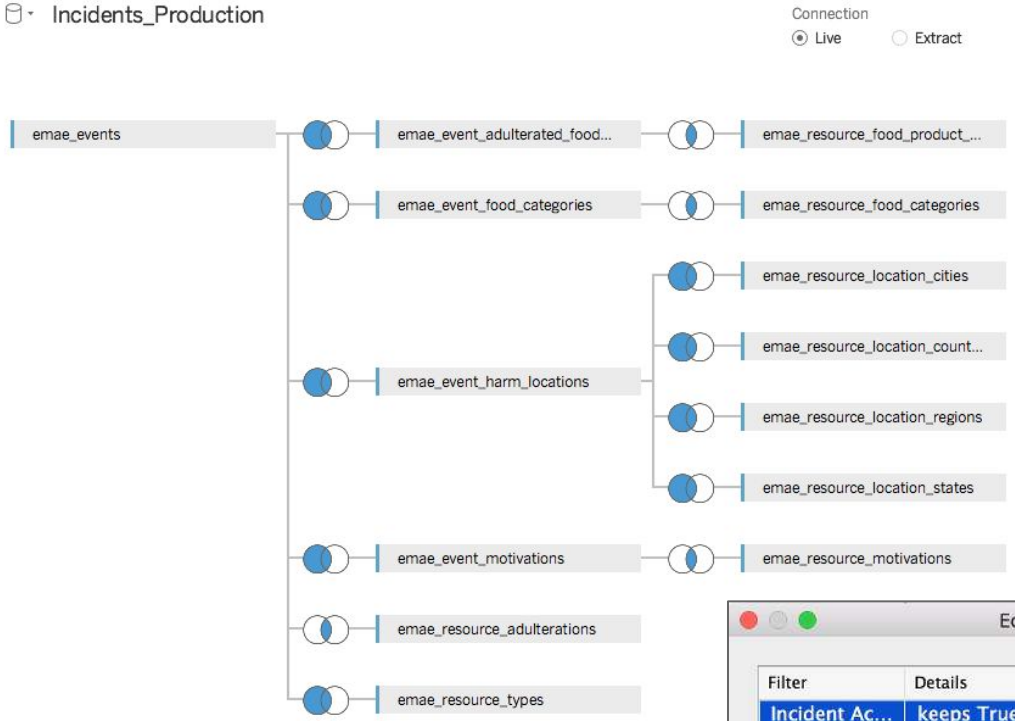
Just changed passwords …

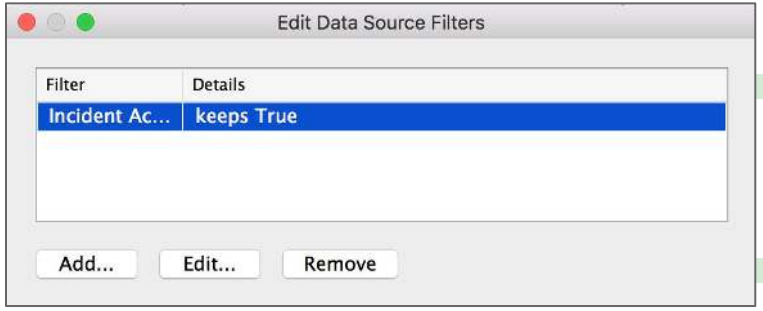

FOOD ADULTERATION INCIDENTS REGISTRY
DASHBOARD
SEARCH
REPORT AN INCIDENT
ABOUT US
Food Category
(All)
Food Product
(All)
Year Ended
1979
2017
Less - to - More
Number of Incidents
669
© OpenStreetMap contributors
Incident Type
Motivation
Adulteration Type
EMA
Terrorism
Sabotage
Economic
Unknown
Revenge
Religiously, Ideologically or Politicall..
Monetary Extortion
Mental Illness
unclear
Dilution
Chemical Hazard
Mislabeling
Artificial Enhancement
Counterfeit
Substitution
Transshipment/Origin Masking
Intentional Distribution of Contaminated ..
Biological Hazard
Theft and Resale

... Backup

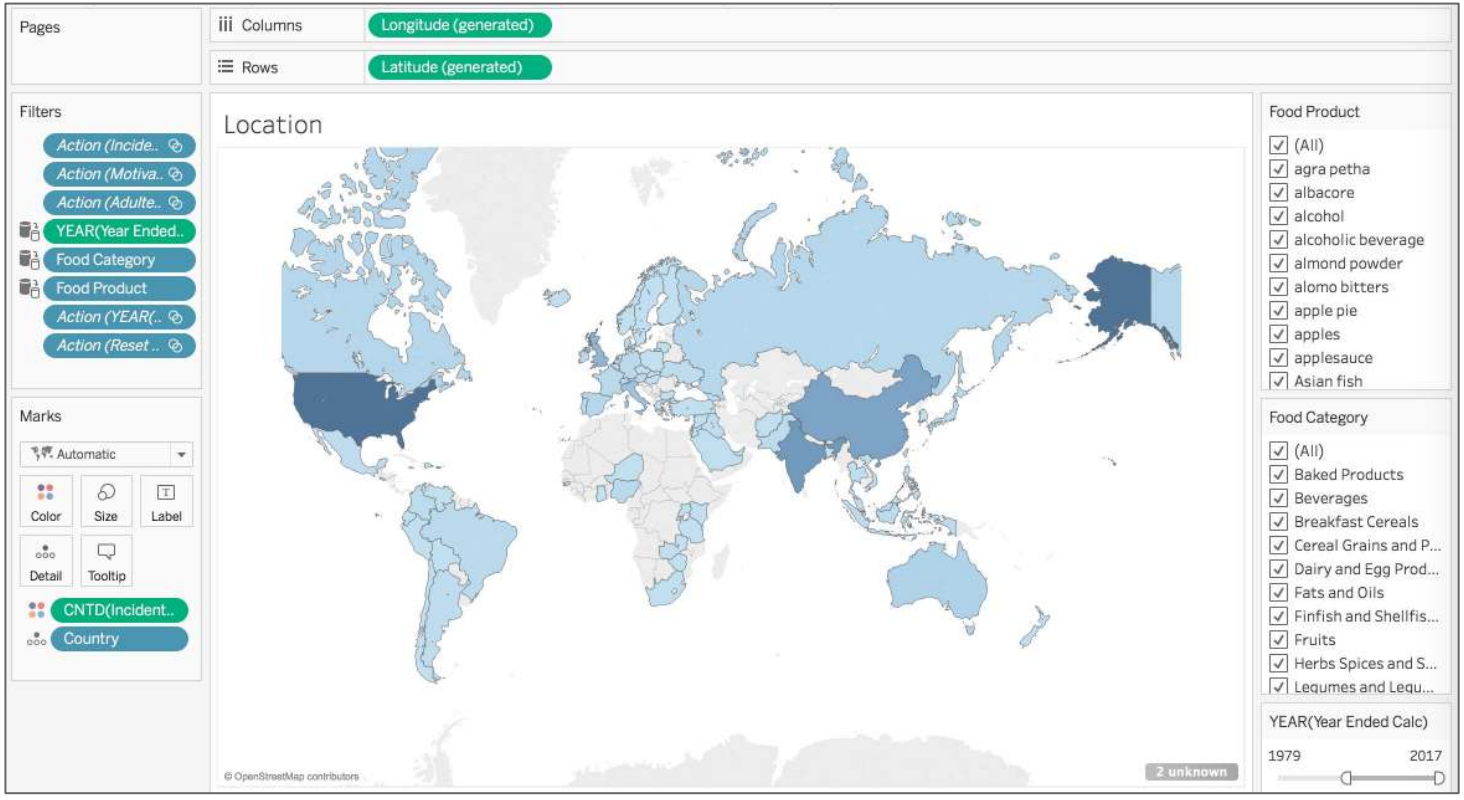
HOW IS IT BUILT?

Agile: Google it, implement, feedback, repeat

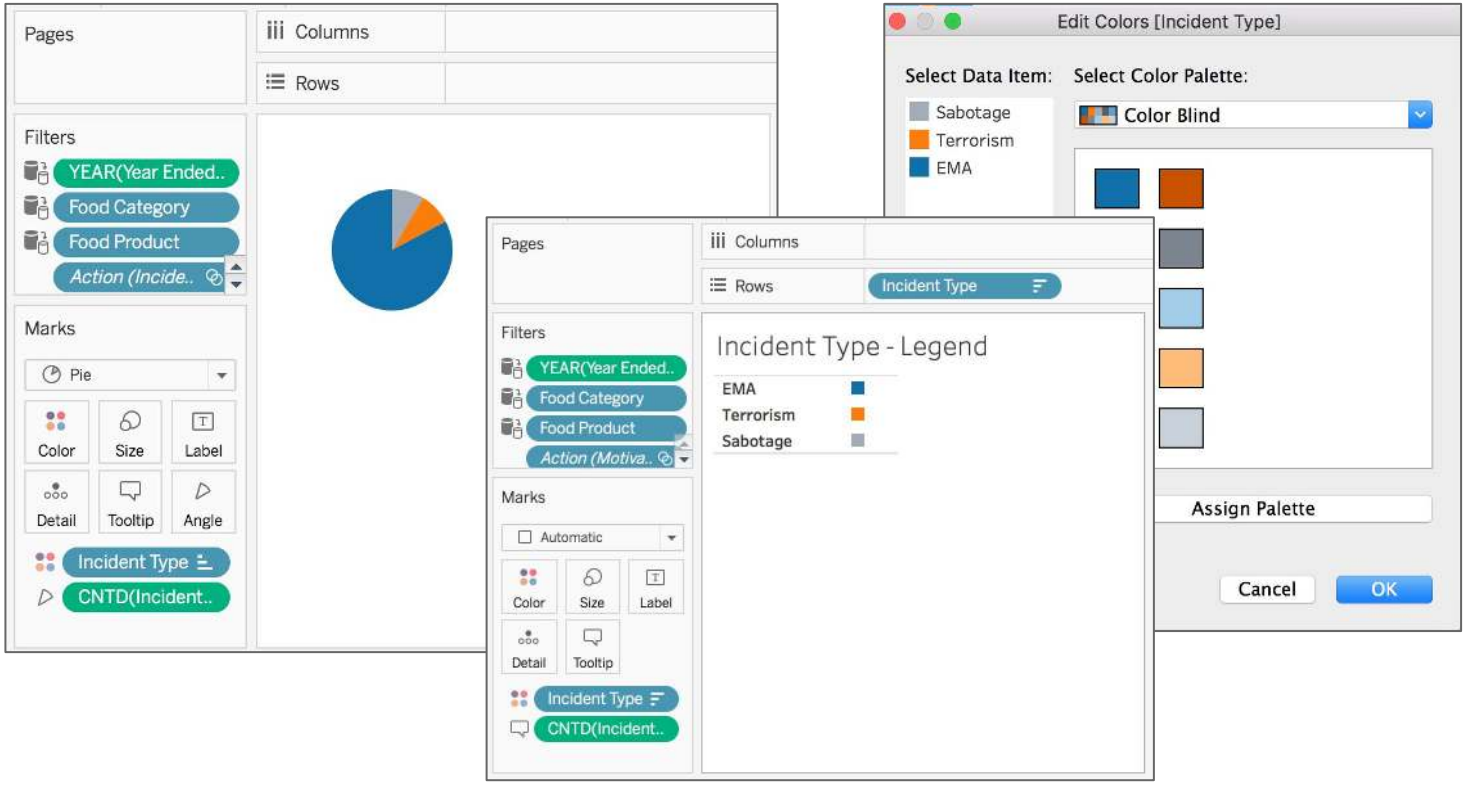




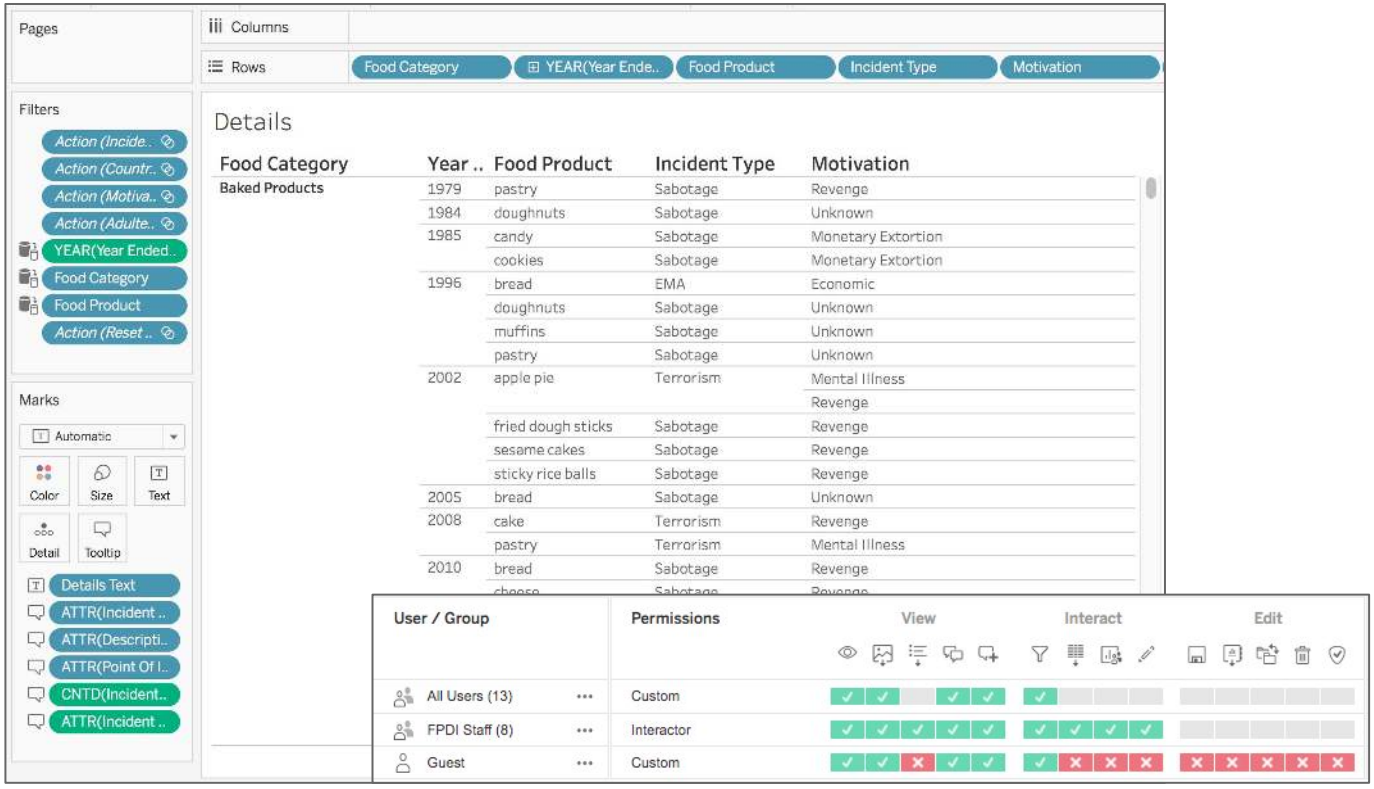
Relational Database Connection



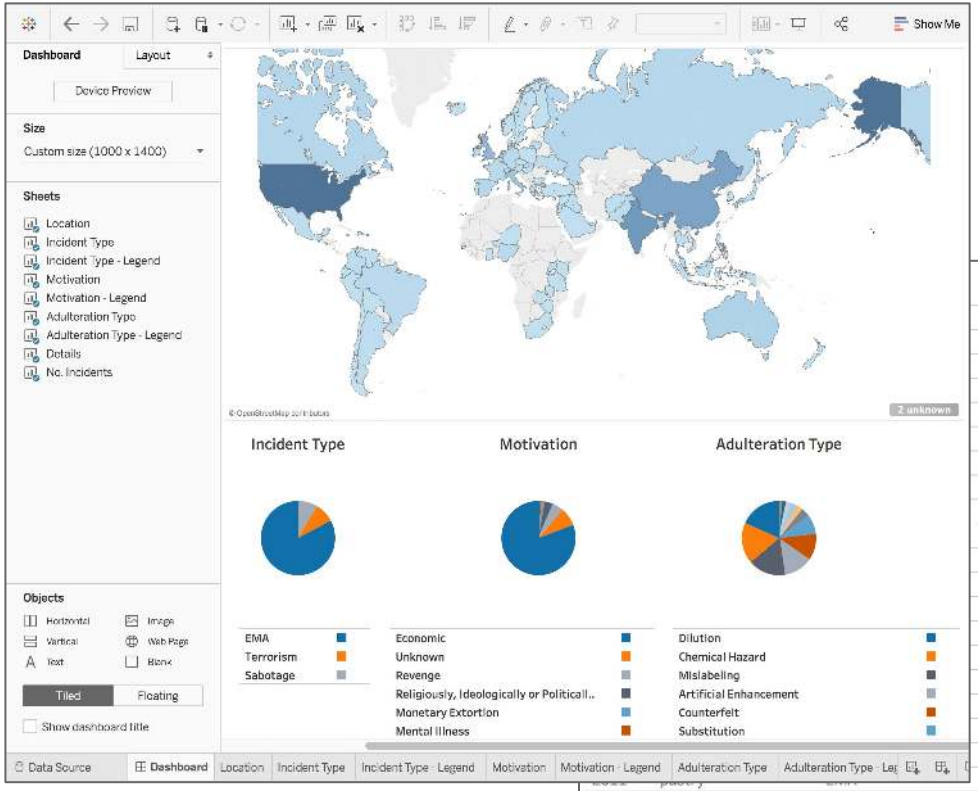
Map of Incidents for global incidents



Individual worksheets for visual appeal & interaction

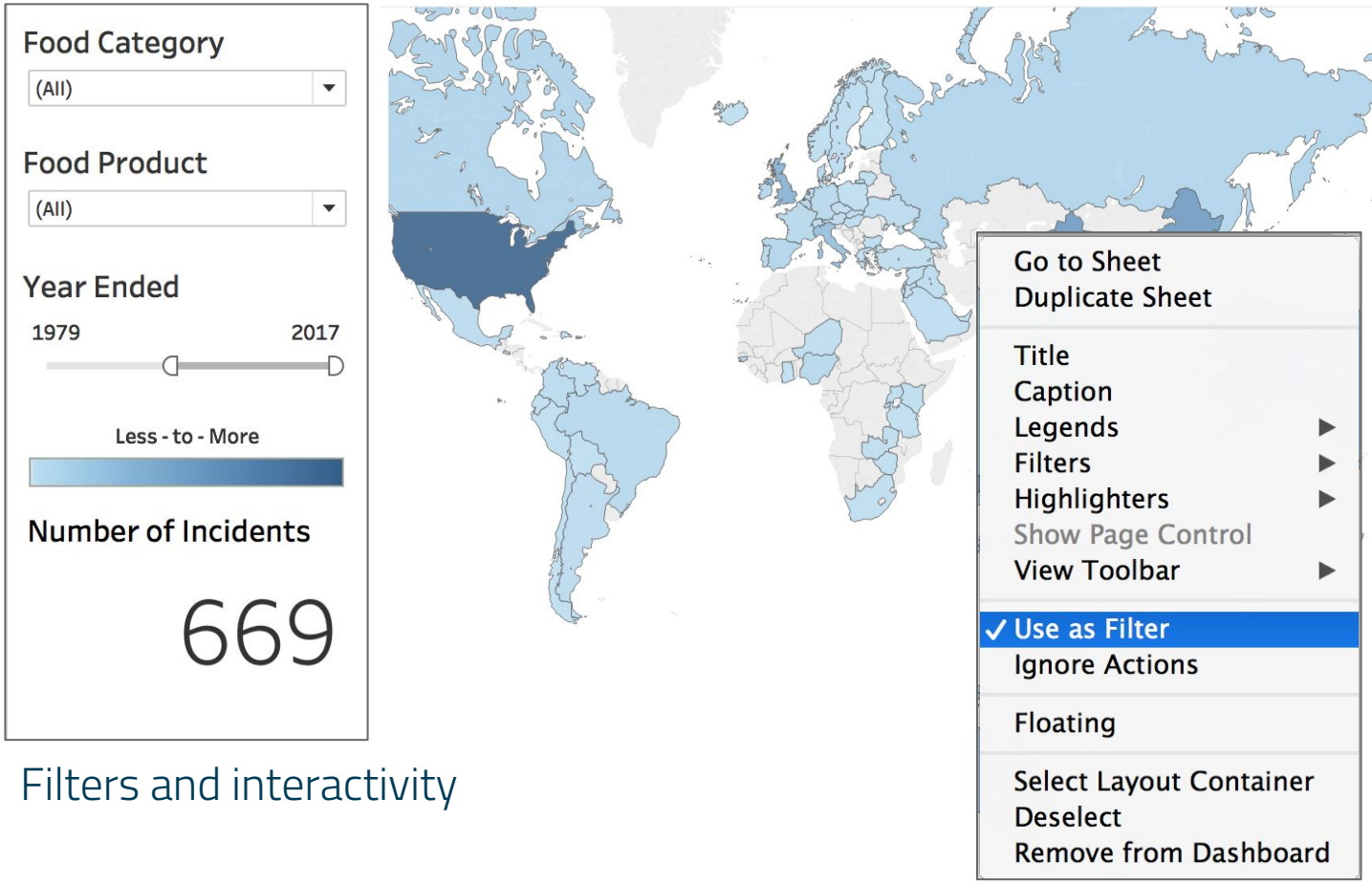


Details Screen for restricted interaction

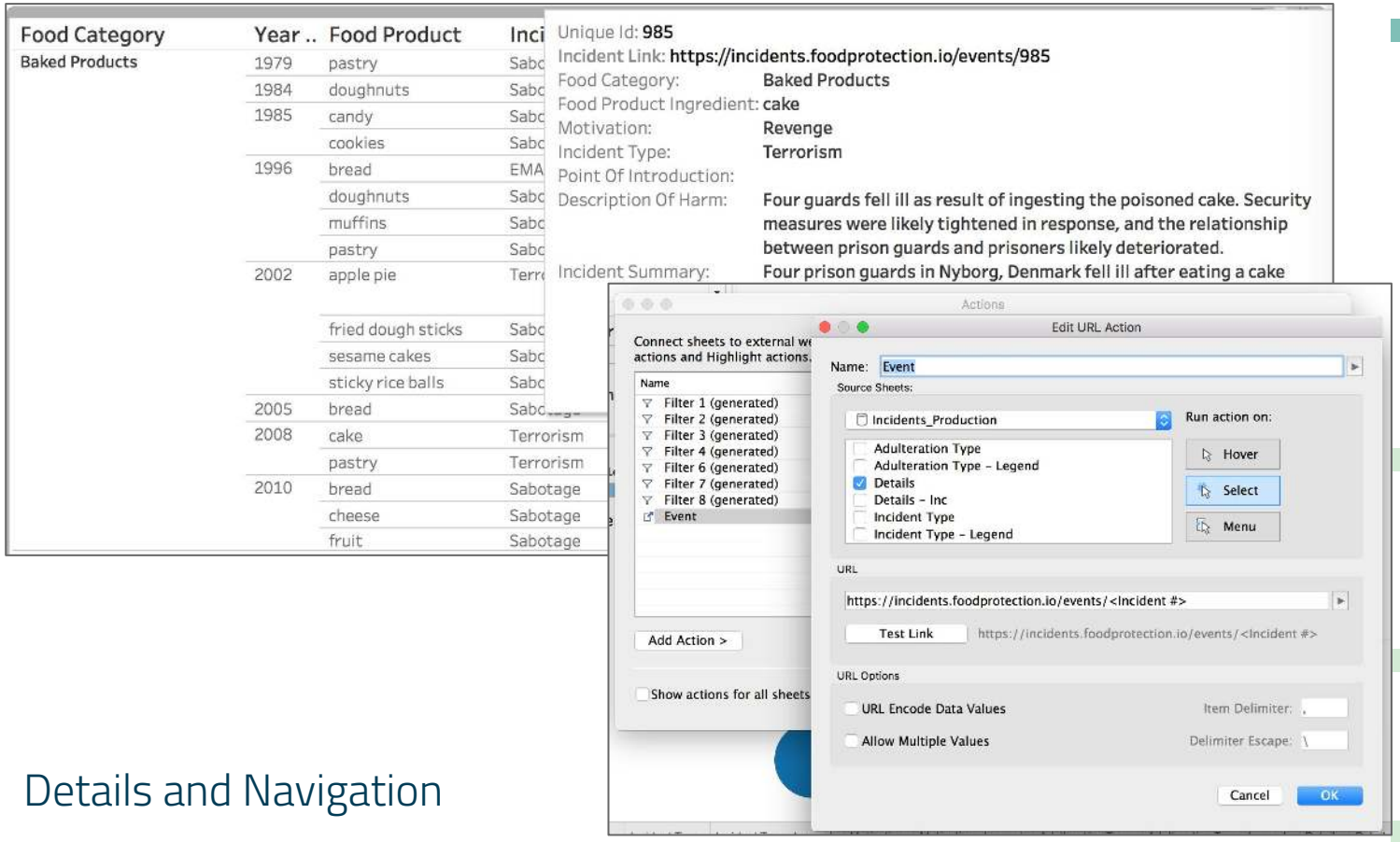


Motivation	Incident #	
Revenge	790	Details
Unknown	777	Details
Monetary Extortion	775	Details
Monetary Extortion	775	Details
Economic	967	Details
Unknown	782	Details
Unknown	782	Details
Unknown	782	Details
Mental Illness	952	Details
Revenge	952	Details
Revenge	851	Details
Revenge	851	Details
Revenge	851	Details
Unknown	988	Details
Revenge	985	Details
Mental Illness	1021	Details
Revenge	954	Details
Revenge	954	Details
Revenge	954	Details
Economic	38	Details
2013	bread	Terrorism	Unknown	813	Details
2014	bread	EMA	Economic	527	Details
	chicken nuggets	Sabotage	Revenge	756	Details
	pancakes	Sabotage	Revenge	756	Details
	pizza	Sabotage	Revenge	756	Details

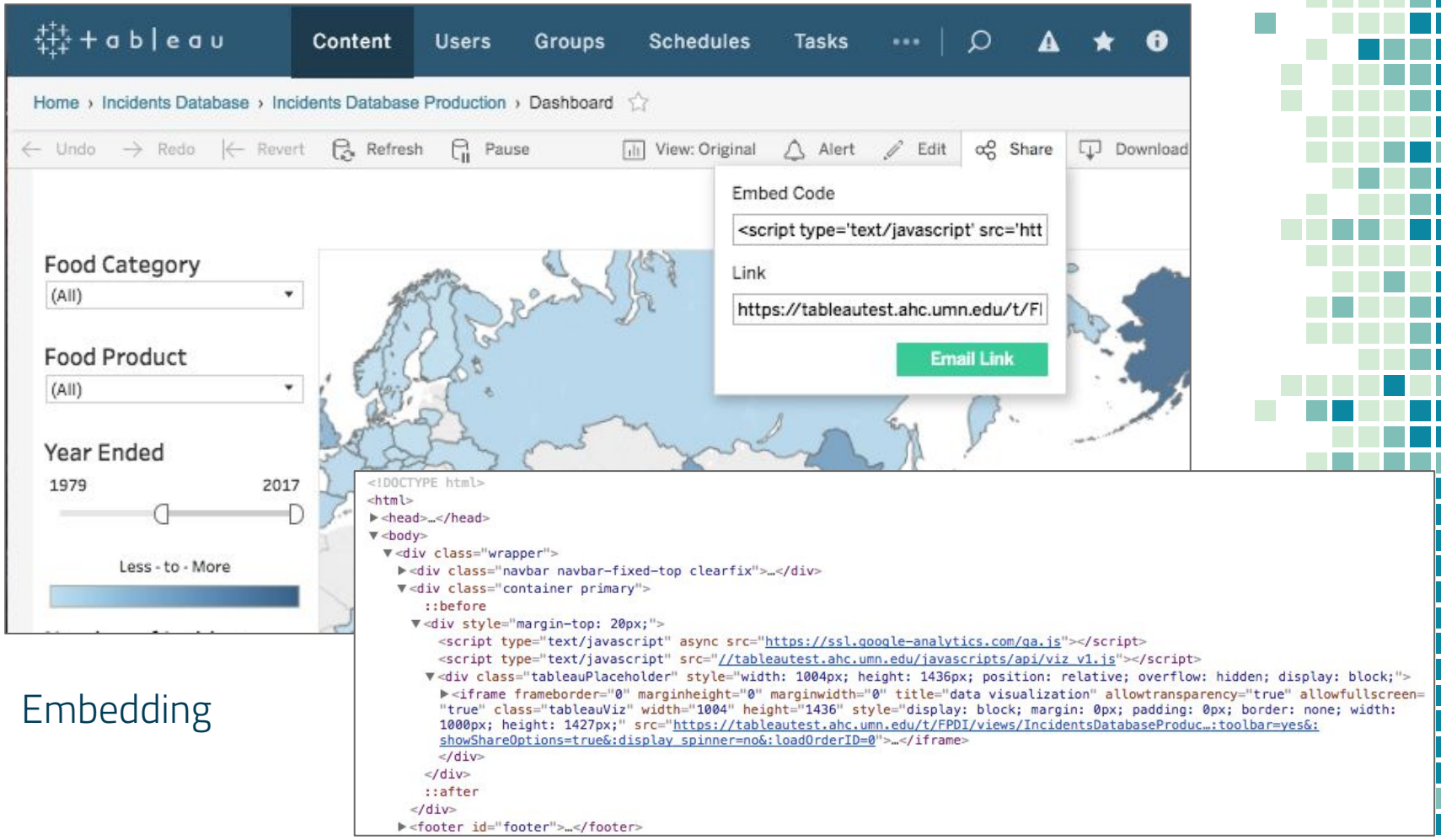
That's a lot of sheet(s)



Filters and interactivity



Details and Navigation


+ableau
Content
Users
Groups
Schedules
Tasks
Home › Incidents Database › Incidents Database Production › Dashboard
Undo
Redo
Revert
Refresh
Pause
View: Original
Alert
Edit
Share
Download
Embed Code
<script type='text/javascript' src='htt
Link
https://tableautest.ahc.umn.edu/t/FI
Email Link
Food Category
(All)
Food Product
(All)
Year Ended
1979
2017
Less - to - More
<!DOCTYPE html>
<html>
<head>…</head>
<body>

…

::before

<script type="text/javascript" async src="https://ssl.google-analytics.com/ga.js"></script>
<script type="text/javascript" src="//tableautest.ahc.umn.edu/javascripts/api/viz_v1.js"></script>

<iframe frameborder="0" marginheight="0" marginwidth="0" title="data visualization" allowtransparency="true" allowfullscreen="true" class="tableauViz" width="1004" height="1436" style="display: block; margin: 0px; padding: 0px; border: none; width: 1000px; height: 1427px;" src="https://tableautest.ahc.umn.edu/t/FPDI/views/IncidentsDatabaseProduc…:toolbar=yes&:showShareOptions=true&:display_spinner=no&:loadOrderID=0">…</iframe>

::after

<footer id="footer">…</footer>

Embedding

LESSONS LEARNED

Data Source Changes

Are rather difficult. Once published, you can't make changes to a data source without downloading, updating overwriting, etc.

Embedding Limitations

Application based authentication isn't built in. iFrame generates Cross Site Forgery warning with certs. Proxy??

Unique Record Counts

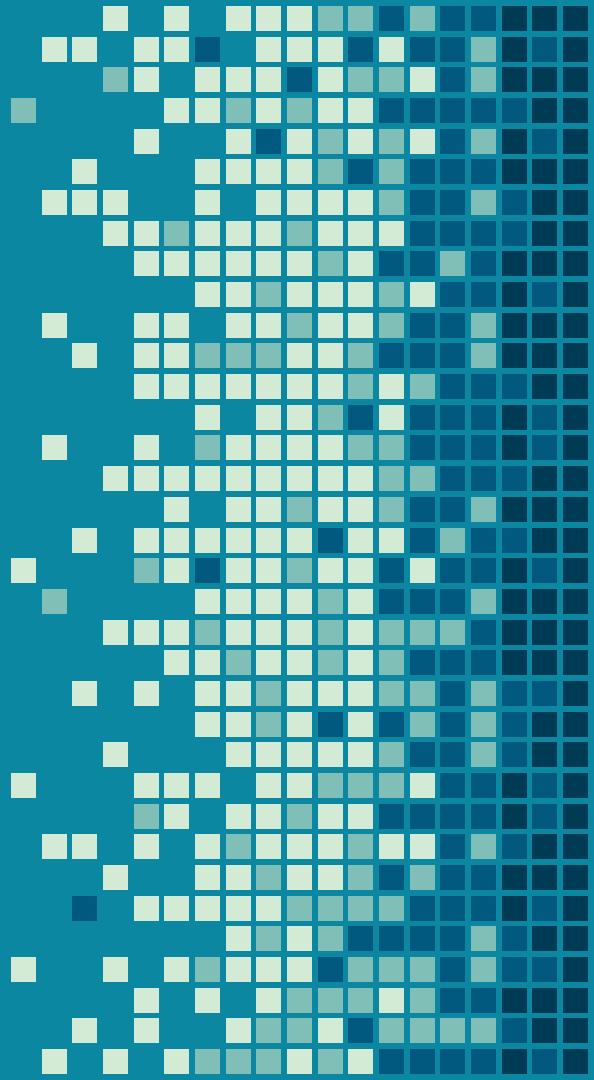
Getting the right incident record counts across all worksheets proved to be more difficult than expected.

DATA LAKE

large-scale repo & processing engine

> *An architectural approach that allows you to store massive amounts of data in a central location, so it's readily available to be categorized, processed, analyzed and consumed by diverse groups within an organization.*

\- *Amazon*



WHAT SOLUTIONS AND WHY?

Tableau

Market leader with the highest rated ability to execute for business intelligence and analytics platforms.

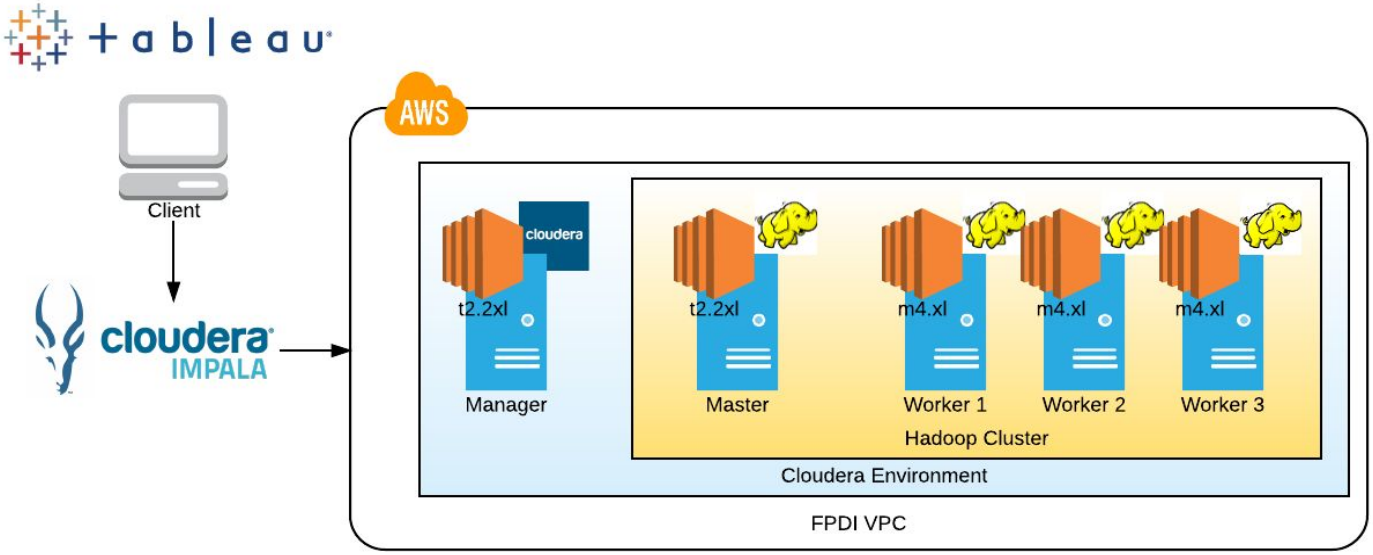
Cloudera

Strong market presence, cloud technical support. **Components** for data governance, cluster administration, and SQL access.

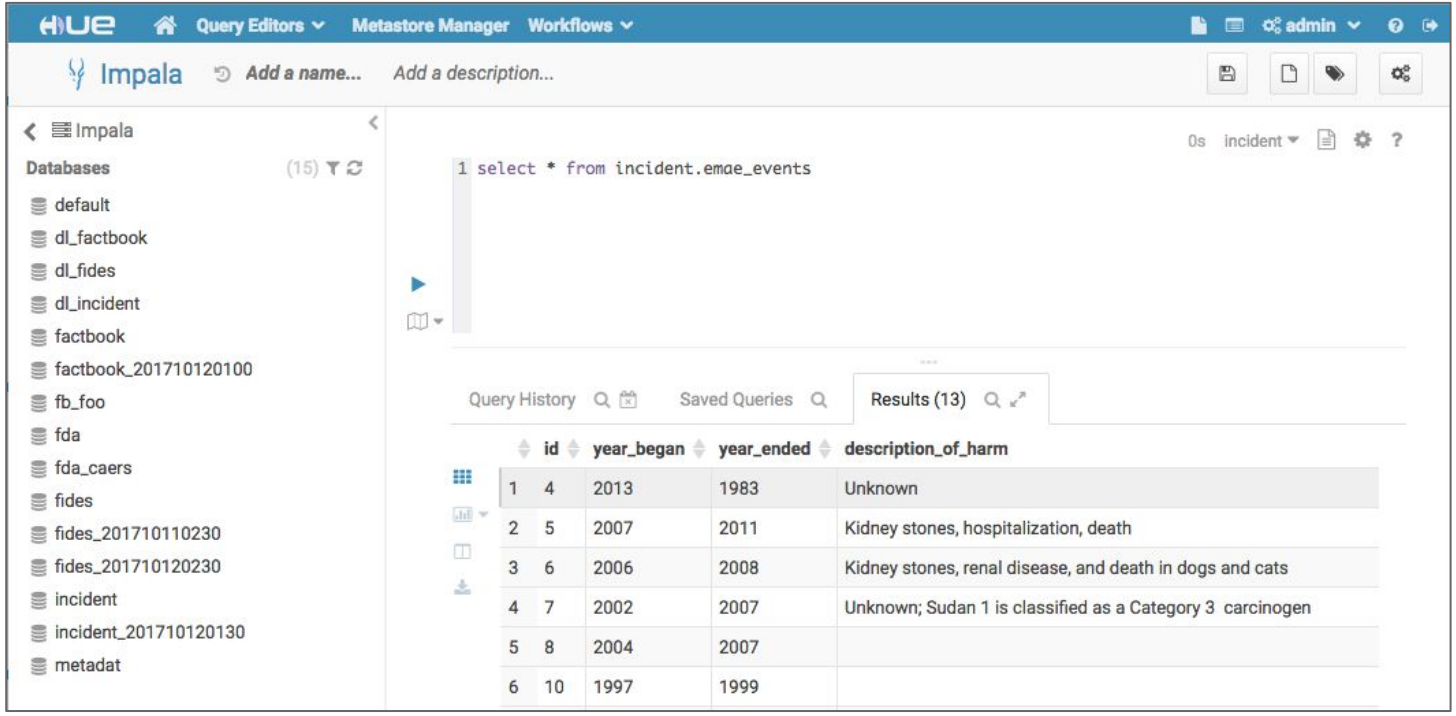
Amazon Web Services

Cloud for **elasticity** and virtual vs. physical hardware acquisition. AWS is market leader in execution and vision.

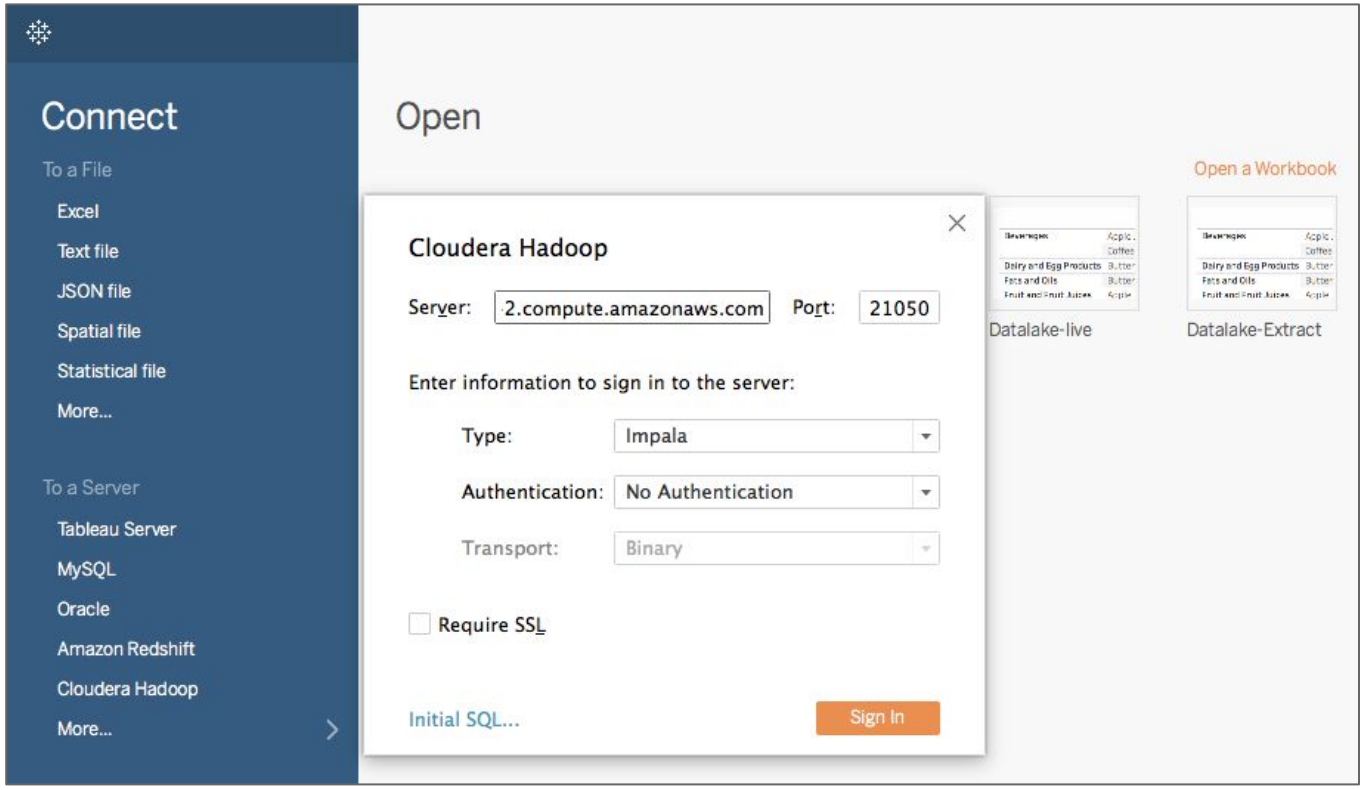
– *Gartner 2017*



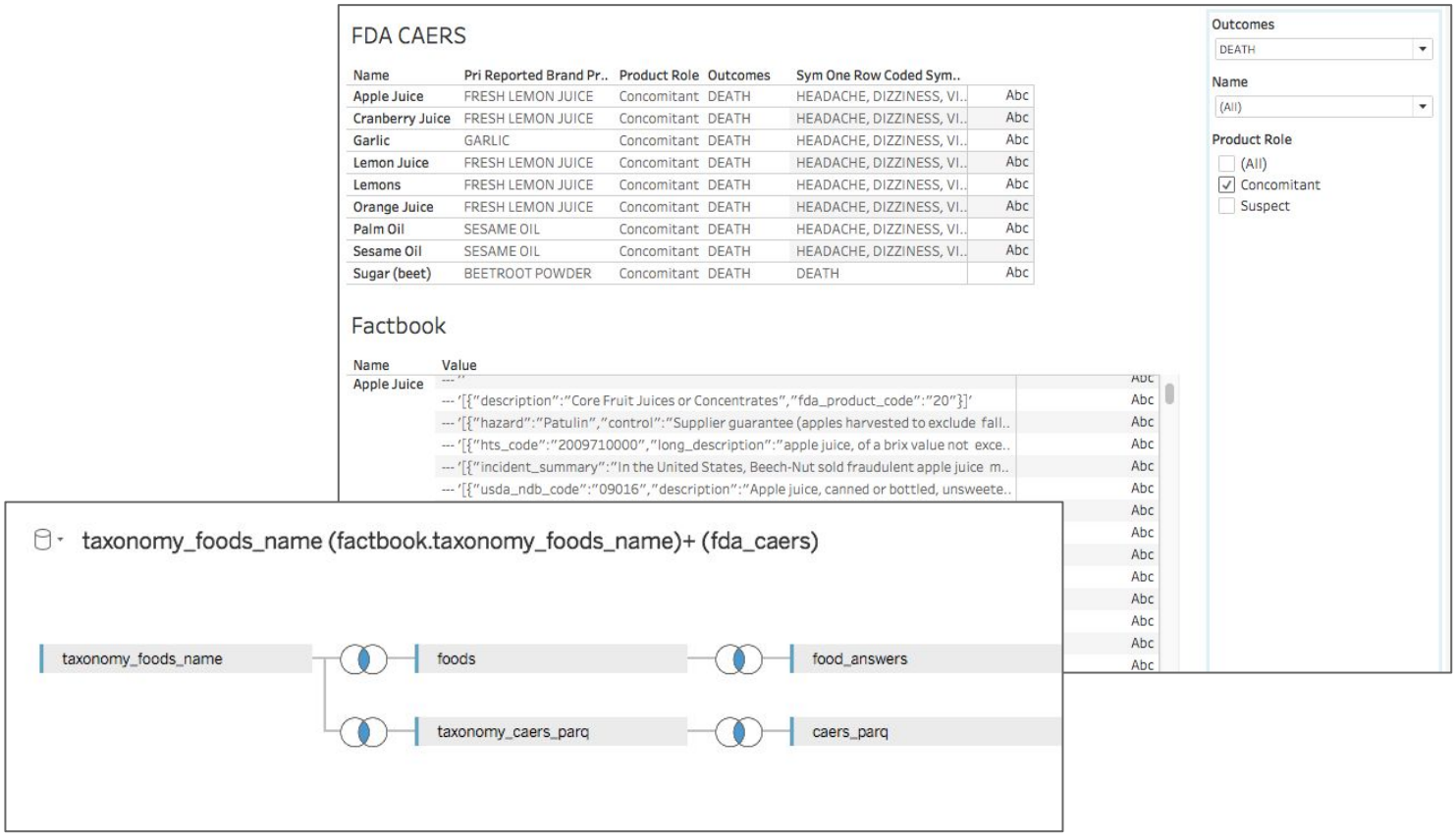
How does it work?



Impala SQL Interface



Impala ODBC Connector



Joining on Common Taxonomy

WHATS NEXT?

Data Visualizations

More dashboards using the data lake as the source of data.

More Data

Upload, categorize and generally make the data available.

Governance

Need to figure out how to reliably provide access to sponsors.

THANKS!

Any questions?

You can find me at:

tsteinma@umn.edu

References

[1] A. (FPDI) Kircher, “Data Lake and Interface,” *Homeland Security University Programs*, 2016. [Online]. Available: http://projects.hsuniversityprograms.org/search/1628/print.pdf.

[2] “phData: Data Lake Consultation.” [Online]. Available: https://www.phdata.io/solutions/project-delivery/. [Accessed: 13-Oct-2017].

[3] “Cloudera ODBC Driver for Apache Hive.”

[4] “Cloudera Impala.” [Online]. Available: https://www.cloudera.com/products/open-source/apache-hadoop/impala.html. [Accessed: 12-Oct-2017].

[5] N. Heudecker, M. Adrian, and A. Jain, “Market Guide for Hadoop Distributions,” *Gart. Res.*, no. February, pp. 1–11, 2017.

[6] “What is a Data Lake?,” *Amazon*, 2017. [Online]. Available: https://aws.amazon.com/big-data/data-lake-on-aws/.

[7] “Apache Impala.” [Online]. Available: https://impala.apache.org/.

[8] Bigdata, “How-to Implement Role-based Security in Impala using Apache Sentry,” *Big Data Analytics News*, 2014. [Online]. Available: http://bigdataanalyticsnews.com/implement-role-based-security-impala-using-apache-sentry/.

References cont...

[9] Tableau Software, “Building a Dashboard.”

[10] FPDI, “Food Adulteration Incidents Registry Dashboard.” [Online]. Available: https://incidents.foodprotection.io/dashboard.

[11] Tableau Software, “Visual Analysis Best Practices: Simple Techniques for Making Every Data Visualization Useful and Beautiful,” 2014. [Online]. Available: http://www.tableausoftware.com/learn/whitepapers/tableau-visual-guidebook.

[12] Tableau Software, “Dashboard Layouts and Formatting.”

[13] Cloudera, “Apache Impala (incubating): Analytic Database for Apache Hadoop,” 2016.

[14] Tableau Software, “Dashboard Interactivity using Actions.” [Online]. Available: https://www.tableau.com/learn/tutorials/on-demand/dashboard-interactivity-using-actions.

[15] “Food Protection and Defense Institute.” [Online]. Available: https://foodprotection.umn.edu/.

[16] Tabaleau Software, “Getting Started with Dashboards and Stories,” 2017. [Online]. Available: https://www.tableau.com/learn/tutorials/on-demand/getting-started-dashboards-and-stories.

[17] S. Sicular, “Three Architecture Styles for a Useful Data Lake,” *Gart. Res.*, no. July, 2016.

[18] Gartner, “Magic Quadrant for Cloud Infrastructure as a Market Definition / Description,” *Gart. Res.*, no. May, pp. 1–26, 2014.

References cont...

[19] J. Parenteau, R. L. Sallam, C. Howson, J. Tapadinhas, K. Schlegel, and T. W. Oestreich, “Magic Quadrant for Business Intelligence and Analytics Platforms,” *Gart. Res.*, vol. 1, no. February, pp. 1–78, 2016.