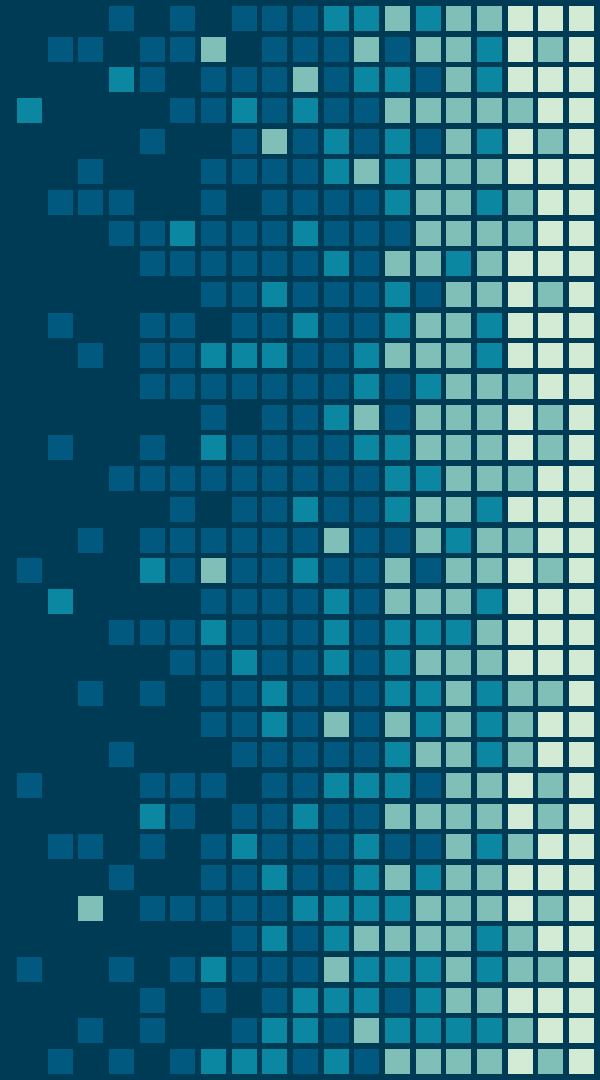


# DASHBOARDS AND DATA LAKES WITH TABLEAU



# BACKGROUND

## Food Protection and Defense Institute (FPDI)

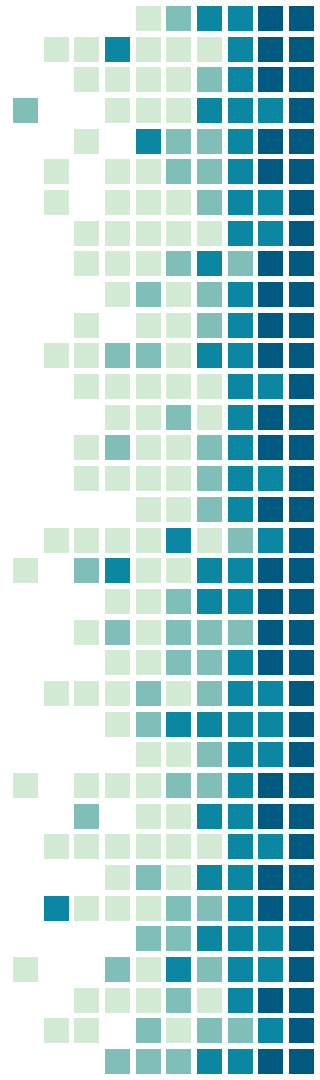
Research Institution located at the University of Minnesota. Started by Department of Homeland Security to look at threats of **"Intentional contamination"** of the Food Supply. FPDI is a part of the Academic Health Center (AHC).

## 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

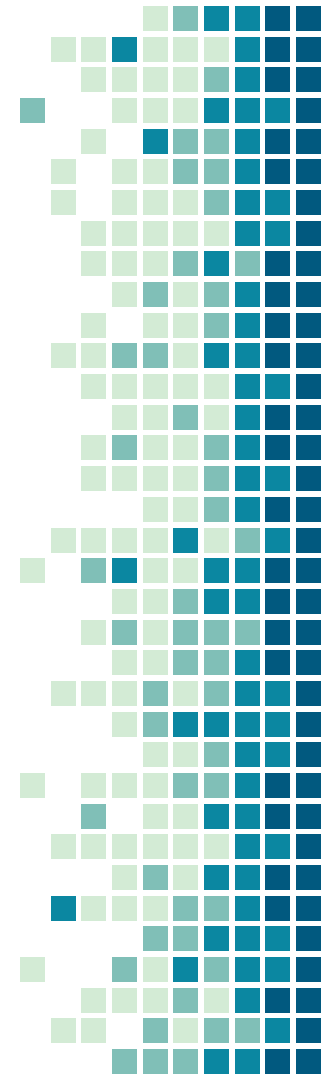
## FOOD SECURITY

"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

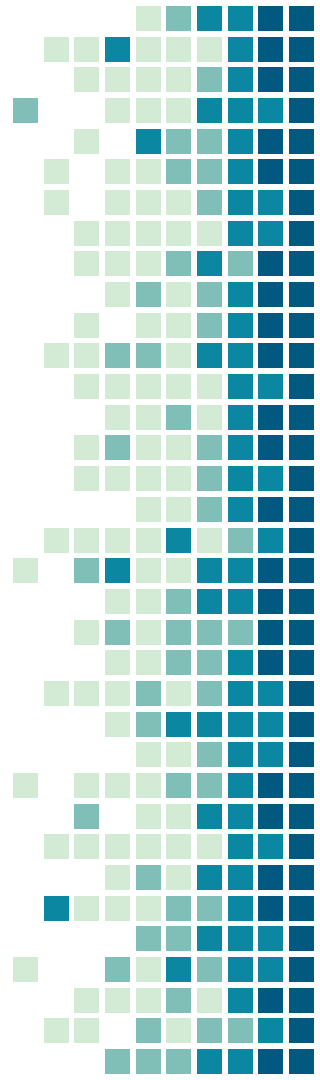
Reducing the impact of **system attacks** by assuring the stability or continuity of the food supply.



# 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.





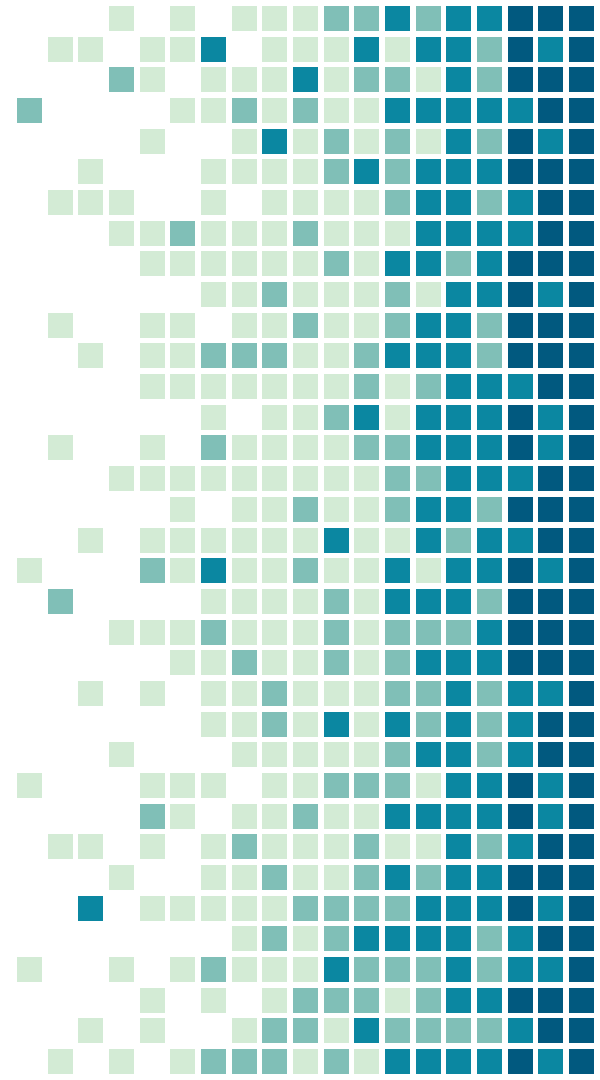
# MELAMINE IN MILK

Used for making whiteboards

Added to milk to improve appearance

# DASHBOARDS

Embedded in web applications

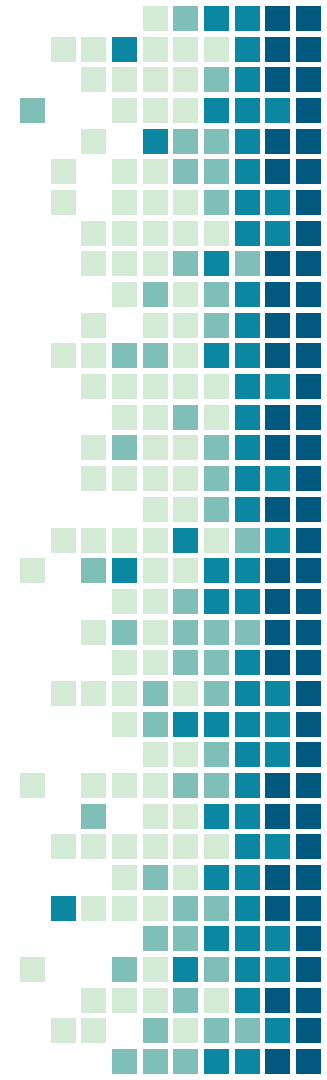


“ *Dashboards increase the analytical power of your visualization by showing multiple perspectives in the same location*

- *Tableau*

# DEMO

Just changed passwords ...







# FOOD ADULTERATION INCIDENTS REGISTRY

[DASHBOARD](#)[SEARCH](#)[REPORT AN INCIDENT](#)[ABOUT US](#)

Food Category

(All)

Food Product

(All)

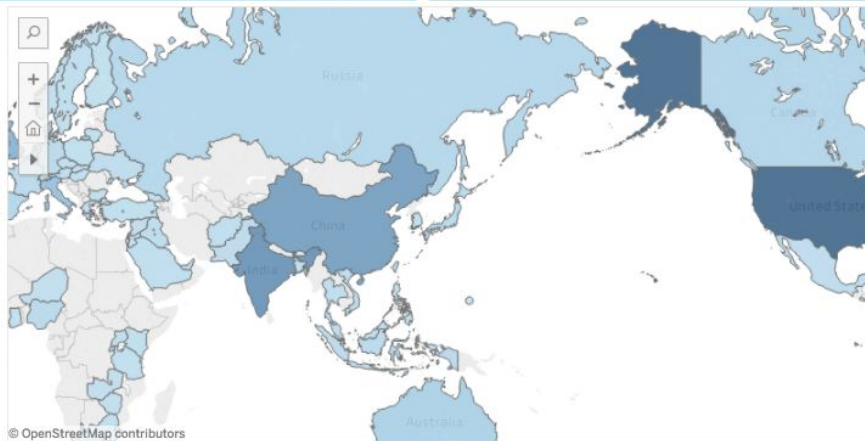
Year Ended

1979 2017

Less - to - More

Number of Incidents

669



Incident Type



EMA  
Terrorism  
Sabotage

Motivation



Economic  
Unknown  
Revenge  
Religiously, Ideologically or Politicall..  
Monetary Extortion  
Mental Illness  
unclear

Adulteration Type

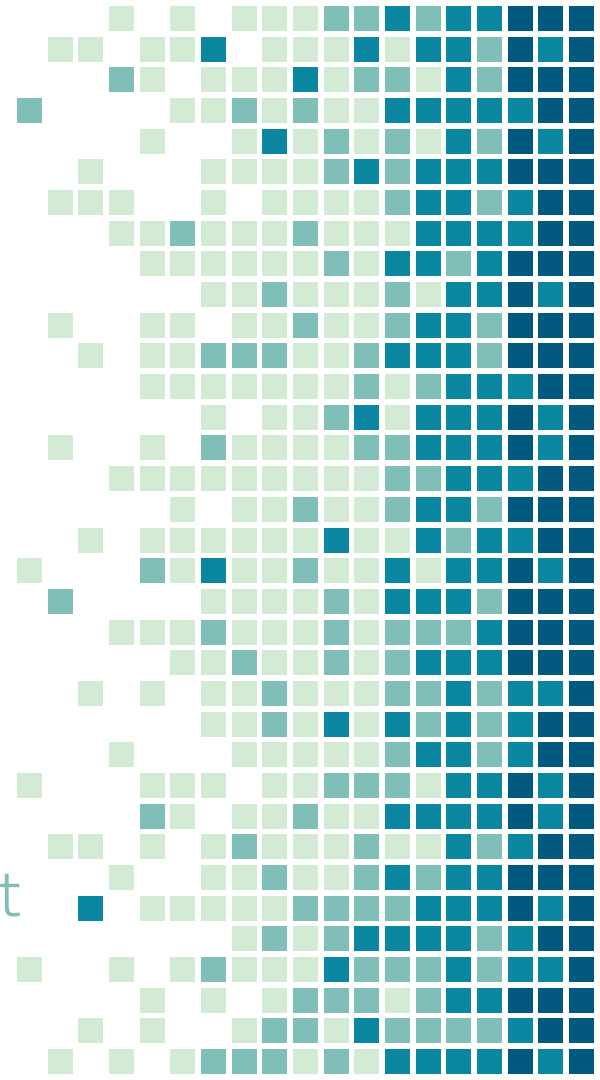


Dilution  
Chemical Hazard  
Mislabeling  
Artificial Enhancement  
Counterfeit  
Substitution  
Transshipment/Origin Masking  
Intentional Distribution of Contaminated ...  
Biological Hazard  
Theft and Resale

... Backup

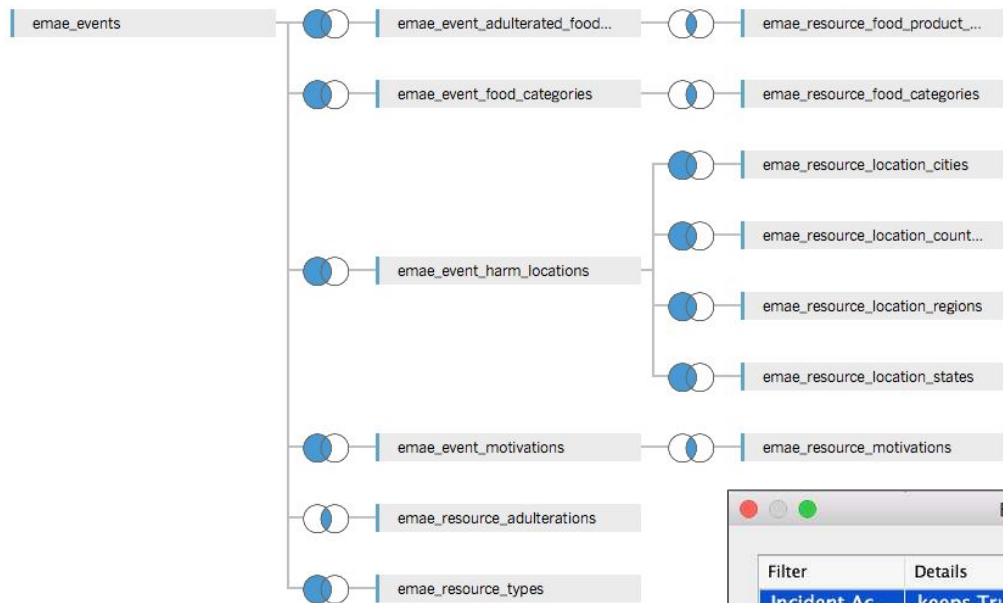
# HOW IS IT BUILT?

Agile: Google it, implement, feedback, repeat



Incidents\_Production

Connection  
☒ Live ☐ Extract

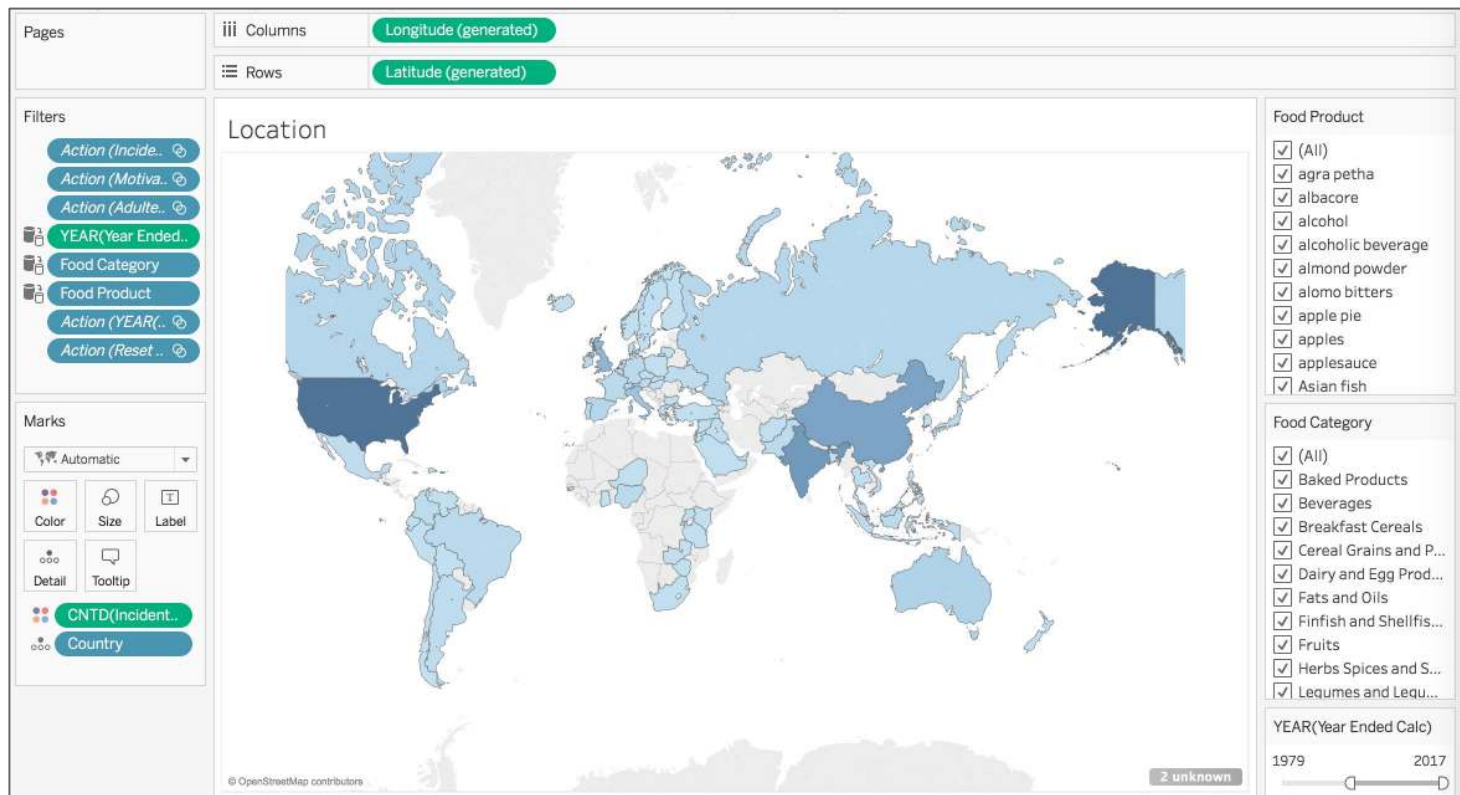


Edit Data Source Filters

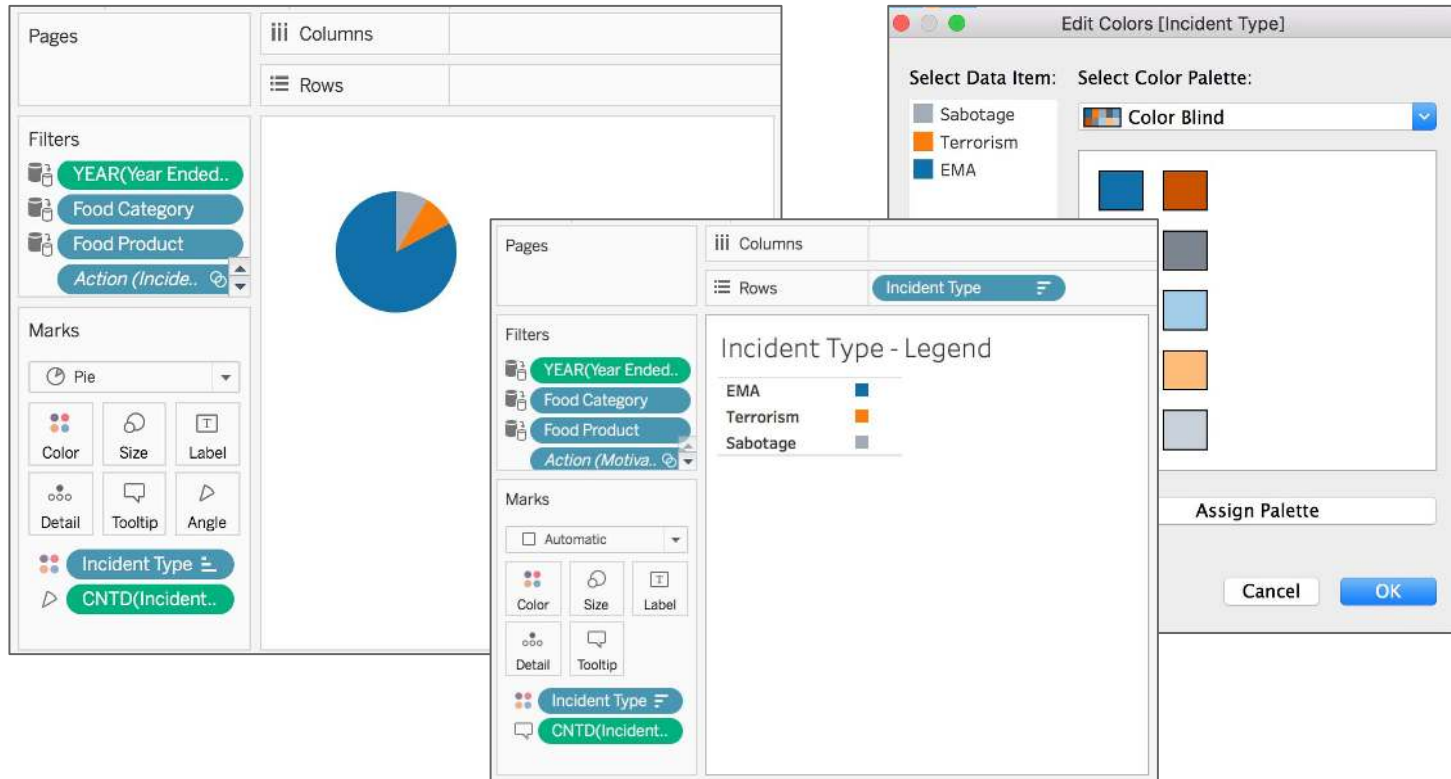
Filter	Details
Incident Ac...	keeps True

Add... Edit... Remove

Relational Database Connection



Map of Incidents for global incidents



Individual worksheets for visual appeal & interaction

Pages

Filters

Action (Incide..)

Action (Countr..)

Action (Motiva..)

Action (Adulte..)

YEAR(Year Ended..)

Food Category

Food Product

Action (Reset..)

Marks

Automatic

Color

Size

Text

Detail

Tooltip

Details Text

ATTR(Incident ..)

ATTR(Descripti..)

ATTR(Point Of I..)

CNTD(Incident..)

ATTR(Incident ..)

Columns

Rows

Food Category

YEAR(Year Ende..)

Food Product

Incident Type

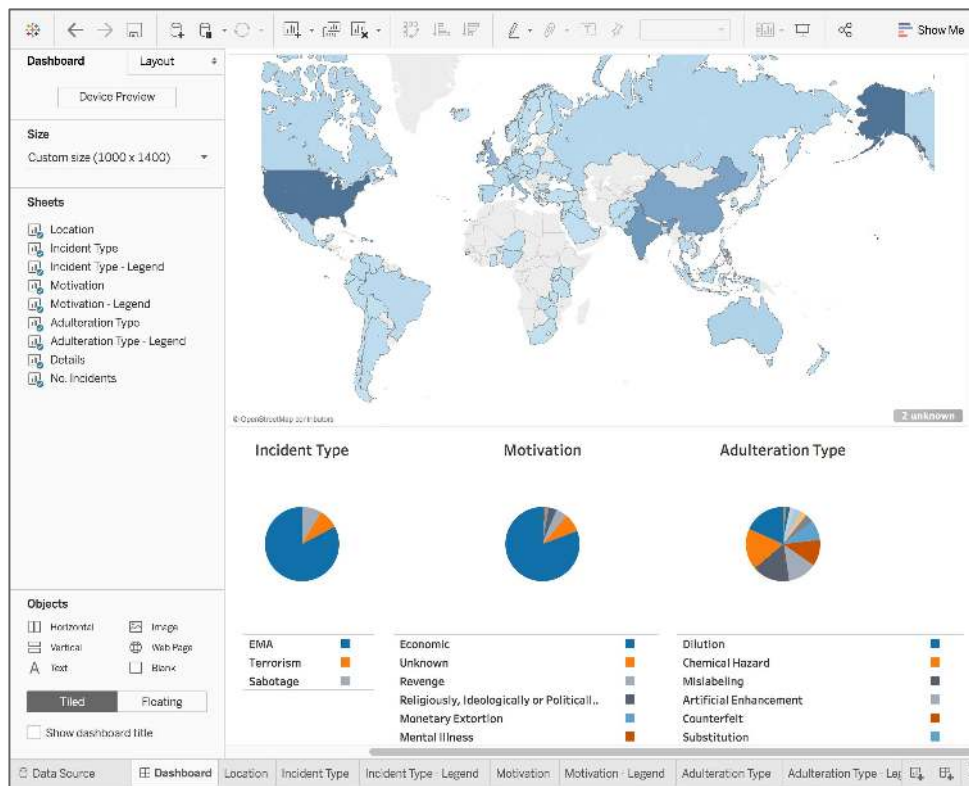
Motivation

Details

Food Category	Year ..	Food Product	Incident Type	Motivation
Baked Products	1979	pastry	Sabotage	Revenge
	1984	doughnuts	Sabotage	Unknown
	1985	candy	Sabotage	Monetary Extortion
		cookies	Sabotage	Monetary Extortion
	1996	bread	EMA	Economic
		doughnuts	Sabotage	Unknown
		muffins	Sabotage	Unknown
		pastry	Sabotage	Unknown
	2002	apple pie	Terrorism	Mental Illness
				Revenge
		fried dough sticks	Sabotage	Revenge
		sesame cakes	Sabotage	Revenge
		sticky rice balls	Sabotage	Revenge
	2005	bread	Sabotage	Unknown
	2008	cake	Terrorism	Revenge
		pastry	Terrorism	Mental Illness
	2010	bread	Sabotage	Revenge
		cheese	Sabotage	Revenge

User / Group	Permissions	View	Interact	Edit	
All Users (13)	Custom				
FPDI Staff (8)	Interactor				
Guest	Custom				

## Details Screen for restricted interaction



That's a lot of sheet(s)

Year	Item	Motivation	Incident #	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

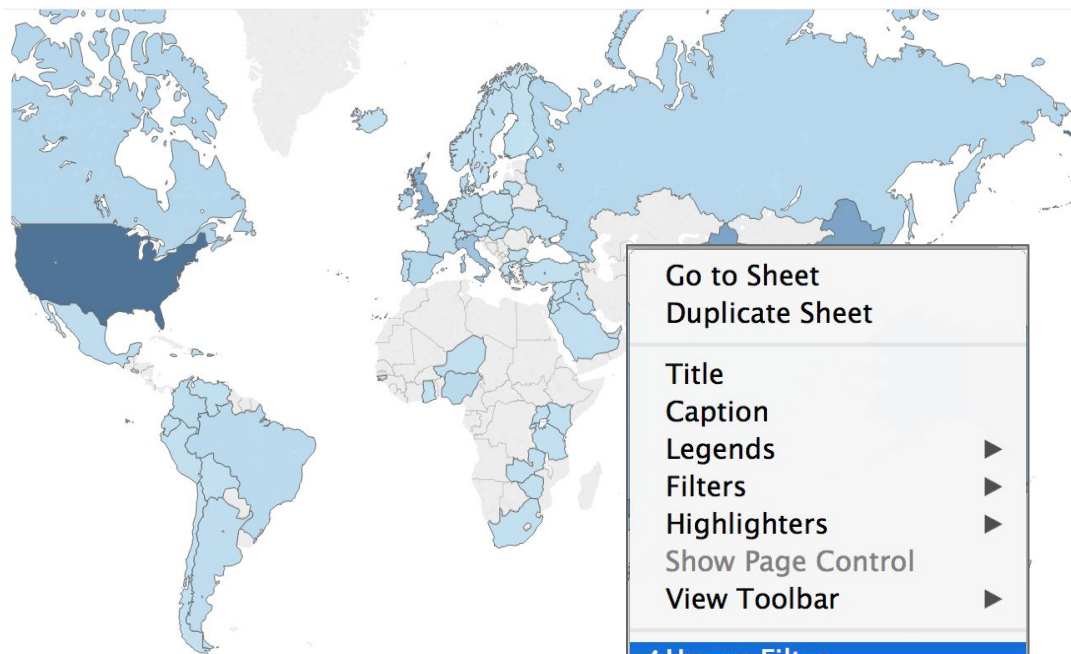
Motivation	Incident #	Details
Revenge	790	Details
Unknown	777	Details
Monetary Extortion	775	Details
Monetary Extortion	775	Details
Economic	967	Details
Unknown	782	Details
Unknown	782	Details
Mental Illness	952	Details
Revenge	952	Details
Revenge	851	Details
Revenge	851	Details
Revenge	851	Details
Unknown	988	Details
Revenge	985	Details
Mental Illness	1021	Details
Revenge	954	Details
Revenge	954	Details
Revenge	954	Details
Economic	38	Details
Unknown	813	Details
Economic	527	Details
Revenge	756	Details
Revenge	756	Details
Revenge	756	Details

**Food Category**  
(All) ▼

**Food Product**  
(All) ▼

**Year Ended**  
1979 2017  
Less - to - More

**Number of Incidents**  
**669**



Go to Sheet  
Duplicate Sheet

Title  
Caption  
Legends  
Filters  
Highlighters  
Show Page Control  
View Toolbar

✓ Use as Filter  
Ignore Actions

Floating

Select Layout Container  
Deselect  
Remove from Dashboard

Filters and interactivity



Food Category	Year ..	Food Product	Incident
Baked Products	1979	pastry	Sabotage
	1984	doughnuts	Sabotage
	1985	candy	Sabotage
		cookies	Sabotage
	1996	bread	EMA
		doughnuts	Sabotage
		muffins	Sabotage
		pastry	Sabotage
	2002	apple pie	Terrorism
		fried dough sticks	Sabotage
		sesame cakes	Sabotage
		sticky rice balls	Sabotage
	2005	bread	Sabotage
	2008	cake	Terrorism
		pastry	Terrorism
2010	bread	Sabotage	
	cheese	Sabotage	
	fruit	Sabotage	

Unique Id: 985

Incident Link: <https://incidents.foodprotection.io/events/985>

Food Category: Baked Products

Food Product Ingredient: cake

Motivation: Revenge

Incident Type: Terrorism

Point Of Introduction:

Description Of Harm: Four guards fell ill as result of ingesting the poisoned cake. Security measures were likely tightened in response, and the relationship between prison guards and prisoners likely deteriorated.

Incident Summary: Four prison guards in Nyborg, Denmark fell ill after eating a cake

Connect sheets to external web actions and Highlight actions:

Name

- Filter 1 (generated)
- Filter 2 (generated)
- Filter 3 (generated)
- Filter 4 (generated)
- Filter 6 (generated)
- Filter 7 (generated)
- Filter 8 (generated)
- Event

Add Action >

Show actions for all sheets

Actions

Edit URL Action

Name: Event

Source Sheets:

Incidents\_Production

Run action on:

- Adulteration Type
- Adulteration Type - Legend
- ☒ Details
- Details - Inc
- Incident Type
- Incident Type - Legend

URL

<https://incidents.foodprotection.io/events/<Incident #>>

Test Link <https://incidents.foodprotection.io/events/<Incident #>>

URL Options

- ☐ URL Encode Data Values
- ☐ Allow Multiple Values

Item Delimiter: ,

Delimiter Escape: \

Cancel OK

## Details and Navigation

Tableau interface showing the Incidents Database Production Dashboard. The dashboard includes filters for Food Category (All), Food Product (All), and Year Ended (1979 to 2017). The main visualization is a map of the United States. A modal window for embedding the dashboard is open, displaying the Embed Code and Link.

**Embed Code**

```
<script type='text/javascript' src='https://tableaustest.ahc.umn.edu/t/FPI/views/IncidentsDatabaseProduction?embed=1'></script>
```

**Link**

<https://tableaustest.ahc.umn.edu/t/FPI>

**Email Link**

## Embedding

# LESSONS LEARNED

## Data Source Changes

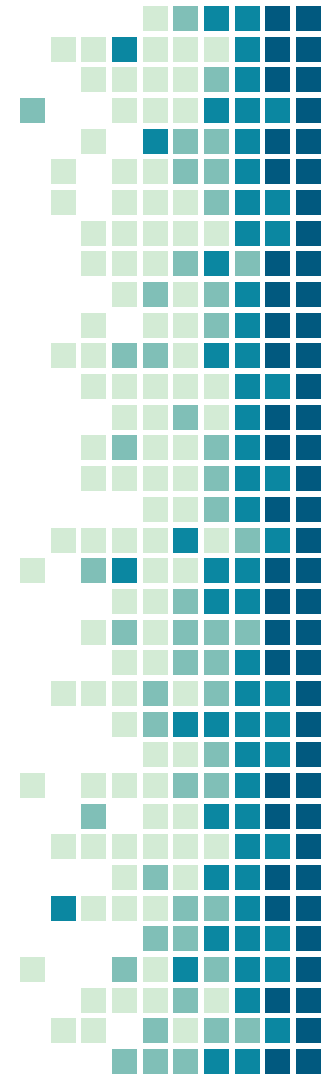
Are rather difficult.  
Once published, you  
can't make changes to  
a data source without  
downloading, updating  
overwriting, etc.

## 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

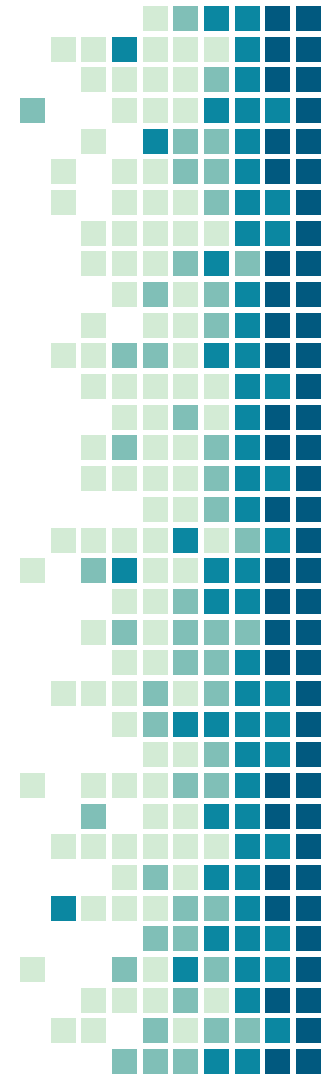
Strong market presence, cloud technical support.

**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*

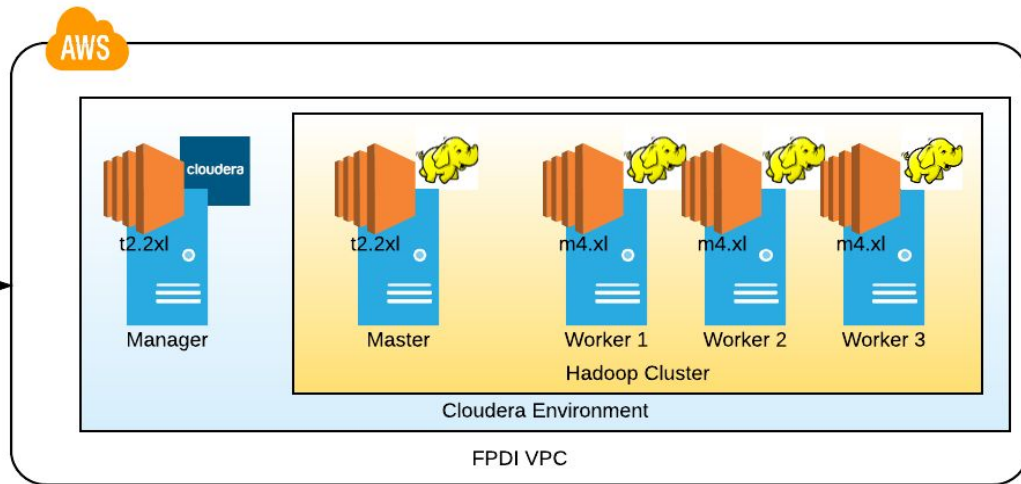




Client



cloudera  
IMPALA



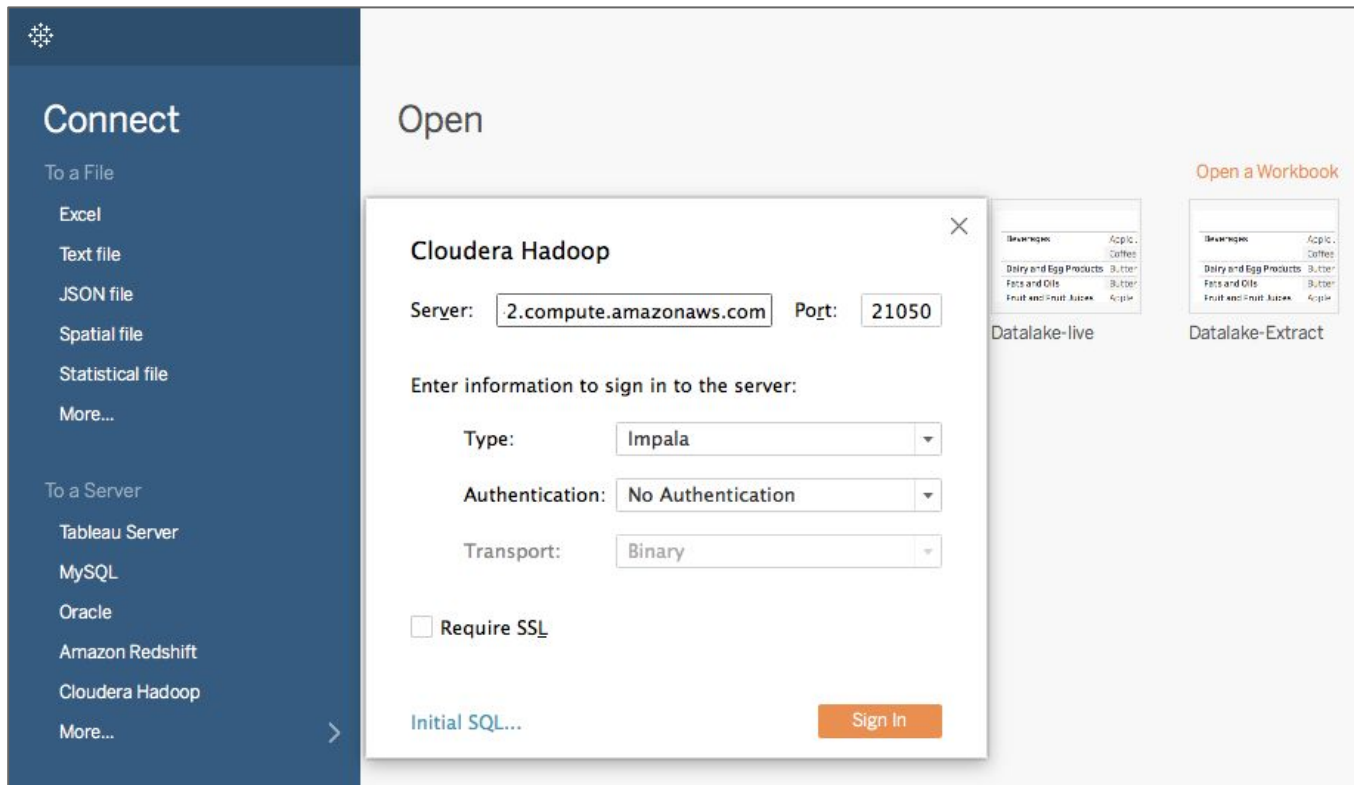
How does it work?

The screenshot displays the Hue web interface for the Impala database. The top navigation bar includes 'HUE', 'Query Editors', 'Metastore Manager', and 'Workflows'. The main header shows 'Impala' with options to 'Add a name...' and 'Add a description...'. On the left, a sidebar lists databases: default, dl\_factbook, dl\_fides, dl\_incident, factbook, factbook\_201710120100, fb\_foo, fda, fda\_caers, fides, fides\_201710110230, fides\_201710120230, incident, incident\_201710120130, and metadat. The central query editor contains the SQL statement: `1 select * from incident.emae_events`. Below the editor, the 'Results (13)' tab is active, showing a table with 6 rows and 5 columns: id, year\_began, year\_ended, and description\_of\_harm. The table data is as follows:

	id	year_began	year_ended	description_of_harm
1	4	2013	1983	Unknown
2	5	2007	2011	Kidney stones, hospitalization, death
3	6	2006	2008	Kidney stones, renal disease, and death in dogs and cats
4	7	2002	2007	Unknown; Sudan 1 is classified as a Category 3 carcinogen
5	8	2004	2007	
6	10	1997	1999	

## Impala SQL Interface





## Impala ODBC Connector

## FDA CAERS

Name	Pri Reported Brand Pr..	Product Role	Outcomes	Sym One Row Coded Sym..	
Apple Juice	FRESH LEMON JUICE	Concomitant	DEATH	HEADACHE, DIZZINESS, VI..	Abc
Cranberry Juice	FRESH LEMON JUICE	Concomitant	DEATH	HEADACHE, DIZZINESS, VI..	Abc
Garlic	GARLIC	Concomitant	DEATH	HEADACHE, DIZZINESS, VI..	Abc
Lemon Juice	FRESH LEMON JUICE	Concomitant	DEATH	HEADACHE, DIZZINESS, VI..	Abc
Lemons	FRESH LEMON JUICE	Concomitant	DEATH	HEADACHE, DIZZINESS, VI..	Abc
Orange Juice	FRESH LEMON JUICE	Concomitant	DEATH	HEADACHE, DIZZINESS, VI..	Abc
Palm Oil	SESAME OIL	Concomitant	DEATH	HEADACHE, DIZZINESS, VI..	Abc
Sesame Oil	SESAME OIL	Concomitant	DEATH	HEADACHE, DIZZINESS, VI..	Abc
Sugar (beet)	BEETROOT POWDER	Concomitant	DEATH	DEATH	Abc

## Factbook

Name	Value	
Apple Juice	---	Abc
	--- [{"description": "Core Fruit Juices or Concentrates", "fda_product_code": "20"}]	Abc
	--- [{"hazard": "Patulin", "control": "Supplier guarantee (apples harvested to exclude fall.."}]	Abc
	--- [{"hts_code": "2009710000", "long_description": "apple juice, of a brix value not exce.."}]	Abc
	--- [{"incident_summary": "In the United States, Beech-Nut sold fraudulent apple juice m.."}]	Abc
	--- [{"usda_ndb_code": "09016", "description": "Apple juice, canned or bottled, unsweete.."}]	Abc

## Outcomes

DEATH

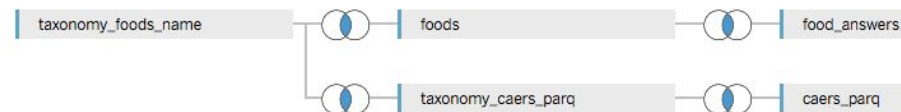
## Name

(All)

## Product Role

- ☐ (All)  
☒ Concomitant  
☐ Suspect

taxonomy\_foods\_name (factbook.taxonomy\_foods\_name)+ (fda\_caers)



# Joining on Common Taxonomy

# WHATS NEXT?

## Data Visualizations

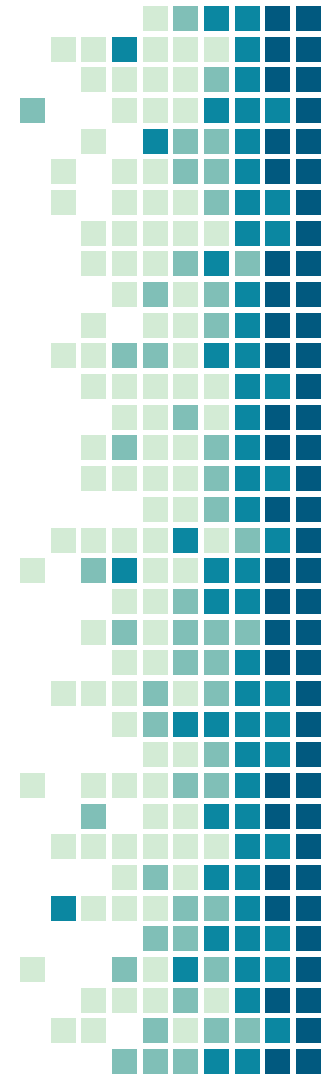
More dashboards using the data lake as the source of data.

## More Data

Upload, categorize and generally make the data available.

## Governance

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



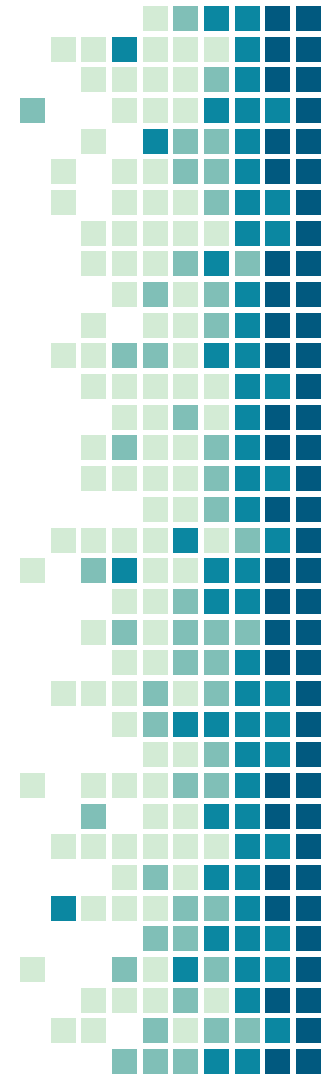
# THANKS!

Any questions?

You can find me at:  
[tsteinma@umn.edu](mailto: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] FPDl, “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] Tableau 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.

