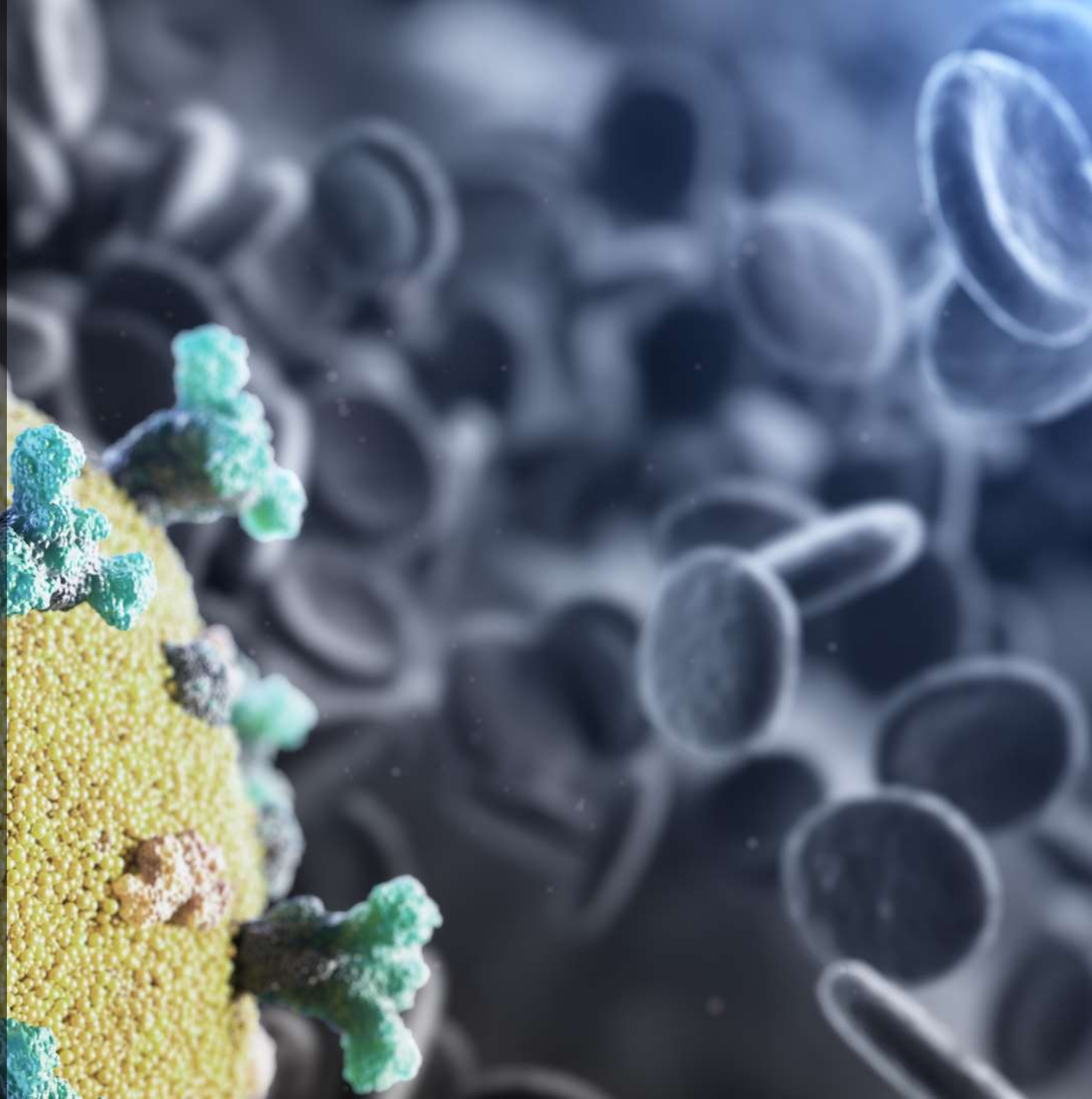
Open COVID-19 Task Force



The Open Coronavirus 2020 (COVID-19) Task Force is a consortium of industry experts working with state and local government(s) to build an analytics platform focused on eliminating the spread of the coronavirus.

Jeff Dailey Vexcel CTO

Jeffda@microsoft.com

704-466-6942

Overview

Objective: Produce new insights that allow targeting different segments of our population with information, aid, and testing derived from a deeper understanding of causality via correlation of data. The core objective Is to reduce the spread, flatten the curve and save lives!

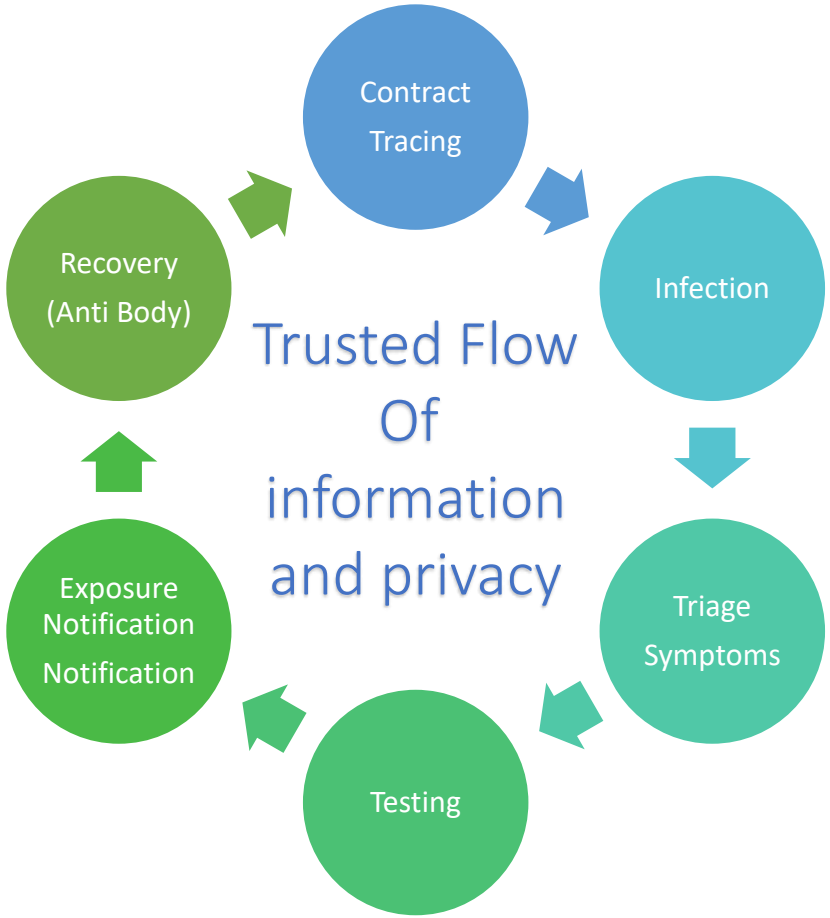
How: We will ingest COVID 19 test information, Antibody, Contract Tracing information along with data from other sources such as census/demographic data, weather data, etc. The data will be hosted in an open data lake - and indexed in a search and mapping platform - to allow data scientists and researchers to run machine learning and artificial intelligence (AI) analytics to enable deep insights and correlation. A key focus will be paring data scientists with medical personnel and researchers to yield new insights.

Who: The Open COVID 19 Task Force is a consortium of industry experts and government agencies, focused on test results and analyses of related information. We expect companies such as Elastic Search, Microsoft, ESRI and DataBricks to get involved and donate people and resources. These private sector teams will work in concert with government agencies to organize test data in a platform that will provide state and local government with deeper insights into their data.

The challenge many COVID 19 efforts in silos

Different COVID 19 Efforts			
Triage	COVID 19 Testing	Contact Tracing	Antibody Testing

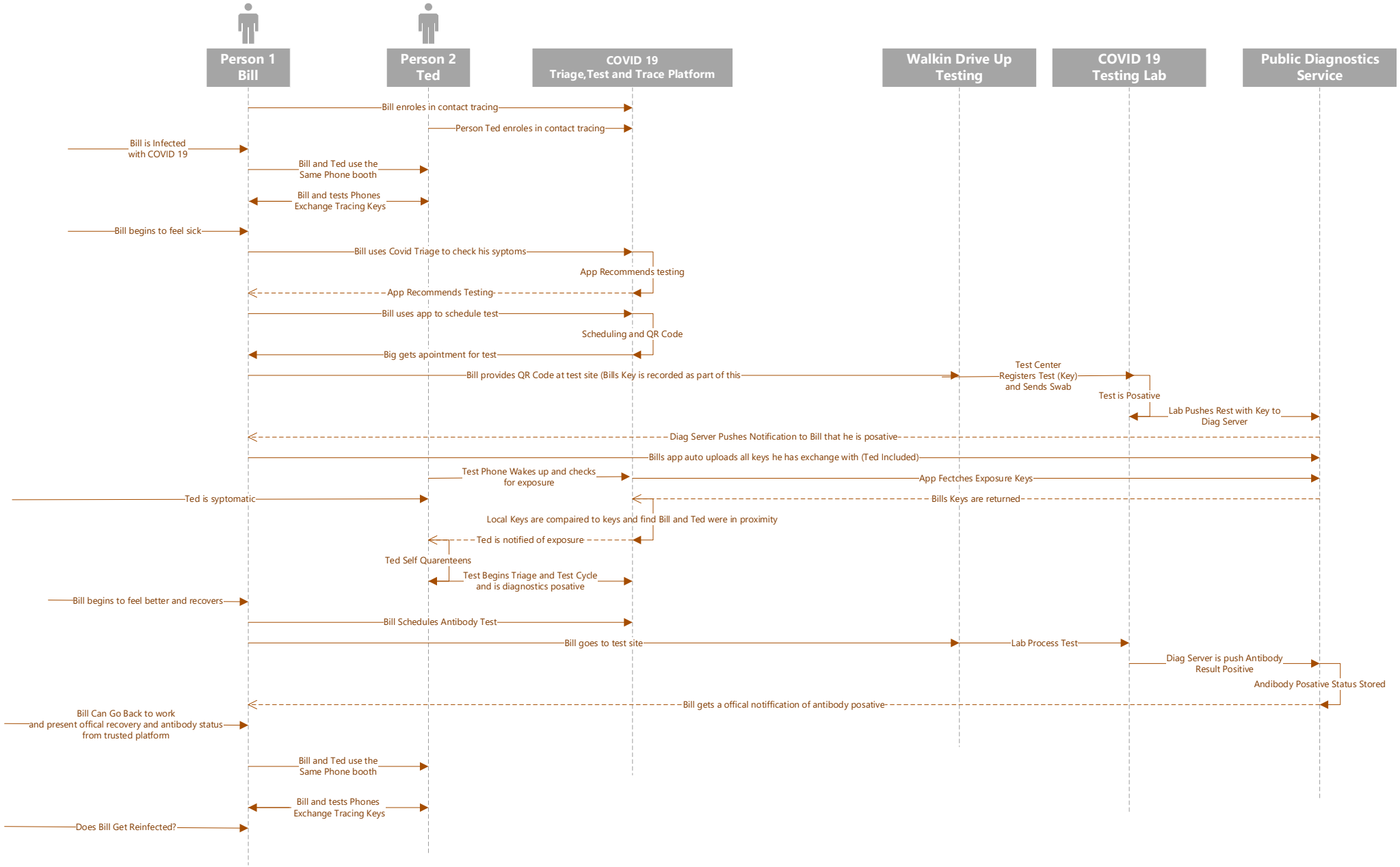
However there is a connected Virtuous Circle of COVID 19 infection and recovery events in insights that are needed



Key Points and challenges

- False positive self reporting. A unified platform would enable official test results to analyses push notifications from the lab to the contact tracing diagnostics service. i.e. The test application is also the tracing application. This will eliminate false positive self reports
- Providing an official antibody status from a trusted source via the app? If the antibody test app is tied into the lab the result could be used for return to work.
- What do we know about antibody carrying people and reinfection due to exposure. If the tracing app also knew about people with the antiboidy analytics comparing exposure rate of antibody and non antibody carrying people could be conducted anonymously .

Future state would be a Unified COVID 19 Platform to bring it all together

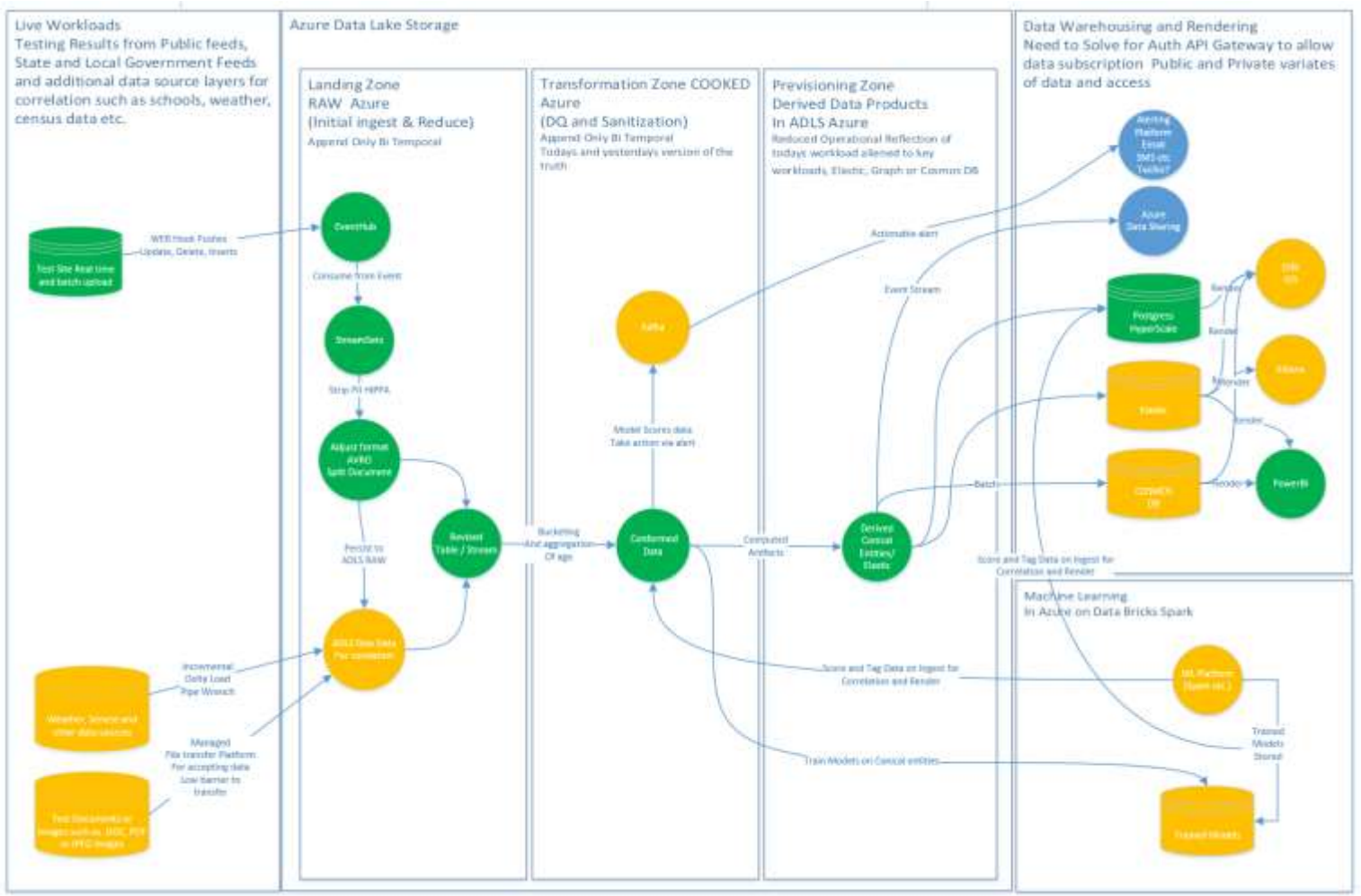


Progress so far focuses on data ingest and analytics (5/6/2020)

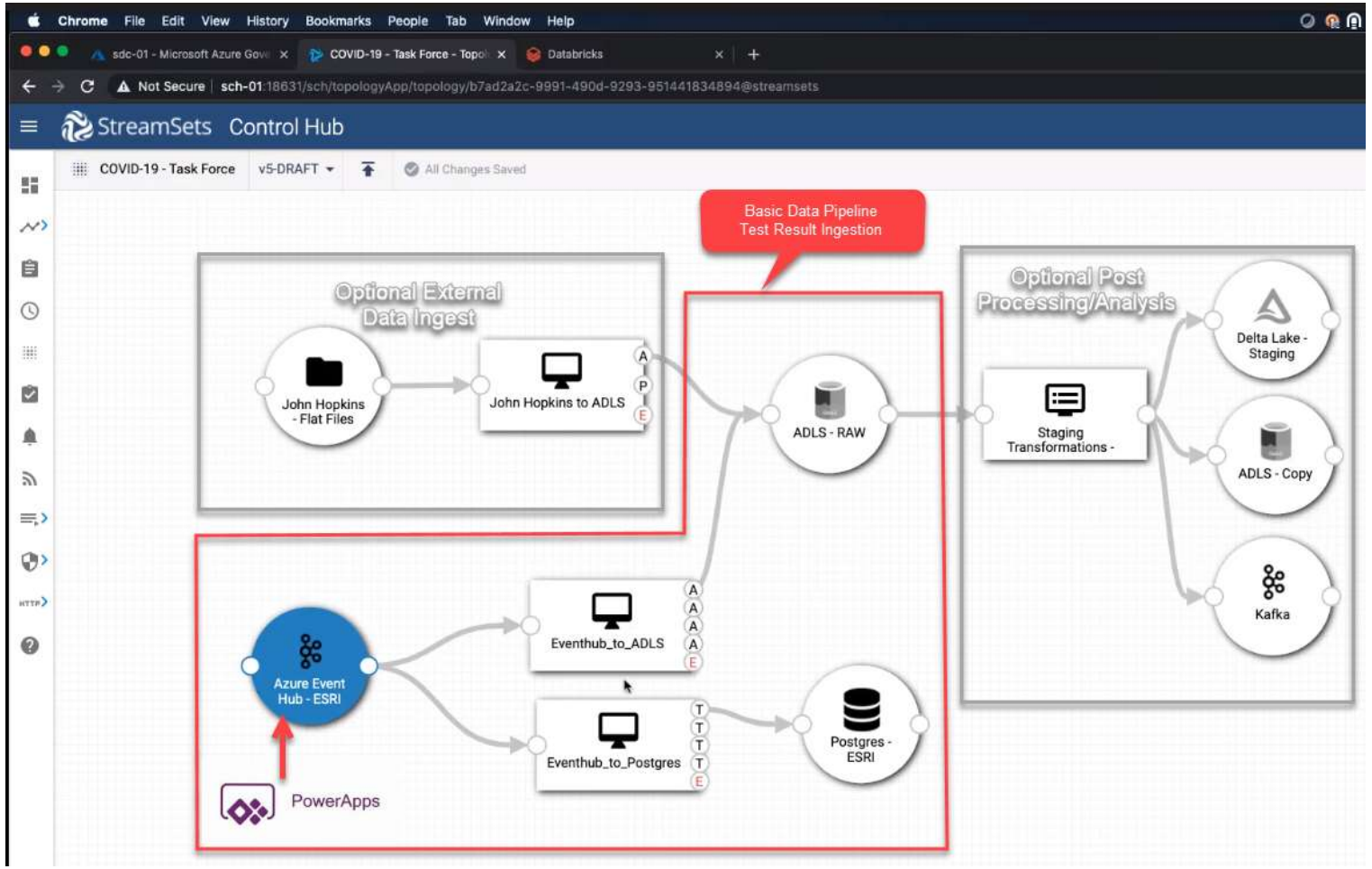
- DataBricks, Elastic, ESRI, StreamSets and Microsoft (Azure) onboarded and executive sponsors in place
- Add Partners agree to six month free subscription to governments or countries fighting covid 19
- Stood up Azure Subscription
- Created end-to-end solution specification
- Engaged COVID 19 Test Platform Data Import Pipeline Built
- Open GitHub Repo Built https://github.com/jeffdailey/OpenCovid19Team
- Engaged NYC State Microsoft Stakeholder to acquire New York state data
- Azure DevOps (ADO) workspace created
- Power BI Integration complete (Sean Donavin)
- ESRI Geo Analytics Integration Complete (Kritica ESRI)
- Data Bricks Analytics Ingegration Complete (Scot Black DataBricks)
- Elastic Analytics Integration complete
- Epics / Workstreams Defined and tasks created to fulfill solution architecture
    - COVID 19 Test Platform Data Export
    - Raw Data ingest
    - Data Conformance
    - Data Provisioning for Analytics ingest
    - Geopatial Analytics (ESRI)
    - Machine Learning (DataBricks)
    - Search Platform (Elastic)
    - Platform Access Control for read and write

Base Architecture In green

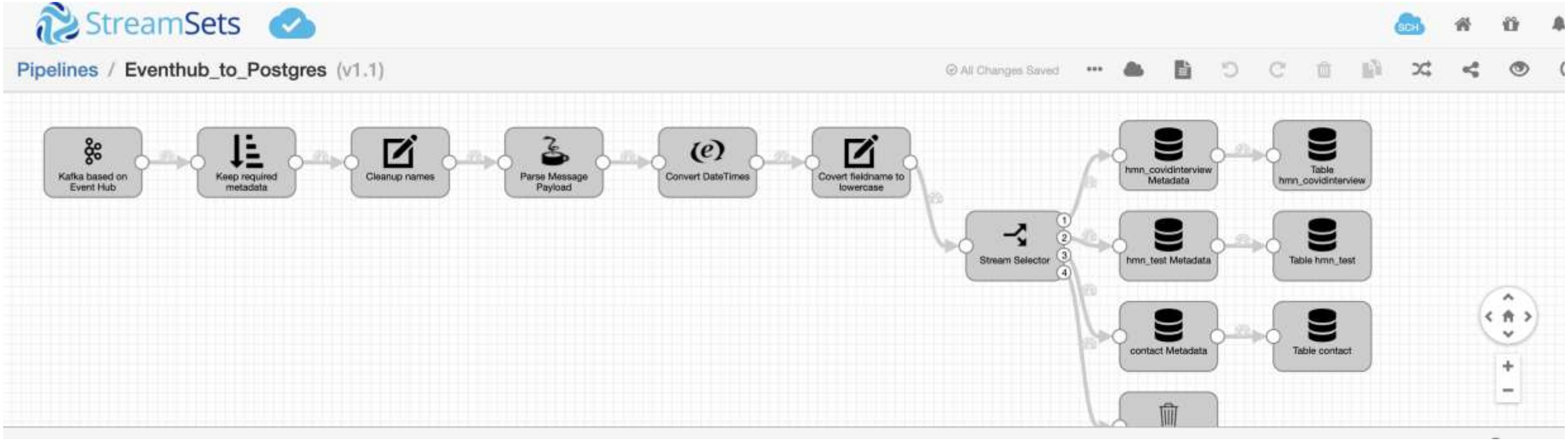
- Enable Real time and Batch ingest from different test platforms and align to standard HIPAA scrubbed format.
- Leverage DataBricks Spark for ML on Azure
- Enable multiple down-stream analytics and rendering platforms such as Elastic and Kibana and/or Cosmos DB and PowerBI



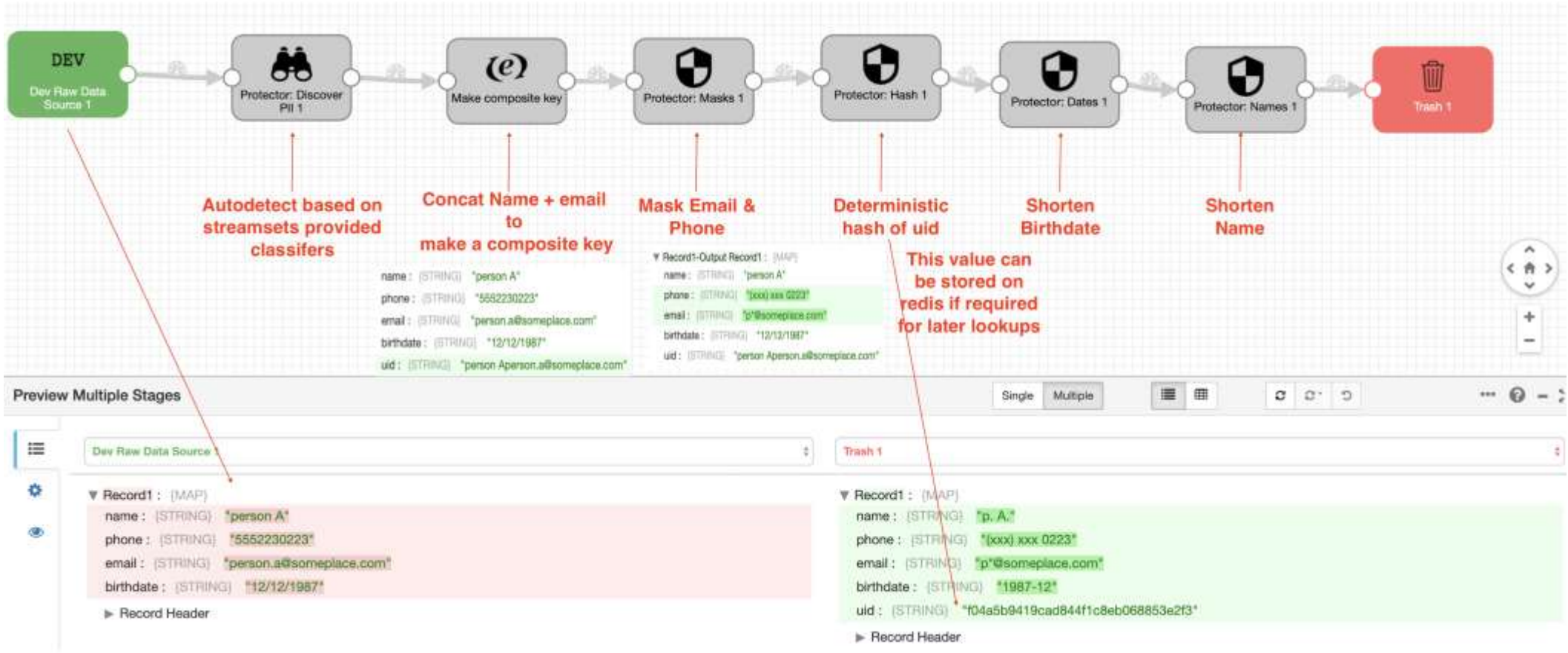
Basic Test Result Ingest



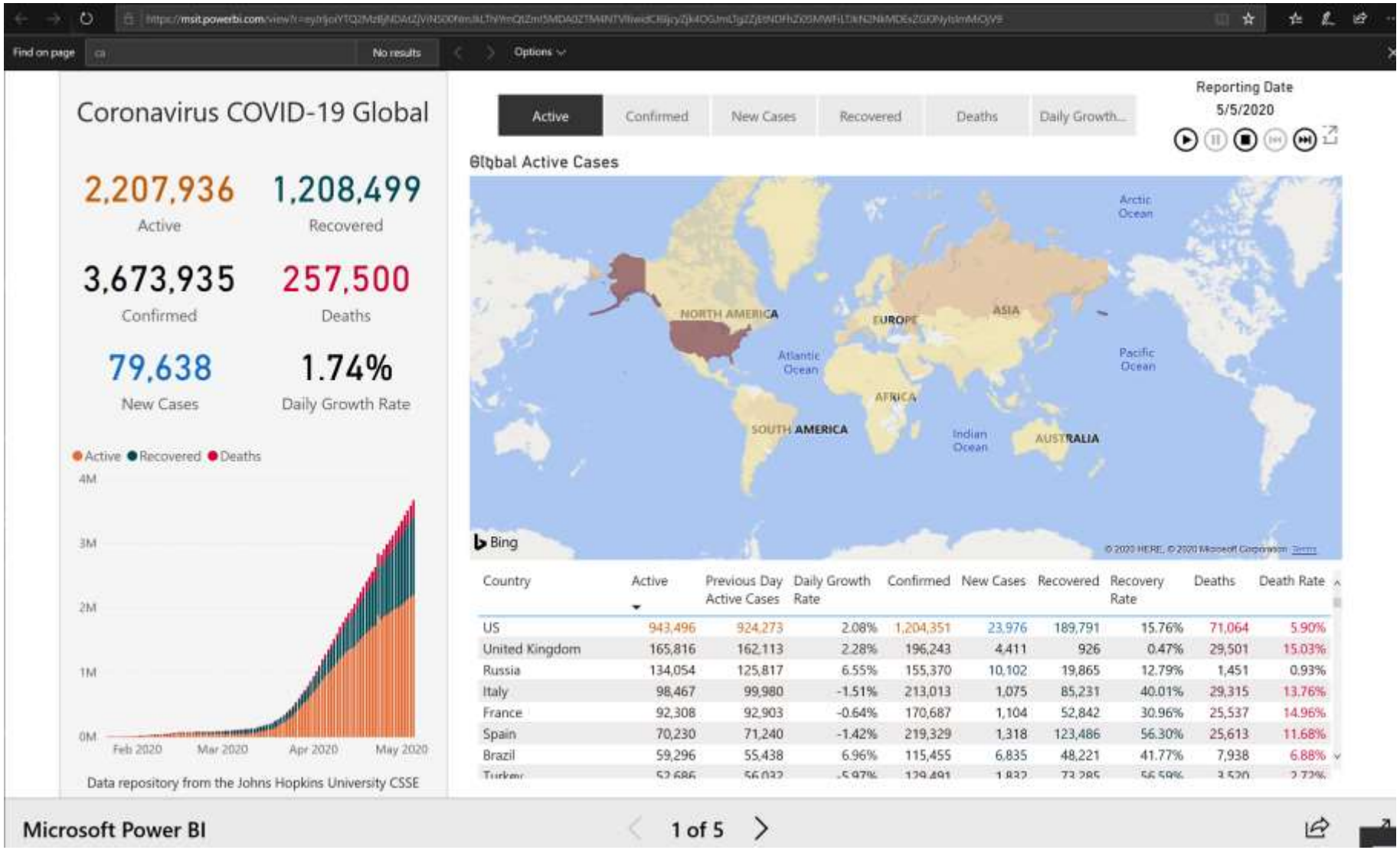
StreamSets Auto Data Drift Management



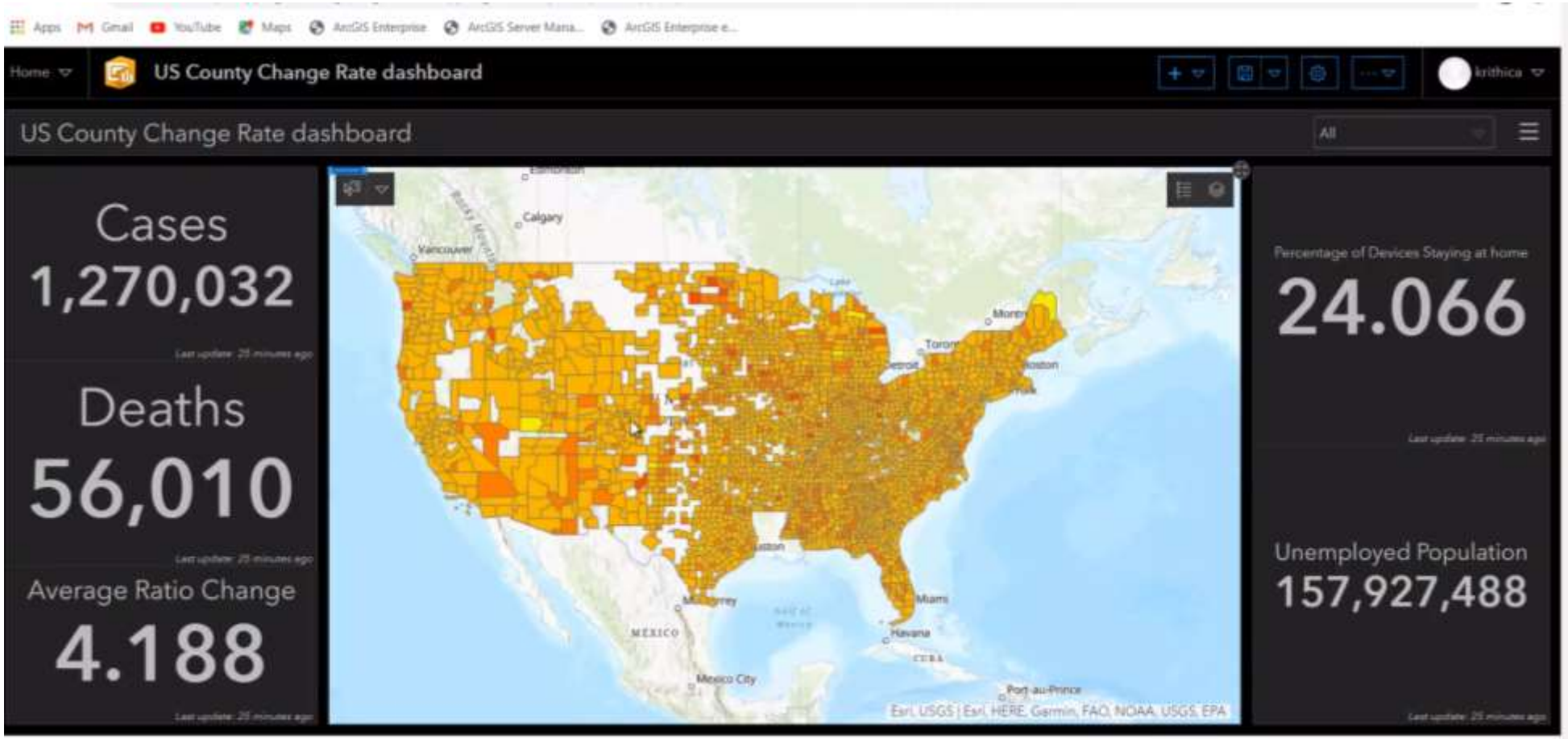
StreamSets Ingest can mask data for analytics



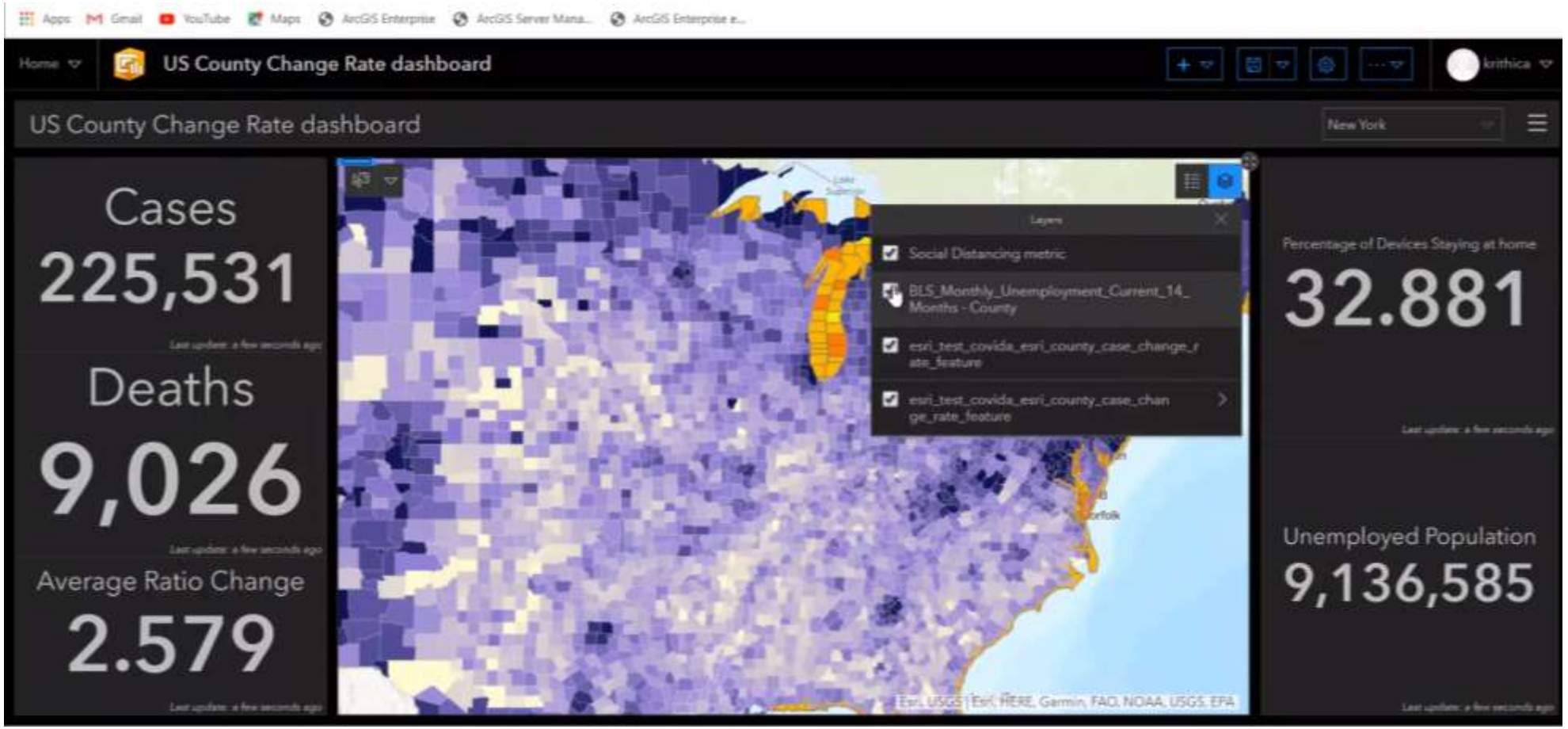
Example PowerBI Covid 19 dashboard



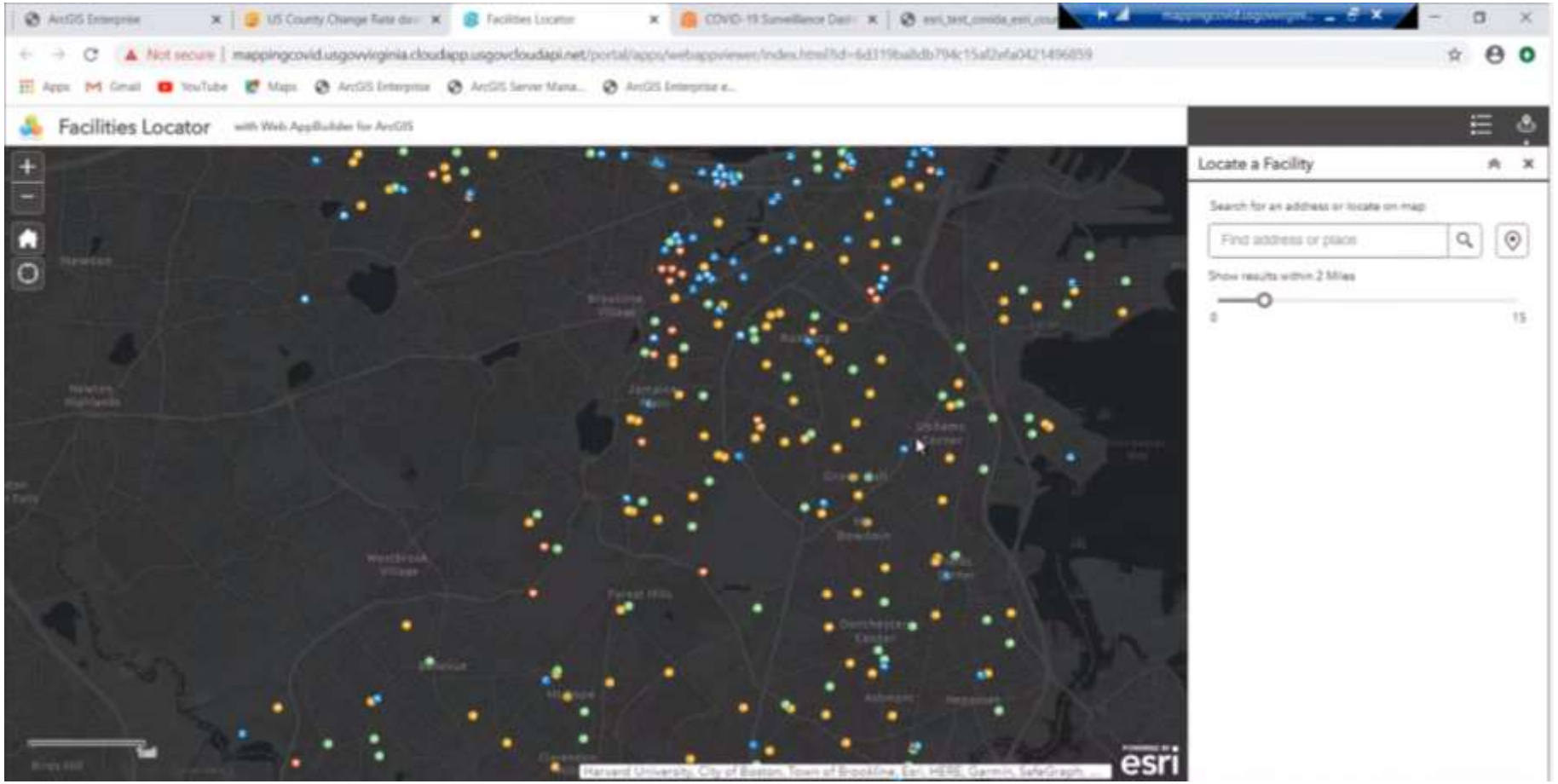
Arc GUS High Level



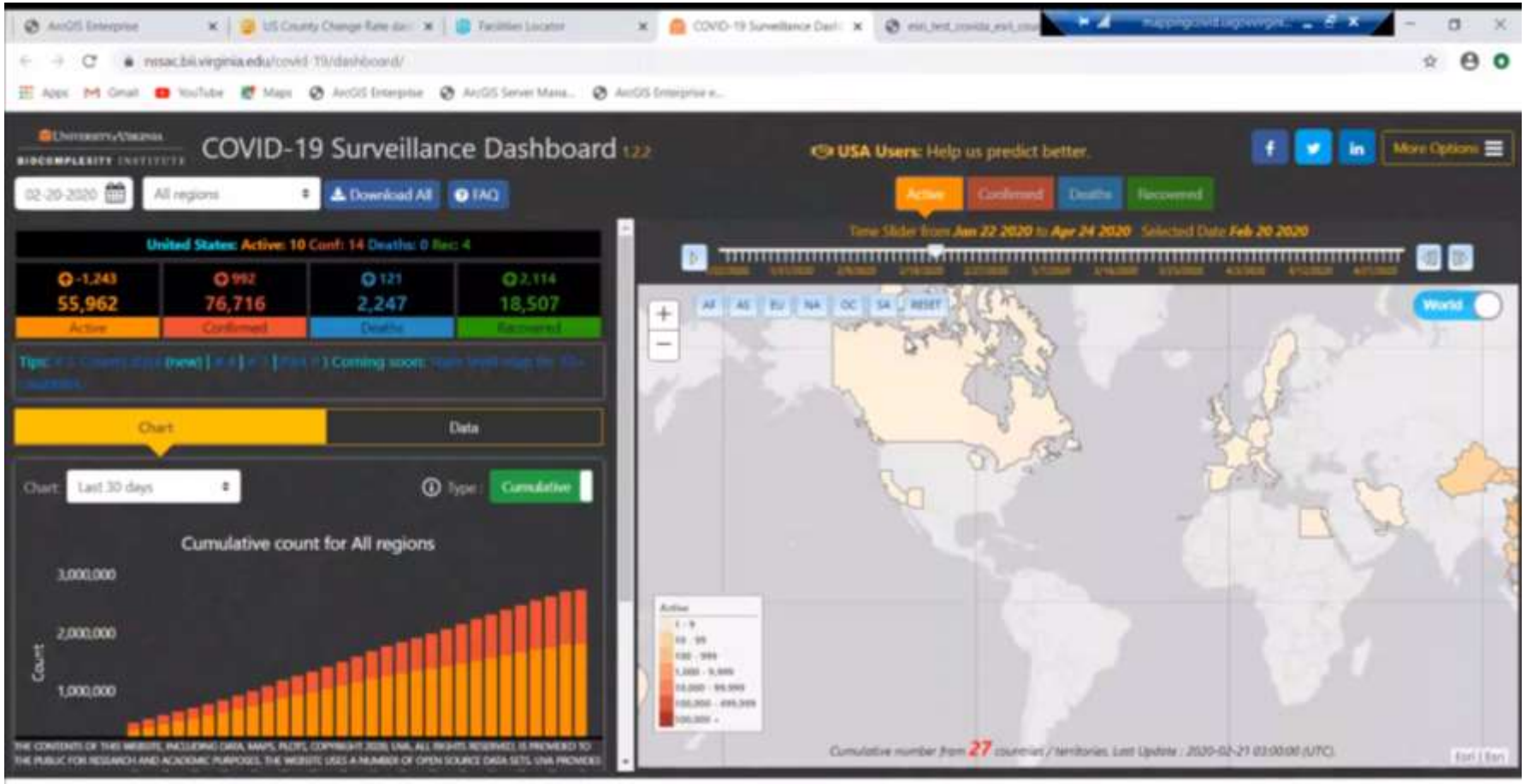
ESRI Arc GIS Drill down



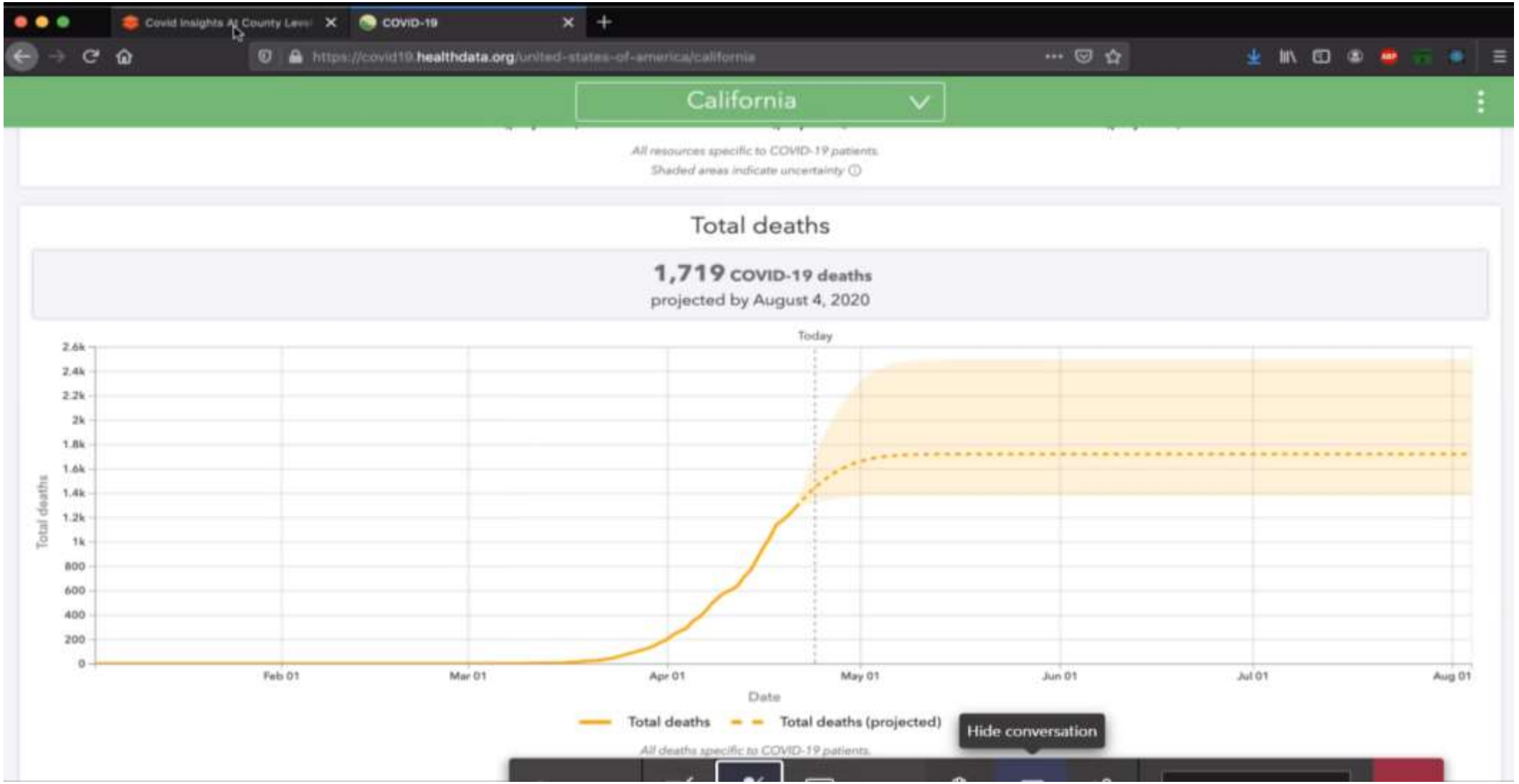
ESRI Arc GIS Points of interest



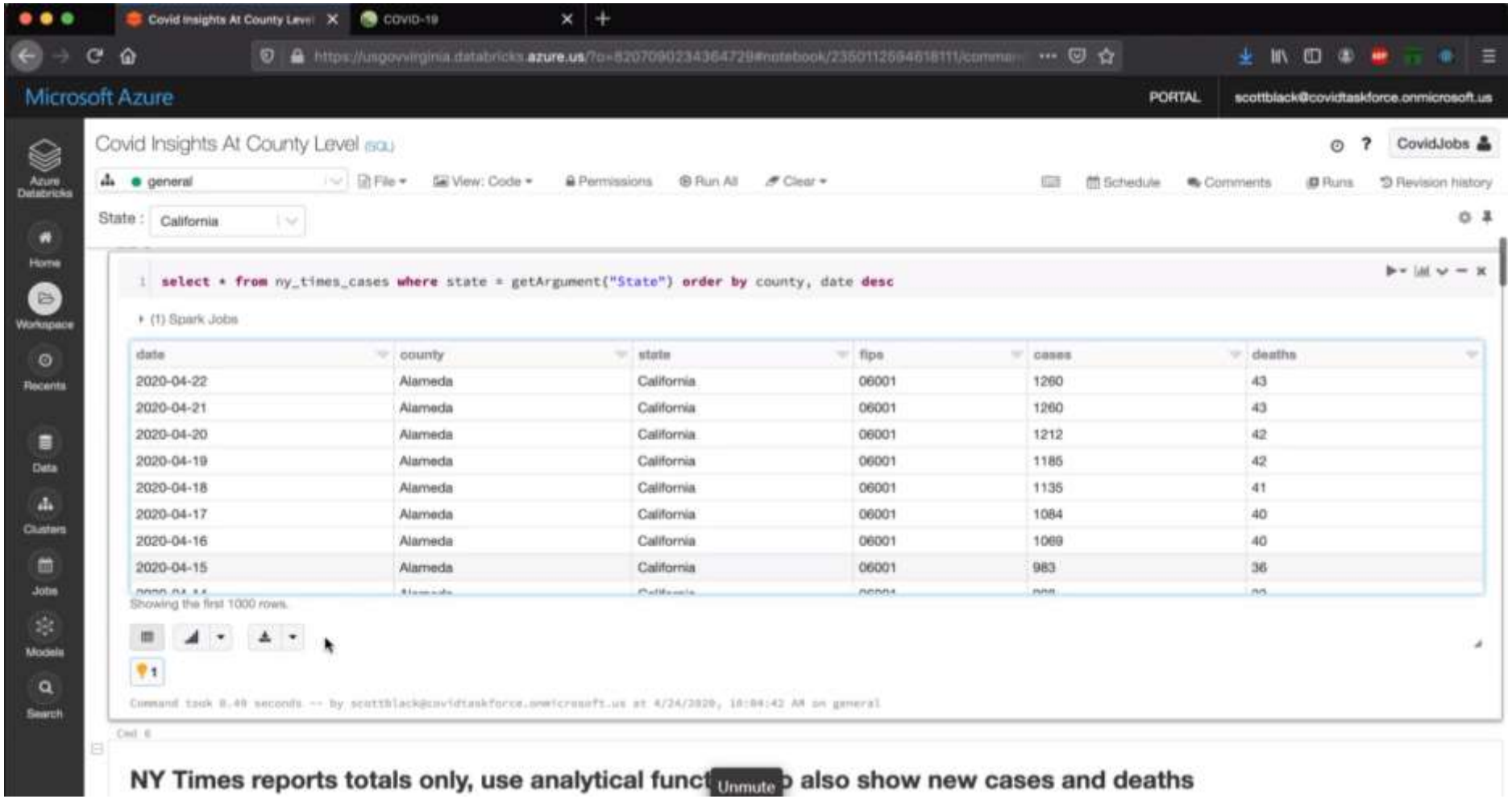
ESRI Arc GIS Analytics View



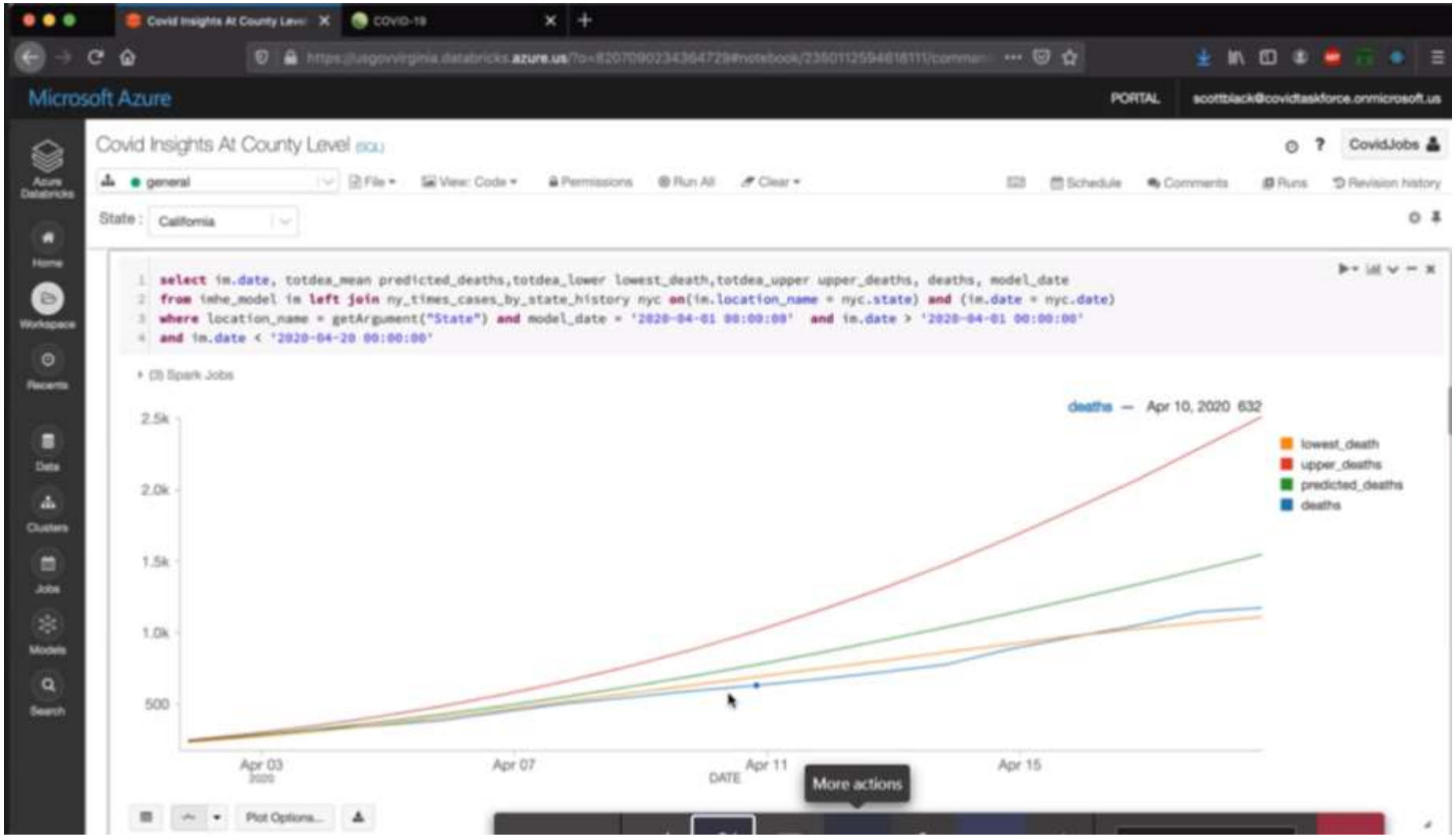
Data Bricks Projection of total death



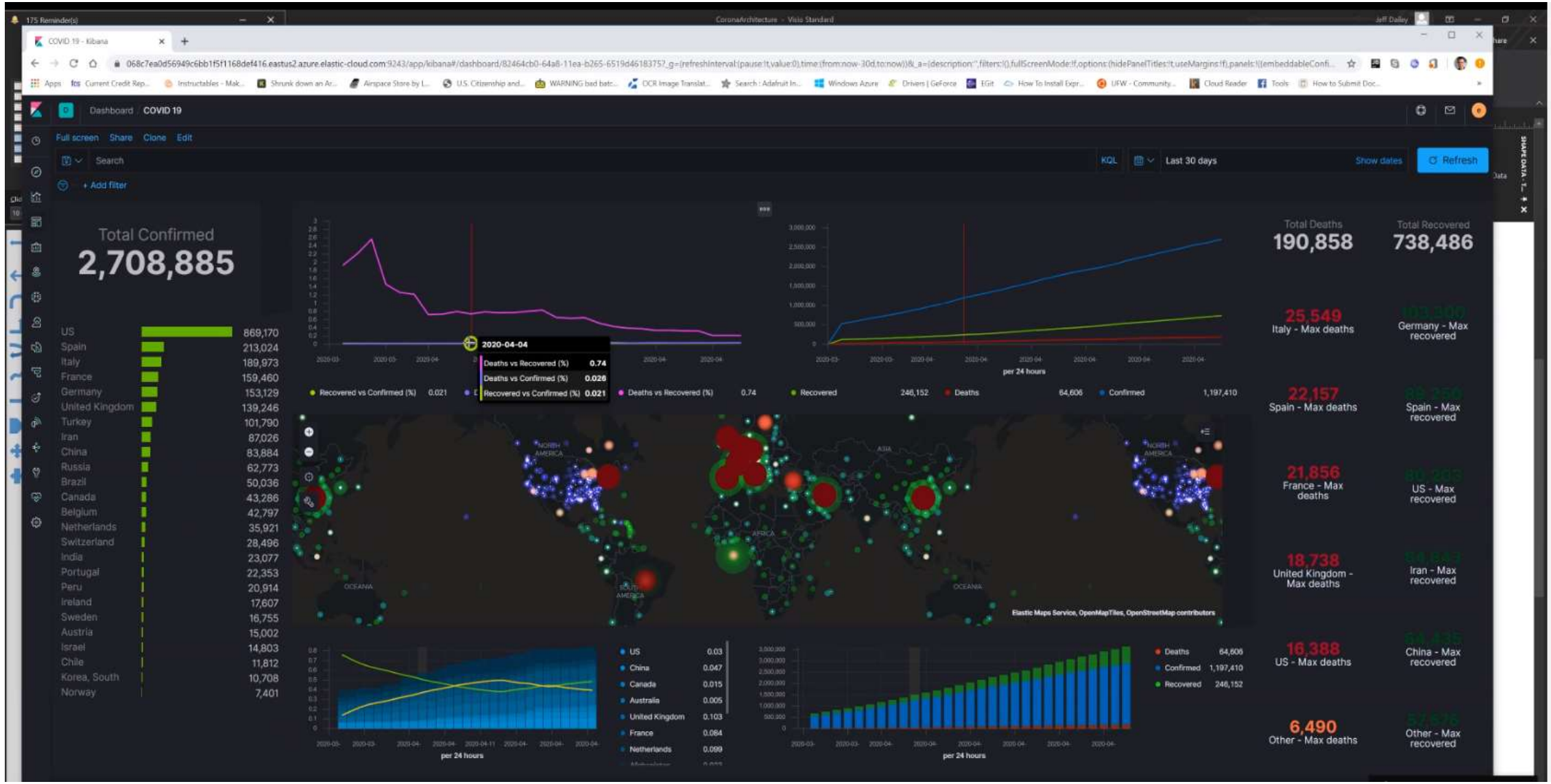
SQL Access to data set via DataBricks



University of Washington Model Compariosn



Elastic Search Kibana Analytics



Appendix

Current/Next Steps and Key Asks

- Decompose specifications into Epics and work items in ADO, and assign to key players
- Deploy and configure Azure Data Lake
- Deploy and configure DataBricks, ESRI and Elastic; work with team on Arm Templates and SMES
- Deploy PostGres HyperScale for the ESRI Platform
- Refine Data Lake Architecture and multi-tenancy HIPAA requirements
- Began working with New York and DC COVID 19 tracking teams to design data integration pipelines and began ingest of COVID 19 test results. (In progress)
  - Target 4/7 for initial ingest of NYC data
- Refine additional data source ingests and schemas (e.g, Census, Weather, School and other data sources).
  - Will have Johns Hopkins and CDC data in lake, along with state and local near real-time feeds.
- Connect with stakeholders at Higher Education and Academic Medical Centers (Johns Hopkins and University of Washington). Align to efforts such as: https://covid19.healthdata.org/
- Need to secure Data Science, Medical and Research stakeholder for design input

Sponsorship Approach

- Reached out to key C-level executives in target companies to enable support through donation of cloud resources, services personnel, program management and engineering. Collaboration with existing state and local government engagements focused on Coronavirus work, such as Texas and New York State, can enable analytics on test results. The key objective is to act locally while thinking globally, allowing the platform to scale worldwide.
- Microsoft Sponsors
  - Azure Resourcing (Mark Russinovich, Azure CTO and Ulrich Homann, Azure CVP)
  - Microsoft Services Data Science (Lenny Fenster MCS Data and AI CTO)
  - Vexcel State and Local Engagement (Jesper Erichsen, GM)
  - GPMO PM Resources (Maurice Magnier GM)
- DataBricks Sponsors
  - Mike Maxwell, Executive Sponsor
  - Tim Boyle, Program Administration
- ESRI Sponsors
  - Mansour Raad
- Gates Foundation
  - TBD
- ElasticSearch
  - Vijay Doshi
- StreamSets

Key Contacts (Ongoing)

Microsoft

COVID Test integration with Analytics platform: Sean McNellis SeanMcn@microsoft.com
COVID Testing Tracker: Drew Gervino drew AndrewG@microsoft.com
(Vexcel CTO): Jeff Dailey JeffDa@Microsoft.com ; Cell 704-466-6942
Microsoft Services Data and AI CTO: Lenny Fenster lfenster@microsoft.com
Azure Government: Zach Kramer Zach.Kramer@microsoft.com

DataBricks

Mike Maxwell, Exec Sponsor: mike.maxwell@databricks.com
Tim Boyle, Program Administration: tim.boyle@databricks.com

ElasticSearch

Vijay Doshi: vijay.doshi@elastic.co

ESRI

Mansour Raad: mradd@microsoft.com
Krithica Kantharaj kkantharaj@esri.com

StreamSets

Asim Mohammad: asim.mohammad@streamsets.com
Shekhar Iyer shekhar@streamsets.com StreamSets President
Tom Ganka thomas@streamsets.com

Databricks Contacts

- Kevin Davis : VP & GM Public Sector (Executive Sponsor)
  - KD@databricks.com
- Mike Maxwell : National Director State & Local Govt's, Higher Ed & Research Institutions
  - mike.maxwell@databricks.com
- Tim Boyle : Director Public Sector Field Engineering
  - Tim.Boyle@databricks.com
- Colin Risler : Mgr Field Engineering, State & Local Govt's, Higher Ed & Research Institutions
  - Colin.risler@databricks.com
- Scott Black : Sr. Solution Architect Public Sector
  - Scott.Black@databricks.com
- Ameet Kini : Regional Leader, Public Sector Resident Solution Architects
  - ameet.kini@databricks.com
- Qing Sun : Public Sector RSA, CMS focus
- Glenn Wilchacky: Enterprise Account Executive, Azure Databricks East
  - glenn.wilchacky@databricks.com
- Abhinav Garg: Principal Solution Architect, Azure Databricks
- Kyle Weller: Program Manager – Azure Databricks US Gov
  - kyweller@Microsoft.com