Data Lakes in the Land of 10,000 Lakes

Miriam Muscoplat, MPH| MIIC Manager

Aaron Bieringer| MIIC Interoperability and Technical Lead

Land Acknowledgement

Every community owes its existence and vitality to generations from around the world who contributed their hopes, dreams, and energy to making the history that led to this moment. Some were brought here against their will, some were drawn to leave their distant homes in hope of a better life, and some have lived on this land for more generations than can be counted. Truth and acknowledgment are critical to building mutual respect and connection across all barriers of heritage and difference.

We begin this effort to acknowledge what has been buried by honoring the truth. The presenter is standing on the ancestral lands of the Dakota people. We want to acknowledge the Dakota, the Ojibwe, the Ho Chunk, and the other nations of people who also called this place home. We pay respects to their elders past and present. Please take a moment to consider the treaties made by the Tribal nations that entitle non-Native people to live and work on traditional Native lands. Consider the many legacies of violence, displacement, migration, and settlement that bring us together here today. Please join us in uncovering such truths at any and all public events.*

*This is the acknowledgment given in the USDAC Honor Native Land Guide – edited to reflect this space by Shannon Geshick, MTAG, Executive Director Minnesota Indian Affairs Council

Overview

- MIIC Technical Landscape
- Why a data lake?
- Process
- Uses
- Lessons Learned

MIIC Technical Landscape

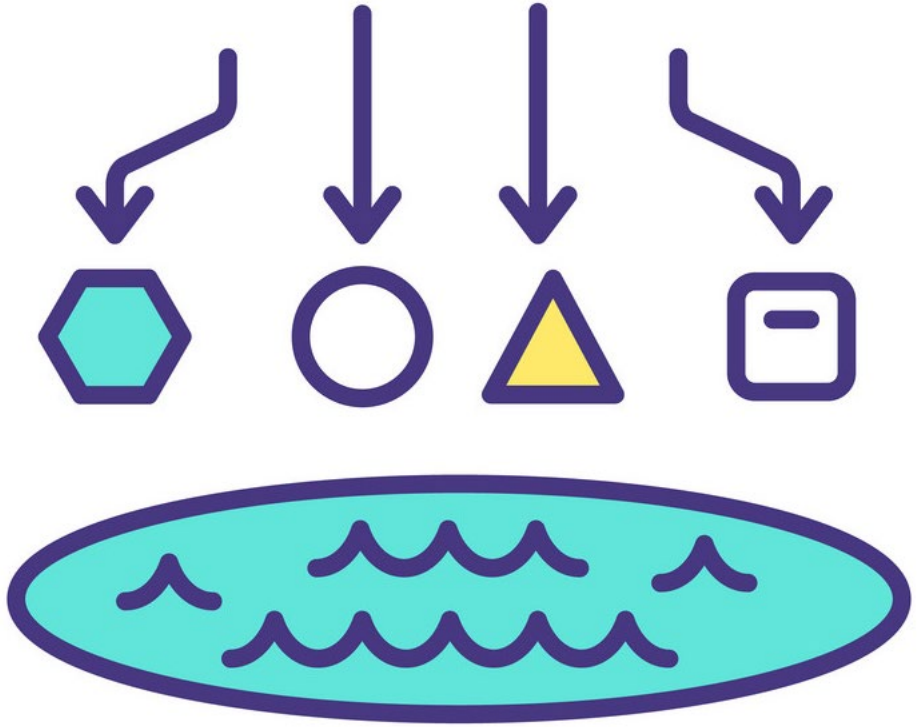
Moving to the Cloud

Before the Cloud

- Oracle database hosted on prem
- Little to no system flexibility during times of increased activity

After the Cloud

- More database resources available to us
- More opportunity for data use outside of MIIC application
- Increased security
- Increased resiliency



DEPARTMENT OF HEALTH

Why a data lake?

What is a data lake?

- What it is:
  - An advanced and flexible set of tools for data management
  - Centralized storage medium used for reporting across systems
  - Storage place for ancillary data
  - A way to automate difficult manual data processes
  - A central access control mechanism for data

- What it is not:
  - A silver bullet for all data management problems
  - A catalog of business data definitions
  - Magic

Advantages

- Automation of workflows
- Ability to remove pressure/traffic from the MIIC application
- Ability to increase or decrease capacity as needed
- Change to input and output parameters does not require dev work
- Future state: easier to work with other data sets within the agency

Use of the MIIC Data Lake

Texting Reminder/Recall

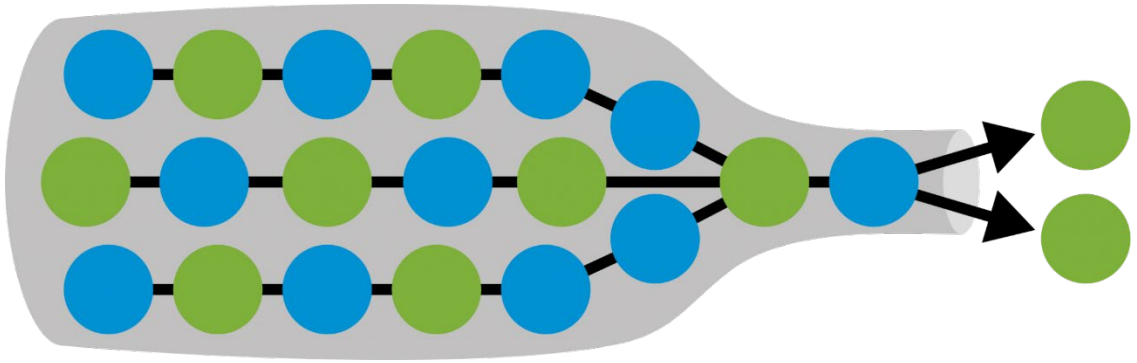
- The data lake allowed us to:
  - Efficiently text without impacting the performance of MIIC
  - Clearly use/manage data for individuals who received texts and across texting campaigns using multiple data sources
  - Pull data using the Amazon Web Services (AWS) texting application
  - Pull data into Tableau to create data visualizations

Data Monitoring

- Vaccination rates in Medicaid population
  - Allows us to easily monitor both COVID and routine vaccination in Minnesota's Medicaid and MNCare populations.
- Future:
  - Participation metrics
  - Quarterly population-based assessment
  - Data Quality metrics
  - Internal and external reporting for text message based reminders

Bulk Query

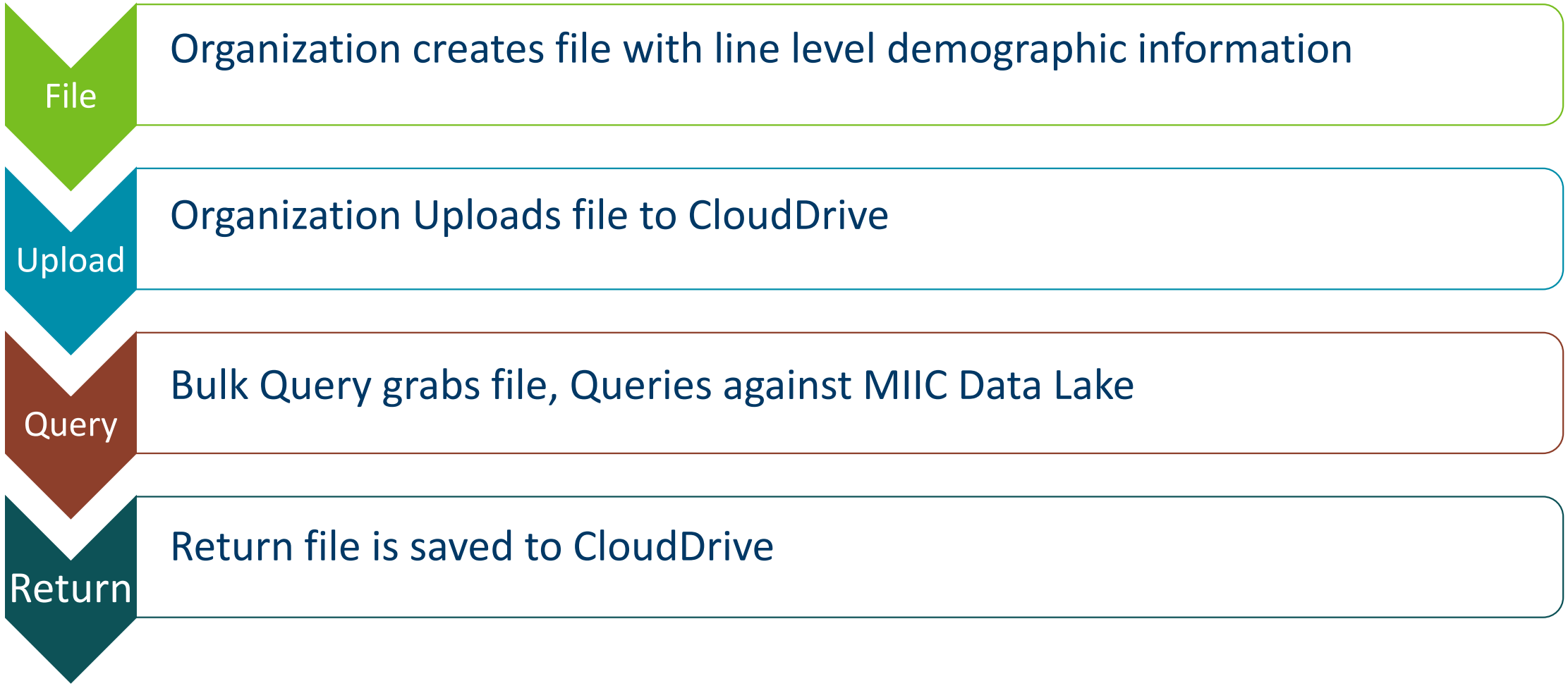
- Healthcare providers and health insurers want updated vaccination data on all their patients/enrollees regularly
  - This could amount to millions of additional queries per week.
  - Already concerned about the impact of additional volume of HL7 updates and point of care queries.
  - Existing large query option in UI is resource intensive and inefficient.



Bulk Query

- Create a copy of our matching algorithm that runs against the data lake to provide this data to providers.
  - No stress on the application or transactional database
  - Multi-threaded, very fast
  - Fully automated, driven by external partner

Bulk Query Process



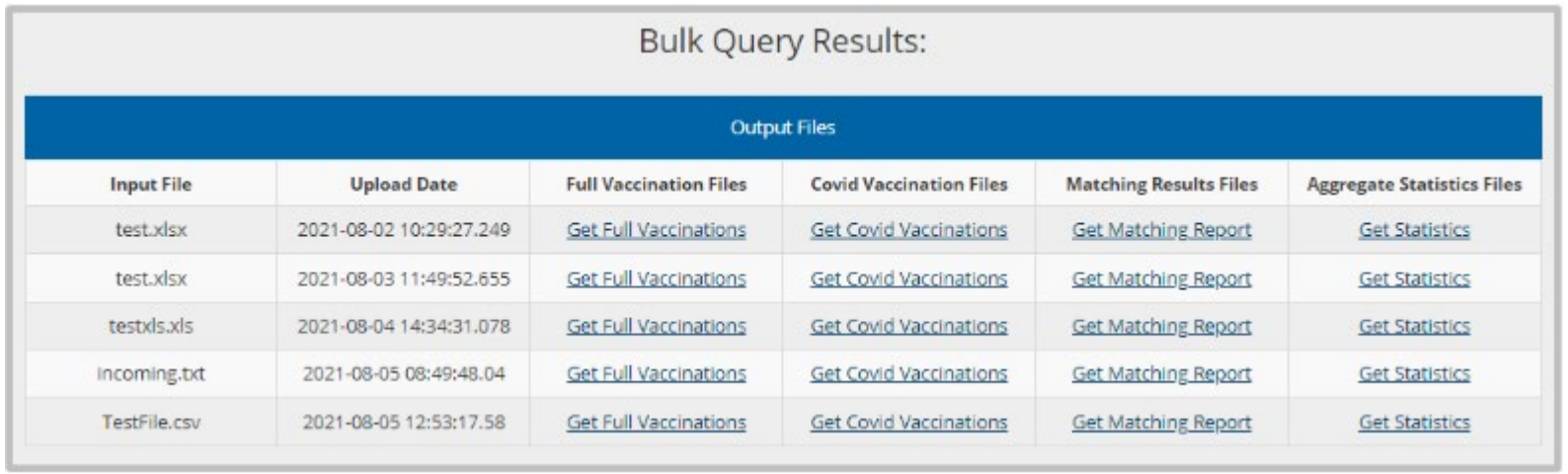
Uses of Bulk Query

Uses of Bulk Query

- Healthcare providers/health insurers
- Matching to COVID case data
- Matching to birth data
- Added to school reporting application
- Employee Vaccine Verification Program

School Reporting Application

- Added query option to school reporting application
- Piloted with 5 school districts with varying levels of internal technical support
- Modified output to provide a COVID only file and a CSV file
- Simple to use and maintain



Bulk Query Results:

Output Files					
Input File	Upload Date	Full Vaccination Files	Covid Vaccination Files	Matching Results Files	Aggregate Statistics Files
test.xlsx	2021-08-02 10:29:27.249	Get Full Vaccinations	Get Covid Vaccinations	Get Matching Report	Get Statistics
test.xlsx	2021-08-03 11:49:52.655	Get Full Vaccinations	Get Covid Vaccinations	Get Matching Report	Get Statistics
testxls.xls	2021-08-04 14:34:31.078	Get Full Vaccinations	Get Covid Vaccinations	Get Matching Report	Get Statistics
incoming.txt	2021-08-05 08:49:48.04	Get Full Vaccinations	Get Covid Vaccinations	Get Matching Report	Get Statistics
TestFile.csv	2021-08-05 12:53:17.58	Get Full Vaccinations	Get Covid Vaccinations	Get Matching Report	Get Statistics

Employee Vaccination Verification Program

- Tool allowing employers to request COVID19 vaccination status of employees with employee consent.
- MIIC queried to obtain COVID19 vaccination status.
- Online survey form created to capture both consent and employee demographic information in one place.
- Simple, secure, automated process

Stats

17 Health Systems/Health Plans

- Over 50,000,000 people queried

29 Schools/School Districts

- Private Schools
- Charter Schools
- Large public-school districts

6 Universities

- Large public universities
- Small liberal arts schools

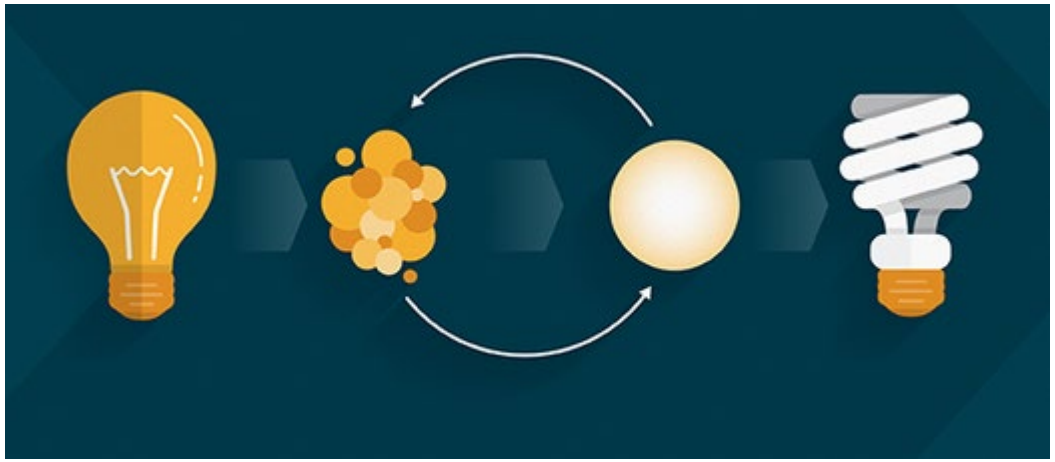
40 employers

- State Agencies
- School Districts
- LTCF

Lessons Learned

Lessons Learned

- Input file format
  - New, pipe delimited file format to meet this need
  - Not standardized
- New IT team
- Easy to change/add output files depending on the audience
- Data lake monitoring



Thank You!

Miriam Muscoplat

Miriam.muscoplat@state.mn.us