



NATIONAL HEAD START ASSOCIATION

# New Resource for Programs: The Head Start Data Lake Collaborative

Wednesday, September 18, 2024

## Raise your hand if...

When OHS announced the Full Enrollment Initiative, you wanted to know how your under-enrollment compared to that of other programs



## Raise your hand if...

Your program is struggling to address challenging behaviors and you're looking for effective solutions





## Raise your hand if...

You wish you could talk to other programs like yours to see what they're doing and what you can learn



## Raise your hand if...

Your program doesn't have a data manager (or similar) role



## Raise your hand if...

You feel like your program collects **reams** of data, but has a hard time actually using it



What you need is...



**the Head Start Data Lake  
Collaborative!**



# Agenda

- 1 What is a data lake?
- 2 What makes the Head Start data lake special?
- 3 Can I see the data lake?
- 4 Is it even possible to safely share data?
- 5 What does the future of the data lake look like?
- 6 How can I join?
- 6 I still have questions...

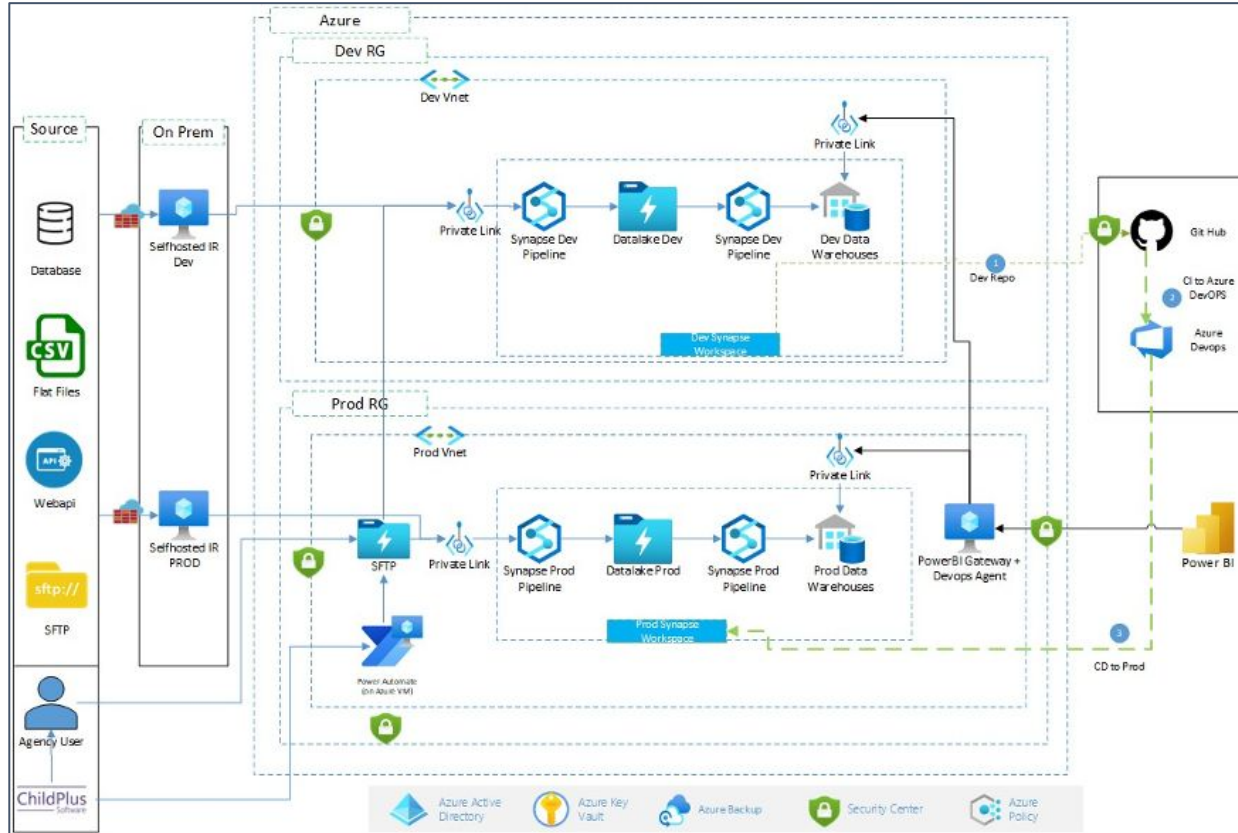


# What is a data lake?

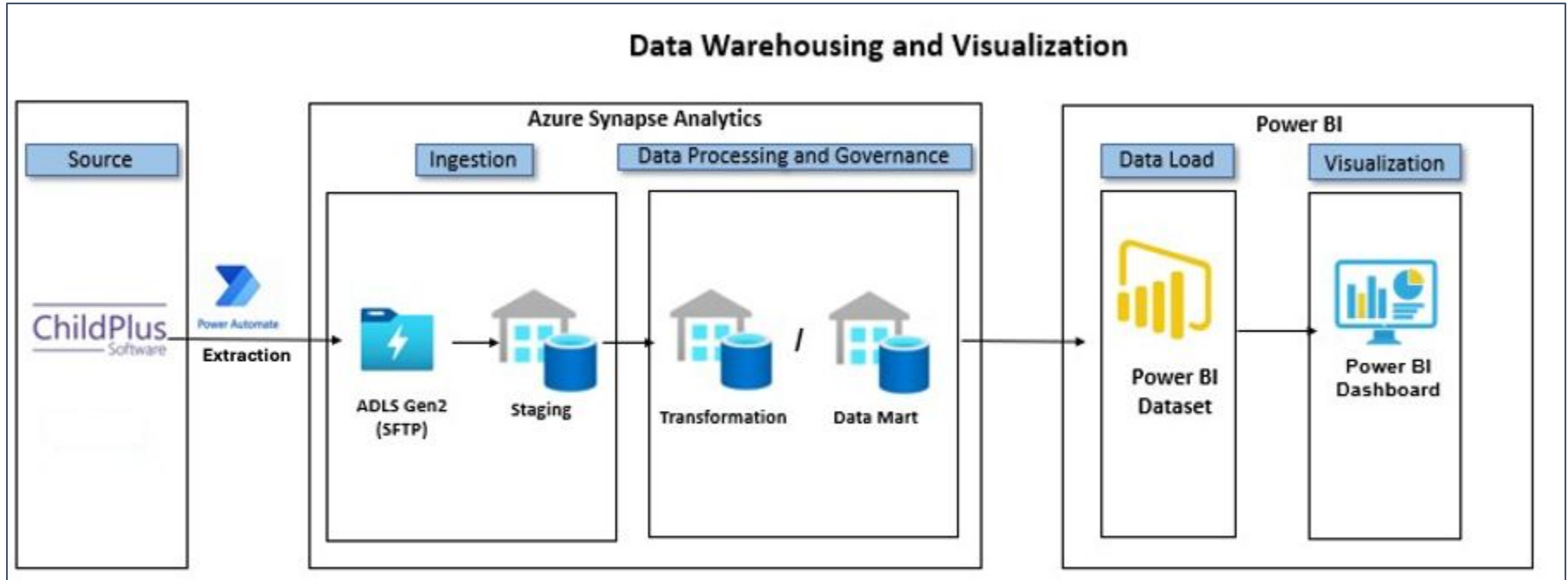


“A data lake is a centralized repository designed to store, process, and secure large amounts of structured, semistructured, and unstructured data.”

# What is a data lake?



# What is a data lake?



# What makes this data lake special?

The Head Start Data Lake will...

- Pull together data from Head Start programs nationwide
- Report data back to programs in the form of usable info
- Identify national trends from aggregated data

... in order to provide actionable information back to programs to improve systems and drive practical improvement.



# Mission Statement



*The Head Start Data Lake advances early childhood practice by delivering actionable information about children, families, staff, and programs to support continuous improvement efforts locally and nationally.*

# Our Model

- **User-centered design**, placing the experience and needs of the Head Start practitioner at the center of our work
- Programs need to feel **safe** to put their data in the lake
- Programs need to feel safe to engage in improvement efforts **without fear**
- NHSA can ensure this is **available to all programs**, no matter their size, budget, or technical capacity
- Expansions and improvements to the data lake will be **driven by programs** (new questions, new data, new analyses)



>Welcome  
TO THE  
LAKE

A wooden sign is mounted on a white brick wall. The sign features the word "Welcome" in a cursive script at the top. Below it, a horizontal wooden cutout of a boat's hull is positioned. The words "TO THE" are cut out of the boat's hull. Below the boat cutout, the word "LAKE" is written in a large, bold, blocky font. The entire sign has a natural wood grain texture.



## National Comparisons

140 programs contributing

### Enrollment

66%

Bucket: 71-80  
Programs: 61

0-10 11-20 21-30 31-40 41-50 51-60 61-70 71-80 81-90 91-100

### Daily Attendance

96%

0-10 11-20 21-30 31-40 41-50 51-60 61-70 71-80 81-90 91-100

### Relative Size of Waitlist

16%

0-10 11-20 21-30 31-40 41-50 51-60 61-70 71-80 81-90 91-100

### Children with an Overdue Screening

24%

0-10 11-20 21-30 31-40 41-50 51-60 61-70 71-80 81-90 91-100

### Children Chronically Absent in the Past Month

18%

0-10 11-20 21-30 31-40 41-50 51-60 61-70 71-80 81-90 91-100



# Is it possible to share data safely?

- Participation in the Collaborative is **purely voluntary**
- Participation should benefit **each member program** and the larger **Head Start community**
- The data and knowledge produced by the Collaborative is **only for supporting improvement**; it is not intended to punish, harm, or embarrass
- All data use within the Collaborative must meet the highest standards for **ethical data use**

# Is it possible to share data safely?

- The data **will not be used for surveillance or enforcement** of statutory or regulatory requirements
- **Privacy, security, and safety** are paramount considerations for all aspects of the Collaborative's development and administration
- The Collaborative should have **minimal barriers to entry** for new programs to participate

# Is it possible to share data safely?

## User Access

- Programs identify the staff authorized to access data within the system
- Users must execute a security pledge before receiving log-in credentials

## Other Uses

- Will be rigorously evaluated by the Council for approval/denial
- Will only use pseudonymous data unless PII is approved for matching
- Nothing can be published unless fully Disclosure-Proofed

# Future of the Data Lake

- Connect to **other MIS** (COPA, GoEngage, CAP60, etc.)
- Expand to **different types of data**
  - Child development data
  - Staff qualifications and PD
  - Family engagement data
  - Publicly-available data
- Develop many, many **new dashboards**



# How can I join?

- Sign **Data Contributor and Participation Agreement**
- Submit request to ChildPlus to **copy live reports into your program's instance** from NHSA's training instance
- Create a **limited-access user account** for the data lake
- **Kickoff** meeting
- **Ta-Da!**

Questions?

