DATA+AI SUMMIT
BY databricks

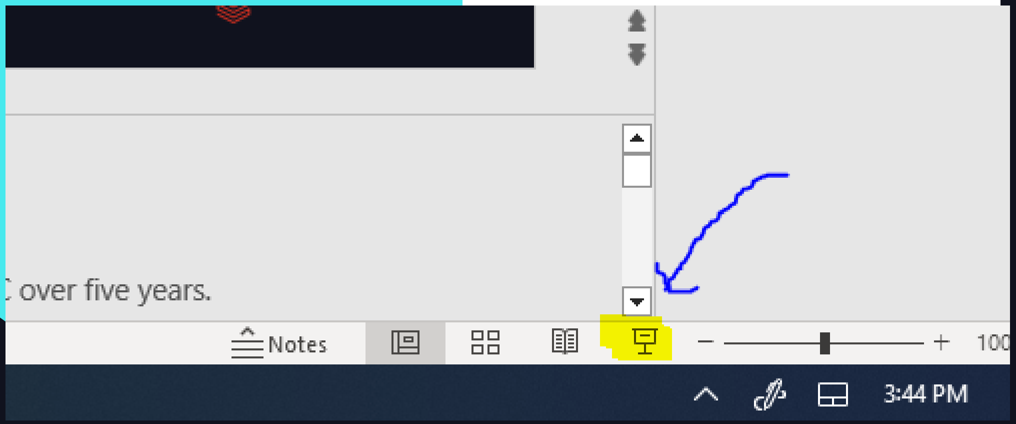
BIG DATA VISUALIZATION IN PUBLIC HEALTH

Exploring an accelerator methodology to build faster and more reliable visualization data products

Data + AI Summit 2024 – San Francisco
John Bowyer
Office of Public Health Data, Surveillance, and Technology
Centers for Disease Control and Prevention
Wed, June 12, 4:00 - 4:40 PM PDT

Per Matt Ritchey: Note on this slide that this is the template used by the conference and they have asked all presenters to use it. The comms group may comment on the appropriateness of using this template.

PLEASE REVIEW IN PRESENTATION MODE TO SEE ANIMATIONS



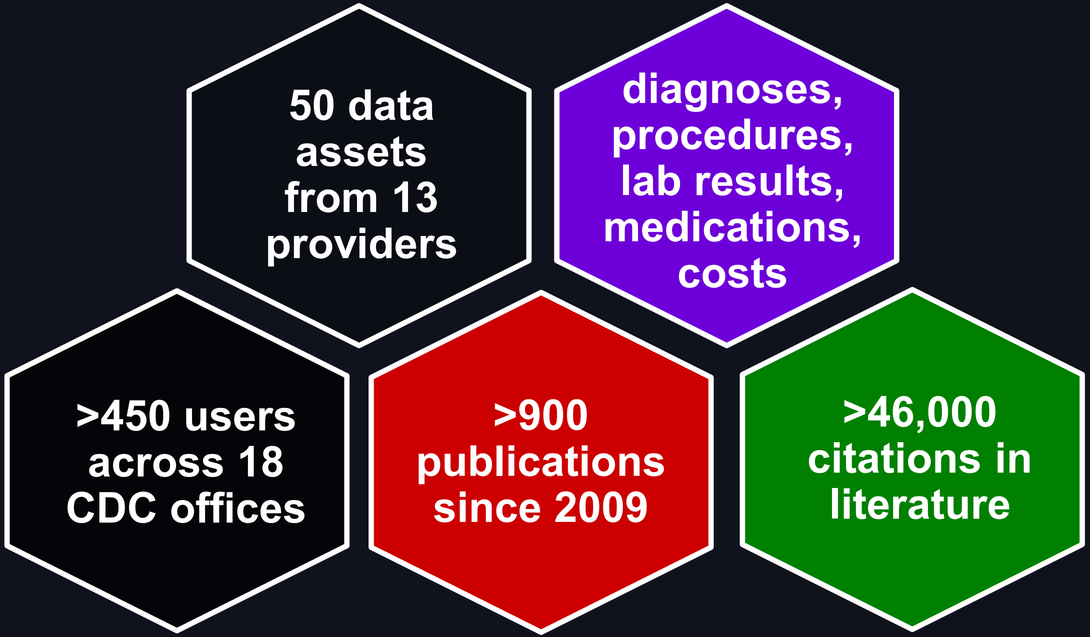
DISCLAIIMER

The findings and conclusions in this presentation are those of the author(s) and do not necessarily represent the views of the Centers for Disease Control and Prevention.

The information shared here is based on implementation experience of a CDC use case and is no way, shape, or form a direct endorsement of the product.

CDC DATA HUB HEALTHCARE DATA ECOSYSTEM

Program acquires, evaluates, manages, and supports the use of healthcare and related data sources for public health research and action—*A "one stop shop" for the agency.* Leverages Databricks within CDC's cloud environment for data management and intelligence capabilities



50 data assets from 13 providers

diagnoses, procedures, lab results, medications, costs

>450 users across 18 CDC offices

>900 publications since 2009

>46,000 citations in literature

PRESENTATION OBJECTIVE AND AGENDA

- Explore an accelerator methodology to build faster and more reliable visualization data products

Agenda

1. **Our Journey in Big Data Visualization in Public Health**

   - **Process** Modernization: Key changes and impacts
   - **Technology** Modernization: Advancements and implementations

2. **Example Dashboard:** Respiratory Virus - Performance Insights

3. **Conclusion:** Achievements and future directions

Hi, I’m John

1988-1992	🎓 Graduated UGA	BA, Fortran Punch Cards
1993-1995	🏭 Southern Company	Mainframe, DB2
1995-1997	🌲 Georgia Pacific	Web, SQL
1997-1998	📶 MCI	Web Portals, SQL
1998-2001	⚡ Clarus Corporation	SQL, .NET, .COM Startup
2001-2007	💊 Pfizer + Pharma (Ctr)	SQL, XML, BizTalk
2007-2018	🤝 Microsoft (Ctr)	Data Architect
2018-Present	🦠 CDC (Ctr -> FTE)	Data Architect

🐘 Tuscaloosa, Alabama, USA — If you know anything about the SEC , it's a rare thing for a UGA grad to find happiness in Crimson Tide country!

CDC DATA HUB

OFFICE OF PUBLIC HEALTH DATA, SURVEILLANCE, AND TECHNOLOGY

OUR JOURNEY IN BIG DATA VISUALIZATION IN PUBLIC HEALTH

Core components of our big data visualization accelerator methodology

Top 5 Process Improvements

- Common Visualization Gallery
- Standardized Agile User Stories and Recipes
- Machine Readable Requirements
- Standardized Data Product Visualizations
- Standardized Project Management

Top 5 Technology Advancements

- Common Data Models
- Standardized Data Conversion
- Data Quality Expectations and Synthea
- Data Product Catalog and Workflows
- Data Product Lifecycle

CDC DATA HUB PROCESS MODERNIZATION

OFFICE OF PUBLIC HEALTH DATA, SURVEILLANCE, AND TECHNOLOGY

CDC DATA HUB VISUALIZATION GALLERY

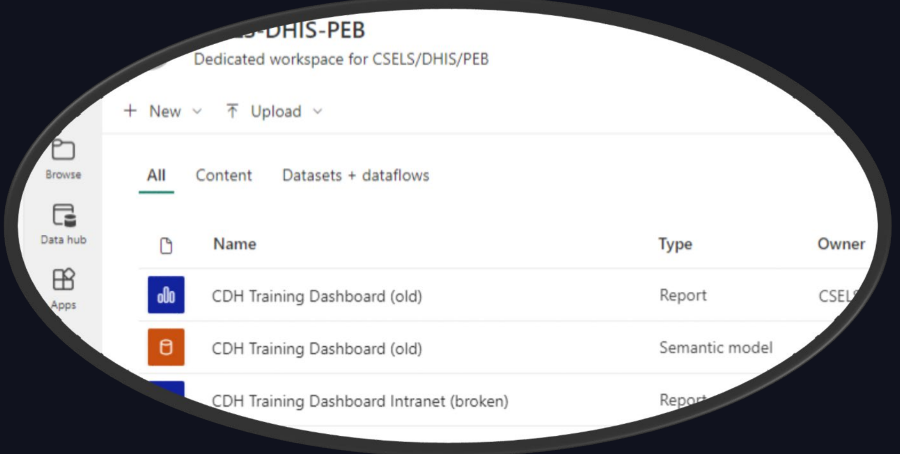
Process Improvement #1: Enhancing Discoverability and Engagement

- No Intranet Dashboards
- Unmonitored Usage
- Limited Offline Formats

Engagement Challenge

Befor
e

Disconnected Power BI Silos



…S-DHIS-PEB

Dedicated workspace for CSELS/DHIS/PEB

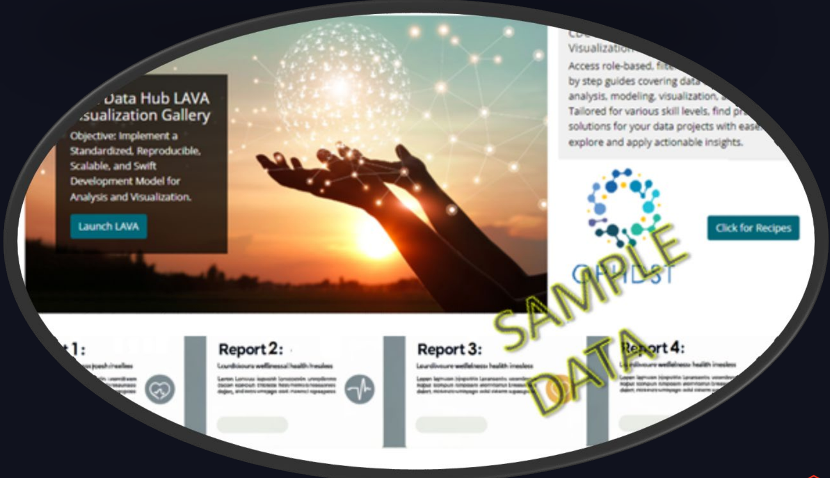
Name	Type	Owner
CDH Training Dashboard (old)	Report	CSEL
CDH Training Dashboard (old)	Semantic model	
CDH Training Dashboard Intranet (broken)	Report	

- Dashboard Intranet Listings
- Usage Monitoring & Optimization
- PDF, DOC, XLSX Export

Engagement Solution

Discoverable, cost-effective, integrated Power BI

After



Data Hub LAVA
sualization Gallery

Objective: Implement a Standardized, Reproducible, Scalable, and Swift Development Model for Analysis and Visualization.

Visualization

Access role-based, filt… by step guides covering data… analysis, modeling, visualization, … Tailored for various skill levels, find pr… solutions for your data projects with ease… explore and apply actionable insights.

- Higher User Adoption
- Better Data Accessibility
- Data-Driven Decisions

User Outcome

STANDARDIZED AGILE USER STORIES AND RECIPES

Process Improvement #2: Agile User Story and Recipe Authoring

- Old Method: 500-Page SOP
- Complex Information
- Poor Discoverability
- Expert/Non-expert Gap

Communication Challenge

- **Recipe Format:** Bite-sized, Visual
- **Training:** Regular Workshops

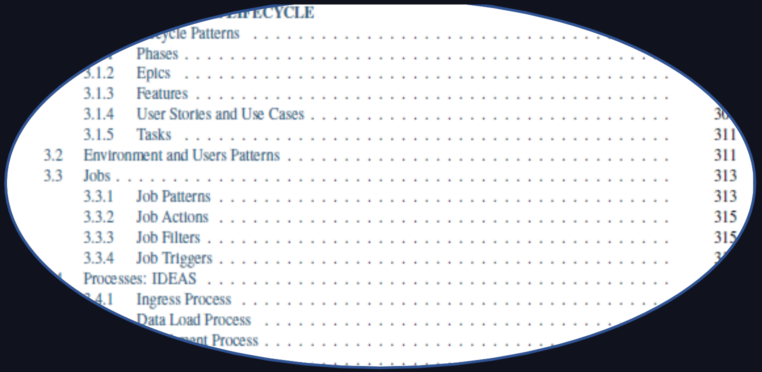
Communication Solution

- User-Crafted Stories
- Less Tech Dependency
- Better Process Ownership

User Impact

Before

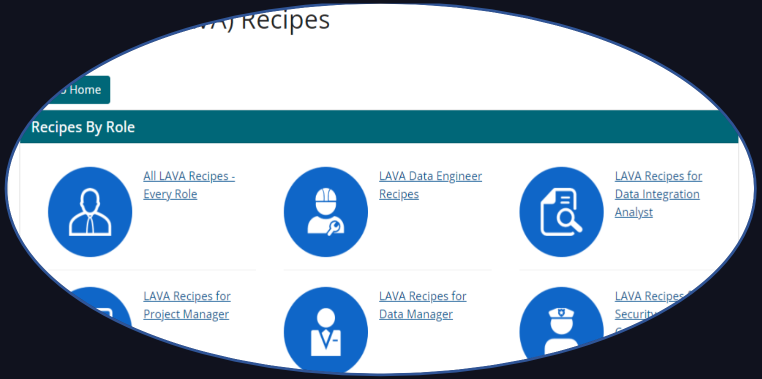
500+ Page Document



LIFECYCLE
cycle Patterns . . . .
Phases . . . .
3.1.2 Epics . . . .
3.1.3 Features . . . .
3.1.4 User Stories and Use Cases 30
3.1.5 Tasks 311
3.2 Environment and Users Patterns 311
3.3 Jobs 313
3.3.1 Job Patterns 313
3.3.2 Job Actions 315
3.3.3 Job Filters 315
3.3.4 Job Triggers 3
Processes: IDEAS . . . .
3.4.1 Ingress Process . . . .
Data Load Process . . . .
ment Process . . . .

After

Quick ~15-Minute Guided Recipes



A) Recipes

Home

Recipes By Role

All LAVA Recipes - Every Role

LAVA Data Engineer Recipes

LAVA Recipes for Data Integration Analyst

LAVA Recipes for Project Manager

LAVA Recipes for Data Manager

LAVA Recipes Securit

Hide

View

CROSSING THE CHASM: MACHINE READABLE REQUIREMENTS

Process Improvement #3: Requirements Authoring Tools

- **Old Methods:** Sticky Note Documentation
- **Issues:** Inaccessible and hardcoded

Requirement Challenge

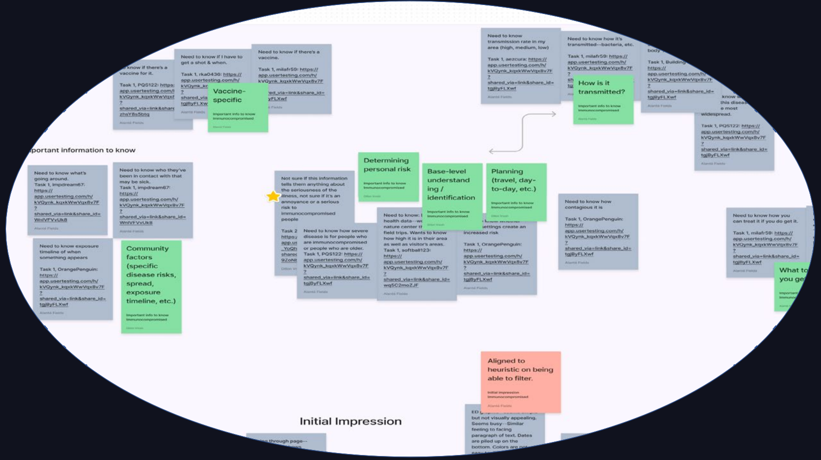
- Machine-Readable Specs
- Accessibility: Non-technical Reader
- Version Control

Requirement Solutions

- **Empowerment:** Autonomy with Easier Tools

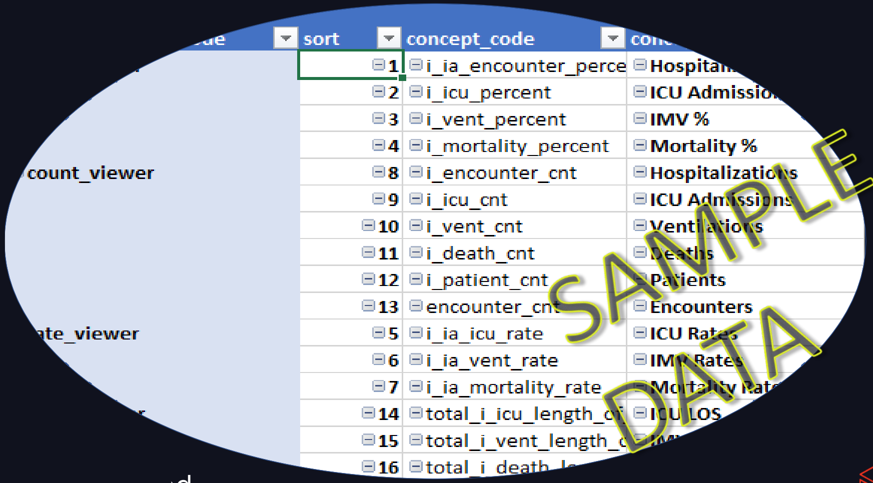
User Impact

Human-Readable Requirements



Vaccine-specific
How is it transmitted?
Determining personal risk
Base-level understanding / identification
Planning (travel, day-to-day, etc.)
Community factors (specific disease risks, spread, exposure timeline, etc.)
Aligned to heuristic on being able to filter.
Initial Impression

Human- and Machine-Readable Requirements



…de	sort	concept_code	conc…
	1	i_ia_encounter_perce	Hospital…
	2	i_icu_percent	ICU Admissio…
	3	i_vent_percent	IMV %
	4	i_mortality_percent	Mortality %
count_viewer	8	i_encounter_cnt	Hospitalizations
	9	i_icu_cnt	ICU Admissions
	10	i_vent_cnt	Ventilations
	11	i_death_cnt	Deaths
	12	i_patient_cnt	Patients
	13	encounter_cnt	Encounters
…ate_viewer	5	i_ia_icu_rate	ICU Rates
	6	i_ia_vent_rate	IMV Rates
	7	i_ia_mortality_rate	Mortality Rat…
	14	total_i_icu_length_o…	ICU LOS
	15	total_i_vent_length_…	IM…
	16	total_i_death_l…	

STANDARDIZED DATA PRODUCT VISUALIZATIONS

- Process Improvement #4: Streamlined and Consistent Dashboard Interface

Metric

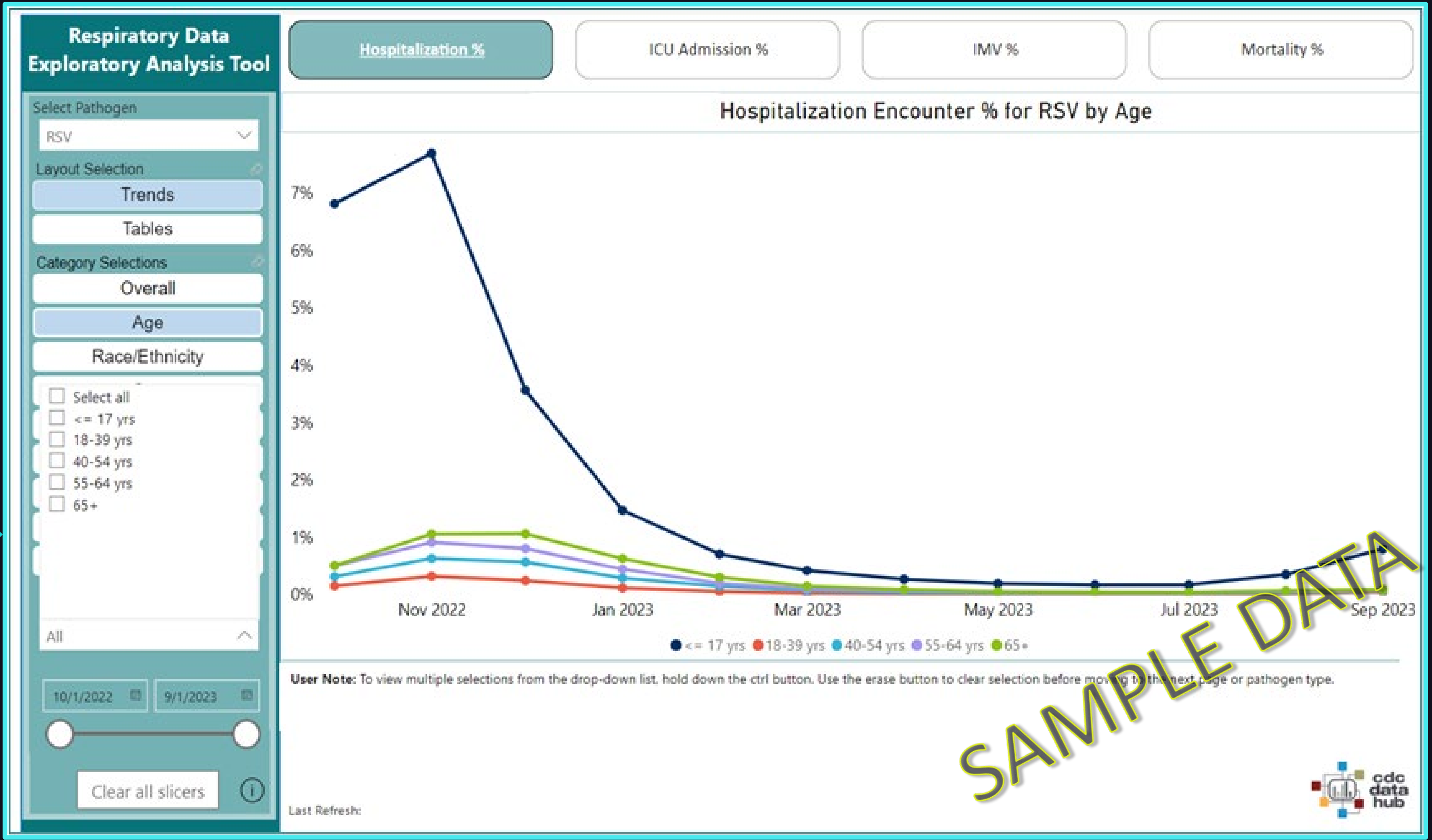
App

Subcategory

View By

View By Group

Date



Respiratory Data Exploratory Analysis Tool

Hospitalization % | ICU Admission % | IMV % | Mortality %

Select Pathogen

RSV

Layout Selection

Trends

Tables

Category Selections

Overall

Age

Race/Ethnicity

☐ Select all
☐ <= 17 yrs
☐ 18-39 yrs
☐ 40-54 yrs
☐ 55-64 yrs
☐ 65+

All

10/1/2022 9/1/2023

Clear all slicers

Hospitalization Encounter % for RSV by Age

User Note: To view multiple selections from the drop-down list, hold down the ctrl button. Use the erase button to clear selection before moving to the next page or pathogen type.

Last Refresh:

©

View

Hide

STANDARDIZED PROJECT MANAGEMENT

Process Improvement #5: Introducing Agile Plan Authoring Templates and Tools

- Details missing
- Started each project from

Planning Challenge

- Reused templates
- Reused naming
- Reused stories

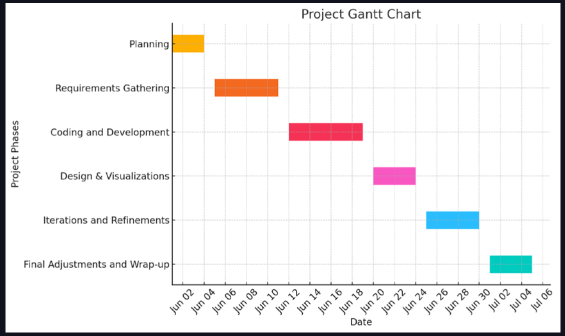
Planning Solution

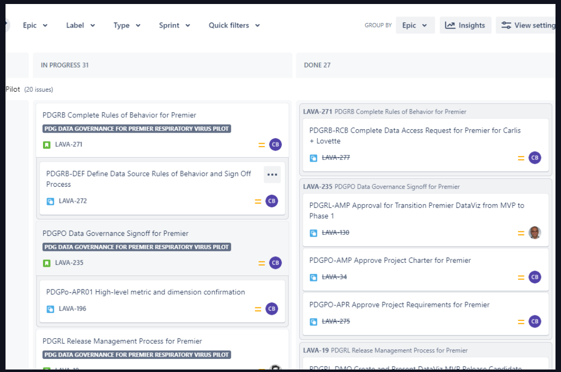
- Clarity
- Confidence
- Repeatability

User Impact

Befor e

Inconsistent Project Reporting ≠





IN PROGRESS 31

DONE 27

Pilot (20 issues)

PDGRB Complete Rules of Behavior for Premier
PDG DATA GOVERNANCE FOR PREMIER RESPIRATORY VIRUS PILOT
LAVA-271 CB

PDGRB-DEF Define Data Source Rules of Behavior and Sign Off Process
LAVA-272 CB

PDGPO Data Governance Signoff for Premier
PDG DATA GOVERNANCE FOR PREMIER RESPIRATORY VIRUS PILOT
LAVA-235 CB

PDGPo-APR01 High-level metric and dimension confirmation
LAVA-196 CB

PDGRL Release Management Process for Premier
PDG DATA GOVERNANCE FOR PREMIER RESPIRATORY VIRUS PILOT

LAVA-271 PDGRB Complete Rules of Behavior for Premier

PDGRB-RCB Complete Data Access Request for Premier for Carlis + Lovette
~~LAVA-277~~ CB

LAVA-235 PDGPO Data Governance Signoff for Premier

PDGRL-AMP Approval for Transition Premier DataViz from MVP to Phase 1
~~LAVA-130~~

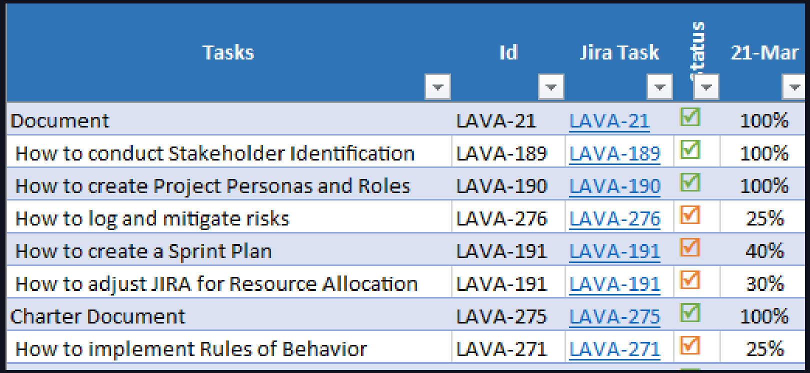
PDGPO-AMP Approve Project Charter for Premier
~~LAVA-34~~ CB

PDGPO-APR Approve Project Requirements for Premier
~~LAVA-275~~ CB

LAVA-19 PDGRL Release Management Process for Premier

One Source of Truth – Office Linked to Jira =

After



Tasks	Id	Jira Task	Status	21-Mar
Document	LAVA-21	LAVA-21	☑	100%
How to conduct Stakeholder Identification	LAVA-189	LAVA-189	☑	100%
How to create Project Personas and Roles	LAVA-190	LAVA-190	☑	100%
How to log and mitigate risks	LAVA-276	LAVA-276	☑	25%
How to create a Sprint Plan	LAVA-191	LAVA-191	☑	40%
How to adjust JIRA for Resource Allocation	LAVA-191	LAVA-191	☑	30%
Charter Document	LAVA-275	LAVA-275	☑	100%
How to implement Rules of Behavior	LAVA-271	LAVA-271	☑	25%

CDC DATA HUB TECHNOLOGY MODERNIZATION

OFFICE OF PUBLIC HEALTH DATA, SURVEILLANCE, AND TECHNOLOGY

ACCELERATING COMMON DATA MODELS

Technology Advancement #1: Multiple Subject Areas using Standardized Schemas

- Inconsistent and Wide Data Schemas
- Data Accessibility
- Privacy Concerns

Data Model
Challenges

- Adopting Common Schema
- Harmonizing Multiple Standards

Data Model
Advancements

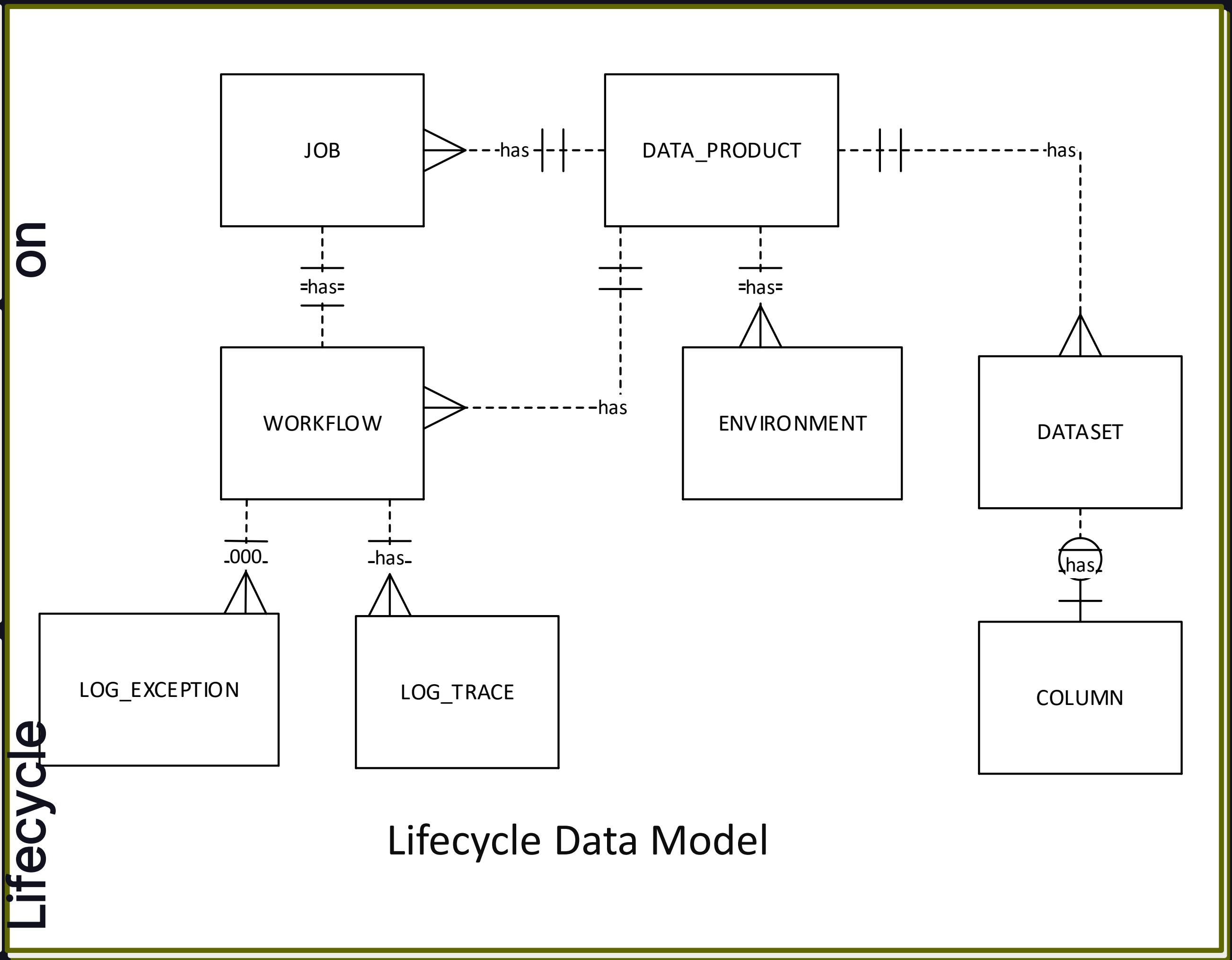
- Enhanced Data Usability
- Streamlined Data Analysis
- Governance Compliance

User
Impact

Visualizati on

Health

Data Lifecycle



Lifecycle Data Model

ACCELERATING STANDARDIZED DATA CONVERSION

Technology Advancement #2: Global Reference Data and Mapping Tools

- Custom Data Values
- Inconsistent and Custom Logic

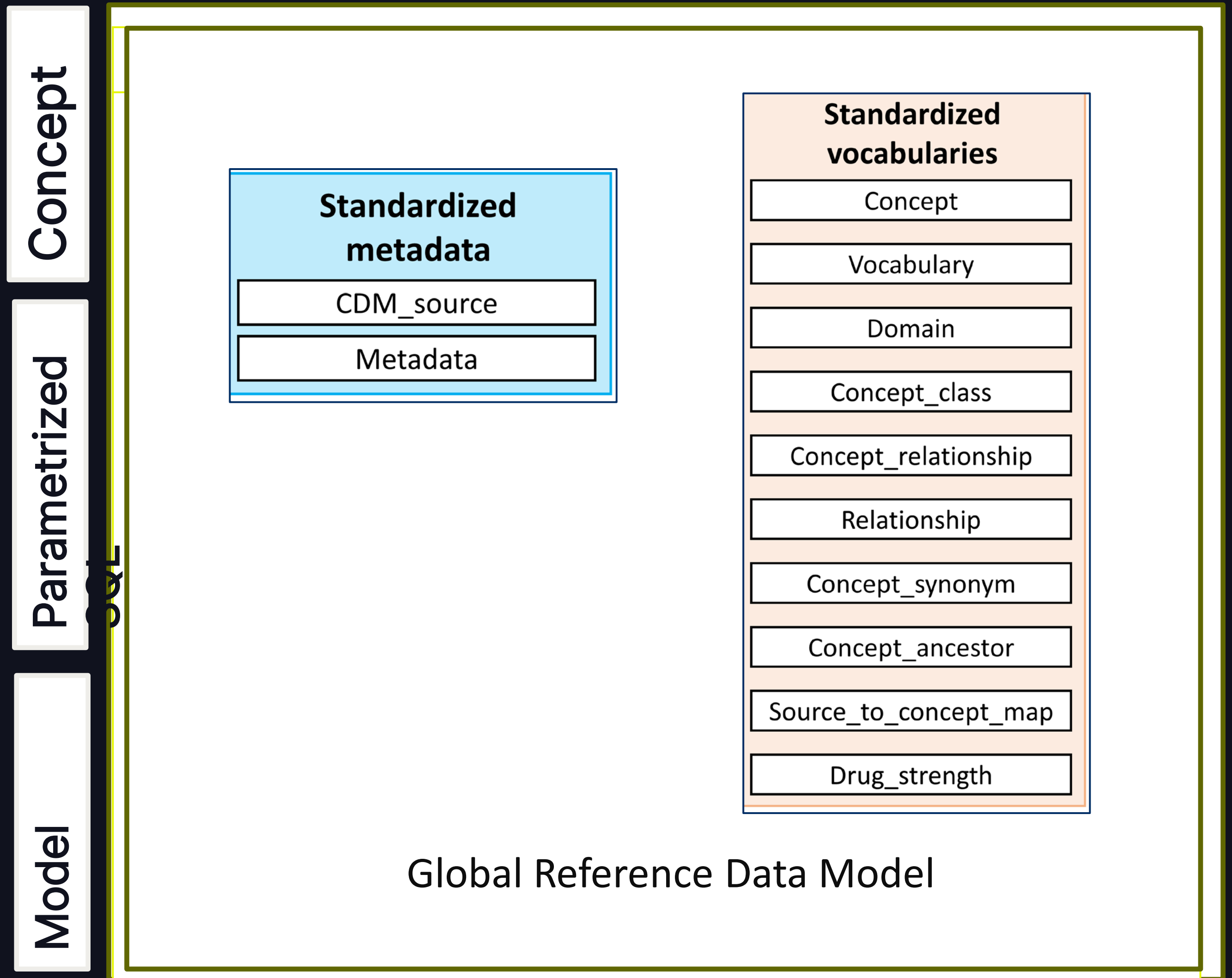
Conversion Challenges

- Universal Conversion Adapters
- Third Party Data Mapping Providers
- Parameterized SQL

Conversion Advancements

- Increased Research Collaboration
- Enhanced Data Reliability

User Impact



DATA QUALITY EXPECTATION TRACKING WITH TEST DATA

Technology Advancement #3: Data Governance Expectation Tracking Tools

- Restrictive Data Access
- Large-scale Dataset Complexity
- Lack of visibility into data quality

Quality Challenge

- Synthetic & Real-World Data Testing
- Schema Monitoring with Unity Catalogs

Quality Advancements

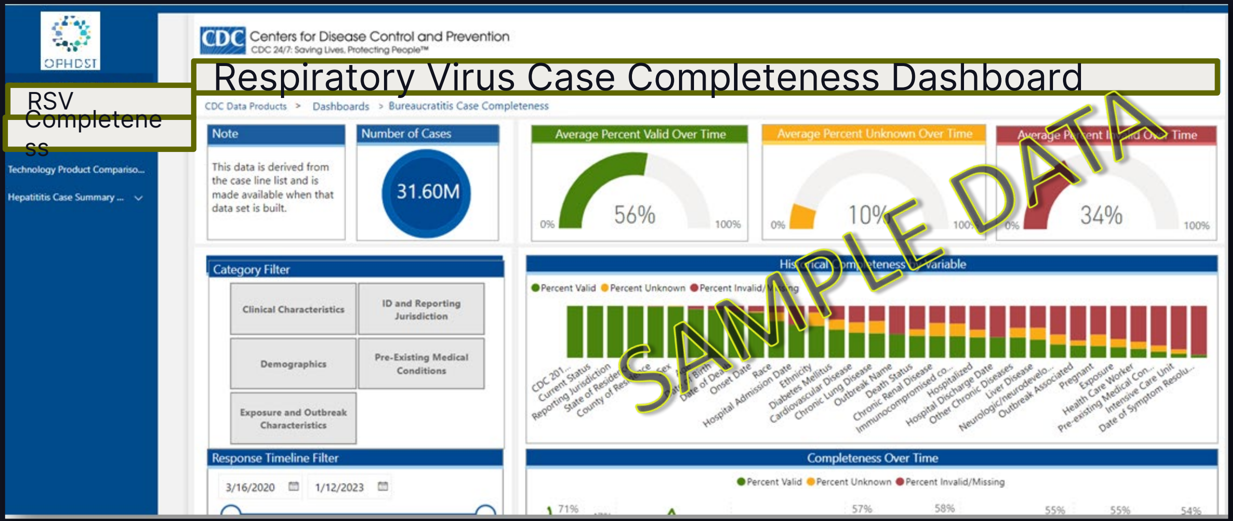
- Free, Immediate Data Access
- Enhanced Trust in Data
- Transparent Quality Insights

User Impact

Synthea Generated Test Data

Real-World Actual Test Data

Expectation Checks



CDC Centers for Disease Control and Prevention
CDC 24/7: Saving Lives, Protecting People™

Respiratory Virus Case Completeness Dashboard

CDC Data Products > Dashboards > Bureaucratitis Case Completeness

RSV Completeness

Technology Product Compariso...

Hepatitis Case Summary ...

Note

This data is derived from the case line list and is made available when that data set is built.

Number of Cases

31.60M

Average Percent Valid Over Time

Average Percent Unknown Over Time

Average Percent Invalid Over Time

Category Filter

Clinical Characteristics

ID and Reporting Jurisdiction

Demographics

Pre-Existing Medical Conditions

Exposure and Outbreak Characteristics

Historical Completeness by Variable

Response Timeline Filter

3/16/2020 1/12/2023

Completeness Over Time

a Quality Insights

DATA PRODUCT CATALOG AND WORKFLOWS

Technology Advancement #4: Prescriptive Guidance for Defining Data Products

- Lack of Guidance on data products
- No Analytical Cataloging
- Limited separation of billing and logging

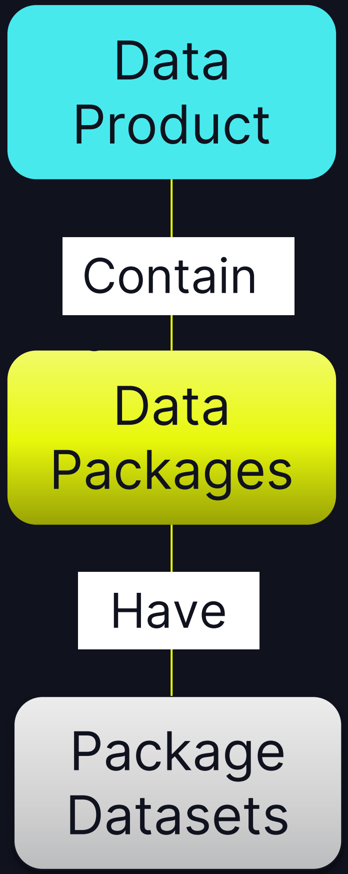
Data Product Challenges

- Established Governance
- API Calls: Pass data product ID
- Partitioned Logic: Shared and

Data Product Advancements

- Findable & Accessible
- Interoperable & Reusable (FAIR)
- Multi-tenant Billing & Logging

User Impact



- Comprehensive set of data & metadata
- Includes multiple Data Packages
- Often maps 1-1 to a database

- Collection of datasets/tables
- Metadata + Actual data
- Often maps 1-1 to a schema

- Metadata in dictionary format
- Often maps 1-1 to a table or file

STANDARDIZED DATA PRODUCT LIFECYCLE (SQL + PYTHON)

Technology Advancement #5: Data Lifecyle Management and Tracking

- Distinct processes per project for collecting data, grouping and basic analysis, and advanced analysis or ML techniques.

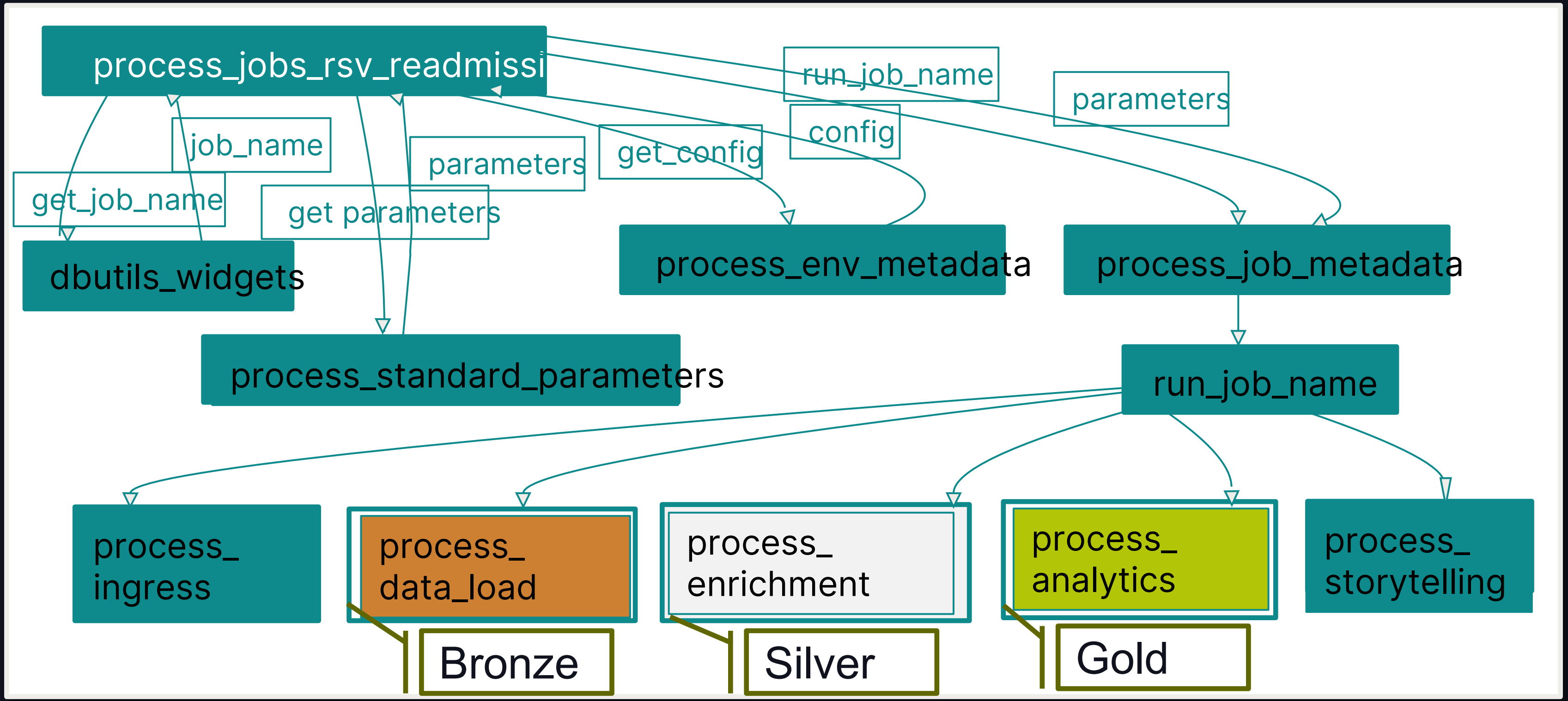
Data Management Challenges

- Shared data ingestion framework
- Standardized 5 step IDEAS process

Data Management Advancements

- Timely data for critical decision-making
- Reduced complexity in data

User Impact



Workflows and Orchestration

Medallion Architecture

CDC DATA HUB CONCLUSION AND FUTURE

OFFICE OF PUBLIC HEALTH DATA, SURVEILLANCE, AND TECHNOLOGY

DATA PROCESS

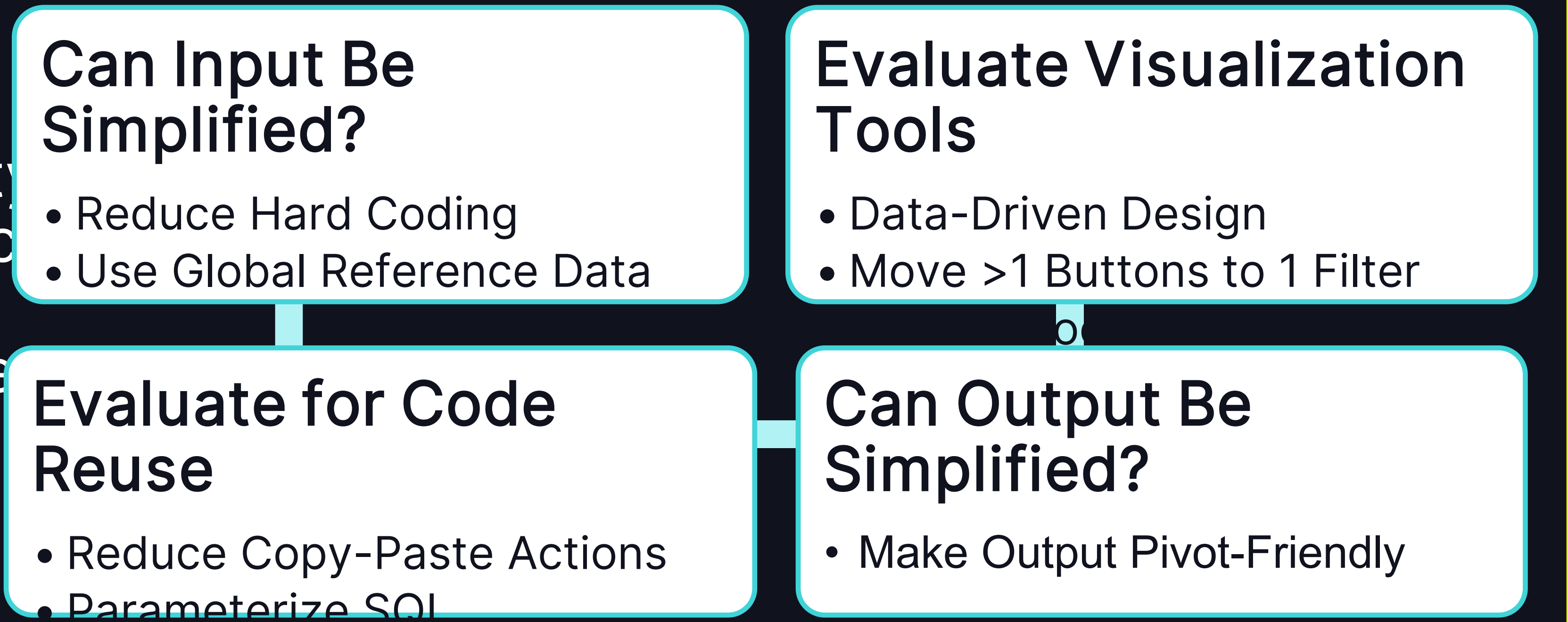
Data Hub Process Improvements

CDC Data Hub Results

- Increased data qualit
- Easier user experienc

Based on Simplifications

- Data Structure Reuse
- Code Reuse
- External Data Reuse



Can Input Be Simplified?

- Reduce Hard Coding
- Use Global Reference Data

Evaluate Visualization Tools

- Data-Driven Design
- Move >1 Buttons to 1 Filter

Evaluate for Code Reuse

- Reduce Copy-Paste Actions
- Parameterize SQL

Can Output Be Simplified?

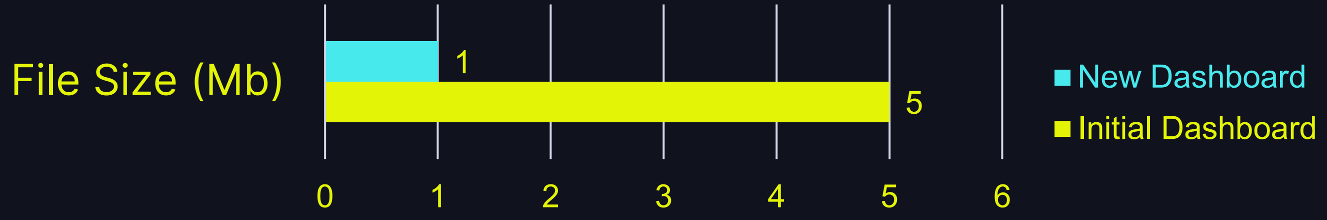
- Make Output Pivot-Friendly

DATA EFFICIENCY

Data Hub Efficiency Gains

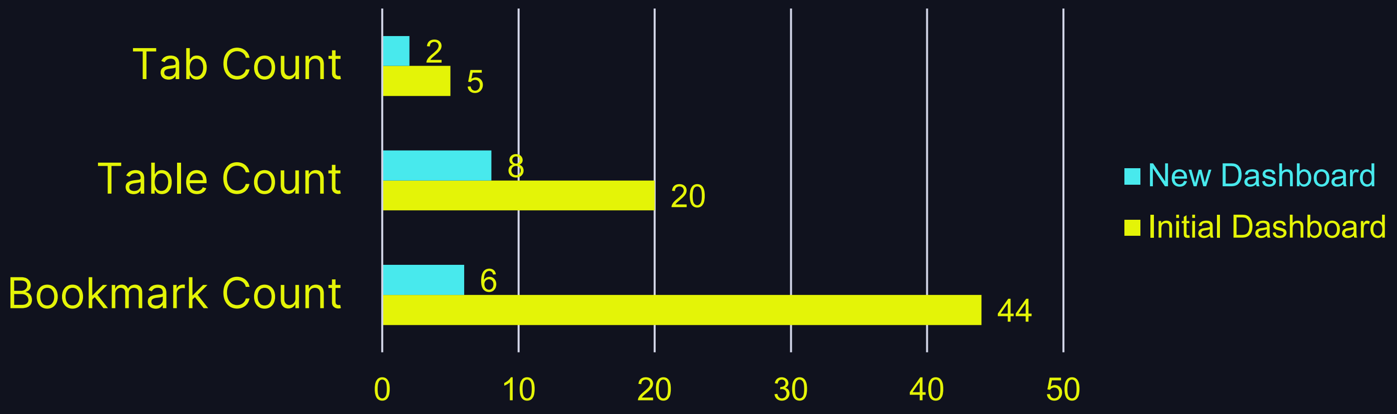
Despite expanding from 4 to over 20 metrics and expanding from 4 to over 10 dimensions, we achieved

SMALLER: Up to 5x File Size Reduction



As you can see, the new dashboard performances captures the same data (and then some) but with less processing power, allowing the dashboard to run faster and smoother than before.

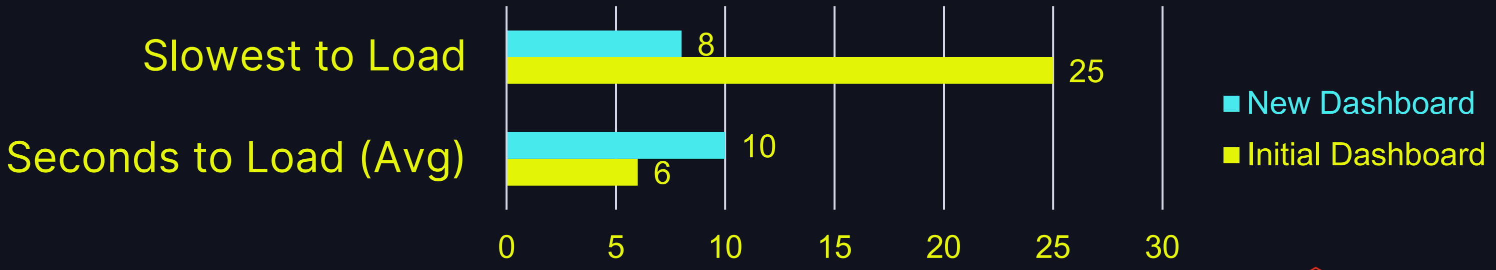
SIMPLER: Up to 7x Less Components



Over **10x**

performance increase per metric as demonstrated in the demo

FASTER: Up to 3x Time to Load Reduction



CONCLUSION: KEY TAKEAWAYS

Demonstrated an accelerator methodology to build faster and more reliable visualization data products

Our Current Status

Process Modernization

- Highlighted modernization solutions in CDC Data Hub visualization processes enhancing data handling efficiency.

Technology Modernization

- Discussed advancements in CDC Data Hub's technology stack improving overall performance.

Respiratory Virus Example Dashboard

- Reviewed findings showing up to a 10x improvement in data visualization performance.

FUTURE IDEAS FOR BIG DATA VISUALIZATION IN PUBLIC HEALTH

Core components of our big data visualization accelerator methodology

Future Potential

Future Opportunities

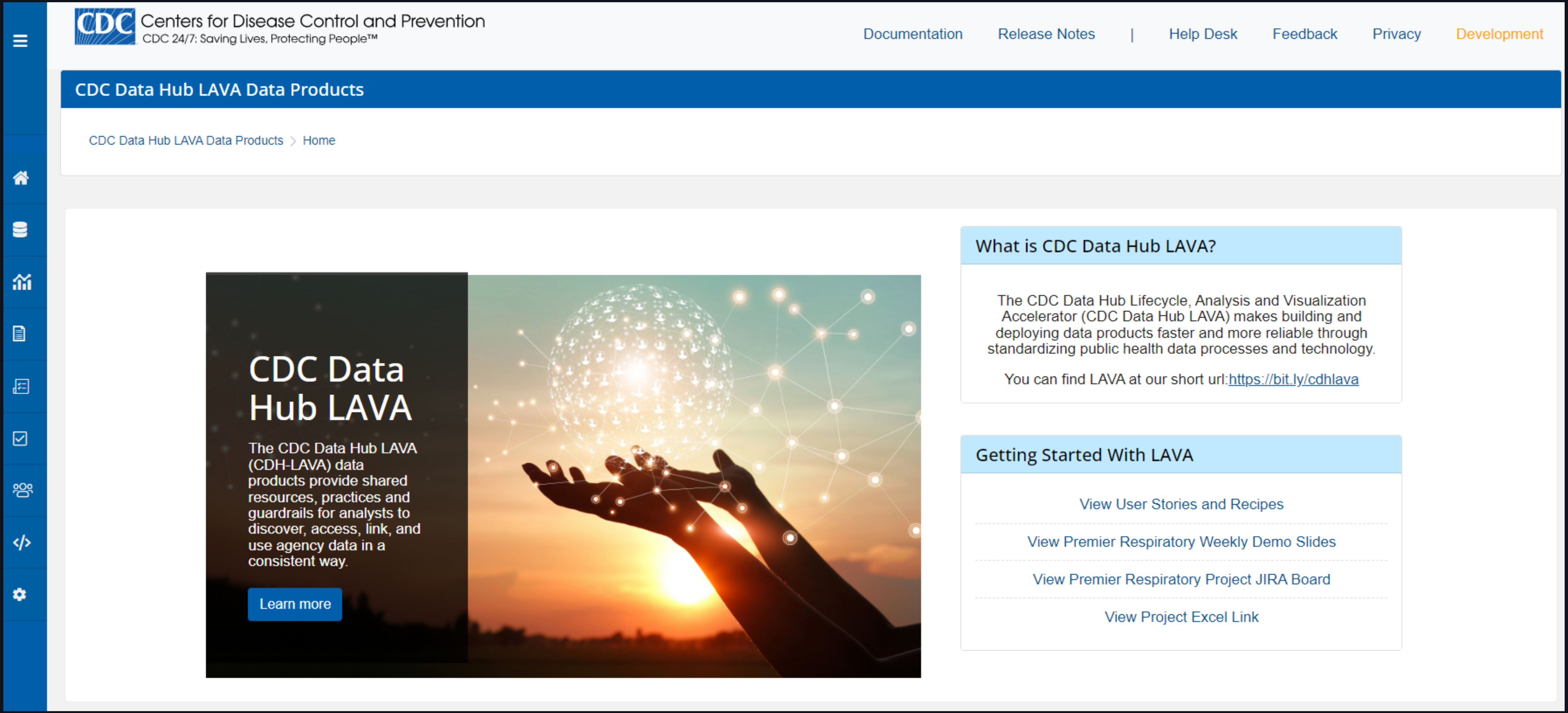
- Prescriptive Analytics
- Put Conclusion in Headlines
- Animate It (Charts and Maps)
- Model Explanations / Driver Analysis
- Make It About Them (Personalization)
- AI/ML and LLM with Forecasting
- Data Sharing

DEMONSTRATION

OFFICE OF PUBLIC HEALTH DATA, SURVEILLANCE, AND TECHNOLOGY

LAVA Portal

Welcome – Introduction to What is LAVA - https://bit.ly/cdhlava



CDC Centers for Disease Control and Prevention
CDC 24/7: Saving Lives, Protecting People™
Documentation
Release Notes
|
Help Desk
Feedback
Privacy
Development

CDC Data Hub LAVA Data Products

CDC Data Hub LAVA Data Products > Home

CDC Data Hub LAVA

The CDC Data Hub LAVA (CDH-LAVA) data products provide shared resources, practices and guardrails for analysts to discover, access, link, and use agency data in a consistent way.

What is CDC Data Hub LAVA?

The CDC Data Hub Lifecycle, Analysis and Visualization Accelerator (CDC Data Hub LAVA) makes building and deploying data products faster and more reliable through standardizing public health data processes and technology.

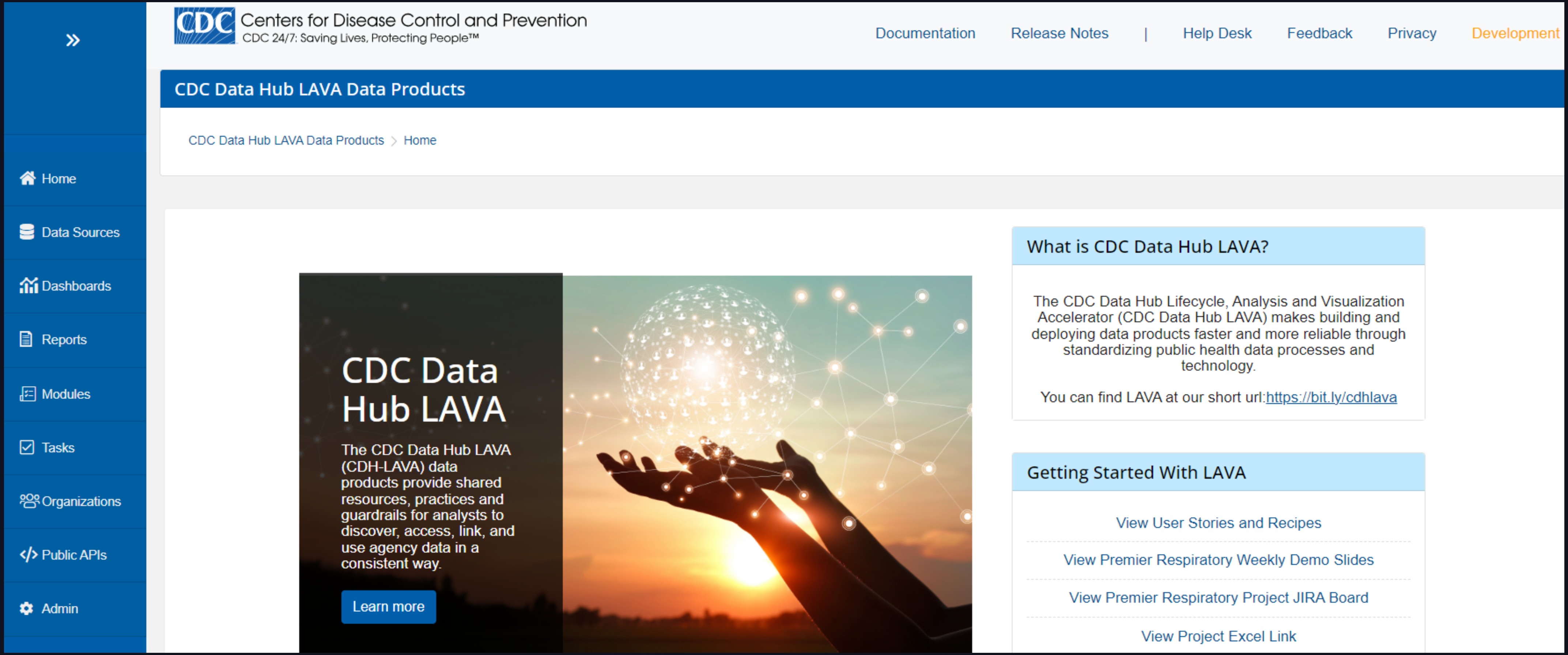
You can find LAVA at our short url:https://bit.ly/cdhlava

Getting Started With LAVA

View User Stories and Recipes
View Premier Respiratory Weekly Demo Slides
View Premier Respiratory Project JIRA Board
View Project Excel Link

LAVA Portal

Portal Sidebar



CDC Data Hub LAVA Data Products

CDC Data Hub LAVA Data Products > Home

CDC Data Hub LAVA

The CDC Data Hub LAVA (CDH-LAVA) data products provide shared resources, practices and guardrails for analysts to discover, access, link, and use agency data in a consistent way.

What is CDC Data Hub LAVA?

The CDC Data Hub Lifecycle, Analysis and Visualization Accelerator (CDC Data Hub LAVA) makes building and deploying data products faster and more reliable through standardizing public health data processes and technology.

You can find LAVA at our short url:https://bit.ly/cdhlava

Getting Started With LAVA

View User Stories and Recipes

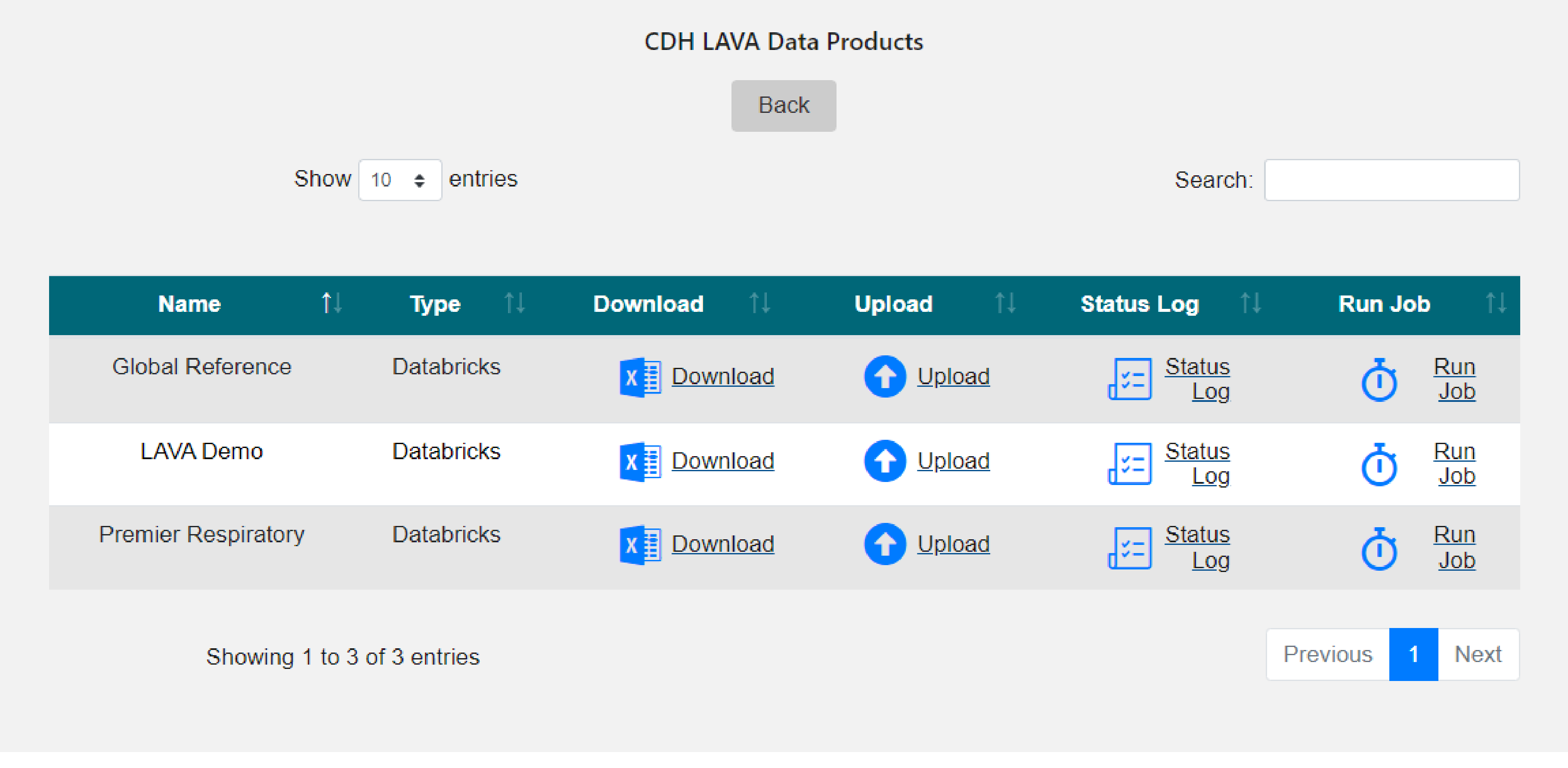
View Premier Respiratory Weekly Demo Slides

View Premier Respiratory Project JIRA Board

View Project Excel Link

LAVA Portal

Data sources



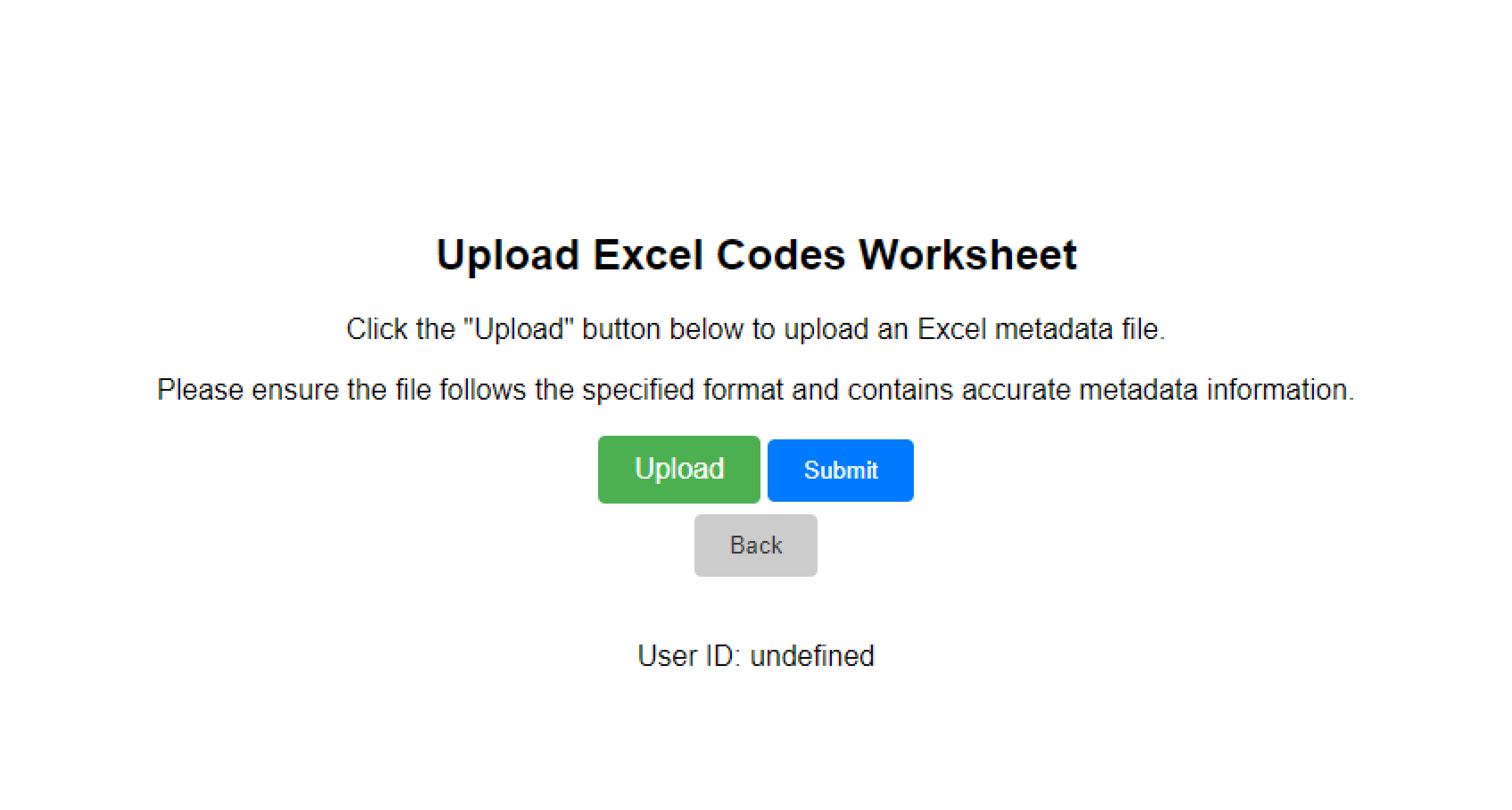
CDH LAVA Data Products

Name	Type	Download	Upload	Status Log	Run Job
Global Reference	Databricks	Download	Upload	Status Log	Run Job
LAVA Demo	Databricks	Download	Upload	Status Log	Run Job
Premier Respiratory	Databricks	Download	Upload	Status Log	Run Job

Showing 1 to 3 of 3 entries

LAVA Portal

Upload



Upload Excel Codes Worksheet

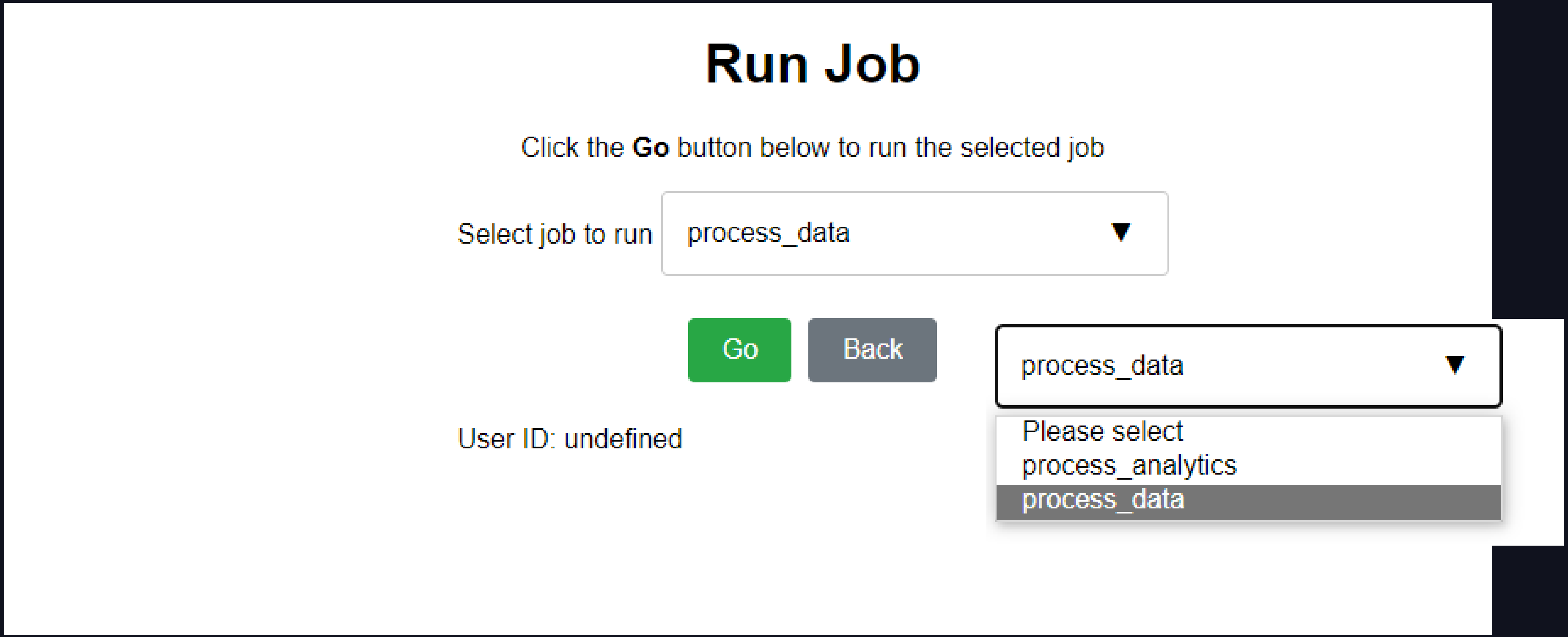
Click the "Upload" button below to upload an Excel metadata file.

Please ensure the file follows the specified format and contains accurate metadata information.

User ID: undefined

LAVA Portal

Run Jobs
Online Option



Run Job

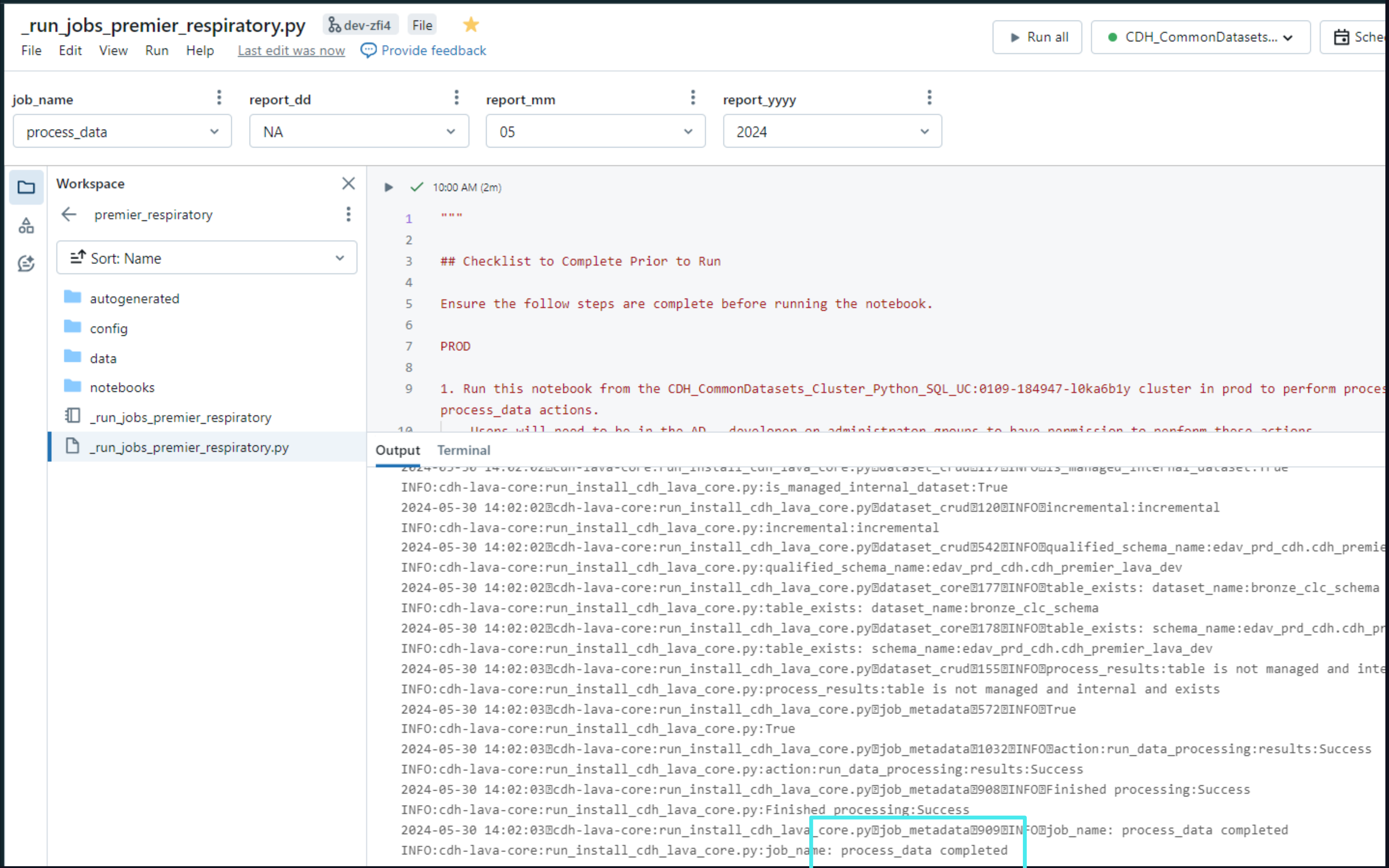
Click the **Go** button below to run the selected job

Select job to run: process_data ▼

User ID: undefined

LAVA Portal

Run Jobs
Notebook Option



job_name: process_data

report_dd: NA

report_mm: 05

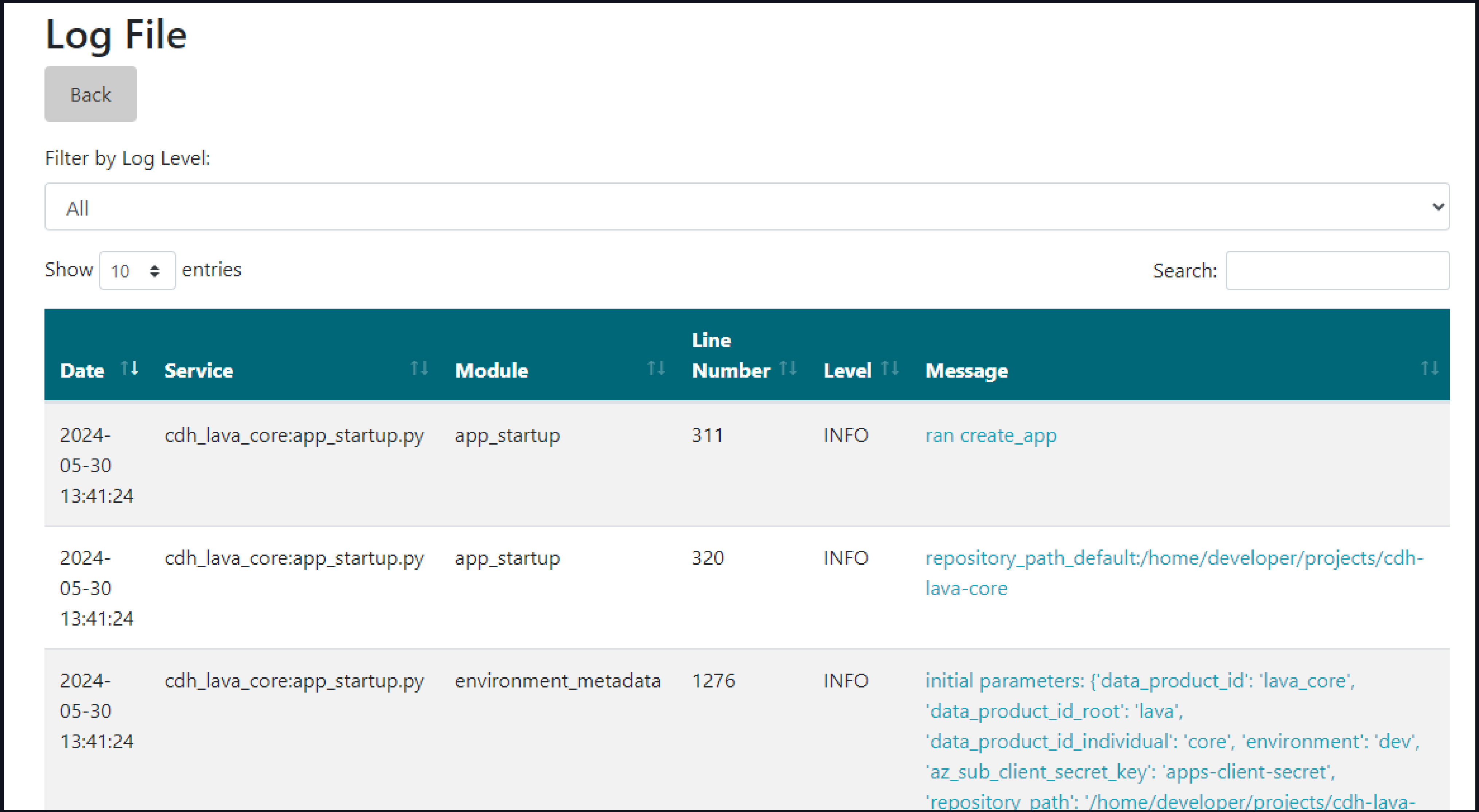
report_yyyy: 2024

```
"""

## Checklist to Complete Prior to Run

Ensure the follow steps are complete before running the notebook.

PROD

1. Run this notebook from the CDH_CommonDatasets_Cluster_Python_SQL_UC:0109-184947-l0ka6b1y cluster in prod to perform proces
process_data actions.
   Users will need to be in the AD  developer or administrator groups to have permission to perform these actions
```

```
2024-05-30 14:02:02cdh-lava-core:run_install_cdh_lava_core.pydataset_crud117INFOis_managed_internal_dataset:True
INFO:cdh-lava-core:run_install_cdh_lava_core.py:is_managed_internal_dataset:True
2024-05-30 14:02:02cdh-lava-core:run_install_cdh_lava_core.pydataset_crud120INFOincremental:incremental
INFO:cdh-lava-core:run_install_cdh_lava_core.py:incremental:incremental
2024-05-30 14:02:02cdh-lava-core:run_install_cdh_lava_core.pydataset_crud542INFOqualified_schema_name:edav_prd_cdh.cdh_premie
INFO:cdh-lava-core:run_install_cdh_lava_core.py:qualified_schema_name:edav_prd_cdh.cdh_premier_lava_dev
2024-05-30 14:02:02cdh-lava-core:run_install_cdh_lava_core.pydataset_core177INFOtable_exists: dataset_name:bronze_clc_schema
INFO:cdh-lava-core:run_install_cdh_lava_core.py:table_exists: dataset_name:bronze_clc_schema
2024-05-30 14:02:02cdh-lava-core:run_install_cdh_lava_core.pydataset_core178INFOtable_exists: schema_name:edav_prd_cdh.cdh_pr
INFO:cdh-lava-core:run_install_cdh_lava_core.py:table_exists: schema_name:edav_prd_cdh.cdh_premier_lava_dev
2024-05-30 14:02:03cdh-lava-core:run_install_cdh_lava_core.pydataset_crud155INFOprocess_results:table is not managed and inte
INFO:cdh-lava-core:run_install_cdh_lava_core.py:process_results:table is not managed and internal and exists
2024-05-30 14:02:03cdh-lava-core:run_install_cdh_lava_core.pyjob_metadata572INFOTrue
INFO:cdh-lava-core:run_install_cdh_lava_core.py:True
2024-05-30 14:02:03cdh-lava-core:run_install_cdh_lava_core.pyjob_metadata1032INFOaction:run_data_processing:results:Success
INFO:cdh-lava-core:run_install_cdh_lava_core.py:action:run_data_processing:results:Success
2024-05-30 14:02:03cdh-lava-core:run_install_cdh_lava_core.pyjob_metadata908INFOFinished processing:Success
INFO:cdh-lava-core:run_install_cdh_lava_core.py:Finished processing:Success
2024-05-30 14:02:03cdh-lava-core:run_install_cdh_lava_core.pyjob_metadata909INFOjob_name: process_data completed
INFO:cdh-lava-core:run_install_cdh_lava_core.py:job_name: process_data completed
```

LAVA Portal

Status Log



Log File

Filter by Log Level:

All

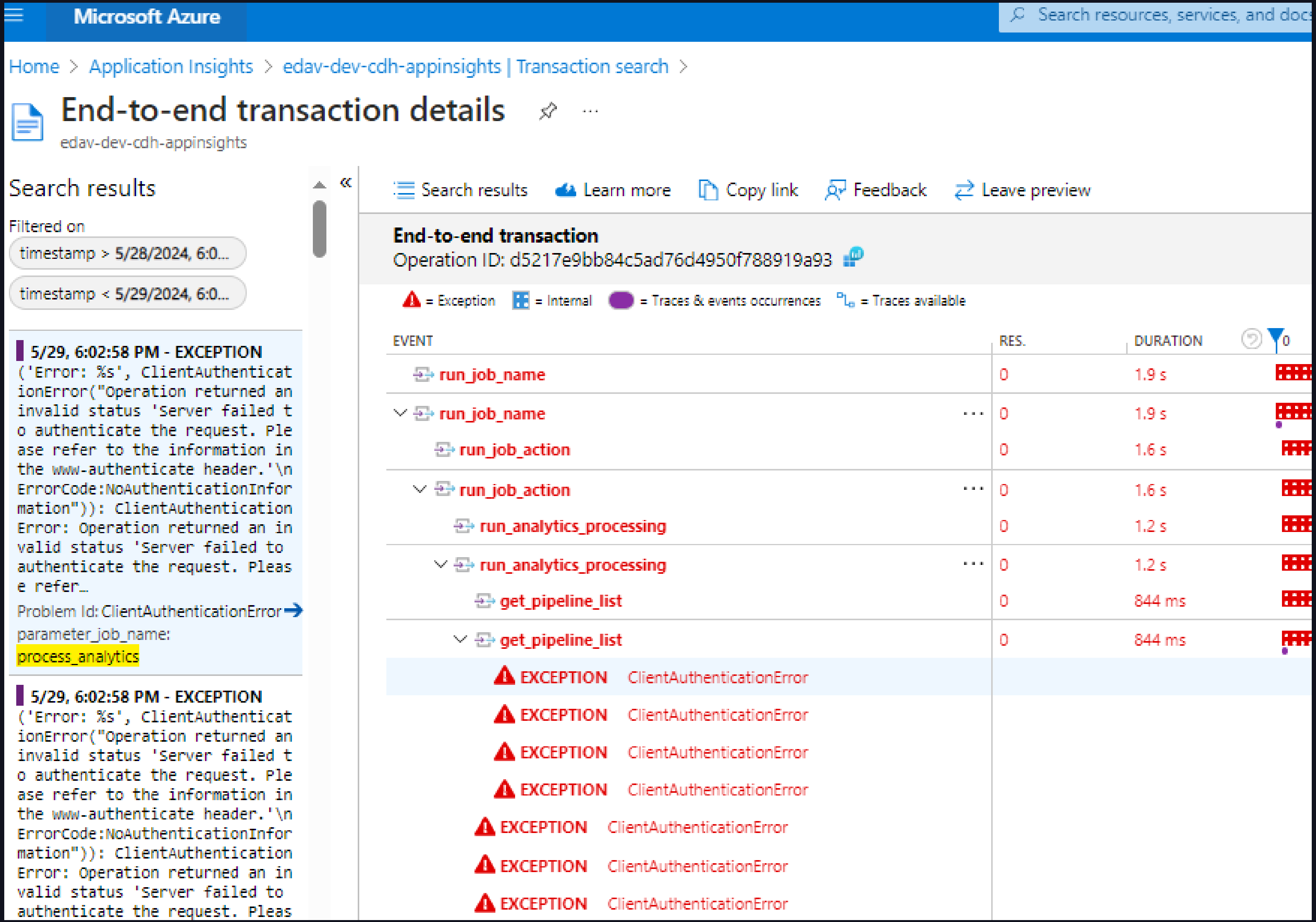
Show 10 entries

Search:

Date	Service	Module	Line Number	Level	Message
2024-05-30 13:41:24	cdh_lava_core:app_startup.py	app_startup	311	INFO	ran create_app
2024-05-30 13:41:24	cdh_lava_core:app_startup.py	app_startup	320	INFO	repository_path_default:/home/developer/projects/cdh-lava-core
2024-05-30 13:41:24	cdh_lava_core:app_startup.py	environment_metadata	1276	INFO	initial parameters: {'data_product_id': 'lava_core', 'data_product_id_root': 'lava', 'data_product_id_individual': 'core', 'environment': 'dev', 'az_sub_client_secret_key': 'apps-client-secret', 'repository_path': '/home/developer/projects/cdh-lava-

LAVA Portal

Log Search
Application Insights



Home > Application Insights > edav-dev-cdh-appinsights | Transaction search >

End-to-end transaction details

edav-dev-cdh-appinsights

Search results

Filtered on

timestamp > 5/28/2024, 6:0...

timestamp < 5/29/2024, 6:0...

5/29, 6:02:58 PM - EXCEPTION

```
('Error: %s', ClientAuthenticat
ionError("Operation returned an
invalid status 'Server failed t
o authenticate the request. Ple
ase refer to the information in
the www-authenticate header.'\n
ErrorCode:NoAuthenticationInfor
mation")): ClientAuthentication
Error: Operation returned an in
valid status 'Server failed to
authenticate the request. Pleas
e refer...
```

Problem Id: ClientAuthenticationError →

parameter_job_name:
process_analytics

5/29, 6:02:58 PM - EXCEPTION

```
('Error: %s', ClientAuthenticat
ionError("Operation returned an
invalid status 'Server failed t
o authenticate the request. Ple
ase refer to the information in
the www-authenticate header.'\n
ErrorCode:NoAuthenticationInfor
mation")): ClientAuthentication
Error: Operation returned an in
valid status 'Server failed to
authenticate the request. Pleas
```

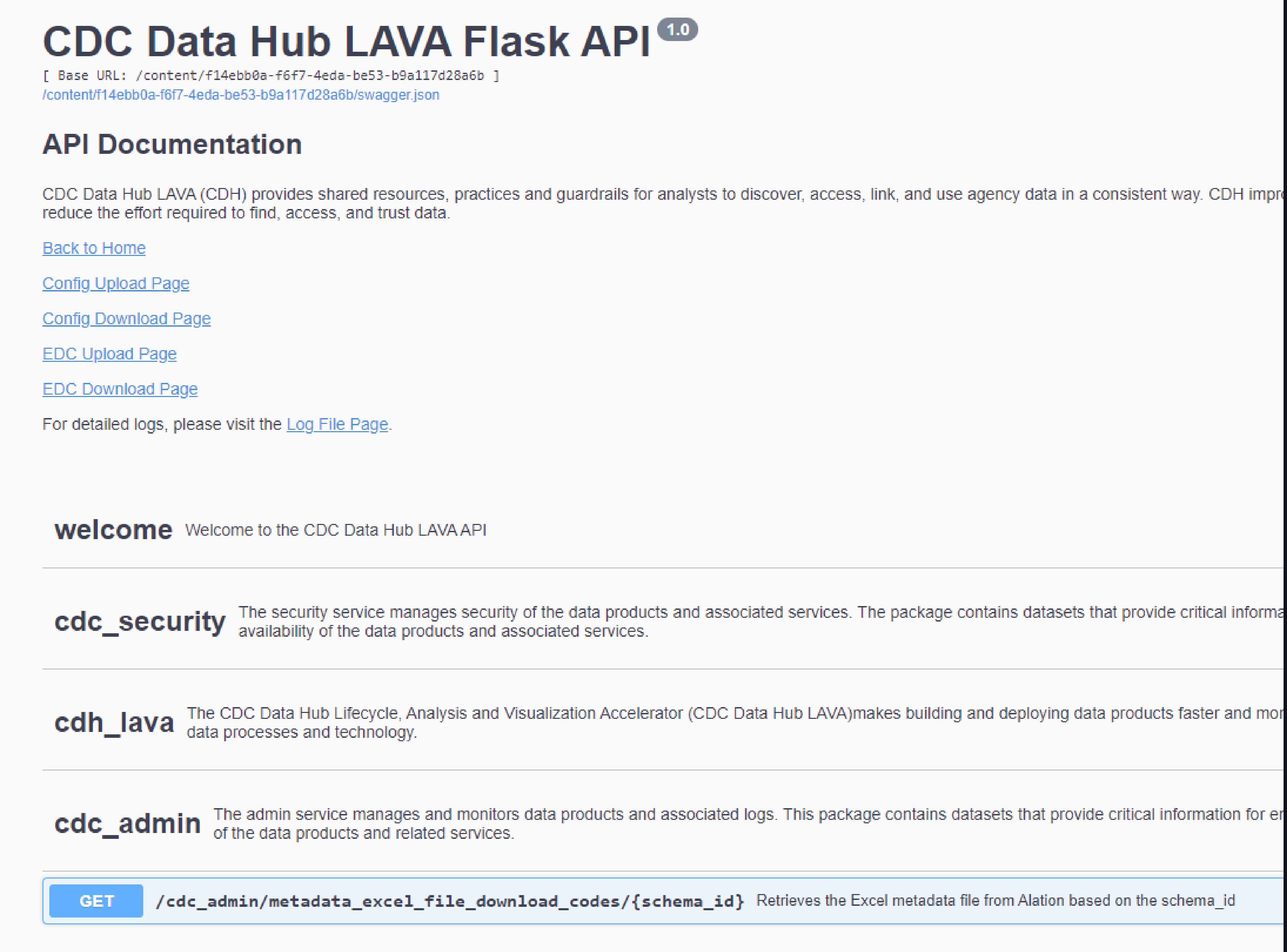
End-to-end transaction
Operation ID: d5217e9bb84c5ad76d4950f788919a93

▲ = Exception ▦ = Internal ● = Traces & events occurrences ⊡ = Traces available

EVENT	RES.	DURATION
run_job_name	0	1.9 s
run_job_name	0	1.9 s
run_job_action	0	1.6 s
run_job_action	0	1.6 s
run_analytics_processing	0	1.2 s
run_analytics_processing	0	1.2 s
get_pipeline_list	0	844 ms
get_pipeline_list	0	844 ms
▲ EXCEPTION ClientAuthenticationError		
▲ EXCEPTION ClientAuthenticationError		
▲ EXCEPTION ClientAuthenticationError		
▲ EXCEPTION ClientAuthenticationError		
▲ EXCEPTION ClientAuthenticationError		
▲ EXCEPTION ClientAuthenticationError		
▲ EXCEPTION ClientAuthenticationError		

LAVA Flask

Open API Specificatio



CDC Data Hub LAVA Flask API 1.0

```
[ Base URL: /content/f14ebb0a-f6f7-4eda-be53-b9a117d28a6b ]
```

/content/f14ebb0a-f6f7-4eda-be53-b9a117d28a6b/swagger.json

API Documentation

CDC Data Hub LAVA (CDH) provides shared resources, practices and guardrails for analysts to discover, access, link, and use agency data in a consistent way. CDH impro
reduce the effort required to find, access, and trust data.

Back to Home

Config Upload Page

Config Download Page

EDC Upload Page

EDC Download Page

For detailed logs, please visit the Log File Page.

welcome Welcome to the CDC Data Hub LAVA API

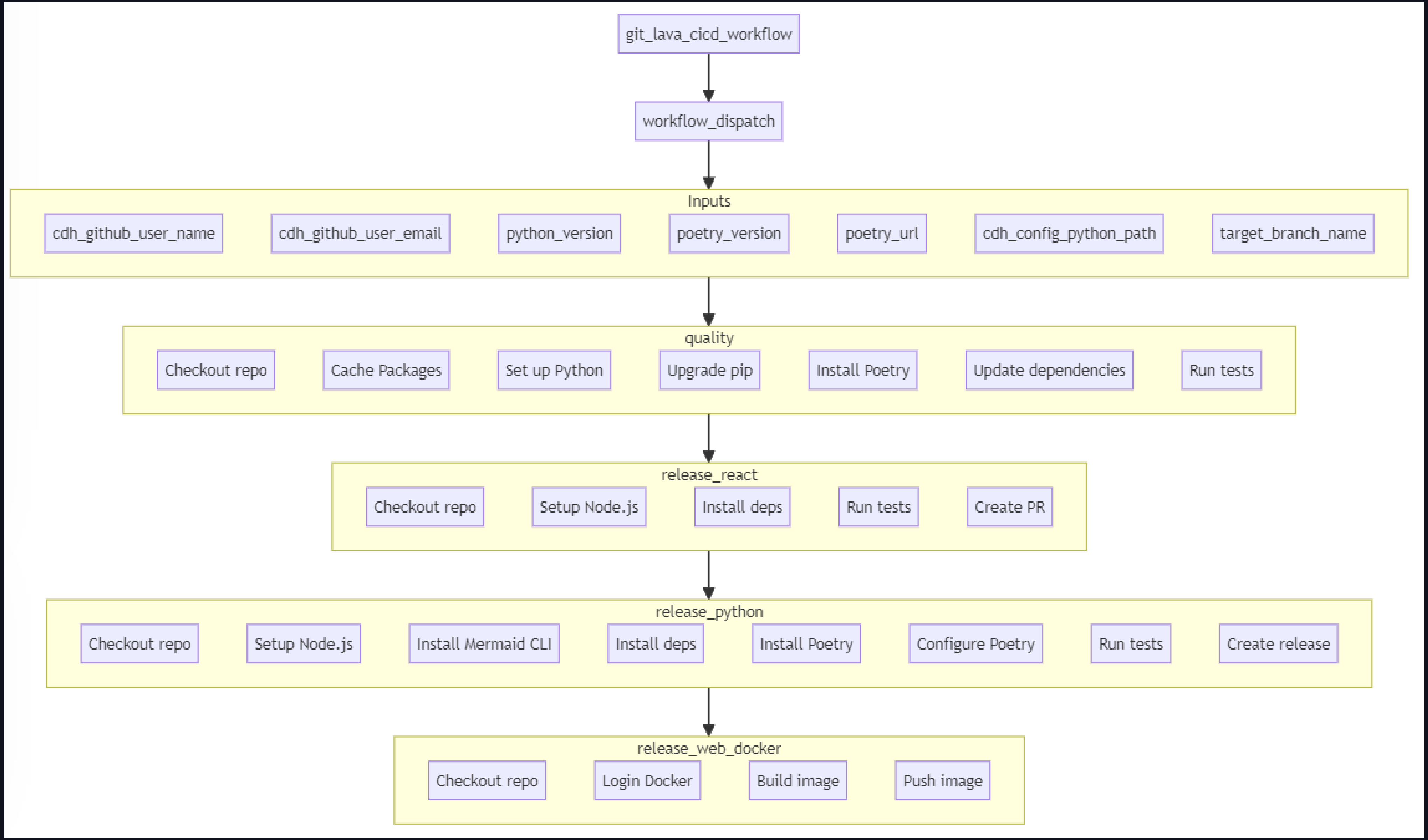
cdc_security The security service manages security of the data products and associated services. The package contains datasets that provide critical informa
availability of the data products and associated services.

cdh_lava The CDC Data Hub Lifecycle, Analysis and Visualization Accelerator (CDC Data Hub LAVA)makes building and deploying data products faster and mor
data processes and technology.

cdc_admin The admin service manages and monitors data products and associated logs. This package contains datasets that provide critical information for e
of the data products and related services.

GET `/cdc_admin/metadata_excel_file_download_codes/{schema_id}` Retrieves the Excel metadata file from Alation based on the schema_id

LAVA DEVSECOPS



git_lava_cicd_workflow

workflow_dispatch

Inputs
- cdh_github_user_name
- cdh_github_user_email
- python_version
- poetry_version
- poetry_url
- cdh_config_python_path
- target_branch_name

quality
- Checkout repo
- Cache Packages
- Set up Python
- Upgrade pip
- Install Poetry
- Update dependencies
- Run tests

release_react
- Checkout repo
- Setup Node.js
- Install deps
- Run tests
- Create PR

release_python
- Checkout repo
- Setup Node.js
- Install Mermaid CLI
- Install deps
- Install Poetry
- Configure Poetry
- Run tests
- Create release

release_web_docker
- Checkout repo
- Login Docker
- Build image
- Push image

WHAT PROBLEMS ARE WE SOLVING

Top 10 Pitfalls that challenge data product efficiency and quality

Dashboard Drifters

- Can't find dashboards.

Death By Docs

- Documentation is too big and unwieldy

Vague Vortex

- Requirements are not translatable to machines.

Molasses Matrix

- Report navigation is slow and difficult

Phantom Facts

- Planning summaries are not tied to details.

The Pancake Stack

- Inconsistent and wide data schemas

Conversion Chaos

- Inconsistent conversion logic.

Opaque Oracle

- Undefined quality expectations

Metadata Mystery

- No analytical cataloging.

Reinvention Rocket

- Little to no ETL reuse.

OUR JOURNEY IN BIG DATA VISUALIZATION IN PUBLIC HEALTH

Core components of our big data visualization accelerator methodology

Top 5 Process Improvements

- Common Visualization Gallery
- Standardized Agile User Stories and Recipes
- Machine Readable Requirements
- Standardized Data Product Visualizations
- Standardized Project Management

Top 5 Technology Advancements

- Common Data Models
- Standardized Data Conversion
- Data Quality Expectations and Synthea
- Data Product Catalog and Workflows
- Data Product Lifecycle

> “We can complete our 20+ page Feedback Packet in 10 minutes....”

-Matt Cole
Epidemiologist and Data Scientist
First Project Completed in 2021

“The new process is 1000x faster for creating standard pivots.”

-Stacey Adjei
Epidemiologist and Data Scientist
Latest Project Completed in 2024