University Analytics

The Maturation of Data Workflows for THECB Reporting

Presenters:

Bobby Childress

Jeremy Davis

Sasha Naourizbaeva

Tarak Myneni

THE UNIVERSITY OF TEXAS AT ARLINGTON

Setting the Stage
Size and Scope of the UTA

- **The University of Texas at Arlington**
  - R1, minority-serving institution with 43,963 CB-reported students enrolled in Fall 2025 (GRAD: 11,084 and UGRD: 32,879) in 194-degree programs in 9 colleges
  - Second-largest academic institution in UT System

- **University Analytics**
  - 29 staff serving the campus through fulfillment of ad hocs as well as data for compulsory submissions and reports – 4,200 data projects in FY2025
    - THECB, IPEDS, UT System, SACSCOC, academic program reviews, enrollment planning, revenue and tuition planning, strategic planning, etc.
    - Benchmarked by International Institute of Analytics (IIA) to guide growth and improvement

A Peek Behind the Curtain
The Early Days of Our Data Warehouse

- **2006 to 2017**
  - Oracle database on prem based on PeopleSoft extracts and transformations
  - All ETL and data tables managed by OIT
    - Modeled tables populated by old SQR code written when UTA switched to PS
  - THECB reports ran from these tables
    - Flat files generated on OIT servers
    - Any report changes required liaising with OIT to modify SQR
    - No historical data to reference and many times didn't match to THECB dashboards and publications
 - Difficult and ineffective data environment!

UTA

A Peek Behind the Curtain
Transitioning to SAS

- **2017 to 2025**
  - University Analytics engaged with SAS to modernize data warehouse
    - Enterprise Guide (EG) and Data Integration Studio (DI) main tools
    - Partnered with SAS for assistance with initial migration

  - Brought ETL control and oversight in-house to University Analytics
    - Established cloud-based infrastructure (AWS) supported by third-party vendor
    - Still extracting from PS but now with direct control of modeled tables
    - Any troubleshooting or modifications to tables managed by our own data engineering team
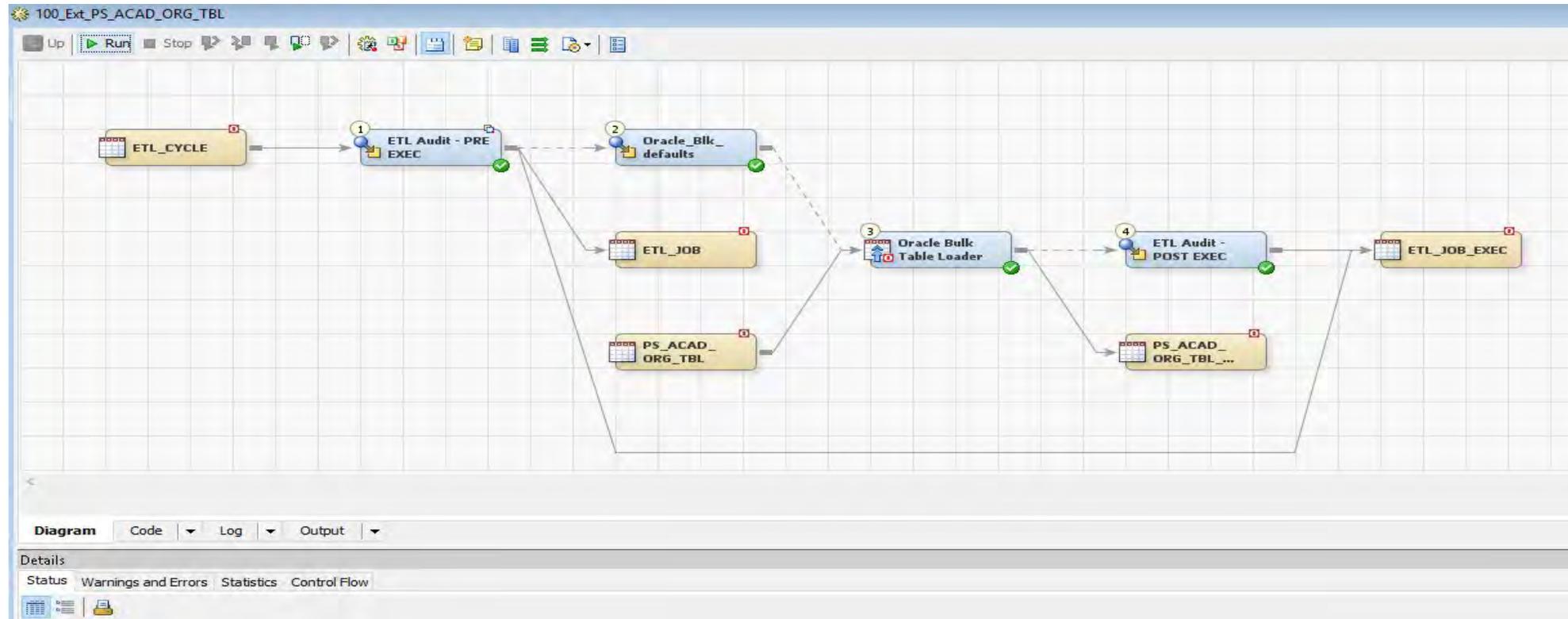
UTA

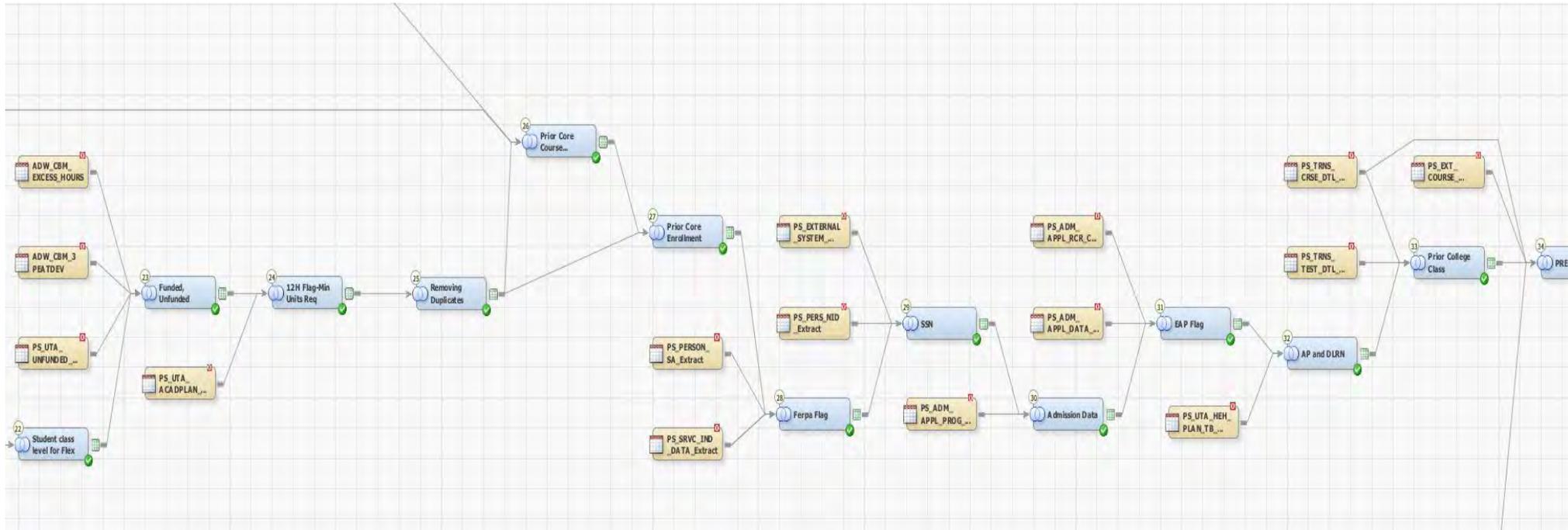# A Closer Look Under The Hood
A High-Level Perspective of Data Environment
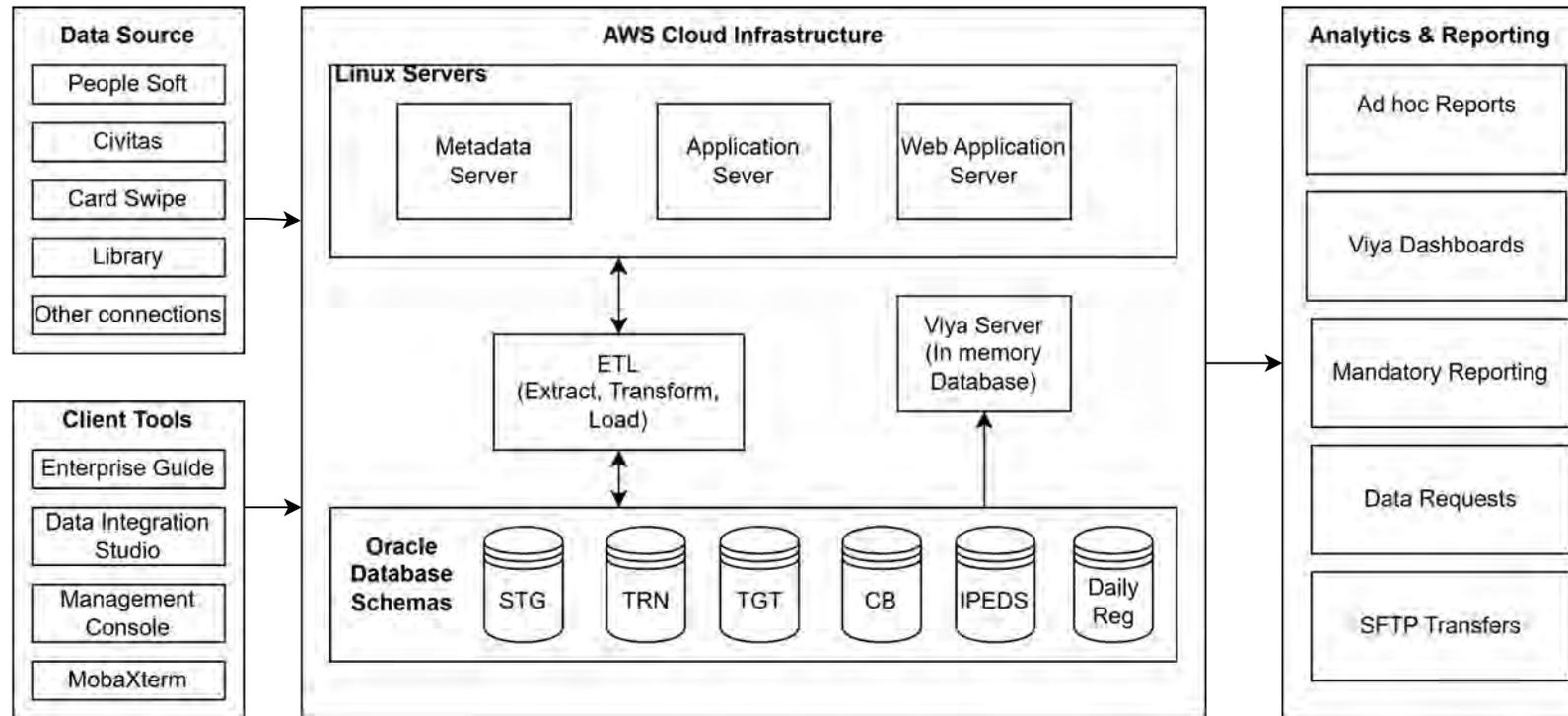
A Closer Look Under the Hood
Simple DI Studio Job

A Closer Look Under the Hood
Complex DI Studio Job

How SAS Enabled Greater Success
Inside the Daily Routine of a Data Quality Analyst

- **Improving Data Quality**
  - Daily runs of quality assurance programs
    - Increases awareness of potential data issues
    - Enables identification of issues prior to report submission
    - Provides opportunity to leverage existing relationships with data owners across campus to correct erroneous data

- **Improving Efficiencies**
  - "Two button process"
  - Streamlines reporting process to regain available bandwidth
  - Enables access to quality preliminary and certified data

How SAS Enabled Greater Success
An Example of a Data Quality Automated Report

How SAS Enabled Greater Success
Data Preservation (Extension to Two Buttons)
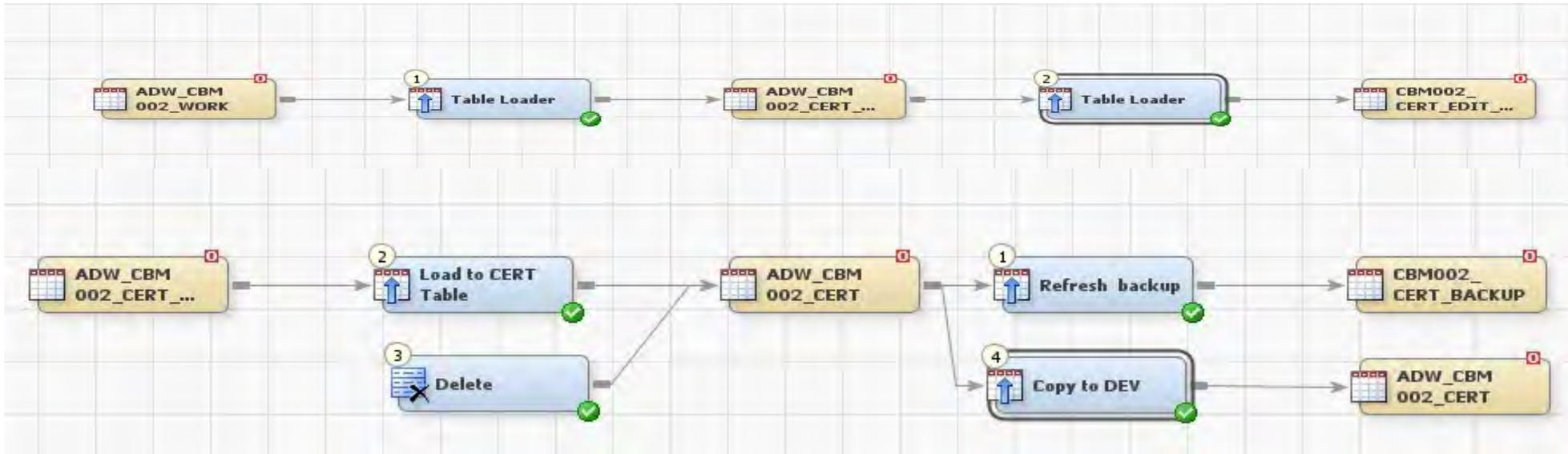
2025 Fall CBM008,CBM002,CBM00S,CBM0E1 preliminary data has been loaded on 02/12/2026 and is yet to be certified.

Summary table below shows the most recent status of all THECB certified reports.

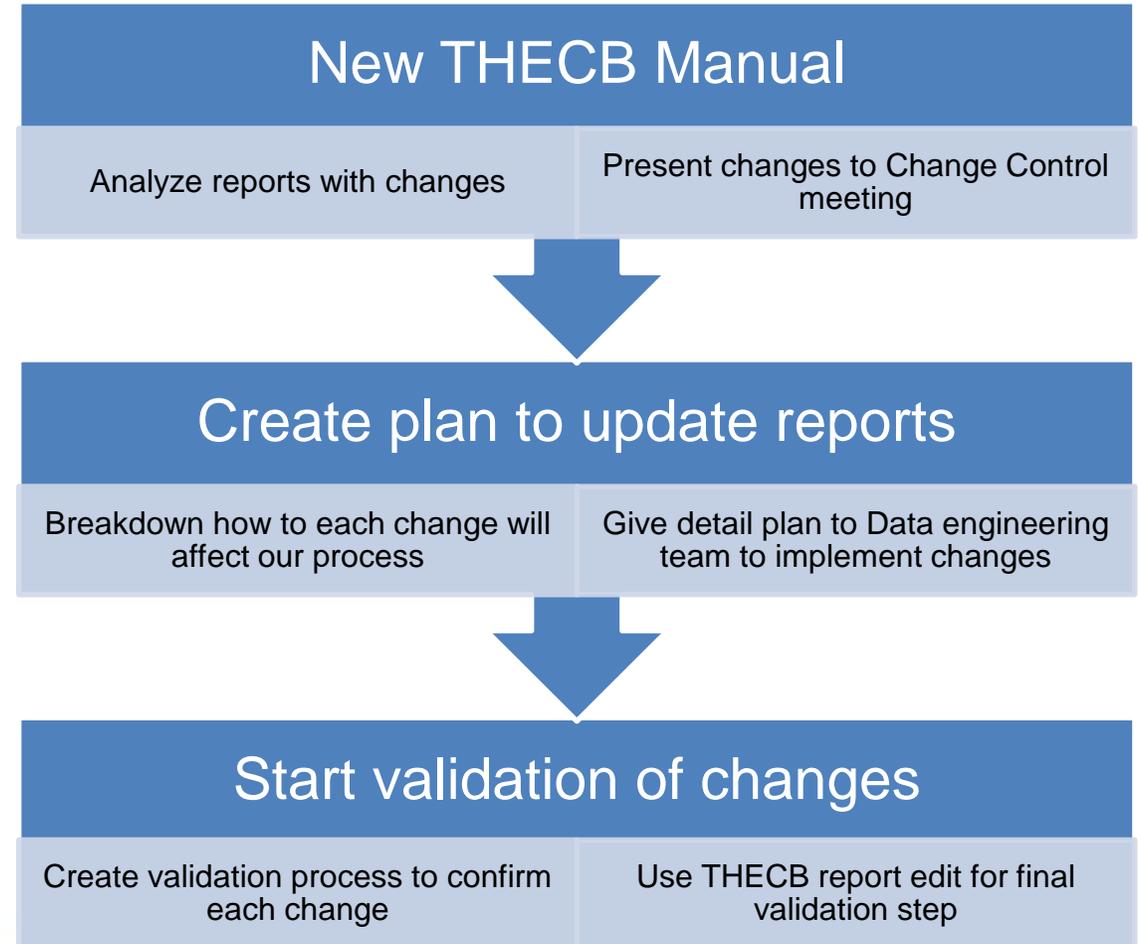| ReportName	LastCertifiedTerm	LastCertifiedTermDescription	LastCertifiedStatus	LastPreliminaryStatus	LastLoadedDate
CBM001	2258	2025 Fall	Y	N	18NOV2025:11:15:00
CBM002	2258	2025 Fall	N	Y	12FEB2026:09:45:15

How SAS Enabled Greater Success
Facilitating Updates Through Change Management

- Established Process for Fixes and Changes
  - Inexplicable anomalies sometimes creep in
    - Node by node running of DI job to identify issue
    - Liaise with data engineering team to implement fix

  - Unexpected changes to CB reports
    - Read and understand updates in new manual
    - Coordinate with data engineering team to facilitate changes

  - Detailed process to allow thorough discussion and documentation of changes
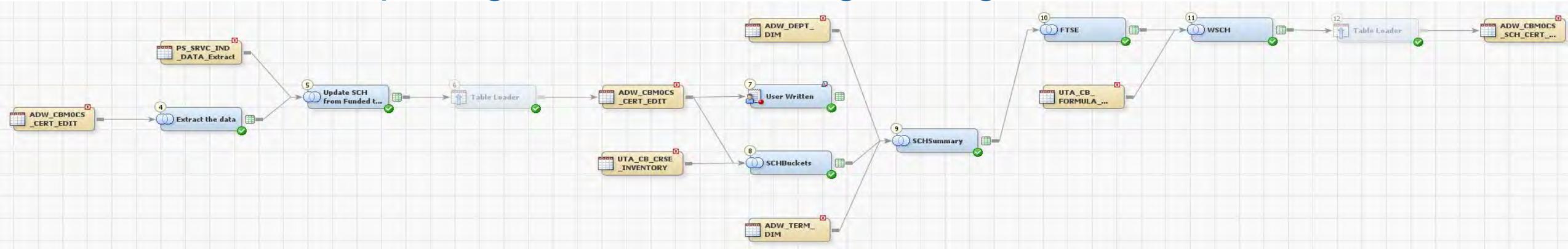
How SAS Enabled Greater Success
Validation and Report Update Workflows

THECB Reports Data Validation

PeopleSoft Data entry issues

Check Quality of Data

Data Extract/Manipulation Logic updates

New THECB Manual

| Analyze reports with changes | Present changes to Change Control meeting |

Create plan to update reports

| Breakdown how to each change will affect our process | Give detail plan to Data engineering team to implement changes |

Start validation of changes

| Create validation process to confirm each change | Use THECB report edit for final validation step |

How SAS Enabled Greater Success
Completing and Documenting Changes in DI Studio

SAS User Experience
How Campus Constituents Interact with SAS

- **Dashboards and Stored Processes**
  - o Most users leverage these for their data needs
  - o University Analytics website has links and thumbnails of dashboards
    - Selected examples include Daily Registration, Admissions Pipeline, FTIC Persistence Prediction, Major Migrations, Student Debt, and Rankings
  - o Website also has most popular stored processes along with link to full inventory of predefined reports

- **Enterprise Guide**
  - o Select users have permissions to develop SAS code using data tables
  - o Such users undergo SAS training similar to University Analytics analysts

SAS User Experience
## Example of Developed Dashboard

This dashboard is composed of several reports that provide **daily registration totals** leading up to **census** date. **Comparison** of **registration** and **SCH to same day** in **previous year** is available by college, department, career, cohort entry status, AP status, CB reported flag, gender, ethnicity/race, and student load.

SAS User Experience
## Example of Developed Dashboard
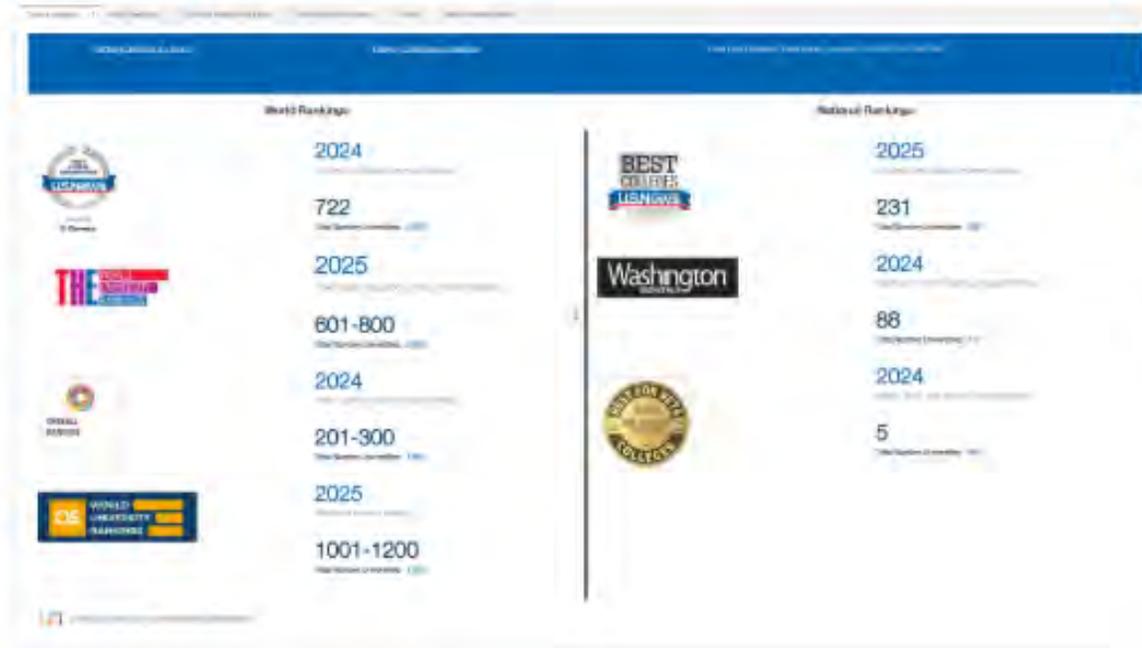
This dashboard displays **UTA's current position in key national and global rankings that the university actively monitors** and participates in. The Trends page provides a **historical view of UTA's outcomes in a selected national or international ranking**, along with descriptions of the fundamental ranking criteria.

SAS User Experience
Example of Developed Stored Process

Weekly Funnel Reports

Summer 2022: Undergraduate Application Pool by U.S. State
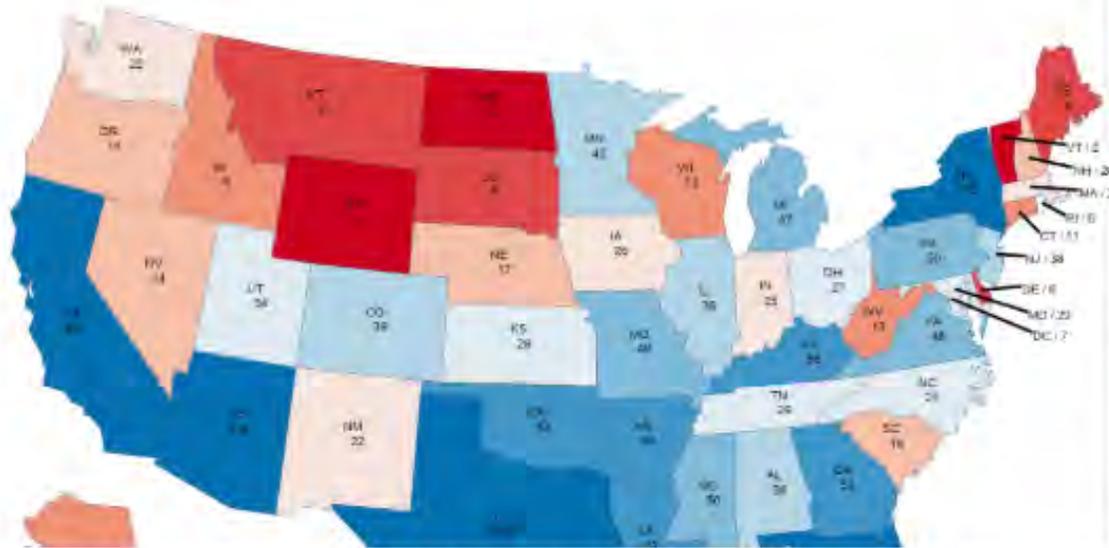
This is a downloadable PDF report that provides **an overview of university funnel statistics** (applied, admitted, matriculated, and enrolled). **Comparison of funnel stats with the same day in the previous year** are also available. Undergraduate and graduate demographics are provided by US state(map), ethnicity/race and gender. **Accelerated online vs traditional program funnel and cancellation statistics** are available **by entry level** (freshman, transfer, doctoral, master's). Additional crosstab views include transfer student funnels **by transfer institution type**, and intent to enroll stats by undergraduate entry level. Graduate funnel provides a program breakdown and college funnel summary by accelerated online program indicator and residency.

The Pursuit of (Data) Happiness
Transitioning from SAS to Databricks

Why Modernize?

- **Databricks One:** Establishes a **single data portal** for the entire campus, allowing users to access dashboards, reports, and answers in one unified web gateway.

- **Cost Efficiency: EC2 Compute** to **Serverless Compute**, a pay-for-what-you-use model that eliminates idle infrastructure costs entirely.

- **Real-Time Data:** Data is processed continuously (e.g., live Card Swipe or Enrollment, Admissions), ensuring dashboards always reflect **current activity**.

- **Future-Proof Scalability:** Built to handle any data format — structured tables, unstructured text, images, or video — ensuring the platform grows with the university's needs.

- **High-Performance Execution:** Built on **Apache Spark**, it uses the **Photon Engine** to deliver blazing speed, processing massive datasets far faster than legacy serial jobs.

The Pursuit of (Data) Happiness
Transitioning from SAS to Databricks

- **Open Standards Architecture:** Supports **SQL, Python, R and Git** natively, replacing proprietary SAS functions/PROC's.

- **Lakeflow Connect:** Automatically ingests data from PeopleSoft, Civitas, and external databases with built-in **Change Data Capture (CDC).**

- **Databricks AI/BI Genie:** An AI-powered tool that allows users to ask data questions in plain English (e.g., "Show me Fall enrollment trends") and get instant, accurate answers.

- **Mosaic AI & Assistant:** Accelerates development with **Databricks Assistant** (context-aware code fixes) and **AutoML** (rapid predictive modeling).

- **Unified Governance & Security:** Uses **Unity Catalog** to enforce consistent security policies across all data assets while also providing **automated lineage.**

UTA

Thank you for joining!

Any questions?

Bobby Childress – childress@uta.edu

Sasha Naourizbaeva – saulen@uta.edu

Jeremy Davis – jeremy.davis@uta.edu

Tarak Myneni – tarakaram.myneni@uta.edu

www.uta.edu/analytics

universityanalytics@uta.edu

UTA