Connecting Colleges to Data: CCC Data Warehouse Upgrades and Updates

Hello
my name is

Mark Cohen

Product Manager, Infrastructure and Data Services

CCC Technology Center

mcohen@ccctechcenter.org

California Community Colleges

Agenda

- **CCC Data Overview**
- **The Platform**
- **The Data**
- **"Student Journey"**
- **Project Update**
- **Accessing the DW**
- **Discussion**

California Community Colleges
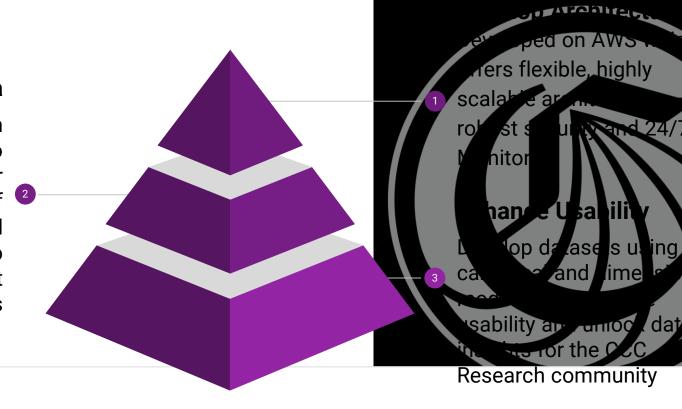
CCC Data Overview

Supported by the Data Management Grant (DMG) initiative from the California Community Colleges Chancellor's Office, CCC Data provides our Colleges with access to critical data used to support institutional decision-making aligned with the CCCCO Vision 2030.

California Community Colleges Chancellor's Office

California Community Colleges

Developing a Data Platform

Aggregate Data

Aggregated data from disparate sources to bring together multiple sources of systemwide data and enable access to college and district Researchers

1

2

3

Design Architecture
Developed on AWS which offers flexible, highly scalable architecture, robust security and 24/7 Monitoring

Enhance Usability
Develop datasets using a canonical and dimensional model to enhance usability and unlock data insights for the CCC Research community

How CCC Data is Utilized

District and College Researchers

CCC Data Warehouse access to District and College Researchers and supporting IT staff.

- Connect with systemwide data sources
- Connect BI tools
- Produce reporting

CCCCO Researchers and Consultants

CCC Data Warehouse and select Data Lake access to CCCCO Researchers and Consultants.

- Access CCCApply and other systemwide data
- Combine datasets to answer questions
- Produce reporting

Select CCC Data Partners

Limited access authorized by CCCCO to identified Systemwide Initiatives and CCC Data Partners.

- CVC Report supporting MIS submission
- Facilitate secure data transfer in production of dashboards

California Community Colleges

CCC Data Platform

ETL's
- Several proven methods available to integrate data with CCC Data
- Extract, Transform, and Loads data
- Runs nightly (or more frequently)

CCC Data Lake
- Collects data, preserving changes.
- Used to power Data Warehouse, Data-Marts, and analytics.
- Built on Amazon AWS S3

DW Connections
- Colleges may connect local BI/ analytics tools or use the DW Report Server
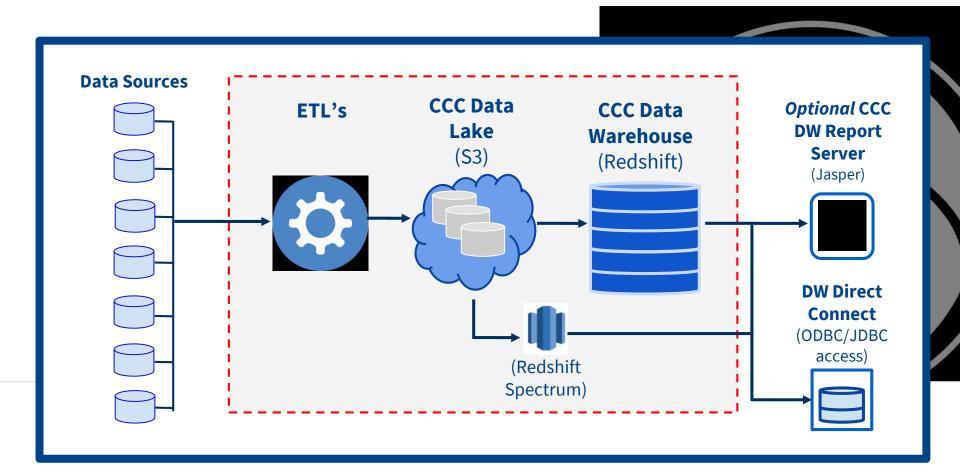 - Direct DW Connection
  - CCC Data Report Server

CCC Data Warehouse
- Structured source of master data segmented by MIS code
- Connects data to generate reports and analytics for end-users
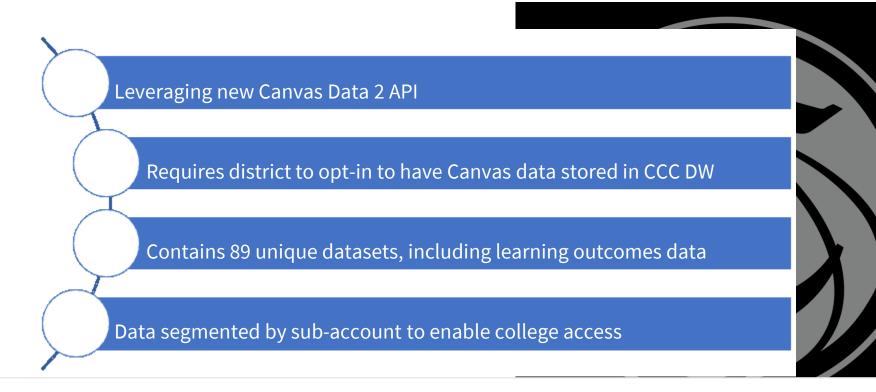- Built on Amazon AWS Redshift

California Community Colleges

CCC Data Infrastructure

The Data

|		Direct Connect	Report Server	Comments
1	**CCCApply Standard and Noncredit Application**	✓	✓	- Secure access to PII, including fields encrypted in CCCApply Report Center - One of top 3 viewed datasets in DW
2	**CCCApply International Application**	✓	✓	Data from the International application
3	**CCCApply Promise Grant Application**	✓	✓	Data from the Promise Grant application
4	**Multiple Measures Placement Service (MMPS)**	✓	✓	Placement recommendations, independent of utilization of MMPS
5	**MyPath**	✓	✓	- Advisor card data - Advisor usage data - Audit table

The Data
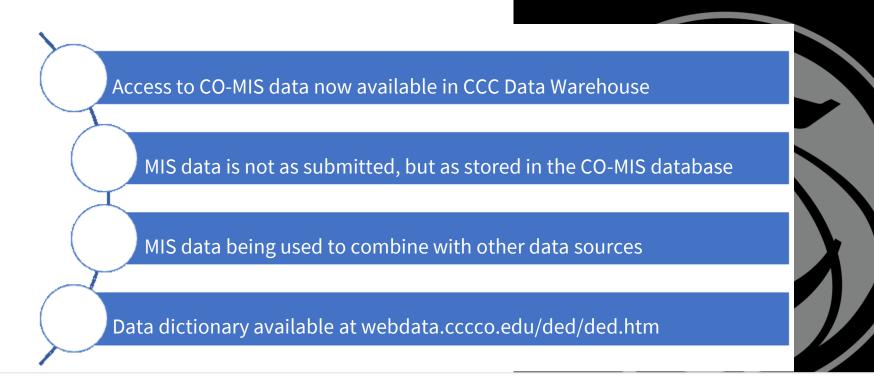
		Direct Connect	Report Server	Comments
6	**Chancellor's Office Curriculum Inventory (COCI)**	✓	✓	• Data from the Curriculum Inventory System
7	**Course ID Numbering System (C-ID)**	✓	✓	• Data from Course ID Numbering system
8	**Canvas**	✓	✗	• Requires district to "opt-in" • One of top 3 viewed datasets in DW • Leveraging Instructure CD2 API
9	**CO-MIS**	✓	✓	• Made available to colleges with CCC Data 3.0 release • Post processed MIS data
10	**Student Journey**	✓	✗	• Brings together tables from CCCApply & CO-MIS • Linked by a common identifier

Canvas Data (CD2)

- Leveraging new Canvas Data 2 API

- Requires district to opt-in to have Canvas data stored in CCC DW

- Contains 89 unique datasets, including learning outcomes data

- Data segmented by sub-account to enable college access

California Community Colleges

Management Information System (CO-MIS)

- Access to CO-MIS data now available in CCC Data Warehouse
- MIS data is not as submitted, but as stored in the CO-MIS database
- MIS data being used to combine with other data sources
- Data dictionary available at webdata.cccco.edu/ded/ded.htm

California Community Colleges

Student Journey Data

- Unified Data Source

- Canonical/Dimensional Modeling Approach

- Adaptive & Research-Driven

- Enhanced Usability & Accessibility
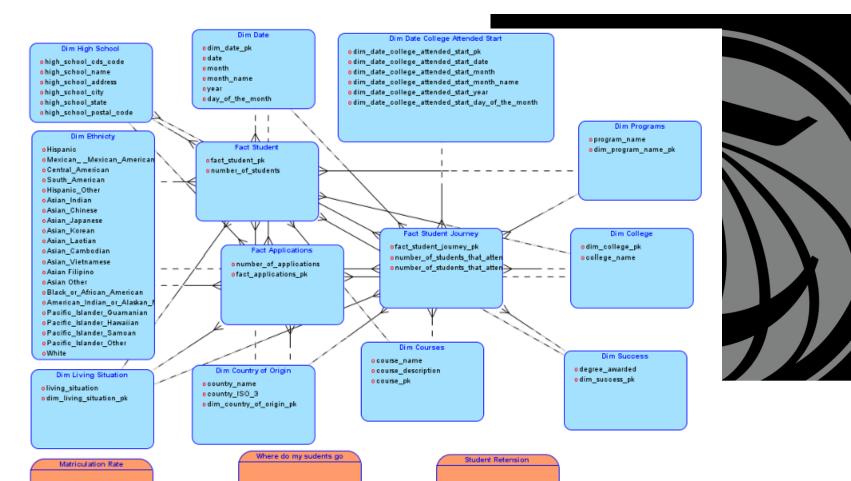
- Deeper Insights & Correlations

California Community Colleges

Attributes "Canonicalized"

source	table_name	number_of_attributes
stnd_application	apply__personal_sociodemographic	40
stnd_application	apply__contact_information	24
stnd_application	apply__education	19
intl_application	apply__education	4
intl_application	apply__personal_information	7
stnd_application	apply__personal_information	3
bogfw_application	apply__personal_sociodemographic	25
bogfw_application	apply__contact_information	10
intl_application	apply__contact_information	43
bogfw_application	apply__personal_information	3
intl_application	apply__personal_sociodemographic	8
stnd_application	apply__college_attended	13

California Community Colleges

Transformation to a Star Schema

Integration Support

AWS Federated Querying Supports *queries between the CCC DW a Colleges local DW*

CCC Data Platform

AWS Athena

AWS Athena

VPN Connect...

College/District Researcher

AWS Lambda

Amazon Virtual Private Cloud

Amazon Redshift

CCC Data Warehouse

Local Data Warehouse (College or District)

Project Update

Evolve the Student Journey dataset

- Integrate additional tables based on Researcher feedback
- Enable data across colleges at an aggregated-level *(no PII)*
- Create associated entities between sources

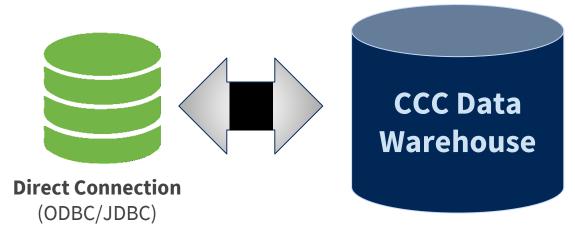## Enhance accessibility and integration

- Support integration of CCC Data Warehouse to local warehouses
- Create associated entities between CCC DW & local data warehouses
- Increased training support for Data Warehouse

Expand available data

- Enable access to additional data within the CCC Data Warehouse
- Identify new data sources to be brought into Data Warehouse

Upgrade the CCC Data platform

- Integrating SuperSet to augment or replace Jasper Reports
- Provide API access as option for Direct Connection to CCC DW

Accessing the DW

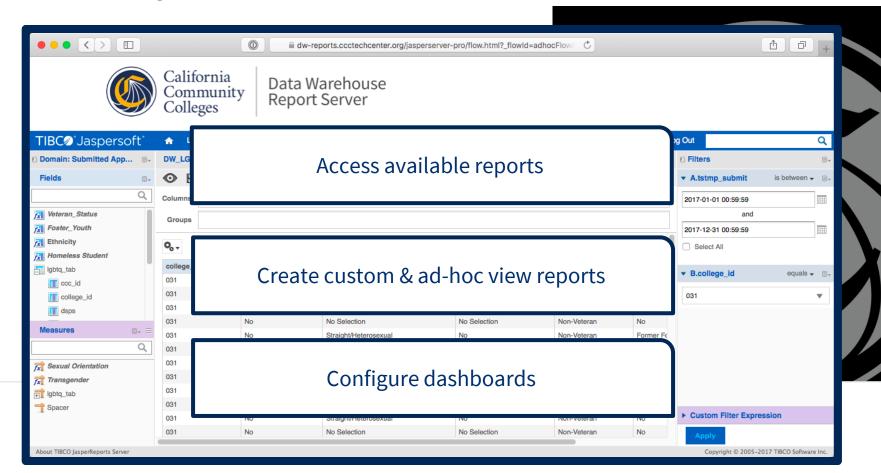CCC Data available to approved college and district staff, through:

Direct Connection
(ODBC/JDBC)

CCC Data Warehouse

DW Report Server
(Jaspersoft)

California Community Colleges

DW Direct Connect

- Supports colleges and district offices connecting to data in CCC Data Warehouse, with:
  - Available BI/analytic applications, such as Tableau and Power BI
  - SQL Workbench for direct queries
  - Local data warehouse, connected as external tables
- Connection requires:
  - Installation of custom site-to-site VPN
  - Data Warehouse login credentials

California Community Colleges

DW Report Server

Access available reports

Create custom & ad-hoc view reports

Configure dashboards

Supporting Documentation

CCC Data Program Overview

A part of the Data Management Grant from the California Community Colleges Chancellor's Office, CCC Data provides the necessary infrastructure to the California Community College System to aggregate data across disparate systems to an enterprise data warehouse.

Access to these data is provided to institutional researchers, college and district administrators, and other decision makers at the 116 California Community Colleges, district offices, and the Chancellor's Office, where these critical data may be used to support instructional and institutional decision-making aligned with the Chancellor's Vision for Success.

Data Available in the CCC Data Warehouse

Source Database	College/District Access	
	DW Direct Connection	DW Report Server
California College Promise Grant Application	✅	✅

Data Warehouse Resources

- **DW Direct Access User Guide**
- **DW Report Server User Guide**
- **Student Journey User Guide**
- **Data Dictionaries**
- **CCC Data Support Site**
- **Jaspersoft (DW Report Server) Tutorials** *external links*
  - Tibco Jaspersoft explainer video
  - Jaspersoft Workshop series: "From 0 to Dashboard"
  - Jaspersoft Tutorials
- **Canvas Data 2 FAQs**

Release Notes

https://cccnext.jira.com/wiki/spaces/DWPub/overview

Requesting Access & Support Resources

- To request access to the Data Warehouse, adopt or upgrade a produ[ct] or schedule a training session: cems@ccctechcenter.org

- Release alerts and peer support for products, visit www.ccctechnology.info

- Production issues, create a staff supp[ort] ticket by emailing us at staffsupportccctc@openccc.zendesk.com

https://ccctechcenter.org/resources/ess

California Community Colleges

Discussion/Questions

mcohen@cccTechCenter.org

Connecting Colleges to Data: CCC Data Warehouse Upgrades and Updates