ITS Data & Analytics

ITS
SEMINOLE
SHOWCASE

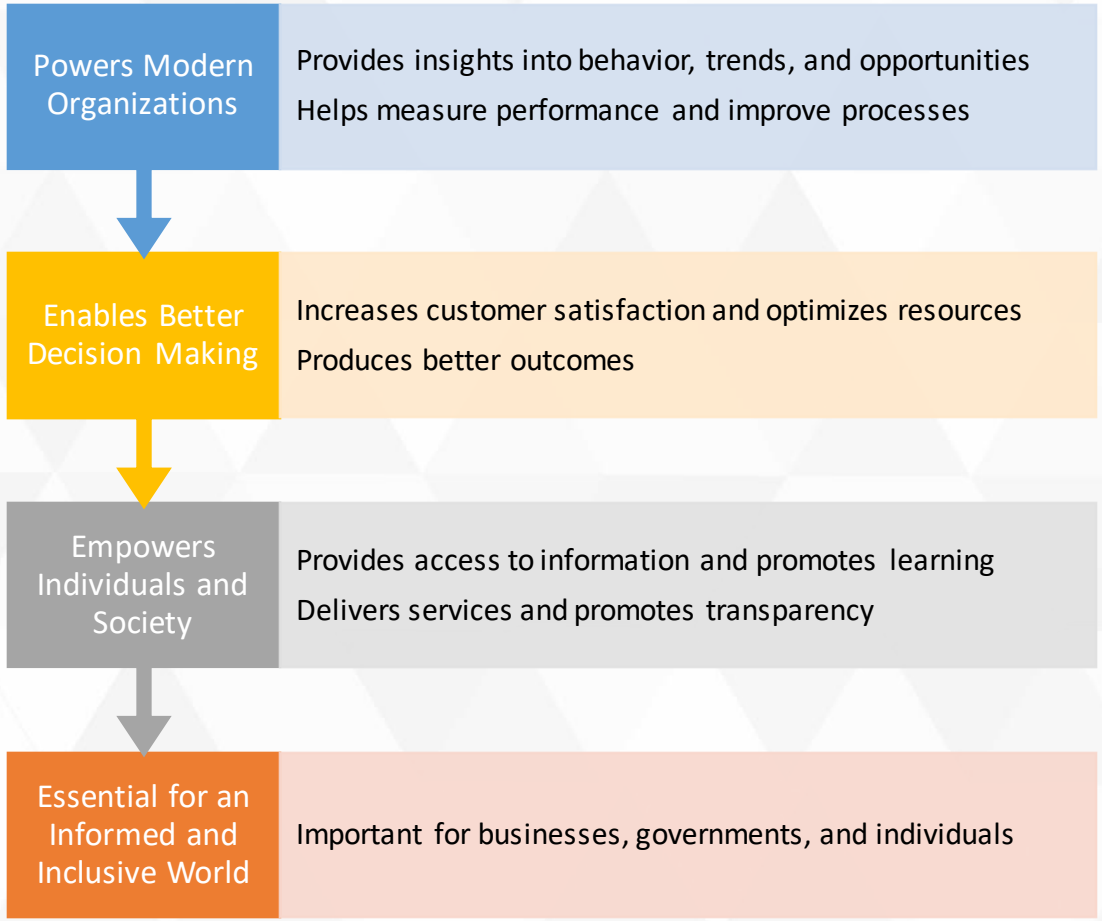
Agenda

- Intro and general overview
- Data Integration
- Data Repositories
- Analytics and Visualizations
- Data Catalog
- Follow up and next steps

Intro and General Overview

- Who are we?
  - The Data & Analytics group at FSU is a team of professionals who provide data management, analysis, and reporting services.
- What do we do?
  - Collect, validate, transform, and integrate data from various sources.
  - Develop and maintain data warehouses, replicas of production data, and the enterprise data lakehouse.
  - Manage analysis and visualization products.
  - Provide training, consultation, and guidance on data quality, governance, and best practices.
- Our goal
  - To foster a culture of data-driven decision making and innovation across the university community.

Why Data?


Powers Modern Organizations
Provides insights into behavior, trends, and opportunities
Helps measure performance and improve processes
Enables Better Decision Making
Increases customer satisfaction and optimizes resources
Produces better outcomes
Empowers Individuals and Society
Provides access to information and promotes learning
Delivers services and promotes transparency
Essential for an Informed and Inclusive World
Important for businesses, governments, and individuals

Data Integration

REDUCE DATA SILOS AND FRAGMENTATION

IMPROVE DATA QUALITY AND ACCURACY

ENHANCE DATA GOVERNANCE AND SECURITY

FACILITATE DATA SHARING AND COLLABORATION

ENABLE DATA-DRIVEN DECISION MAKING AND INNOVATION

ETL (Extract Transform Load)

Batch-oriented approach

- Reliable Reporting & Analysis of historical data
- Scheduled basis - Daily, weekly, or monthly

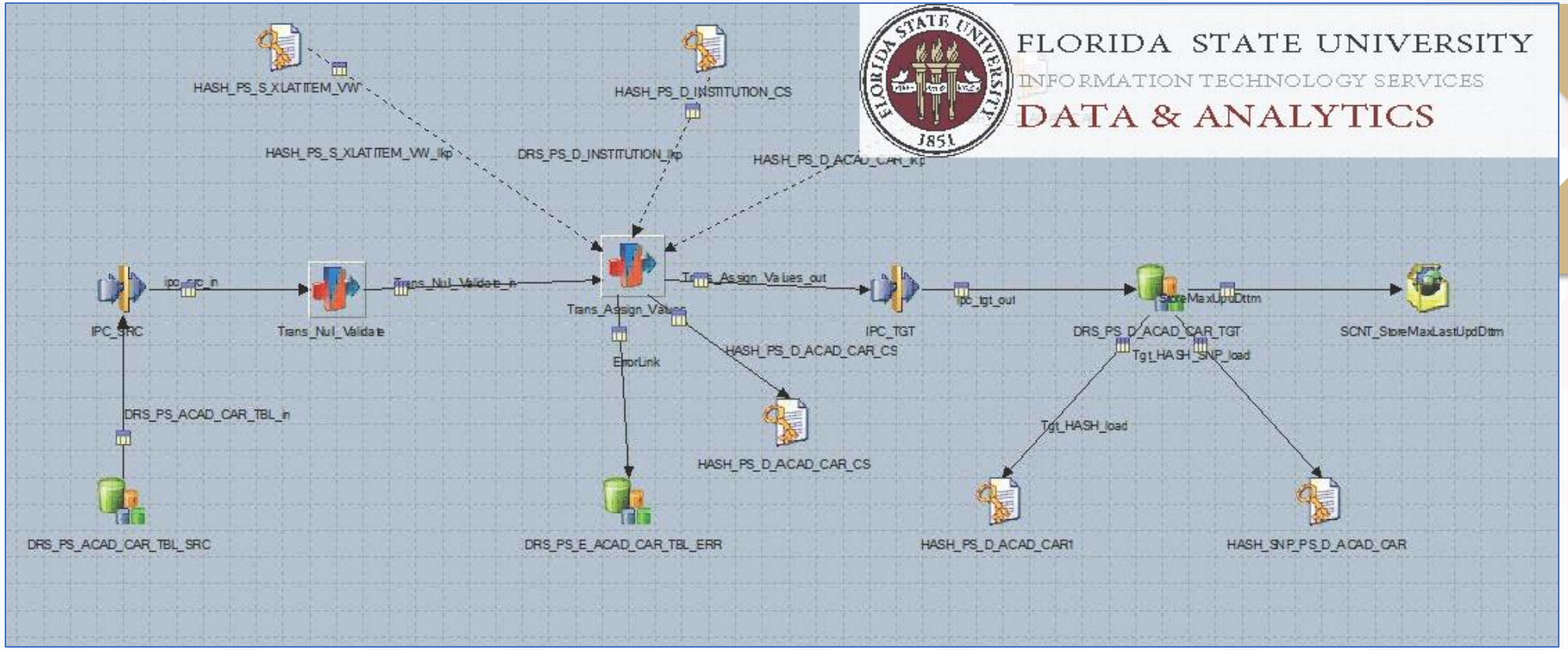
Valuable

- Large volumes - structured and semi-structured data efficiently and reliably
- Data quality and integrity - applying business rules, validations, and transformations

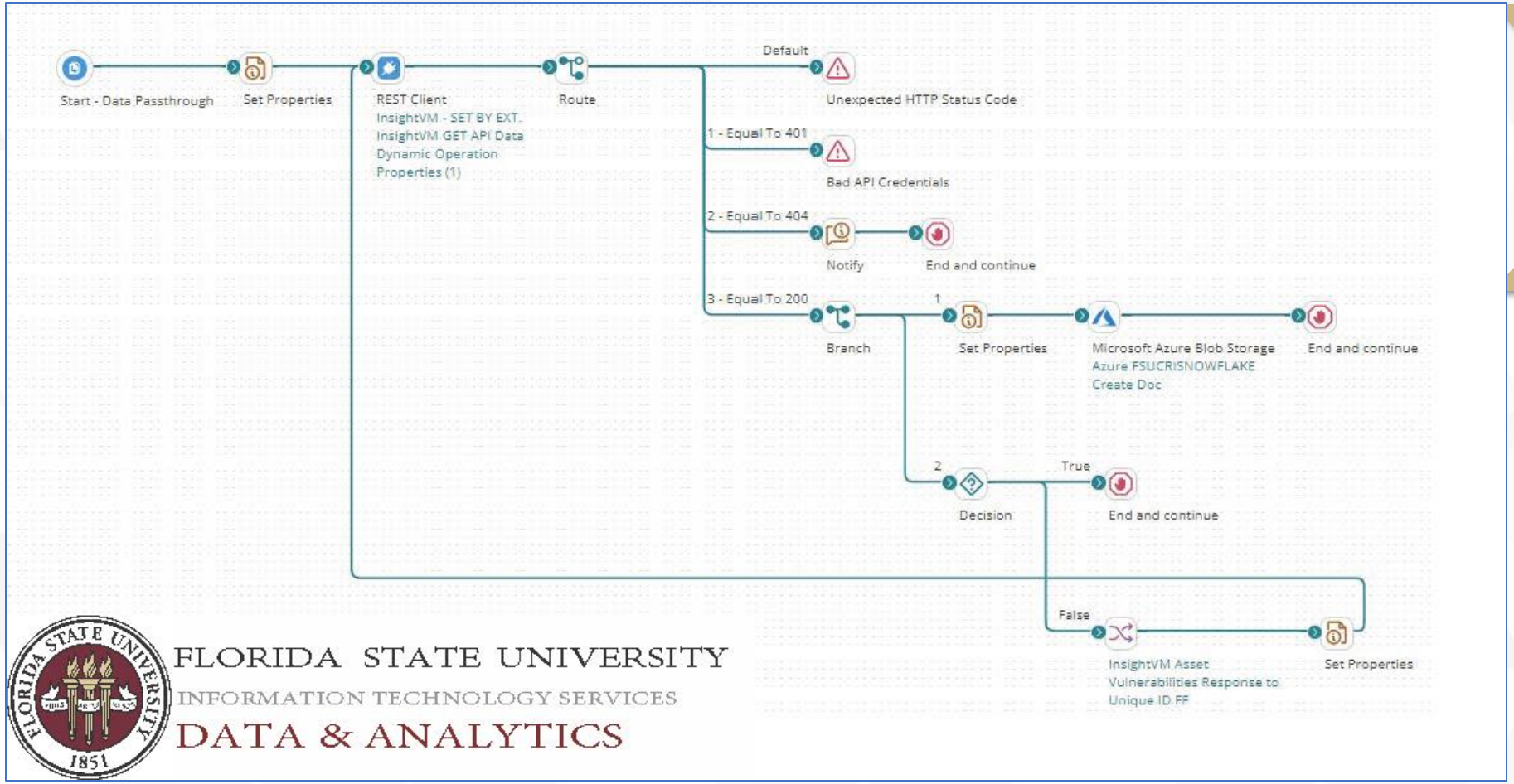
Tools

- DataStage: IBM - Graphical interface for building data integration pipelines
- Boomi: Cloud-based platform - connects data and applications across different environments

DataStage


FLORIDA STATE UNIVERSITY
INFORMATION TECHNOLOGY SERVICES
DATA & ANALYTICS
HASH_PS_S_XLATITEM_VW
HASH_PS_D_INSTITUTION_CS
HASH_PS_S_XLATITEM_VW_lkp
DRS_PS_D_INSTITUTION_lkp
HASH_PS_D_ACAD_CAR_lkp
ipc_src_in
Trans_Nul_Validate_in
Trans_Assign_Values_out
IPC_SRC
Trans_Nul_Validate
Trans_Assign_Values
ErrorLink
IPC_TGT
ipc_tgt_out
DRS_PS_D_ACAD_CAR_TGT
Tgt_HASH_SNP_load
SCNT_StoreMaxLastUpdDttm
HASH_PS_D_ACAD_CAR_CS
DRS_PS_ACAD_CAR_TBL_in
Tgt_HASH_load
HASH_PS_D_ACAD_CAR_CS
DRS_PS_ACAD_CAR_TBL_SRC
DRS_PS_E_ACAD_CAR_TBL_ERR
HASH_PS_D_ACAD_CAR1
HASH_SNP_PS_D_ACAD_CAR

Boomi

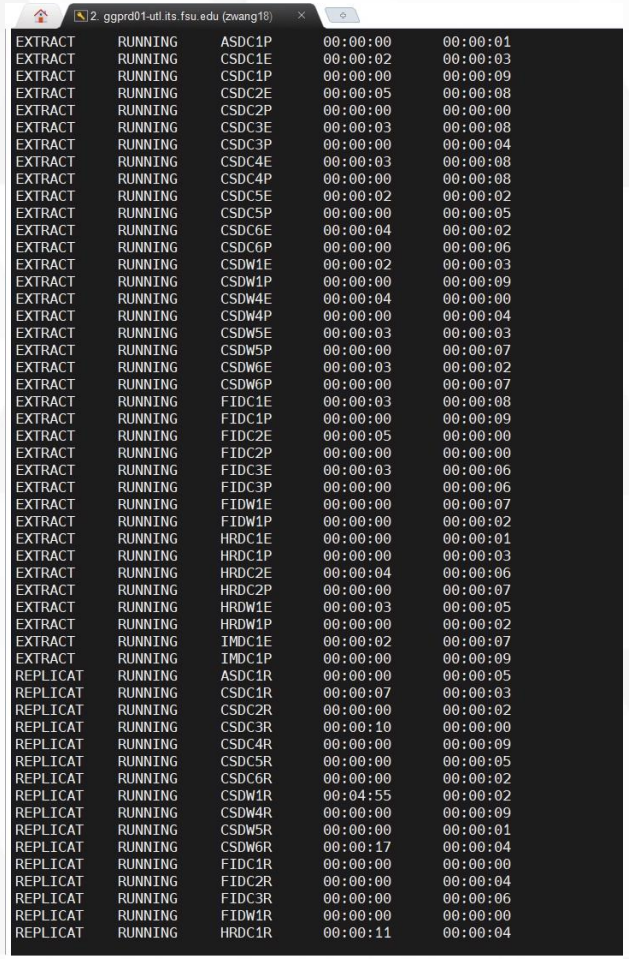


FLORIDA STATE UNIVERSITY
INFORMATION TECHNOLOGY SERVICES
DATA & ANALYTICS

Changes Happen (Change Data Capture)

- CDC captures and delivers changes made to data sources in real time or near real time
  - Monitors inserts, updates, and deletes performed on source data
  - Streams changes to target system as events or messages
- Useful for incremental integrations
  - Reduces data volume and frequency of batch loads
  - Captures only changes that occur in source data
- Enables faster and more timely delivery of data to target system
  - Improves responsiveness and accuracy of data-driven applications and decisions
- Suitable for scenarios where data sources are dynamic and unpredictable
  - Latency between data ingestion and delivery is critical
- GoldenGate is an Oracle product that offers a comprehensive solution for implementing CDC processes

Goldengate


2. ggprd01-utl.its.fsu.edu (zwang18)
EXTRACT RUNNING ASDC1P 00:00:00 00:00:01
EXTRACT RUNNING CSDC1E 00:00:02 00:00:03
EXTRACT RUNNING CSDC1P 00:00:00 00:00:09
EXTRACT RUNNING CSDC2E 00:00:05 00:00:08
EXTRACT RUNNING CSDC2P 00:00:00 00:00:00
EXTRACT RUNNING CSDC3E 00:00:03 00:00:08
EXTRACT RUNNING CSDC3P 00:00:00 00:00:04
EXTRACT RUNNING CSDC4E 00:00:03 00:00:08
EXTRACT RUNNING CSDC4P 00:00:00 00:00:08
EXTRACT RUNNING CSDC5E 00:00:02 00:00:02
EXTRACT RUNNING CSDC5P 00:00:00 00:00:05
EXTRACT RUNNING CSDC6E 00:00:04 00:00:02
EXTRACT RUNNING CSDC6P 00:00:00 00:00:06
EXTRACT RUNNING CSDW1E 00:00:02 00:00:03
EXTRACT RUNNING CSDW1P 00:00:00 00:00:09
EXTRACT RUNNING CSDW4E 00:00:04 00:00:00
EXTRACT RUNNING CSDW4P 00:00:00 00:00:04
EXTRACT RUNNING CSDW5E 00:00:03 00:00:03
EXTRACT RUNNING CSDW5P 00:00:00 00:00:07
EXTRACT RUNNING CSDW6E 00:00:03 00:00:02
EXTRACT RUNNING CSDW6P 00:00:00 00:00:07
EXTRACT RUNNING FIDC1E 00:00:03 00:00:08
EXTRACT RUNNING FIDC1P 00:00:00 00:00:09
EXTRACT RUNNING FIDC2E 00:00:05 00:00:00
EXTRACT RUNNING FIDC2P 00:00:00 00:00:00
EXTRACT RUNNING FIDC3E 00:00:03 00:00:06
EXTRACT RUNNING FIDC3P 00:00:00 00:00:06
EXTRACT RUNNING FIDW1E 00:00:00 00:00:07
EXTRACT RUNNING FIDW1P 00:00:00 00:00:02
EXTRACT RUNNING HRDC1E 00:00:00 00:00:01
EXTRACT RUNNING HRDC1P 00:00:00 00:00:03
EXTRACT RUNNING HRDC2E 00:00:04 00:00:06
EXTRACT RUNNING HRDC2P 00:00:00 00:00:07
EXTRACT RUNNING HRDW1E 00:00:03 00:00:05
EXTRACT RUNNING HRDW1P 00:00:00 00:00:02
EXTRACT RUNNING IMDC1E 00:00:02 00:00:07
EXTRACT RUNNING IMDC1P 00:00:00 00:00:09
REPLICAT RUNNING ASDC1R 00:00:00 00:00:05
REPLICAT RUNNING CSDC1R 00:00:07 00:00:03
REPLICAT RUNNING CSDC2R 00:00:00 00:00:02
REPLICAT RUNNING CSDC3R 00:00:10 00:00:00
REPLICAT RUNNING CSDC4R 00:00:00 00:00:09
REPLICAT RUNNING CSDC5R 00:00:00 00:00:05
REPLICAT RUNNING CSDC6R 00:00:00 00:00:02
REPLICAT RUNNING CSDW1R 00:04:55 00:00:02
REPLICAT RUNNING CSDW4R 00:00:00 00:00:09
REPLICAT RUNNING CSDW5R 00:00:00 00:00:01
REPLICAT RUNNING CSDW6R 00:00:17 00:00:04
REPLICAT RUNNING FIDC1R 00:00:00 00:00:00
REPLICAT RUNNING FIDC2R 00:00:00 00:00:04
REPLICAT RUNNING FIDC3R 00:00:00 00:00:06
REPLICAT RUNNING FIDW1R 00:00:00 00:00:00
REPLICAT RUNNING HRDC1R 00:00:11 00:00:04

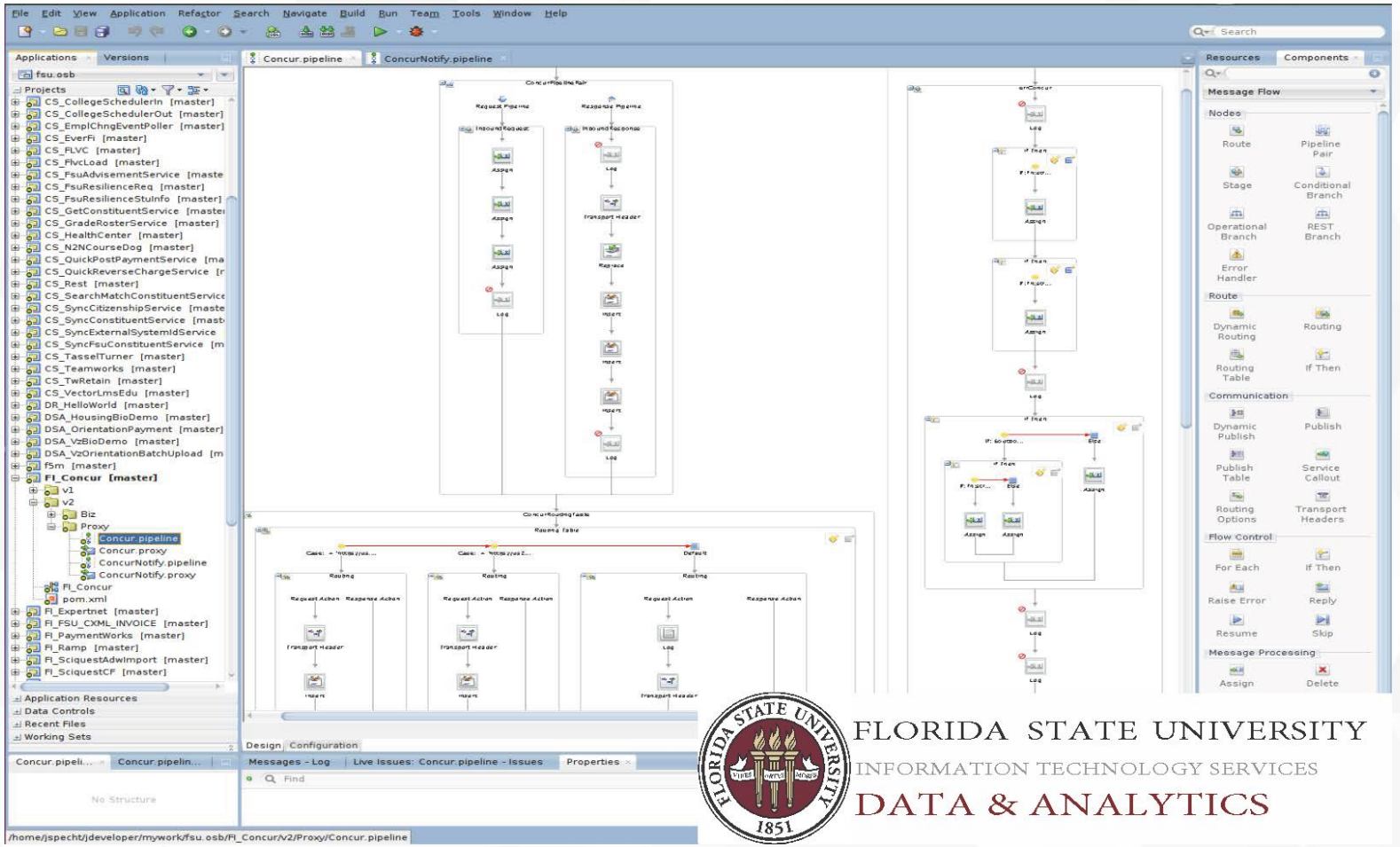
Streaming Data

- Streaming data is continuously generated and transmitted by various sources
  - Processed in real time or near real time to extract insights and trigger actions
- Event-driven architecture (EDA) is a design pattern for streaming data processing
  - Consists of event producers, event consumers, and event brokers
  - Enables decoupling of event producers and consumers
  - Improves scalability, flexibility, and reliability of the system
- Currently use home grown Event Framework
- Piloting IBM Event Automation
  - Cloud-based platform for EDA
  - Uses Apache Kafka and Apache Flink for event streaming and processing
  - Provides scalability, fault tolerance, security, and streaming SQL

APIs, API Management, and ESBs

- APIs allow for the creation of modular and reusable software components
  - They enable the development of new features and innovations
  - They facilitate interoperability and compatibility between systems and devices
- API Management simplifies the creation and deployment of APIs
  - They enforce policies and standards for API quality, security, and performance
  - They monitor and analyze API usage, performance, and errors
- ESBs provide a common communication channel and integration layer for connecting diverse applications and services
  - They route and deliver messages and events between applications and services
  - They orchestrate complex workflows and business processes

Service in OSB configured through JDeveloper



FLORIDA STATE UNIVERSITY
INFORMATION TECHNOLOGY SERVICES
DATA & ANALYTICS

Data Repositories

- Data Repositories: Collections of Data for Analysis and Reporting
  - Can have different levels of transformation
  - Depends on use case and data quality requirements
- Data Lake: Stores Raw and Unstructured Data
  - No predefined schema or format
- Data Warehouse: Stores Structured and Curated Data
  - Processed and organized according to common model and standard
- Data Cache: Stores Frequently Accessed Data
  - Reduces latency and improves performance
- Essential Component of Data and Analytics Solution
  - Enables data access, integration, and governance

Data Warehouse

Stores structured and curated data for analysis and reporting

- Uses dimensional model with facts and dimensions

Integrates data from multiple sources in common format

- Improves data quality and consistency

Improves data performance and scalability

- Handles large volumes of historical and current data

Supports data governance and security

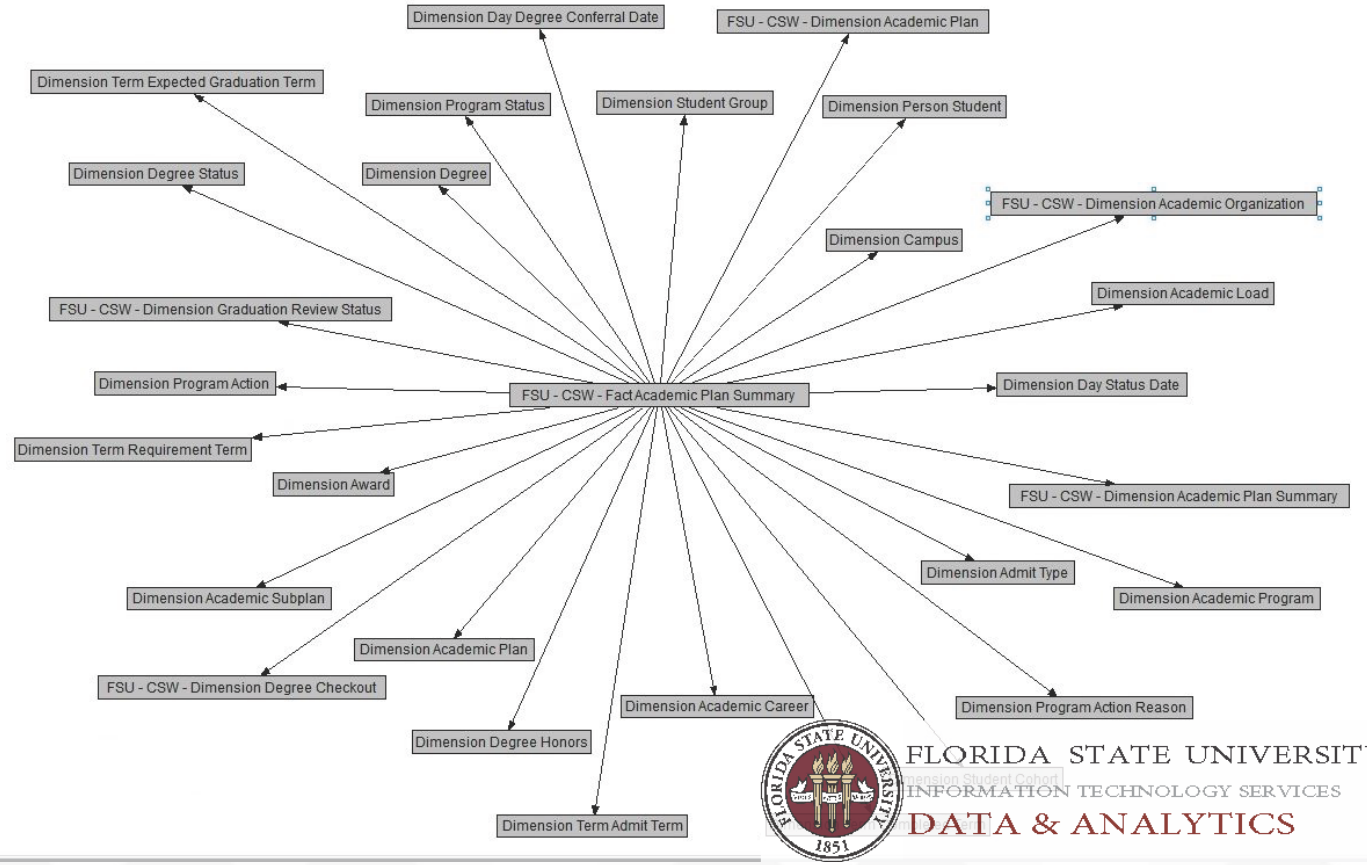
- Access control, auditing, data lineage, and metadata management

Data is slightly stale due to daily loading

- Good source for common business facts

Data accessed through OBIEE or Power BI

Subject Area in OBIEE from DW

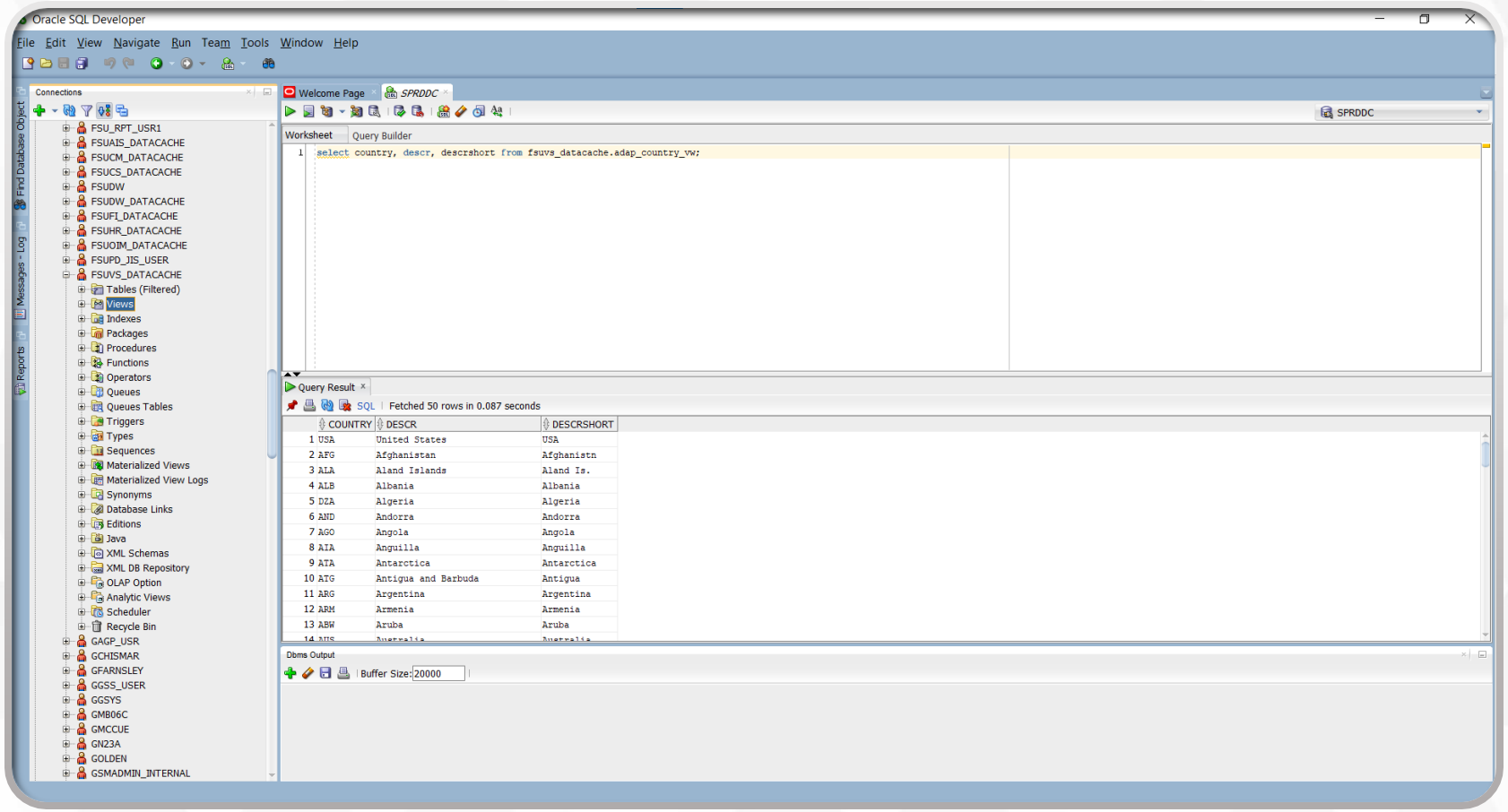


FLORIDA STATE UNIVERSITY
INFORMATION TECHNOLOGY SERVICES
DATA & ANALYTICS

Data Cache

- Near real-time replica of commonly used data from various systems
- Accessed via views, which act like virtual tables and can join and alter data
- Enables faster and easier access to relevant and frequently used data

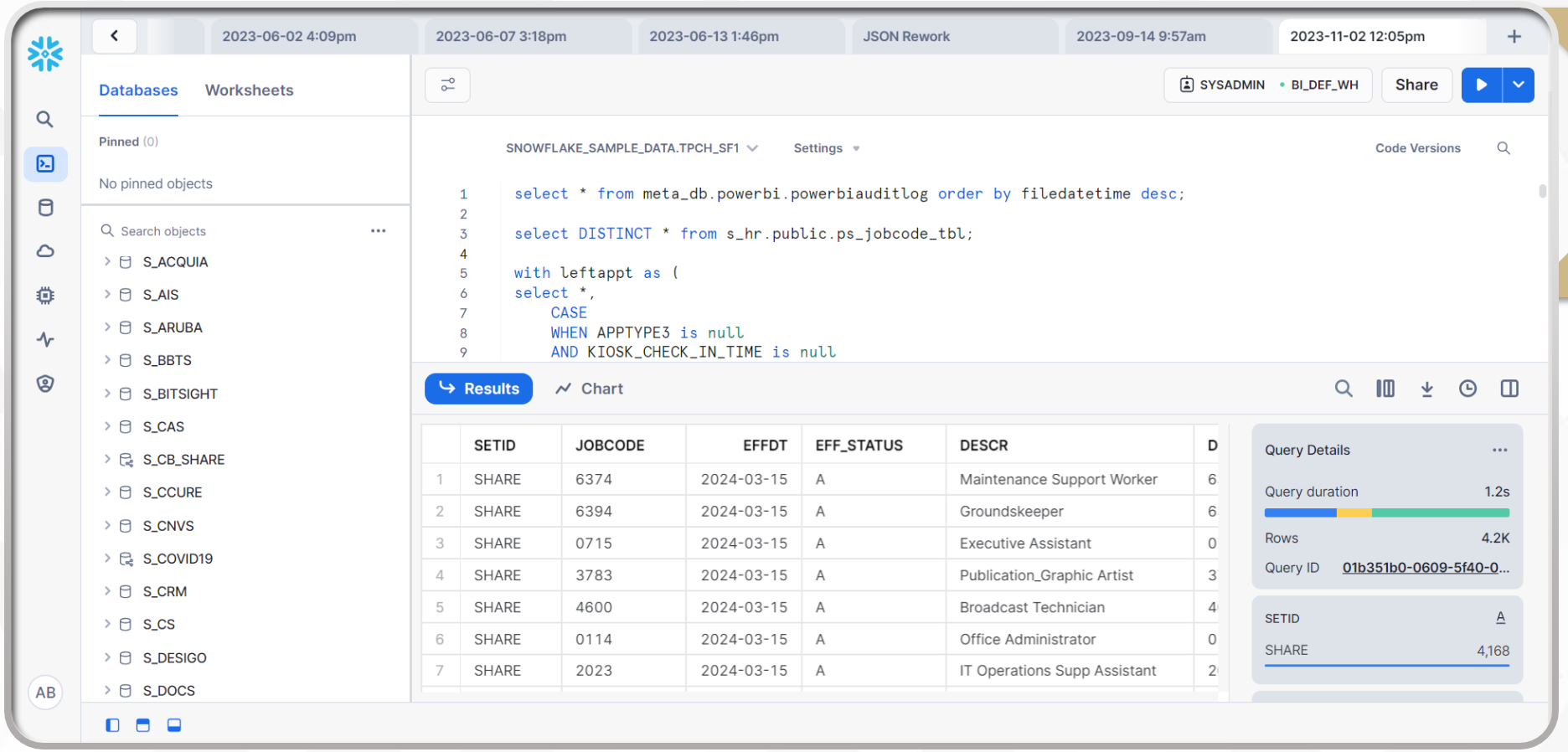
SQL Developer showing a query in DC



Snowflake

- Snowflake is a cloud-based data platform that allows for storage, processing, and analysis of large volumes of structured and semi-structured data.
  - It is both the name of the product and the data lakehouse built on it.
  - A data lakehouse combines the scalability and flexibility of a data lake with the performance and reliability of a data warehouse.
- Data is captured from various source systems and replicated into databases in Snowflake as close to the raw as possible.
- Data from the source systems is made available through the ODS database, which stands for operational data store.
- Snowflake offers scalability, security, and simplicity.

Worksheet in Snowflake



Analytics and Visualizations

- Tools for exploring and understanding data
  - Perform different types of analysis
  - Present results in visual or textual formats
- Help users discover patterns, trends, outliers, or anomalies
  - Generate insights and recommendations
  - Inform decision-making and action-taking

MyFSUBI (OBIEE)

OBIEE is a comprehensive suite of business intelligence tools

Provides interactive and flexible ways to access, analyze, and share data

Supports various types of analytics

Allows users to create and modify their own data models and visualizations

Used at FSU since 2010

FSU modified and enhanced many of the delivered models and reports

Provides more accurate, relevant, and timely information to users

Widely used by numerous groups around campus

Enables users to access and analyze data from various sources

Helps users monitor performance indicators and support strategic planning

Facilitates collaboration and communication among users

Power BI

Power BI is a cloud-based business intelligence platform that allows users to connect, explore, and visualize data from various sources.

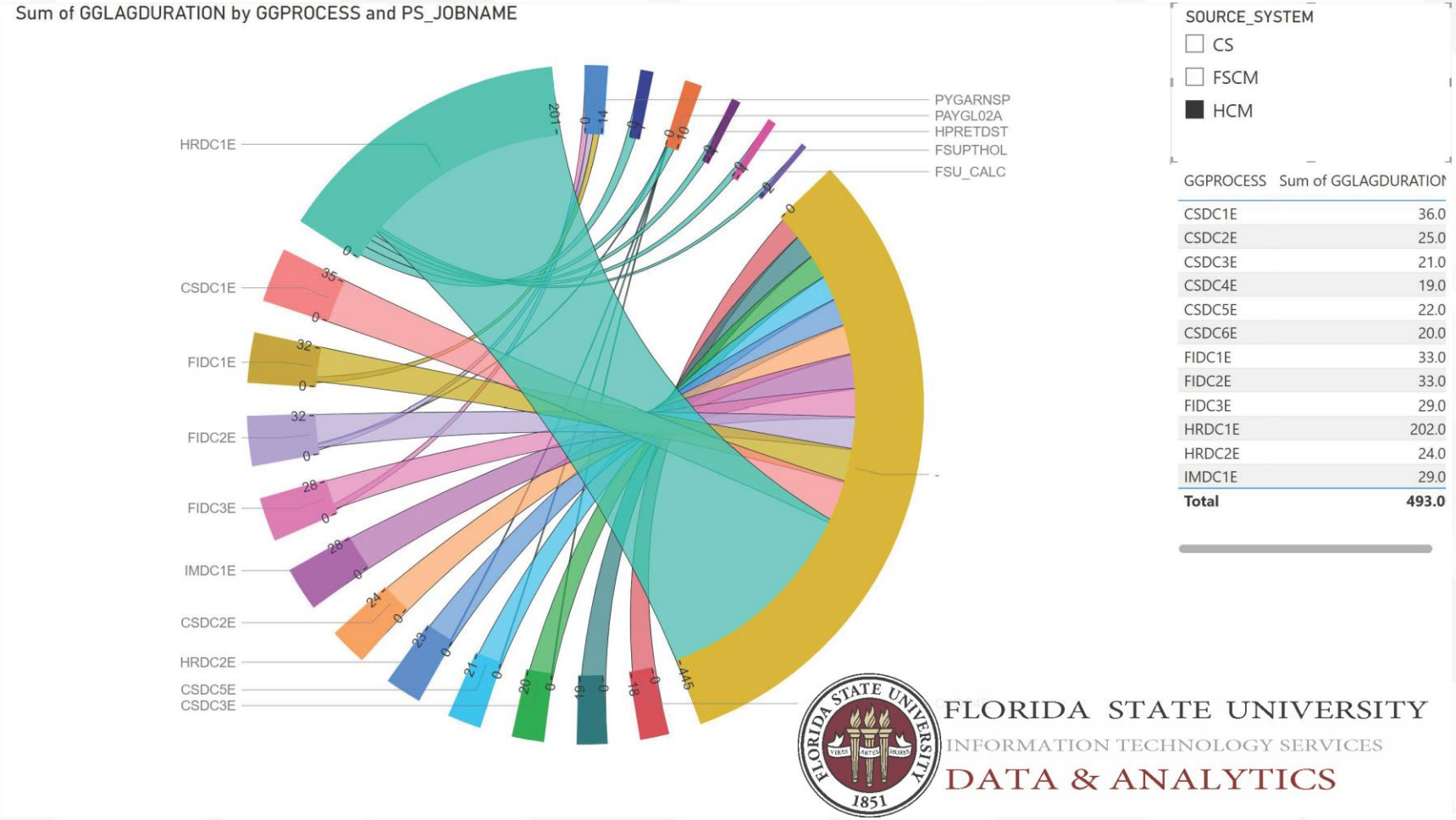
Users can create and customize their own data models, reports, dashboards, and apps, using a simple and intuitive interface.

Power BI supports a wide range of data sources, from structured databases and files, to unstructured web pages and social media, to real-time streaming data and APIs.

Power BI provides a rich set of data transformation and modeling tools, such as Power Query and Power Pivot.

Power BI offers a variety of data visualization options, such as charts, maps, tables, gauges, and KPIs.

Power BI Visualization

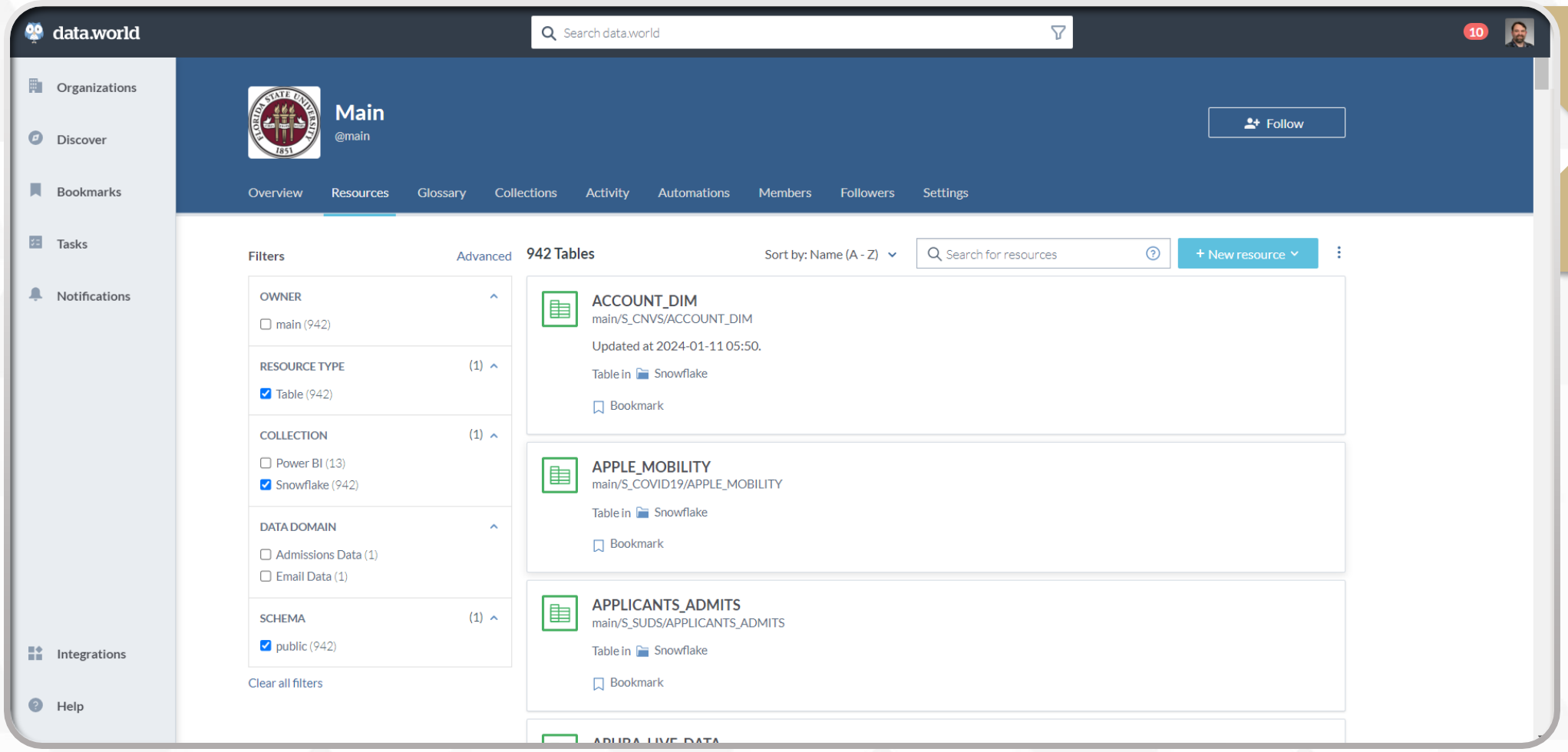


FLORIDA STATE UNIVERSITY
INFORMATION TECHNOLOGY SERVICES
DATA & ANALYTICS

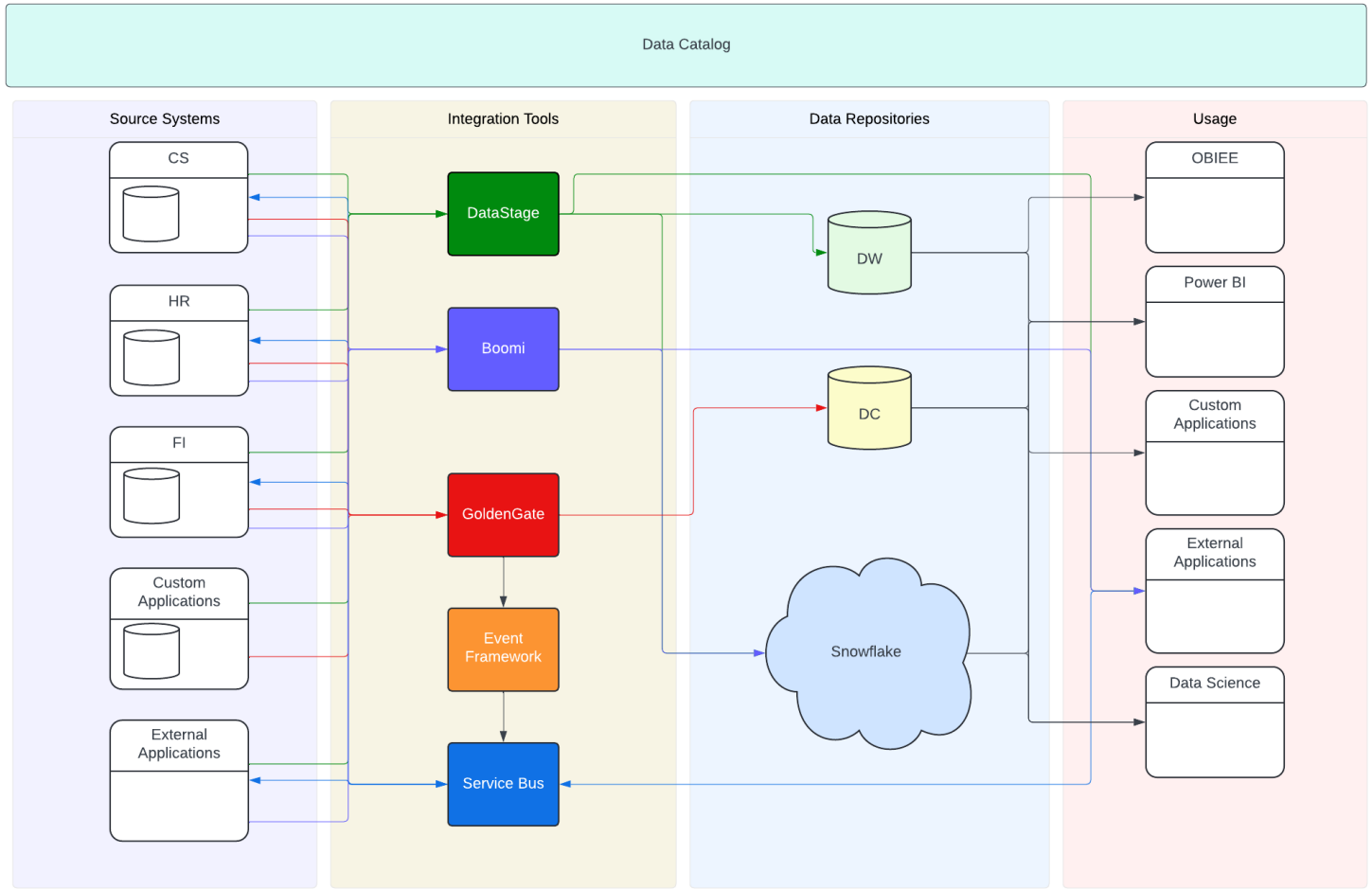
Data Catalog

- Data discovery enables users to explore data from various sources, identify patterns and trends, and gain insights that can drive action.
  - Empowers users to take control of their data analysis and make informed decisions.
  - Self-service functionality allows users to easily access information about the data they need.
- Challenges of data discovery include data quality, security, governance, and compliance.
  - Active metadata management addresses these challenges by providing a unified and dynamic view of the data across the enterprise.
- data.world is a tool that supports active metadata management.
  - Allows users to create and join data projects, access information about data from various sources, and collaborate with other users.

Search in data.world



Putting It Together



Follow up and next steps

- Questions or Specific Data Needs?
  - Reach out to Data & Analytics team for assistance – Salesforce or ITS-BIAnalyticsTeam@fsu.edu
- Connect with other data enthusiasts and experts
  - Join data-centric groups and communities at the university
    - Data Users Group
      - OBI Community of Practice
      - Power BI Community of Practice
    - DSA Data & Donuts

Please Provide Feedback!