Microsoft

Open Education Analytics
a modern education data estate

Dispersed data prevents strategic use of data and AI

For the Use Cases You Need

Personalized Learning Pathways

Real Time Data Dashboards

Open Data for Research and Evaluation

Identifying and supporting at risk students

State, Province and National Longitudinal Reporting

2

Data Engineers · Data Scientists · Contracted Data Scientists · Analysts · Learners, Educators · School Leaders, Researchers

1) Authentication/Authorization — Azure Active Directory

2) Data Processing Tools — Azure Synapse Analytics

3) Storage — Azure Data Lake Storage

On-premises data · Cloud data · SaaS data

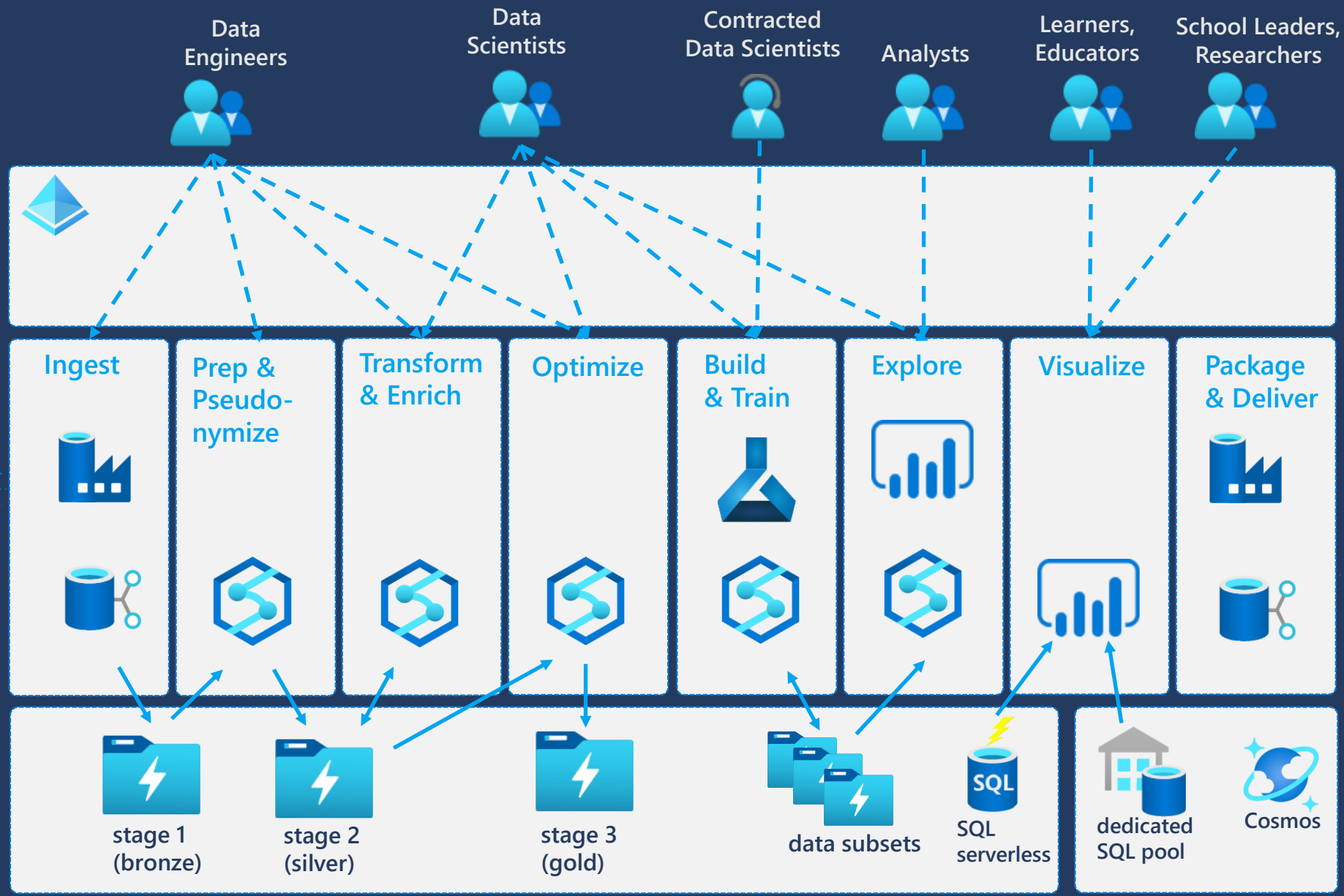Personalized Learning Pathways · Realtime Data Dashboards · Open Data for Research and Evaluation · State, Province and National Reporting

1) Collect · 2) Compute · 3) Communicate

Studio

A single place for Data Engineers, Data Scientists, and IT Pros to collaborate on enterprise analytics

Synapse Studio

Synapse Studio divided into **Activity hubs**.

These organize the tasks needed for building analytics solution.

Microsoft Azure | Synapse Analytics ▸ internalsandboxwe5

prlangad@microsoft.com
MICROSOFT

Home
Data
Develop
Orchestrate
Monitor
Manage

Synapse workspace
intern
New ⌄

Resources
Recent   Pi
NAME
CopyPi

Overview
Quick-access to common gestures, most-recently used items, and links to tutorials and documentation.

Data
Explore structured and unstructured data

Develop
Write code and the define business logic of the pipeline via notebooks, SQL scripts, Data flows, etc.

Orchestrate
Design pipelines that that move and transform data.

Monitor
Centralized view of all resource usage and activities in the workspace.

Manage
Configure the workspace, pool, access to artifacts

On-prem data

Cloud data

SaaS data

Ingest

Azure Data Factory

Azure Data Share

1) Collect

On-prem data

Cloud data

SaaS data

Ingest

stage 1

1) Collect

Microsoft Azure | Synapse Analytics ▸ syn-oea-cisd3demo4

Search

admin@contosoisd3.onmicrosoft.com
TEST_TEST_CONTOSO

Synapse live   Validate all   Publish all

Home

Data

Develop

Integrate

Monitor

Manage

Data

Workspace   Linked

Filter resources by name

▲ Azure Data Lake Storage Gen2   1

  ▲ syn-oea-cisd3demo4 (Primary - sto...

      public

      stage1

      stage2

      stage3

      synapse (Primary)

      test-env

  ▶ Integration datasets   8

Get_schools_classes...   stage1   ✕

New SQL script   New notebook   New data flow   New integration dataset   ··· More

stage1 ▸ AAD ▸ classes

Name   Last Modified   Content Type   Size

data_24d8bb08-003e-4989-975e-996746266ef9_37c0be75-10b2-443b-8b69-e725b52e9cf0....   6/11/2021, 4:53:30 PM   159.9 KB
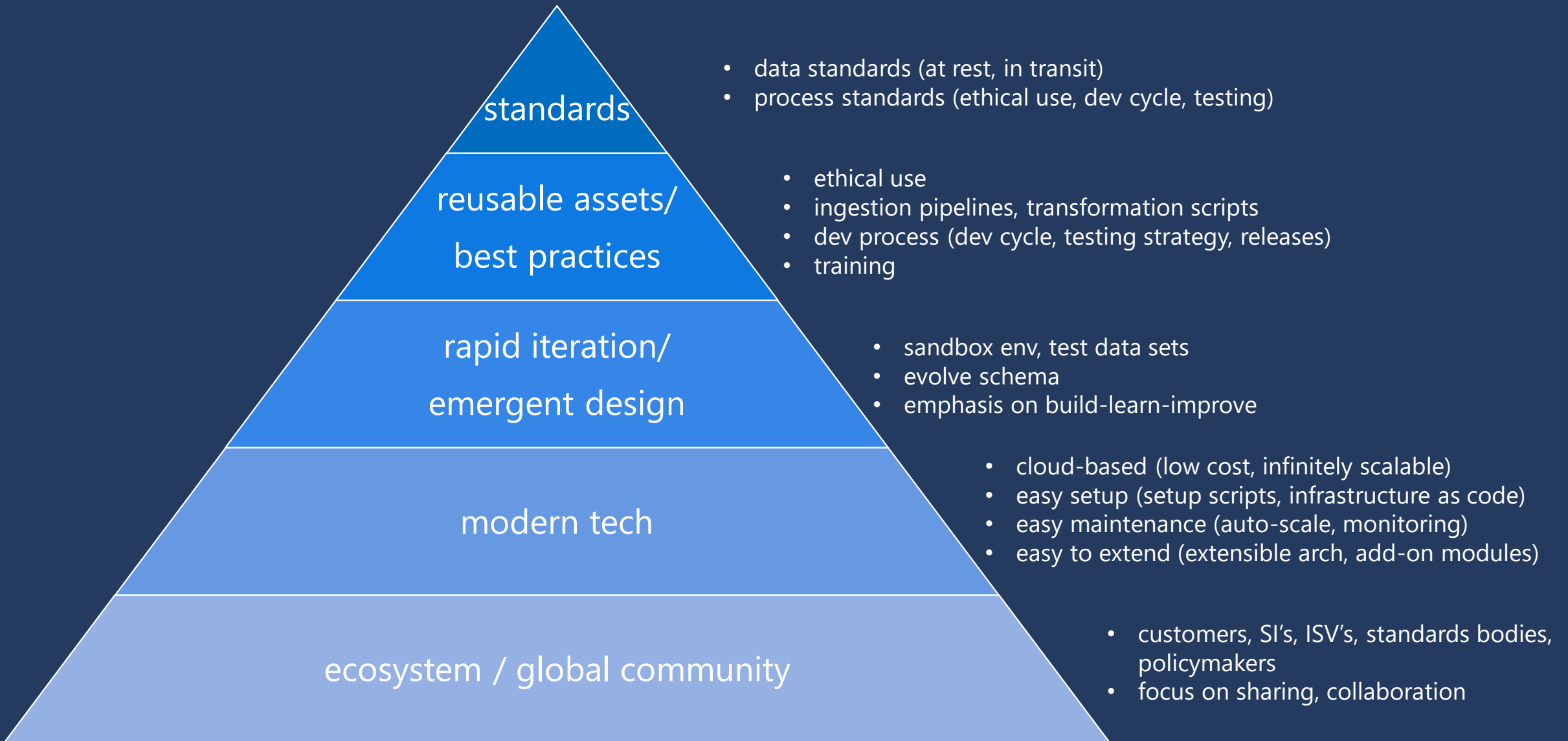
8

OEA Focus

- data standards (at rest, in transit)
- process standards (ethical use, dev cycle, testing)

standards

- ethical use
- ingestion pipelines, transformation scripts
- dev process (dev cycle, testing strategy, releases)
- training

reusable assets/ best practices

- sandbox env, test data sets
- evolve schema
- emphasis on build-learn-improve

rapid iteration/ emergent design

- cloud-based (low cost, infinitely scalable)
- easy setup (setup scripts, infrastructure as code)
- easy maintenance (auto-scale, monitoring)
- easy to extend (extensible arch, add-on modules)

modern tech

- customers, SI's, ISV's, standards bodies, policymakers
- focus on sharing, collaboration

ecosystem / global community

Search or jump to...

Pull requests   Issues   Marketplace   Explore

Code   Issues 1   Pull requests   Discussions   Actions   Projects   Wiki   Security   Insights   Settings

main   2 branches   2 tags

Go to file   Add file   Code

genegc Merge branch 'main' of https://github.com/microsoft/OpenEduAnalytics   b009e63   5 days ago   115 commits

.vs	add to readme	4 months ago
docs	Add files via upload	last month
modules	Update readme.md	10 days ago
packages	updated setup scripts.	last month
CODE_OF_CONDUCT.md	Initial CODE_OF_CONDUCT.md commit	7 months ago
CONTRIBUTING.md	Create CONTRIBUTING.md	last month
LICENSE	Initial LICENSE commit	7 months ago

Cell 3

```
1    # Process Roster data from stage 1 to stage 2
2    #
3    # Sets up the edu_dl (stage 2 data lake) with whatever data is found in the DIP inbound folder.
4    # This includes:
5    # - adding column names
6    # - casting values into a schema
7
8    # Calendar
9    df = spark.read.csv(stage1_path + '/Roster/Calendar.csv', header='false')
10   if (df.count() > 0):
11       sqlContext.registerDataFrameAsTable(df, 'Calendar')
12       df = spark.sql("select _c0 Id, _c1 Name, _c2 Description, cast(_c3 as int) SchoolYear, cast(_c4 as b
13       df.write.format("parquet").mode("overwrite").save(f'{stage2_path}/Calendar')
14   # Course
15   df = spark.read.csv(stage1_path + '/Roster/Course.csv', header='false')
16   if (df.count() > 0):
17       sqlContext.registerDataFrameAsTable(df, 'Course')
18       df = spark.sql("select _c0 Id, _c1 Name, _c2 Code, _c3 Description, _c4 ExternalId, to_timestamp(_c5
```
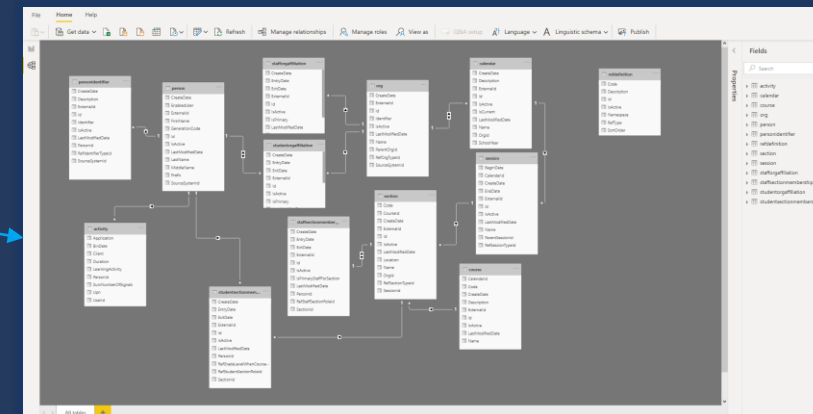
Microsoft

Open Edu Analytics
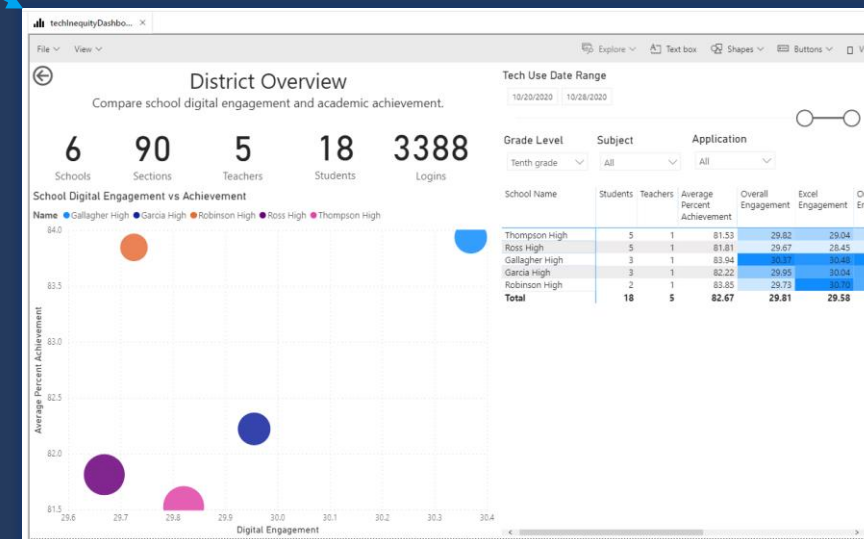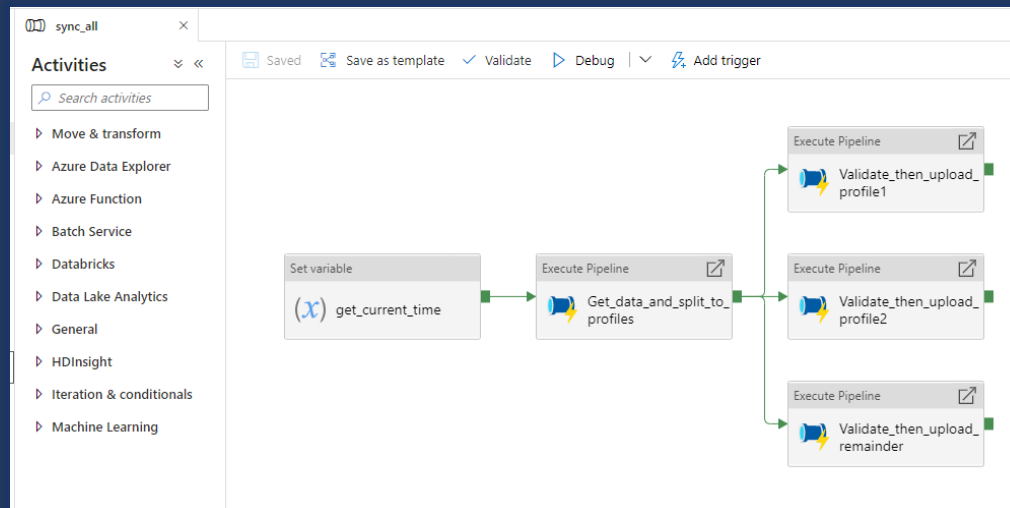Implementation Guide

Published: October, 2020

Introduction 1
1) Setup Data Share in School Data Sync 1
2) Setup Resources in Azure Portal 2
3) Setup Synapse 5
4) Setup Power BI 7

sync_all

Activities

Search activities

Move & transform
Azure Data Explorer
Azure Function
Batch Service
Databricks
Data Lake Analytics
General
HDInsight
Iteration & conditionals
Machine Learning

Saved   Save as template   Validate   Debug   Add trigger

Set variable
get_current_time

Execute Pipeline
Get_data_and_split_to_profiles

Execute Pipeline
Validate_then_upload_profile1

Execute Pipeline
Validate_then_upload_profile2

Execute Pipeline
Validate_then_upload_remainder

techInequityDashbo...

File   View

District Overview
Compare school digital engagement and academic achievement.

Tech Use Date Range
10/20/2020   10/28/2020

6 Schools   90 Sections   5 Teachers   18 Students   3388 Logins

Grade Level: Tenth grade   Subject: All   Application: All

School Digital Engagement vs Achievement

Education Analytics and Insights
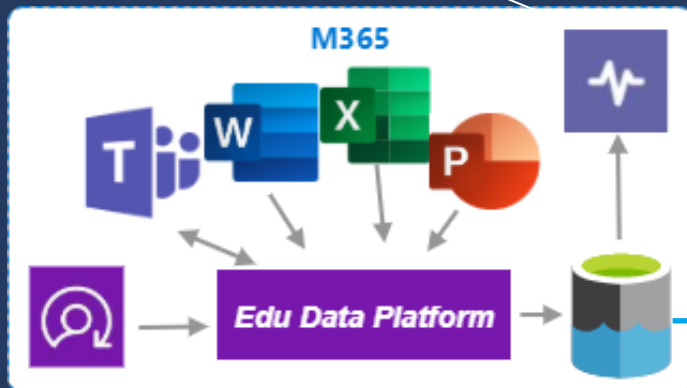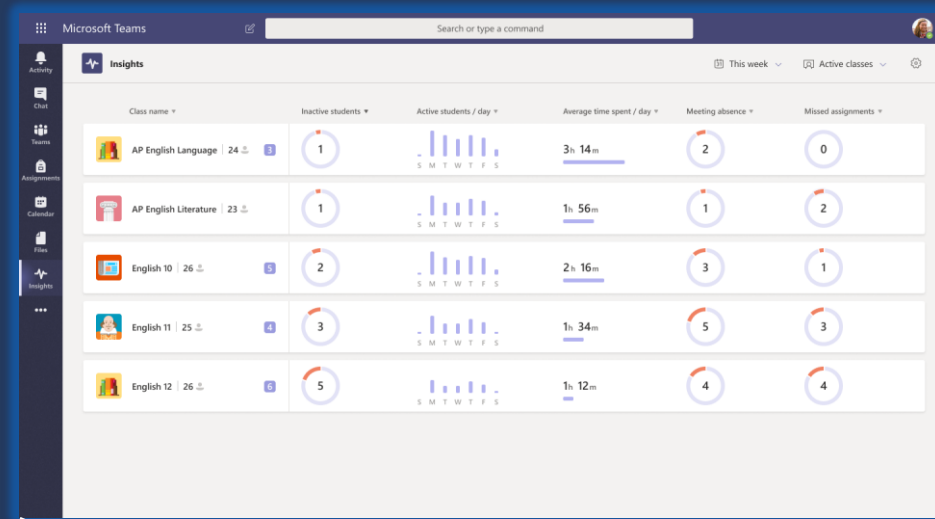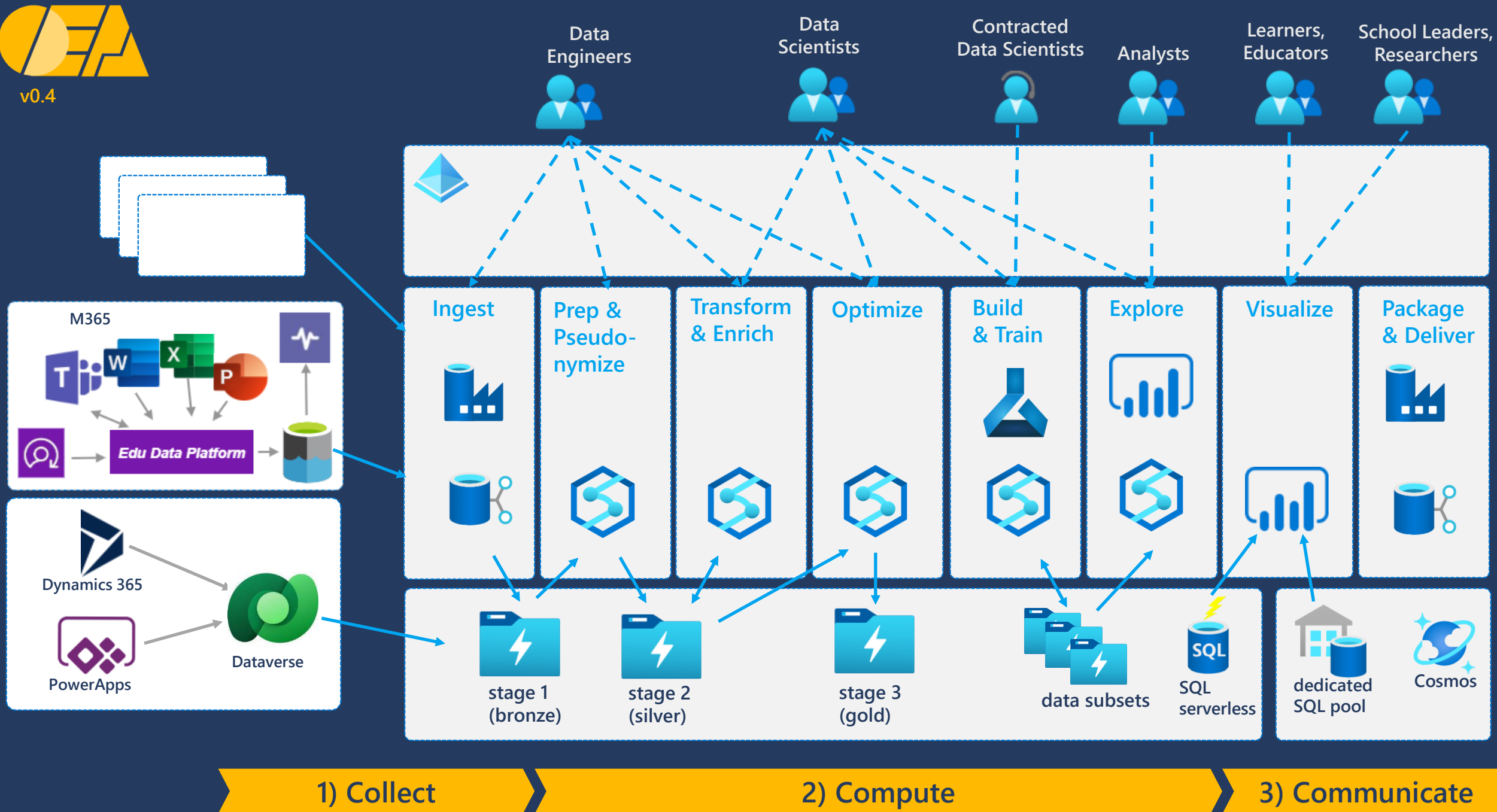
Multi-tenant architecture