UC San Diego

THE UNBEARABLE UNIFICATION OF EVERYTHING

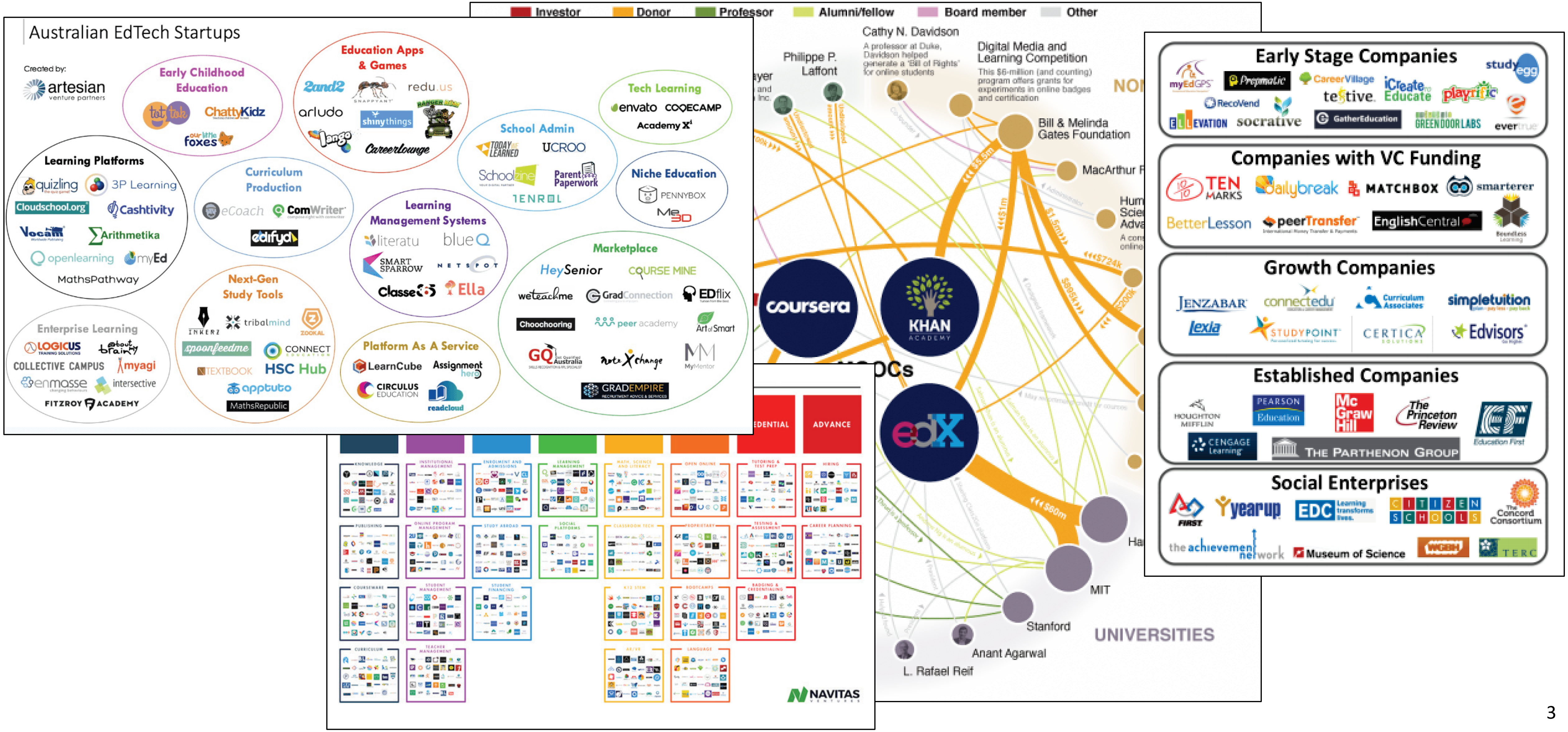
Vince Kellen, PhD

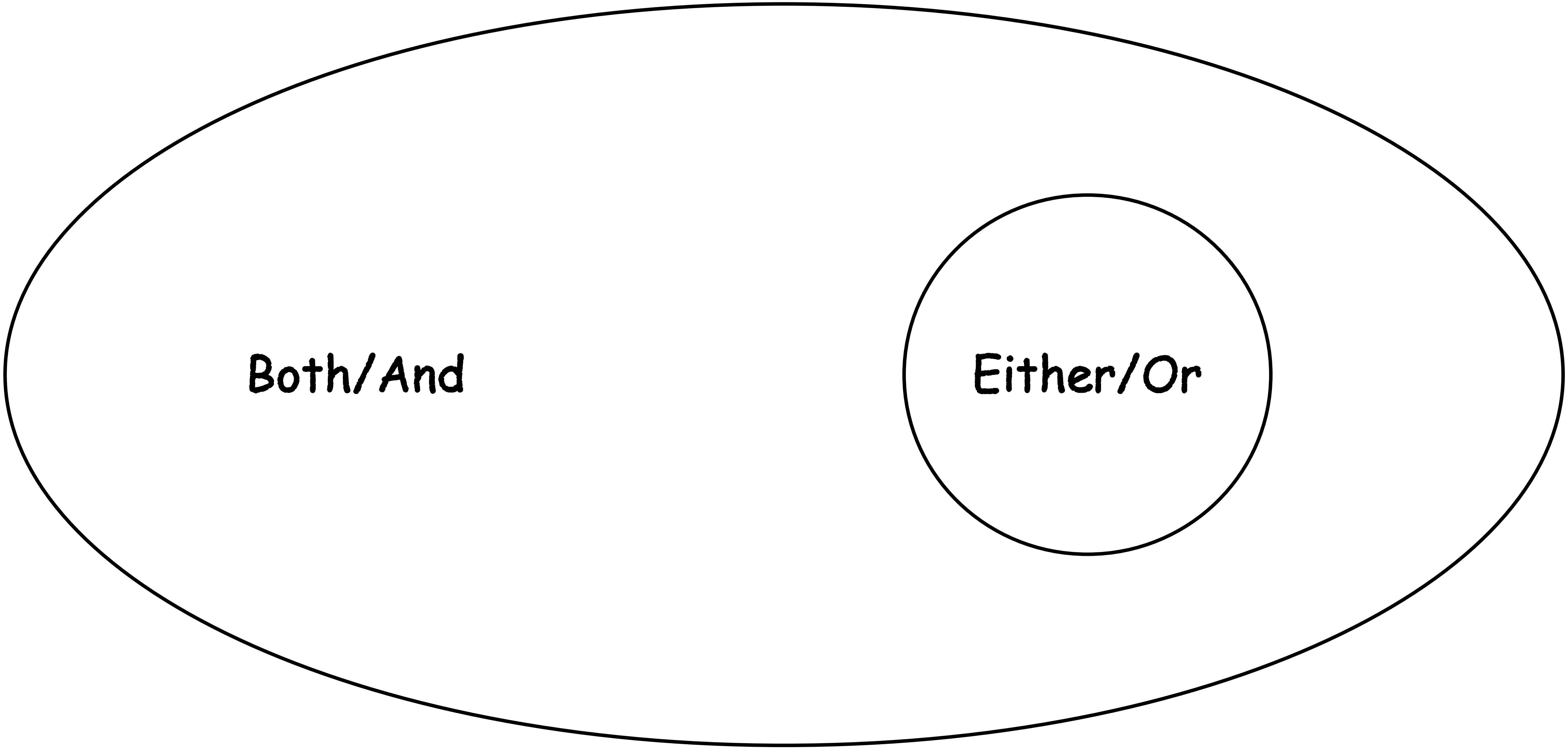
Chief information Officer

March 20, 2019

Multiverse

EdTech Ecosystem: a universe of universes





Both/And

Either/Or

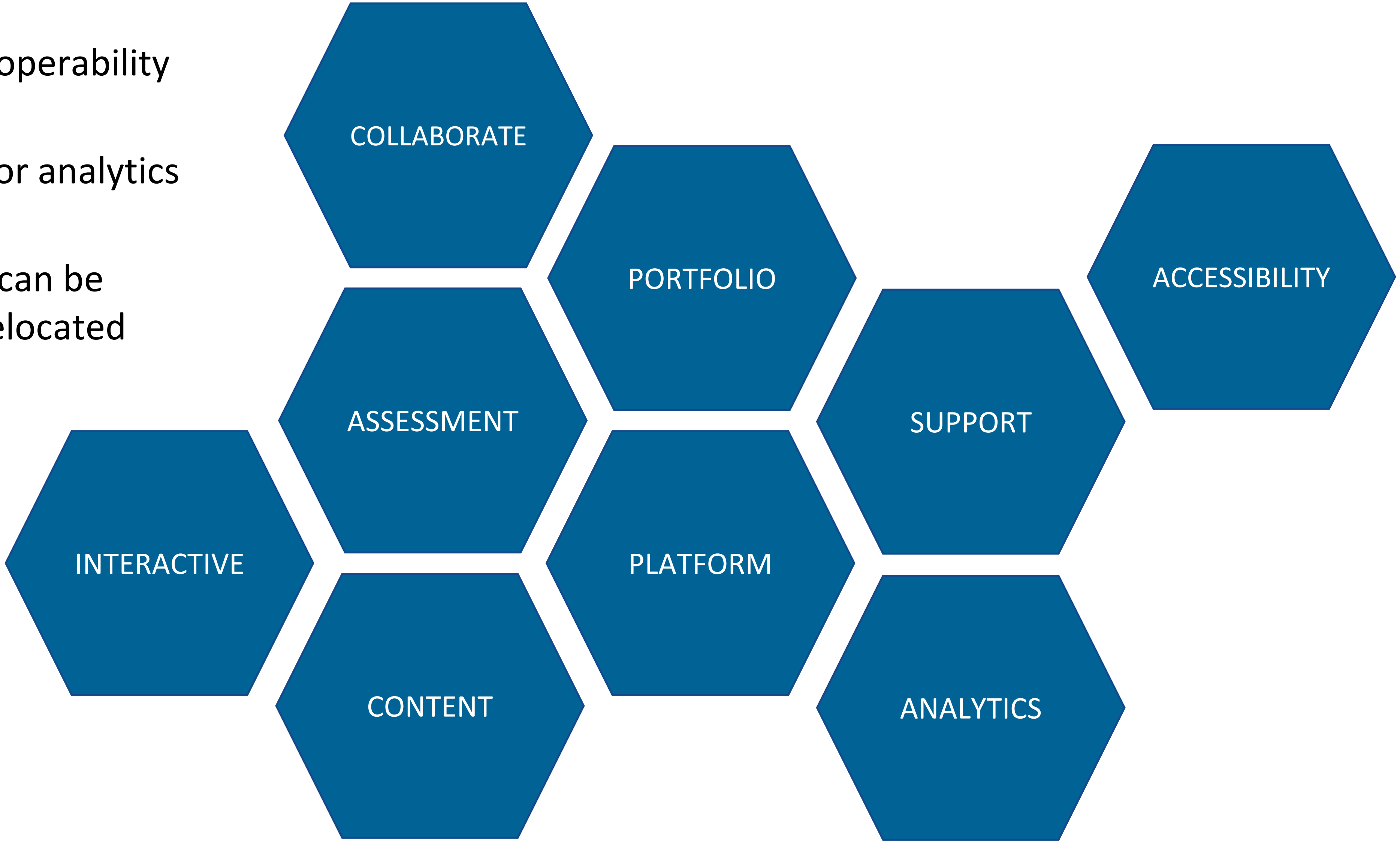
UCSD Online platform/ecosystem considerations

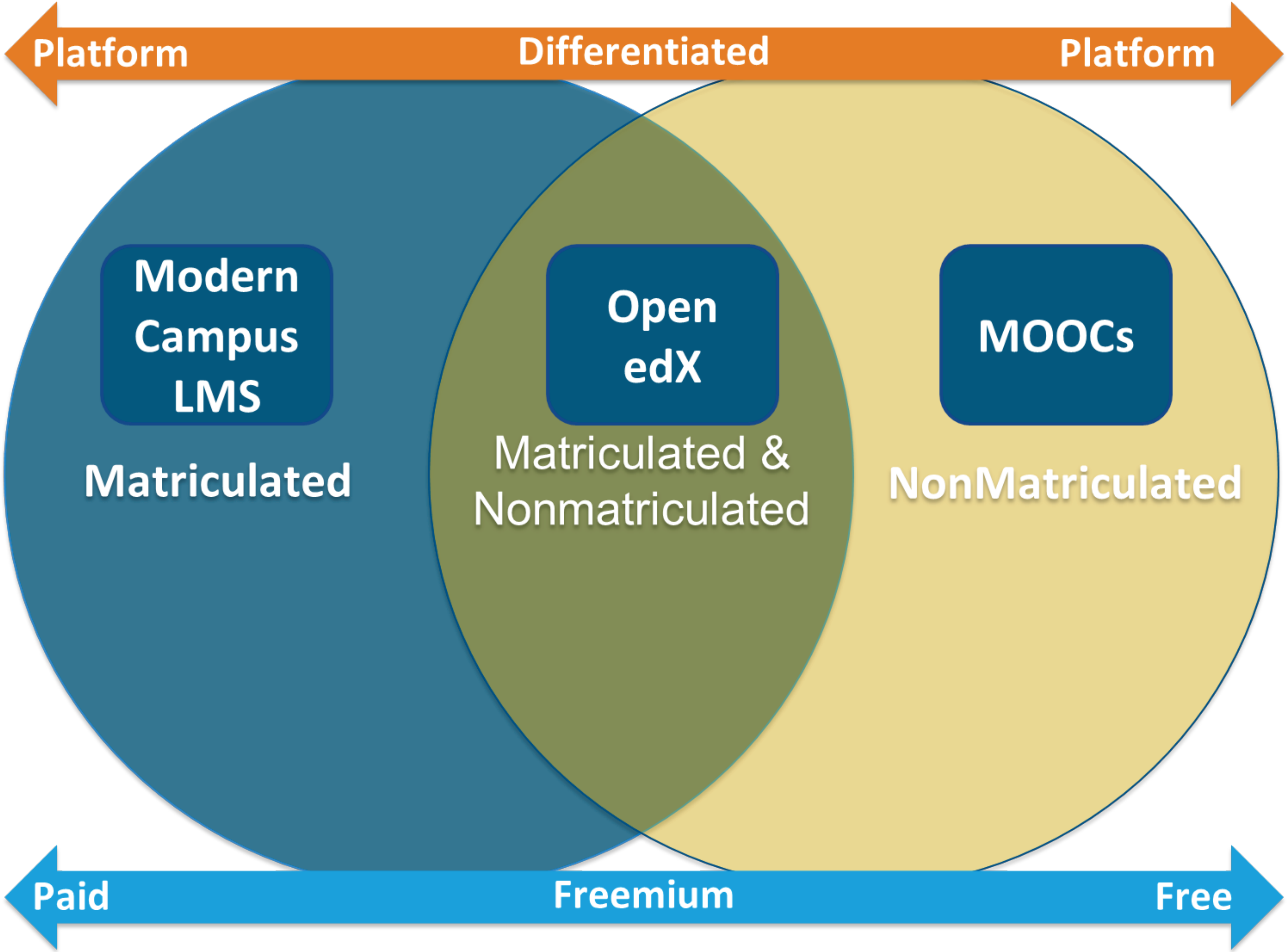
- UCSD should deploy offerings across multiple platforms to allow for a portfolio of online offerings, from Free to Freemium to Paid
- UCSD should choose, deploy offerings expecting EdTech to be a very dynamic market
- UCSD should take a modular, ecosystem approach based on industry standards, particularly IMS Global standards such as LTI Advantage and Caliper.
- An ecosystem approach is desirable in order to facilitate multiple architectures best suited for particular programs and offerings.
- Platform and ecosystem choices should privilege a real-time analytics capability and roadmap that works with UCSD’s Student Activity Hub analytics approach.
- Content should be centrally managed and distributed to platforms and tools within the UCSD Online ecosystem.
- The best disciplinary pedagogies and supporting technologies for UCSD Online learner experiences should drive platform and ecosystem decisions.

Q.E.D. - Use a modular, ecosystem approach

Use industry interoperability standards to:

- Integrate data for analytics and messaging
- Ensure content can be easily reused, relocated





Platform Differentiated Platform

Modern Campus LMS

Matriculated

Open edX

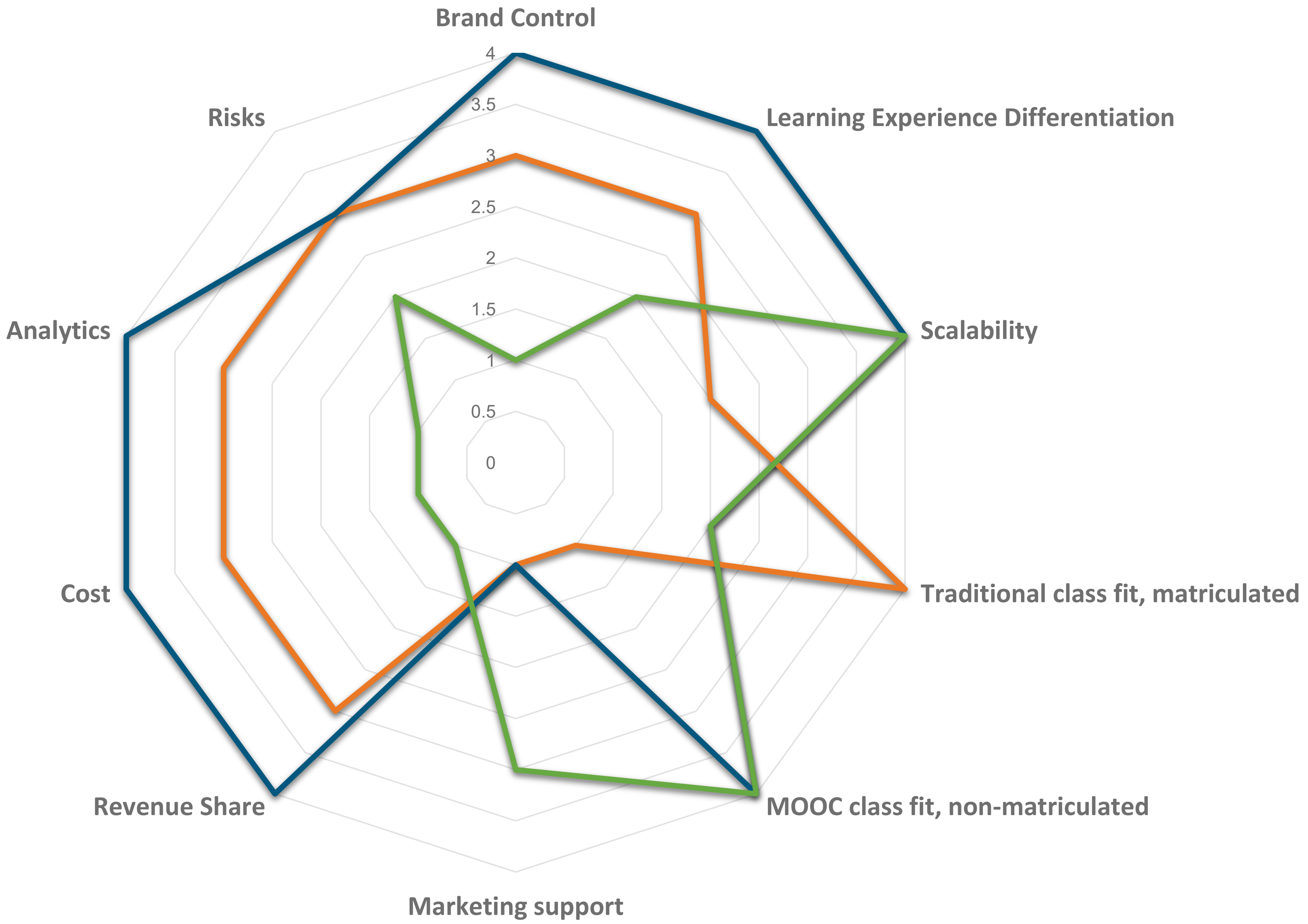
Matriculated & Nonmatriculated

MOOCs

NonMatriculated

Paid Freemium Free

Platform Comparison



Proposed UCSD Online Platform Rationale

Modern Campus LMS

- Modern, evolving, and easy-to-use interface
- Designed to serve traditional pedagogies in traditional course structures at standard enrollment scale for matriculated students
- Mobile-first development model aligns with student expectations and needs
- Strong user community to influence product roadmap
- Standards-based to permit easy integration of third-party solutions and tools

- Online Master's Programs
- One-year Master's programs
- Online Credit Courses
- Concurrent Enrollment (Cohort & Non-Cohort)
- Summer Credit (PS and Secondary visitors)

Open edX

- Gives UCSD control of their brand and content
- Rationalizes data, analytics, and content within agreed standards and interoperability
- Scales teaching, learning, and student success practices from one classroom to a global audience
- Supports non-conventional learning experiences (e.g., WeAreTritons and related compliance training)
- Scales to support distinct approaches to teaching and learning
- Designed to support offerings at MOOC scale in non-traditional structures for matriculated and non-matriculated students

- Professional/Continuing Ed (Certificates)
- Stackable Certificates/Micromasters (Long-term)
- Online Master's Programs (Experience Differentiated)

MOOCs

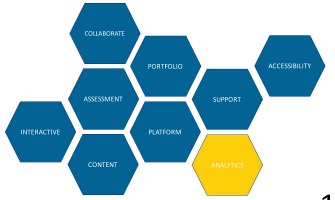
- Vendor-maintained platforms that allow UCSD to focus on course content
- Marketing boost from affiliation with a global cohort of large universities
- Certificates that are consistent across programs and institutions

- Stackable Certificates/Micro-masters (Short-Term)
- Funnel to attract learners to other UCSD Online Offerings via MOOCs

Online Learning: Analytics

Benefits of robust learning analytics capability

- **Student Success Support –** robust analytics combined with multi-channel messaging with “nudges” for students is likely to improve student success in online programs. Real-time events, alert and analytics can help coordinate student advisors, coaches and faculty
- **Learning Research –** investments in learning analytics capabilities will provide a wealth of data to faculty researchers seeking to understand best practices in online and hybrid teaching and learning
- **Personalized Learning –** by better understanding points of student need or excellence in real-time, where needed, a personalized learning experience can be developed for learners with just-in-time supplemental support or additional challenge materials.



Online Learning: Analytics

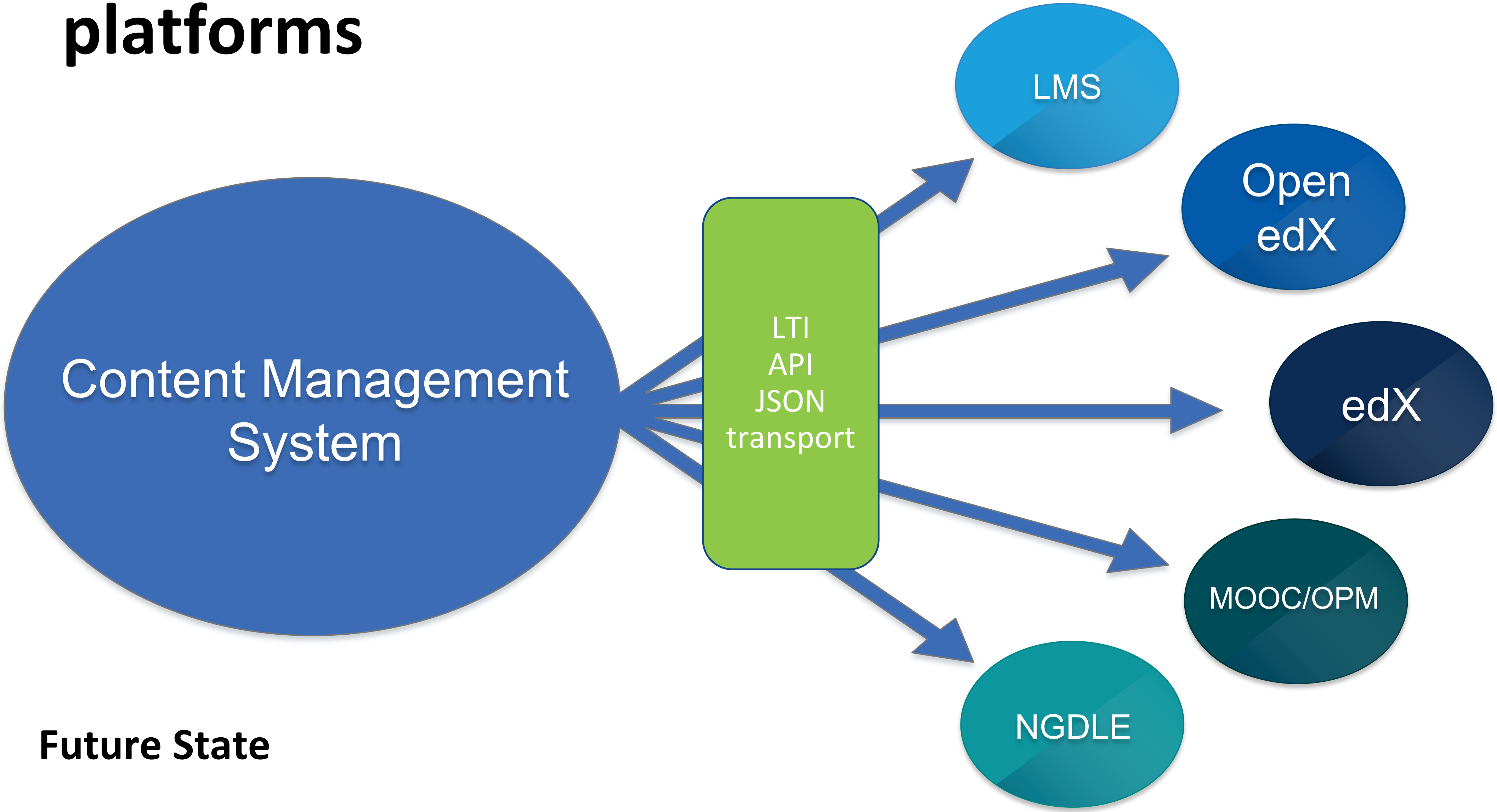
Standards for content and data integration

- The development of a robust learning platform analytics environment depends upon vendors adopting data and content interoperability standards and enforcing them within the UCSD ecosystem
- For learning event (clickstream) data, Caliper, the standard with the best functionality and growing higher education vendor adoption, should be the preferred standard
- LTI and LTI Advantage are the preferred standards for content integration
- Where LTI or Caliper are not available, well-developed APIs are another avenue for data consumption and provision
- Data Integration Design Principles:
  - All ecosystem components should be LTI & Caliper Compliant
  - For learning event data, real-time Incremental data streaming is to be preferred
  - Bidirectional API for most (or all) core data should be available
  - Nightly data dumps/loads should be a last resort, used only when required

Wherever possible, select tools with Level 2 and 3 of architectural capability

	EdTech Data Standards	EdTech Content Standards	APIs	Realtime, Incremental Streaming	Bi-directional Data & Content (Batch Loads)	Cloud Capable
Level 1	None	None	None	None	Available	Yes
Level 2	Caliper	LTI	None	Limited	Available	Yes
Level 3	Caliper	LTI	90% data and content coverage; bidirectional	Available for 100% of streaming-eligible data	Available	Yes

Ease of content relocation is a critical factor in providing online offerings across multiple platforms

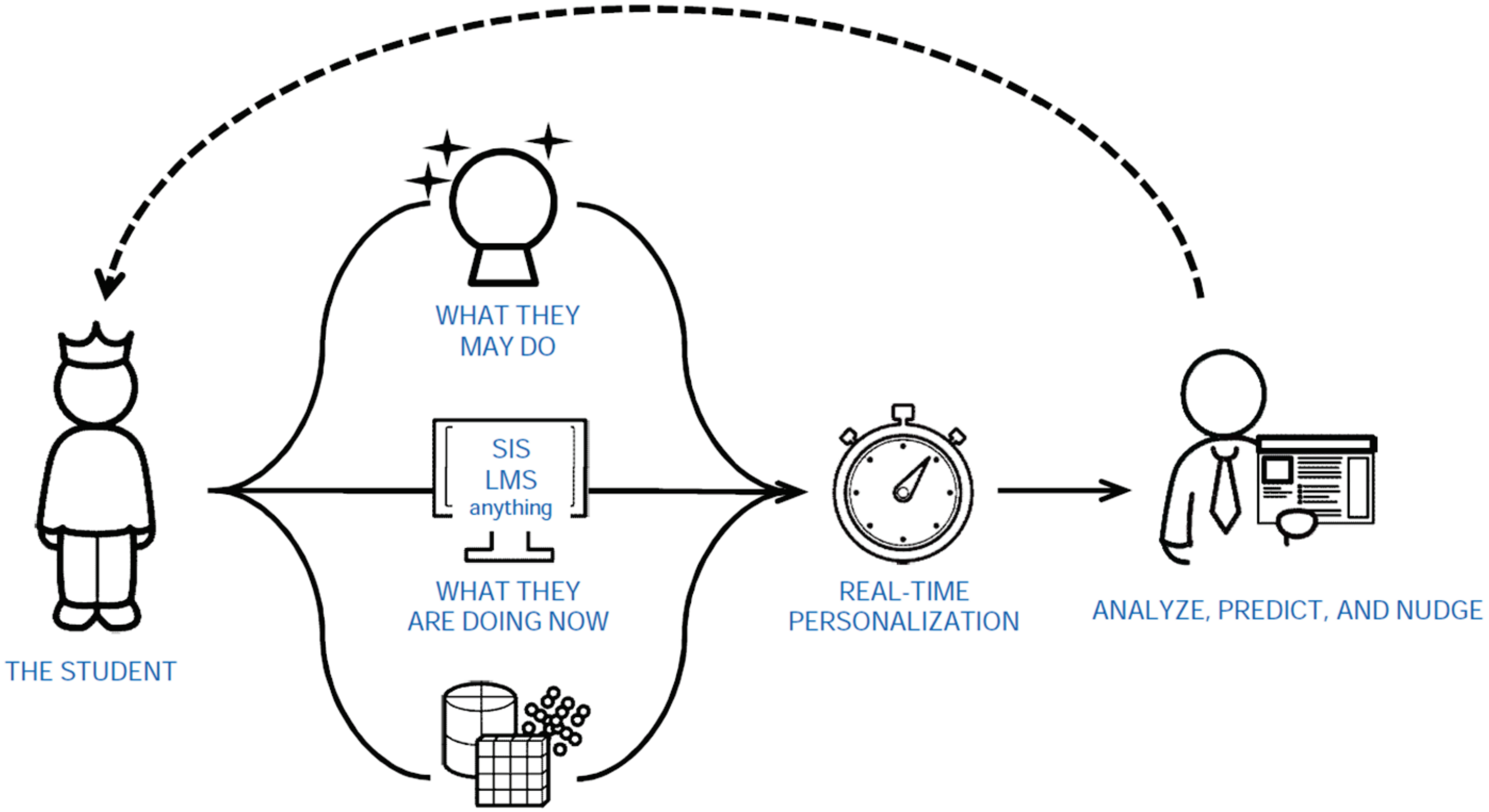


Future State

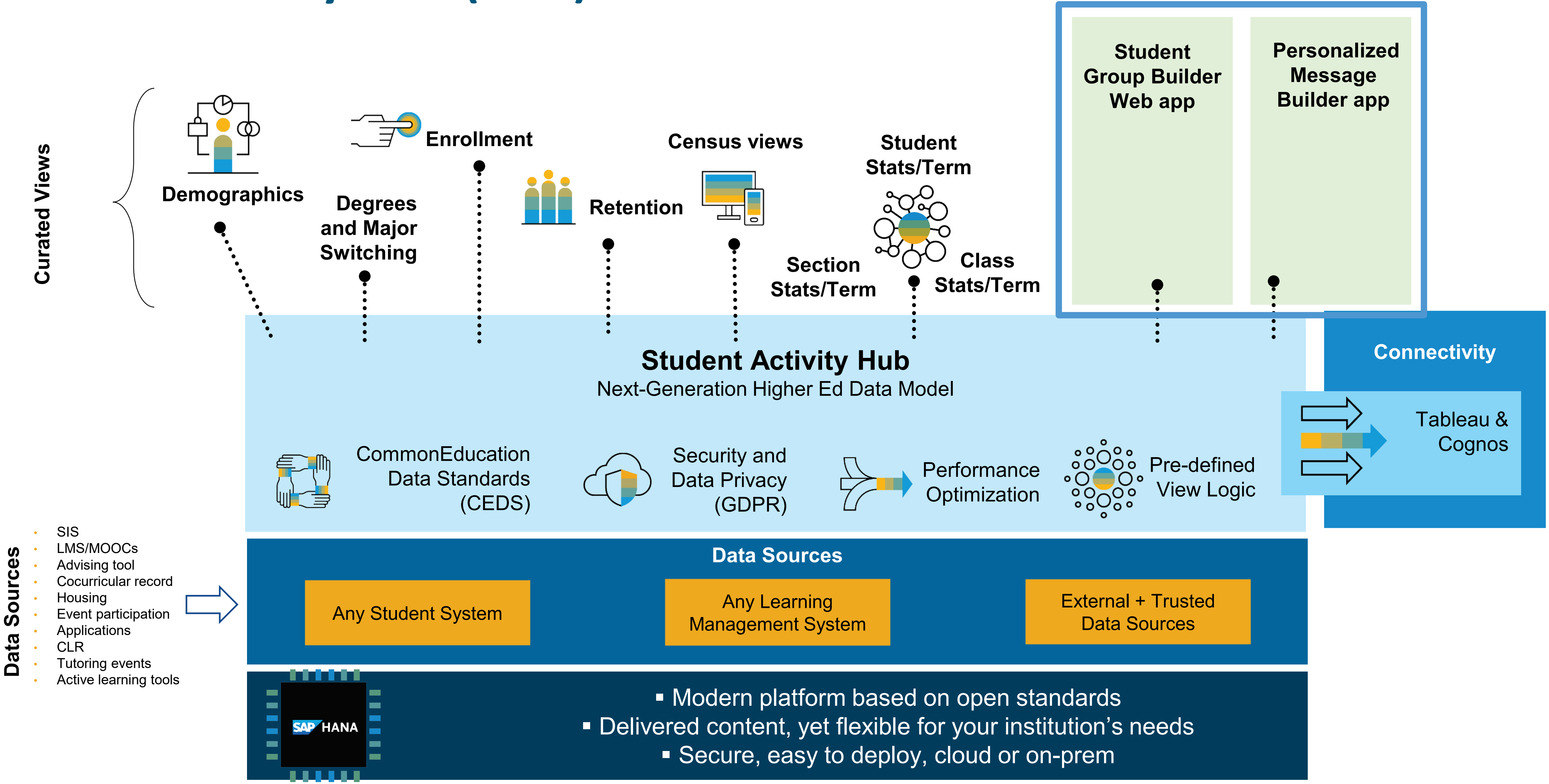
Analytics: Jupiter Mission

Student analytics scope

1. Give analysts access to anonymized views
2. Enable real-time, personalized mobile messaging, alerting, etc.
3. Allow for rich, comprehensive, large scale learning analytics



Student Activity Hub (SAH) Platform Overview



Curated Views

Demographics

Degrees and Major Switching

Enrollment

Retention

Census views

Student Stats/Term

Section Stats/Term

Class Stats/Term

Student Group Builder Web app

Personalized Message Builder app

Student Activity Hub

Next-Generation Higher Ed Data Model

CommonEducation Data Standards (CEDS)

Security and Data Privacy (GDPR)

Performance Optimization

Pre-defined View Logic

Connectivity

Tableau & Cognos

Data Sources

- SIS
- LMS/MOOCs
- Advising tool
- Cocurricular record
- Housing
- Event participation
- Applications
- CLR
- Tutoring events
- Active learning tools

Data Sources

Any Student System

Any Learning Management System

External + Trusted Data Sources

SAP HANA

- Modern platform based on open standards
- Delivered content, yet flexible for your institution’s needs
- Secure, easy to deploy, cloud or on-prem

SAH: Group and message builder

Student group builder

Analyze student and learning activities to uncover trends

Filter and group students according to different attributes

Explore (and save) results in graphs and list format

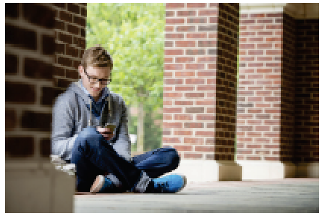
Group management

Store groups – including static and dynamic groups

Track group membership over time

Compare and analyze groups

Use groups as “attributes” in BI tools



Personalized messaging

Automatically generate user-defined messages

Use message templates and embed variables

Tie message recipients to student groups

Group builder and message builder tools interact. Group builder allows for:

- ✓ Grouping students together via any combination of fields and selection criteria (full set operations and Boolean logic)
- ✓ Changes in group membership creates events (“added to group”, “removed from group”) that can trigger messages, emails or workflow
- ✓ Groups also integrate with all analytics, allowing analysis to quickly compare and contrast different subpopulations of students. Subpopulations can be overlapping
- ✓ Groups are reusable and sharable and can be easily referenced within all workbooks and reports

“Curated views” of the data, de-identified

Demographics
Residency, SAT/ACT and other entrance test scores, academic status, etc.

Enrollment
Enrollment counts by class, departments, divisions/schools, colleges, including course grades

Major/Minors (wide and narrow)
Degrees, Programs, switching of majors, etc.

Retention (wide and narrow)
Cohort, retention and graduation rates, etc.

Admissions
Applicants, Applications, Test Scores, Scholarships

Student Statistics Per Term
Dozens of common student statistics, term-by-term for examining progression

Class and Section Stats Per Term
Dozens of class and section statistics, term by term for course and section planning, instructor load, etc., course performance correlations

Continuing education students (Extension, other)
Demographics, enrollment, credentials

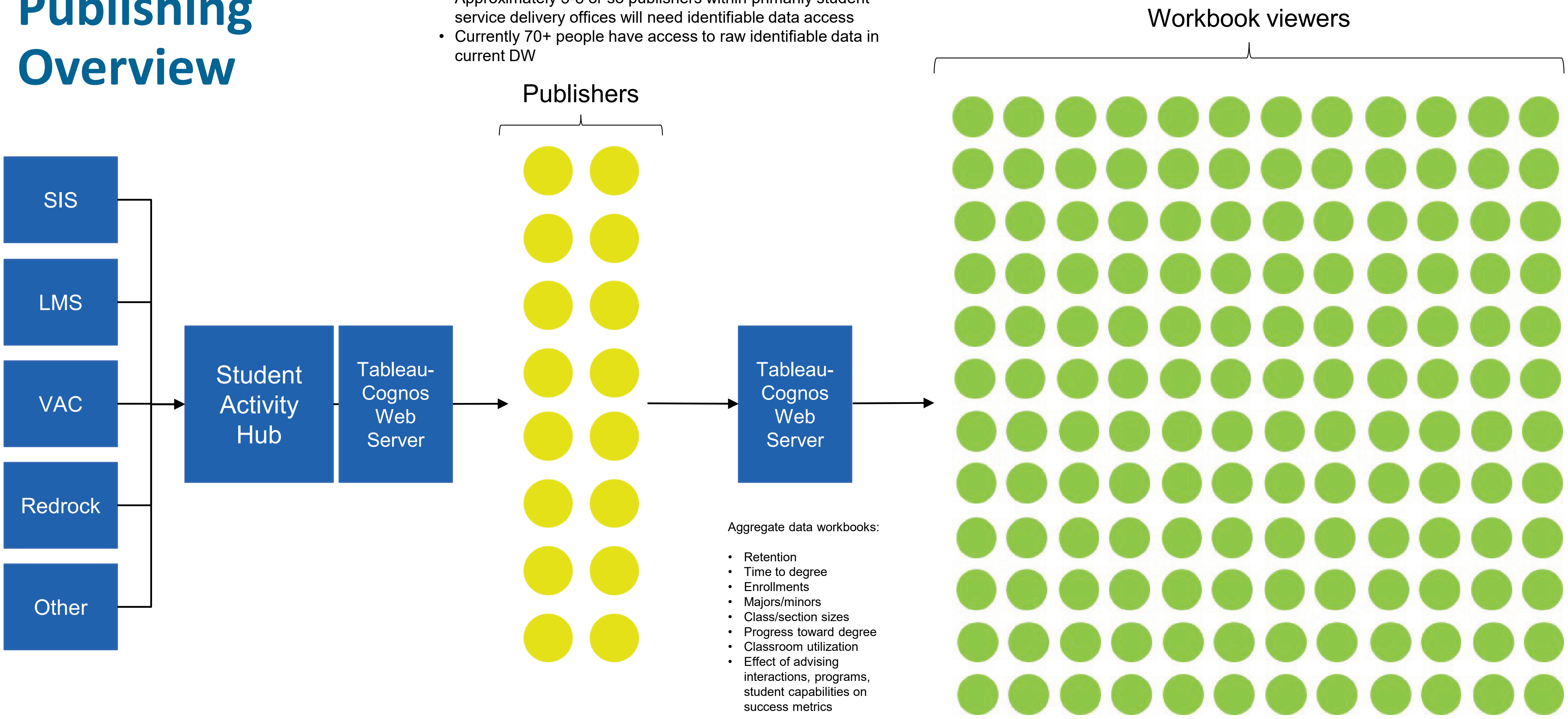
Learning analytics
Learning events, grading events, comprehensive learner record

Student engagement
Advising, tutoring, co-curricular events

Student Activity Hub (SAH) Data Publishing Overview

- Legitimate educational interest only; skilled analyst
- Using Tableau Desktop, other authoring tool, secure access
- Creates dashboards, interactive analytic screens, reports
- Access to granular, de-identified data only, control small cell size if needed
- Approximately 30-40 split between central and distributed groups
- Approximately 5-8 or so publishers within primarily student service delivery offices will need identifiable data access
- Currently 70+ people have access to raw identifiable data in current DW

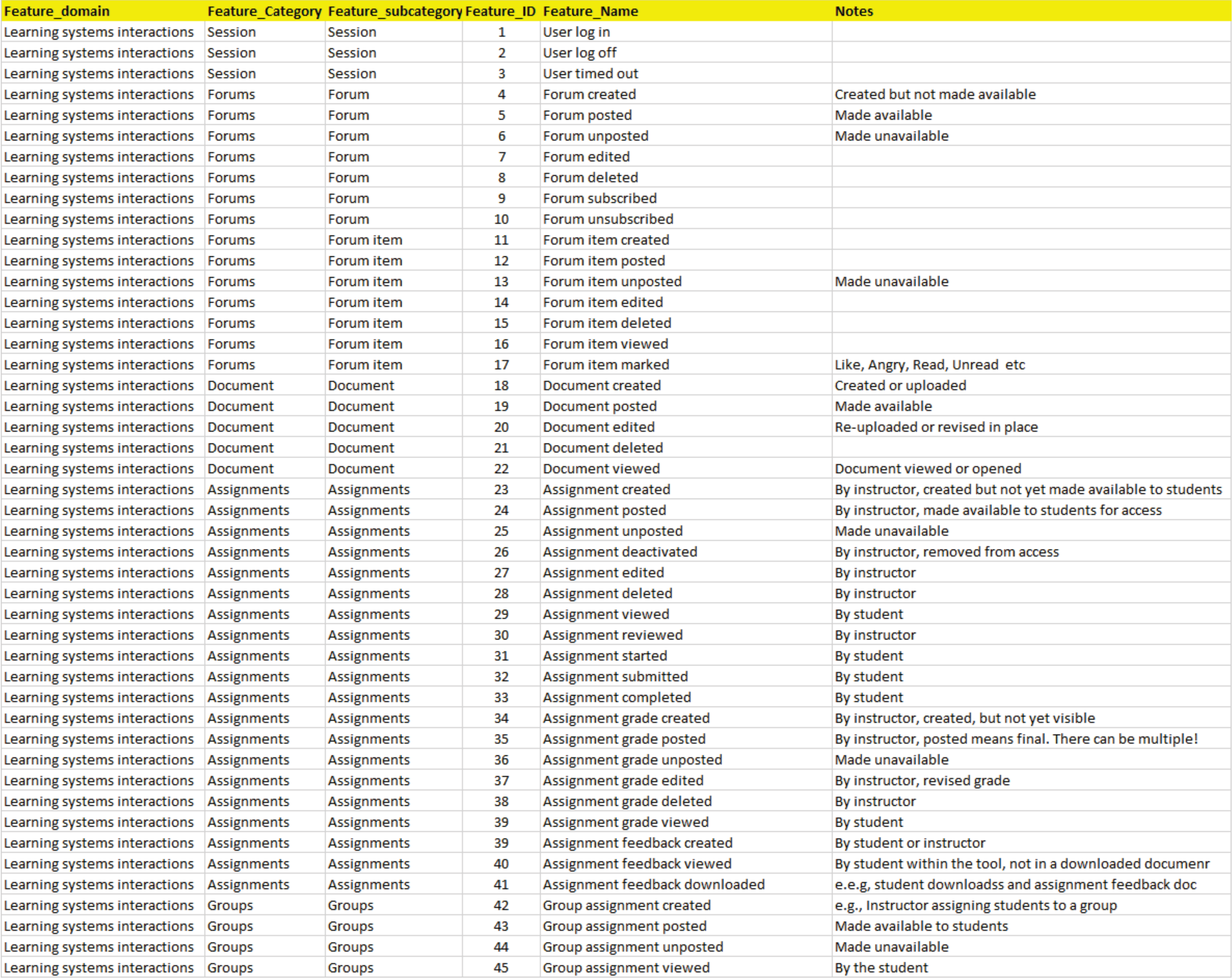
- Legitimate interest only; staff, faculty with secure UCSD credentials
- Accesses published workbooks via the web
- No direct data access, no identifiable data, no downloading of data
- Can manipulate the data in the workbook only to the degree the publisher allows
- Access to identifiable data, lists of students, etc. is only through the VAC or an authorized report



Aggregate data workbooks:

- Retention
- Time to degree
- Enrollments
- Majors/minors
- Class/section sizes
- Progress toward degree
- Classroom utilization
- Effect of advising interactions, programs, student capabilities on success metrics

Master map of learning/other events



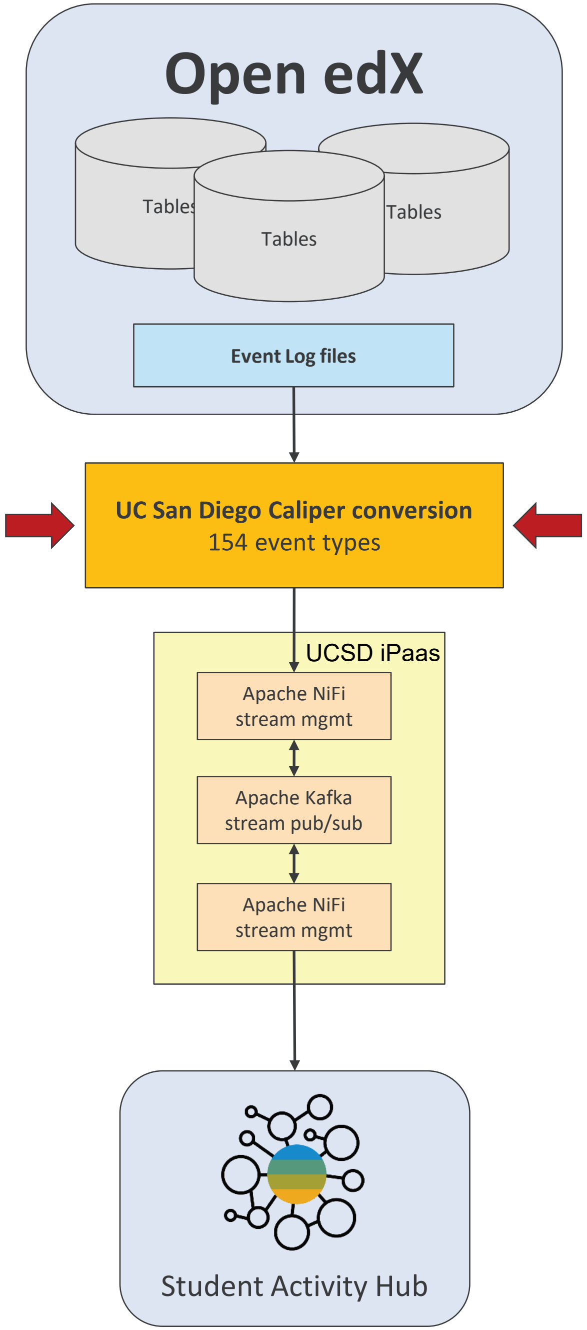
Feature_domain	Feature_Category	Feature_subcategory	Feature_ID	Feature_Name	Notes
Learning systems interactions	Session	Session	1	User log in	
Learning systems interactions	Session	Session	2	User log off	
Learning systems interactions	Session	Session	3	User timed out	
Learning systems interactions	Forums	Forum	4	Forum created	Created but not made available
Learning systems interactions	Forums	Forum	5	Forum posted	Made available
Learning systems interactions	Forums	Forum	6	Forum unposted	Made unavailable
Learning systems interactions	Forums	Forum	7	Forum edited	
Learning systems interactions	Forums	Forum	8	Forum deleted	
Learning systems interactions	Forums	Forum	9	Forum subscribed	
Learning systems interactions	Forums	Forum	10	Forum unsubscribed	
Learning systems interactions	Forums	Forum item	11	Forum item created	
Learning systems interactions	Forums	Forum item	12	Forum item posted	
Learning systems interactions	Forums	Forum item	13	Forum item unposted	Made unavailable
Learning systems interactions	Forums	Forum item	14	Forum item edited	
Learning systems interactions	Forums	Forum item	15	Forum item deleted	
Learning systems interactions	Forums	Forum item	16	Forum item viewed	
Learning systems interactions	Forums	Forum item	17	Forum item marked	Like, Angry, Read, Unread etc
Learning systems interactions	Document	Document	18	Document created	Created or uploaded
Learning systems interactions	Document	Document	19	Document posted	Made available
Learning systems interactions	Document	Document	20	Document edited	Re-uploaded or revised in place
Learning systems interactions	Document	Document	21	Document deleted	
Learning systems interactions	Document	Document	22	Document viewed	Document viewed or opened
Learning systems interactions	Assignments	Assignments	23	Assignment created	By instructor, created but not yet made available to students
Learning systems interactions	Assignments	Assignments	24	Assignment posted	By instructor, made available to students for access
Learning systems interactions	Assignments	Assignments	25	Assignment unposted	Made unavailable
Learning systems interactions	Assignments	Assignments	26	Assignment deactivated	By instructor, removed from access
Learning systems interactions	Assignments	Assignments	27	Assignment edited	By instructor
Learning systems interactions	Assignments	Assignments	28	Assignment deleted	By instructor
Learning systems interactions	Assignments	Assignments	29	Assignment viewed	By student
Learning systems interactions	Assignments	Assignments	30	Assignment reviewed	By instructor
Learning systems interactions	Assignments	Assignments	31	Assignment started	By student
Learning systems interactions	Assignments	Assignments	32	Assignment submitted	By student
Learning systems interactions	Assignments	Assignments	33	Assignment completed	By student
Learning systems interactions	Assignments	Assignments	34	Assignment grade created	By instructor, created, but not yet visible
Learning systems interactions	Assignments	Assignments	35	Assignment grade posted	By instructor, posted means final. There can be multiple!
Learning systems interactions	Assignments	Assignments	36	Assignment grade unposted	Made unavailable
Learning systems interactions	Assignments	Assignments	37	Assignment grade edited	By instructor, revised grade
Learning systems interactions	Assignments	Assignments	38	Assignment grade deleted	By instructor
Learning systems interactions	Assignments	Assignments	39	Assignment grade viewed	By student
Learning systems interactions	Assignments	Assignments	39	Assignment feedback created	By student or instructor
Learning systems interactions	Assignments	Assignments	40	Assignment feedback viewed	By student within the tool, not in a downloaded documenr
Learning systems interactions	Assignments	Assignments	41	Assignment feedback downloaded	e.e.g, student downloadss and assignment feedback doc
Learning systems interactions	Groups	Groups	42	Group assignment created	e.g., Instructor assigning students to a group
Learning systems interactions	Groups	Groups	43	Group assignment posted	Made available to students
Learning systems interactions	Groups	Groups	44	Group assignment unposted	Made unavailable
Learning systems interactions	Groups	Groups	45	Group assignment viewed	By the student

- Four level hierarchy
- At the level of granularity or lower than Caliper, xAPI
- Can map to Caliper, xAPI or future standards
- Can extend and define our learning events as needed without waiting for standards
- Can map post-hoc to standards as they evolve
- Extendible domains
  - Learning systems interactions
  - Advising interactions
  - Co-curricular interactions
  - Academic interactions
  - Advising interactions
- We are also maintaining a “Tool Hierarchy” to categorize EdTech ecosystem tools and provide a simple containership model

UC San Diego’s Open edX real-time events

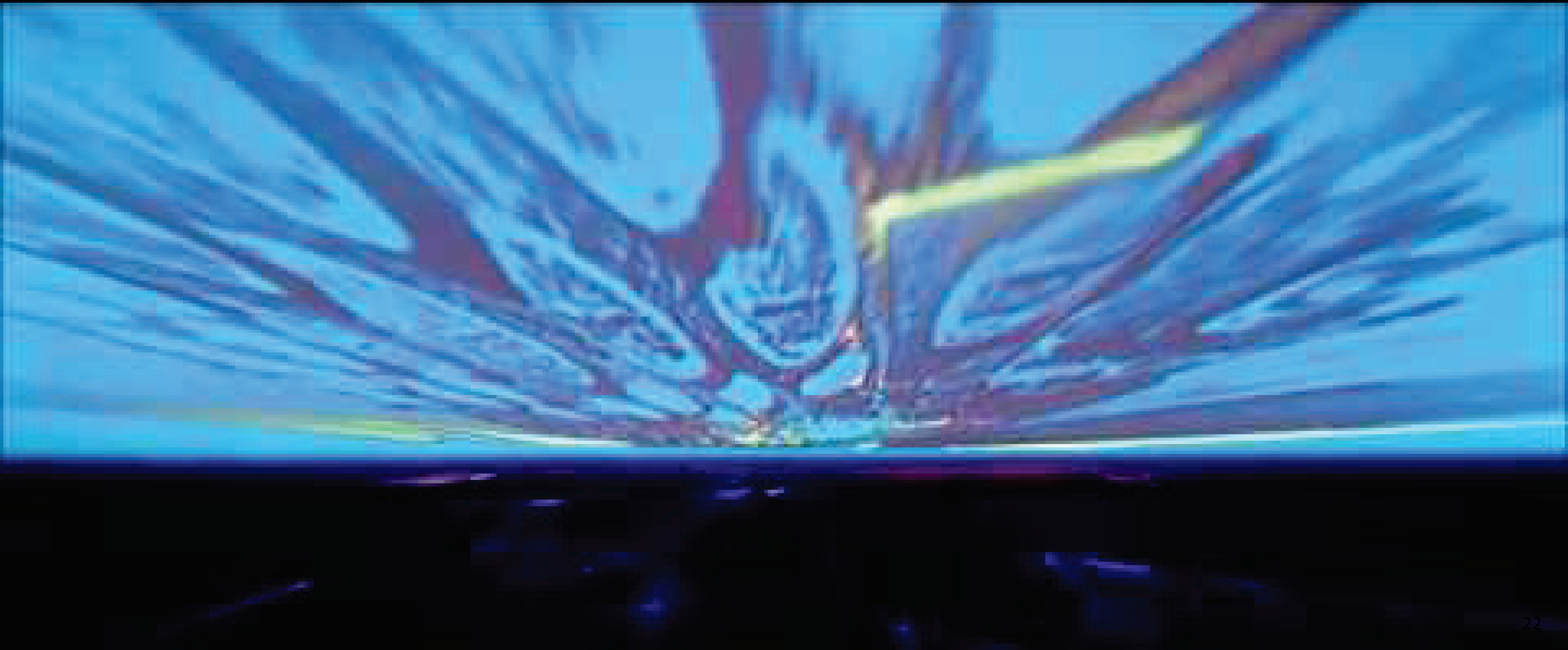
- ✓ Allows real-time collection of course/learner activity to flow into analytics tools, which can then be used to more effectively design classes and boost student success
- ✓ Maps to the current Caliper events where there is a clean match. Event carries additional attributes to identify the open EdX event where the event type does not match Caliper
- ✓ 154 events types in total
- ✓ Can be used with open source integration tools (Apache Kafka, NiFi) to tail the log file for real-time ingestions
- ✓ At UC San Diego, we ingest directly into the SAH learning events table (with data tiering!) Curated views transform into consumable views
- ✓ Learning event data will be consumed by Group Builder / Message Builder for integrated “nudging” and personalized messaging
- ✓ Full event log can be analyzed in to the HANA directly or through other tools

More info here. Source code here. For details email Amin Qazi, amqazi@ucsd.edu

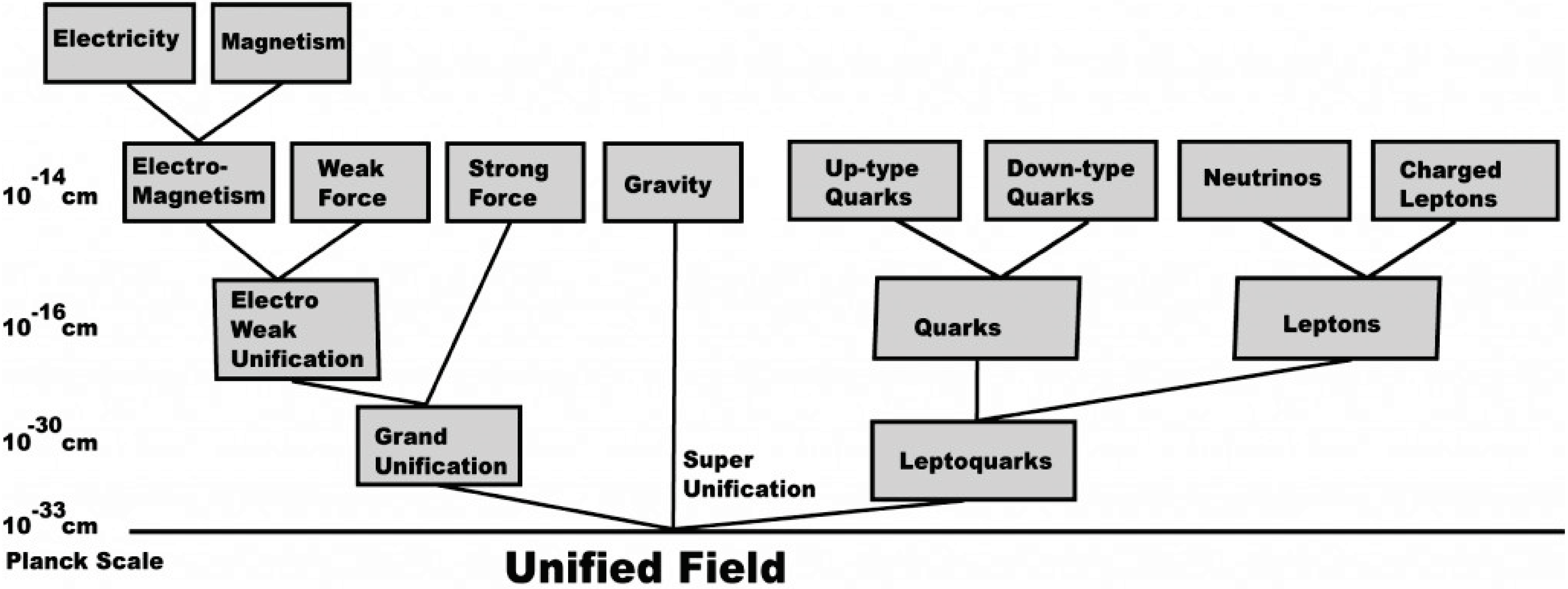


JUPITER

AND BEYOND THE INFINITE



Unified Field Theory



Electricity

Magnetism

10^{-14} cm

Electro-Magnetism

Weak Force

Strong Force

Gravity

Up-type Quarks

Down-type Quarks

Neutrinos

Charged Leptons

10^{-16}cm

Electro Weak Unification

Quarks

Leptons

10^{-30} cm

Grand Unification

Super Unification

Leptoquarks

10^{-33}cm

Planck Scale

Unified Field

New rules

1. **Everything is a verb**
   - All data are loaded into a very long, very wide, insert-only activity table. Relevant changes/deletions are new rows. Idempotency
   - Streaming is the new dominant way to move data in/out
2. **Express maximum semantic complexity**
   - All data (attributes, rows) are added ahead of actual use
   - No aggregates. All data is stored in and processed at its lowest level of granularity
3. **Build provisionally**
   - Curated views are designed for specific analysis needs (vignettes), can come and go
   - No “permanent” dimensional modeling. Analytic views contain a simple list of attributes for analysts
4. **Design for the speed of thought**
   - Sub second analyst click response. Real-time data where needed
   - Curated views must make it very easy for analysts to manipulate
   - Push logic (set and Boolean) to the back-end, free the front end for visualization
5. **Waste is good**
   - No need to conserve space. Curated views can be overlapping and duplicative, data can be exploded
   - A hierarchy of reusable SQL code results in an OO-like, environment
6. **Democratize the data**
   - Make it easier to understand, consume and use
   - Enable the community to share, encourage de-centralized, bottom-up data analysis and use

Overview of the next generation data warehouse

Employee Activity Hub

Sources

- PPS
- UCPath
- Hire Online
- JD Online
- UC Learning
- Kuali Protocols
- Employee LMS
- CITI
- Performance Management

Uses

- Positions
- Pay
- Training
- Performance
- Engagement

Project Activity Hub

Sources

- Jira
- Confluence
- SeviceNow

Uses

- IT project analysis
- Portfolio analysis
- Resource allocations

Research Activity

Sources

- Interfolio
- PPS appointment
- Kuali
- PPM
- Role/affiliates system
- Research scholar
- Oracle Finance /

Uses

- Faculty appointment and roles
- Sponsored project financial analysis
- Research compliance analysis
- Research portfolio analysis

Common Tables

People / Identity
Organization
Facilities
Common Activities
Hierarchies

Common Tools

Tableau / Cognos
SPSS / R
API Access
Mobile Messaging

Embedded Platform Tools

Statistical & Predictive
Machine Learning
Graphing Algorithms
Spatial Analysis
Text Mining
R

Student Activity Hub

Sources

- SIS
- LMS
- VAC
- Redrock
- Student Event Management
- DARS
- ProSam
- Slate
- Co-curricular record
- Extension/MOOC

Uses

- Enrollment
- Demographics
- Majors/minors
- Statistics per term, progression
- Retention, graduate rate, time to degree
- Learning analytics
- Student engagement
- Applicants/Applications
- Test scores
- Scholarships
- Non-matriculated progress/success

Facilities Activity

Sources

- Tririga
- CAMS

Uses

- Classroom utilization
- Building utilization
- Walk-time
- IDC Analysis
- Maintenance
- Planning
- Event

Financial Activity

Sources

- ESR Finance
- ESR Budget
- ESR Student

Uses

- Activity pattern analysis
- Ad-hoc analysis
- Multi-fund analysis
- Tuition revenue modeling
- Budgeting and forecasting

Advancement & Alumni Activity

Sources

- BlackBaud
- Alumni iModules

Uses

- Constituent analysis
- Financial analysis
- Campaign analysis

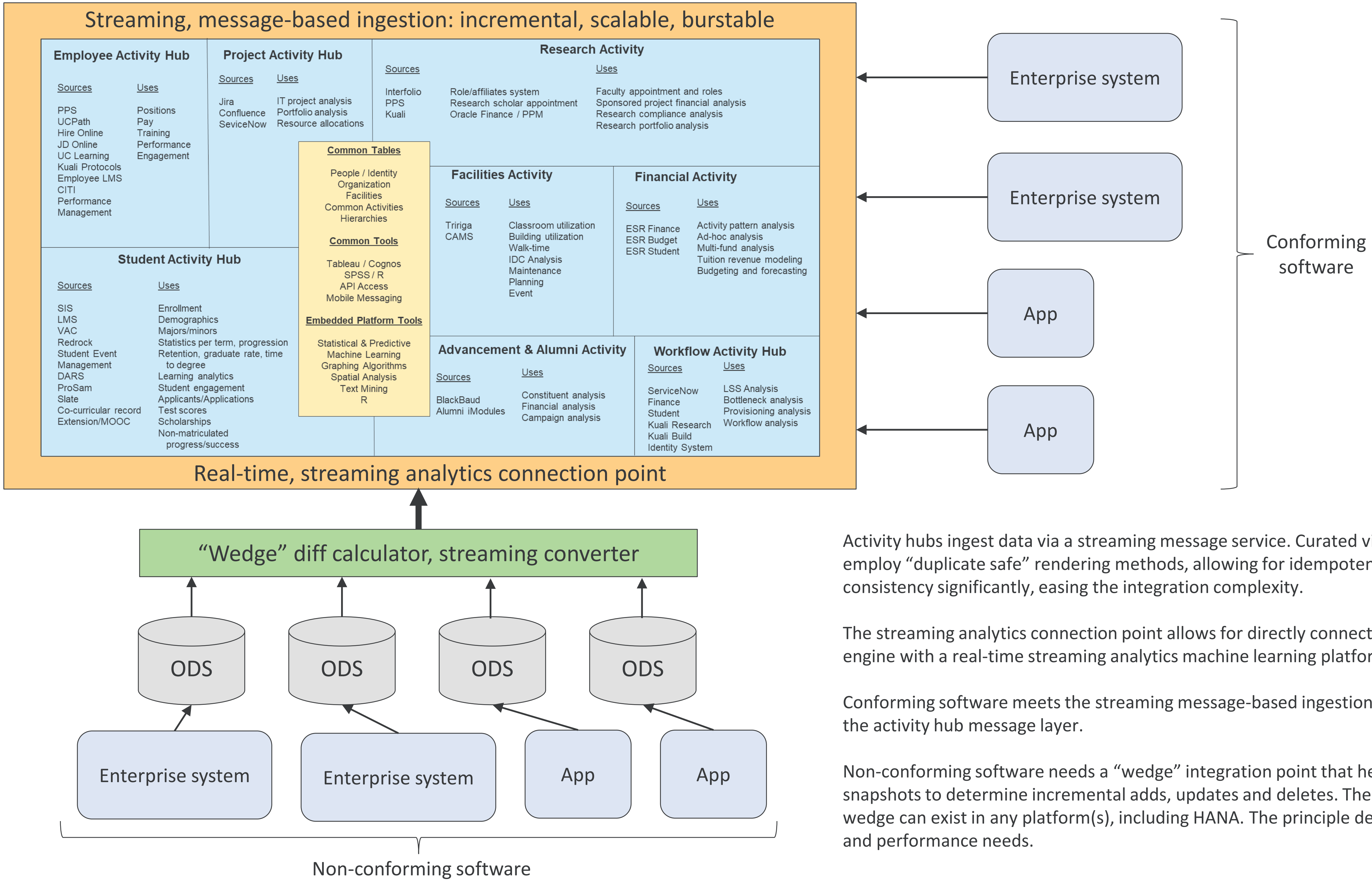
Workflow Activity Hub

Sources

- ServiceNow
- Finance
- Student
- Kuali Research
- Kuali Build
- Identity System

Uses

- LSS Analysis
- Bottleneck analysis
- Provisioning analysis
- Workflow analysis



Streaming, message-based ingestion: incremental, scalable, burstable

Employee Activity Hub

Sources	Uses
PPS UCPath Hire Online JD Online UC Learning Kuali Protocols Employee LMS CITI Performance Management	Positions Pay Training Performance Engagement

Project Activity Hub

Sources	Uses
Jira Confluence SeviceNow	IT project analysis Portfolio analysis Resource allocations

Research Activity

Sources		Uses
Interfolio PPS Kuali	Role/affiliates system Research scholar appointment Oracle Finance / PPM	Faculty appointment and roles Sponsored project financial analysis Research compliance analysis Research portfolio analysis

Common Tables

People / Identity
Organization
Facilities
Common Activities
Hierarchies

Common Tools

Tableau / Cognos
SPSS / R
API Access
Mobile Messaging

Embedded Platform Tools

Statistical & Predictive
Machine Learning
Graphing Algorithms
Spatial Analysis
Text Mining
R

Facilities Activity

Sources	Uses
Tririga CAMS	Classroom utilization Building utilization Walk-time IDC Analysis Maintenance Planning Event

Financial Activity

Sources	Uses
ESR Finance ESR Budget ESR Student	Activity pattern analysis Ad-hoc analysis Multi-fund analysis Tuition revenue modeling Budgeting and forecasting

Student Activity Hub

Sources	Uses
SIS LMS VAC Redrock Student Event Management DARS ProSam Slate Co-curricular record Extension/MOOC	Enrollment Demographics Majors/minors Statistics per term, progression Retention, graduate rate, time to degree Learning analytics Student engagement Applicants/Applications Test scores Scholarships Non-matriculated progress/success

Advancement & Alumni Activity

Sources	Uses
BlackBaud Alumni iModules	Constituent analysis Financial analysis Campaign analysis

Workflow Activity Hub

Sources	Uses
ServiceNow Finance Student Kuali Research Kuali Build Identity System	LSS Analysis Bottleneck analysis Provisioning analysis Workflow analysis

Real-time, streaming analytics connection point

Enterprise system

Enterprise system

App

App

Conforming software

“Wedge” diff calculator, streaming converter

ODS

ODS

ODS

ODS

Enterprise system

Enterprise system

App

App

Non-conforming software

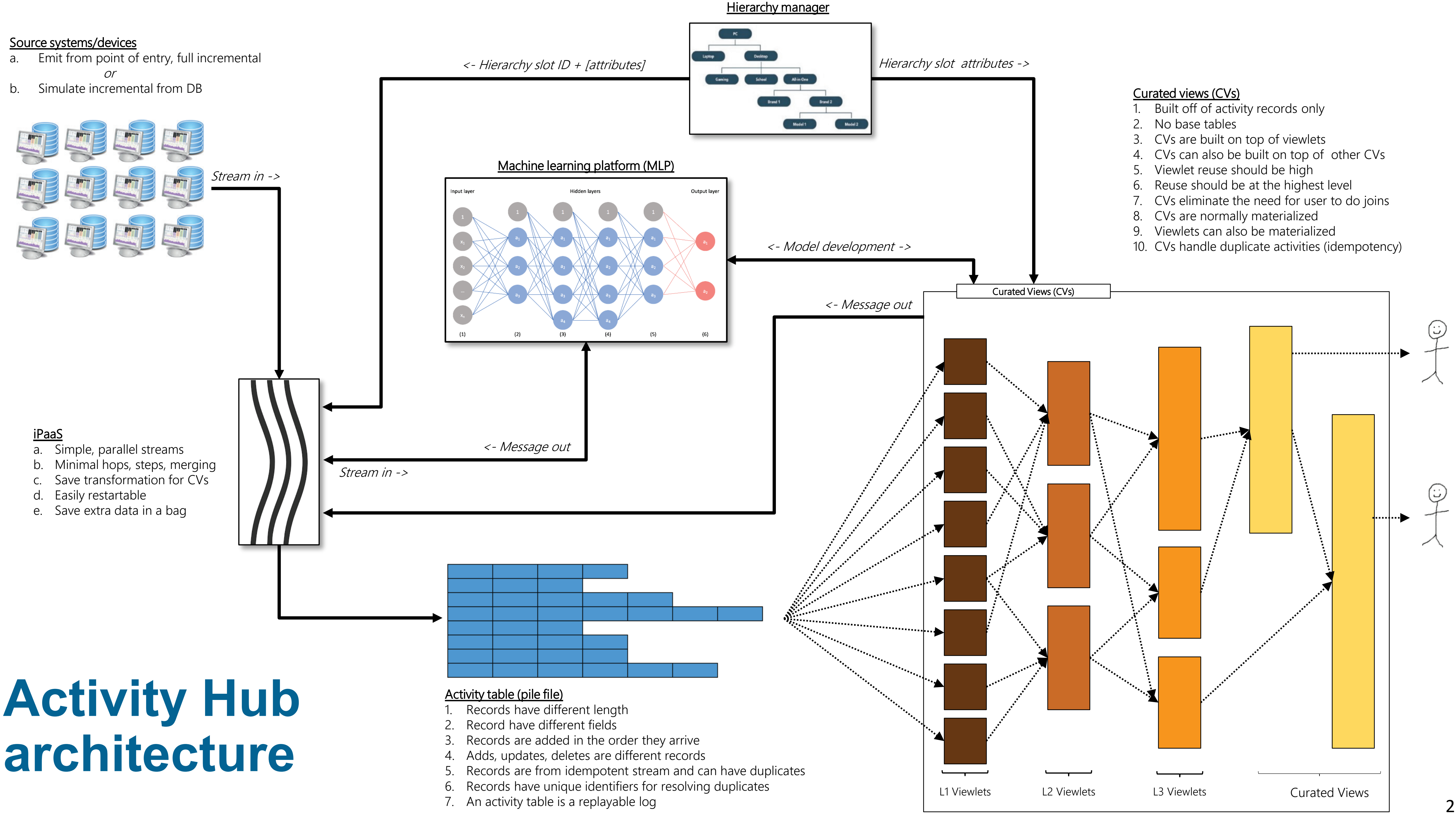
Activity hubs ingest data via a streaming message service. Curated views and activity tables should employ “duplicate safe” rendering methods, allowing for idempotent messages. This relaxes data consistency significantly, easing the integration complexity.

The streaming analytics connection point allows for directly connecting the streaming ingestion engine with a real-time streaming analytics machine learning platform to process inbound messages

Conforming software meets the streaming message-based ingestion method and submit directly to the activity hub message layer.

Non-conforming software needs a “wedge” integration point that helps calculate differences in snapshots to determine incremental adds, updates and deletes. The ODS and other tools for this wedge can exist in any platform(s), including HANA. The principle define choice is long-term cost and performance needs.

Activity Hub architecture



Source systems/devices

a. Emit from point of entry, full incremental

or

b. Simulate incremental from DB

iPaaS

a. Simple, parallel streams
b. Minimal hops, steps, merging
c. Save transformation for CVs
d. Easily restartable
e. Save extra data in a bag

Activity table (pile file)

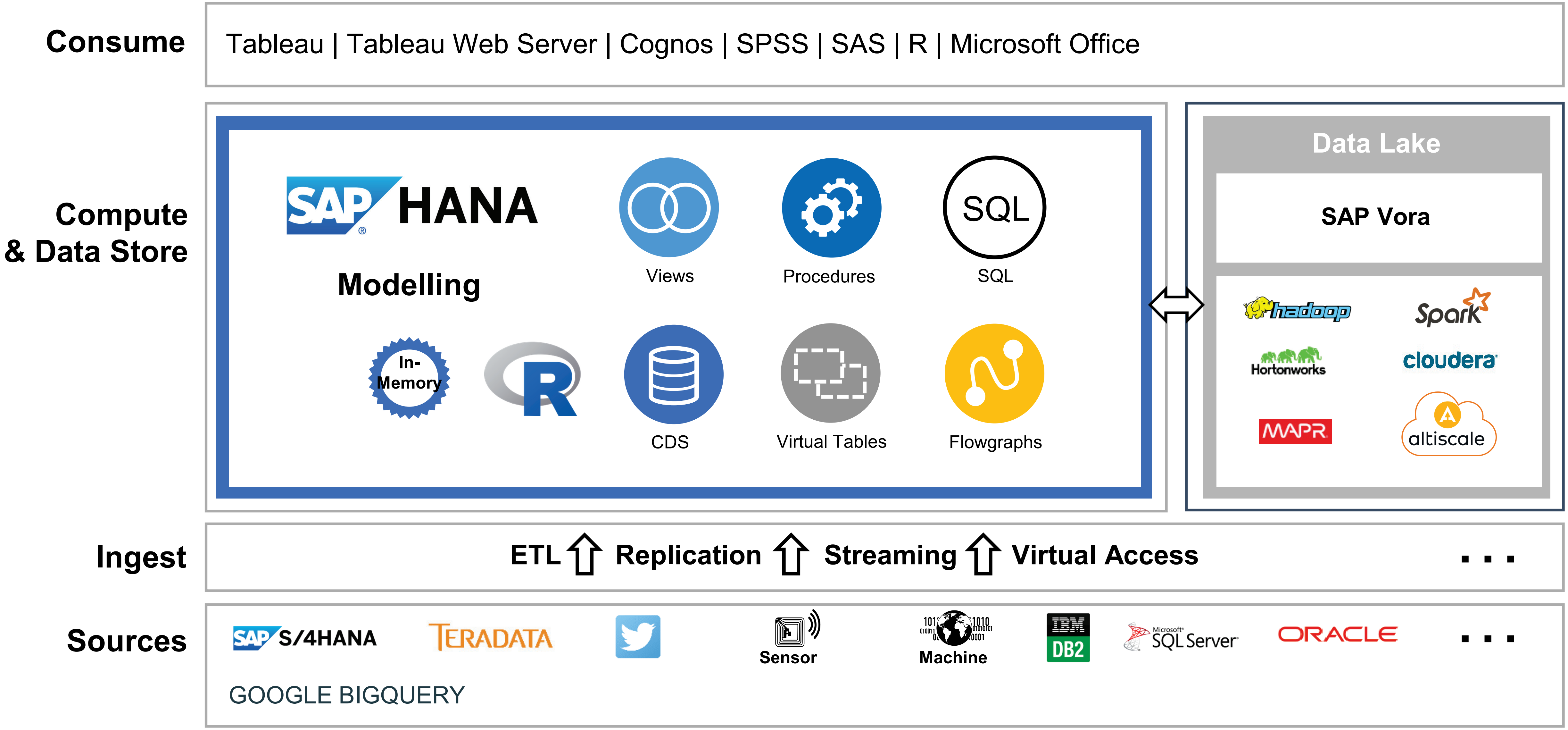
1. Records have different length
2. Record have different fields
3. Records are added in the order they arrive
4. Adds, updates, deletes are different records
5. Records are from idempotent stream and can have duplicates
6. Records have unique identifiers for resolving duplicates
7. An activity table is a replayable log

Curated views (CVs)

1. Built off of activity records only
2. No base tables
3. CVs are built on top of viewlets
4. CVs can also be built on top of other CVs
5. Viewlet reuse should be high
6. Reuse should be at the highest level
7. CVs eliminate the need for user to do joins
8. CVs are normally materialized
9. Viewlets can also be materialized
10. CVs handle duplicate activities (idempotency)

SAP HANA

Data Ingestion and Analytics modelling overview



Consume

Tableau | Tableau Web Server | Cognos | SPSS | SAS | R | Microsoft Office

Compute & Data Store

SAP HANA

Modelling

In-Memory

R

Views

Procedures

SQL

CDS

Virtual Tables

Flowgraphs

Data Lake

SAP Vora

hadoop

Spark

Hortonworks

cloudera

MAPR

altiscale

Ingest

ETL ⇧ **Replication** ⇧ **Streaming** ⇧ **Virtual Access** . . .

Sources

SAP S/4HANA

TERADATA

Sensor

Machine

IBM DB2

Microsoft SQL Server

ORACLE

. . .

GOOGLE BIGQUERY

Platform predictive capabilities

Classification Analysis

- CART
- C4.5 Decision Tree Analysis
- CHAID Decision Tree Analysis
- K Nearest Neighbour
- Logistic Regression Elastic Net
- Back-Propagation (Neural Network)
- Naïve Bayes
- Support Vector Machine
- Random Forests
- Gradient Boosting Decision Tree
- Linear Discriminant Analysis (LDA)
- Confusion Matrix
- Area Under Curve (AUC)
- Parameter Selection/Model Evaluation

Regression

- Multiple Linear Regression Elastic Net
- Polynomial, Exponential, Bi-Variate Geometric, Bi-Variate Logarithmic Regression
- Generalized Linear Model
- Cox Proportional Hazards Model

Cluster Analysis

- ABC Classification
- DBSCAN
- K-Means/Accelerated K-Means
- K-Medoid Clustering
- K-Medians
- Kohonen Self-Organized Maps
- Agglomerate Hierarchical
- Affinity Propagation
- Latent Dirichlet Allocation (LDA)
- Gaussian Mixture Model (GMM)
- Cluster Assignment

Time Series Analysis

- Single/Double/Brown/Triple Exponential Smoothing
- Forecast Smoothing
- Auto – ARIMA/ Seasonal ARIMA
- Croston Method
- Forecast Accuracy Measure
- Linear Regression with Damped Trend and Seasonal Adjustment
- Test for White Noise, Trend, Seasonality
- Fast Fourier Transform (FFT)
- Correlation Function

Association Analysis

- Apriori
- Apriori Lite
- FP-Growth
- KORD – Top K Rule Discovery
- Sequential Pattern Mining

Probability Distribution

- Distribution Fit/Weibull analysis
- Cumulative Distribution Function
- Quantile Function
- Kaplan-Meier Survival Analysis

Outlier Detection

- Inter-Quartile Range Test (Tukey's)
- Variance Test
- Anomaly Detection
- Grubbs Outlier Test

Recommender

- Factorized Polynomial Regression Models

Link Prediction

- Common Neighbors
- Jaccard's Coefficient
- Adamic/Adar
- Katzβ

Statistical Functions

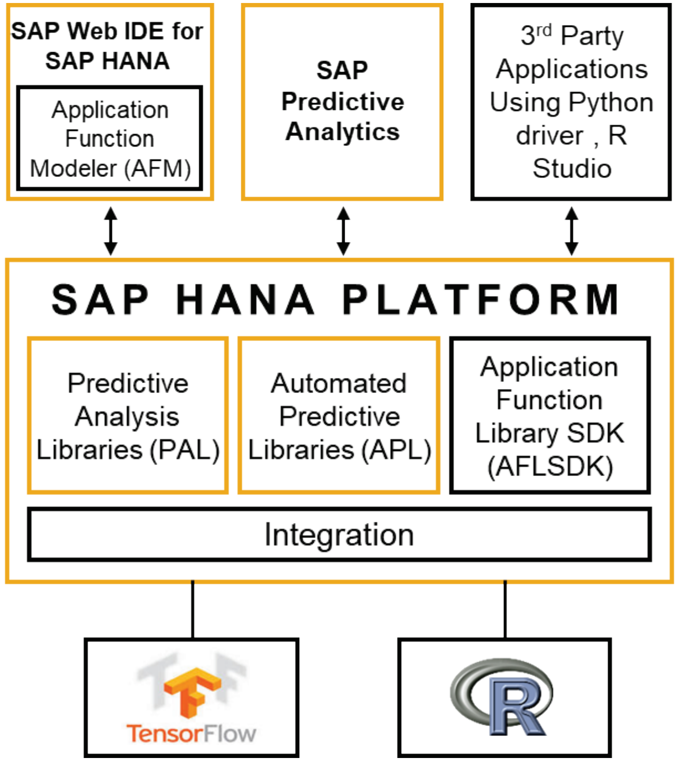
- Mean, Median, Variance, Standard Deviation, Kurtosis, Skewness
- Covariance Matrix
- Pearson Correlations Matrix
- Chi-squared Tests:
  - Test of Quality of Fit
  - Test of Independence
- F-test (variance equal test)
- Data Summary
- ANOVA
- One-sample Median Test
- T Test
- Wilcox Signed Rank Test

Data Preparation

- Sampling
- Binning
- Scaling
- Partitioning
- Principal Component Analysis (PCA)/ PCA Projection

Other

- Weighted Scores Table
- Substitute Missing Values



SAP Web IDE for SAP HANA

Application Function Modeler (AFM)

SAP Predictive Analytics

3rd Party Applications Using Python driver , R Studio

SAP HANA PLATFORM

Predictive Analysis Libraries (PAL)

Automated Predictive Libraries (APL)

Application Function Library SDK (AFLSDK)

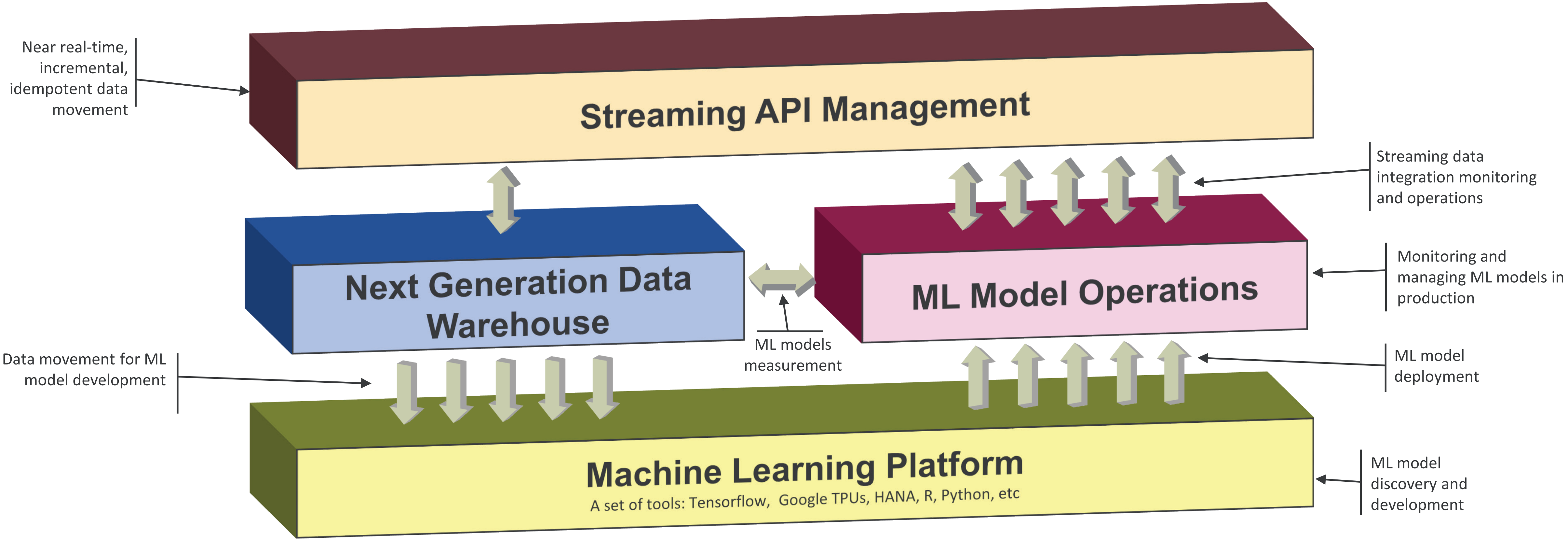
Integration

- 90+ prepackaged machine learning/predictive algorithms
- Supports association, clustering, classification, regression, time series, ...
- Supports different types of data – structured, streaming and series data
- Real-time scoring for several algorithms
- Integrated with open source machine learning libraries – TensorFlow and R

Managing multiple ML models in the next generation analytics

How can we use machine learning to improve administrative processes, student success, research outcomes?

- Multiple models may be active per each business opportunity (e.g., student learning feedback, student success intervention, financial activity fraud detection)
- Multiple models will be developed and trained based on prior streams of data
- Multiple models will be deployed to actively interact with real-time streams of data, interacting with requesting systems and users, activating workflows
- Multiple models can be managed within a ‘single pane of glass.’ Operations can ensure reliability, detect anomalies, bring up and take down models
- Model measurement data feeds back into the next generation data warehouse to guide further model development
- Faculty experts can utilize this infrastructure to help provide needed expertise rather than use consultants
- The data within this environment can serve workbench for data science and research activities
- The next generation data warehouse (SAP HANA) has best-in-class de-identification capabilities transparent to the end-user, enabling safe use for researchers


Near real-time, incremental, idempotent data movement
Streaming API Management
Streaming data integration monitoring and operations
Next Generation Data Warehouse
ML Model Operations
Monitoring and managing ML models in production
ML models measurement
Data movement for ML model development
ML model deployment
Machine Learning Platform
A set of tools: Tensorflow, Google TPUs, HANA, R, Python, etc
ML model discovery and development

Questions?