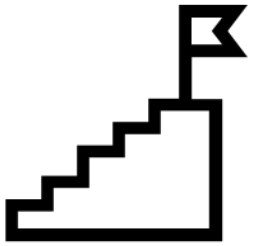
Student Activity Hub (SAH)

IT Community Overview

7/2/2021

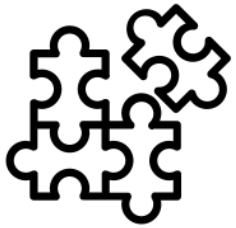
Authors: vkellen, kchou, aqazi, jmwhite, abeecham, others…

SAH: A mission-driven, multi-institution collaboration



MISSION

Advance the state of student data management and student analytics in order to achieve our institutional goals, as diverse as they may be, while protecting institutional autonomy and control over all data.



PROBLEM

The SAH tackles the student data management data and analysis problems directly, giving control back to the institution. Think of SAH as a rich and high performance ‘transmission.’ You can drive it anywhere you like.

SOLUTION

SAH allows for the merging of all kinds of data in one solution. Each institution has its own high-speed, in-memory server environment. With its security, scalability and sophistication, we can integrate any and all student data.

OUTCOME

The goal is modest. We want help institutions who might to leverage a common, but easily tailored or customized solution. Our goal is not to “sell” large numbers of SAH. We just want to make a difference where we can and collaborate with peers.

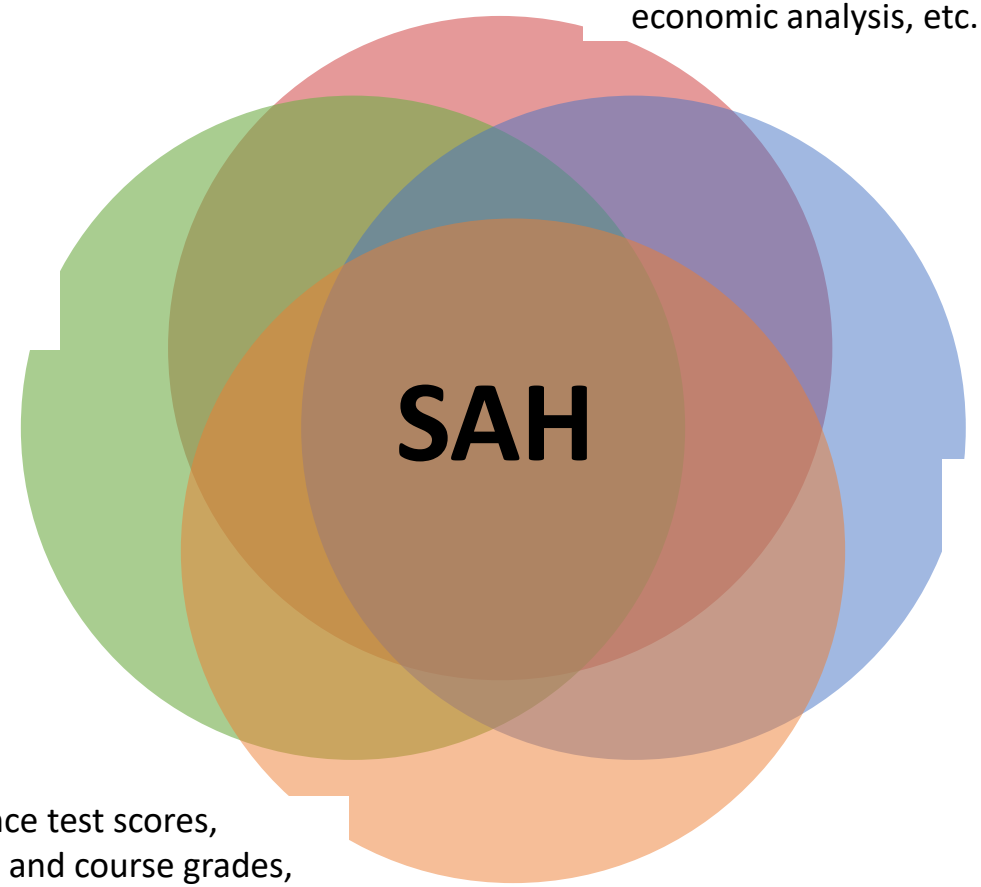
How are we different?

- We are extremely transparent. With prices, our technology approach, with everything
- We are using a wickedly-fast and very powerful analytic platform that is typically only found in larger corporate environments
- We can capture all forms of student data and have exquisite designs for retention, learning and engagement analytics
- We use a very rigorous, disciplined software engineering approach with our core engineers that allows institutions to safely customize as they see fit
- We do not need to make a profit. We want to keep our costs extremely low. We don’t use traditional sales methods
- We are very partner-friendly. We know we can’t do this alone and we prefer not to
- We embrace the IMS Global standards for student data, including Caliper and Edu-API

The student activity hub (SAH) can support various needs

Institutional analytics:
Graduation rates, retention rates, enrollments, demographic, lists of majors/minors, socio-economic analysis, etc.

Engagement analytics: advising interactions, co-curricular activities, degree progress tool use, mobile app interactions, etc.



Learning analytics:
Course engagement, submissions, within-course grades, assignments, discussions, clickstream, page views, video views

Academic analytics: entrance test scores, satisfactory progress, term and course grades, commencement of academic activity, bottleneck course, degree switching etc.

SAH was designed to give institutions full control

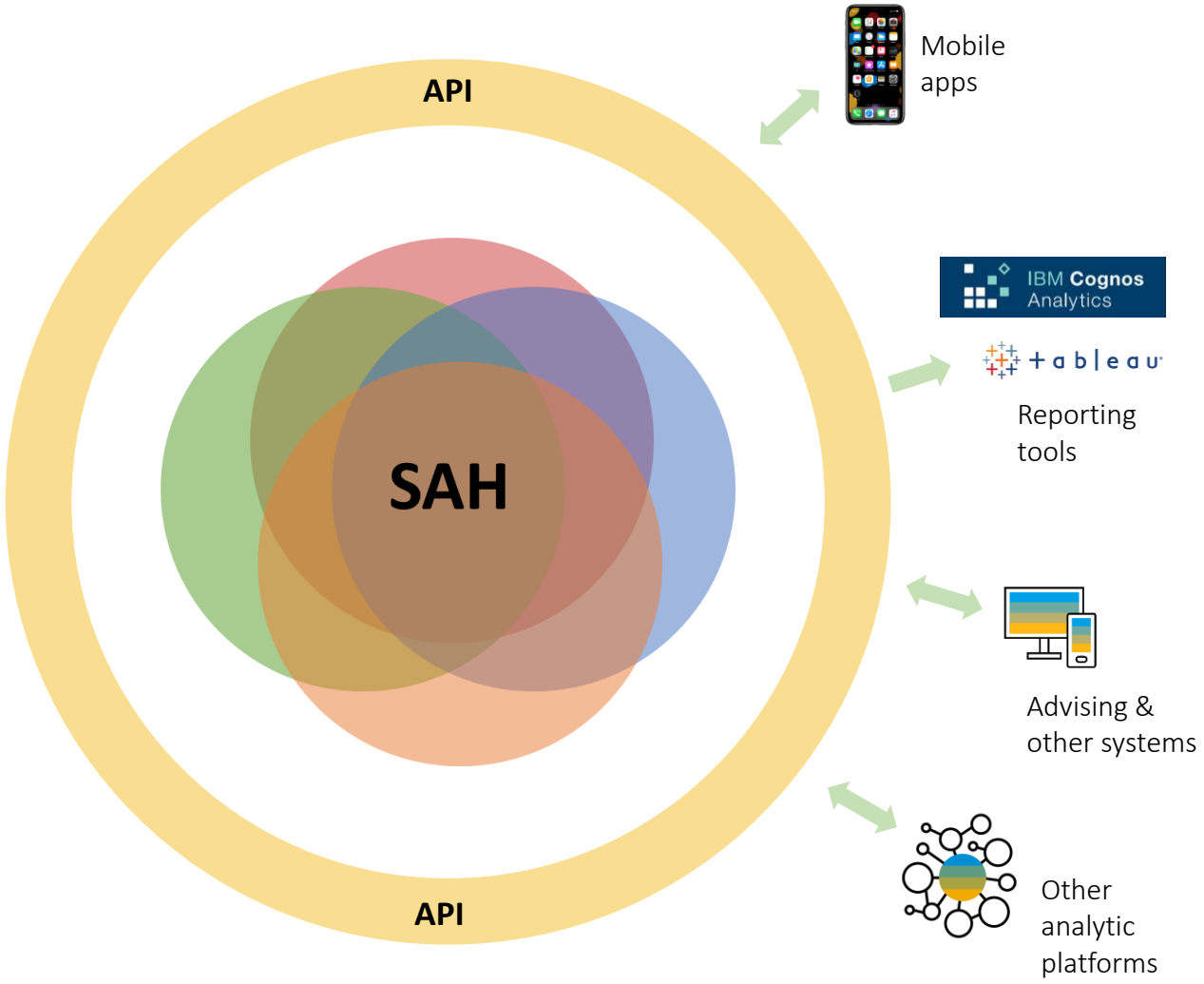
SaaS/IaaS: You can establish the level of control you need. We can operate in a full SaaS or in a full IaaS mode and adjust fees as needed. Items of control include:

- Data integration platform: We use Apache Kafka, Apache NiFi, Go Anywhere and WSO2 API manager. Institutions are free to choose their own integration tools and operate them or let us do it for them
- Custom view construction: The core SAH views are easily 'forkable' enabling institutions to develop their own solutions. We can perform the customization work or the institution can. Either way! All views are 100% ANSI SQL (2016)
- New activity tables: Institutions are free to add their own activity tables (a type of data lake), provided they do not alter the delivered activity tables. Views can freely access data from delivered activity tables or institution customized activity tables
- Metadata management and daily operations: As views get created and modified, we have a metadata administrative console (AH-MAC) tool that enables 'materializations', controls API access for downstream applications, and creation of data groups (Group Builder). These two tools are available to institutions that want full control over their environment. These two tools are written in Python. Institutions can 'fork' their own tools, but will need to manage the change process for new server console tools themselves. Institutions can administer their environment or let us do it for them
- Report building: At the moment, SAH does require each institution to have a reporting strategy. We have a large collection of workbooks in Tableau and Cognos we make available
- Change management: Since each institution has its own data and server instance, each institution can establish its own change processes and also choose when to accept changed or new core view designs

SAH's is completely open for any institutional need

With SAH's open-source API framework, any other institutional application or platform can receive data from or send data to SAH in real-time. You are in control of what data you want to integrate

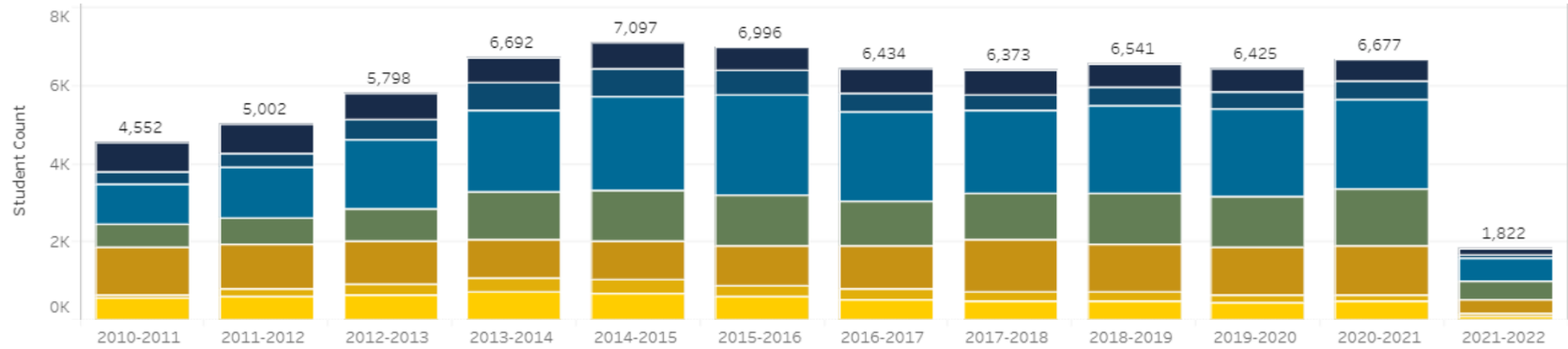
- Institutional innovation, central or distributed
- Advising, case management and student support tools
- Third-party software
- Other analytic platforms, tools or services



FOR INTERNAL USE ONLY. If you receive requests outside of your department, please forward the request to Marcela Alvarez (msalvarez@eng.ucsd.edu).

Jacobs School 3rd Week Enrollment Trends

Final Registration Census Report: UN*



Dept	Major	2010-2011	2011-2012	2012-2013	2013-2014	2014-2015	2015-2016	2016-2017	2017-2018	2018-2019	2019-2020	2020-2021
Grand Total		4,552	5,002	5,798	6,692	7,097	6,996	6,434	6,373	6,541	6,425	6,677
BENG	Total	754	741	680	632	678	626	638	621	602	583	581
	BE02 - Pre-Bioengineering:Pre-Medical	72	8	2								
	BE25 - Bioengineering	222	244	222	186	204	191	174	169	179	187	178
	BE26 - Bioengineering: Pre-Medical	135	96	51	12	2						
	BE27 - Bioengineering (Biotechnology)	285	308	296	237	236	204	210	207	221	206	219
	BE28 - Bioengineering: Bioinformatics	40	85	109	148	158	140	135	122	103	92	104
	BE29 - Bioengineering: BioSystems				49	78	91	119	123	99	98	80
CENG	Total	312	374	513	686	708	627	471	377	439	439	450
	CE25 - Chemical Engineering	312	374	513	686	708	627	471	377	439	439	450
CSE	Total	1,033	1,277	1,777	2,102	2,390	2,542	2,271	2,147	2,270	2,254	2,279
	CS02 - Pre-Computer Science	1										
	CS25 - Computer Engineering	231	243	350	373	361	322	236	174	157	146	132
	CS26 - Computer Science	683	881	1,221	1,430	1,682	1,873	1,667	1,501	1,686	1,767	1,810
	CS26 - Minor:Computer Science	57	61	94	133	184	240	297	419	394	305	301
	CS27 - Comp Sci w/Spec Bioinformatics	28	52	69	114	111	69	57	41	30	30	35
	CS28 - Computer Science	35	41	44	53	54	43	16	13	4	9	4

Program Level ID: UN

Quarter: Fall Qtr

Dept: (All)

Degree Type: (All)

Ethnicity: (All)

Gender: (All)

Applicant Type: (All)

Pell Recipient Flag: (All)

First Generation Flag: (All)

Military Status: (All)

Citizenship: (All)

CA Resident Flag: (All)

Registration Status: (All)

BENG
CENG
CSE
ECE
MAE
NENG

* Data shows 3rd week (census) enrollment counts for all completed quarters and live enrollment for the current quarter. Current quarter enrollment is final after the 3rd week of instruction. Data includes primary and secondary majors.
Source: SAH-MajorMinorNarrowCensus-View - 7/28/2021 9:04:21 AM

The Six New Rules

1. Everything is a verb, including nouns
2. Express maximum semantic complexity
3. Build provisionally
4. Design for the speed of thought
5. Waste is good
6. Democratize the data

Credit: Gza Blint Ujvrosi / EyeEm / Getty Images © 2019

HANA-> Event streams, re-playable log, all attributes, lowest level of granularity, reusable-overlapping views, sub-second clicks, real-time data, explode data, feral denormalization, equal access for all, ease of use, ethical and fair use

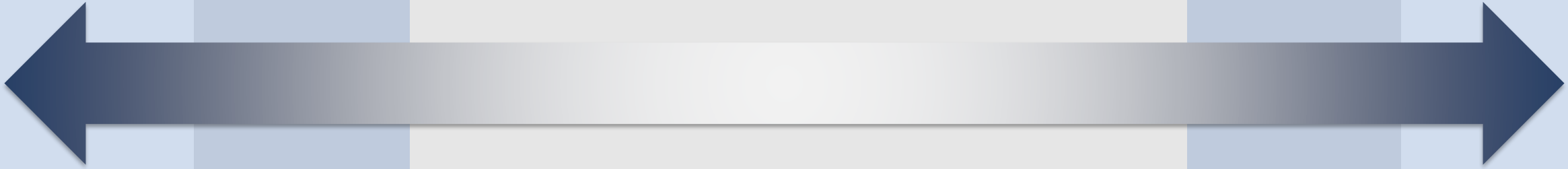
Kellen, V. (2019). 6 New Rules for 21st Century Analytics. Business Technology & Digital Transformation Strategies, Data Analytics & Digital Technologies. Cutter Consortium. https://www.cutter.com/article/6-new-rules-managing-21st-century-analytics-502286
Kellen, V. (2019). 21st-Century Analytics: New Technologies and New Rules. Educause Review. https://er.educause.edu/articles/2019/5/21st-century-analytics-new-technologies-and-new-rules

Under-the-hood architecture points

- Very narrow core software engineering technical skill set (by design)
  - SQL views on top of views. 100% ANSI SQL (2016), as SAP HANA complies with it.
  - Use of stored procedures (freezing data and snapshotting, enabling controlled API access to all data), and a few functions, all ANSI SQL
  - Three server-side console apps in Python: AH Metadata management, Group Builder and Message Builder (coming soon)
- Other aspects
  - High performance SAP architecture enables simpler designs
  - We have strict design standards and guidelines for safe and ‘forkable’ view development. We use these techniques ourselves to be flexible
  - All views are evaluated against the design standards and guidelines before inclusion in the core
  - Collaborating institutions can create new views, using UC San Diego views (modular, building block approach) on their own!

SAH: both a data lake and a data warehouse

Scalable, fast, and 'tiered' technologies allow us to cover a wider range of possibilities. Large amounts of inert data can be pushed, automatically to lower cost storage. We can accommodate a reasonable range of ingestion methods, but with a strong inclination towards streaming approaches. We prefer predefined schemas but also typically stored JSON strings for future 'unpacking'



Data Lake

- Structured and unstructured data
- Flat vs. hierarchical
- Data stored for later processing
- Heterogenous ingestion
- Fast onboarding of new data
- Low-cost storage
- No predefined schemas
- Low level of detail
- Mostly data scientist accessing

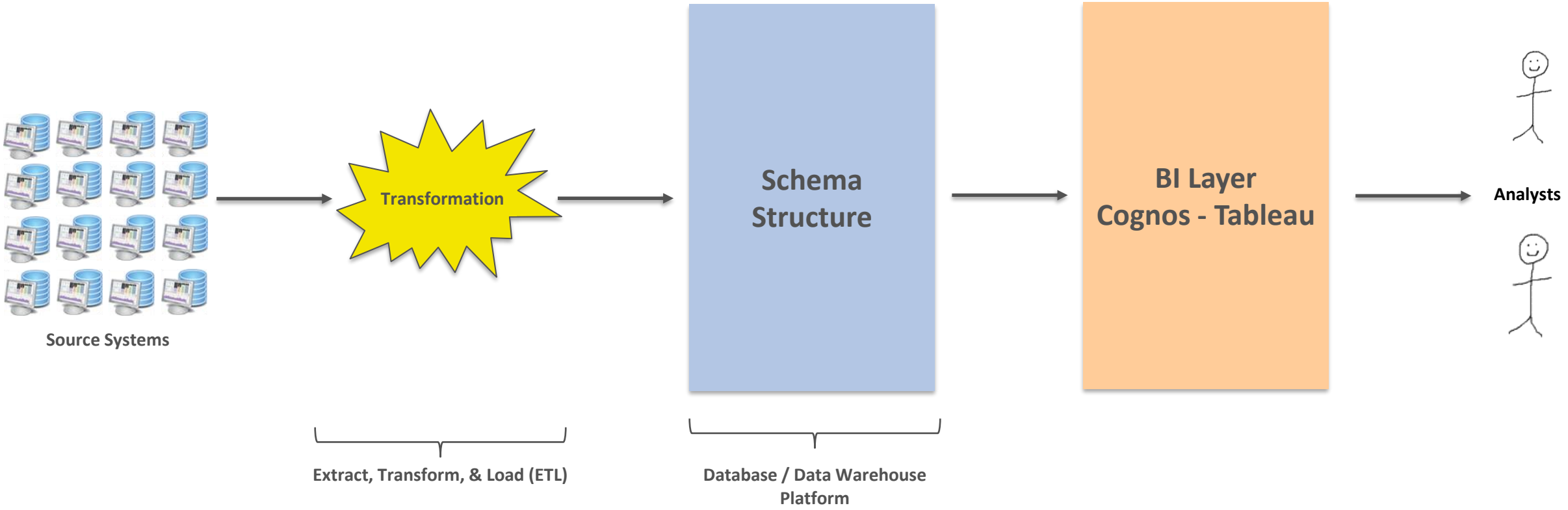
New Rules

- Structured and Highly structured data
- Flat or very hierarchical
- Can be lightly or highly processed
- Conformed real-time and batch ingestion
- Reasonably fast onboarding of new data
- Mix of both high and low cost storage
- More easily defined schemas
- Aggregates for logical convenience only
- Business users and data scientists accessing

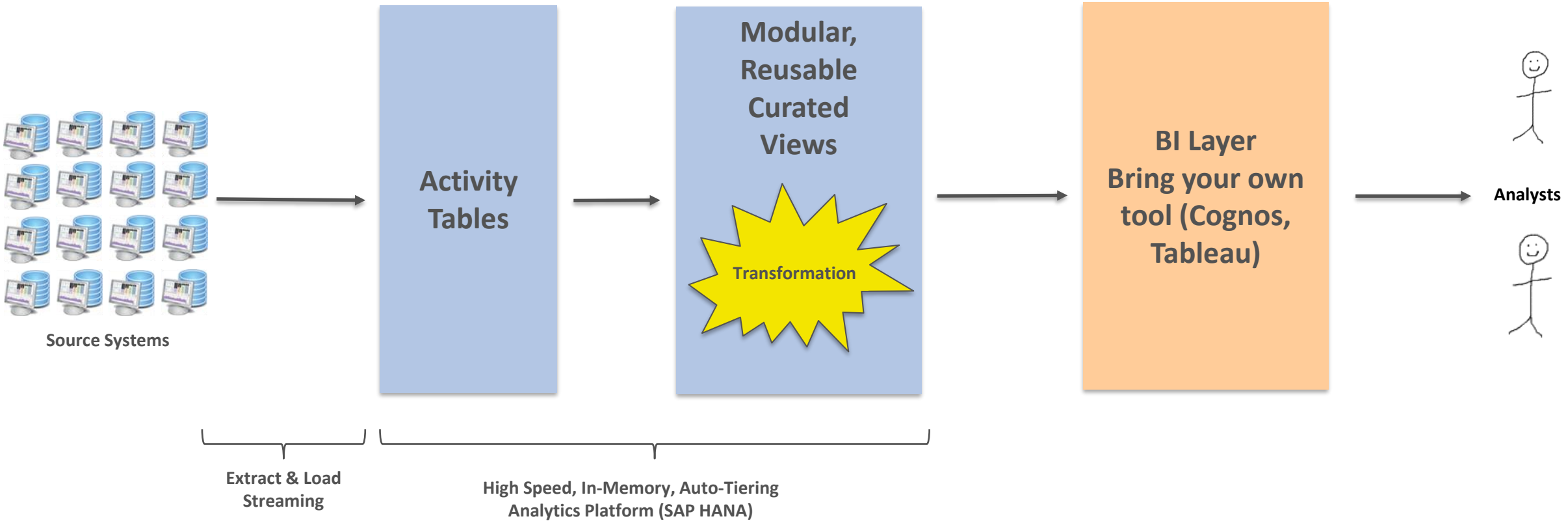
Data Warehouse

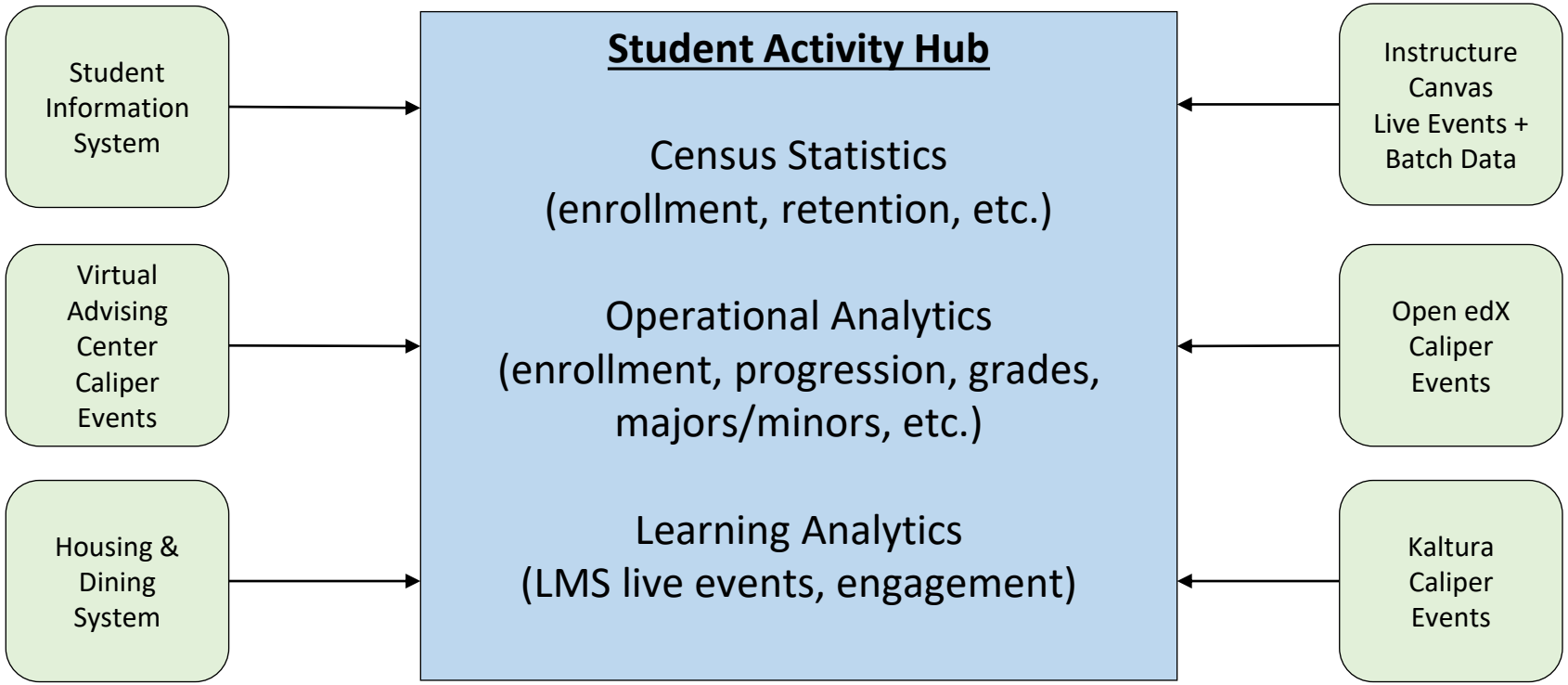
- Highly structured data
- Can be very hierarchical
- Highly processed data
- Conformed ETL for ingestion
- Slower onboarding of new data
- High cost storage/memory
- Predefined schemas
- Reliance on aggregates
- Business users accessing

Moving transformation – typical approach



Moving transformation – SAH approach





Critical features:

- One integrated data model uniting learning analytics & institutional & operational data real-time
- Modular, lego-style reusability of view components
- High-speed, in-memory analytics (SAP HANA), high-availability, auto tiering to warm and cool storage, all AWS (or GCP or Azure)
- ~500 million rows per year of Caliper events + ~170 million rows 'base data', (~100 million rows of Canvas base data)
- Congruent ontology governing classification of events, academic hierarchy, programs, majors, minors, etc.
- Learning analytics can access student retention/progression stats, and vice versa etc.
- Tableau and Cognos secure web access for visualizations
- k-anonymity, l-anonymity and differential privacy tools available

Activity tables and view types

Activity Table

All data streamed in via Apache NiFi / Kafka

Data replication services via SAP SDI

Activity hubs can have many activity tables

Activity table match common ingestion patterns

Three types of activity tables:

1. IoT style (e.g., Caliper event steams)
2. Table replication (e.g., Canvas Batch)
3. Table incremental replication (e.g., UC Path)

Base Views

Marks and/or remove duplicates

If the activity table is incremental replication, removes deletes

Manages type conversions as needed

If needed, renames columns that reflect source system

Creates reusable column segments used more widely within IVs or CVs

Contains view-localized column segments that are not shared

Can reference other BVs

Intermediate Views

Combines data from other BVs or IVs

Are typically either wide (repeated columns) or narrow (repeated rows instead of repeated columns)

Adds in more extended calculations, aggregations, complex where clauses, complex joins (e.g., business logic, or logic to enhance materialization, snapshot performance)

Can perform type conversions as needed

Can rename columns to user-friendly and highly conformed names

Curated Views

Combines data from other BVs and IVs

Normally do not reference each other, but can if needed

Transforms column names into user-friendly names, replaces underscores in column names with spaces

Can filter data through WHERE or JOIN clauses

Can integrate GB_GROUP_CONTENTS keys for needed

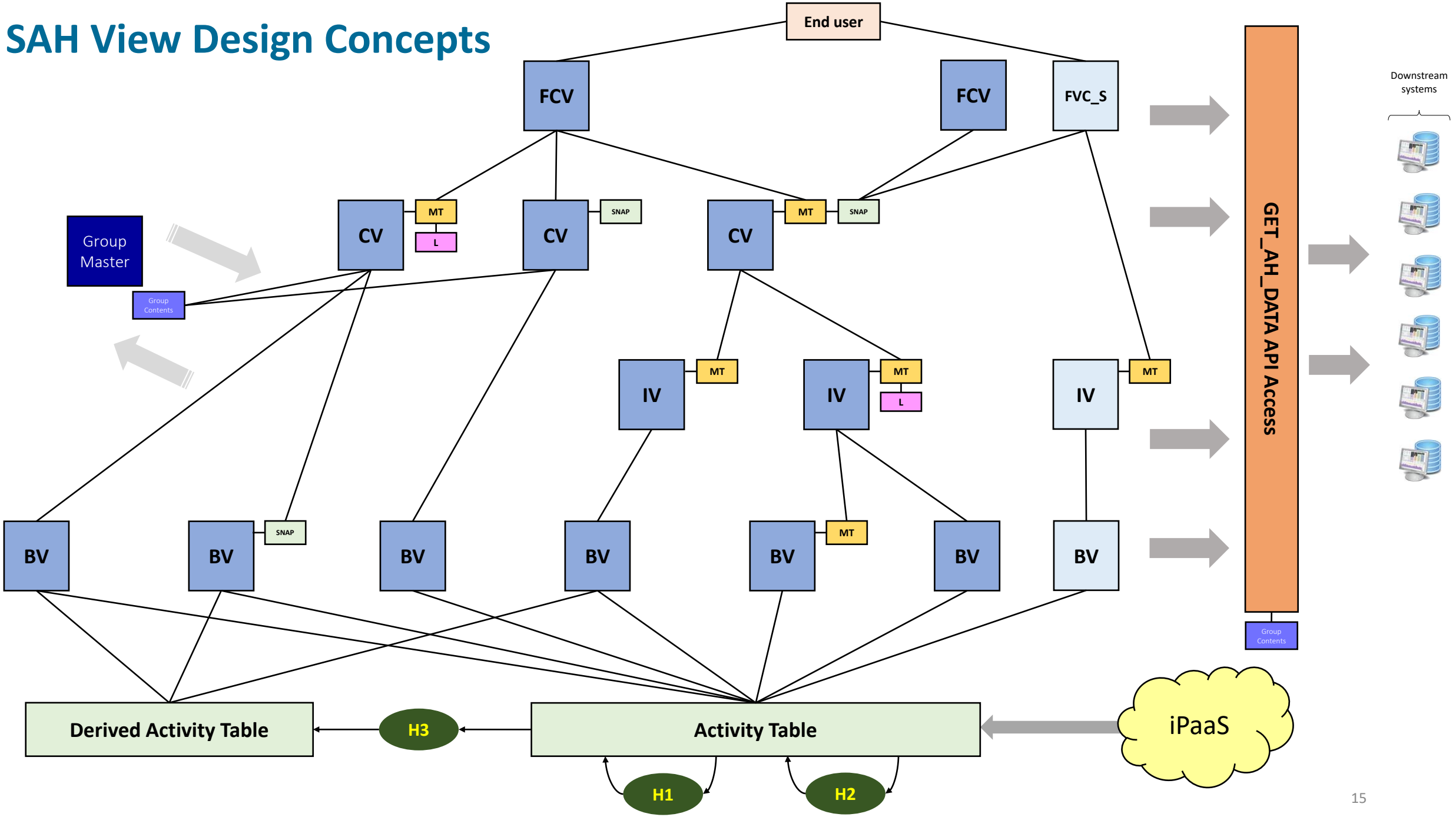
Final Curated Views

Combines data from CVs, Ivs or BVs as needed, fulfilling on an analysis ‘vignette’ or common need

Serve as Tableau and Cognos data sources

Can service API requests via column clause groups (in GET_AH_DATA)

SAH View Design Concepts



“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, schools, colleges, including course grades. Census and operational metrics

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

Student Statistics Per Term
Dozens of common student statistics, term-by-term for examining progression. Census and operational versions

Retention
Cohort, retention and graduation rates, etc. Census and operational metrics

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

Admissions
Applicants, Applications, Test Scores, Scholarships

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

LMS and other learning analytics
Canvas, OpenEdX, Kaltura. Canvas specific views and general learning event views

Student Activity Hub Views – Canvas & SIS

Canvas Batch

- SAH_FCV_CB_LMS_SUBMISSION
- SAH_FCV_LMS_CB_DISCUSSION_TOPIC_ENTRY
- SAH_FCV_LMS_CB_QUIZ_QUESTION_ANSWER
- SAH_FCV_LMS_CB_WIKIPAGE
- SAH_FCV_LMS_CB_ASSIGNMENT
- SAH_FCV_LMS_CB_CONVERSATION_MESSAGES
- SAH_FCV_LMS_CB_EXTERNAL_TOOL_ACTIVATION
- SAH_FCV_LMS_CB_STUDENT_STATS_PER_TERM
- SAH_FCV_LMS_CB_STUDENT_STATS_PER_WEEK
- SAH_FCV_LMS_CB_STUDENT_COURSE_STATS_PER_TERM
- SAH_FCV_LMS_CB_COURSE_STATS_PER_TERM
- SAH_FCV_LMS_CB_GROUPS
- SAH_FCV_LMS_CB_GROUP_MEMBERS
- SAH_FCV_LMS_CB_ENROLLMENTS

Canvas Caliper / Batch Real time

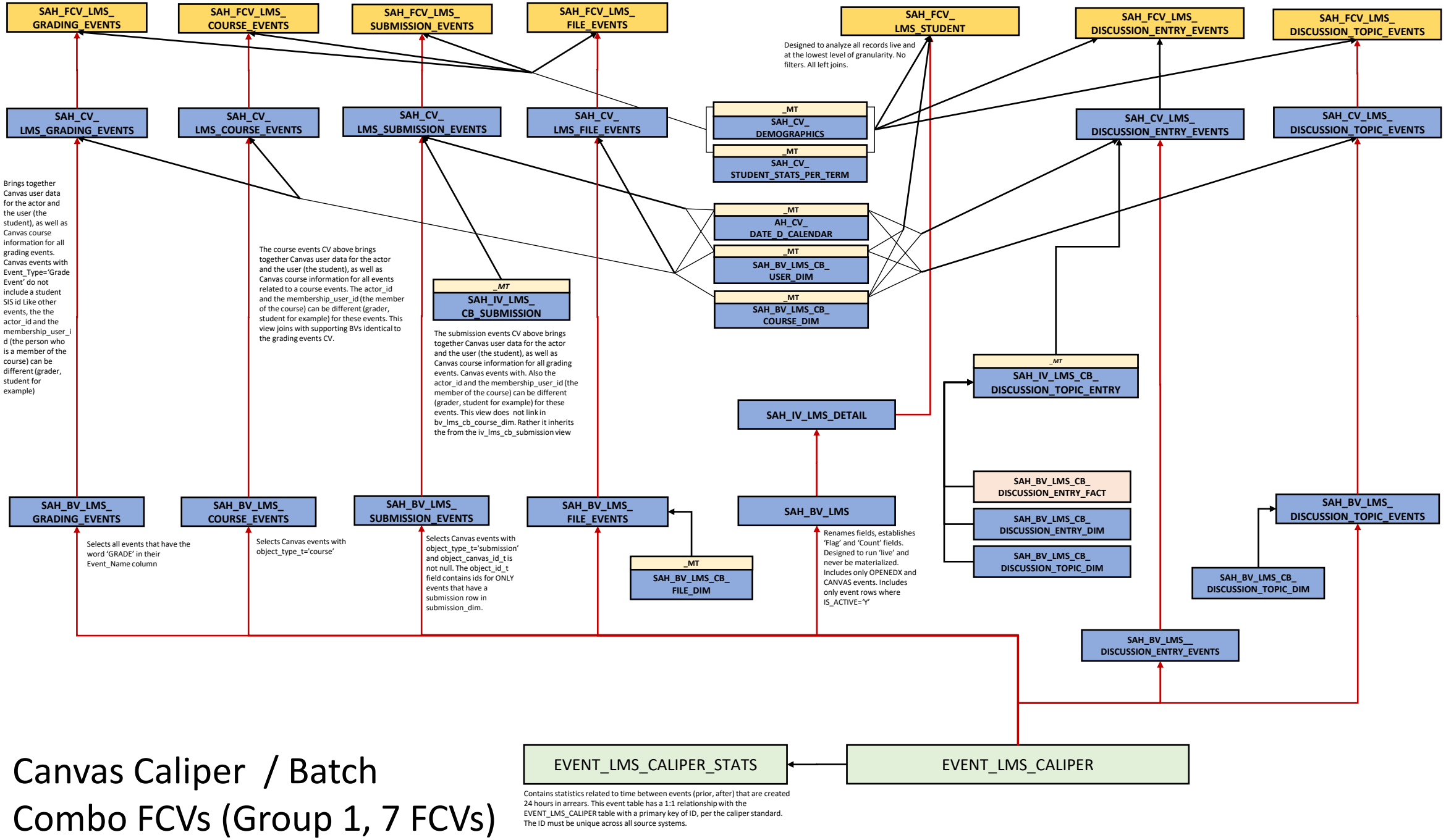
- SAH_FCV_LMS_STUDENT
- SAH_FCV_LMS_COURSE_EVENTS
- SAH_FCV_LMS_GRADING_EVENTS
- SAH_FCV_LMS_SUBMISSION_EVENTS
- SAH_FCV_LMS_FILE_EVENTS
- SAH_FCV_LMS_DISCUSSION_TOPIC_EVENTS
- SAH_FCV_LMS_DISCUSSION_ENTRY_EVENTS
- SAH_FCV_LMS_QUIZ_EVENTS
- SAH_FCV_LMS_WIKI_PAGE_EVENTS
- SAH_FCV_LMS_ASSIGNMENT_EVENTS
- SAH_FCV_LMS_EXTERNAL_TOOL_EVENTS
- SAH_FCV_LMS_GROUP_EVENTS
- SAH_FCV_LMS_ENROLLMENT_EVENTS

Canvas Caliper Stats

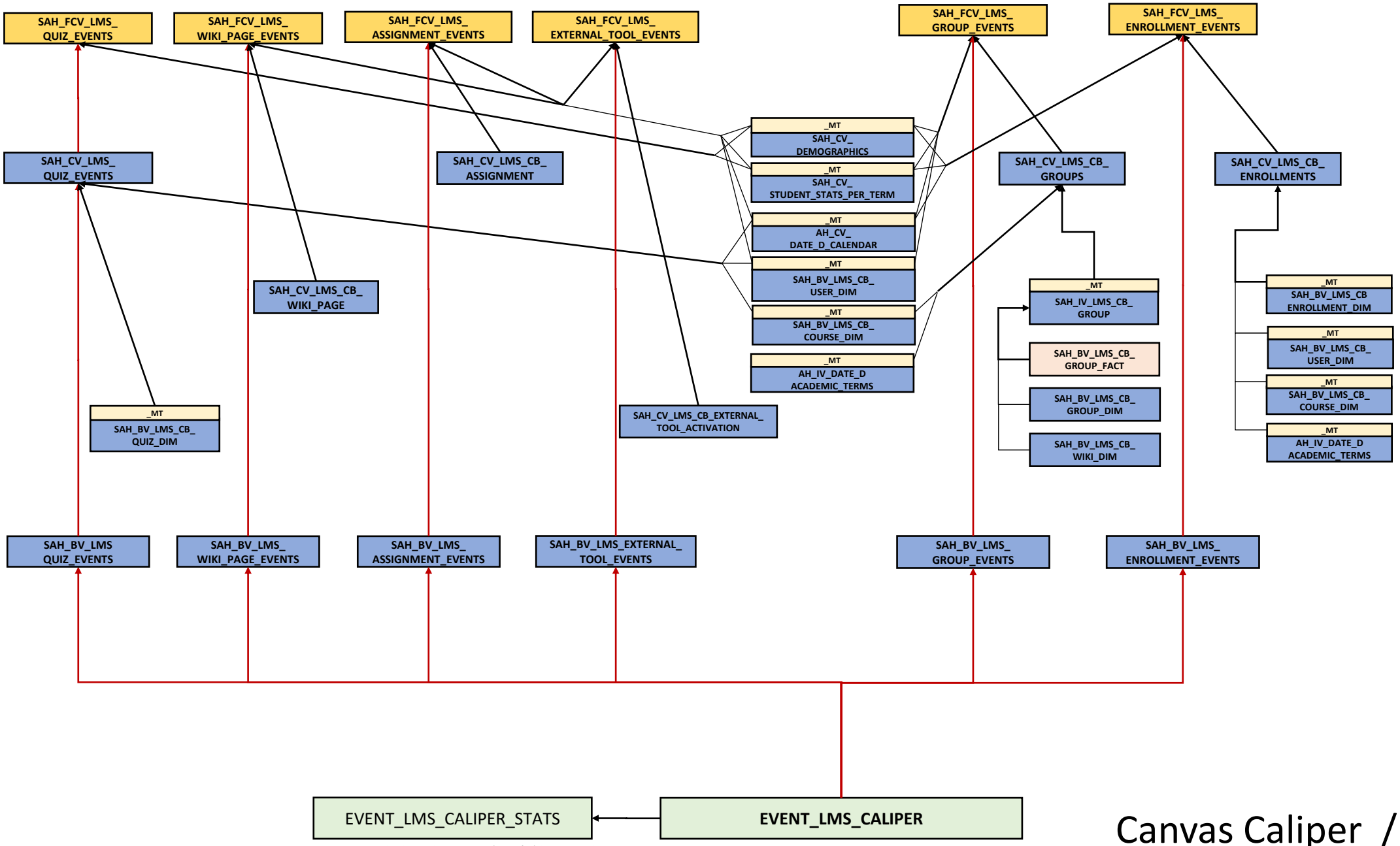
- SAH_FCV_LMS_STUDENT_STATS_PER_DAY
- SAH_FCV_LMS_STUDENT_STATS_PER_TERM
- SAH_FCV_LMS_STUDENT_STATS_PER_TERM_NARROW
- SAH_FCV_LMS_COURSE_STUDENT_STATS_PER_TERM
- SAH_FCV_LMS_SECTION_STUDENT_STATS_PER_TERM

Student System

- SAH_FCV_ADMISSION
- SAH_FCV_COURSE_STATS_PER_TERM
- SAH_FCV_DEGREE
- SAH_FCV_DEMOGRAPHICS
- SAH_FCV_ENROLLMENT
- SAH_FCV_ENROLLMENT_CENSUS
- SAH_FCV_HOUSING
- SAH_FCV_MAJOR_MINOR
- SAH_FCV_MAJOR_MINOR_DETAIL
- SAH_FCV_MAJOR_MINOR_NARROW
- SAH_FCV_MAJOR_MINOR_NARROW_CENSUS
- SAH_FCV_RETENTION
- SAH_FCV_RETENTION_CENSUS
- SAH_FCV_RETENTION_DETAIL
- SAH_FCV_RETENTION_DETAIL_CENSUS
- SAH_FCV_STUDENT_STATS_PER_TERM
- SAH_FCV_STUDENT_STATS_PER_TERM_CENSUS



Canvas Caliper / Batch Combo FCVs (Group 1, 7 FCVs)

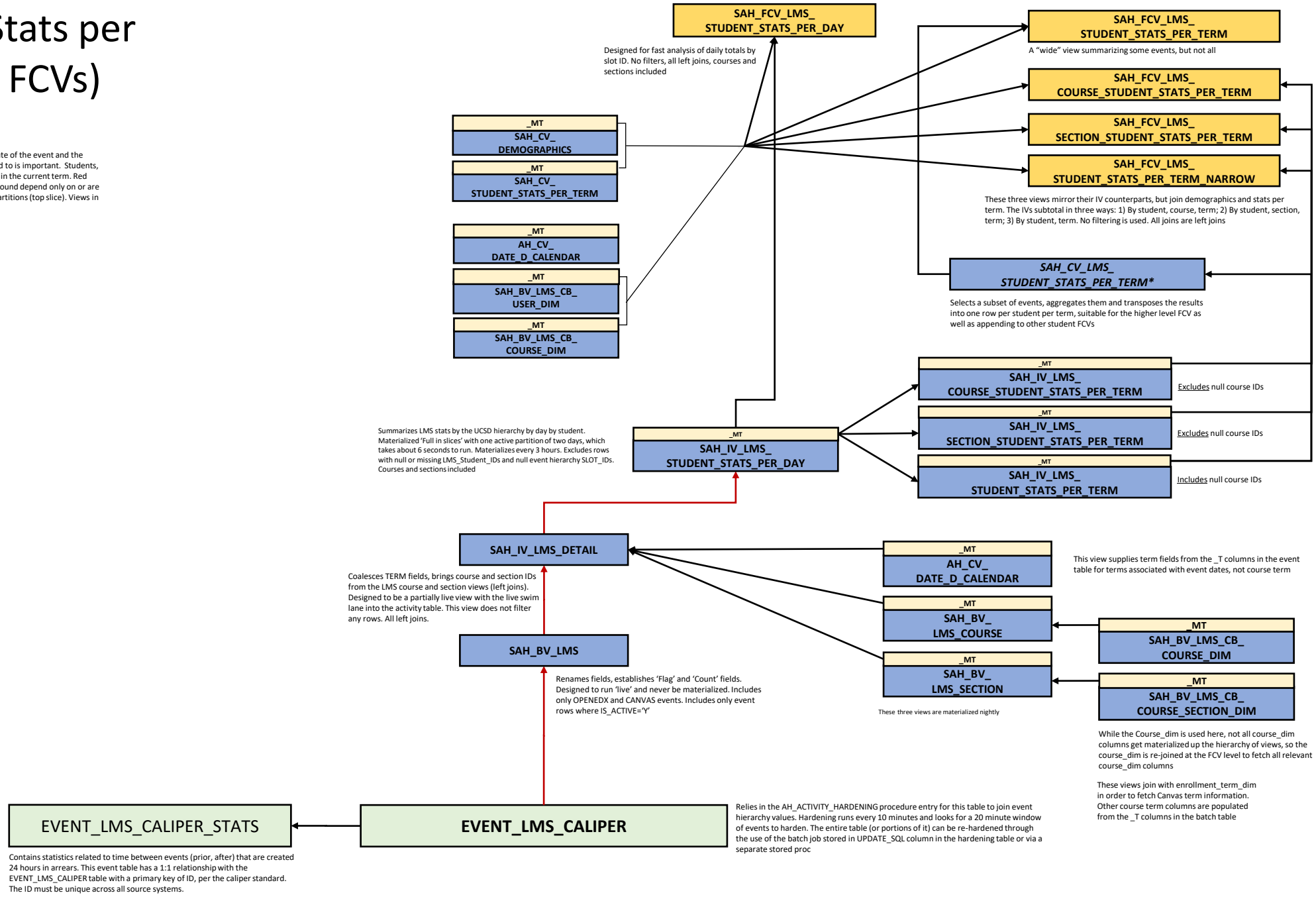


Canvas Caliper / Batch Combo FCVs (Group 2, 6 FCVs)

Canvas Caliper Stats per day, per term (5 FCVs)

Notes:

In order to analyze LMS event data, the difference between a) the date of the event and the 'calendar' term it falls into and b) the term that the course is assigned to is important. Students, instructors and TAs frequently have events coming from courses not in the current term. Red arrows indicate "live" real-time swim lanes. Views in the grey background depend only on or are hardened views that are rapidly incrementally with recently active partitions (top slice). Views in italics and marked with an * have not been constructed yet.



API access

- AH_GET_DATA stored procedure
  - Column clauses
  - Blanket, group, individual security
  - Linkage to GroupBuilder for additional filtering
  - All logged, elapsed time, rows returned, original SQL
- SAH_PUT_DATA
  - JSON parameter, write to any activity table
  - Controlled approach depending on transaction design style

Three server console applications

- AH Metadata Administration Console (AH-MAC) Q4-2021
  - Manage the metadata, automate aspects of view development, automate view migration, logging of everything
- Group Builder (GB) – done
  - Multi-pass, single-pass and pass-through queries
  - Multi-domain and multi-domain cross-walking
  - Single-pass web interface suitable for non-IT analysts (2022Q3)
- Message Builder (MB) – coming 2022Q1
  - Works with GB and prepares personalized content to be handed over to a CRM or message distribution service
- All Python apps

Activity Hubs: Synoptic Ontological Congruence

UC San Diego Activity Hubs

For trained decision makers and analysts across campus

Designed for "blended" analysis and data curation

Works in tandem with app-specific reports

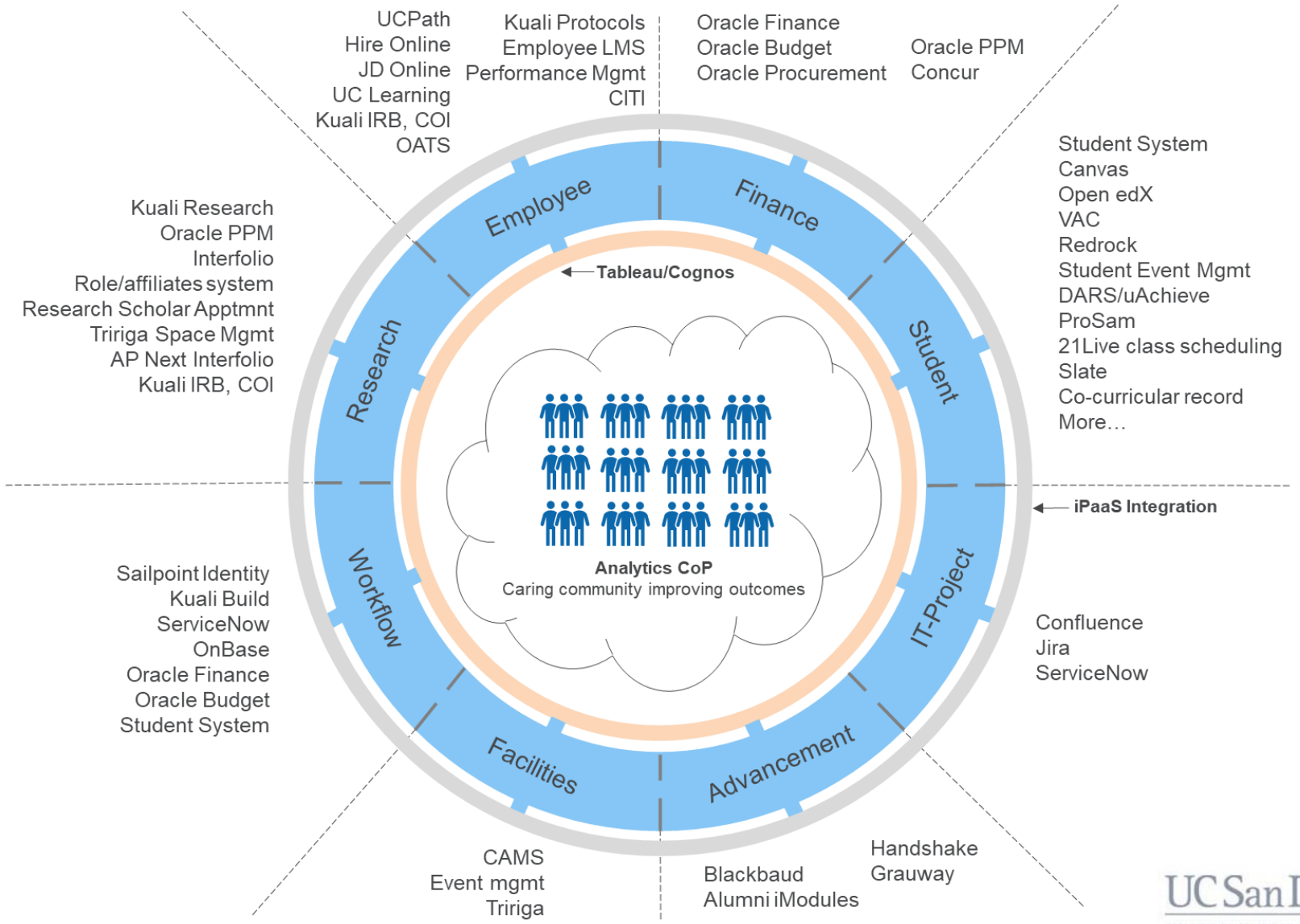
Community-driven environment

Secure. Private. Safe.

Includes data from all major systems

Local systems access via secure integration (iPaaS)

Integrated with mobile and advanced analytic & ML technologies



UC San Diego
INFORMATION TECHNOLOGY SERVICES

What kinds of collaborations are available?

Institutions can choose the level of collaboration they need and increase or decrease their level of collaboration. Our goal is to support whatever the institution needs

1. **Use SAH - Partnering institution.** Determine the level of control and services you wish and start using the platform. We let institutions choose between a full-service relationship (we do everything) to a low-service relationship (you do everything) or anything in between, adjusting prices accordingly. Anything is possible.
2. **Share views with each other - SAH developer marketplace:** Institutions can share locally developed views and other code with each other for free in a view sharing marketplace web site. SAH was designed to support community development without compromising the core view software engineering quality
3. **Help co-develop new core views - SAH co-development institutions:** Institutions that develop views or other software can bring them to the SAH co-development partners for inclusion into the core product. On a going-forward basis, a portion of the SAH licensing fees can be shared with co-development partners, recovering institutional development costs

What are the 3rd party partnering opportunities?

1. **Consulting partner:** Strategy, organizational development, governance, change management, implementation planning
   - Embargoed
2. **Implementation partner:** Implementation planning, implementation, transition management
   - Slower, Inc., Embargoed
3. **Adapter development:** Write adaptors for specific source systems (e.g., PS 9.2, Banner, etc…)
   - Embargoed
4. **Service delivery:** Help with ongoing managed services and service delivery
   - Slower, Inc.
5. **Platform development:** Help with platform enhancements, including ML and advanced analytics
   - SAP
6. **Cloud providers:** Host the SAH environment
   - AWS

Roadmap ahead: 2022 and beyond

- Data management views (DM)
  - These views measure different aspects of data quality and data management processes, organizational ability to ensure high quality data
- Machine Learning Platform
  - Advanced analytics, predictive analytics, ANNs, etc.
- Degree modeling language (DML)
  - Graph theory applied to validation of degree completion rules and simulation of major and minor choices, replacing conventional degree audit and planning tools
- Database application development
  - Use SAP HANA as a transaction environment that reads/writes activity tables, eliminating integration