Americas' SAP Users' Group

THE AI-DRIVEN UNIVERSITY: DREAM OR HALLUCINATION?

Amin Qazi, UC San Diego | Amass.co

Session ID 84317

About the Speaker

Amin Qazi

- Enterprise Architect,
  UC San Diego

- Principal, Amass

Key Outcomes/Objectives

1. Understand AI, and data and data warehousing roles in supporting AI
2. Show how capturing events in near-real time can manage risk
3. Where to embed AI, including data streams

Agenda

- What is a university?
- UC San Diego's data warehouse & AI strategies
- Use cases that inform thinking, design, and implementation

Summary

- AI is:
  - important and challenging
  - changing how you collect, store, and use data
  - critical

- UC San Diego has taken an aggressive approach to change our thinking, and adopt new processes to prepare for this change

- Accomplishing the same is possible, but may be hindered by strategy, leadership, and resources

Universities are complex, multi-faceted ~~organizations~~ businesses.

- Finance
- HR
- Education
- Research
- Hospitality
- Medical
- Athletics
- Facilities
- Retail
- Logistics
- Investments

They are also mired in outdated ways of thinking.

- Finance
- HR
- Education
- Research
- Hospitality
- Medical
- Athletics
- Facilities
- Retail
- Logistics
- Investments

AI drives benefits across the organization.

Organizations implementing AI report the following benefits:

Influencing Sales

3 in 4 increase sales of new products and services by more than 10%

Boosting Operations

78% increase operational efficiency by more than 10%

Engaging the Customer

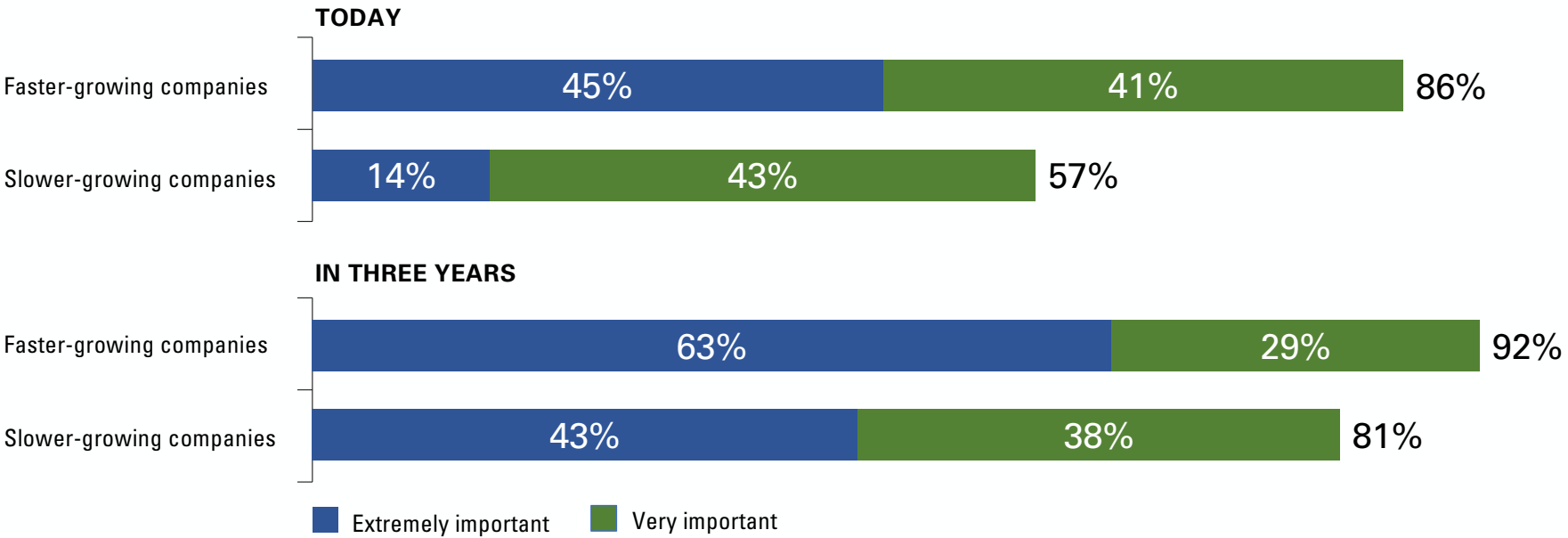
75% enhance customer satisfaction by more than 10%

Generating Insights

79% generate new insights and better analysis

AI is already embraced by fast-growers*.

Organizations' rating of the importance of AI today and in three years.



* Fast-growing companies are defined as those whose revenue growth is "far above" industry average.
n = 975 executives in Europe and the U.S.

UC San Diego | AVIASS.co

Source: Cognizant

ASUG

Can we get to the AI Promised Land?

Can we get to the AI Promised Land?

Spoiler

- Yes
- And no
- And maybe

Can we get to the AI Promised Land?

Spoiler

- Yes
- And no
- And maybe

Out of scope

- Operations
- Talent management
- Regulations
- Ethics
- etc.

What is AI?

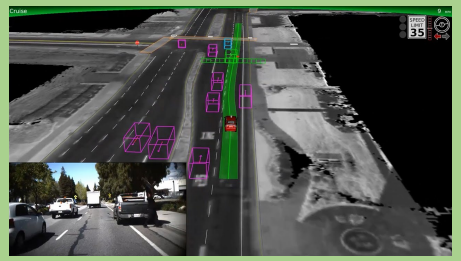
Artificial Intelligence "the science of getting computers to act without being explicitly programmed" (Stanford University)

Machine Learning "ability to learn without being explicitly programmed" (Arthur Samuel)

- **Reinforcement Learning**
- **Supervised Learning**
- **Predictive Learning**

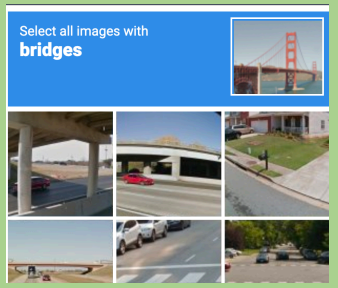
What is AI?

Reinforcement Learning



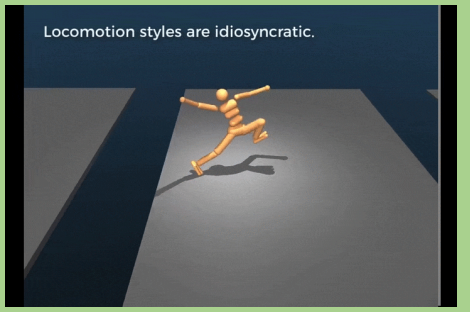
Control
(no target variable)

Supervised Learning



Classification
(categorical target variable)

Predictive Learning



Control
(categorical target variable)

UC San Diego | AVIASS.co

ASUG

Where was AI in 2016?

" it makes suggestions that no human would ever make, like adding milk chocolate to a clam linguine or mayonnaise to a Bloody Mary."

Hoof-n-Honey Ale (by IBM Watson)

- 4 oz. India pale ale
- 1 oz. veal stock
- 2 slices peach
- 4 oz. Burgundy wine
- 1 egg white
- 1 tsp. sugar
- 1 piece grilled beef
- Honey
- Warm water

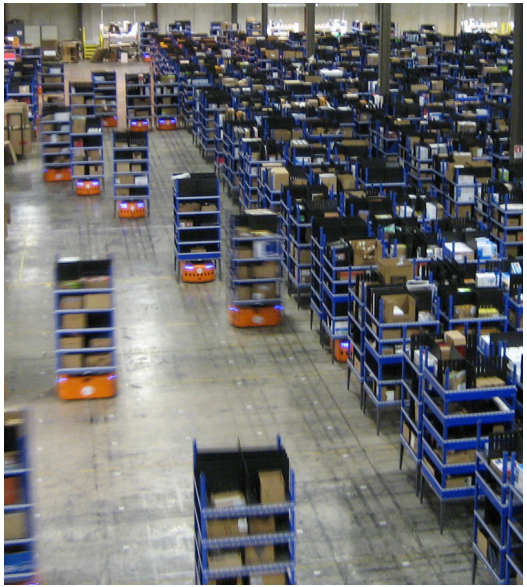
UC San Diego | AVIASS.co

Source: The New Yorker, November 20, 2016

ASUG

Where is AI today?

warehouses



aws

capacity planning

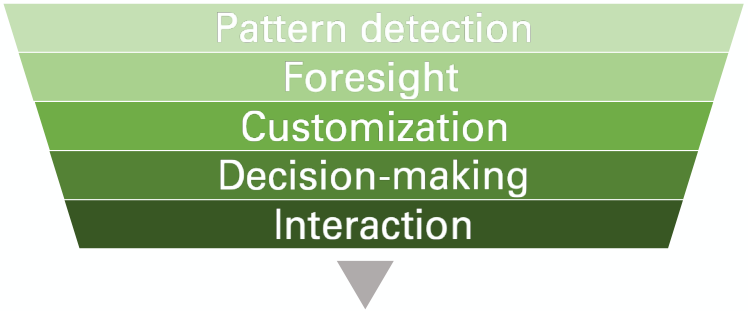
visual assessment

UC San Diego | AVIASS.co

Source: The Economist, April 13-19, 2019

ASUG

What is AI?



Doing the same things, better — *Doing something radically different*

Leaner, faster operations
- Using automation to improve the efficiency of business-as-usual processes
- Reducing the cost of simple, routine processes, while maintaining or improving quality of experience

Tailored products and advice
- Personalizing interactions to more closely meet the unique needs of customers
- Providing convenient, high-quality service, while maintaining scalability

Ubiquitous presence
- Making products and services available to customers in their preferred format and channel
- Expanding the reach of institutions' channels and offerings geographically and across customer segments

Smarter decision-making
- Using advanced data science to optimize business outcomes (e.g., higher retention rates)
- Integrating large volumes of data to derive better insights across the university (e.g. better space management)

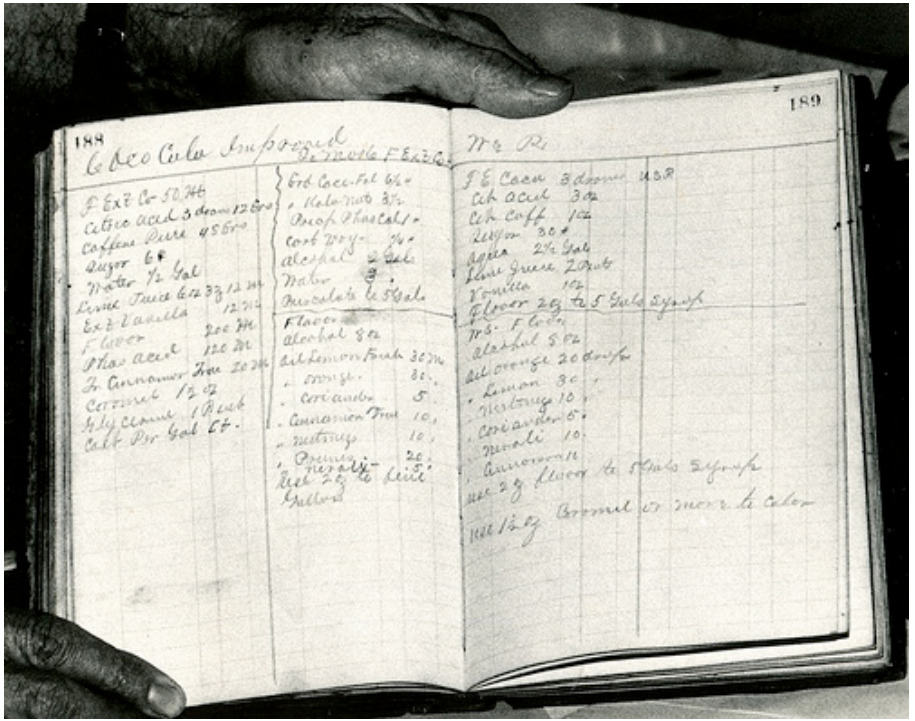
New value propositions
- Differentiating offerings through new operating models and ways of working
- Building brand new pedagogies, products, services and business models that use AI at the core

UC San Diego | AVIASS.co

Source: World Economic Forum

ASUG

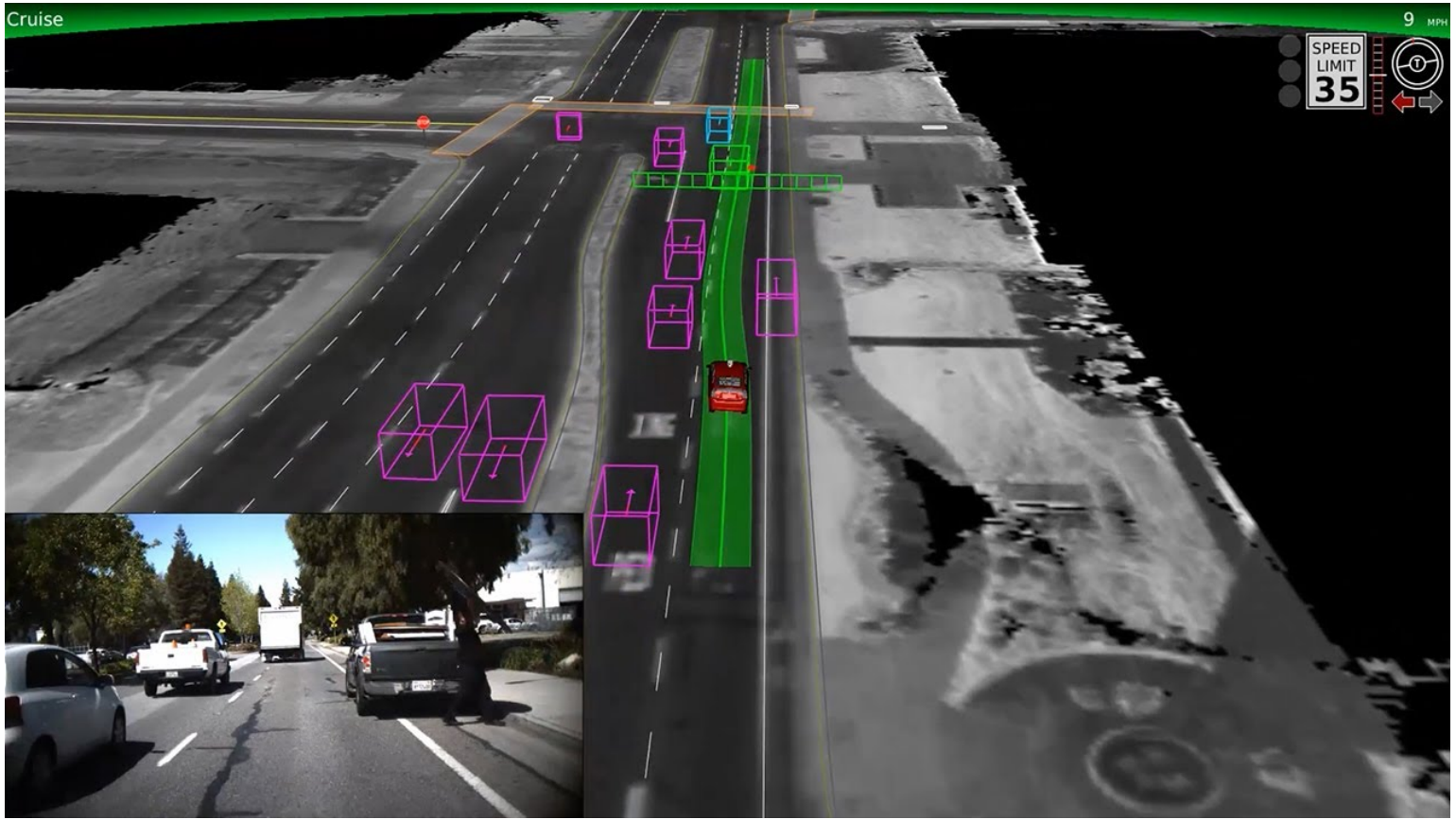
Data are priceless.



Data collection is changing.

Not more of
the same.

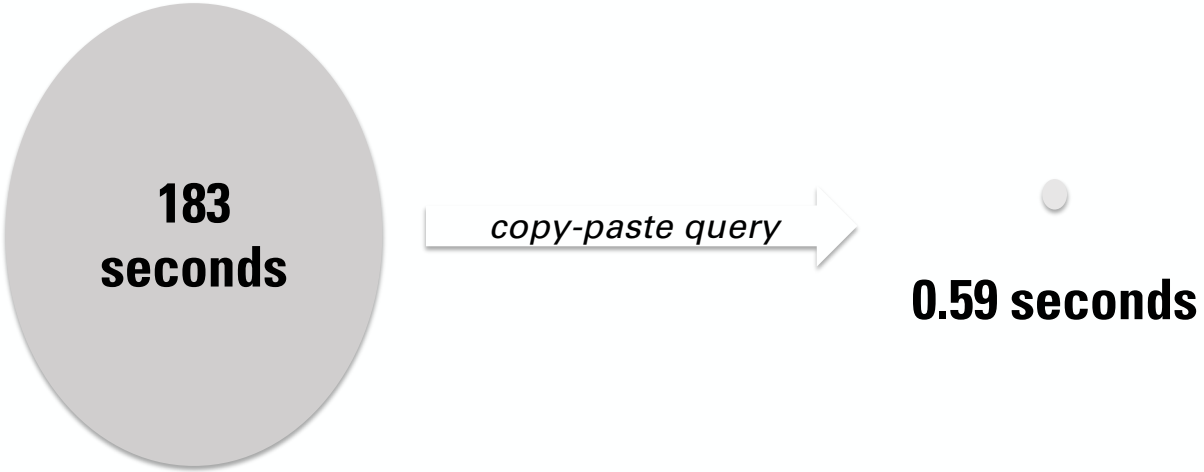
Different.



UC San Diego | AVIASS.co

ASUG

Technology affects the realm of possibilities.



unnamed
(traditional database)

SAP HANA
(columnar in-memory database)

UC San Diego | AVIASS.co

ASUG

Data are different.

Data are different.

Data management needs to be different.

Data processing will be different.

Again, not more of the same.

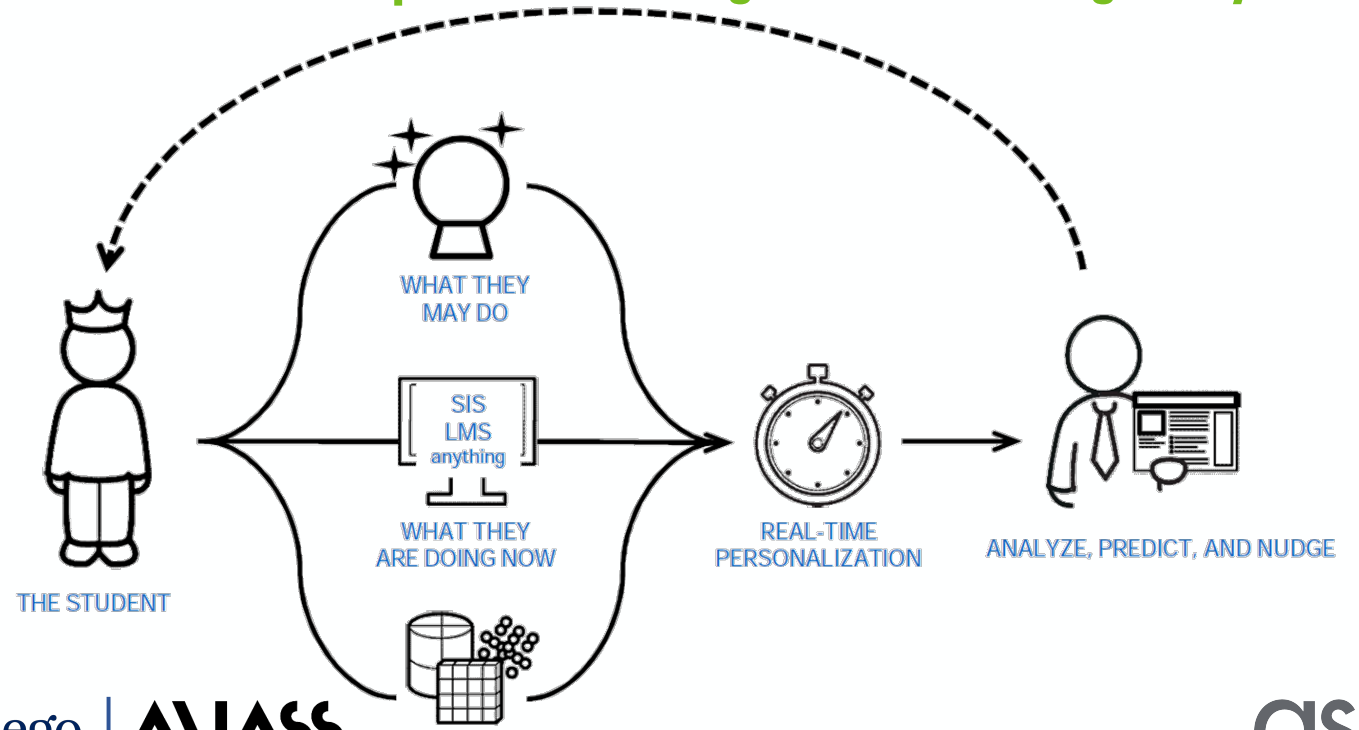
New reference architecture for data to support AI.

Design Principles

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. Curated views
   - No dimensional modeling. No joins! Curated views present a long list of attributes for analysts to choose from
   - Curated views are designed for specific analysis needs (vignettes)
4. Speed and ease is of the essence
   - 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. Redundancy and data explosion are good
   - No need to conserve space. Curated views can be overlapping and duplicative
   - A hierarchy of reusable SQL code results in an OO-like, highly reusable environment
6. Democratize it
   - Make it easier to understand, consume and use
   - Enable the community to share, encourage bottom-up data analysis and use

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



UC San Diego | AVIASS.co

ASUG

UC San Diego's next generation data warehouse uses

Employee Activity Hub

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

Project Activity Hub

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

Research Activity Hub

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

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

Financial Activity Hub

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

Student Activity Hub

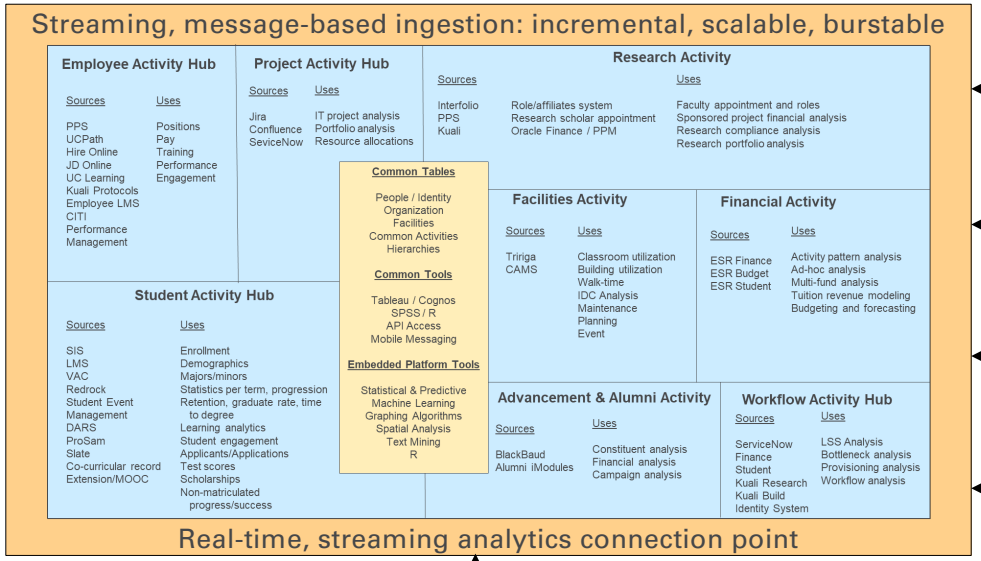
Sources	Uses
SIS	Enrollment
LMS	Demographics
VAC	Majors/minors
Redrock	Statistics per term, progression
Student Event Management	Retention, graduate rate, time to degree
DARS	Learning analytics
ProSam	Student engagement
Slate	Applicants/Applications
Co-curricular record	Test scores
Extension/MOOC	Scholarships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

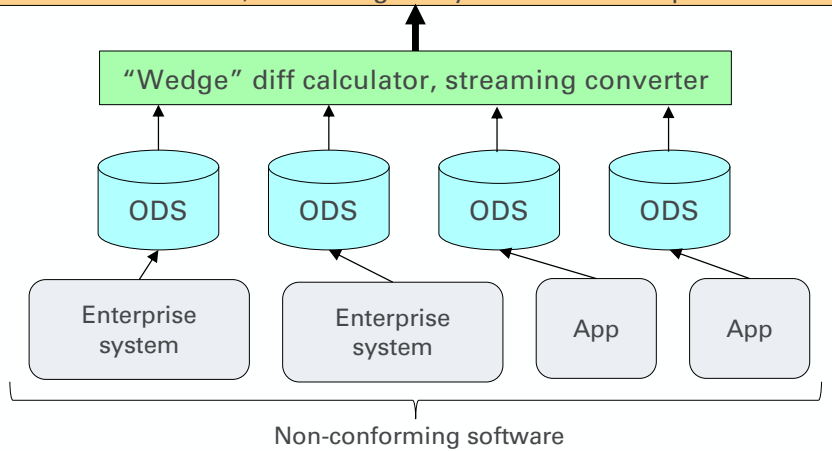
Advancement & Alumni Activity Hub

Sources	Uses
BlackBaud	Constituent analysis
Alumni	Financial analysis
iModules	Campaign analysis

Workflow Activity Hub

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





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.

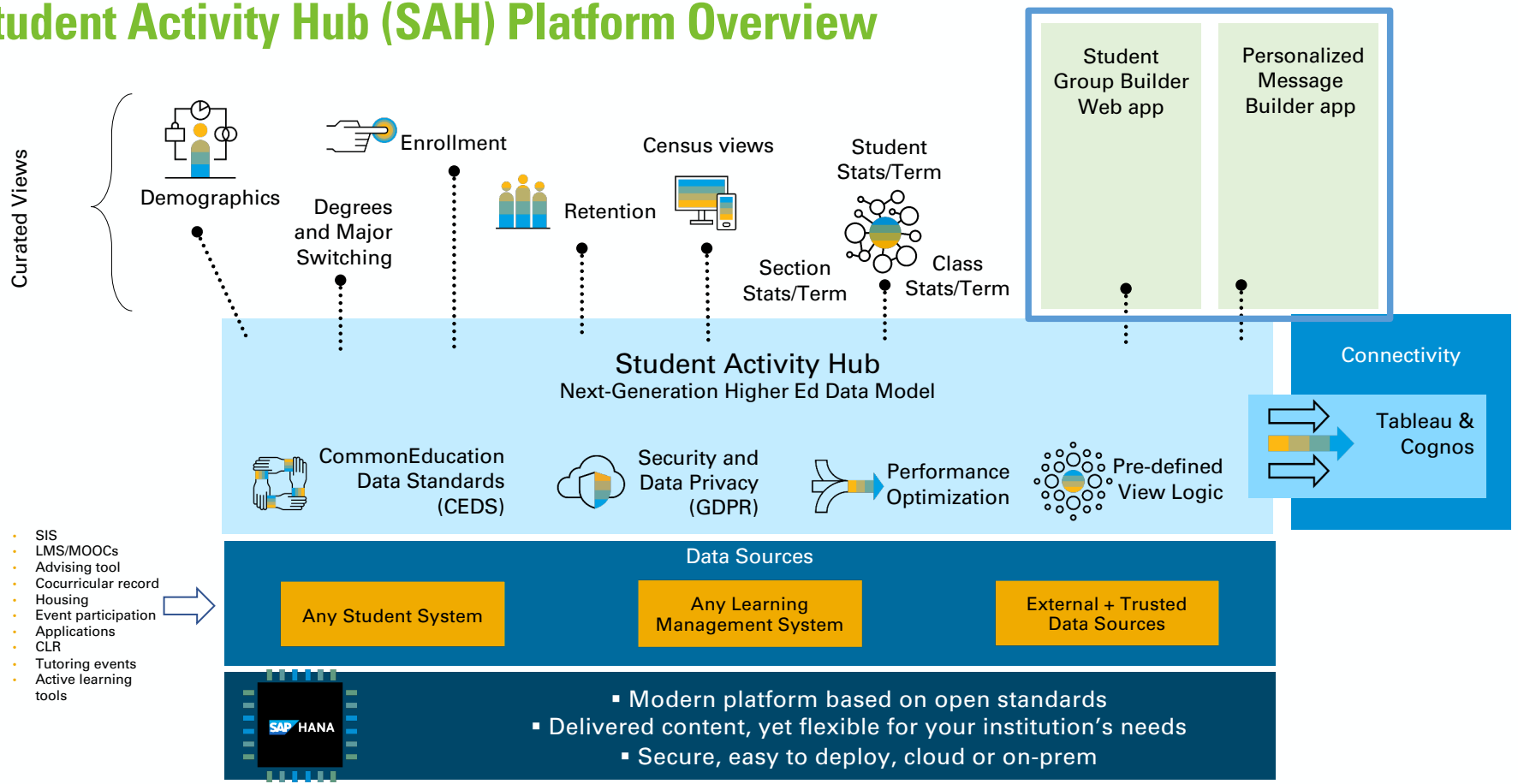
The streaming analytics connection point allows for directly connecting the streaming ingestion engine with a real-time streaming analytics machine learning platform to process inbound messages

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.

ASUG

Student Activity Hub (SAH) Platform Overview



UC San Diego | AVIASS.co

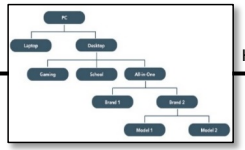
ASUG

Activity Hub architecture

Source systems/devices
a. Emit from point of entry, full incremental
or
b. Simulate incremental from DB

Stream in ->

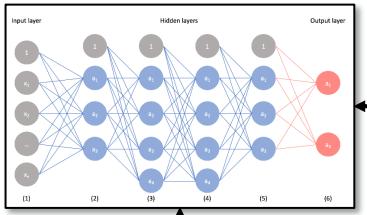
Hierarchy manager



<- Hierarchy slot ID + [attributes]

Hierarchy slot attributes ->

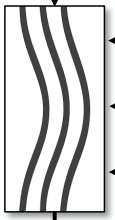
Machine learning platform (MLP)



<- Model development ->

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)

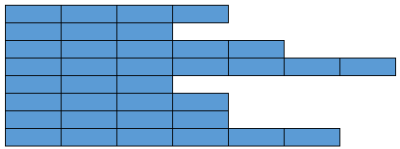
<- Message out



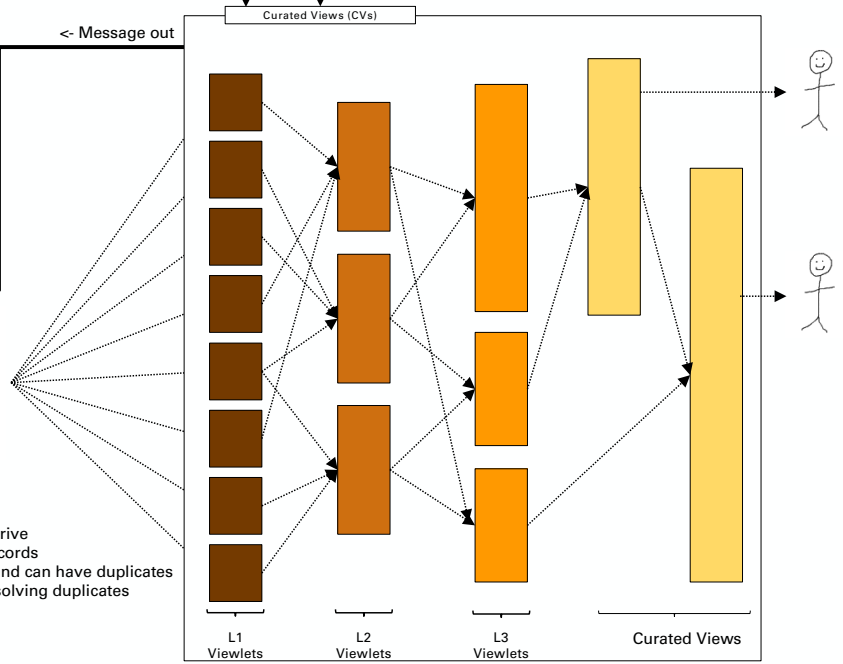
iPaaS
a. Simple, parallel streams
b. Minimal hops, steps, merging
c. Save transformation for CVs
d. Easily restartable
e. Save extra data in a bag

<- Message out

Stream in ->

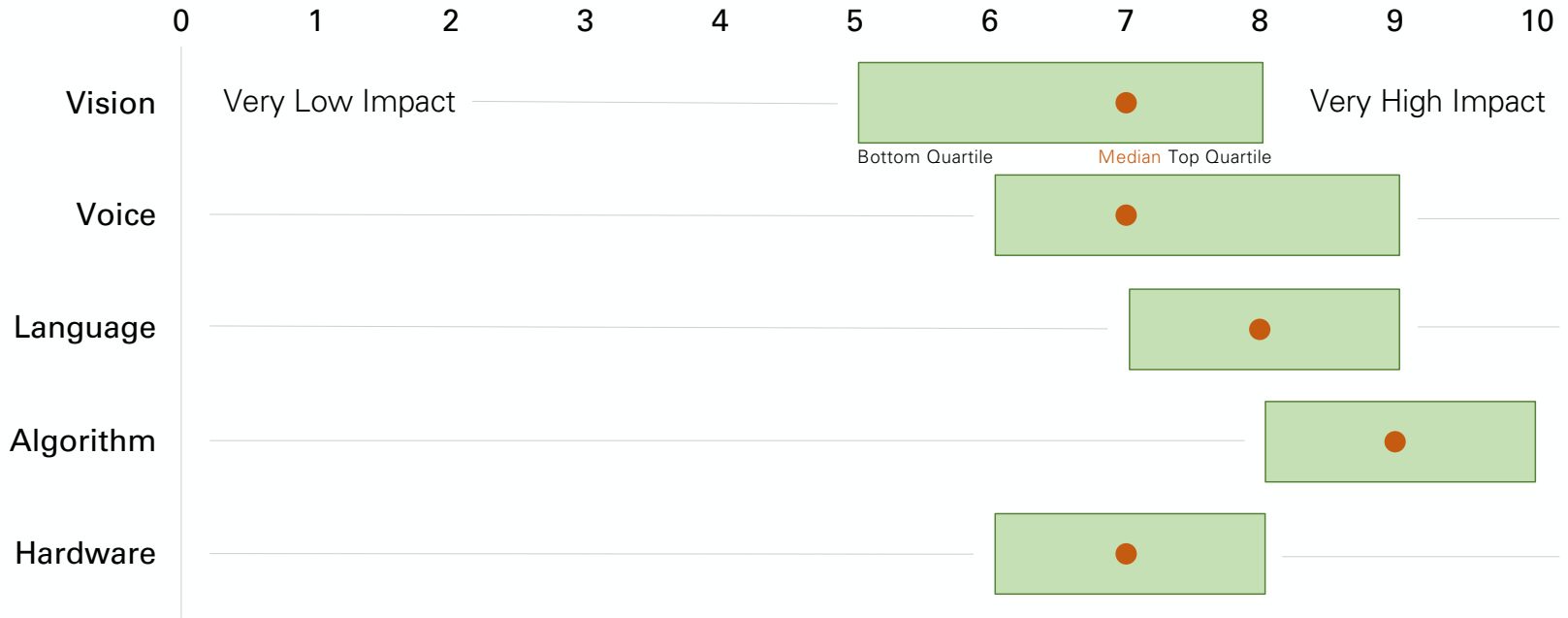


Activity table (pile file)
1. Records have different length
2. Record have different fields
3. Records are added in the order they arrive
4. Adds, updates, deletes are different records
5. Records are from idempotent stream and can have duplicates
6. Records have unique identifiers for resolving duplicates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



UC San Diego | AVIASS.co

There are very high potential and impact from algorithms; hardware and vision, less so.



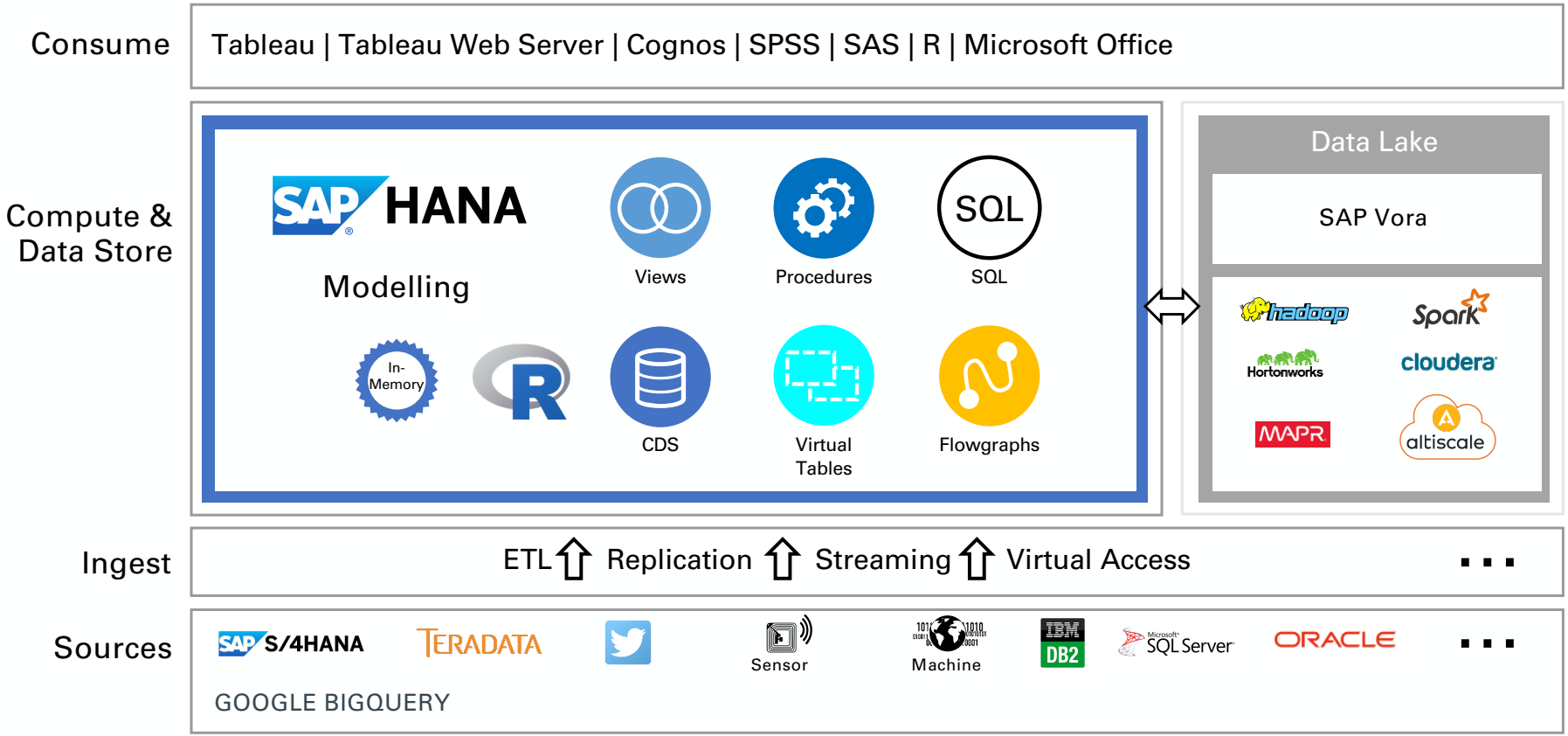
UC San Diego | AVIASS.co

Source: HolonIQ Global Executive Panel, April 2019, n=377

ASUG

SAP HANA

Data Ingestion and Analytics modelling overview



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

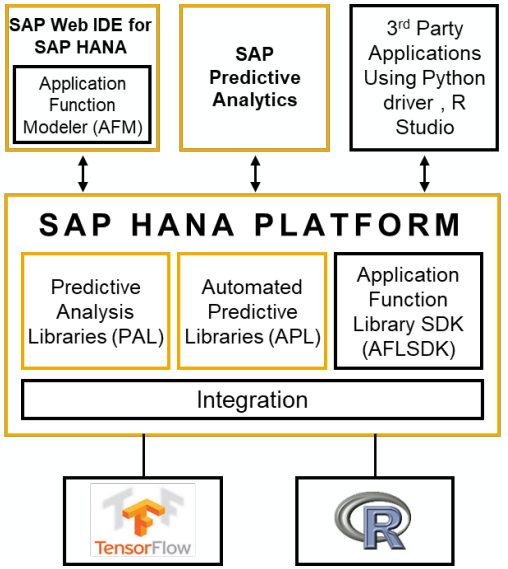
Recommender
- Factorized Polynomial Regression Models

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



- 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

UC San Diego | AVASS.co

ASUG

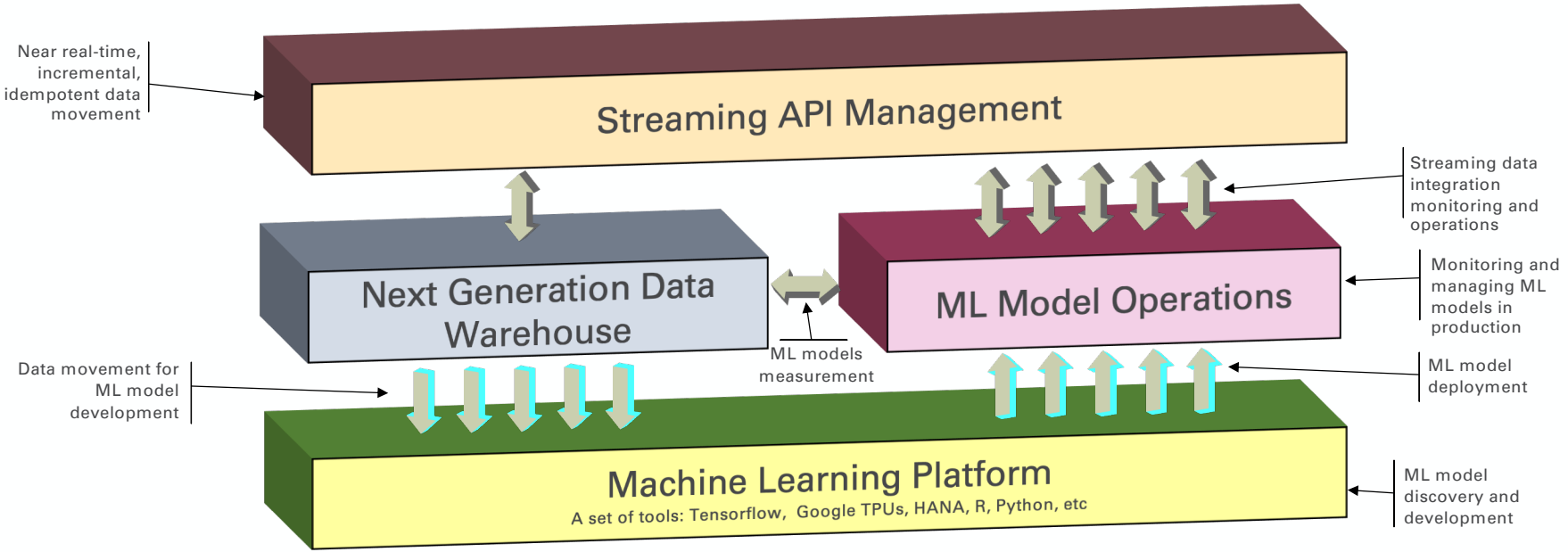
Managing multiple ML models in the next generation analytics

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

Managing multiple ML models in the next generation analytics

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



UC San Diego | AIASS.co

ASUG

What are our ingredients for technology?

- Data warehouse
- Data management
- Machine Learning
- Analysis
- Application / Interaction / Intervention

Back to how ML/AI might apply at a university.

- Hospitality
- Medical
- Education
- Athletics
- Facilities
- Retail
- Logistics / Transportation
- Investments

Without a clear goal, AI initiatives may be doomed from the start.

	Defined Success Criteria
Finance	+
HR	+
Education	-
Research	+
Hospitality	-
Medical	+
Athletics	+
Facilities	+
Retail	+
Logistics	+
Investments	+

A common ontology greatly supports organization of data.

	Defined Success Criteria	Common Ontology
Finance	+	+
HR	+	-
Education	-	-
Research	+	-
Hospitality	-	-
Medical	+	+
Athletics	+	-
Facilities	+	+
Retail	+	-
Logistics	+	+
Investments	+	+

In many areas, data are either not collected, or are scattered, de-centralized, and disorganized.

	Defined Success Criteria	Common Ontology	Structured Data
Finance	+	+	+
HR	+	-	+
Education	-	-	-
Research	+	-	-
Hospitality	-	-	-
Medical	+	+	-
Athletics	+	-	-
Facilities	+	+	+
Retail	+	-	+
Logistics	+	+	+
Investments	+	+	+

Many data are subjective, making machine learning more challenging.

	Defined Success Criteria	Common Ontology	Structured Data	Data Indep. of Human Variables
Finance	+	+	+	+
HR	+	-	+	-
Education	-	-	-	-
Research	+	-	-	+
Hospitality	-	-	-	-
Medical	+	+	-	-
Athletics	+	-	-	-
Facilities	+	+	+	+
Retail	+	-	+	+
Logistics	+	+	+	+
Investments	+	+	+	+

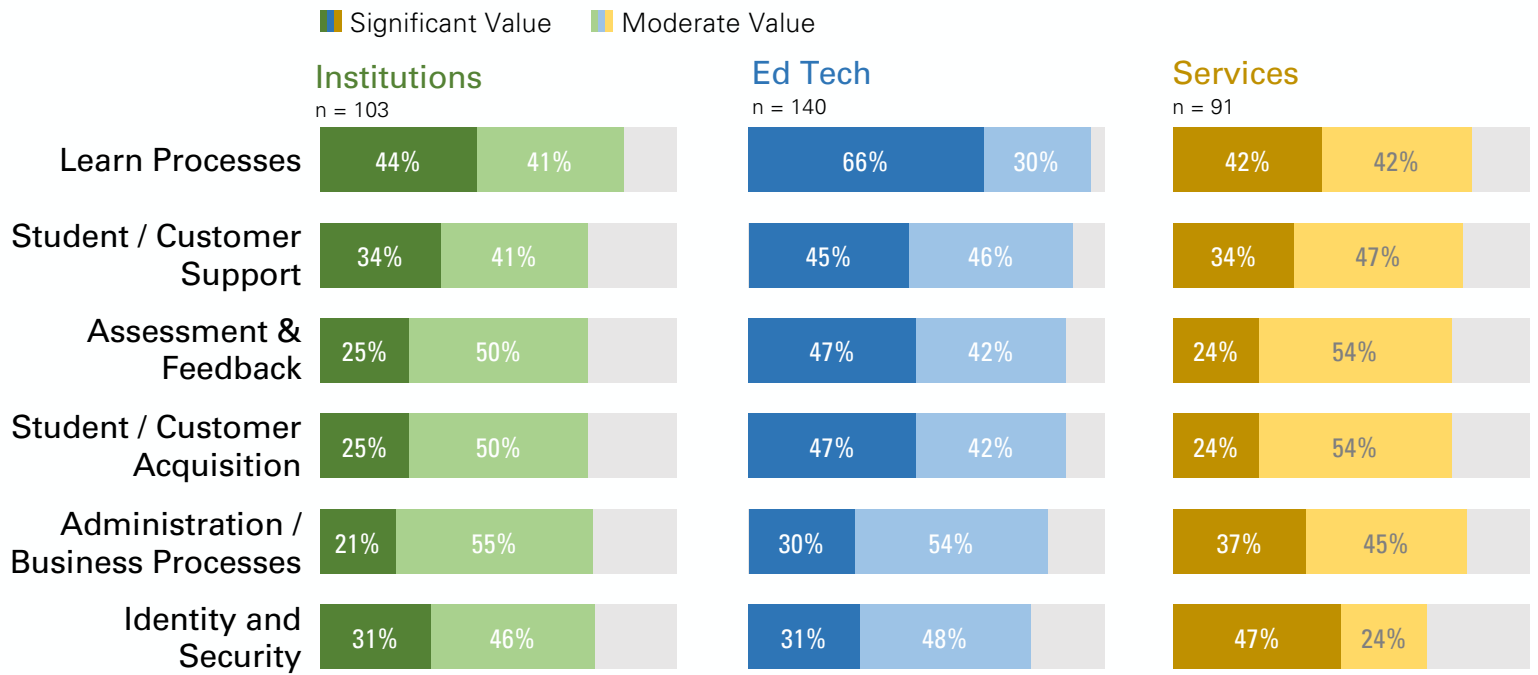
UC San Diego | AVIASS.co

ASUG

Much of a university's operations can benefit from the use of AI.

	Defined Success Criteria	Common Ontology	Structured Data	Data Indep. of Human Variables	Overall
Finance	+	+	+	+	+
HR	+	-	+	-	+
Education	-	-	-	-	-
Research	+	-	-	+	-
Hospitality	-	-	-	-	-
Medical	+	+	-	-	+
Athletics	+	-	-	-	-
Facilities	+	+	+	+	+
Retail	+	-	+	+	+
Logistics	+	+	+	+	+
Investments	+	+	+	+	+

There are still opportunities to extend the value of AI in many domains.



UC San Diego | AIASS.co

Source: HolonIQ Global Executive Panel, April 2019, n=334

ASUG

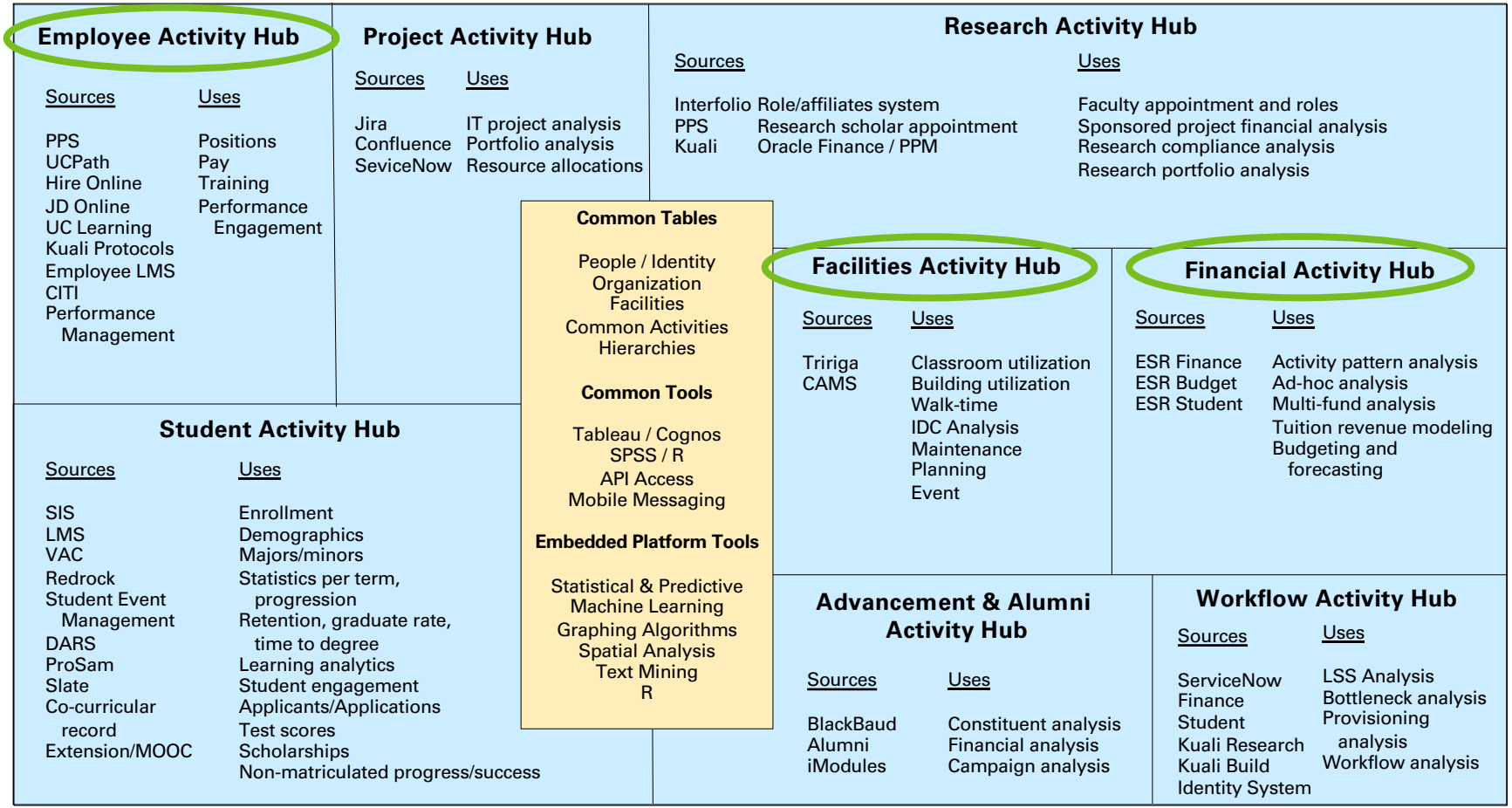
Much of a university's operations can benefit from the use of AI.

	Defined Success Criteria	Common Ontology	Structured Data	Data Indep. of Human Variables	Overall
Finance	+	+	+	+	+
HR	+	-	+	-	+
Education	-	-	-	-	-
Research	+	-	-	-	-
Hospitality	-	-	-	-	-
Medical	+	+	-	-	+
Athletics	+	-	-	-	-
Facilities	+	+	+	+	+
Retail	+	-	+	+	+
Logistics	+	+	+	+	+
Investments	+	+	+	+	+

UC San Diego | AIASS.co

ASUG

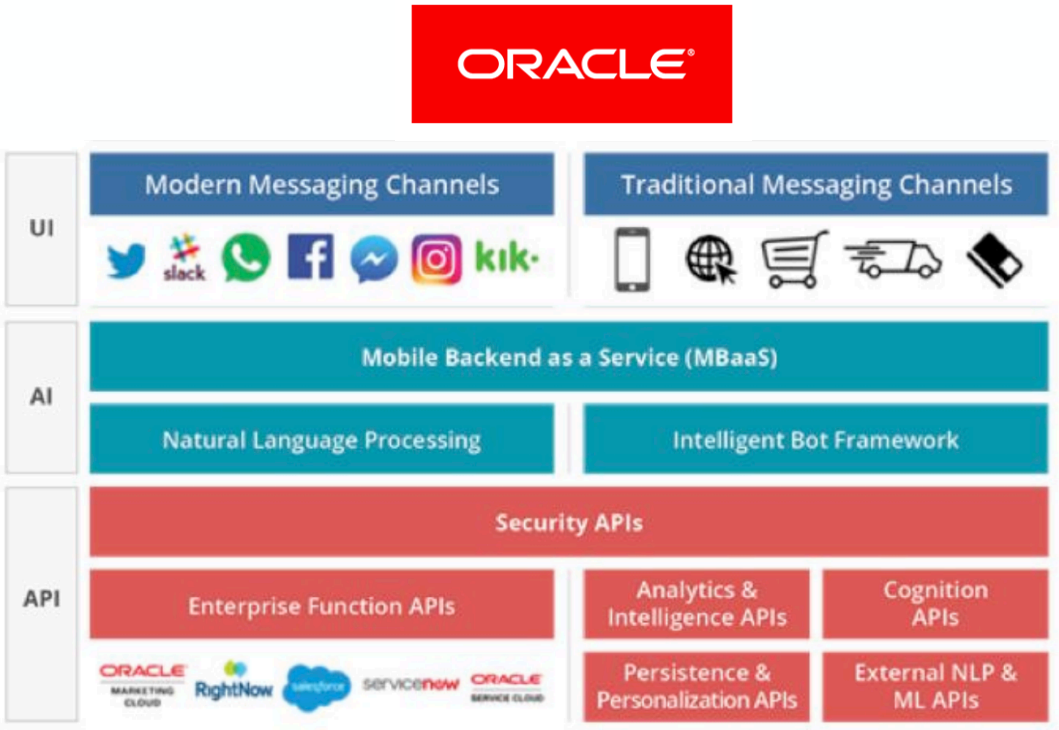
UC San Diego's next generation data warehouse uses



Streamlining student interactions with AI.

University of Adelaide used Oracle Intelligent Bots to reduced call center wait times by 97% and resulted in 60% "Awesome" rating from students.

"The AI/natural language means the bot infers what a user means *without having to hard-code every possible question*. This vastly improves the bot's ability to provide a correct answer."



UC San Diego | AIASS.co

Source: Rubicon Red, 2018

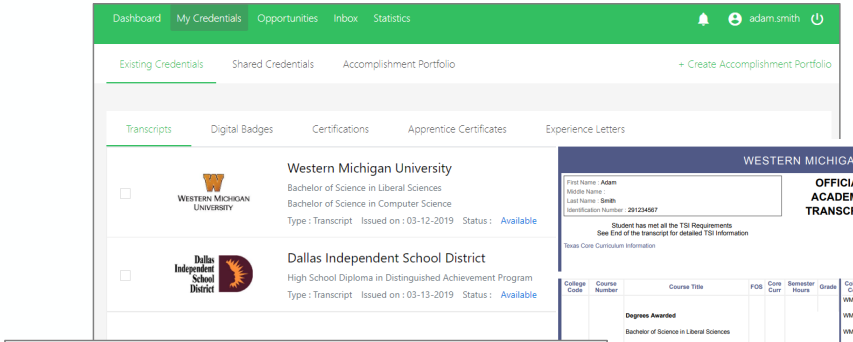
ASUG

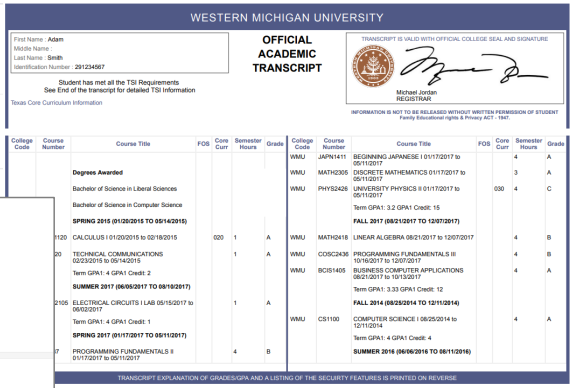
Matching student credentials with job descriptions using AI.

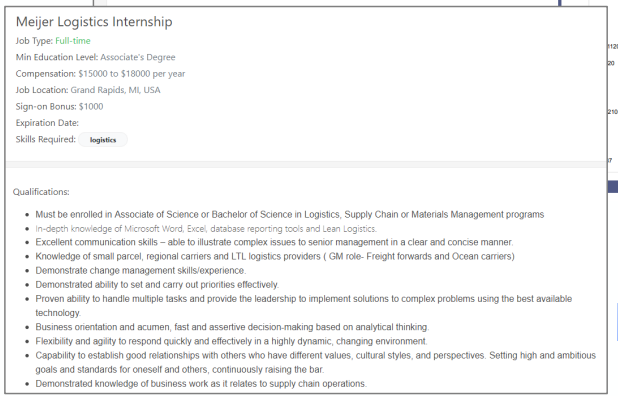
GREENLIGHT

Greenlight (glcredentials.com) provides **secure, simple, and instant sharing and validation of candidate records** for academic admissions and transfers, internships, scholarships, and job applications.

They streamline manual processes using distributed ledger technology, or blockchain, allowing employers to digitally receive records that are secure and instantly tamper evident.





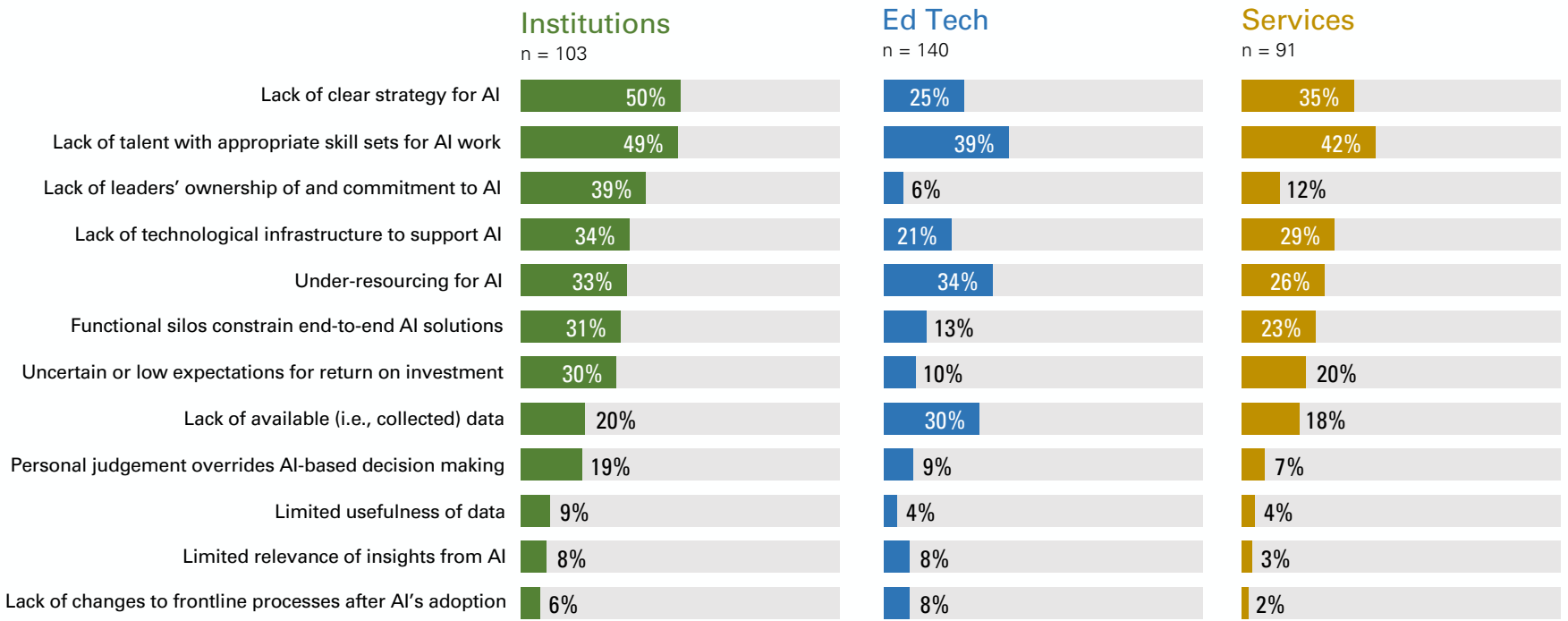


AI to match student academic records and professional interests to open opportunities.

UC San Diego | AVIASS.co

ASUG

Institutions struggle most with basics: strategy, talent, and leadership.

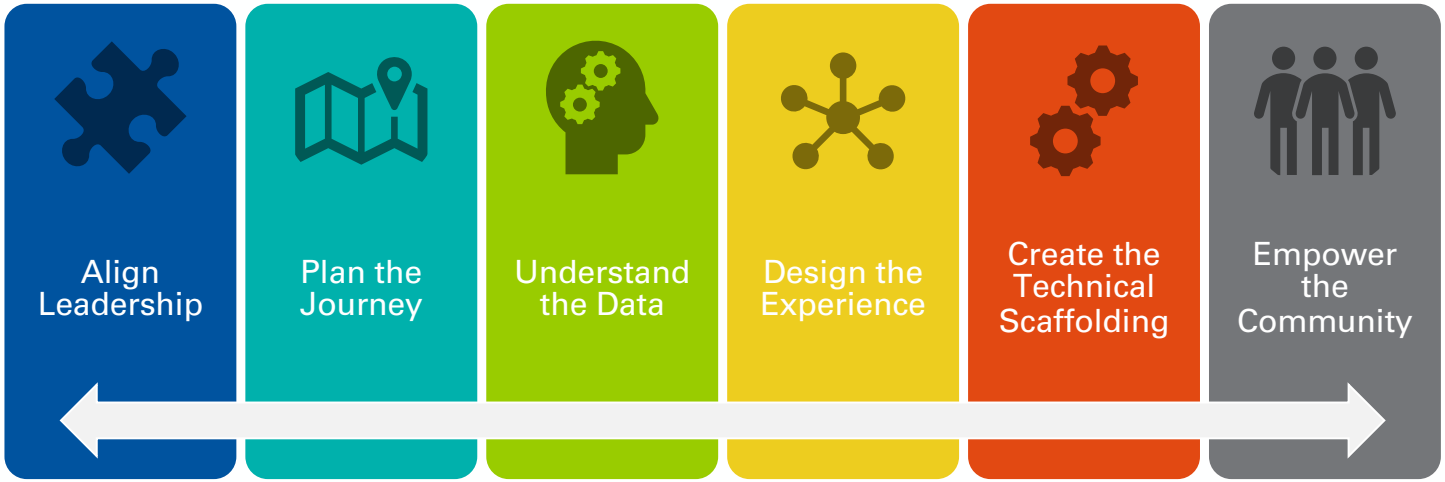


UC San Diego | AVIASS.co

Source: HolonIQ Global Executive Panel, April 2019, n=334

ASUG

The foundation for this transformation is exemplified in six key capabilities.

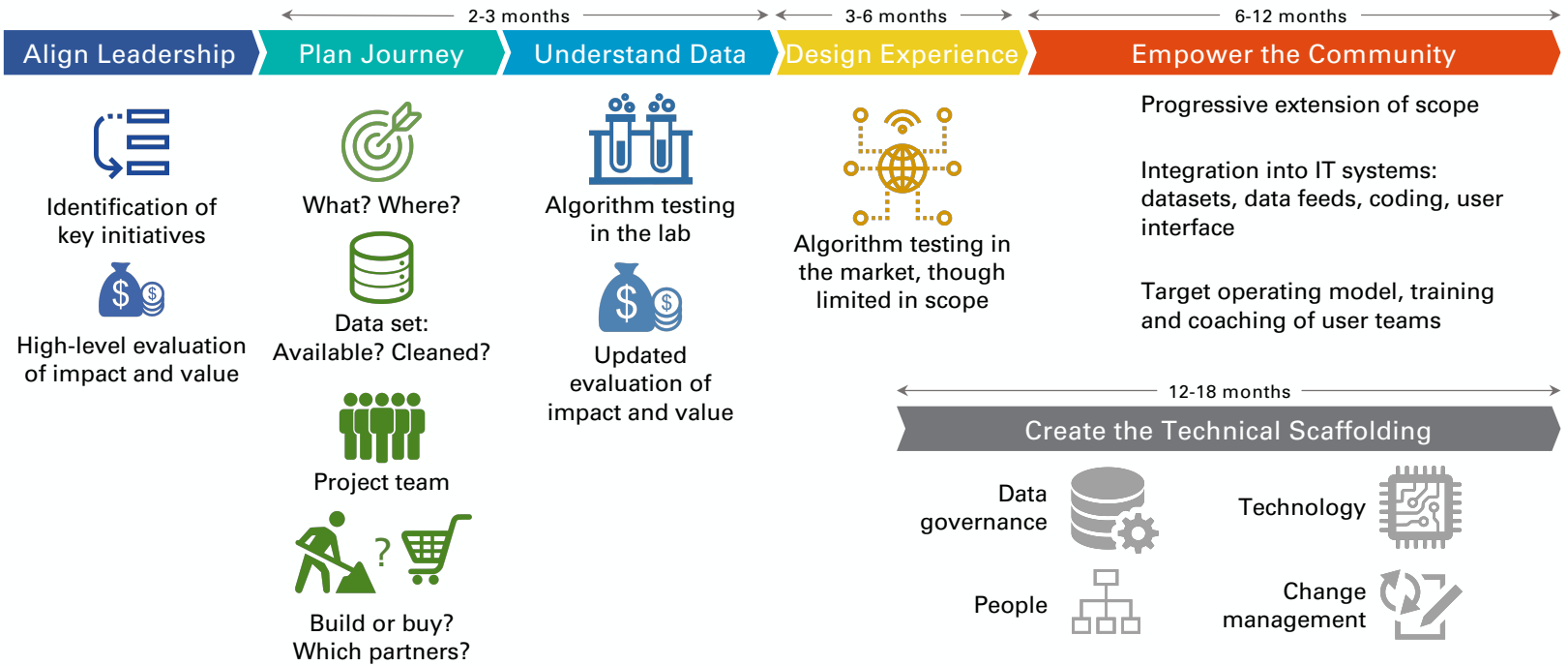


UC San Diego | AMASS.co

Source: Amass

ASUG

But it's possible to start now.



Source: BCG analysis & Amass

It's not about rich; it's about light.

- Users should be able to decide how they want to use data to meet their needs
- The technology should be easy to replace
- Your solution needs to be bespoke: balance re-inventing the wheel with following the herd

Summary

- AI is:
  - important and challenging
  - changing how you collect, store, and use data
  - critical

- UC San Diego has taken an aggressive approach to change our thinking, and adopt new processes to prepare for this change

- Accomplishing the same is possible, but may be hindered by strategy, leadership, and resources

Thanks

Take the Session Survey.

We want to hear from you! Be sure to complete the session evaluation on the SAPPHIRE NOW and ASUG Annual Conference mobile app.

UC San Diego | AVIASS.co

ASUG

Presentation Materials

Access the slides from 2019 ASUG Annual Conference here:
http://info.asug.com/2019-ac-slides

Q&A

For questions after this session, contact me at
amin@amass.co
or
amqazi@ucsd.edu.

Let's Be Social.

Stay connected. Share your SAP experiences anytime, anywhere.
Join the ASUG conversation on social media: **@ASUG365 #ASUG**