



# Accelerate and Scale Cross-Border Threat Assessment Analytics

## Delivery

4/28/2021

# Agenda

- Texas A&M Strategic Goals
- Future State & Usage Scenarios
- Dataiku Value and Example
- POC Process Overview
- Budgetary Pricing
- Dataiku Academic Programs

# Strategic Goals

Deliver strategic risk awareness across borders (US, Mexico, Canada, Central America) with increasingly complex and changing sources

Adopt a data science platform that will provide better collaboration across several joint research and development centers

Drive time to value and scale projects with small team to delivery 3-5X the number of projects per year

Enable every team member to contribute, leveraging their own skill sets while maintaining full transparency and collaboration

<https://www.whitehouse.gov/briefing-room/statements-releases/2021/02/24/fact-sheet-securing-americas-critical-supply-chains/>



data  
iku

# Current State Challenges

Several research teams across US and Mexico each model and deliver in separate environments and tools without collaboration or easy transparency. Team hitting resource scale bottleneck and struggling with inclusivity of subject matter experts to ensure model understanding, quality improvements and liability management (through auditability).

## Time to value and complexity management challenges:

- Each team uses their own tools and code libraries
- Diverse teams spread across North America
- Lack of collaboration or easy approval from subject matter experts (SMEs)
- Low explainability, consensus or auditability creates unknown liabilities
- Lengthy time from inception to operationalized data and model products
- COVID creating rapid changes and threats that must be nimbly addressed

Driving for deep collaboration, speed to value and continual improvement. Must compete for regular grants and projects; need to scale quickly past just 10 data flows and lower post production maintenance to bid for & deliver additional projects

## Environment & Environment Flux:

- On-prem 15 node unkerberized HDP cluster
- On-prem HPC cluster
- AWS reporting and output product delivery
- Low transparency use of R & python
- SSO

# Future State

**Rapidly deliver insights and new models with diverse group of remote research teams.**

**Inclusive Data Science culture with Subject Matter Experts interpreting and auditing models for external audiences. Scale number and delivery speed of models to extend to additional countries and threat areas.**

**Create transformative culture of collaboration that improves insights, models and time to value.**

## Remote collaborative participation:

- Quickly prototype and deploy data science projects
- Enable tools, libraries and languages of choice with transparency and appropriate IT controls
- 10+ Data Scientists & External Application Audience
- EDA, Explainability & Traceability are needed
- Collaboration and cross-team approvals critical
- Reuse of models and flows
- Improved predictive capabilities

## Use cases include :

- Supply-chain early warnings (gas shortage)
- Reuters threat awareness (NLP)
- Panjiva commercial import/export forecasts & warnings

Support and abstract the technical complexity of a wide range of technologies including HDP, python, R with high explainability and documentation that enables audit and limits liability.

## Frictionless data science with strong CI/CD/MLOPs:

- Versioning (incl. Git), audit trail, RBAC
- Seamless deployment process
- Continual monitoring and improvement lifecycle
- Batch & Real-time
- Seamless insights and analysis output via elastic AWS capabilities

# A few soundbites from our customers

## At the use case level

- **77% accuracy** in detecting potential churners - *Showroomprivé, E-commerce*
- **+ 11% efficiency** of anti churn outbound call campaigns - *Coyote, IOT*
- **Lift 5X better** on propensity model for newspaper subscription - *Publishing company*
- Number of sales call **from 10 to 4** to identify a sales opportunity - *Manufacturing*
- JTI won the European Association of Corporate Treasured Award. **Accuracy increase by 20% and workload reduced by 80%**

## At the analytical output level ...

- Big initiative (+2000 users - '00 projects - SC to sales & marketing) - **Several 100 millions \$ of benefits** - *GE Aviation*
- Customer DS Team **is 5x more efficient** in developing data projects from design to production with Dataiku - *Aviva, Insurance*
- It took us **3 hours instead of 3 months** to deploy the model - *Buildertrend, Residential Construction*
- Time to production was **reduced by a factor of 3** - *Marketing Services company*



# Recent Customer Business Value Example

## Strategic Goals



**Improve productivity**



**Streamline + govern projects**



**Produce better results + explainability**

## Key Solution Capabilities

Exploratory data analysis and AutoML help identify most predictive variables, etc.

100+ pre-built data prep and transformation

Embedded collaborative documentation: wiki, to-do list, tags, comments, discussions

Duplicate projects in whole or part for immediate reusability. Re-use existing models

Schedule and monitor scenarios including metrics, checks

Govern projects including versioning and audit trail, multi-persona reviews, and explainability

Build/Use plugins to extend Dataiku DSS capabilities if needed

## Business Impact

**25% productivity increase** on average for DSS users (low end of customer estimates)

**10% avg better results** (i.e. 10% higher project value)

**Delivery of projects previously not economically viable**

**2.5x increase in project throughput** (low end of customer estimates)

**Lowered risk through governance**

## Joint Value Estimation

**\$11.5M/y productivity gain**  
Assumes 150 users at \$300k fully burdened cost

**~\$500k/y incremental value**  
On ~50 projects per year x \$100k avg value

**\$5M/y unlocked opportunity cost**  
The value of "next projects" that otherwise go undone. Assumes incremental 50 projects a year and \$100k project value.

**\$TBD lowered compliance and reputational risk**

# POV Process

Scope, Success Criteria &  
Participants





# Success Criteria

- Vendor with strong expertise and leadership in the data science platform space
- Supports all stages of DS project (data prep, modeling, testing, deployment, monitoring)
- Allows ML/AI models to be operationalized and deployed with ease
- EDA, AutoML, collaboration, explainability, and code-your-way options
- Easy deployment of batch, API's and Analytic Apps

## Expected benefits / Success Criteria

- Efficiency - Deploy models faster with more business inclusivity
  - Easy reuse
  - Support for multiple languages and skills with traceability & control
  - AutoML for productivity and explainability
  - Automate deployment and monitoring (MLOps)
- Multi-Persona Collaboration
  - Expand ML/AI impacts
  - Lower risk by bringing more points of views and areas of expertise
  - Increase number of projects and quality
  - Create 'continuous improvement' ML lifecycle

# Roles & Persona(s) in POV

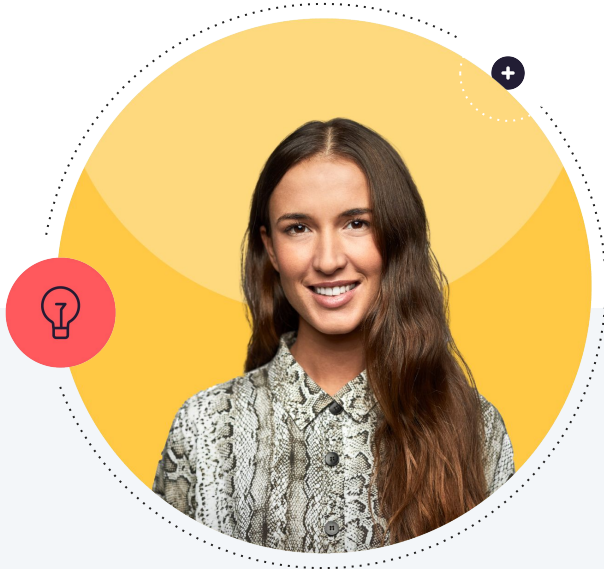
Name	Title	Function / Keys to Success
Zenon Medina	Dr	PI, U.S. – Mexico R-13 Project Associate Professor, Zachry Career Development Professor II, Civil & Environmental, Petroleum, Ocean & Geography, Texas A&M University
Guillermo Duran-Sierra	PhD student	SGL Team Lead, U.S. Mexico R-13 Project
??		<b>Dr Medina will identify 6 - 8 participants across disciplines to participate</b>
??		

**Method: Identify 1 to 2 high value use cases and success criteria; estimate time taken before Dataiku and with Dataiku to quantify speed and process improvements.**

# Dataiku Users & Budgetary Pricing



# Dataiku User Types



**Designer**

for low-code Citizens  
& pro-code Engineers



**Explorer**

for no-code Analysts &  
subject matter experts

# Dataiku Features by User Type




### Designers

Retrieve and process data, build and deploy models

### Explorers

Build and share dashboards, interact and collaborate with teams



**Features**

User licenses are named licenses, they can be reassigned if an employee leaves the company. Individual user rights and functionality may be configured by the administrator.

	Designers	Explorers
<b>Collaboration</b>	✓	✓
<b>Use Applications</b>	✓	✓
<b>Create and Read Insights &amp; Dashboards</b>	✓	✓
<b>Visual Data Preparation</b>	✓	—
<b>Visual Recipes &amp; SQL Notebooks (including Hive / Impala)</b>	✓	—
<b>Coding Recipes &amp; Advanced Notebooks (including R / Python / Scala)</b>	✓	—
<b>Machine Learning Wizard &amp; Recipes</b>	✓	—

# Business Edition



Business Edition
\$111,000/year
Research-Industry Partnership \$66,000 (40% Discount)



Users	
Designers (20 max)	Explorers
\$49,800/year for 10 users (\$4,980/user/year)	\$9,960 / year for 20 (\$498/user/year)
\$39,840/year (20% Discount)	\$0/year up to 50 (100% Discount)
Including <b>One</b> discovery training session (for 10 users)	Up to 8 Days Professional Services \$0

**BUDGETARY PRICING ONLY**

Total Annual License
\$170,760/year
Research-Industry Partnership Rate <b>\$106,440.00</b>
Plus Teaching Only Usage Free

## Potential Add-Ons:

- External Facing API Capacity
- + \$33,00.00 (40% Discount)
- + Enterprise Edition
- + Unlimited Design, Automation & API Nodes

# Dataiku Business

## Package Overview



Features

### Dataiku Custom Business Edition

1 Design Node

1 Automation Node

1 Deployer Node

5 maximum API services

20 maximum number of users

**Deployment Type:** Custom, install, and deploy Dataiku from installation package

**Available on:** Any customer environment (on-premises, in customer's cloud, or hybrid)

**Customer Success Manager** to support optimal value delivery

**Delivery of initial implementation plan** developed by a dedicated Implementation Manager (up to 8 days of professional services)

\$111,000 / year



Setup and onboarding  
services



# Dataiku Business

## Implementation Services included with Initial Subscription



Setup and onboarding services

### Up to 8 days of Implementation Services

Assisted **installation** for standard setup

Assisted **operationalization** and **deployment** of first DSS projects

**Coaching** hours for **core user** group on all aspects of Dataiku

Global implementation **coordination**

Up to 1 **onsite** visit



**Dataiku Implementation Manager**  
to define implementation plan  
with content and timeline

Additional services (extra time, project delivery)  
to be covered in a separate, dedicated SOW







### "Business" support

Error Level	P1	P2	P3	P4
<b>Description</b>	Critical error. "System is down" or software is completely inoperative, or a condition severely and significantly impacting the API nodes, design nodes, or automation nodes. No procedural work around exists. For example, all API nodes crash, or the design or automation node crashes and the software does not restart after the crashes.	Serious or high-impact business condition affecting substantial number of users. Involves the software operating in a severely restricted/reduced capacity causing significant impact to portions of the customer's business. No procedural work around exists.	Medium to low-impact problem. Involves partial non-critical functionality loss. May be a minor issue with limited or no loss of functionality or impact to the Customer's operation for which there is a workaround available. For example, a data transformation function does not work as expected, and a workaround in code is required.	Software is functional. Relates only to proposed feature enhancements or proposed modifications. No material impact on quality, performance or functionality of Software. For example, errors in documentation, or a proposed feature enhancement.
<b>Target Initial Response</b>	8 Business Hours	2 Business Days	2 Business Days	2 Business Days
<b>Target Update Frequency</b>	Every 4 hours Escalation in 72h to VP of Support	Every Business Day	As needed	As needed

\***Support Hours** are 9am - 5pm local time on any business day in New York, Paris or Singapore, as determined by Dataiku based on the Customer's location.



# Much more than software

## From Usage to Value

Services available to all users



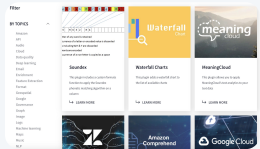
### Academy.dataiku.com

On-demand training and certification



### Community.dataiku.com

Peer-to-peer support and best practices sharing



### Online Repository

Free plugins extending Dataiku DSS features

## Supporting Company Success

Services included in Subscription



### Customer Success

Named orchestrator of your success with Enterprise AI



### Implementation services

Architecture, deployment, set-up and training

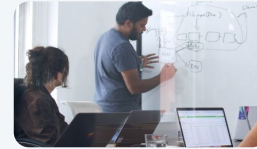


### Customer support

Quick answers through the Dataiku support portal

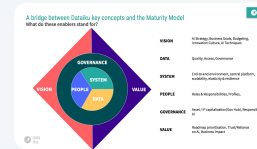
## Transforming your Organization

Ad-Hoc Engagements



### Data Science Services

Hands-on assistance for your projects



### Strategic Advisory

AI Maturity Assessments and Acceleration of your organizational transformation through best practices and tailored recommendations



### Business Value Consulting

Measuring and driving the value of your AI initiatives

# About Dataiku Academic Programs



# Data Science Studio

---

## Research/Industry Partnerships

Research/Industry Partnerships  
Leverage Dataiku Enterprise Edition to propel research & development opportunities with your industry partners.

**Contact us to learn more about purchasing options.**

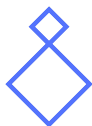
## Free Enterprise Edition for Teaching and Learning

Comes with full collaboration features. Our most powerful form of DSS to give your students a real-world experience.

On-campus server install or Dataiku host options available. **Contact us to learn more.**

# Teaching Resources

---



## Demos & workshops

Attend live academic webinars or contact our team to organize customized sessions.



## Online learning & certifications

Explore the Dataiku Academy course catalogue and learning paths that lead to certifications.



## Dataiku community

Bringing together DSS users and data science professionals for learning, peer-to-peer support, and best-practice sharing.



## Data science content

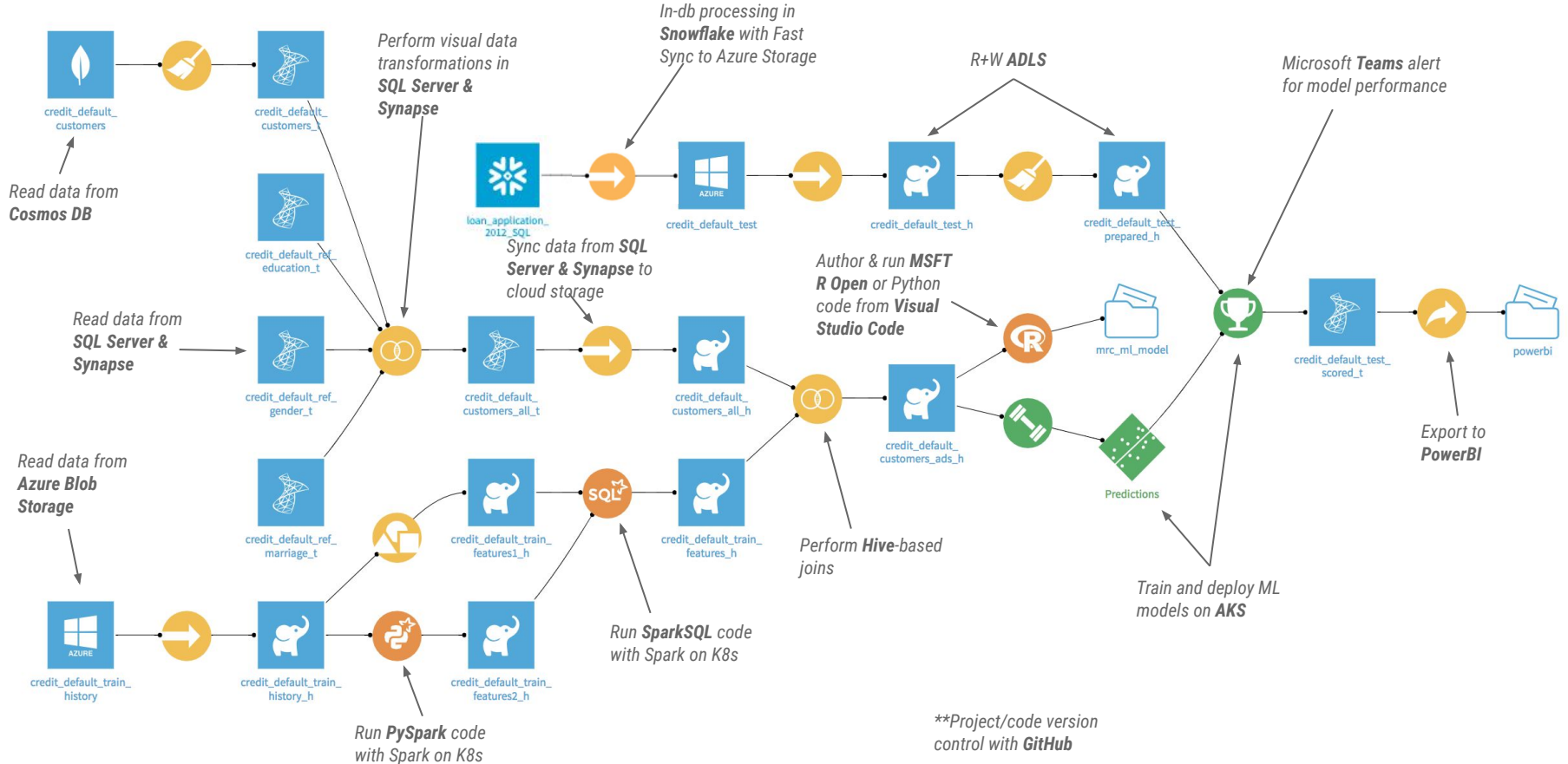
Explore a wealth of data science content, from industry and DSS webinars to white papers, blogs, podcasts, use cases, etc.

# Thank You!

---

# Example Dataiku Flow on Azure

## Pro-Code and Low-Code users easily collaborate in a single project



# Price Structure

Description		Amount
Core Annual Pricing		Chosen Package Price
Plus Users	• Designer Users	+ \$4,980.00/user
	• Explorer Users	+ \$498.00/user
		Annual Price
Possible Add-ons	• External Facing API Capacity (for Enterprise and Business editions)	+ \$55,200.00
	• Additional Continent (for Enterprise editions)	+ \$74,400.00
		Final Annual Price