

SDSC Summer Institute 2020

Title: CloudBank: Managed Services to Simplify Cloud Access for Computer Science Research and Education

Instructors: Shava Smallen

Date: August 7, 2020

Location: Main Room Session

Gitter (session support): Main Session

Gitter (general system support): Help Desk





Managed Services to Simplify Cloud Access for
Computer Science Research and Education

Summer Institute 2020
August 7, 2020

UC San Diego

Mike Norman, PI
Shava Smallen, Co-PI
Vince Kellen, Co-PI

SDSC SAN DIEGO SUPERCOMPUTER CENTER

W
UNIVERSITY of
WASHINGTON
Ed Lazowska, Co-PI

Berkeley
UNIVERSITY OF CALIFORNIA

David Culler, Co-PI



NSF Award: 1925001
Deep Medhi, PO



UC San Diego

CloudBank project team

- UC San Diego/SDSC: Mike Norman (PI), Vince Kellen (Co-PI), Shava Smallen (Co-PI), Brian DeMeulle, Declan Fleming, Paul Hoover, Kate Kaya, Nick Marangella, Sonia Nayak, Amin Qazi, Jessica Respicio, Shawn Strande, Mary Thomas, Nicole Wolter, Choonhan Youn, Frank Wuerthwein
- University of Washington: Ed Lazowska (Co-PI), Rob Fatland, Sarah Stone, Amanda Tan
- UC Berkeley: David Culler (Co-PI), James Percy
- Strategic Blue: James Mitchell, Jon Bryant, Frank Contrepois, Steve Old, Emma Sampayo, Hitesh Shah, James Smith
- External Advisory Board: Dana Brunson (Chair), Maytal Dahan, Dennis Gannon, Tracy Camp, Tracy Futhey, Michael Franklin, Jason Liu
- Cloud providers: AWS, GCP, Azure. IBM coming early 2021.
- National Science Foundation: Deep Medhi

CloudBank 101

- 5-year **pilot** project, funded by via a \$5M NSF Cooperative Agreement
- Goal: Make it *easier for CISE researchers and educators* to use the public cloud

The project is organized into 3 main efforts



User Portal
UCSD/SDSC



Education & Training
UW & UCB



Financial Management
UCSD/ITS & Strategic Blue

CloudBank is focused initially on specific NSF CISE solicitations

- **IMPORTANT SCOPE:** Only specific NSF solicitations are eligible for cloud resources through CloudBank. The solicitation will state eligibility explicitly.
- There have been 4 such solicitations so far:
 - Smart and Connected Communities (Closed), Cyber-Physical Systems
 - **New:** Computer and Information Science and Engineering: Core Programs, Cyberinfrastructure for Sustained Scientific Innovation: Elements and Framework Implementation
- PIs request cloud funding. Without CloudBank, institutions would add cloud funding directly to their budget and many would pay indirect costs.
- The total of your formal budget request and the cloud request must not exceed the total allowable request for the solicitation.
- Cloud resources are free of indirect costs. I.e., are treated like equipment

CloudBank seeks to address pain points of researchers and educators using public cloud

"Setting up your first experiment in the cloud is too time-consuming, and different for each cloud platform."

"I need documentation specific to my research."

"Provisioning all students in a class with their own cloud credits and dealing with their cost overruns is a major pain. Financial losses can be substantial if students are given free reign."

"There is no simple way to run multiple experiments with software already installed in one instance."

"Research credits are difficult for us to keep track of."

"We need billing help. It's opaque as to what is going on and I don't have the time to wade through this."

"Cloud is just not designed for typical use case of a small research group."

"We need an expedited process for setting up experiments very quickly. Often we use cloud resources when we're on a deadline so quick setup is critical."

"I just don't have enough time to learn all this. Can you give me a jump start?"

"It's a Hobson's choice between giving everyone root access or losing my life in fine-tuning access."

"Cloud is a serious learning curve. Just too much documentation, and of the wrong type. Students get lost."

CloudBank seeks to address pain points of researchers and educators using public cloud

“Setting up your first experiment in the cloud is too time-consuming, and different for each cloud platform.”

“I need documentation specific to my research.”

“Provisioning all students in a class with their own cloud credits and dealing with their cost overruns is a major pain. Financial losses can be substantial if students are given free reign.”

“There is no simple way to run multiple experiments with software already installed in one instance.

“Research credits are difficult for us to keep track of.”

“We need billing help. It’s opaque as to what is going on and I don’t have the time to wade through this.”

“Cloud is just not designed for typical use case of a small research group.”

“We need an expedited process for setting up experiments very quickly. Often we use cloud resources when we’re on a deadline so quick setup is critical.”

“I just don’t have enough time to learn all this. Can you give me a jump start?”

“It’s a Hobson’s choice between giving everyone root access or losing my life in fine-tuning access.”

“Cloud is a serious learning curve. Just too much documentation, and of the wrong type. Students get lost.”

CloudBank seeks to address pain points of researchers and educators using public cloud

“Setting up your first experiment in the cloud is too time-consuming, and different for each cloud platform.”

“I need documentation specific to my research.”

“Provisioning all students in a class with their own cloud credits and dealing with their cost overruns is a major pain. Financial losses can be substantial if students are given free reign.”

“There is no simple way to run multiple experiments with software already installed in one instance.

“Research credits are difficult for us to keep track of.”

“We need billing help. It’s opaque as to what is going on and I don’t have the time to wade through this.”

“Cloud is just not designed for typical use case of a small research group.”

“We need an expedited process for setting up experiments very quickly. Often we use cloud resources when we’re on a deadline so quick setup is critical.”

“I just don’t have enough time to learn all this. Can you give me a jump start?”

“It’s a Hobson’s choice between giving everyone root access or losing my life in fine-tuning access.”

“Cloud is a serious learning curve. Just too much documentation, and of the wrong type. Students get lost.”

CloudBank seeks to address pain points of researchers and educators using public cloud

“Setting up your first experiment in the cloud is too **time-consuming**, and different for each cloud platform.”

“I need **documentation** specific to my research.”

“Provisioning all students in a class with their own cloud credits and dealing with their cost overruns is a **major pain**. Financial losses can be substantial if students are given free reign.”

“There is **no simple way** to run multiple experiments with software already installed in one instance.

“Research credits are **difficult** for us to keep track of.”

“We need billing help. It’s opaque as to what is going on and I don’t have the time to wade through this.”

“Cloud is just **not designed** for typical use case of a small research group.”

“We need an expedited process for setting up experiments very quickly. Often we use cloud resources when we’re on a deadline so **quick setup is critical**.”

“I just don’t have enough **time** to learn all this. Can you give me a jump start?”

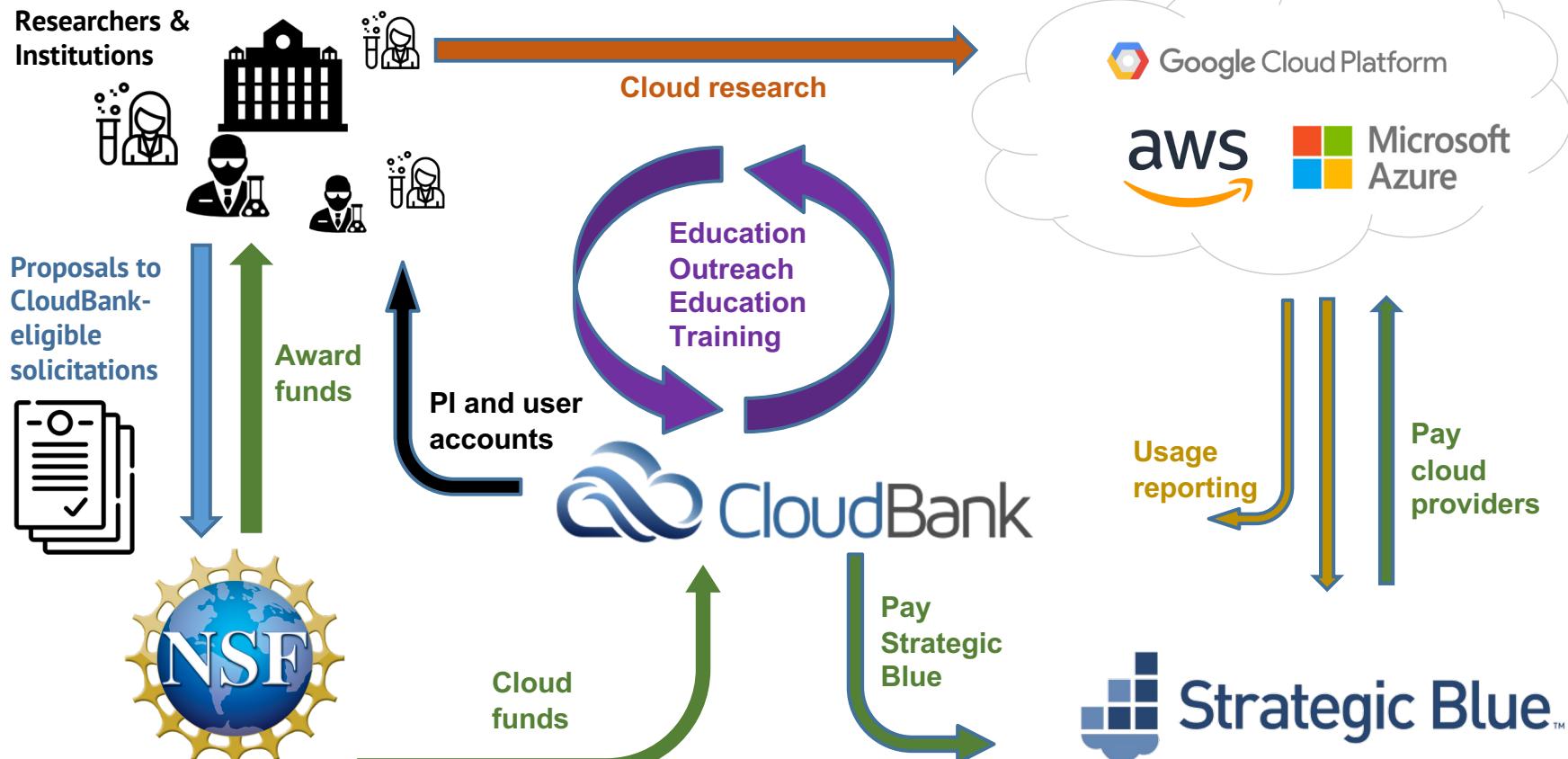
“It’s a Hobson’s choice between giving everyone root access or **losing my life** in fine-tuning access.”

“Cloud is a serious **learning curve**. Just too much documentation, and of the **wrong type**. Students get lost.

CloudBank provides managed services to address these pain points

- Cloud solution consulting, proposal assistance, cloud training
- Onboarding, account setup, user management, off-boarding
- Tools for PIs to grant permissions to research group members and students
- Unified reporting across cloud providers
- Bundling multiple small requests into a bulk request to cloud providers, providing leverage to secure discounts
- Pass along savings to researchers

CloudBank workflow



SDSC SAN DIEGO
SUPERCOMPUTER CENTER

CloudBank

UC San Diego

You request cloud resources via a Supplementary document in your NSF proposal

From the d

“If requeste

your request

(1)title of t

(2)the tota

(3)which p

(4)a techn

with how the cost was estimated.

That sounds easy, right?

But there are hundreds of
offerings across the providers
and figuring out what to use for
your research can be confusing!

CloudBank can help you understand the public cloud, with training modules, if you still need help, the CloudBank support team can help.

help@cloudbank.org

And if we can't help, there are options:

- You may have people from providers that work with your campus or we can work with cloud providers on your behalf
- CloudBank Community Forum
- Groups like Campus Champions are a great resource

Once you figure out what you want to put in your proposal, you will use the public cloud calculators to determine your costs

You can request more than 1 provider in your proposal.

If you have room in your supplementary document, consider adding a screen shot of the calculator output.

Cloud offerings evolve rapidly, and you can switch to a different provider after your award.

CloudBank for Instruction

<https://www.cloudbank.org/training/cloudbank-community>

The screenshot displays three windows illustrating the CloudBank for Instruction platform:

- Left Window (Web Browser):** Shows the Data Science 8 course page (www.data8.org). The page title is "Data 8: The Foundations of Data Science". It includes sections for "Offerings" (listing Summer 2017 through Fall 2015), "Materials" (listing textbook, assignments, and lecture materials), and a "Calendar" showing weekly sessions from Wed 1/18 to Wed 2/8.
- Middle Window (GitHub):** Shows the "Data Science 8" organization page on GitHub. It lists pinned repositories: "datascience" (A Python library for introductory data science), "textbook" (The textbook Thinking: The Fundamentals of Data Science), "jupyterhub-k8s" (Data 8's deployment of JupyterHub on Kubernetes), and "jupyterhub" (JupyterHub configuration). The GitHub interface shows pull requests, issues, and marketplace items.
- Right Window (Jupyter Notebook):** Shows a Jupyter Notebook session titled "Census" with the last checkpoint at 8:18 AM. The notebook cell toolbar is set to "None". The code cell contains:

```
In [1]: # HIDDEN
from datascience import *
import matplotlib-inline
import matplotlib.pyplot as plt
plt.style.use('fivethirtyeight')
```

The output cell shows a table of census data:

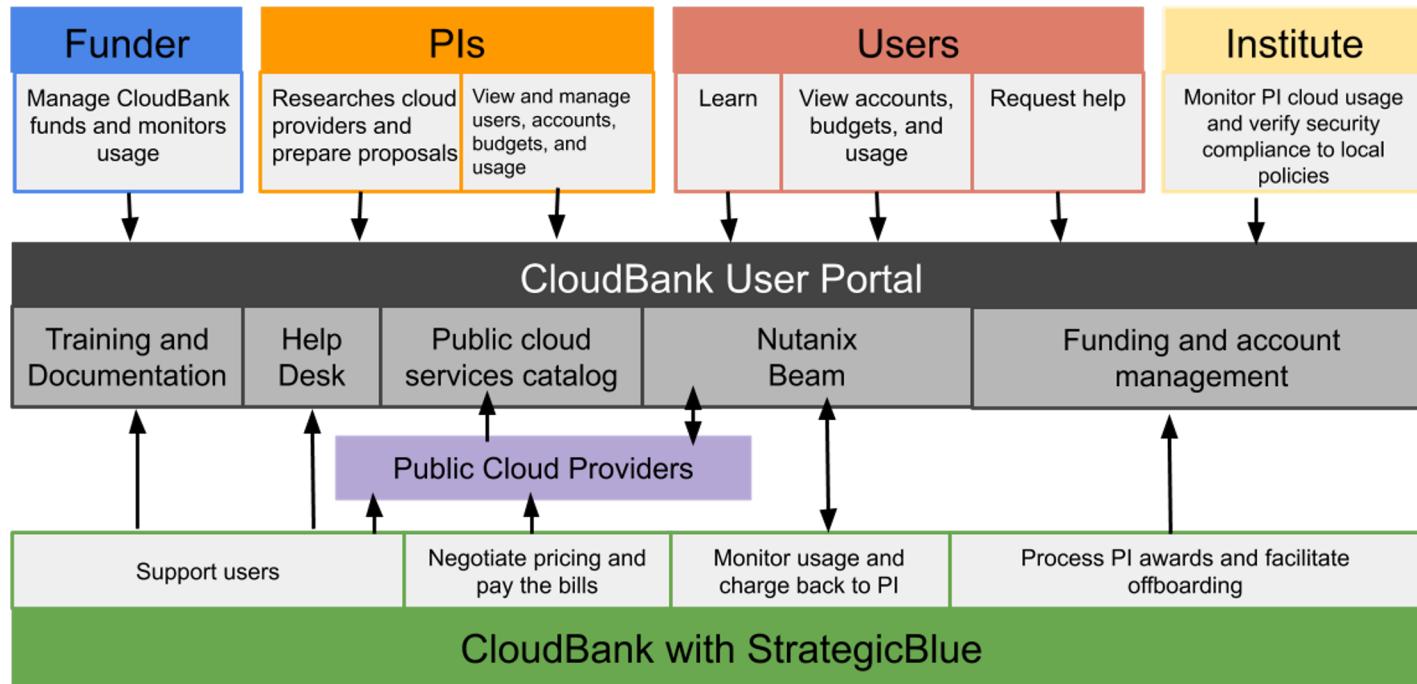
SEX	AGE	CENSUS2010POP	ESTIMATESBASE2010	POPESTIMATE2010	POPESTIMATE2011	POPESTIMATE2012	POPESTIMATE2013
0	0	3944153	3944160	3961330	3963071	3966665	3945610
0	1	3978070	3978090	3957888	3966510	3978008	3943077
0	2	4096829	4096939	4090662	3971573	3979952	3992690
0	3	4119640	4119651	4111920	4102501	3983049	3992425
0	4	4063170	4063186	4077552	4122303	4112638	3994547
0	5	4056858	4056872	4064663	4087713	4132210	4123408
0	6	4063631	4066412	4073013	4074979	4097780	4143094
0	7	4030579	4030594	4043047	4083240	4084954	4108615
0	8	4046486	4046497	4029664	4093206	4093212	4095827
0	9	4148353	4148369	4125415	4035769	4063193	4104133

... (296 rows omitted)

```
In [2]: def categorize_sex(s):
    return ['all','male','female'][s]
```

```
In [3]: pre_census = raw_census.select(['SEX', 'AGE', 'CENSUS2010POP', 'POPESTIMATE2014'])
pre_census.rename(columns={'CENSUS2010POP': '2010pop'})
pre_census.rename(columns={'POPESTIMATE2014': '2014est'})
pre_census['CAT'] = pre_census.apply(categorize_sex, 'SEX')
p2_census = pre_census.drop(['SEX'])
p2_census.move_to_start(['CAT'])
```

CloudBank User Portal



Drupal™

Beam

CI Logon

zendesk

**UC Berkeley
Data Stack**

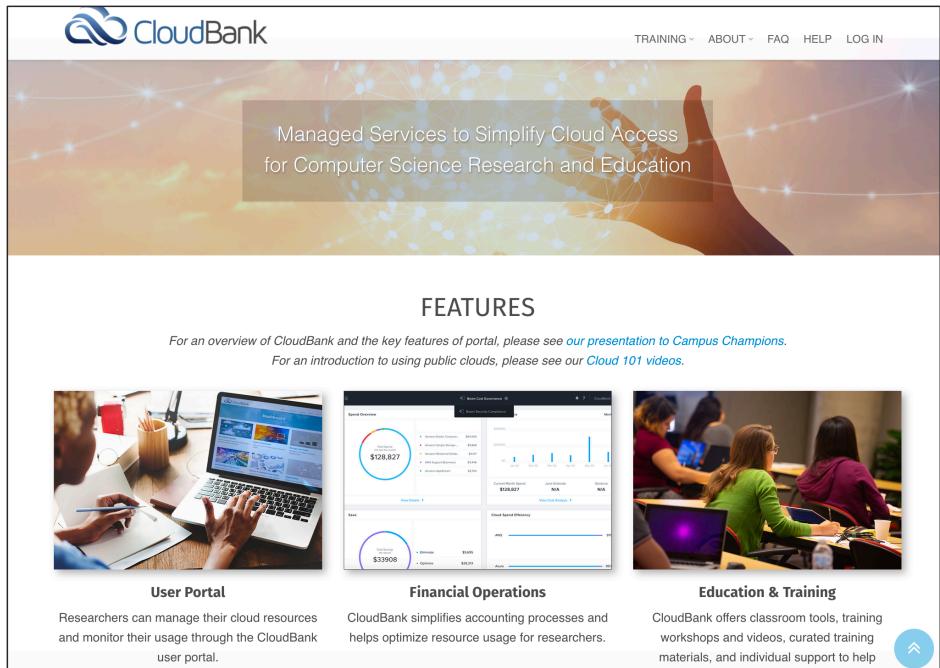
SDSC SAN DIEGO SUPERCOMPUTER CENTER

CloudBank

UC San Diego

Now in production as of August 1, 2020

- Continuing portal and training development
- Continuing early testing – if you are interested in helping out, please contact help@cloudbank.org
- Expecting first production users in September
- September Webinar for PIs preparing proposals for CORE and CSSI



The image shows the CloudBank website homepage. At the top, there is a navigation bar with links for TRAINING, ABOUT, FAQ, HELP, and LOG IN. The main header features the CloudBank logo and the tagline "Managed Services to Simplify Cloud Access for Computer Science Research and Education". Below the header, there is a section titled "FEATURES" with a brief description and links to presentation and video resources. Three main features are highlighted with images: "User Portal" (a person using a laptop), "Financial Operations" (a screenshot of a dashboard showing financial data like \$128,827 and \$33908), and "Education & Training" (two people working at a computer). A blue circular arrow icon is located in the bottom right corner.

Thank you!

Shava Smallen (ssmallen@sdsc.edu)

<https://www.cloudbank.org>
help@cloudbank.org



Consent to Attribute Release



[CloudBank](#) requests access to the following information. If you do not approve this request, do not proceed.

- Your CILogon user identifier
- Your name
- Your email address
- Your username and affiliation from your identity provider

Select an Identity Provider

University of California, San Diego

Remember this selection

Log On

By selecting "Log On", you agree to [CILogon's privacy policy](#).



Google Cloud

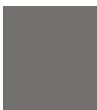
available in 2021

(content in progress)

[DASHBOARD](#) ▾ [TRAINING](#) ▾ [ABOUT](#) ▾ [FAQ](#) ▾ [LOG OUT](#)

Using the CloudBank Catalog: Major categories of services are highlighted below. Click on any category (e.g., Compute) and it will expand to show a table with types of services represented in the rows and public clouds in the columns. Hover over the row header in the leftmost column to see an expanded definition of the type of service.

Compute	
Storage	
Network	
Database Services	
Big Data & Analytics	
Machine Learning	
IoT	
Application Services	
Management Services	
Console & APIs	





DASHBOARD ▾ TRAINING ▾ ABOUT ▾ FAQ ▾ LOG OUT

Database Services

Big Data & Analytics

Big data is a combination of structured, semistructured and unstructured data collected by organizations that can be mined for information and used in machine learning projects, predictive modeling and other advanced analytics applications. [Source: [TechTarget](#)]. Big data analytics is the often complex process of examining large and varied data sets, or big data, to uncover information -- such as hidden patterns, unknown correlations, market trends and customer preferences -- that can help organizations make informed business decisions.[Source: [TechTarget](#)]

	Amazon Web Services	Google Compute Platform	Microsoft Azure	IBM Cloud
<i>Batch Data Processing</i>	Amazon Elastic MapReduce	Google DataProc , Google DataFlow	Azure Batch , HDInsight	
<i>Stream Data Processing</i>	Amazon Kinesis	Google DataFlow	Azure Stream Analytics	
<i>Stream Data Ingest</i>	Amazon Kinesis	Google Cloud Pub/Sub	Azure Event Hubs , Azure Service Bus	
<i>Analytics</i>	Amazon Redshift	Google BigQuery	Azure Data Lake Analytics , Azure Data Lake Store	
<i>Workflow Orchestration</i>	Amazon Data Pipeline , AWS Glue	Google Cloud Composer	Microsoft Flow , Azure Logic Apps	





DASHBOARD ▾ TRAINING ▾ ABOUT ▾ FAQ ▾ LOG OUT

Google BigQuery

HOME / TOOLS

“

A serverless, highly scalable, and cost-effective cloud data warehouse designed to help you make informed decisions quickly, so you can transform your business with ease. Accelerate time-to-value with a fully managed and serverless cloud data warehouse that is easy to set up and manage and doesn't require a database administrator. Jump-start data analysis cost effectively and uncover meaningful insights to stay competitive.



BigQuery

URL: <https://cloud.google.com/bigquery>

Vendor: Google Compute Platform

Categories:

Analytics

Available Training



DASHBOARD ▾ TRAINING ▾ ABOUT ▾ FAQ ▾ ADMIN ▾ SUPPORT LOG OUT

Manage Funds

Fund

An allocation of dollars to a PI and one or more Co-PIs for a specific project or CloudBank fund is created for each of your NSF awards that have a CloudBank

Spend checked hourly but depends on frequency of billing data from each Cloud

Email notifications sent at first 1%, each 10% balance spend, and daily once 20% remains.

for NSF awardees, a
e of the award.

You are a PI or Co-PI of the following 6 fund(s):

FUND ▾	FUNDER	TYPE	START	END	AWARD AMOUNT	BALANCE	PI	CO-PI	MANAGERS
April Early Test Fund 1	CloudBank	Discretionary	2020-04-20	2020-08-28	\$150.00	\$42.50 (28%)	Shava	Smallen	
May Early Test Fund 1	CloudBank	Discretionary	2020-04-20	2020-08-01	\$20.00	\$4.43 (22%)	Shava	Smallen	
May Early Test Fund 2	CloudBank		2020-05-08	2020-08-01	\$20.00	\$7.05 (35%)	Shava	Smallen	



DASHBOARD ▾ TRAINING ▾ ABOUT ▾ FAQ ▾ SUPPORT LOG OUT

ID: 2

Funder: Test
Discretionary
Start: 2020-02-19
End: 2025-07-01
Award Amount: \$50.00
Balance: \$46.59
Balance Percent: 93%
PI: User 1
Co-PI: User 2

User 3

Requested Clouds: Amazon Web Services
Google Cloud Platform
Microsoft Azure

Status

Active

Cloud offerings evolve rapidly, and you can try different providers after your award.

Linked Billing Accounts

Add a new billing account to this Fund.

Can spend award in multiple clouds

ID	STATUS	AWARD AMOUNT	TOTAL SPEND	BALANCE	PUBLIC CLOUD	EDIT
24xxxxxxxx	ACTIVE		\$0.01		Amazon Web Services	edit
24abcd-efgh-ijklmnop-123458	ACTIVE		\$0.24		Microsoft Azure	edit
xxxx-yyy-zzz	ACTIVE		\$3.12		Google Cloud Platform	edit
86xxxxxxxxxx	ACTIVE	\$50.00	\$0.04	\$49.96 (100%)	Amazon Web Services	edit

Click on any linked billing account to manage access

Microsoft Azure Search resources, services, and docs (G+) CloudBank user1@cloudbank.org CLOUDBANK

Home > CloudBank Test Subscription A Manage Cancel subscription Rename Change directory

Subscription ID: 24abcd-efgh-ijklmnop-123458
Subscription name: CloudBank Test Subsc
Directory: CloudBank (cloudbankorg.onmicrosoft.com)
Current billing period: 7/1/2020-7/31/2020
My role: Contributor
Currency: --
Offer: Enterprise Agreement
Offer ID: MS-AZR-0017P
Status: Active

Overview Activity log Access control (IAM) Tags Diagnose and solve problems Security Events

Cost Management

- Cost analysis
- Budgets
- Advisor recommendations

Billing

- Partner information

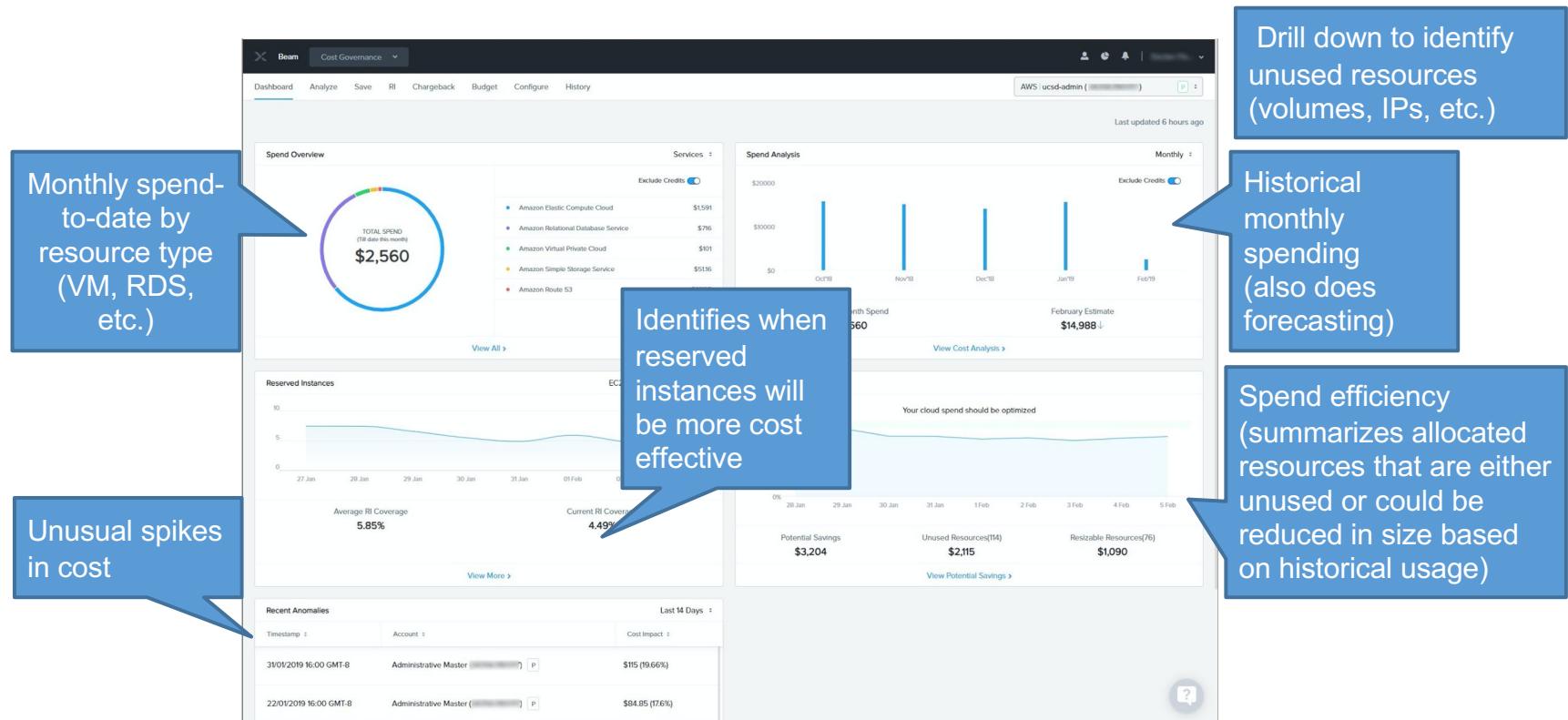
Settings

- Programmatic deployment
- Resource groups
- Resources

User 1 Sign out

user1@cloudbank.org View account Switch directory ...

Sign in with a different account



Beam

Nutanix Beam enables CloudBank to provide PI and user views across **multiple accounts** across **multiple cloud providers**

SDSC SAN DIEGO
SUPERCOMPUTER CENTER



UC San Diego



Beam

Provides views of instances and compliance to security policies.

Suggests fixes to get an instance into compliance

Out of the box security policies (250+ security checks and can also define your own)

The screenshot displays the Beam application interface across three main sections:

- Cloud Security Summary:** A world map showing the distribution of security issues. Total Issues: 262. High Severity: 118 failed. Medium Severity: 126 failed. Low Severity: 18 failed. Specific regions highlighted include US West (Oregon) with 124 issues, CA (Canada) with 8, US East (Ohio) with 12, US West (Portland) with 8, US East (New York) with 3, EU (London) with 8, EU Central (Frankfurt) with 8, Asia Pacific (Seoul) with 8, Asia Pacific (Mumbai) with 8, Asia Pacific (Singapore) with 8, and Asia Pacific (Beijing) with 8.
- Compliance Policies:** A table listing various compliance policies:

Name	Created By	Last Modified At	Description
DR Policy	Product	Nov 30, 2018, 150:21 AM	Policy with Disaster Recovery specific audits
HIPAA (v5010) Policy	Product	Nov 30, 2018, 150:21 AM	Automated Validation of HIPAA technical safeguards for AWS Services
Performance Policy	Product	Nov 30, 2018, 150:21 AM	Policy with Performance specific audits
Security Policy	Product	Nov 30, 2018, 150:21 AM	Policy with Security specific audits
Beam Compliance Policy	Product	Nov 30, 2018, 152:48 AM	Policy with all audits
PCI DSS (v3.0) Policy	Product	Nov 30, 2018, 150:21 AM	PCI DSS Audits
OIS (v0.0) Policy	Product	Nov 30, 2018, 150:21 AM	A set of security configuration best practices for AWS
- Regulatory Compliance View:** A detailed view of AWS CIS (v1.0.0) Policy compliance status. Total Checks: 93. Status: All. Type: All. Automated Check: All. 62.50% Compliant. The view is broken down into five categories: 1. Data Protection, 2. Identity and Access Management, 3. Business Continuity, 4. Event Monitoring and Response, and 5. Audit and Logging. Each category shows the total number of checks, passes, fails, and non-applicable items.

New Compliance view for all Regulatory policies and Standards. Can set Rule-based alerts to get notified on any specific audit non-compliance as and when they occur.