



ENTERPRISE-WIDE DEMOCRATIZATION OF AI ML

JUNE 2021

MEET THE TEAM



SAMMED KUMAR

TECHNICAL PRODUCT MANAGER

Sammed is a technical product manager at D Cube Analytics, has 15+ years of Industry experience in Data and Analytics, Data governance and syndication. Sammed has managed various initiatives in the areas of building Pharma MDM platforms, Data Warehouse, and advanced personalization.



AJAY MEDA

ASSOCIATE PRODUCT ARCHITECT

Ajay is an Associate Product Architect at D Cube Analytics. He brings close to 7+ years of experience in Cloud – DevOps and product planning. He has experience in designing network and security for various applications. He has experience in designing Data Engineering and Pharma MDM platforms.



SAMUEL JAIDEEP

ASSOCIATE PRODUCT ARCHITECT

Samuel Jaideep is an Associate Product Architect at D Cube Analytics, has 7+ years of Industry experience in Web Technologies and Cloud Engineering. Jaideep has been involved in architecting and development of various Web Applications for multiple clients in the course of time.

OUR AGENDA TODAY

1. Introduction – Overview of democratization
2. The need (use cases and why democratization)
 - A. Industry patterns impacting enterprise adoption of AI
 - B. Capabilities needed to address the challenges
 - C. Enabling a democratization platform for a large biopharma – A case study
3. Solution components (Focus areas for implementing governance)
 - A. Tools and software
 - B. People and collaboration
 - C. Asset governance
4. Summary

Introduction – Overview of democratization



Digitization through AI/ML requires the following -

- **Technology and tools**
- **People and collaboration**
- **Data and other assets**

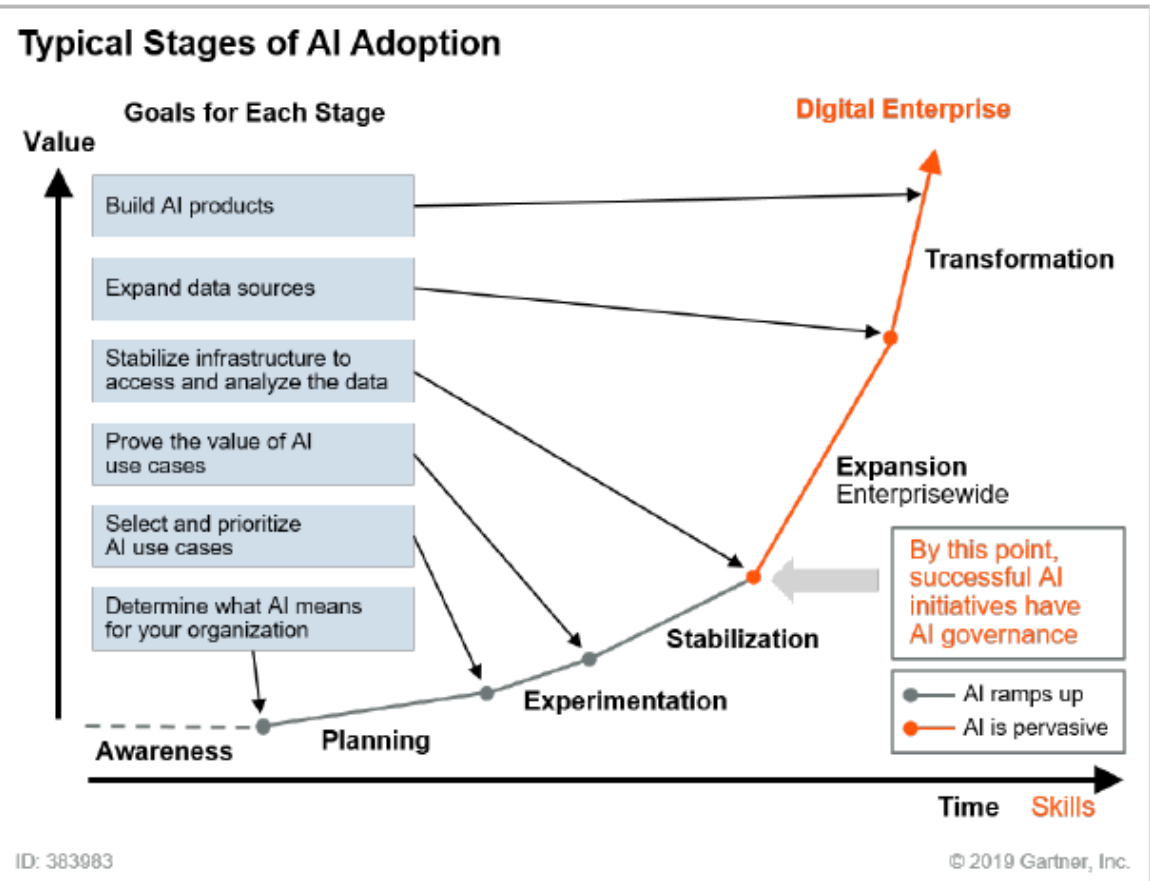
The intent of democratization is to

- Enable AI citizens to fully utilize the benefits of above components
- Enable AI governors to implement well-defined access policies and processes

Overall intent is to increase adoption and participation by each user persona providing

- **Enhanced visibility**
- **Reduced process friction**

Risks of having ungoverned ecosystem



Source: Gartner (April 2019)

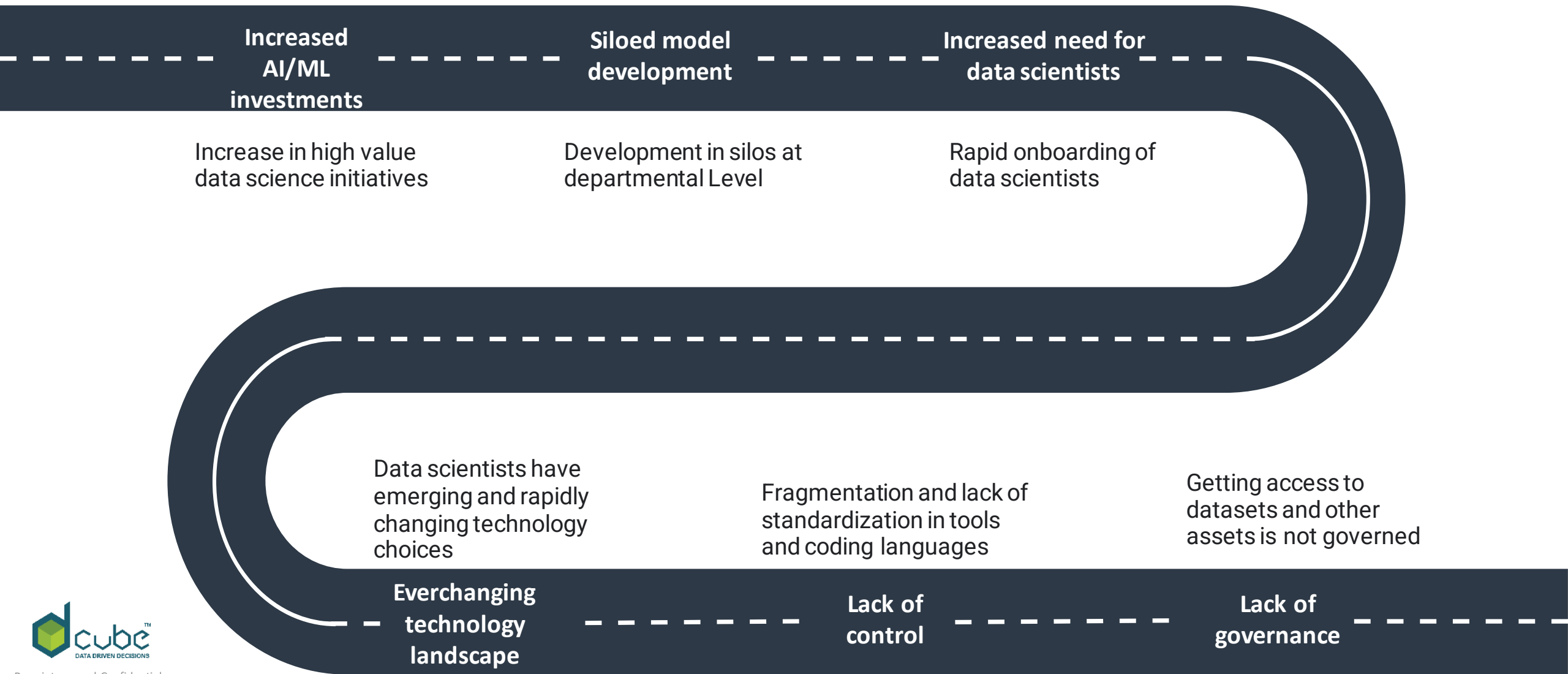


Proprietary and Confidential



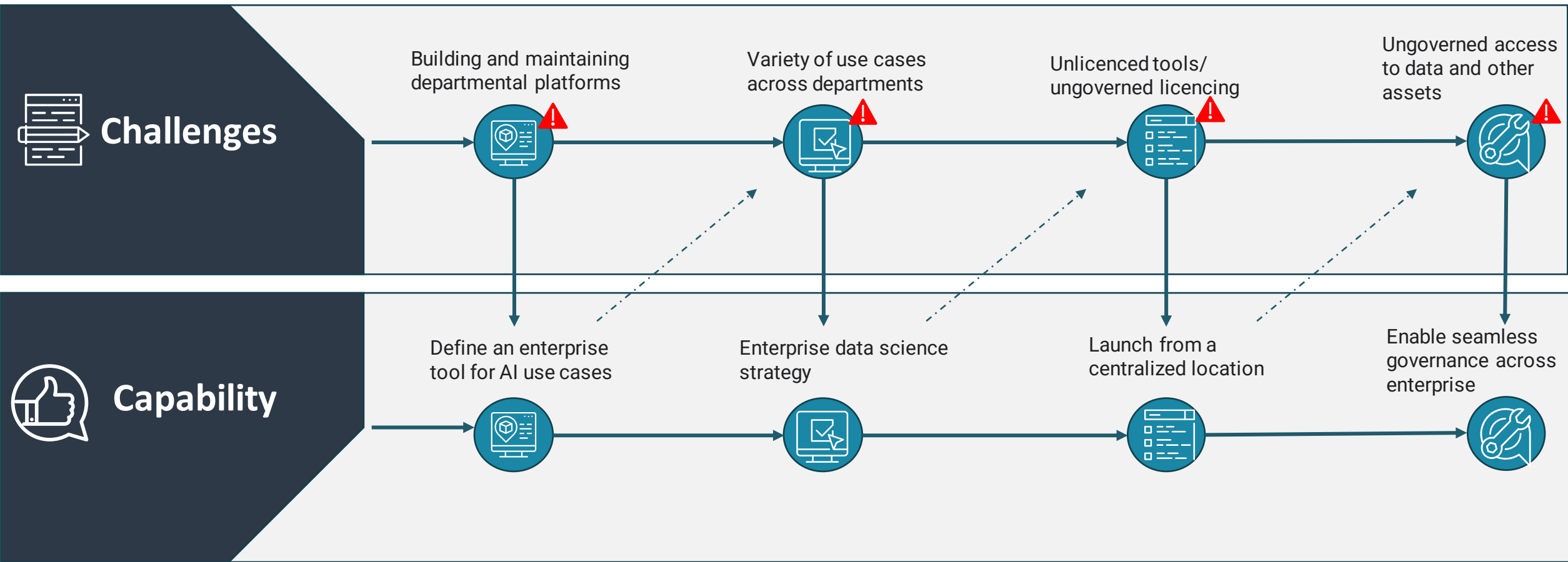
Industry patterns impacting enterprise adoption of AI

Below roadmap shows digitization journey of a typical mid-size/large biopharma organization using AI/ML.



Capability needed to address the challenges

Challenges faced by a mid-size/large biopharma organization trying to adopt enterprise governance.



Case study (1/3) – Enabled a unified data science platform built on cloud for a leading biopharma

Objective & Problem statement

- The Data Science leadership of a leading pharma company was in a need for a variety of data science capabilities across departments.
- Due to explosion in high value Data Science projects and emerging/rapidly changing technology choices, departments started using diverse and fragmented set of tools.
- Industry tools did not provide required capabilities to take models to production and operationalize them.
- There was a need of a streamlined approval process; without which getting access to data and AI tools was time consuming.

Solution approach

- Surveyed the needs of several departments having 600+ Data Scientists across departments and performed market research on this basis.
- Conducted workshop sessions involving relevant stakeholders, domain experts and analysts to brainstorm multiple use cases to capture & identify functional customizations.
- Fast-fail approach was utilized using agile delivery methodology to identify key bottle necks.
- A unified data science experience was built to handle multiple tools including Databricks, R-Studio, Shiny apps deployed on containers.

Outcomes/ Impact

- Delivered high impact by creating a unified data science platform on cloud.
- The platform offered a very contextualized experience including integration to client cost centers .
- Considerable license cost savings were realized with users moving into cloud-based computes through the unified solution.
- Benefits to the various stakeholders are highlighted in subsequent pages.

Case study (2/3) – Leadership and information governor view

The platform integrated governance process to enable frictionless provisioning of the data and other assets. This provided leadership and information governors visibility to the usage and cost of the data science resources across enterprise.

1 Allows departmental heads to make different kind of assets available for its users. Different Assets are governed on the platform.

- Libraries
- Datasets
- Cluster templates
- Container images
- Packages

3 Multiple workflow-based approvals were leveraged such as project access, dataset access, vendor data access and custom cluster provisioning

1

Asset Administration

Departments: R&D, Operations, Marketing

Assets: Containers, Clusters, Computes, Datasets, Libraries, TPA

Asset Instances

Container	Documentation	Owner	Department
RStudio	https://rbe.s3	John William	R&D
RStudio	https://marketing.s3	Luis Alvarez	Marketing
Jupyterhub	https://mcom.s3	Anaximander	Marketing
RStudio + Latex	https://rlap.s3	Mary Annings	Operations
Jupyterhub	https://rlap.s3	Daniel Britto	R&D

Showing 1 to 5 of 8 entries.

General Details

Container Name	Owner	R/Python Version	Description	Documentation Link	Container Link
RStudio	John William	2.7	RStudio for analytics	https://rbe.s3	https://rbe/dds-unity/d3analytics/ecr/amazon.com

Container Details

Department	Approved By	Authentication Enabled
R&D	Meda Krishna	No

2

Container Management

Department: DTESTS

Display Name	Documentation Link	Container Type	R/Python Version	Update Date	Number of Projects	Number of Deployments	Global Status Available	Department Availability	Status
RStudio	https://rbe.s3	RStudio	2.7	05/12/2020	1	1	True	True	✓
Jupyterhub	https://mcom.s3	Jupyter	2.7	05/04/2020	0	0	True	True	✓
RStudio	https://rlap.s3	R Shiny	2.7	12/26/2019	0	0	True	True	✓

Showing 1 to 3 of 3 entries (Filtered from 6 total entries)

Compute Management

Department: DTESTS

Display Name	Environment	Memory	CPU	EBS Volume	Update Date	Number of Projects	Number of Deployments	Global Status Available	Department Availability
Small	project & app	0.512	0.5	1.0	01/27/2020	1	1	True	True
Medium	project & app	2.0	1.0	10.0	01/27/2020	0	0	False	True
Large	project & app	4.0	2.0	1.0	01/27/2020	0	0	False	True
Special	project & app	16.0	8.0	1.0	01/27/2020	0	0	False	True

3

My Approvals

Search... All Statuses

All Types: Project Access, Custom Cluster, Dataset Access, FA Access, TPA Dataset Access

Request ID	Name	Request Type	Requested By	Department	Last Updated
4	demo-fa	FA Access	rohit.singh		12/18/2019
5	demo-fa	FA Access	rohit.singh		12/18/2019
1	DTESTS	FA Access	razia.sultana		04/28/2020
10	DTESTS	FA Access	sammed.kumar		01/22/2020
34	DTESTS	FA Access	ajay.krishna		03/05/2020
6	demo-fa	FA Access	rohit.singh		12/18/2019
9	DTESTS	FA Access	hari.prasad		12/27/2019

Case study (3/3) – Data Scientist/analyst view

The platform provided easy and intuitive tool provisioning to the data scientists along with collaboration capabilities using organized project structure

- 1
- Ability to publish/share reusable components and cite a project based on the level of collaboration
- Provision to restrict access at project level/ department level/ Enterprise level
- Provision to tag the cost center for each project

- 2
- Provision to select the server resources based on project needs. (Predefined or Custom)
- Provision to add datasets for accessing centralized data feeds on S3 Buckets

- 3
- Provision to deploy developed apps for internal/external consumption
- Provision to scale server environment dynamically
- Provision to allow developed apps to be published to various stakeholders

1

Add New Project

Project Info

Department* RESEARCH

Project Group* SDP

Project Name* Next-Best-Engagement

Project Visibility* Enterprise-wide

Code Visibility* Project Member

Project Description*

Chargeback Cost Info

Cost Center* Please Select Cost Center

WBS

Project Tags*

2

Resources

Resource Type* Databricks(Clusters) Container(Single Server) AWS Sagemaker

Cluster type* demo-small Custom

Library

Library Selection

Data

Data Selection

Members

Username Member Name Type Member Role

Additional Options

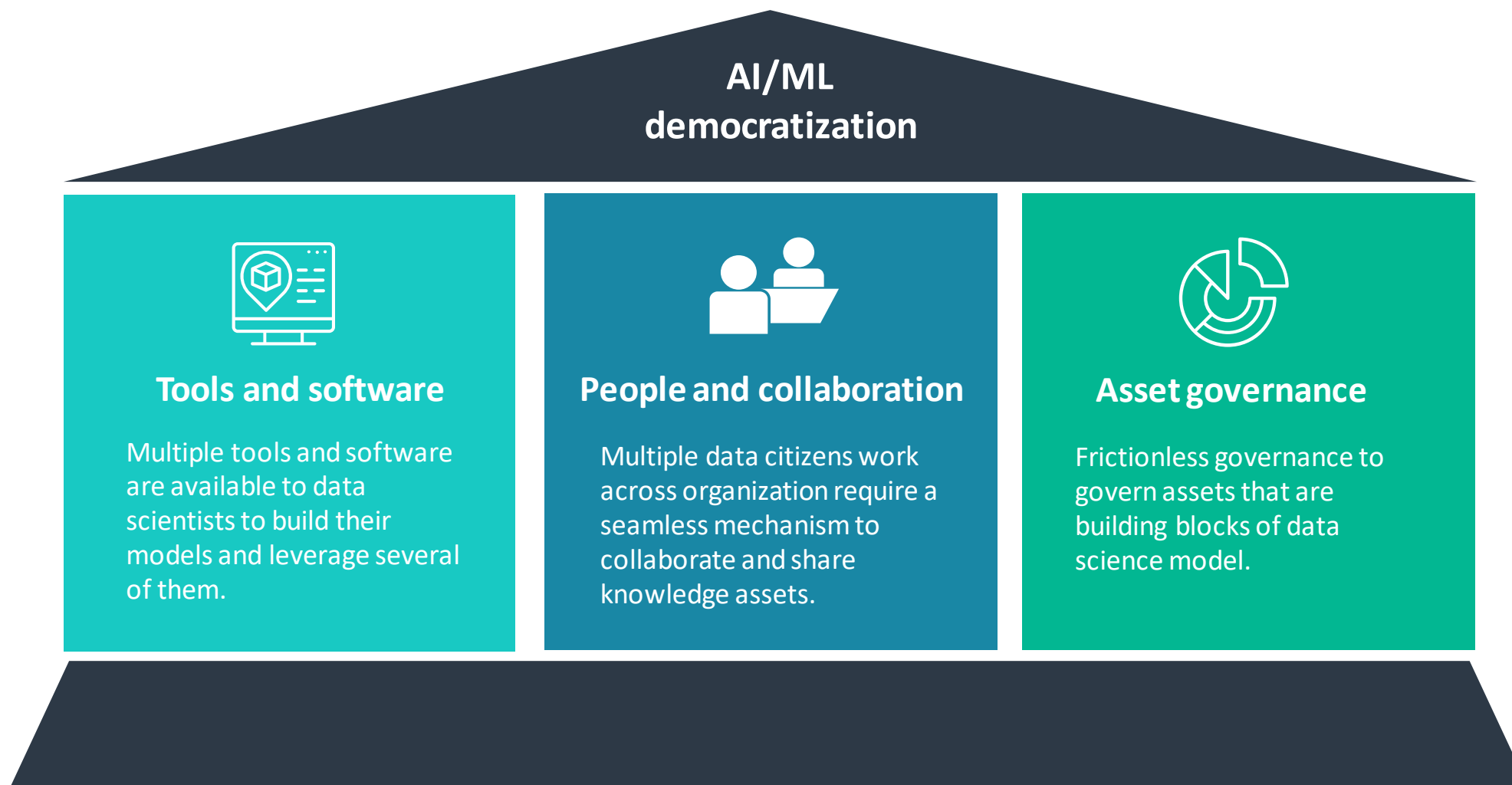
Create QuickStart Notebooks

3

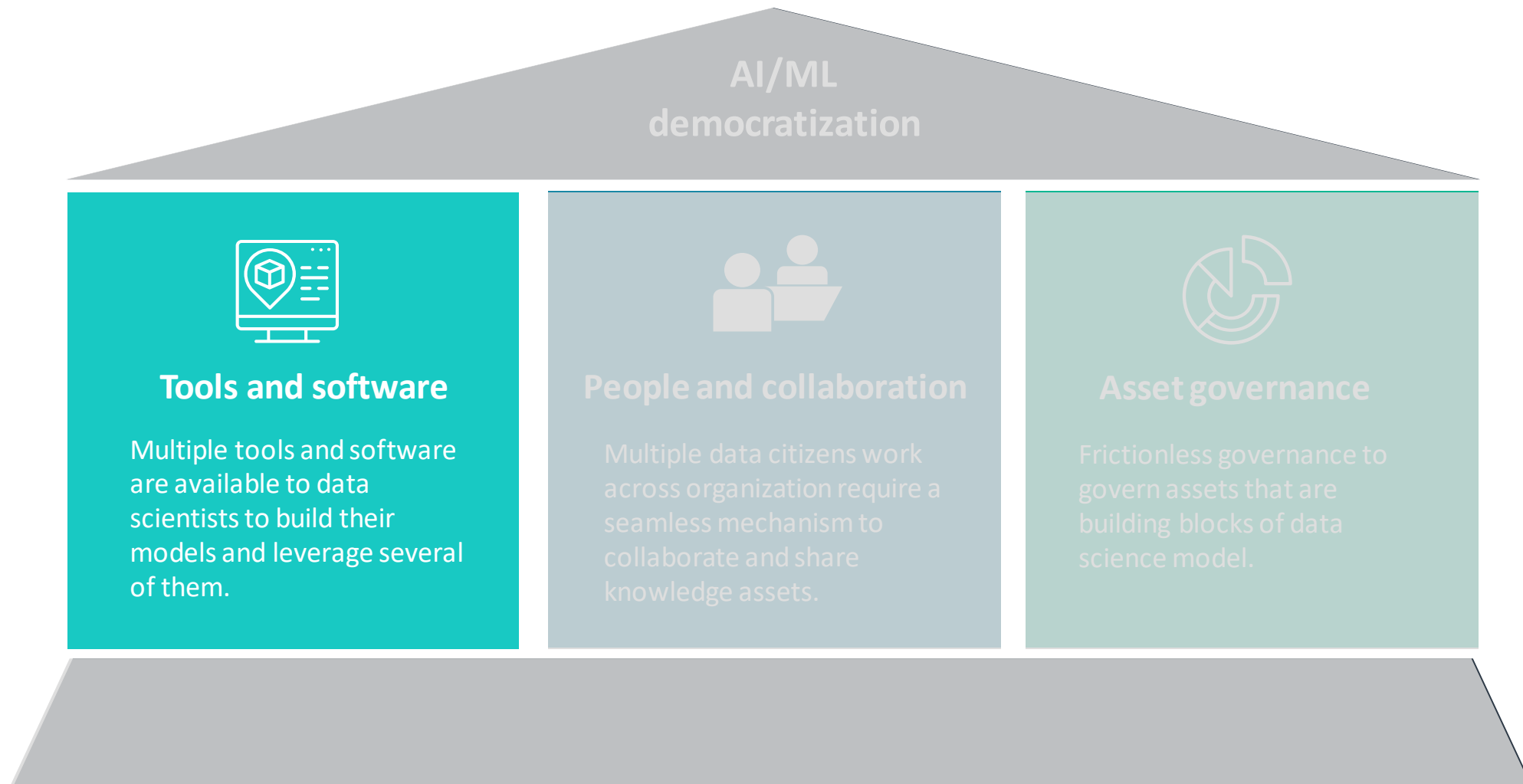
Search Show My Deployments

Project Name	Department	Environment	Service Name/ Endpoint	Started On	Type	Memory	CPU	Started By	Quick Action
Marketing Intelligence	DSTEST5	dev	demofinalcheckss	04/26/2020	r-shiny	0.512	0.5	ajay.krishna	

Solution components *(Focus areas for implementing governance)*



Solution components *(Focus areas for implementing governance)*



Tools and Software

Implement centralized access to tools and software adhering with enterprise data science strategy; while providing distributed accountability and control to the data scientists.

01

Centralized tool access

Launch all software/tools needed from a centralized location. Ability to deploy new clusters along with required libraries for optimal model development environment.

02

Seamless transition to clusters/servers

Provision a dedicated server/cluster to the project team with ability to start/stop based on project use. This would provide data scientist a seamless experience shifting from their local installations.

03

Self-service provisioning and usage

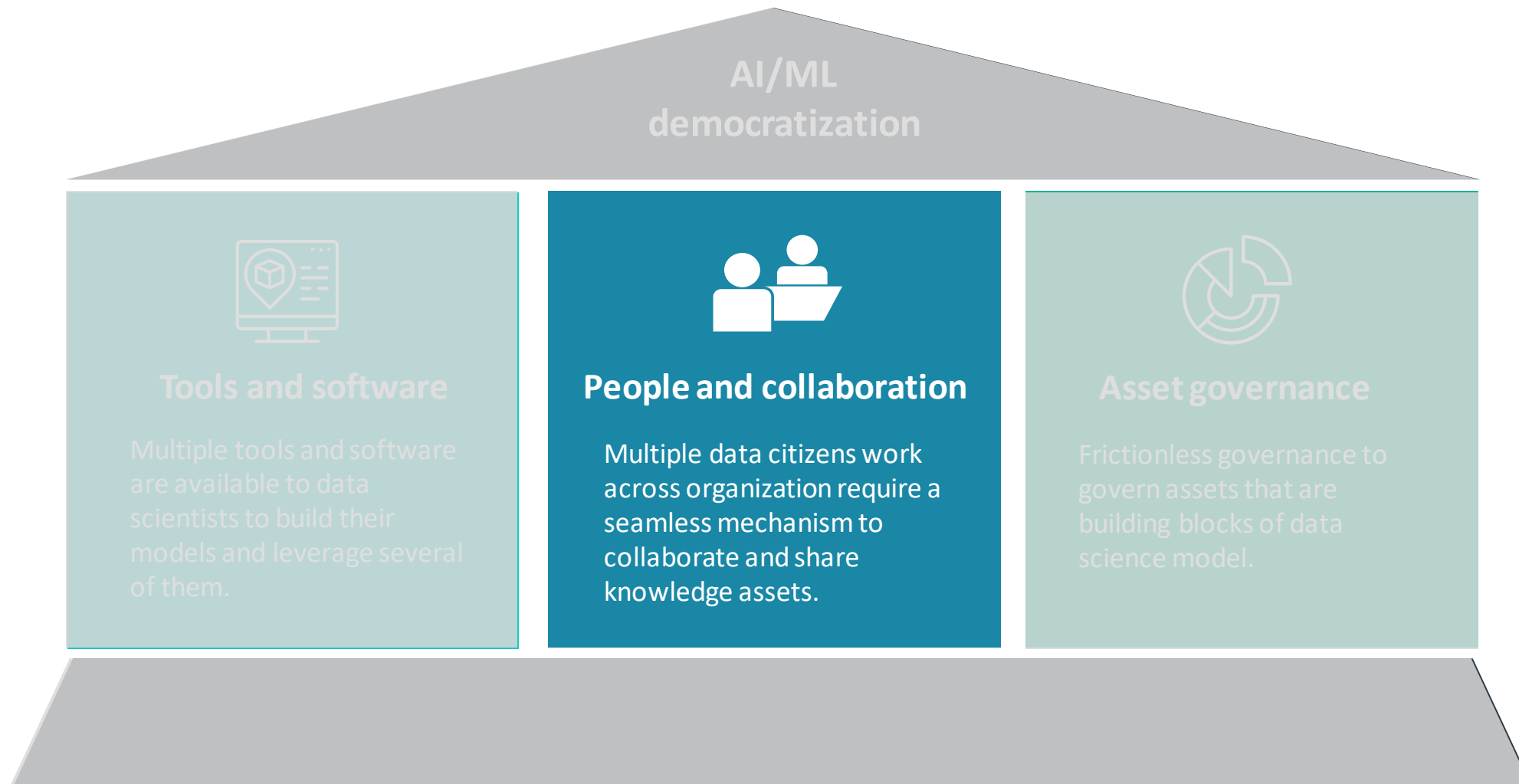
Enable monitoring of workload and server health and usage. Provision to scale up the infrastructure if needed during development and deployment.

04

Regulated provisioning

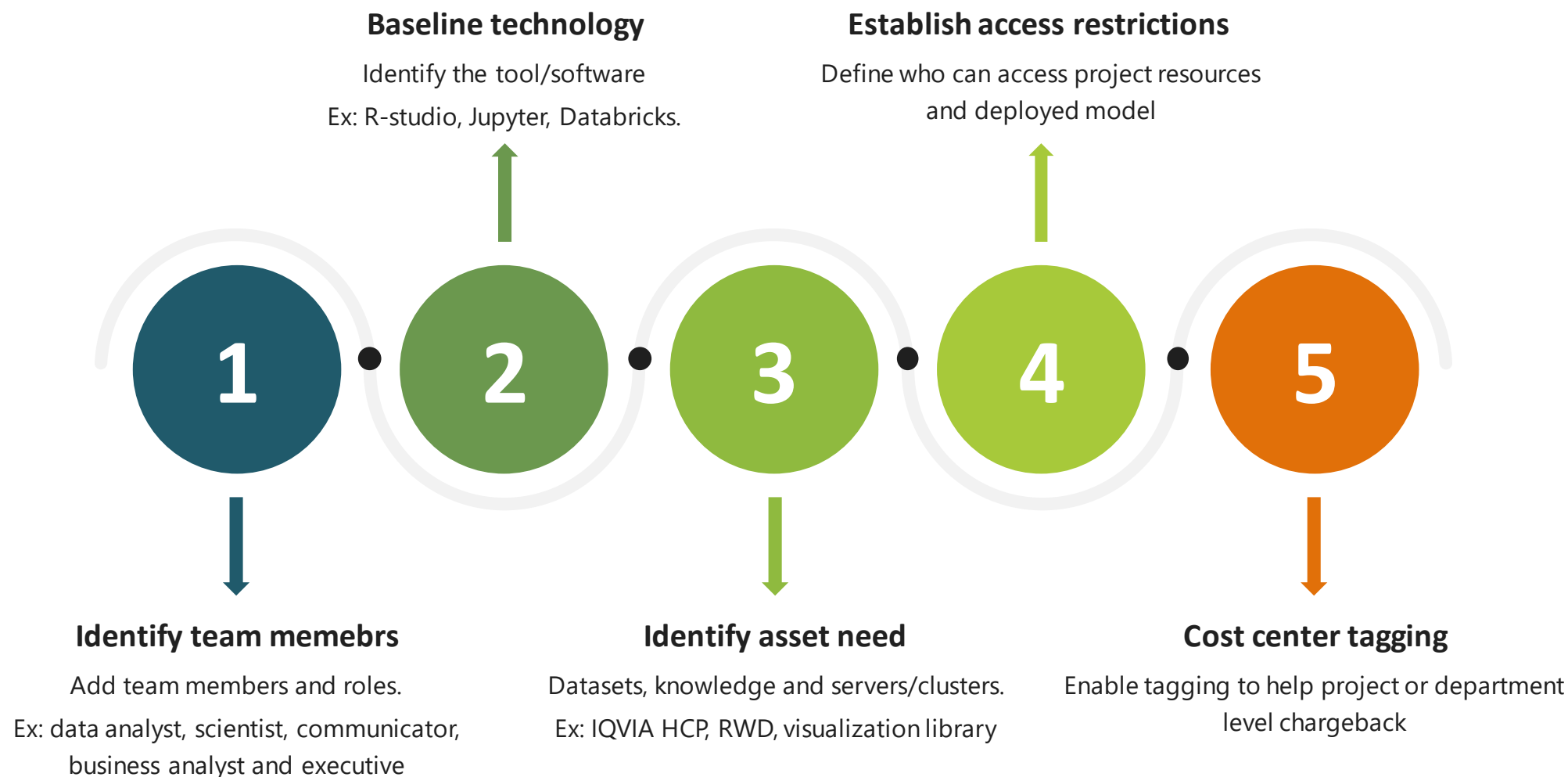
Department admins can limit inventory of tools available in their departments. Any special server configuration needed can be made available through exception approvals.

Solution components *(Focus areas for implementing governance)*

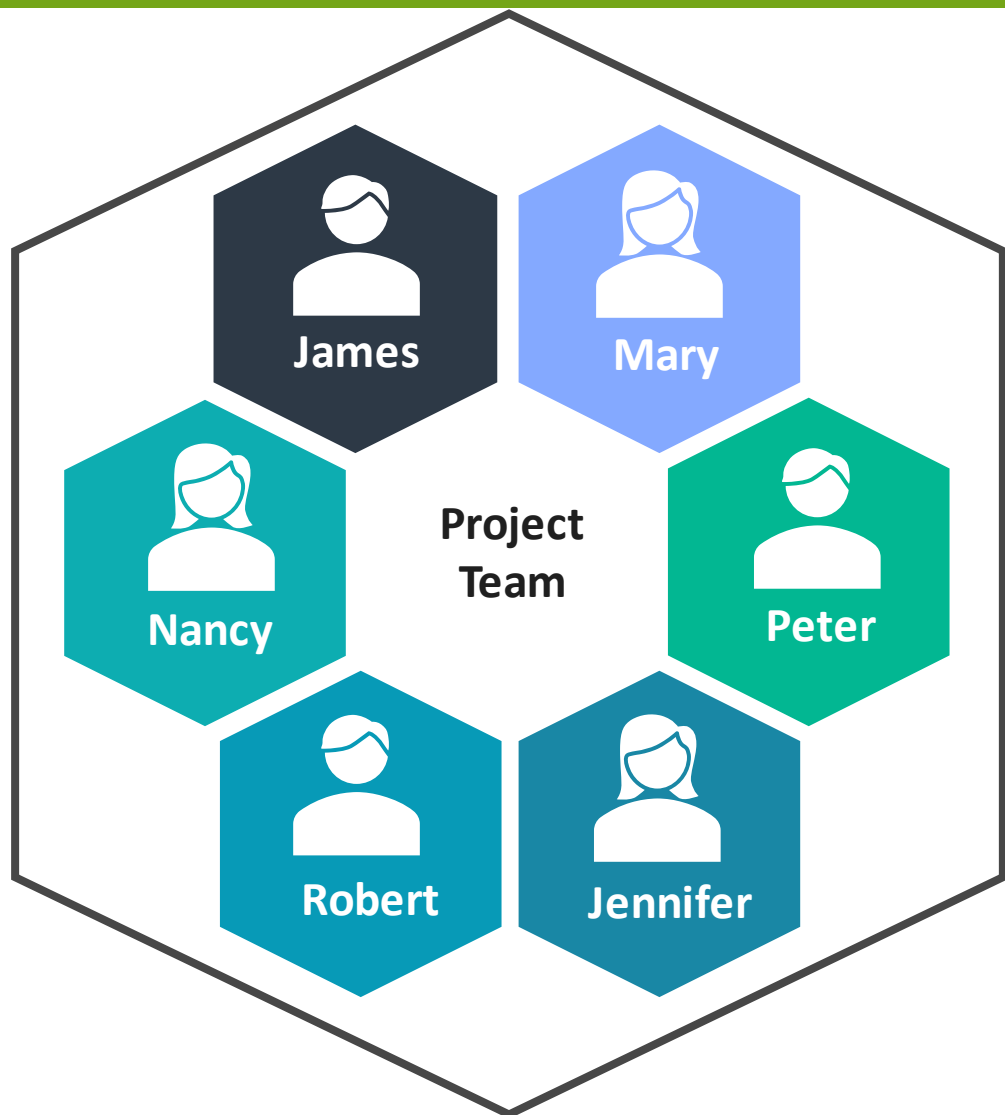


Project structure for collaboration

Project based structure (in-line with enterprise project structure) help regulate the collaboration within the team and across the enterprise. In order to enable democratization, follow below steps during AI/ML project creation.



Advantages of project based structure



Structured provisioning

Resource needs for the project are defined during project initiation. Team members can only be provisioned access based on project context.



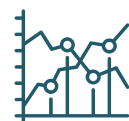
Project workspace

Provides centralised workspace to the team members for model development



Centralised repository

Shared code-base helps the team to align on latest development and versions



Traceability to the resources used

Visibility of all the resources used in the project is available to developers and leadership









Cost centre tracking

Track usage metrics of different cloud-based services

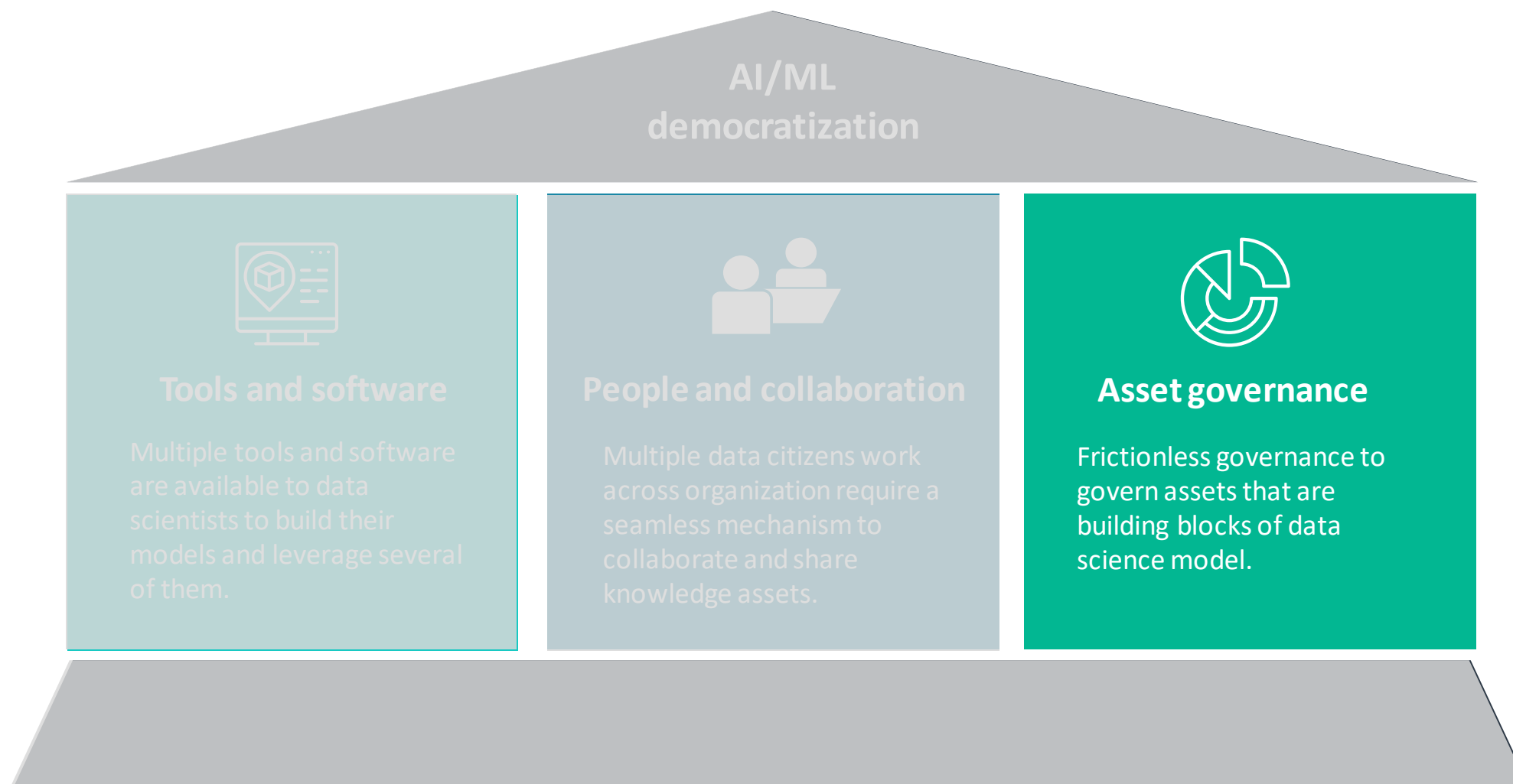
Collaboration across enterprise



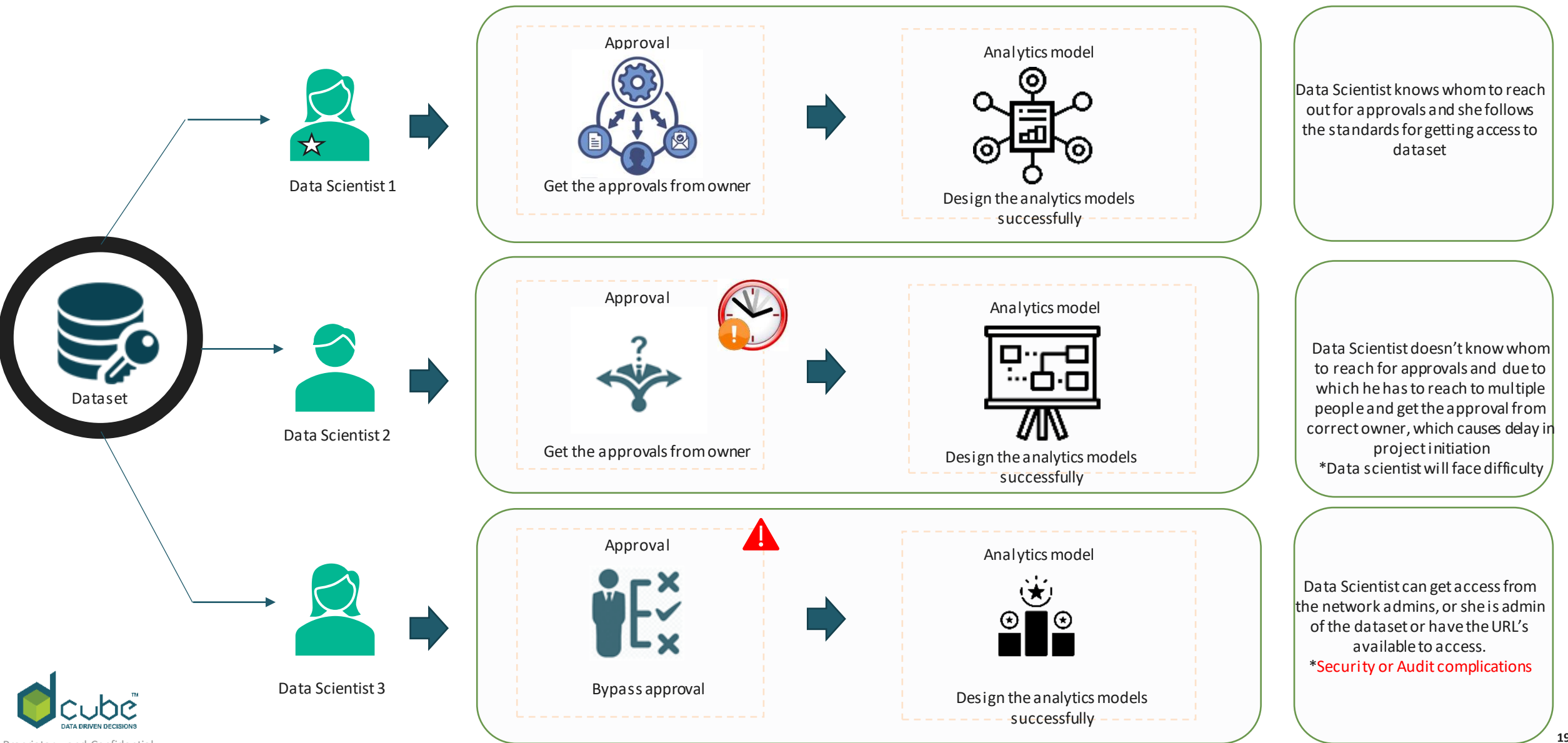
Components requiring collaboration:

-  **Dataset (Vendor, Processed)**
-  **Libraries**
-  **Algorithms**
-  **Packages / Containers Images**
-  **Enterprise Standards for DS**
-  **Visualisation and discovery**

Solution components *(Focus areas for implementing governance)*

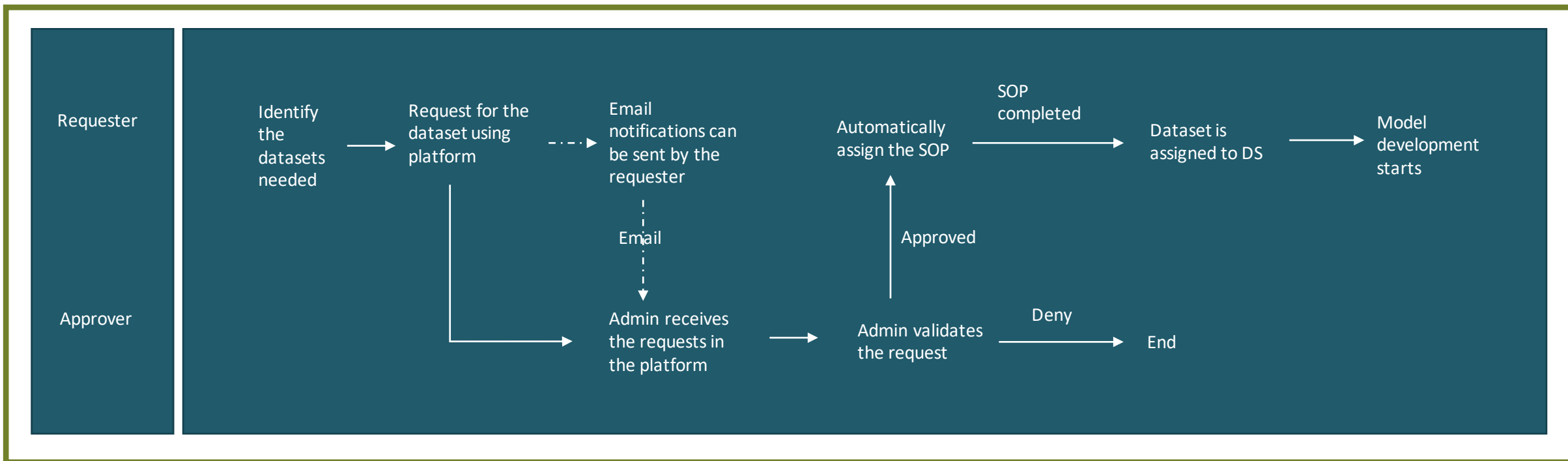


Resource access without governance



Workflow based governance – an illustration

Project



*This is an illustrative flow given for one of the shared assets (dataset).



Datasets



Libraries



Packages / Container Images



Servers/clusters



Algorithms

Workflow based governance provides regulated asset access and visibility

Typical problems with asset access



Delayed project kick off due to provisioning roadblocks



Department level access enforcement



Enforcement of contract limitations



Lack of structured Spend monitoring



Workflow based governance



Zero wait provisioning with well defined workflows



Provision to access based on department/cost centre

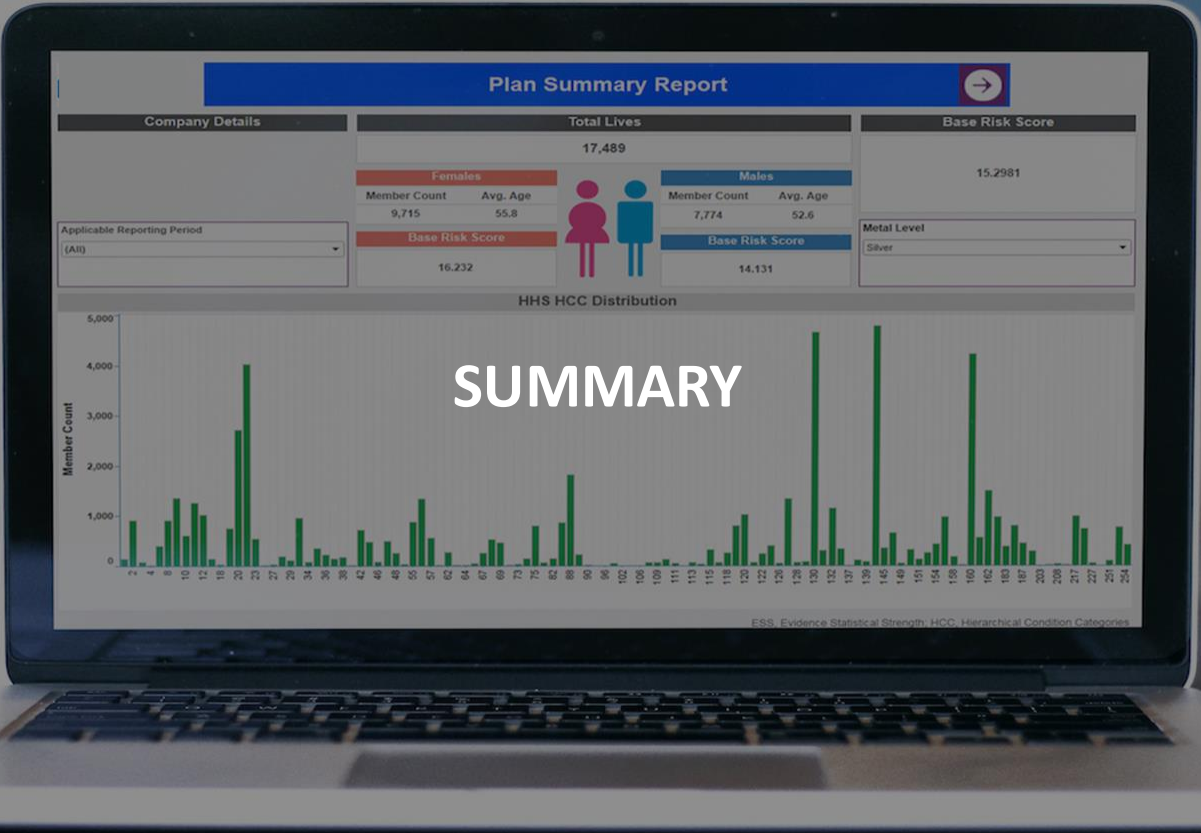


Pharma based workflows with provision for SOPs, Trainings



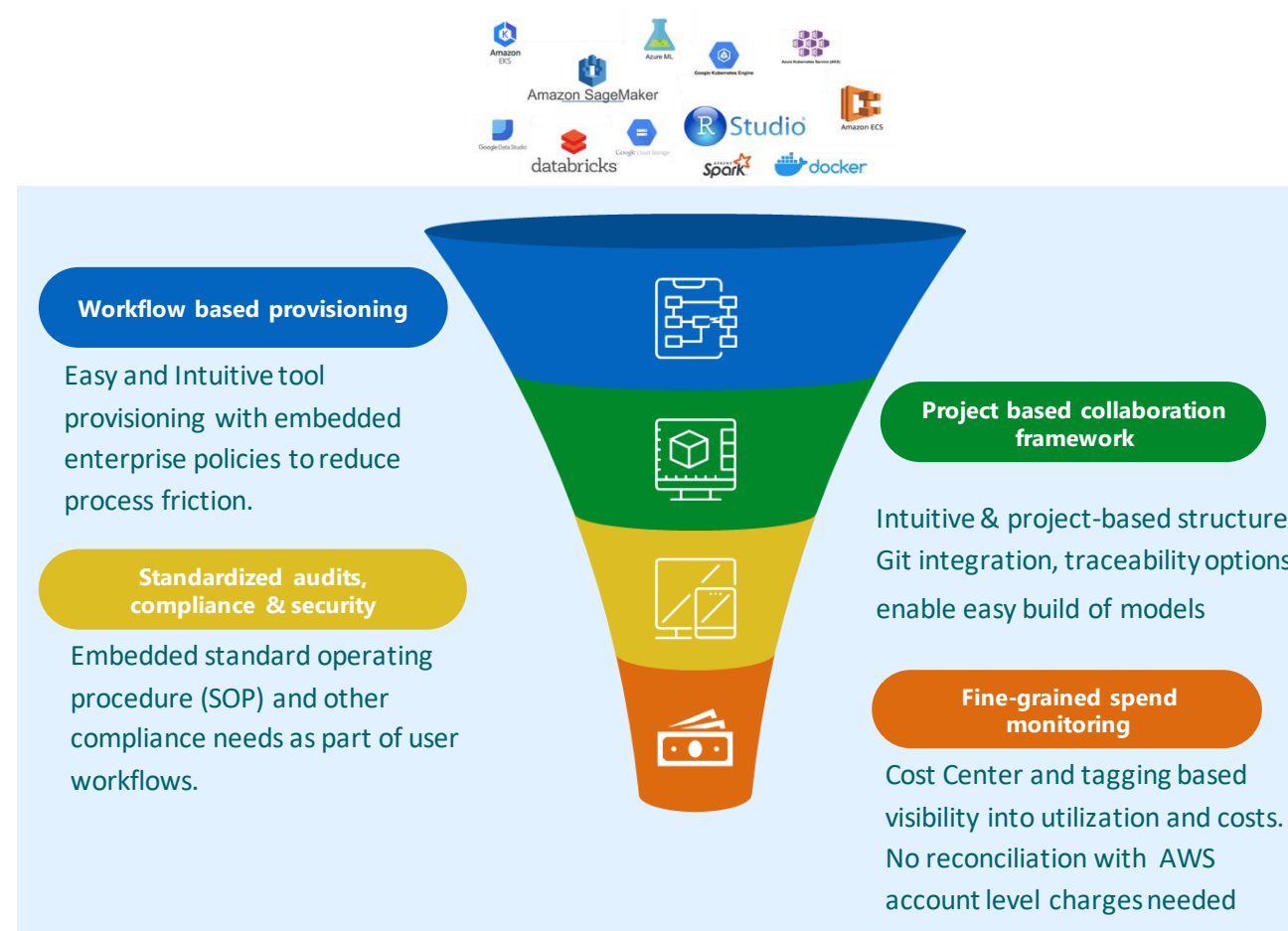
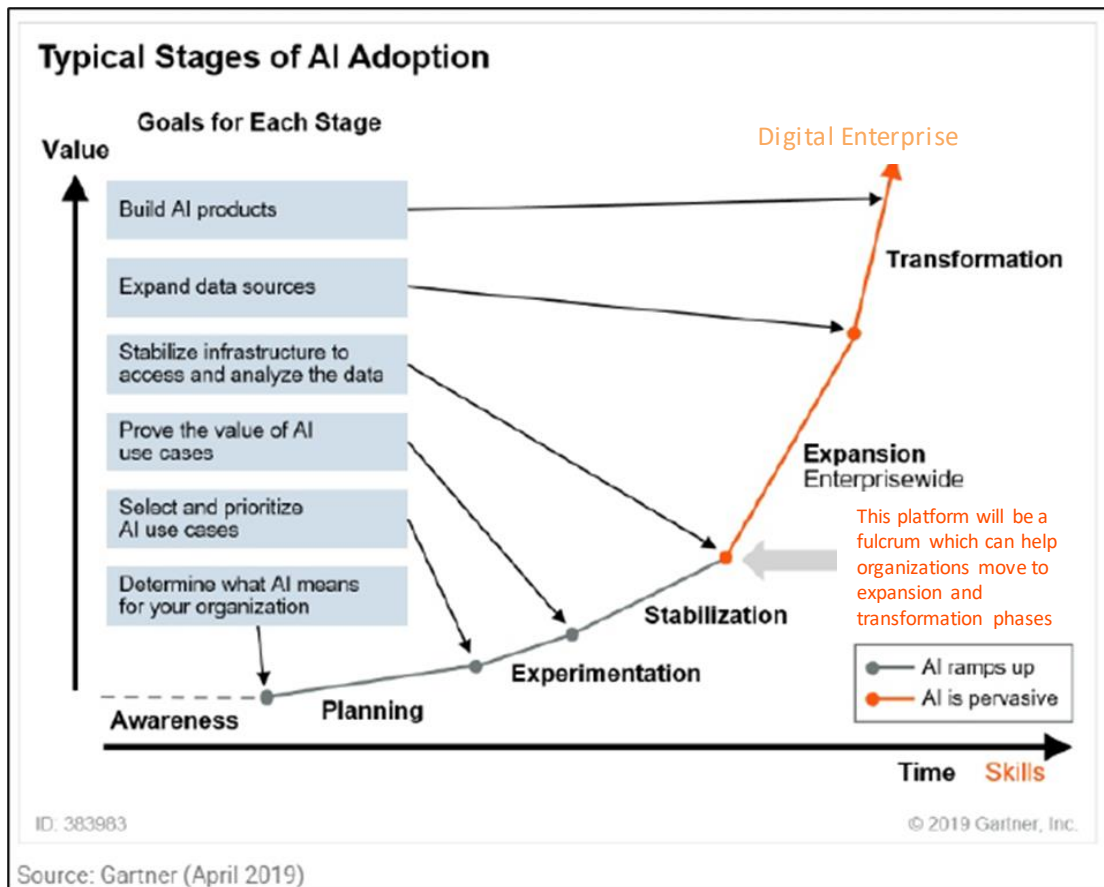
Better visibility during compliance audits and budgeting

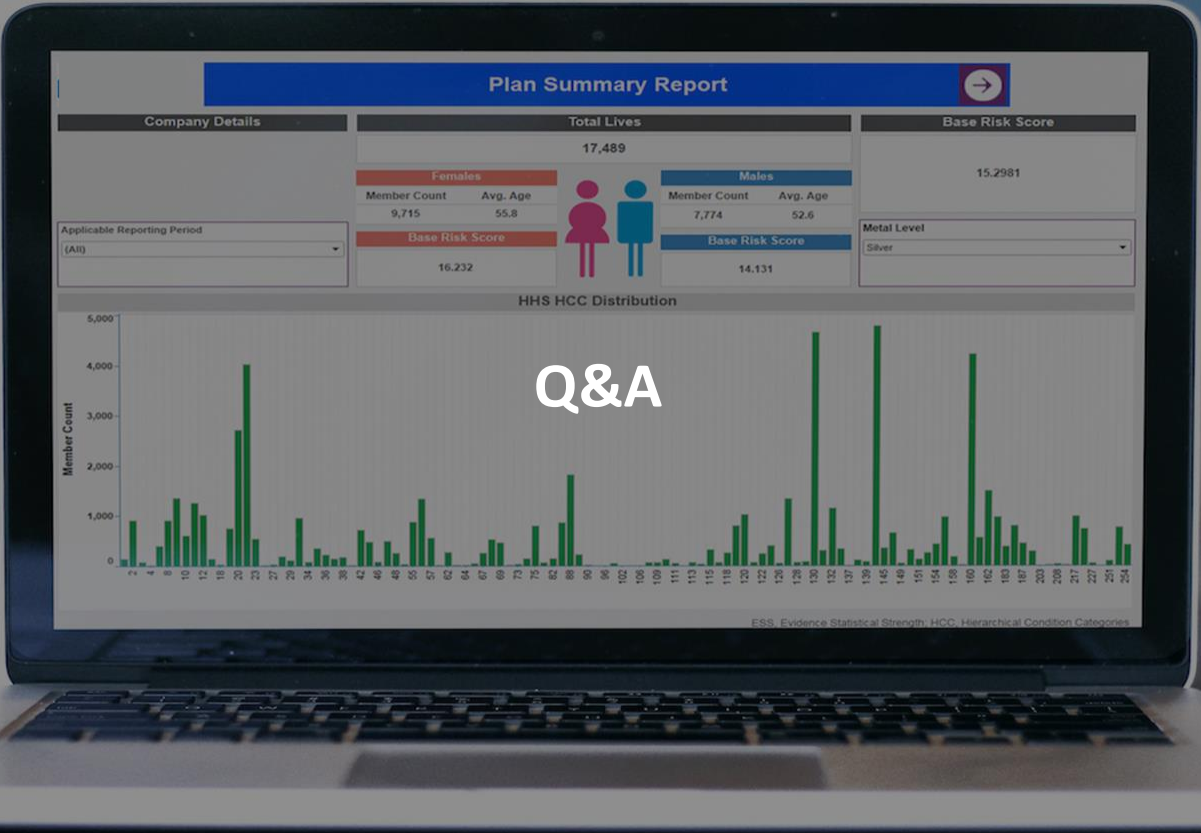
Workflows help in complete
Governance of the platform



Enabling organizations to leapfrog from experimentation to expansion

Enables bio-pharma organizations to democratize analytics ecosystem through a unified platform integrating governance, security & spend monitoring





Know more about democratization of AI/ML

Contact Us

Contact

Email

info@dcubeanalytics.com

Website

dcubeanalytics.com



Contact

Phone

US : +1 847.807.4996

US

Office

D Cube Analytics Inc. 1320 Tower
Road, Schaumburg, Illinois 60173,
USA

References

- Facebook-Cambridge Analytica Data scandal.

https://en.wikipedia.org/wiki/Facebook%E2%80%93Cambridge_Analytica_data_scandal

- Form 10-K: Alphabet Inc.

<https://www.sec.gov/Archives/edgar/data/1652044/000165204419000004/goog10-kq42018.htm>

- AI Governance spotlight – early lessons and next practices

Gartner Publication (published Apr 2019)

- Artificial intelligence maturity model

Gartner Publication (published Mar 2020)

- How Do Data Science Workers Collaborate?

<https://medium.com/ai%C2%B3-theory-practice-business/how-do-data-science-workers-collaborate-c4158d8bd471>

