

The Intelligent Data Fabric

*The Right Data, to
The Right People, at
The Right Time ...*



accelerating Artificial Intelligence / Machine Learning / Decision Making

What is your Data User Experience like?

A



B



C



D



Your AI is only as good as your data.
Is your data house in order?



I know it's someplace in here !



Your data consumers used to just be your analysts and data scientists.

Now, **everyone** needs access to **trusted data** for useful insights.

So, we *copied and pasted* data to give more users access.

The problem is, when you scale that across an enterprise it's *untenable*.

5-7 ERP systems

8 Data lakes

5 Data warehouses

8.6 Clouds¹

364 SaaS apps²

1,000+ Sources³

Despite spending an average of **\$229M¹** on data-related costs, organizations are still fighting a disparate data landscape with:



More data sprawl

Multiple locations and silos



More data complexity

Documents, images, videos



Worse data quality

Stale and inconsistent



Bruce Cerretani

IBM Solutions Architect



My DNA



20 Years



25+ Years



***Cerretani Family
Board of Directors***

Ryan Renee, Chairwoman

Data complexity

inhibits enterprises from becoming data-driven

68%

of enterprise data
is not analyzed



80%

of time is spent
cleaning, integrating,
and preparing data



82%

of enterprises see data
quality as a barrier in
data integration projects

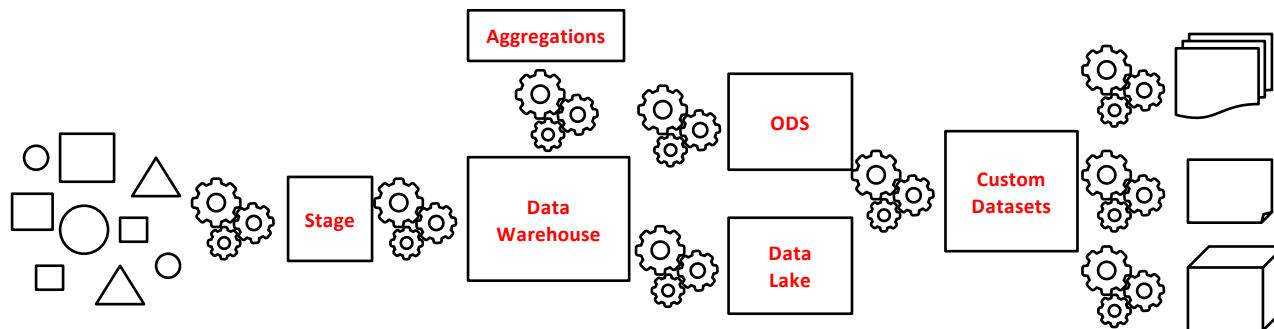


82%

of enterprises
are inhibited by
data silos

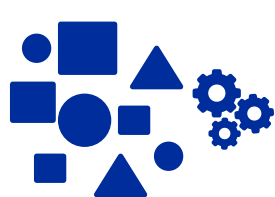
Data First ... 'Old School' vs. 'New School'

Old School

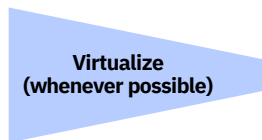


- ✗ Multitudinous sources
- ✗ Complex data integration
- ✗ Complex data movement
- ✗ Complex data storage
- ✗ Distributed, disparate business rules
- ✗ Costly to build, maintain, and support
- ✗ Requires many resources, and different skillsets
- ✗ Involves many different vendors
- ✗ Must accommodate constant change
- ✗ IT-managed

New School



Any source
(Producer)



Raw & Curated data
(copy only when necessary)



Intelligent Integration
(Knowledge Catalog + AI Engine)

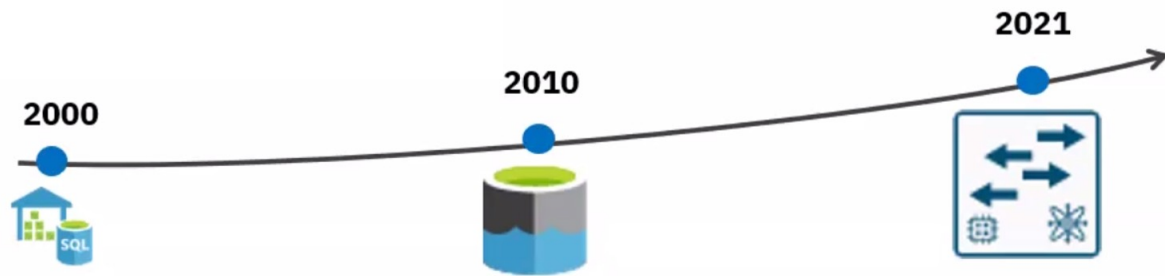


Any target
(Consumer)

- ✓ Any source to any target
- ✓ Automatically detect business rules
- ✓ Automatically detect changes
- ✓ Centralize data management, data architecture, data integration
- ✓ Centralize data governance, privacy, security, and policies
- ✓ Avoid moving, storing, duplicating, and integrating data (minimize or eliminate ETL)
- ✓ Automatic scalability (scale-up) and elasticity (scale-out)
- ✓ Lower cost
- ✓ Significantly fewer vendors
- ✓ Business-managed

Data Fabric vs Previous Data Platforms

A natural evolution from previous data platforms



	Datawarehouse	Data Lake	Data Fabric
Architecture	Centralized	Centralized	Distributed
<u>DataOps</u>	Tool driven manual.	Tool driven. Some automation	AI infused automation
<u>MLOps</u>	Absent	Presence of some elements.	Full lifecycle support
<u>Hybrid Multicloud</u>	On-premise or single cloud.	On-premise or single cloud.	The fabric may span across many clouds and locations
<u>Polyglotism</u>	Low. RDBMS, Star Schema	Moderate. RDBMS, Blob	High. Does not pose any restriction on database type
Consumers	BI	Mostly BI. Some analytics.	BI, Advanced Analytics, System of Record, System of Engagement

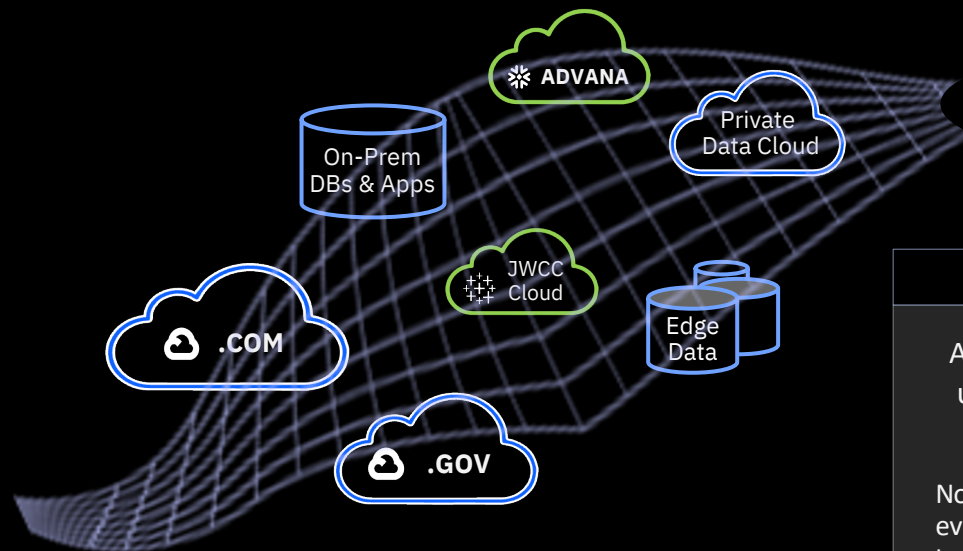
An OPEN Approach to Data

A **Data Fabric** is an architectural pattern, set of services on a platform that integrates data across the enterprise regardless of data location (On-Prem, Cloud, Multi Cloud, Hybrid Cloud).

Data Silo Killer ...

connecting the

Right Data, to the
Right People, at the
Right Time



E Pluribus Unum

(out of many, one)

Data Fabric “Design”

An abstraction layer to share and use trusted data and AI across a hybrid data landscape.

Not a pendulum swing to de-centralize everything, but a balance of what should be de-centralized vs centralized ... either logically or physically.

Using data where it lives when possible, copying data when beneficial.



“

Advana has done a great job of demonstrating the power of that sort of solution.

... Once you've got your data in one place, authoritatively collected, it can be leveraged and all sorts of analytics can be developed on top of that.

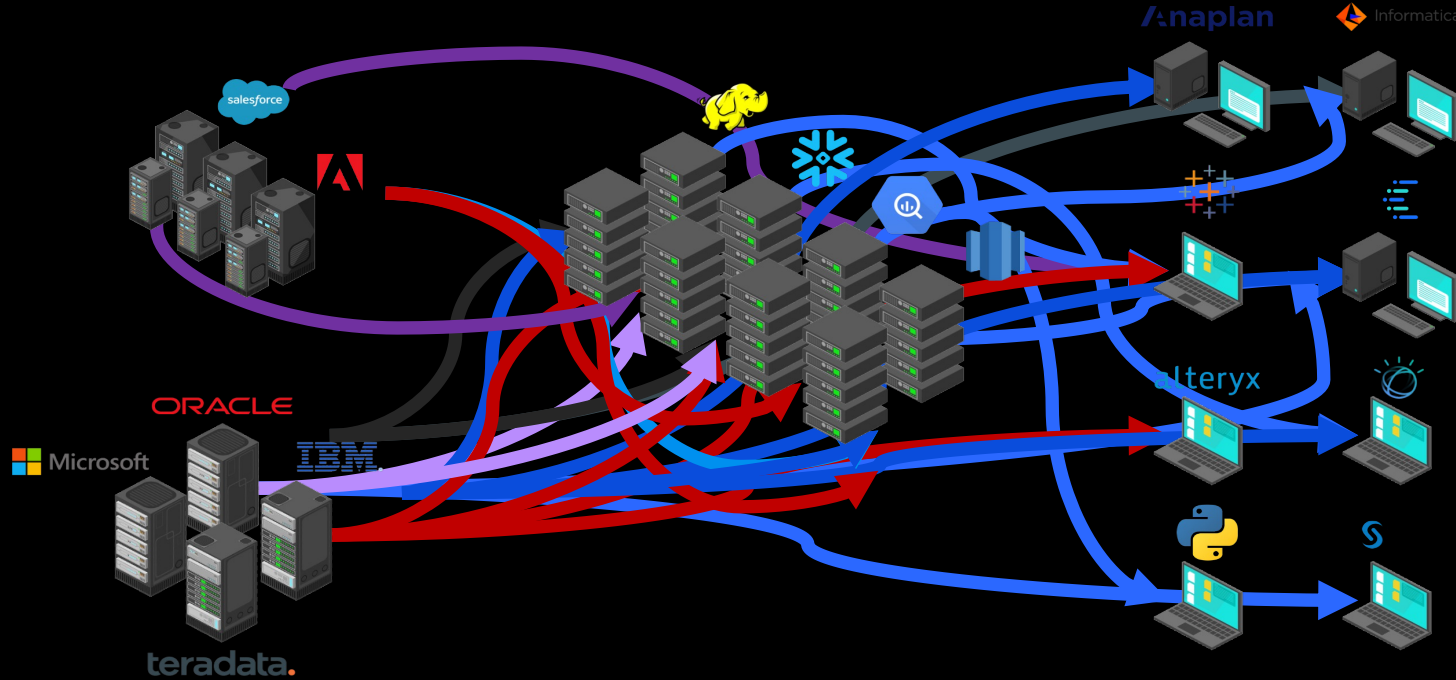
... [looking into] extending that with a federated model that allows us to leverage the data at its source, but still enable sort of self-service access permitted and be a role or attribute-based sort of mechanism that let people compose analytics based on that data.”

”

Bill Streilein, Chief Technology Officer for DoD's Chief Digital and Artificial Intelligence Office (Mar 2023)

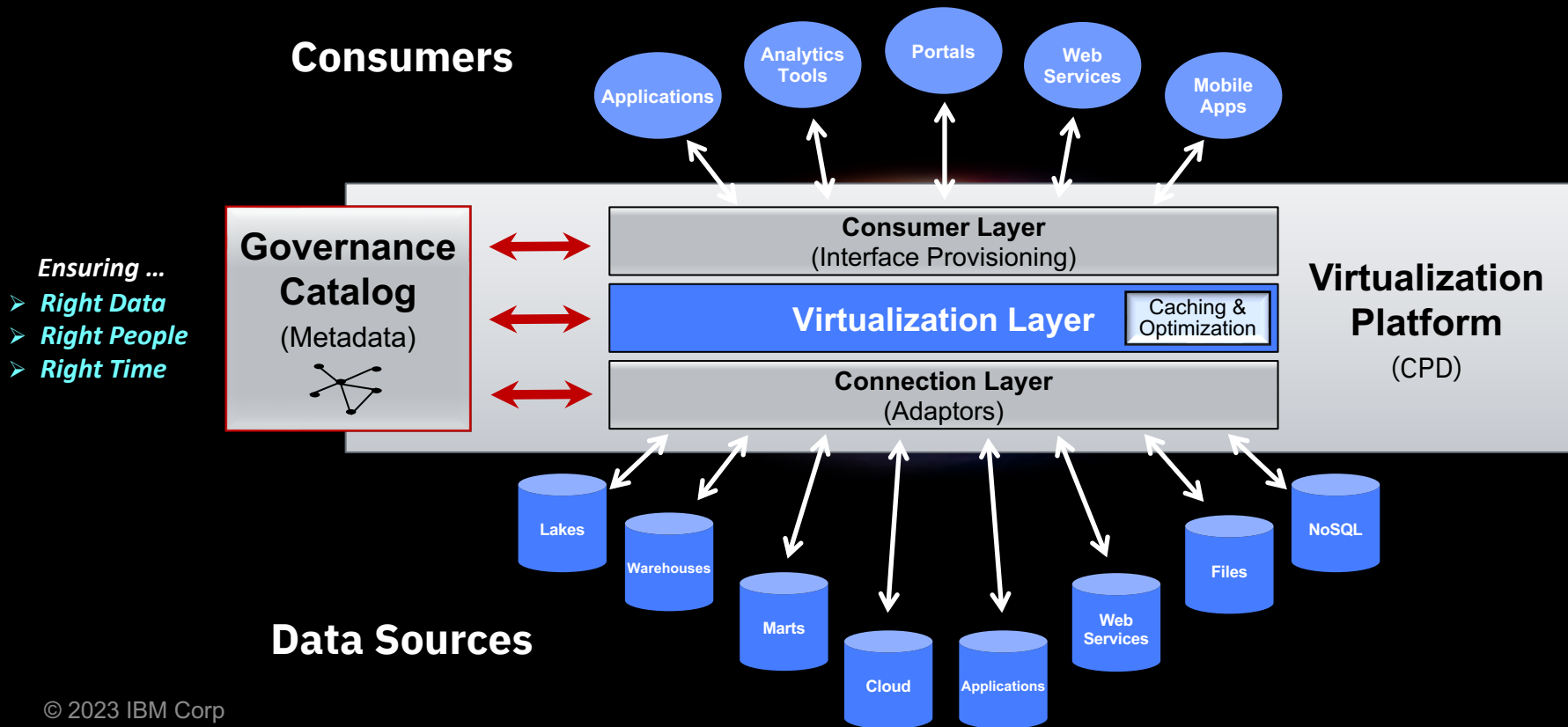
Common reality

Tool oriented, limited value, unnecessarily complex, and time-consuming



Data Virtualization

*Centralized and Trusted Access to Distributed Data ... the **Data Silo Killer***



Data Fabric and Data Mesh

Data Fabric is a technology architecture, while data mesh represents processes and principles – not technology; these are not competing concepts.

As the end-goal of both data fabric and data mesh is to facilitate self-service data consumption, a data fabric architecture can be the implementation layer that supports data mesh processes.

Data Mesh is an approach centered on organizational processes to enable agile, domain-specific ownership and creation of reusable data products. It is technology agnostic and domain owners are responsible for the entire data lifecycle.

Key benefits of a Data Fabric approach

Intelligent Integration

Automate data engineering tasks and augment data integration across hybrid cloud resources

Unified Governance

Automate data governance, protection and security tasks enabled by active metadata

Knowledge Insights

Enables self-service for easy end user data consumption and collaboration

The 4 principles of a Data Mesh

Domain ownership

End-to-end data ownership belongs to the applicable line of business teams and departments

Data as a product

Data needs to be tied to business goals and the value they generate

Self-service data infrastructure

Any data consumer, regardless of role must have easy access to the data they need

Federated governance

Enables self-service for easy end user data consumption and collaboration

Example Use Cases



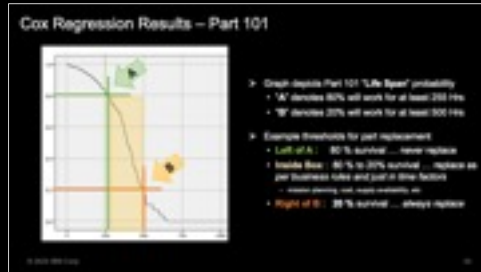
Detect Abnormal Activity



Optimize Supply Chain



Up to Date Site Surveys



Improve Mission Readiness



Predict Critical Events

Turn Data into Business Value



IBM Corp Data Fabric

use case landscape – across the enterprise

Key Business Benefits



10X ROI from Data & AI-based transformation initiatives



More than 50% improvement to cycle time of key business processes



Complex data queries now measured in **seconds** vs. minutes & hours



Achieved GDPR readiness on **Day 1**

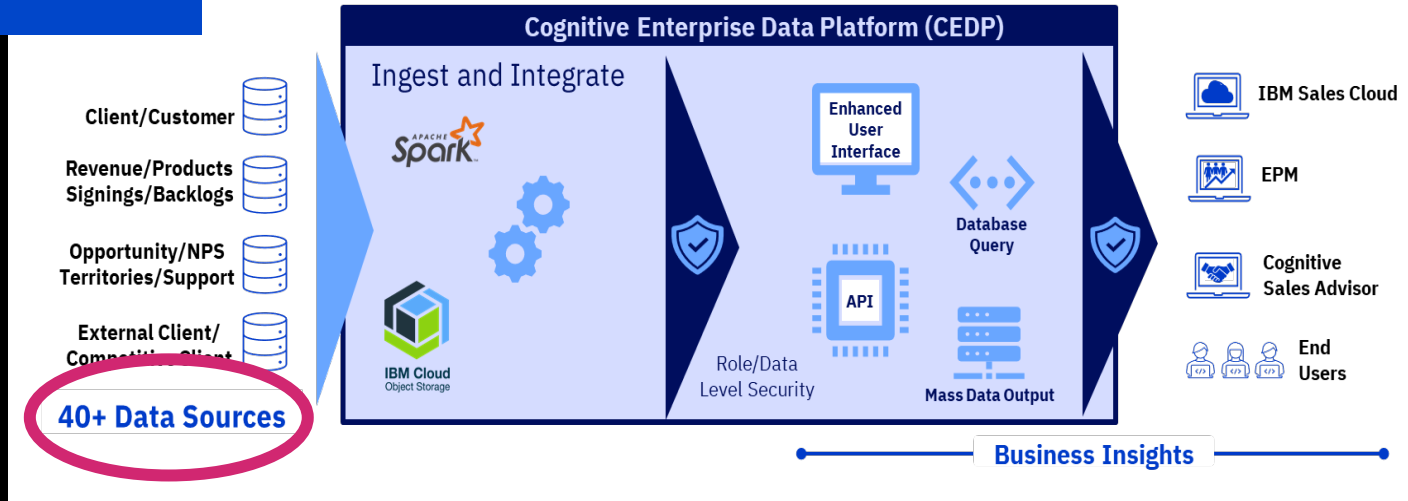


Central platform uptake to **1 trillion records**, 100% quarter over quarter increase

Ongoing 'Discovery' process scales benefits

IBM provided new levels of integrated data for analytics and business insights at scale

Single Virtual Source of
integrated and trusted data



Automation need to Scale



Automated AI

Automate training, predictive accuracy and explainability for models that rely on siloed data sets and time-variant conditions



Automated Privacy

Automate enforcement of universal privacy and usage policies across the ecosystem of sources and consumers



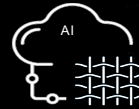
Automated Catalog

Automate data discovery, understanding and enrichment by maintaining real-time knowledge of business-ready data



Automated SQL

Automate access to unify data across distributed data/cloud landscapes without data movement



An intelligent data fabric

- Weave together new insights from unified data
- Democratize data usage across more people
- Unleash productivity and reduce cost/risk
- Make AI trusted, easier, and faster

Infuse

Operationalize AI throughout the enterprise

Analyze

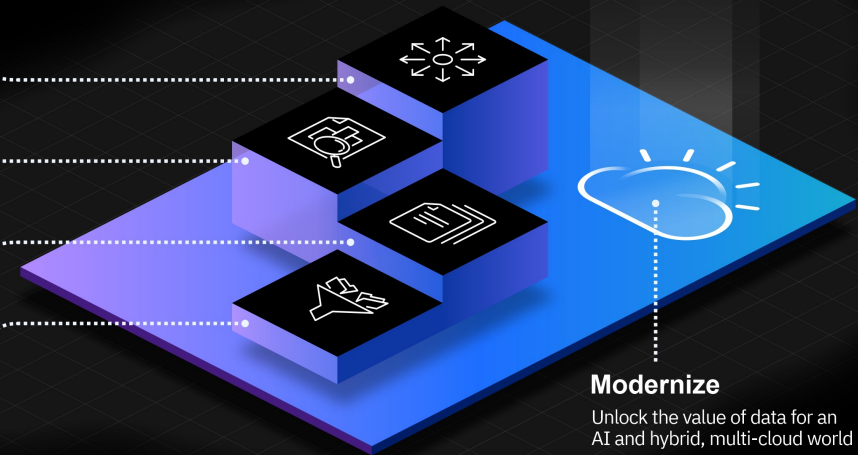
Build and scale AI with trust and transparency

Organize

Create a business-ready analytics foundation

Collect

Make all data simple and accessible

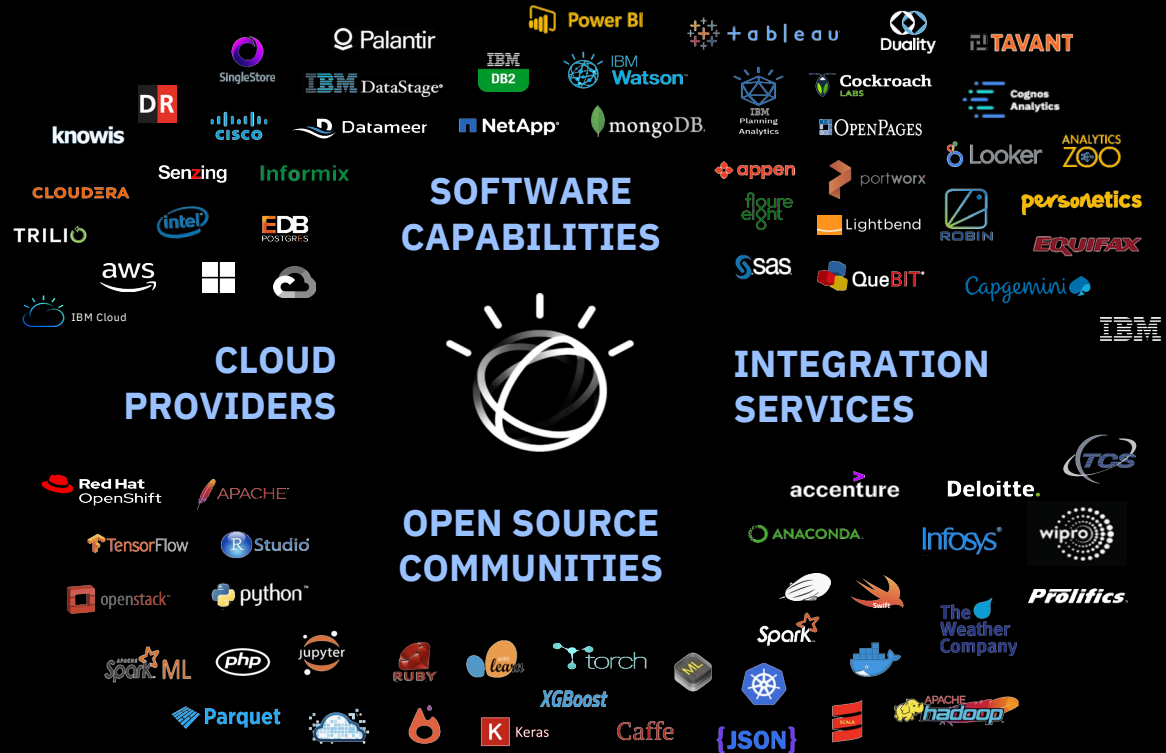


Modernize

Unlock the value of data for an AI and hybrid, multi-cloud world

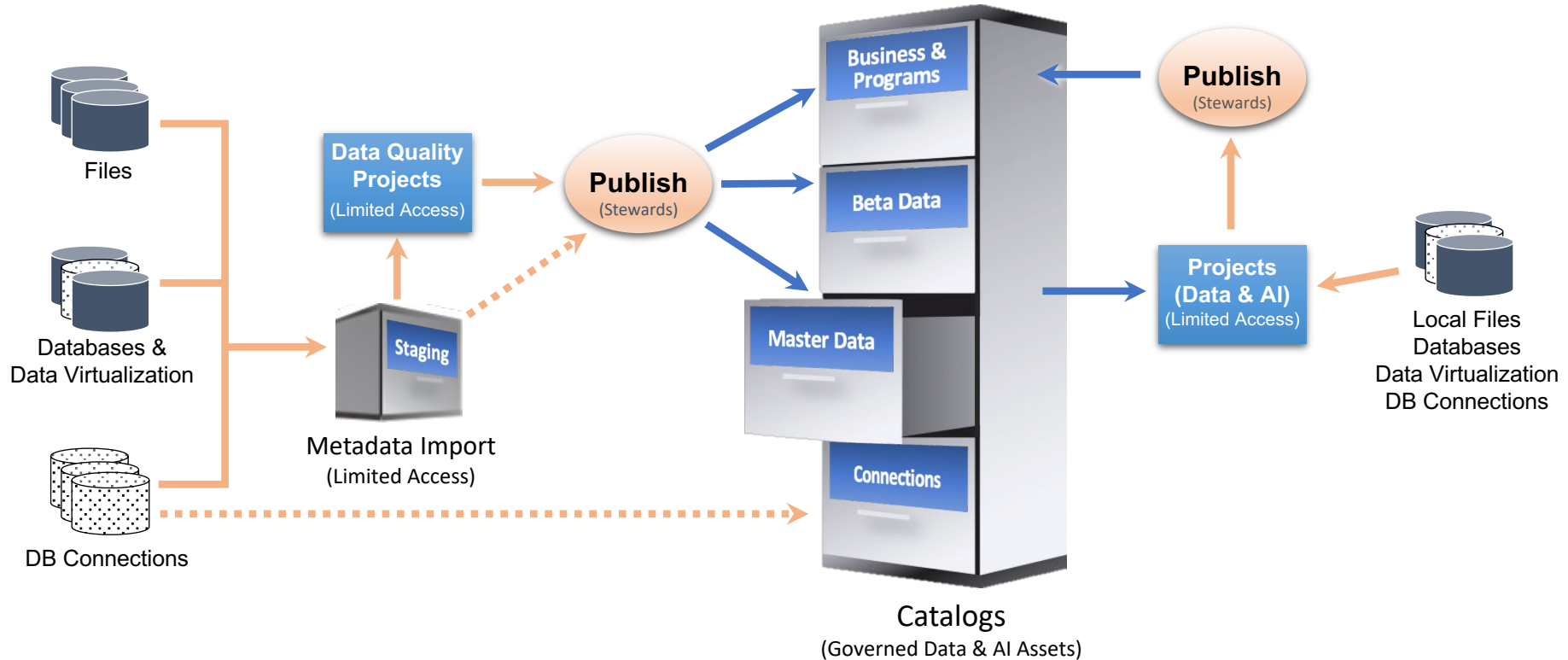
An AI Open Ecosystem is needed that integrates and customizes to your needs

Your Data and AI is only as good as the open ecosystem that supports it



Data Fabric lessons learned:

Establish a Catalog Strategy



Data Fabric lessons learned:

Real world experience

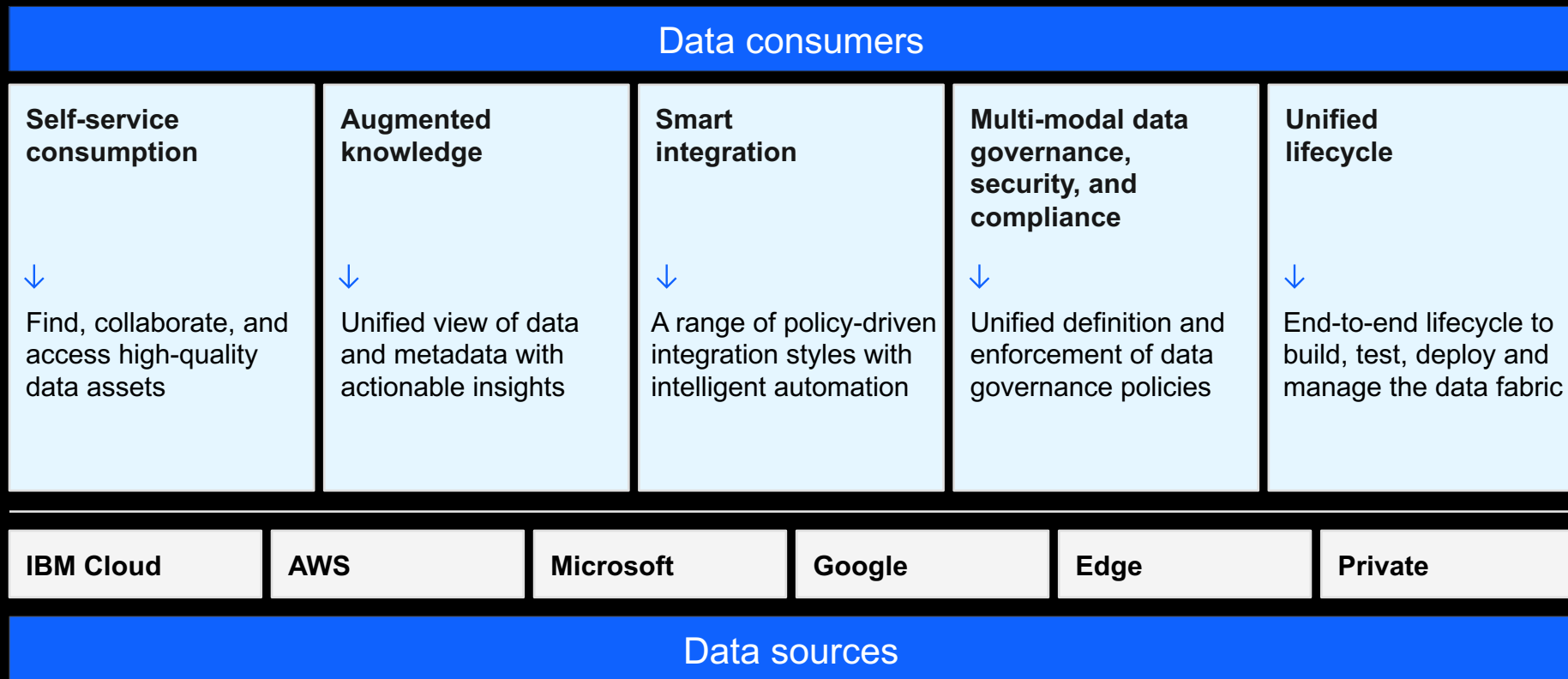
Seek First to Understand	Map Out a Data Strategy	Establish KPIs & Controls	Scale & Expand
Know why & where you are going ... understand your objectives	Know how you are getting there ... share it with key influencers	Know where you are and show value along the way	Adoption is critical ... grow or die
This is hard stuff, and it starts with leadership and vision	Use Functional Design Specifications (FDS) and Detailed Design Specifications (DDS)	Define KPIs that are important to mission success	Promotion and measurable results are key
Engage key influencers and invested stakeholders	Don't do this for the sake of doing it, do it because it makes the job easier and the mission more effective	Build in governance, traceability, and explainability to enable trustworthy AI	Demonstrate value with Data Literacy, Access, Trust, and Understanding at the speed of relevance
Inventory and prioritize data	Simpler is Better	Zero trust is an outcome ... not a blocker	Build upon what you have ... better and faster delivery of new data products

Data Fabric lessons learned:

Shape before you Scale & Expand

	Single project	Multiple projects	Multiple organizations
Use Cases	Identify a use case that delivers value while laying a foundation	Scale to other use cases within the same organization	Scale to other organizations
Data governance and privacy	Catalog is the cornerstone of data fabric and governance	Grow the discovery, mastering and curation of data domains and use of a data marketplace	Resolve cross organizational data sharing as organizational silos are entered into the catalog
Multi-cloud data integration	Virtualize data when possible, copy / move data when beneficial	Multiple integration styles and stores	Catalog the integration patterns and data sources from additional organizations
Customer 360	Connect customer data across disparate sources	Spread across more domains and automate pipelines	Expand across more domains and automate pipelines
MLOps and trustworthy AI	Prepare, train, deploy, and monitor machine learning models	Expand the MLOps practice by applying model lifecycle management, which builds trust and facilitates the sharing of best practices	Regardless of the specific MLOps practices and tools used in other areas, standardize on the model governance tooling

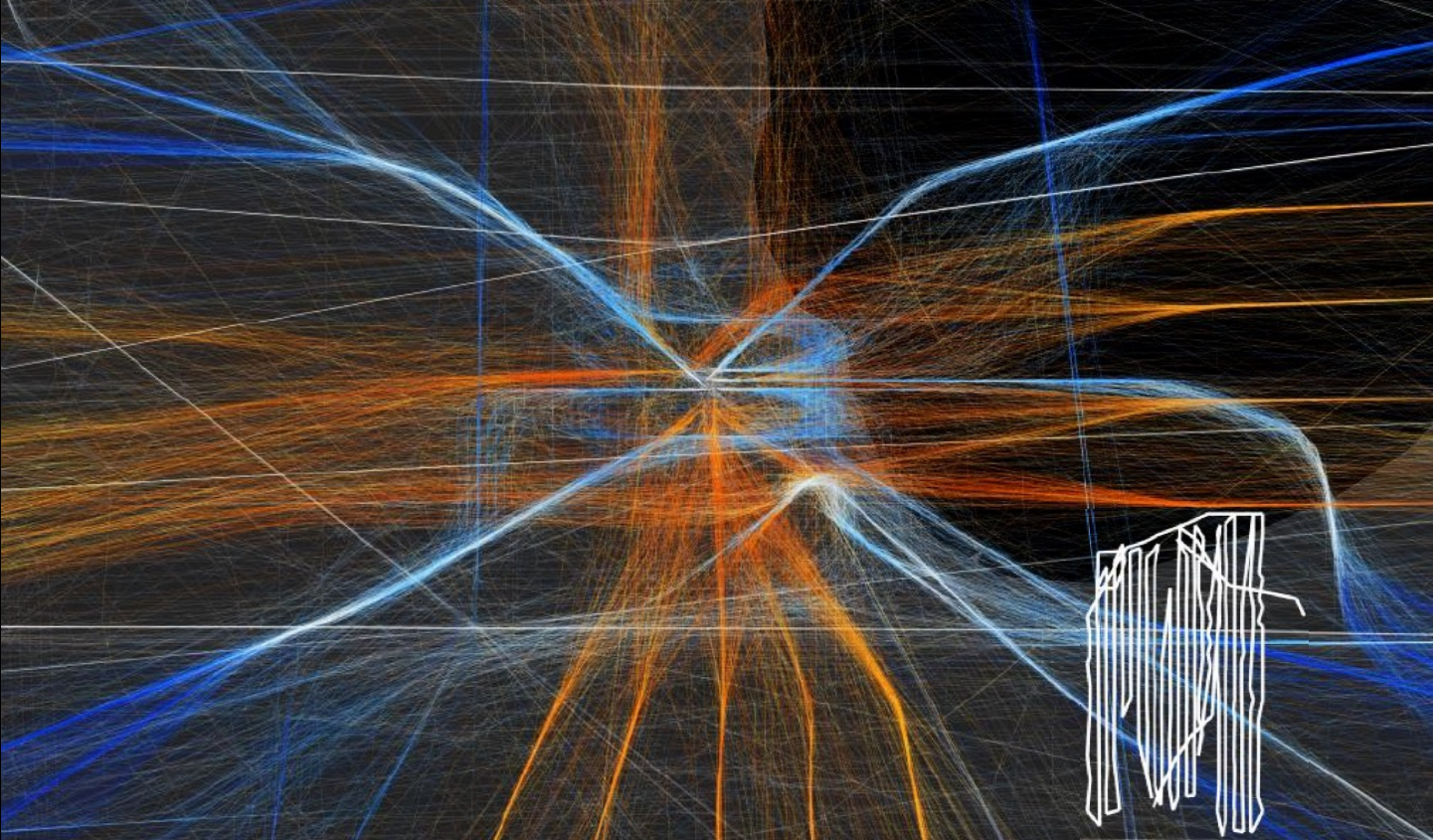
Elements of a **Data Fabric** design for data-driven organizations



Discussion



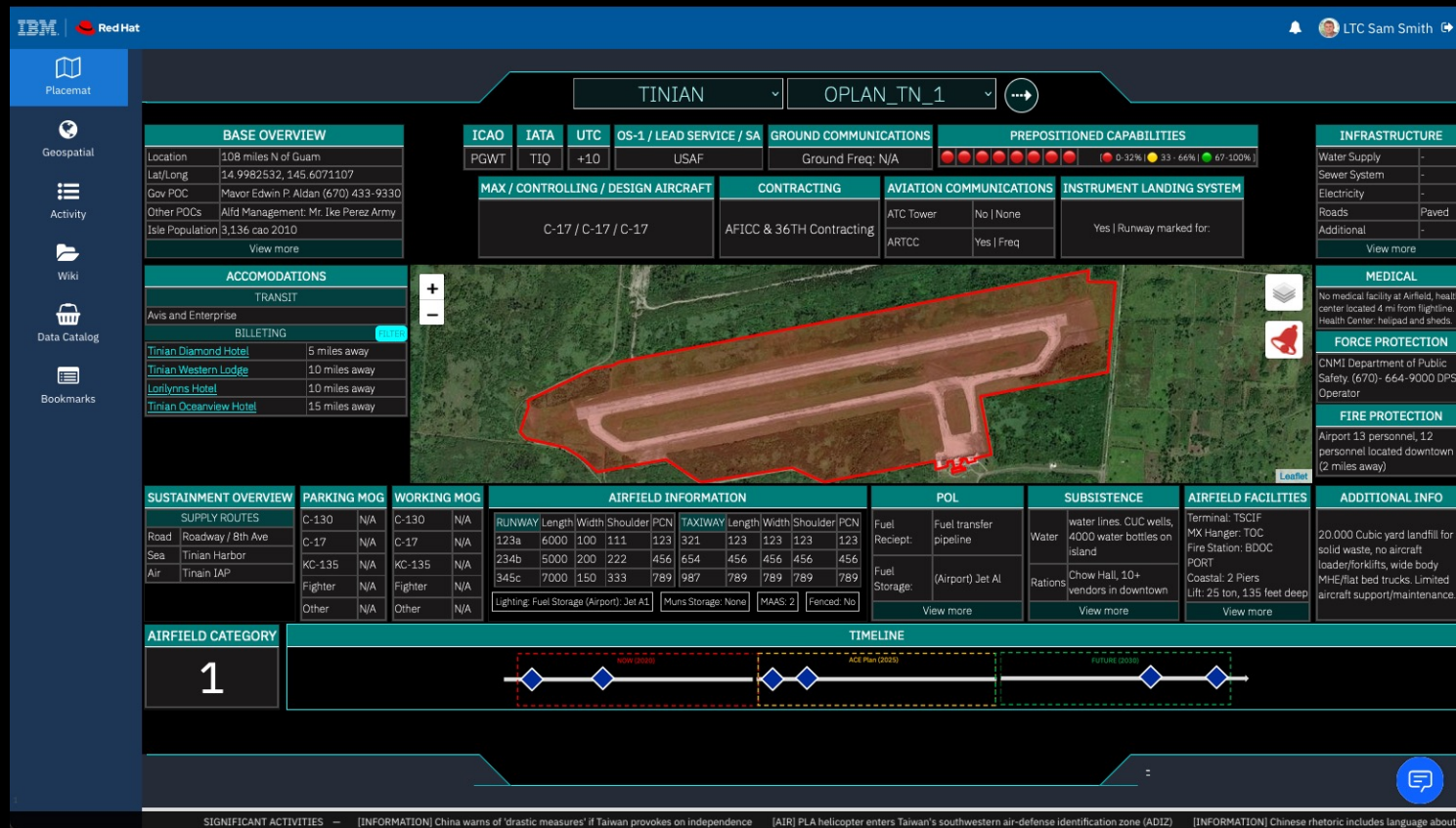
Discover Unknown Air Traffic Patterns *(by Sandia Labs - Tracktable)*



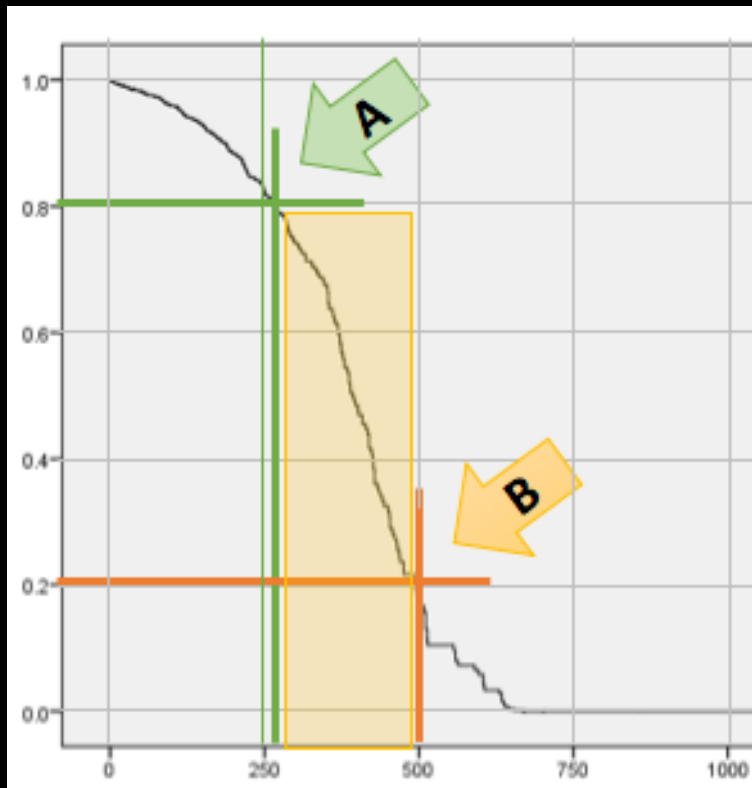
Understand your Inventory and Supply Chain



Deployment Site Ability to Support Mission



Cox Regression Results – Part 101



- Graph depicts Part 101 **“Life Span”** probability
 - **“A”** denotes 80% will work for at least 255 Hrs
 - **“B”** denotes 20% will work for at least 500 Hrs
- Example thresholds for part replacement
 - **Left of A** : 80 % survival ... never replace
 - **Inside Box** : 80 % to 20% survival ... replace as *per business rules and just in time factors*
 - mission planning, cost, supply availability, etc
 - **Right of B** : 20 % survival ... *always replace*

Aircraft Engine Fault Prediction for Fault 111

- Analysis of 17 flights (Tail # 1)
- A high probability of a fault 111 seen 3 flights prior to actual fault event
 - Fault code 112 occurred on flights 15 thru 17

