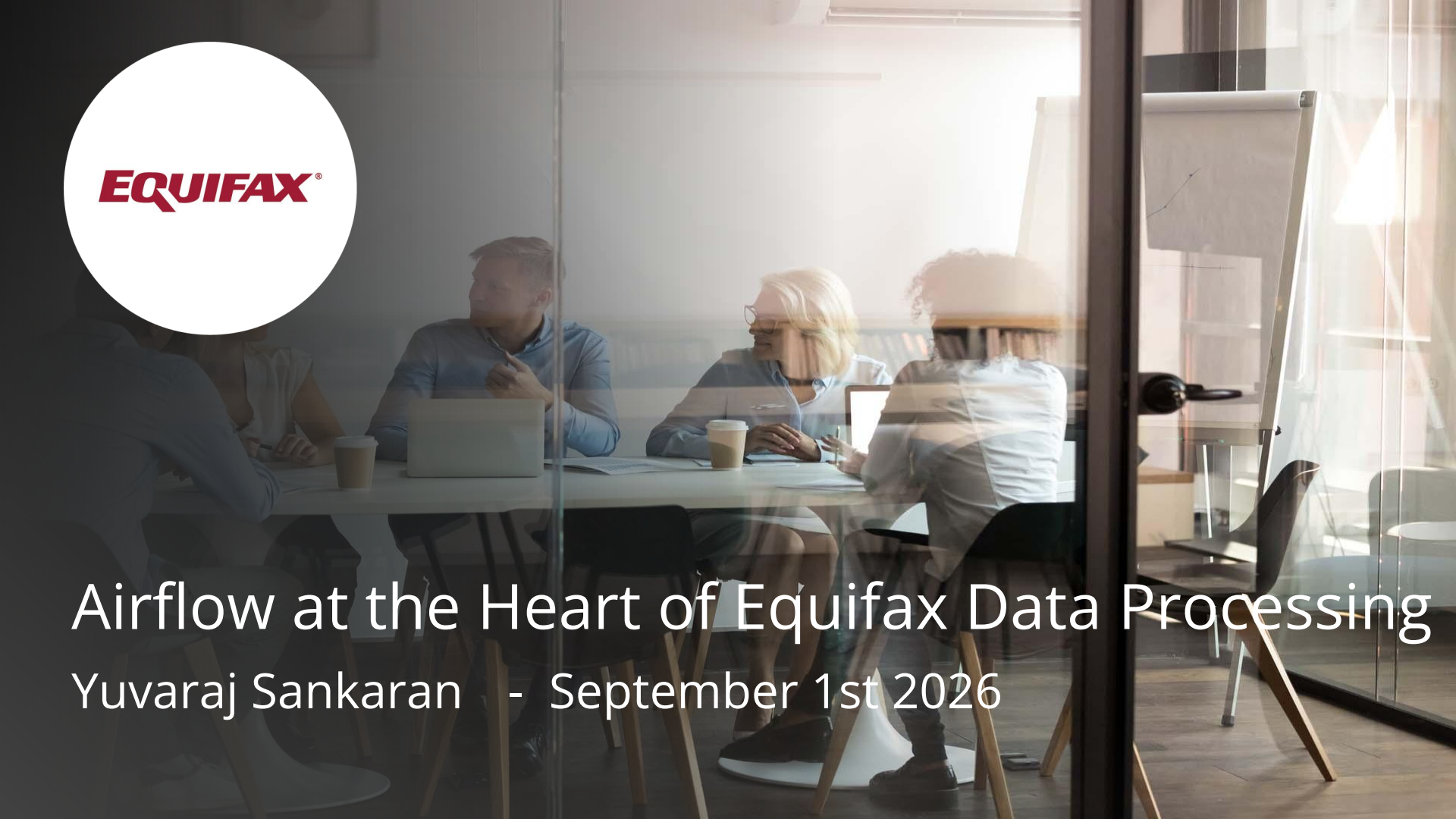




EQUIFAX®

Airflow at the Heart of Equifax Data Processing

Yuvaraj Sankaran - September 1st 2026



We play an essential role in the economy by helping companies in diverse industries such as automotive, communications, utilities, financial services, fintech, healthcare, insurance, mortgage, professional services, retail, e-commerce, and government agencies, make critical decisions with greater confidence.

Our unique blend of **differentiated data, analytics, and technology** lets us create the insights that power decisions to move people forward. We help businesses provide a seamless and positive experience during life's pivotal moments — like applying for a job or mortgage, financing an education, or buying a car.

Equifax, a global data, analytics, and technology company

	15,000 Employees
	24 Countries
	Atlanta, GA, U.S. Headquarters
	Since 1899 Brand Heritage
	EFX NYSE Ticker Symbol
	\$6.1 B USD 2025 Revenue
	40% Diversity In U.S. Employees
	\$2.6M USD 2025 Philanthropic Giving

Part of EFX Global Platforms



Real-time insights and decisions

Multi-data assets at scale

Faster and more efficient products to market

Embedded AI capabilities

Improved speed and performance

Always-on stability

Built-in security

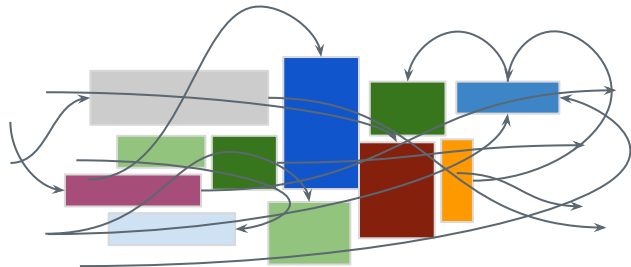
DATA FABRIC

Why Equifax built a new Data Fabric?

And what problems are being solved.

Pre-Data Fabric - Pipeline Jungle

Equifax Data Assets built over last few decades were purpose built for the specific data domains and employed classical technologies. These systems were monolithic, hard to maintain, manage and change to meet today's business needs to build multi-data products, govern and scale in an agile fashion.



From multiple siloed Exchanges and Applications with separate unshared Data stores

Post-Data Fabric

The transformation erased the technical debt and is maximizing the potential of EFX's unique data assets.

Transition from a purpose built multiple all-in-one data platform strategy to a new single data fabric on a highly scalable, secure, fully governed and flexible cloud native environment.

Organize the data based on a standardized enterprise domain model to enable seamless access to all our data assets powering new, compliant, secure Multi-Data Products.

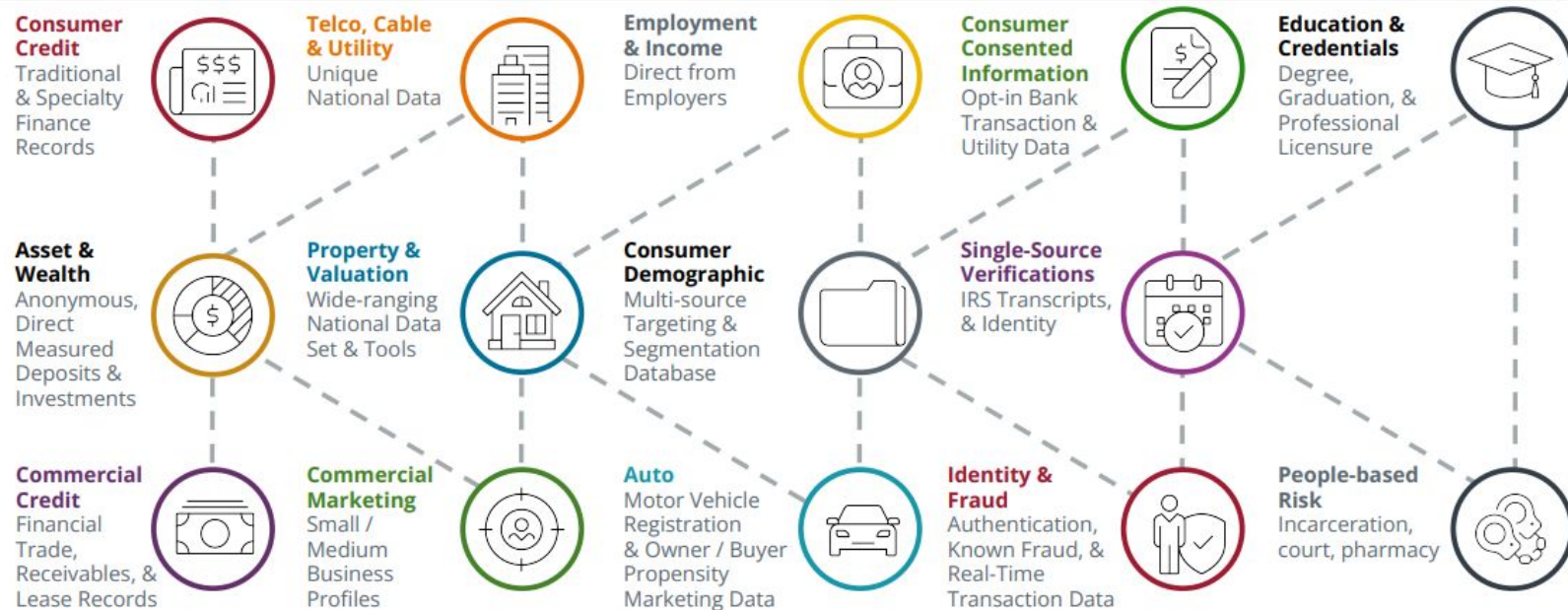


To Single Standardized Data Fabric for all Exchanges

Data Fabric Enables Multi Data Products

A Robust Foundation for Comprehensive Multi-Data Solutions

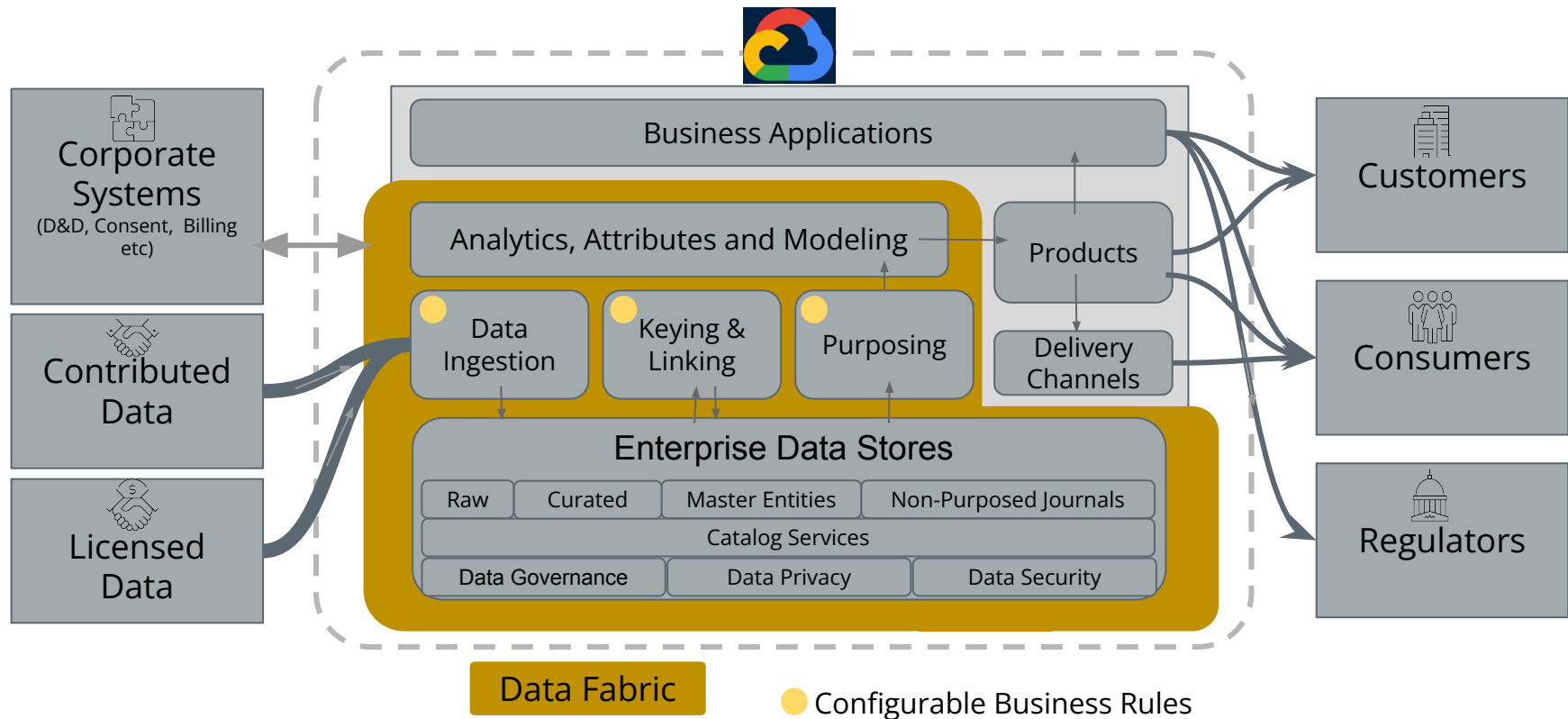
Insights Are Created Through Connected and Differentiated Data



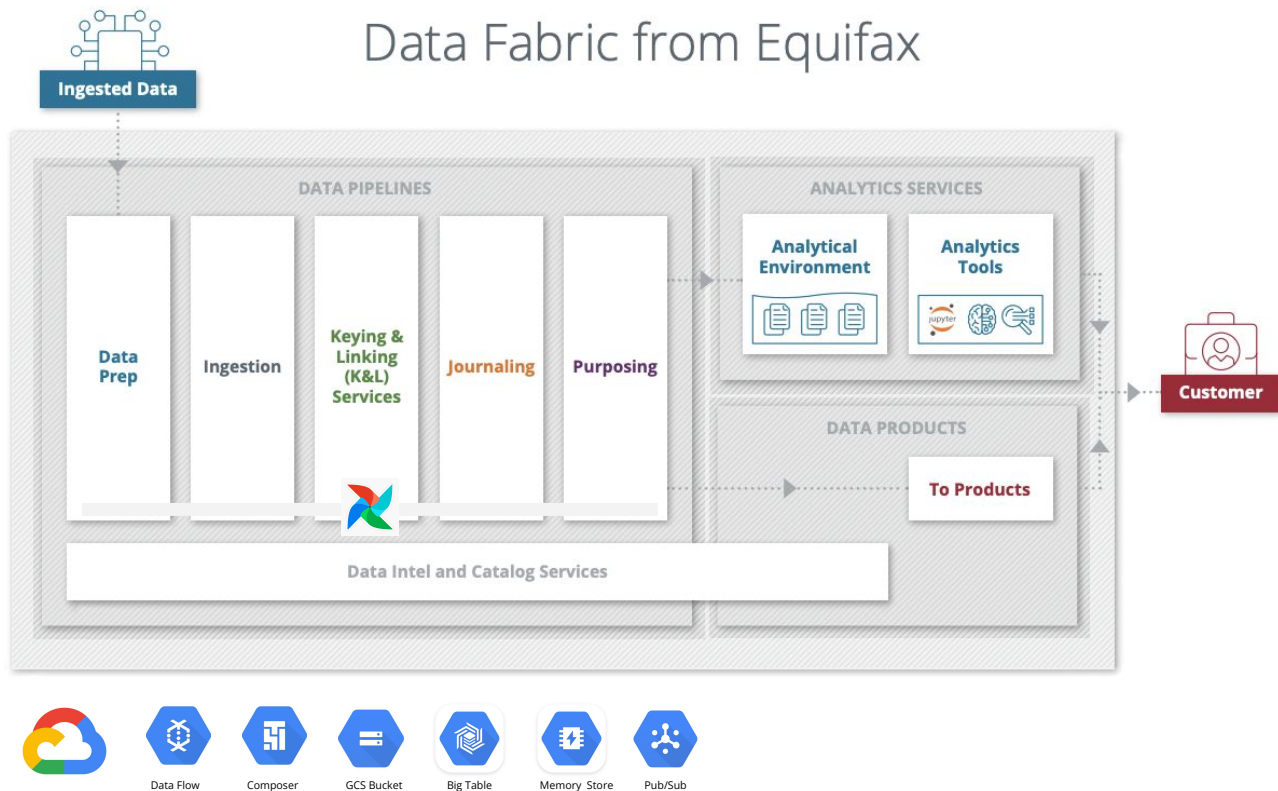
Certain data is ingested into the Equifax environment, and curated, enriched, and purposed. Moving from over 70 data silos, multiple sources of data are organized into a seamless, globally-distributed Data Fabric with logical separation and governing rules.

Data Fabric Functional Overview

Data Fabric manages the entire value chain for monetizable data



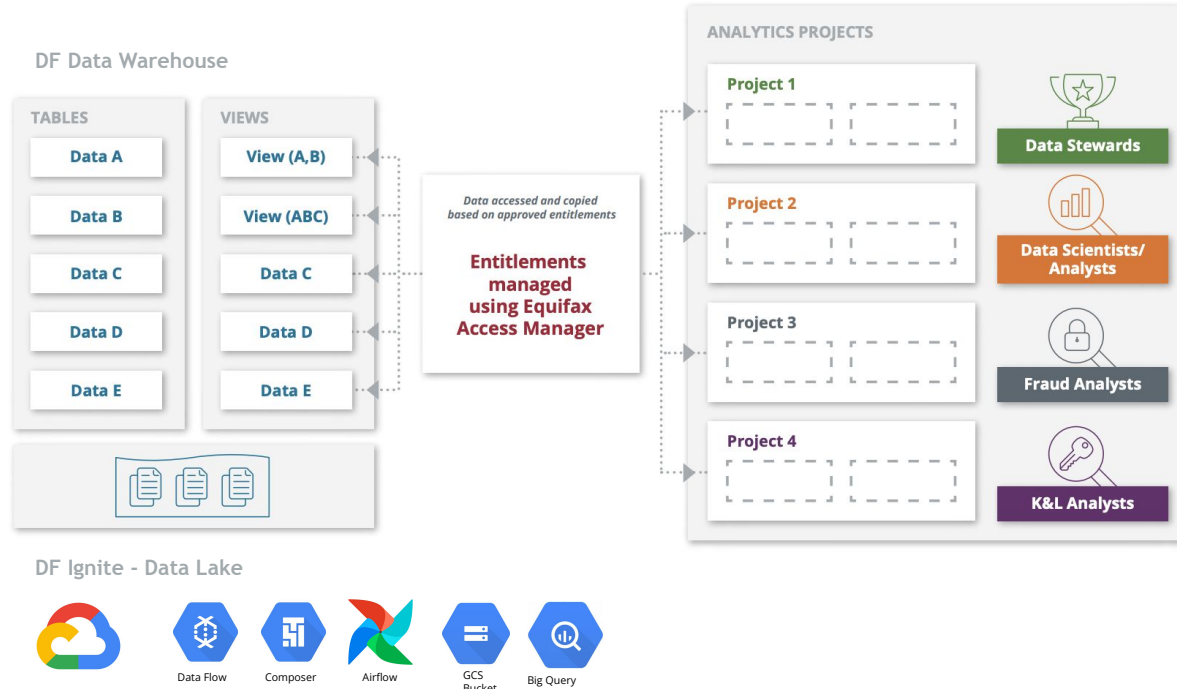
Data Fabric - Manages the EFX Data Lifecycle



- Data Fabric includes Core Systems and Analytical warehouses
- Every Equifax Exchange is a set of configurations on the Fabric
- The Data could be isolated by Source and consumption Domains
- The Data is keyed and Linked for Monetization for downstream products

Data Fabric Analytical Sandbox - Governance

The following diagram illustrates the governed data access in the analytics environment:



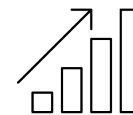
- Ignite Analytical Sandbox allows Internal and External Data Scientists to build Analytical Products using our Enterprise Data Sets in a fully Governed and Compliant Manner with Robust Security and Governance.
- Ignite enables RBAC based access to the different user roles that use enterprise Data Assets

EFX Data Fabric - Business Impact

Scale out Data ecosystem

- Petabyte scale workflow orchestration
 - 50+ Billion ingestions/month, using ~900 pipelines Globally
 - Supporting products ~5+ Billion dollars in revenue
- High throughput Ingestion/Low Latency lookup
 - ~650M+ supplier records Ingested daily
 - Supporting both Batch and Streaming Pipelines
- Business outcomes
 - Every Equifax Credit Report , Employment Verification Report and similar data products delivered by Equifax is orchestrated by Apache Airflow

Equifax Data Fabric in Numbers



>800M

Number of consumer entities



Number of entities Globally in Consumer Credit , Employment, Commercial and Utility Exchanges

6

Number of regions deployed



US, Canada, UK, Australia, India and LATAM

>500B

of Observations in US



Total number of observations in our Journal in the US

>10PB

in Data lake



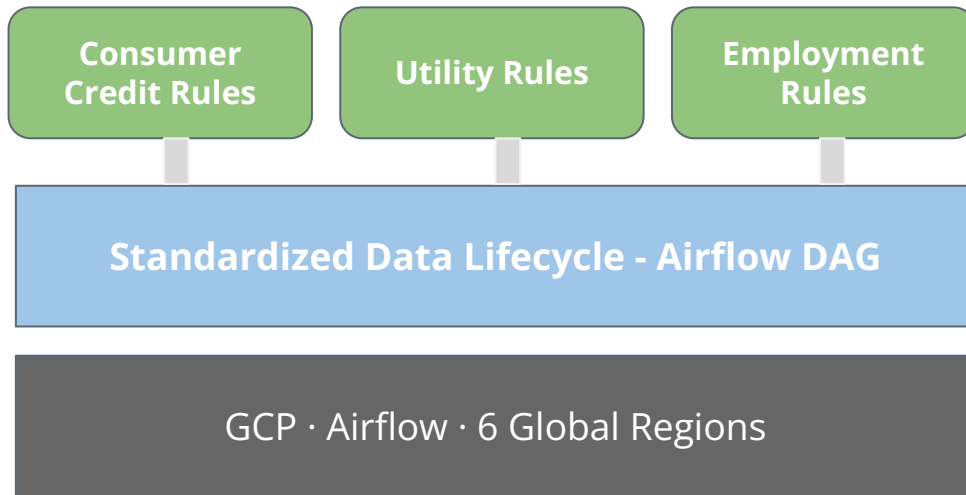
The Size of data in Data Fabric globally



Domain Agnostic DAGs

Orchestration decoupled from Business Logic

- DAGs know nothing about credit, utility, or employment data
- Domain rules plugin as configuration, owned by domain teams
- Platform team owns the pipeline; domains own the rules
- New data exchanges onboard as configs — not new codebases
- Idempotency - Key Design Principle. Allows reruns without side effects





Data Fabric - Airflow Operational Metrics

A scale out platform that is at the core of Data Processing at Equifax

14000+

Approximate
number of DAG runs
executed across all
active Geographies

100,000+

Approximate task
executions
completed per
month across all
regions

15,000+

Data Flow Jobs **Daily**
Across All Regions to
process Streaming and
Batch workloads

10 TB+

The Amount of data
processed **daily**
across all regions.

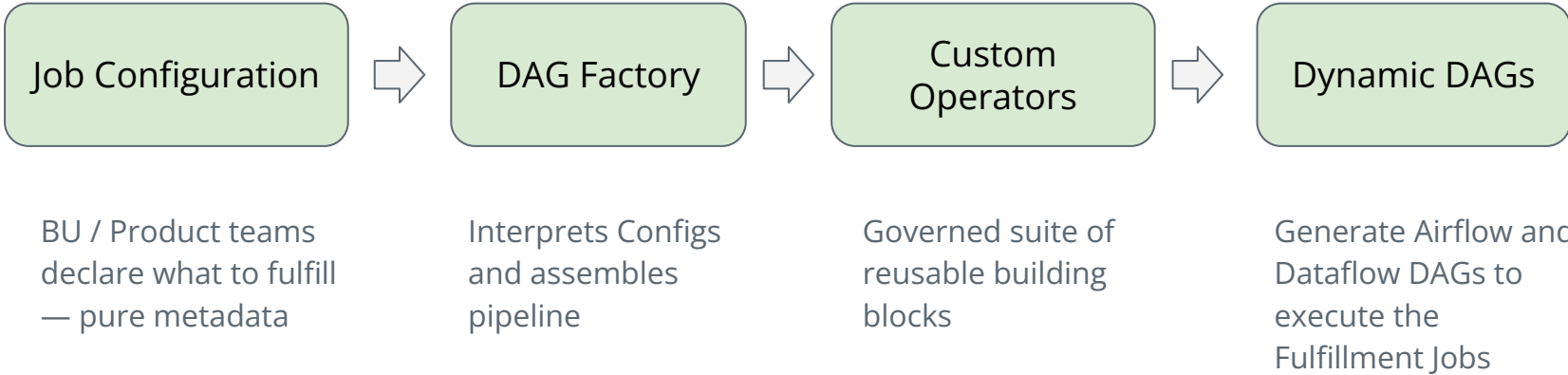
A Cloud Native, Diverse Data ecosystem that supports a large variety and volume of data processing across multiple global regions.

GLOBAL BATCH PLATFORM

EFX Global Batch Platform - Powered by Apache Airflow

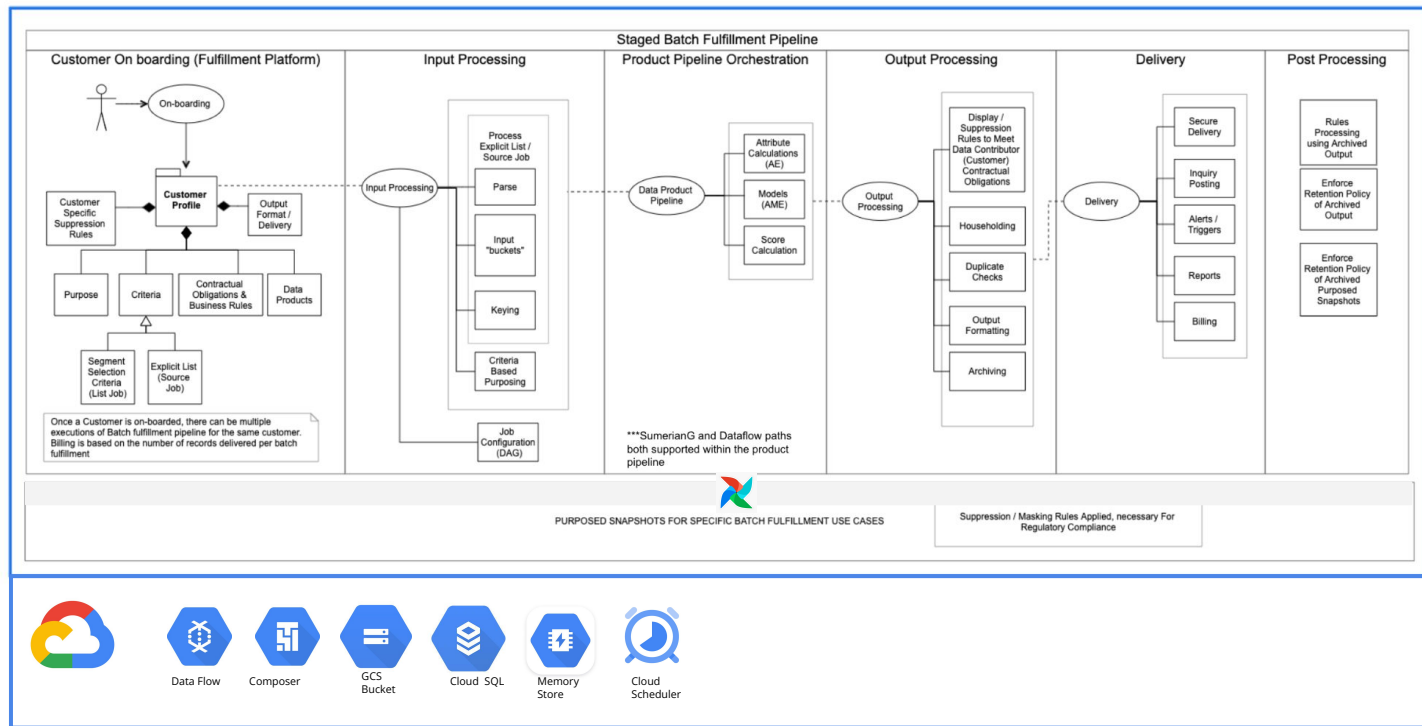


Declarative Config Driven Orchestration - Users never write Raw Dags



Orchestration as a Product: The Platform encodes best practices once - Every Fulfillment Job/Composite Product Inherits it Automatically

Global Batch Platform - Staged Fulfillment



- Petabyte Scale Offline Fulfillment Platform
- Ships Billions of Insights nightly to Customers Globally.
- Self Service Platform for EFX BUs
- Assembles EFX Data Products, Features, Scores, applying Customer Specific Criteria and Formatting
- Utilizes Nightly Snapshots from Data Fabric
- Supports more than \$100 million of Dollars in Revenue



Global Batch Platform - Airflow Operational Metrics

A petabyte scale Fulfillment Platform

32,000+

Approximate number of DAG runs executed across all active Geographies

320,000+

Approximate task executions completed per month across all regions

\$100M+

Annual Revenue generated by GBP across multiple regions

12,000+

No of **SLA compliant** Customer Fulfillment Jobs completed **Monthly**

2 Billion+

Number of **Consumer Insights** delivered monthly across all regions.

A robust scale out revenue generating platform operating under strict customer SLAs , delivering insights Globally to our Customers.

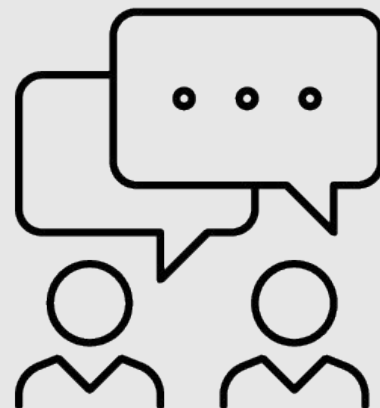
Where we are headed ?

- **Universal Context Layer** - Building the UCL that brings the different contexts together and makes it available to Agents for both onboarding and operating various workflows for internal and customer facing use cases.
- **Agent Native Operations** - That could provide more visibility, intelligent monitoring, alerting , recommendations ; That could help reduce the operational burden of running such large Implementations.
- **Self healing pipelines** - Building pipelines that have the full context of the business function ,SLA commitments, technical metadata of the Data ecosystem and helps heal production issues with minimum HITL.
- **Robust Security , Governance and Compliance around Orchestration** - As orchestration becomes more intelligent and dynamic, building Governance and Security in this layer becomes even more critical.

Airflow isn't just running our ETL. It's running the operational core of a regulated financial enterprise.

Questions?

Answers

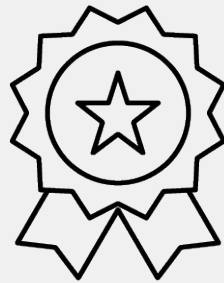


Appendix

Learnings and Best Practices



- **Deploy unified Data Fabric architecture across the Enterprise** - Resist Organizational pressure to build LOB specific siloes.
- **Don't constraint Data Fabric platform only to Analytical workloads** - Core Data systems that manage the Entire Data Lifecycle can also be part of the Enterprise Data Fabric.
- **What Data Belongs in the Data Fabric ?** - Enterprise Data Assets . Don't make the Data Fabric the database of the company. Application Data can remain within the Application.
- **Treat data as an Enterprise Asset.** Be prepared to have tough conversations on domain ownership, especially when transforming existing systems that may be biased either towards centralization or decentralization. Challenge the status quo.
- **Standard Security, Governance and Operational Disciplines** Key to maximizing reuse of the Fabric.
- **Embrace Data Regulations** - Different Data Regulations for Different Data domains could exist, however adopt the strictest regulation for all the domains.
- **Build robust Multi-tenancy Concepts into the Platform.** The Platform should be able to provide logical and physical separation of the Data throughout the lifecycle



DPKL AIRFLOW MONITORING

alarm Performance Contracts (SLAs & Timeouts)

Define strict execution expectations instead of basic failure checks:

Service Level Agreements (SLAs)

Define a `sla` parameter on tasks (e.g., must finish within 2 hours of DAG start). Missing it triggers an `SLA Miss` record in the metadata DB and can launch an `on\_sla\_miss\_callback` alert.

Execution Timeouts

Enforce hard limits with `execution\_timeout`. If a task hangs (e.g., frozen queries), Airflow forcefully fails the task, releases the worker slot, and prevents cluster resource starvation.

bar_chart Built-in Monitoring Visuals

Leverage powerful visual charts within the Airflow Web UI to spot degradation:

Gantt Charts

Provides absolute visualization of task execution sequence and duration within a DAG run. Essential for locating exact operational bottlenecks.

Task Duration Over Time

Tracks run times across multiple execution cycles. Clearly highlights gradual performance decay (e.g., query creep from 5 to 45 mins).

Landing Times

Maps total time elapsed from scheduled DAG start to individual task completion, serving as a primary metric for downstream delivery targets.



| Powering the World with Knowledge™

equifax.com