Data Driven Organization

The Modern Data Community

Steve Cooper, Worldwide Lead Data-Driven Everything (D2E)
“An organization that harnesses data as an asset, to drive sustained innovation and create actionable insights to supercharge the experience for their customers so they demand more.”
Agenda

• The Modern Data Community – breaking the monolith
• Data Producers – data product owners
• Platforms – from maintenance to customer experience
• Consumers – diversity and high velocity decision making
• Automation – invent and simplify a balance between governance and agility
• Tenets
• Getting Started
The Modern Data Organization

Breaking the monolith
A shift to **microservices**

- Decoupled architecture of single-purpose services
- Business logic and data only accessible through hardened APIs
- Increased speed, agility, and innovation
Create a Community with a Data Marketplace

Data-driven organizations enable agility by pushing responsibility to the edges, to the producers and consumers of data.

- **Producers**: “Teams that want to share data”
  - Domain expertise
  - Data ownership and governance
  - Data quality
  - Metadata Management

- **Lake House Platform**: “Team that runs the marketplace”
  - Build security controls
  - Build and run the platform
  - Simplify on-boarding
  - Enterprise datasets
  - Training and community

- **Consumers**: “Teams that want to use data”
  - Execute business priorities
  - Business analytics development
  - Data Discovery
  - Data pipeline development
  - Creation of new insights

Level of decentralization depends on maturity of skills, complexity of business, domain knowledge required, and pace of tech change.
ENGIE builds the Common Data Hub on AWS, accelerates zero-carbon transition

**Challenge**
ENGIE’s decentralized global customer base had accumulated lots of data, and it required a smarter, unique approach and solution to align its initiatives and to efficiently provide data across its global business units.

**Solution**
ENGIE built its Common Data Hub data lake on AWS, enabling the company’s business units to collect and analyze data to support a data-driven strategy and to lead the zero-carbon transition.

**Result**
- Collected 95 TB of data across 351 projects
- Automated energy predictions
- Maximized wind farm energy production

**Benefits**
Since implementing the CDH, ENGIE’s renewable fleet of wind farms, solar farms and hydroelectric dams is significantly more efficient. If you improve the availability and performance of an asset that’s worth $100 million or $500 million by just 1% because you tap into the right data—well, I’ll let you do the math.

Yves Le Gélard CDO and CIO.
Data Producers

Data product owners
Producers
"Teams that want to share data"

- Domain expertise
- Data ownership and governance
- Data quality
- Metadata Management

- Motivations
- Domain knowledge
- Metadata
- Quality and reliability
- Data access
- New skills
Product-Oriented Operating Models

**1. Re-envision the world as products**

FROM Systems TO Products

**2. Organize teams around products.**

Activity-Based Teams

- Business
- PMO
- Design
- Dev
- Mgmt
- Ops

**3. Bring the work to the teams.**

Bring Team to the Work

**4. Reduce risk through iteration.**

Large Batch

- Risk
- Jan
- Dec

Small Batch

- Risk
- Jan
- May
- Sept

© 2021, Amazon Web Services, Inc. or its Affiliates.
• Operates **thousands of micro-services** to serve **millions of customers**.
• 50 petabytes of data, 600,000 user analytics jobs each day.
• **1,800** teams producing data, **3,300** teams analyzing and acting on data.

**Challenge**
• On-prem Oracle did not scale, difficult to maintain, and costly.
• Limited businesses ability to generate insights and deploy ML.

**Solution**
• Built a data marketplace on AWS.
• Doubled the data stored (100PB), lowered costs, gained insights faster.
• **Move data across the business quickly and easily.**
• **Easily discover data** with **reduced latency** for analytics results.
Platforms

From maintenance to customer experience
Lake House platform

“Team that runs the marketplace”

- Build security controls
- Build and run the platform
- Simplify on-boarding
- Enterprise datasets
- Training and community

- Life after maintenance
- Customer centricity
- Abstract complexities
- Community building
- New skills
Transparency and predictive intelligence in production plant

Challenge
• Increased complexity in production.
• Factory operations need to be adaptable to changing sales and supply demands.
• Avoid points of failure, to fulfil customer orders consistently.

Solution
• Total Intelligent Manufacturing product providing transparency into the end-to-end production operations in real time through a single pane of glass.
• Predicts points of failure in the production process ahead of time and integrates into digital plant simulations.
• Creates a broader data collaboration capability connecting manufacturing with supply chain and customer retail.

Planned Benefits
• Increased plant productivity and steady state daily throughput.
• Equipment effectiveness with reduced outages and quality issues.
• Process agility and enhanced customer experience.
Consumers

Diversity and high velocity decision making
Consumers
“Teams that want to use data”

- Diverse personas
- High velocity decisions
- Data discovery
- Native access
- New skills

- Execute business priorities
- Business analytics development
- Data Discovery
- Data pipeline development
- Creation of new insights
BMW Group uses AWS-based data lake to unlock the power of data

Challenge
BMW Group’s rigid on-premises data lake was challenging their ability to scale to meet demand and accessing siloed data required long lead times.

Solution
BMW Group decided to re-architect and move its on-premises data lake to the AWS Cloud—using a serverless architecture that offered agility, flexibility, and a modern web portal give users across the globe access to data.

Result
• Democratizes data usage at scale
• Processes terabytes of telemetry data from millions of vehicles daily
• Resolves issues before they impact customers
• Accelerates innovation
• Training 5,000 software engineers and applying AWS Working Backwards methodology
• Identification of business challenges and develop new cloud-enabled solutions

“To stay innovative, we are focusing on creating new digital and connected experiences and driving change in our value chain by enabling data-driven decisions.”

Kai Demtröder, BMW Group vice president of data, AI
Automation

Invent and simplify a balance between governance and agility
- Balancing governance and agility
- Ingest at speed
- Detecting quality issues
- Understanding data at scale
- Helping consumers find data
Brent Shafer, Chairman and CEO of Cerner, Talks About Using AWS to Transform Healthcare

Overview
Cerner Corporation delivers healthcare technology globally to 3 million healthcare professionals and innovates to create a seamless and connected world in which everyone thrives. The company has spent the last four decades digitizing healthcare data and ridding clinician's offices of manila folders and filing cabinets. Now, the collaboration of Cerner and AWS will deliver data that is more accessible and actionable and uses AWS AI and machine learning technologies to predict and potentially prevent health problems. Also, as part of its effort to modernize how it delivers solutions and improves patient outcomes, Cerner has been migrating its privately hosted platforms to AWS. One of the company's goals is to bring more joy to the practice of medicine—to that end, Cerner is testing its Virtual Scribe technology using speech recognition and Amazon Transcribe Medical to dramatically reduce manual data entry and give doctors more time to spend with patients.

“
We’re excited about how this collaboration helps us move closer to Cerner’s vision. Our vision is a seamless and connected world where everyone thrives.

Brent Shafer, Chairman and CEO
Tenets
Tenets for a Modern Data Community

• We enable highly agile organizations by empowering at the edges. Empowered organizations require players to accept greater responsibility.

• Domain-relevant, high quality, discoverable, and trustworthy data is the basis of successful communities. This is the responsibility of Data Producers.

• Innovation is organic. It requires connecting ideas, data, tooling, and know-how. The Platform teams need to abstract complexity from this equation.

• High velocity decisions sustain organizations. This requires Data Consumers to “experiment patiently, accept failures, plant seeds, protect saplings, and double down when you see customer delight.”*

Getting Started
Recommendations

• Think big, start small, scale fast.
• Work backwards from customer challenges.
• Form a multi-disciplinary teams including business, technology, and data skills.
• Incentivize your data producers by creating metrics on the availability and completeness of their data.
• Build a community, celebrate success by publishing blogs and writing stories about what you’re doing.
• Automate tasks to increase adoption.
Want to build a data vision and strategy?

- Joint engagements with business and technology stakeholder alignment
- Create an organizational vision for innovation with data to drive business outcomes
- Define the first pilot, learn, and build

Jumpstart the data flywheel

Have a strategy and need help executing it?

- Joint engineering engagements between customers and AWS technical resources
- Create tangible deliverables to accelerate strategic databases, analytics, and ML initiatives
- Leave with an architecture, working prototype, path to production, and deeper knowledge of AWS services

Come with an idea, leave with a solution
Thank you
Appendix
Improving Clinical Trials to Cure Cancer: Fred Hutchinson Podcast on Amazon Comprehend

**Challenge**
The mission of Fred Hutchinson Cancer Research Center is the elimination of cancer and related diseases as causes of human suffering and death. For cancer patients and the researchers dedicated to curing them, time is the limiting resource. The process of developing clinical trials and connecting them with the right patients requires research teams to sift through and label mountains of unstructured clinical record data.

**Solution**
With Amazon Comprehend Medical's Entity extraction API built for Health, Fred Hutch can extract disease conditions, medications, treatment outcomes, or PHI from medical records for each patient to measure a operational metric.

“When I think about the real win of natural language processing and machine learning in the clinical domain, it’s really about speeding things up.”

Emily Silgard, Data Science Manager

Company: Fred Hutchinson Cancer Research Center
Country: US
Employees: 3,500
Website: FredHutch.org

About Fred Hutch
At Fred Hutchinson Cancer Research Center, home to three Nobel laureates, interdisciplinary teams of world-renowned scientists seek new and innovative ways to prevent, diagnose and treat cancer, HIV/AIDS and other life-threatening diseases. Fred Hutch’s pioneering work in bone marrow transplantation led to the development of immunotherapy, which harnesses the power of the immune system to treat cancer.
JP Morgan Chase: Enterprise wide innovation with data + control

“Most modern organizations recognize that their data benefits their entire enterprise. Data has value to the individual business process that produces it, but data’s additional potential can be realized when it’s combined with other data assets.”

Anu Jain – Head of Enterprise Data Technology

- JPMC is comprised of multiple lines of business (LoBs) and corporate functions (CFs) that span the organization.
- The regulated nature of the industry requires effective data risk management with controls to mitigate exposure.

Challenge

- Needed to enable data consumers across JPMC’s LoBs and CFs to more easily find and obtain the data they need.
- Whilst maintain control and visibility of data usage.

Outcome

- Enabled easy discoverability and data sharing across the enterprise
- Gave data owners control and visibility to managing their data effectively
- Cataloguing provides a single point of visibility for where data is used

Balance governance with Agility

A health IT company providing solutions to empower clinicians, and patients, is a great example of modern data governance automation.

Solution

• Ingest widely distributed patient and hospital data
• Near-real-time predictions about patient care and hospital operations e.g. hospital capacity and length of patient stay
• Data privacy is central to the solution, allowing researchers to work on de-identified patient health data
• HIPAA compliance to adequately safeguard protected health information (PHI).

Managing 27,000 facilities
Optimize unique needs of 150M individuals
Aggregating EHR, Claims, and Personal Data to create a consolidated, longitudinal personal record

Personalized reminders
Pace of innovation reduced from 6 months to 4 weeks