Real World Big Data Use Cases – What other CDO’s are already doing with Hadoop

Chris Moore, Principal Solution Engineer, Trifacta
Steven Totman, Industry Leader Financial Services, Cloudera
What is this?
Data is your most important asset

...only if...

your data can help achieve your goals and objectives.

Asking Bigger Questions
Data is Transforming Business
How do you want to leverage your data assets?

Increase Revenue
How do I build a 360 degree picture of my customer to deliver new revenue streams?

Reduce Costs
How can I build better data-driven products and services, at lower cost?

Manage Risk & Compliance
How can I meet compliance regulations and preserve data security to minimize our corporate risk profile?
Adoption of Cloudera Enterprise in Financial Services

Stages of Adoption

Modernize Architecture
- ETL Offload
- Operational Data Store
- Storage/Archive Optimization
- New Data Sources - Structured / Unstructured

Lower Business Risks
- Risk Aggregation and Modeling
- Regulatory Compliance
- Fraud
- Trade Surveillance
- Algo back-testing

Improve Product & Services Efficiency
- Trade & Investment Research
- Tick Data Storage
- Internet of Things

Drive Customer Insights
- Proactive Care/Customer Journey Science
- Churn Analysis
- Personalization
- Up-sell/Cross-Sell
- Next Best Offers
- Contextual Offers & Services

New & Rapidly Growing
- Cyber Security

Value Drivers
- Initial Adoption/Use Case
- Expansion into Strategic Initiatives
- Continued Expansion into Strategic Initiatives
- Growing Focus on Revenue Generation
- C Level / Board Level Priority

All Segments of Fin Serv
- Investment Banks, Asset Managers, Exchanges, Retail Banks & Payment Processors
- Investment Banks, Asset Managers, Brokerages, Exchanges
- All Segments of Fin Serv

© Cloudera, Inc. All rights reserved.
What Happens Today With New Customers

Drive Customer Insights
- Customer Journey

Operationalize

Modernize Architecture
- Collect, Create

Explore, Enrich, Analyze

Improve Product & Services Efficiency

Lower Business Risks
Where is Your Data?

Traditional Data Plus New Data Sources

Key Structured Data Sources/ Systems
- Core banking Systems
- Risk Data
- CRM
- Website and log data
- Payment/transactions
- ATM Data
- Other/Legacy Platforms

Key Un/Semi-structured Data Sources
- Clickstream Data
- Social Media
- Customer Sentiment
- Location/e-Comms Data
- Machine Data
- Apps Data
- Documents, Videos, Call Center Records
Putting Big-Data to Work Across Financial Services
Key Use Cases for Cloudera Enterprise Built on Hadoop

- Drive Customer Insights (Customer Journey / 360)
  - Up-sell/Cross-sell
  - Next Best Offer
  - Personalization
  - Customer Churn Analysis
  - Proactive Customer Care

- Risk & Compliance
  - Risk Aggregation & Modeling (VaR...)
  - Algo Testing
  - CCAR, KYC, BCBS239, OATS...

- Fraud
  - AML Payments
  - Fraud
  - Insider Threats
  - Cyber Security

- Product & Services Efficiency
  - Investment Research
  - Wealth Management
  - Trading

- Modernize Architecture
  - ETL Offload, Archive & Storage Optimization

© Cloudera, Inc. All rights reserved.
Customer 360
Example: Customer 360 / Journey Profile

- Who are you?
- Where are you?
- What have you purchased?
- Channel usage analysis: credit or debit?
- Payment mode (ATM, cards)?
- What can you afford?
- What wallet share do we own?
- What is your next big purchase/move (car, new business, home, college)?
Customer 360 / Journeys

Raw Data
- Unstructured IVR Logs
- Web Logs
- Structured Agent Logs
- Retail Desktop Data
- Mobile, Social etc.
- Segmentation, DQW, Transactional Data
  - Product, Location, Tenure

Events
- Instant Transform
  - IVR Prompts
  - Transaction Data
  - Agent Steps
  - Retail Data

Paths
- All Events
  - Models representing Events and the traffic of activity between Events are generated

Tasks
- Completions &
- Paths to Outcomes
  - Journeys
    - Enrollment Journeys
      - Kiosk Visit
      - Failed Web Pay
        - Enrollment
      - IVR Call
      - Transfer to Agent
      - Web Pay
        - Success
      - Web Pay
        - Confirmed
      - CSAT & NPS Score
      - Retail Store Visit
    - Churn Journeys

© Cloudera, Inc. All rights reserved. 11
Increase Customer Satisfaction, Loyalty, Upsell Rates

Challenge:
- Disparate view of customers
- Unable to correlate data across omni-channel touch points to understand the customer journey

Solution:
- Reduces customer complaints by 25%
- Analyzes 100% of the issues and complaints, customer transactions and behavior across omni-channel touch points
- Increases customer satisfaction by 2X
- Saved RBS 7.5 Million pounds
Fraud
Real-Time Fraud Detection & Prevention with Cloudera

- Apply Machine Learning Techniques
- Run Natively on Hadoop
- Enable Real-Time Fraud Analytics
- Integrate Threat Visualization Tools
- Reduce False Positives
Risk & Compliance
Risk & Compliance Current Data Architectures
Risk & Compliance Enterprise Data Hub
Unlimited data. Diverse access. One platform.
Opportunity

*Drive faster regulatory reporting & improve data lineage visibility*

- Citco’s regulatory reporting team is responsible for generating reports that prove compliance with relevant laws and regulations
- Their team evaluated new technology & process to improve the efficiency of creating these reports & provide greater transparency into the data behind them

Challenge

*Scaling processes while reducing potential liability issues*

- Citco had been working in Excel, which was inefficient and required lots of manual intervention - making it very difficult to show how final numbers were calculated
- Citco relied on one person to manage highly-sensitive data, which was a huge liability for the company

Solution

*Trifacta accelerates report creation & more transparent data lineage*

- With Trifacta, Citco has scaled their regulatory reporting across the team, leading to faster insights and clear lineage for audits
- Adopting Cloudera simultaneously means they now have a centralized architecture to store, access, and analyze a variety of client data
Builds Holistic Picture of US Market By Looking at 30BN Events/Day

Challenge:
- Overseeing transactions from more than 4,100 firms incl. exchanges, brokers-dealers & trade reporting facilities
- Difficult and costly to aggregate and analyze increasing volume of data from numerous sources incl. orders, quotes and trades

Solution:
- Built market event graph database using EDH
- Provides interactive access to graph data for investigations
- Using EDH on-premise and in the cloud
- Monitoring and analyzing transactions to detect fraud, insider trading, short sale, best execution
- Savings of $10-20M annually
Enabling the Analytical Community
Over the past few years, **Self-Service Data Preparation** has emerged as a stand-alone category.

Not simply a feature of Business Intelligence tools.
Self-Service is Now Expected

How can I access the raw data?

I can’t tell you until I see the data – let me see the data first.

What do you want to analyze?

I can’t just point you to the raw data – you’ll need to tell me.
What’s the Cause? Existing Tools...

Hand-Coding

Mapping-Based ETL
What’s Required to Bridge the Gap?

Empower Users

Govern Data & Processes
Trifacta’s Approach: Experience Matters
Interactively Explore Diverse Data
Intelligence to Guide Transformations
Maintain Governance While Empowering Users

**ORGANIZATIONAL REQUIREMENTS**

- Metadata & Lineage
- Operationalization

**TECHNOLOGY REQUIREMENTS**

- Security
Cyber
Forbes

Federal Union Says OPM Data Breach Hit Every Single Federal Employee

U.S. investigating report email account linked to CIA director hacked

The Telegraph

Vatican investigates mystery over hacked computer belonging to finance chief

BBC

Sony pays up to $8m over employees' hacked data
The environment has changed...

- Hackers are more sophisticated
- Attacks are more frequent
- The world is hyper connected
Powering the next generation of cybersecurity

- Aggregated Events
- User Data
- Raw System Logs
- Network Flows/DNS
- Full Packet Capture
- Video, Text, Images

Data Types (MBs>PBs)

Time (Months)

Apache Hadoop Based Applications (PBs)

SIEM (TBs)

Native Analytics
(Basic > Advanced)

- Search
- Correlations
- SQL
- Advanced Statistics
- Machine Learning

Based Applications
(PBs)

Apache Hadoop Base
Join the community today:
Open-Network-Insight.org
Thank you