

Mainframe + Hadoop

Bridging the Gap Between Big Iron and Big Data

Matt Brandwein, Director, Product Marketing, Cloudera

Jorge A. Lopez, Director, Product Marketing, Syncsort



Mainframes | A Critical Source of Big Data

71%
Fortune 500

30 Billion
Bus. Transactions / day

Top 25
World Banks



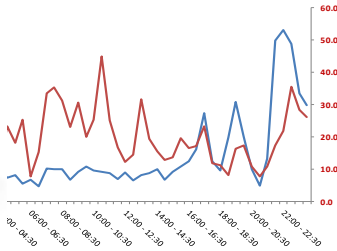
9 of World's
Top Insurers



23 of Top 25 US
Retailers



How Mainframes Work



Expensive Processing

- Every month you pay based on CPU utilization, commonly measured in MIPS
- Any MIPS reduction = Instant OpEx savings



Expensive Storage

- Storage can reach up to \$100K/TB
- Tape technology is often present
- Uses compressed data types (EBCDIC, packed decimal)



\$15.7 Million

Typical MIPS costs for the “average” \$10B organization

The Hadoop Opportunity

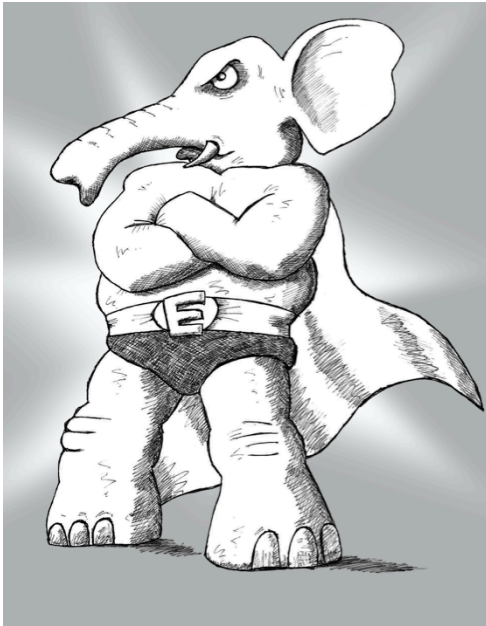
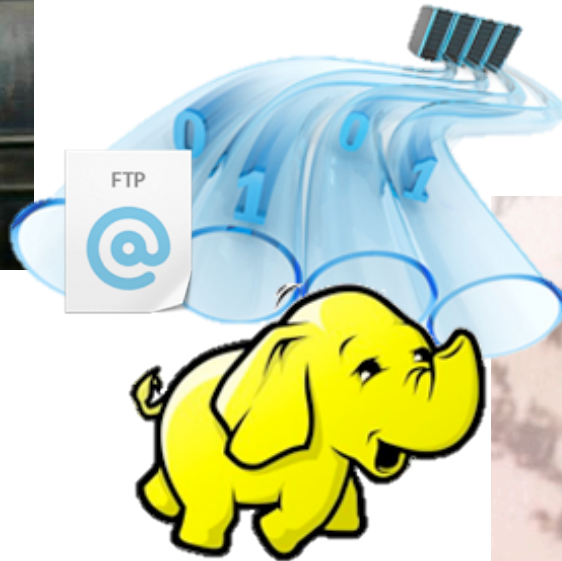


Image: quirkyjoe.com

Produce **POWERFUL INSIGHTS** by combining behavioral data with granular mainframe data

Increase **BUSINESS AGILITY & REDUCE COSTS** by offloading mainframe data & batch processing to **HADOOP**

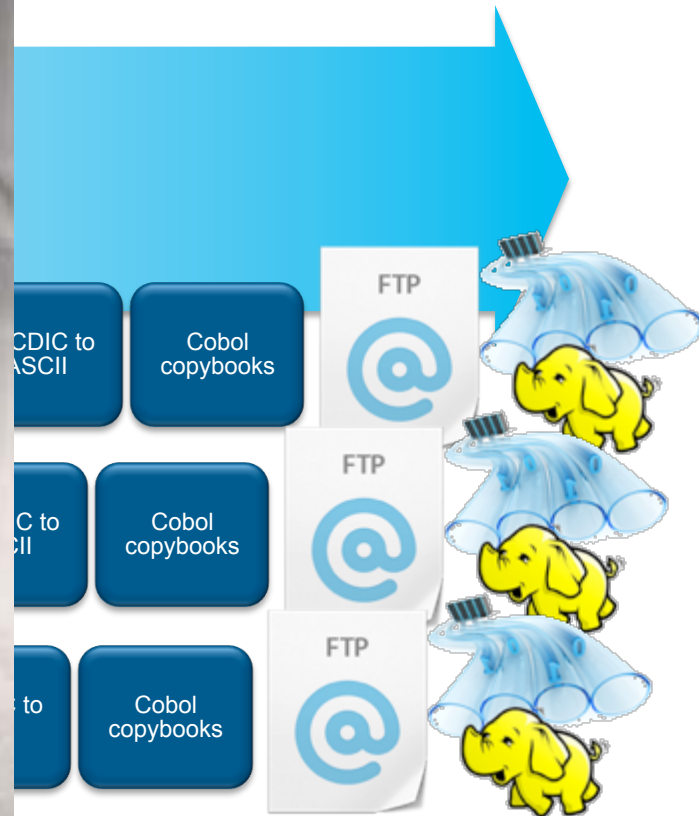
Perception: Just Call the Mainframe Guy...



Images: <http://monkeestv.tripod.com/BatMonkee/>

Reality: Far Away... So Close!

Reality



Every Change = Time, Cost

Suits & Hoodies – The Most Unlikely Duo



Integration Gaps

- Connectivity
- Data conversion (EBCDIC vs ASCII)

Expertise Gaps

- COBOL appeared in 1959, Hadoop in 2005
- Mainframe & Hadoop skills shortage

Security Gaps

- Hosts mission critical sensitive data
- Very difficult to install new software on the MF

Costs Gaps

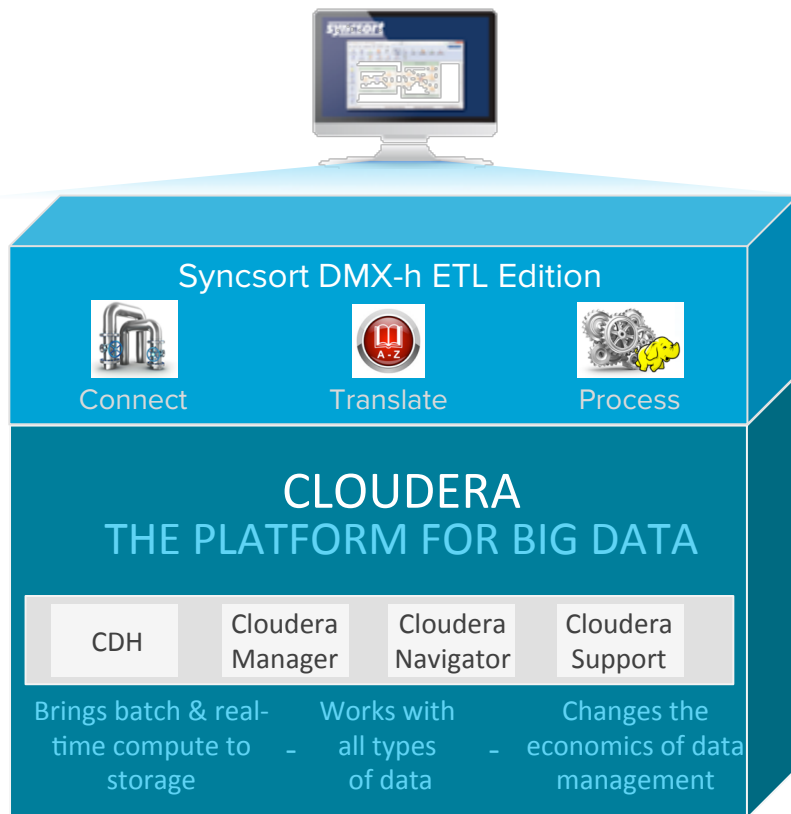
- Mainframe data is (expensive) Big Data
- Even FTP costs CPU cycles (MIPS)

Bridging the Gap Between Big Iron and Big Data

cloudera
Ask Bigger Questions



INTEGRATING BIG DATA... SMARTER
syncsort

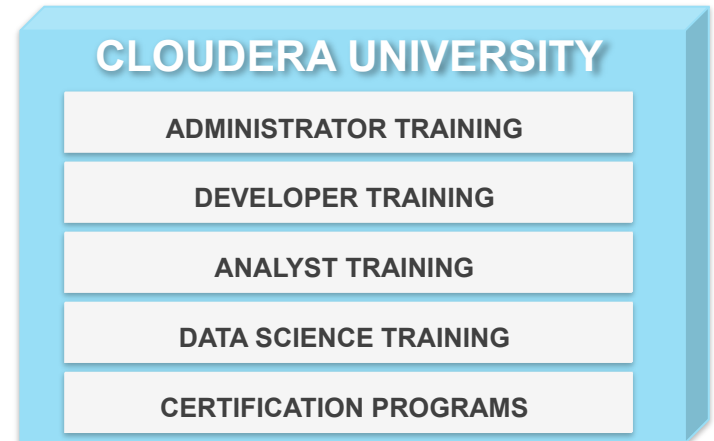
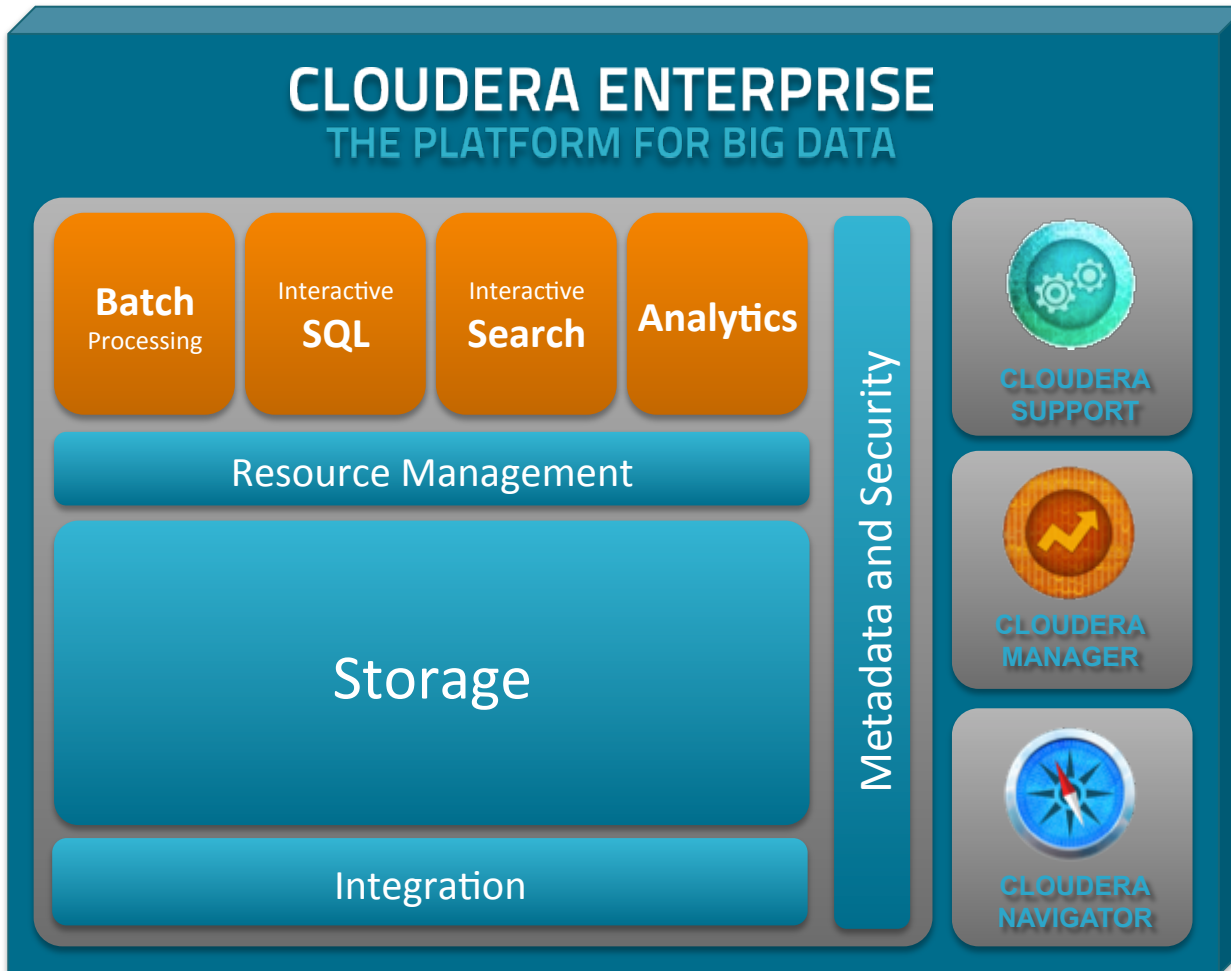


A SMARTER APPROACH TO BIG MAINFRAME DATA!

- ✓ Zero-MIPS Connectivity
- ✓ Painless Integration & Translation
- ✓ Mainframe-like Performance & Reliability
- ✓ Massively Affordable Scalability
- ✓ Support for Offloading Batch Cobol & JCL Processing to Hadoop
- ✓ Iron-clad Security
- ✓ Decades of Proven Mainframe Expertise
- ✓ Easy Deployment, Monitoring & Admin

Cloudera

The Standard for Apache Hadoop™ in the Enterprise



Why Cloudera?

We Deliver Customer Success with Hadoop in the Enterprise

- ✓ **Assurance** – Remove Risk
- ✓ **Expertise** – Maximize Value
- ✓ **Influence** – Build the Future

- **Customers:** Over 50% of the Fortune 50 and 65% of the Fortune 500 plus top US intelligence and defense agencies
- **Partners:** 700+ in hardware, software, and services
- **Education:** 15,000+ trained annually; developers, admins, analysts, data scientists
- **Community:** Founders and top supporters of the Hadoop open source ecosystem working for you

Meets Enterprise Requirements

| | | |
|-------------------------|------------------------|---|
| Data Access | Interactive SQL | ✓ |
| | Interactive Search | ✓ |
| | SAS, R Integration | ✓ |
| | Resource Management | ✓ |
| Enterprise Capabilities | Security | ✓ |
| | Highly Availability | ✓ |
| | Disaster Recovery | ✓ |
| | Audit and Lineage | ✓ |
| | Online Upgrades | ✓ |
| | Change Mgmt & Rollback | ✓ |

Why Syncsort?

For 40 years we have been helping companies solve their big data issues...
even before they knew the name Big Data!

Integrating Big Data... Smarter!

- 50% of all mainframes run Syncsort
- 1,500 Mainframe Customers: Most used & trusted 3rd party mainframe software
- Speed leader for ETL & Sort
- A history of innovation
 - 25+ Issued & Pending Patents
- Large global customer base
 - 15,000+ deployments in 68 countries
- First-to-market, fully integrated approach to Hadoop ETL

Our customers are achieving the impossible, every day!



Key Partners



Smart Contributions to Improve Hadoop

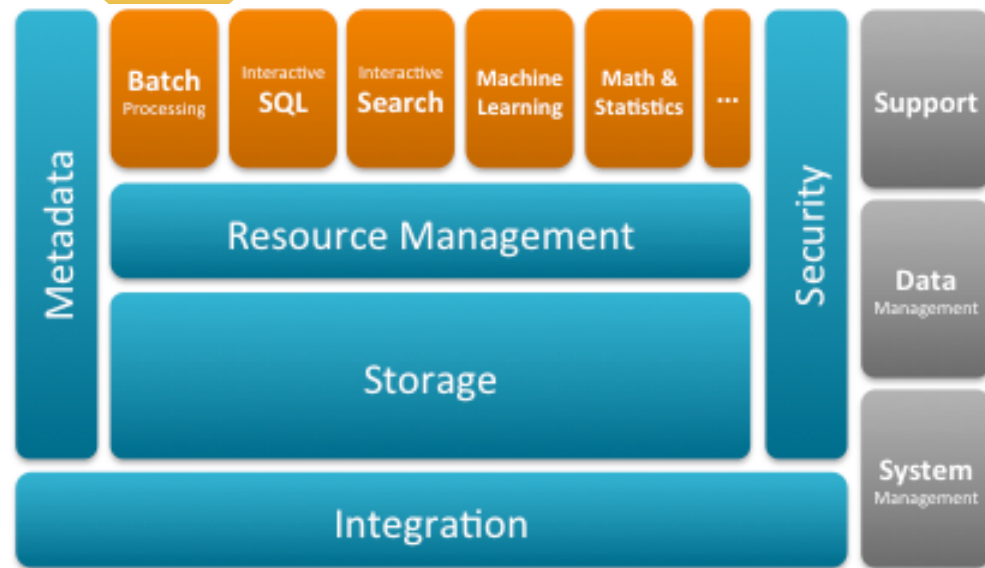
Augmenting Critical Batch Processing Capabilities



| JIRA | Description | Status |
|------|---|--------|
| 2452 | Allow External Sorter Plug-in for Map | ✓ |
| 4808 | Allow Reduce-side merge to be pluggable | ✓ |
| 4809 | Make classes required for 2452 public | ✓ |
| 4812 | Create reduce input merger plug-in | ✓ |
| 4842 | Shuffle race can hang reducer | ✓ |



...and more!!

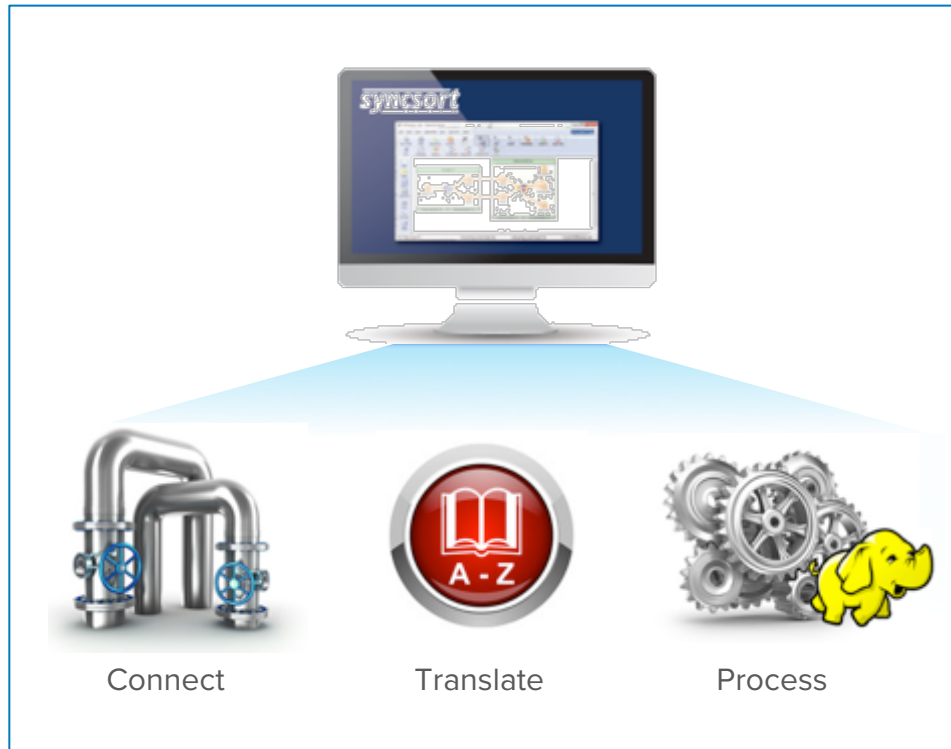


cloudera®

Cloudera Enterprise™ | The Platform for Big Data

Plugin Shipping on **CDH 4.2** and later

The Smarter Approach to Hadoop ETL... and Mainframe



- ✓ **Connect** – One tool to connect all your data
- ✓ **Translate** - Best in class mainframe data access with seamless data translation & COBOL Copybooks support
- ✓ **Process** – Hadoop ETL without coding. Develop, test & debug locally in Windows; deploy on Hadoop

PLUS...

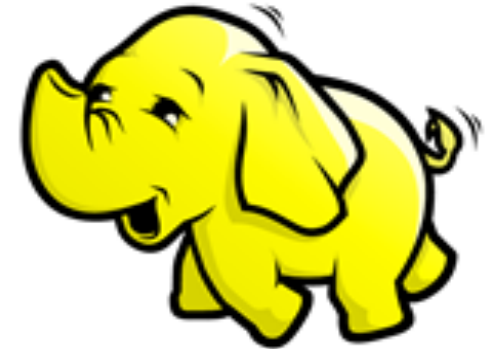
- ✓ **Enterprise-grade security**
- ✓ **Smarter deployment, monitoring & administration**
- ✓ **Disruptive cost-structure**
- ✓ **Decades of Mainframe expertise**

Integration - Why is Working with Mainframe Data So Hard?



```
01 STUDENT.
  10 FIRST_NAME PIC X(30).
  20 LAST_NAME  PIC X(30).
  30 DATE_OF_BIRTH PIC X(8) COMP.
  40 MARSH_COURSES PIC X(4) COMP.
  50 COURSES
    51 COURSE OCCURS 8 TIMES DEPENDING ON MARSH_COURSES.
    60 COURSE_TITLE PIC X(40).
    70 ENROLLMENT_ID PIC X(4).
    80 MARSH_ASSIGNMENTS PIC X(4) COMP.
    90 ASSIGNMENTS OCCURS 8 TIMES DEPENDING ON MARSH_ASSIGNMENTS.
    91 ASSIGNMENT_TITLE PIC X(40).
    92 SUB_DATE      PIC X(4) COMP.
    93 GRADE         PIC X(2).
  60 MARSH
    71 BOOK OCCURS 3 TO 1 TIMES DEPENDING ON MARSH_BOOKS.
    80 TERM        PIC X(2) COMP.
    90 RETURN_DATE PIC X(8) COMP.
```

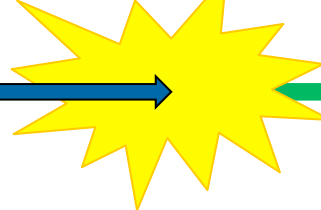
File Definitions (Metadata) in Cobol Copybooks



EBCDIC

A B Y Z
x'C1' x'C2' x'E8' x'E9'

Conversion



ASCII

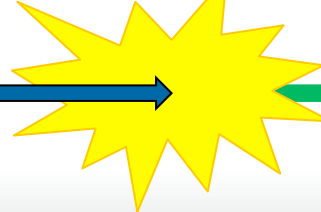
A B Y Z
x'41' x'42' x'59' x'5A'

Packed Decimal

x'12843154976C'
6 bytes

Non-Viewable, Saves space

Conversion



Numeric

12,843,154,976
14 bytes
Viewable, WYSIWYG

Cloudera + Syncsort: Smarter Connectivity... Also for Mainframe

Because Mainframe Is Big Data Too!

Connect

- Read files directly from mainframe
- No software required on mainframe
- Already installed on 50% of mainframes

Translate

- Parse & transform: packed decimal, EBCDIC/ASCII, multi-format
- No coding required

Load & Process

- Load directly to HDFS
- Offload batch data processing
- Find more insights



Iron-clad Security. Zero Pain. Zero Mainframe Footprint



Mainframe Provides Iron-clad Security...
So, Why compromise?

Other solutions

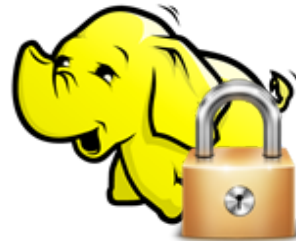


Requires you to install un-proven, untested software on **Your Mainframe**



Adopt their own security model and constantly **sync** with your own

Cloudera + Syncsort



- **Nothing** to install on **mainframe**
- Painless support for **Kerberos & LDAP**
- **User-level security** using authentication protocol

- Secure **data loads & extracts**
- Secure **job execution**

The Economics of Data

Cost of managing 1TB of data

\$20,000 – \$100,000



Mainframe

\$15,000 – \$80,000



EDW

\$250 – \$2,000



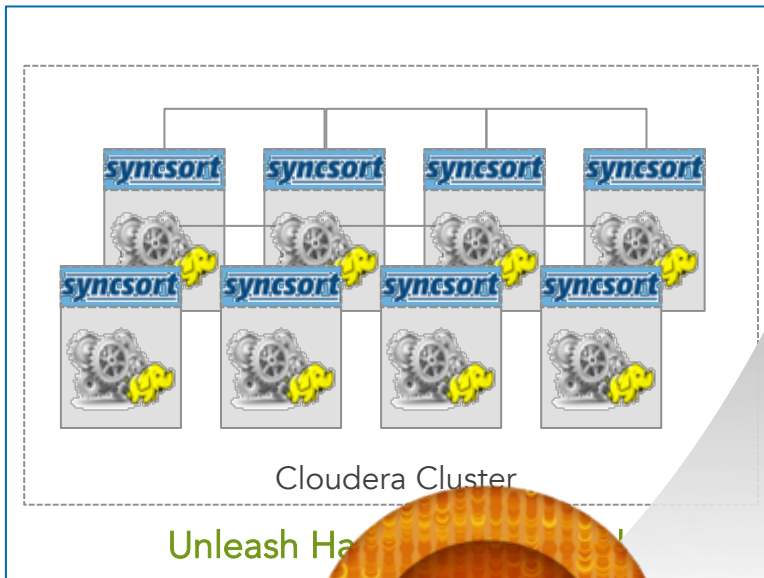
Hadoop

But there's more...

| | | | |
|---------------|---|---|---|
| Scalability | ✓ | ● | ✓ |
| Performance | ✓ | ● | ✓ |
| Reliability | ✓ | ● | ✓ |
| Agility | ● | ● | ✓ |
| Skills Supply | ● | ✓ | ● |

Smarter Deployment, Monitoring & Administration...

...through Cloudera Manager



1 Manage
Easily deploy, configure & optimize clusters

2 Monitor
Maintain a central view of all activity

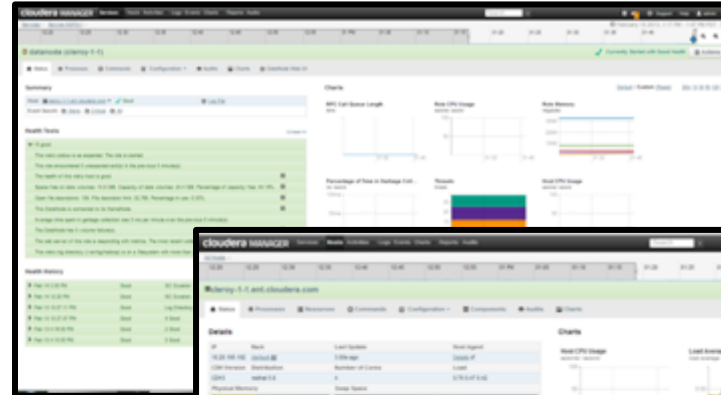
3 Diagnose
Easily identify and resolve issues

4 Integrate
Use Cloudera Manager with existing tools

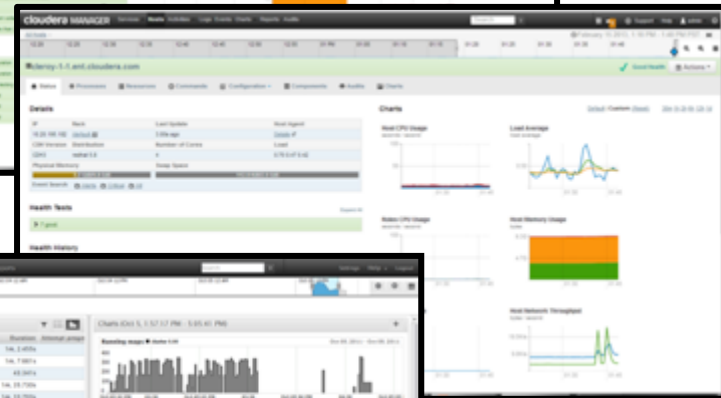
cloudera
Ask Bigger Questions

Get a 360° View of Your Cluster, Including DMX-h Logs

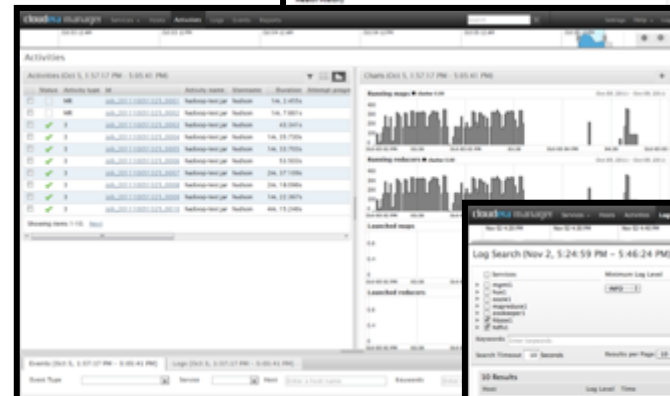
View service health & performance



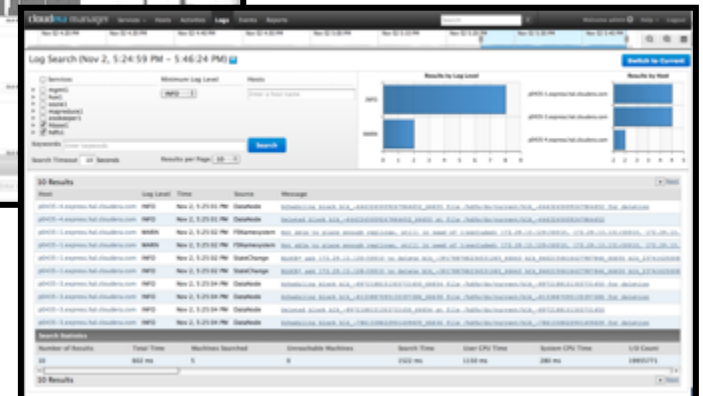
Get host-level snapshots



Monitor & diagnose workloads



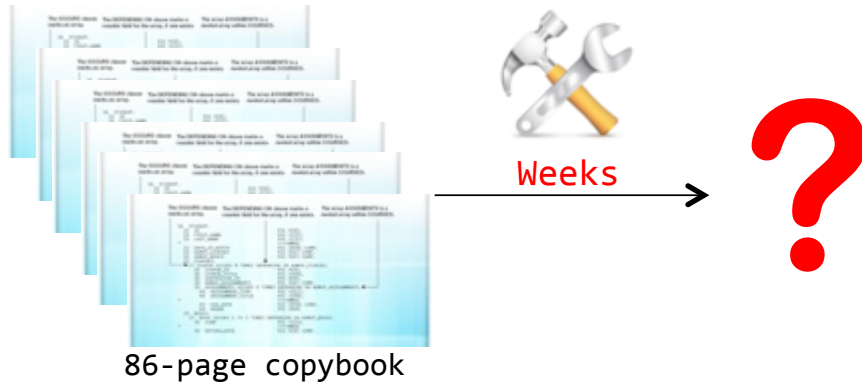
Gather, view & search Hadoop & DMX-h logs



...And more!!

Understanding Mainframe Data at Major US Bank

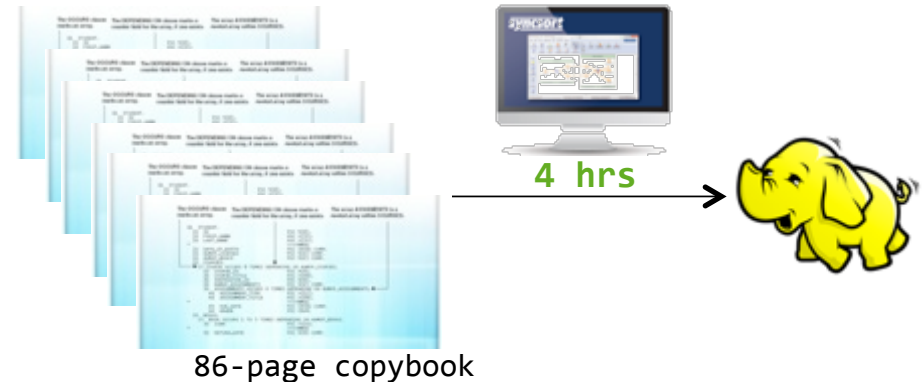
Before: Manual Effort



Customer hit a wall after months of manual effort migrating Mainframe data

- Difficult to find data errors. No Mainframe application logic that matches Copybook
- Large and complex Copybooks
- Depends on Mainframe team to provide data
- Very manual-intensive ; inadequate documentation
- Not scalable. Only a few Java + Mainframe experts could do the work

After: DMX-h + CDH



- Easy to validate Copybooks and find data errors
- Ability to pull data directly from Mainframe without relying on Mainframe team
- No coding. No scripting. Easier to document, maintain & reuse
- Enables developers with a broader set of skills to build complex migration jobs.

Three Quick Takeaways



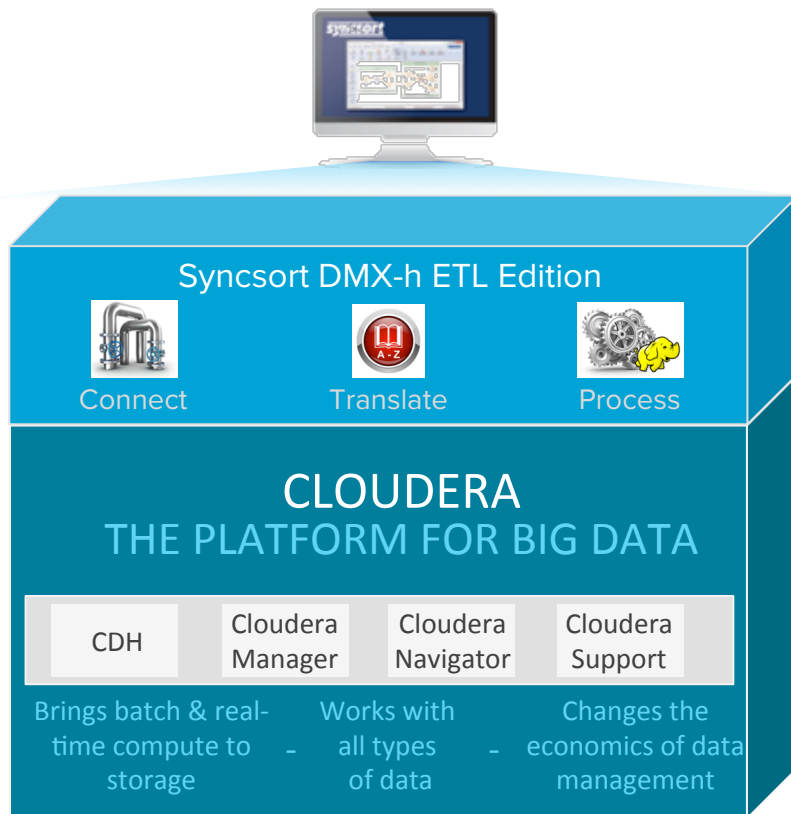
1. Bring Suits & Hoodies together early in the process
 - Build cross-organizational teams
 - Understand mutual concerns
 - Identify critical data & applications
2. Clearly define business & IT objectives
 - Reduce costs
 - Uncover new insights
3. Create a roadmap that gradually builds the skills of your organization
 - Copy → Migrate → Offload

Bridging the Gap Between Big Iron and Big Data

cloudera
Ask Bigger Questions



INTEGRATING BIG DATA... SMARTER
syncsort



A SMARTER APPROACH TO BIG MAINFRAME DATA!

- ✓ Zero-MIPS Connectivity
- ✓ Painless Integration & Translation
- ✓ Mainframe-like Performance & Reliability
- ✓ Massively Affordable Scalability
- ✓ Support for Offloading Batch Cobol & JCL Processing to Hadoop
- ✓ Iron-clad Security
- ✓ Decades of Proven Mainframe Expertise
- ✓ Easy Deployment, Monitoring & Admin

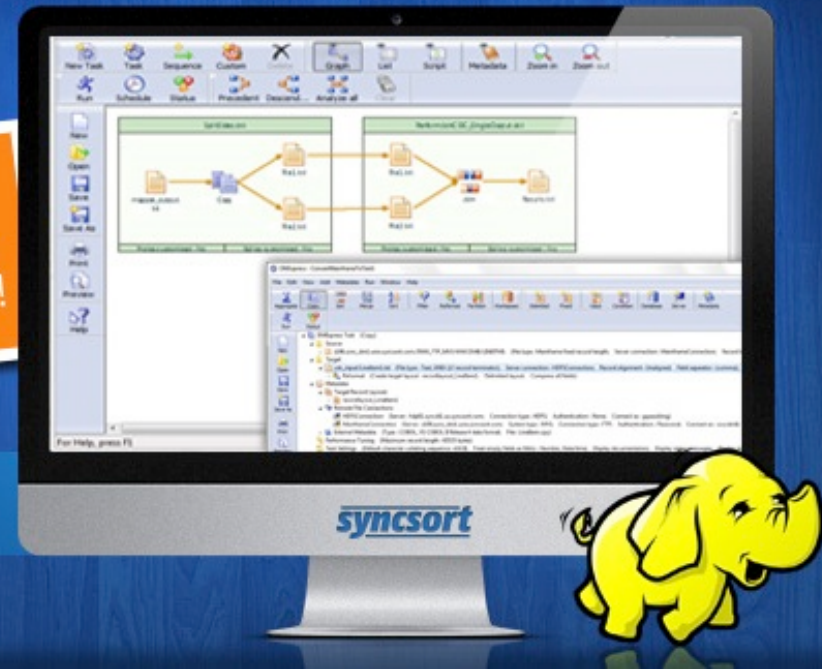
Test Drive DMX-h:

*BRIDGE THE GAP BETWEEN
BIG IRON & BIG DATA!*

- Self-contained image
- Use case accelerators for
- mainframe, Hadoop and more!

RUNNING ON CDH

Try It
Free for
30 Days!



(cloudera[®] + *INTEGRATING BIG DATA... SMARTER* syncsort)
Ask Bigger Questions

Try it **FREE** at: syncsort.com/try



Stop by booth **#304**
Get a demo | Get a shirt