

Schnelle und einfache Datenvorbereitung für Business Analytics

Dominik Claßen
Director of Sales Engineering EMEA & APAC

Big Data is a Big Opportunity



The amount of all data generated annually is **doubling every two years** ¹

Nearly **80% of companies** are investing in big data or planning to do so.²



Average salaries for top skills: ³

MapReduce: **\$127,315**

HBase: **\$126,369**

Pig: **\$124,563**

Flume: **\$123,186**

Hive: **\$120,873**

1) IDC, Digital Universe Research, 2014; 2) tdwi.org, "Hadoop for the Enterprise," Q2 2015; 3) InfoWorld, "How to make \$120,000: Get a job in big data," May 2015

But Big Data is HARD

- ① New skills necessary
- ② High effort and risk
- ③ Continuous change

"Through 2018, 70% of Hadoop deployments will not meet cost savings and revenue generation objectives due to skills and integration challenges." -- Gartner¹

1) Gartner Analyst, Nick Heudecker; infoworld.com, Sept 2015

What Does Failure Look Like?



Major government agency

Implemented Hadoop

Wanted to report on enrollment for new healthcare program

Needed to process diverse data from website



Fast forward

9 months of development with legacy ETL tool

Staff inexperience shows, tool is too complex to maintain

Adds dedicated engineer from vendor



Bad to worse...

Handles just 10% of data, and doesn't fully process it

Removes tool, decides to take a custom approach

Implementation takes 16 hrs to run

...Think about the implications for:
Staffing costs | ROI | people's jobs

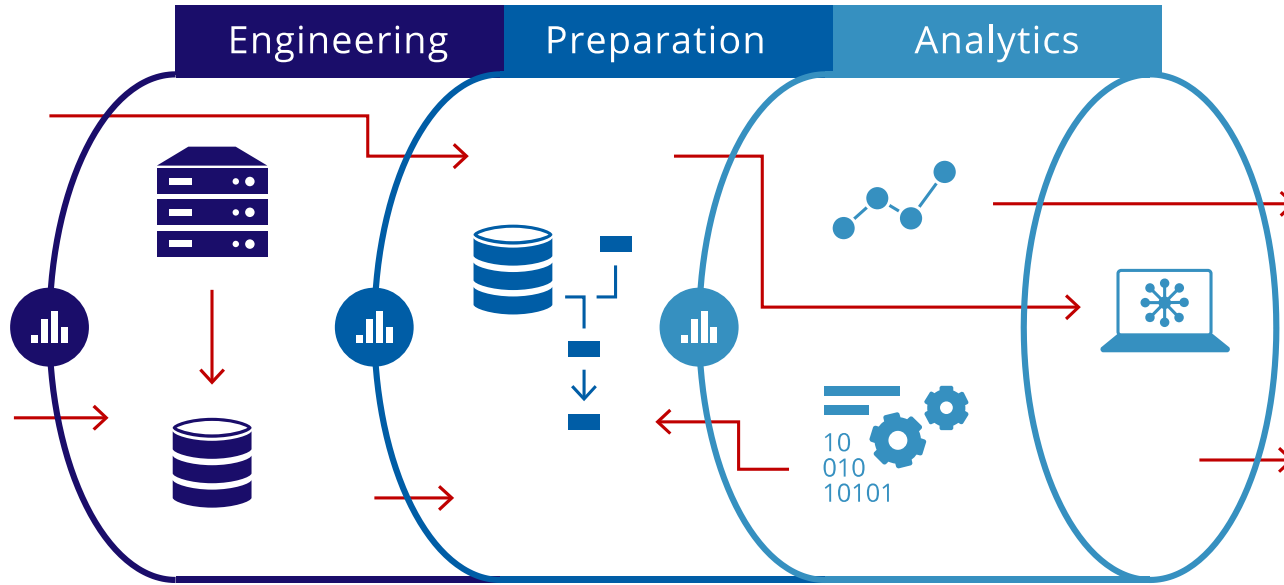


This is still one of the most significant big data challenges:

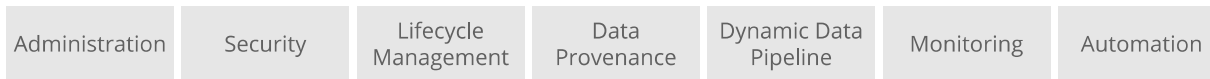
“Determining how to get value from Hadoop” – Gartner¹

1) “Gartner Survey Highlights Challenges to Hadoop Adoption”, gartner.com, May 2015

Die Data Pipeline



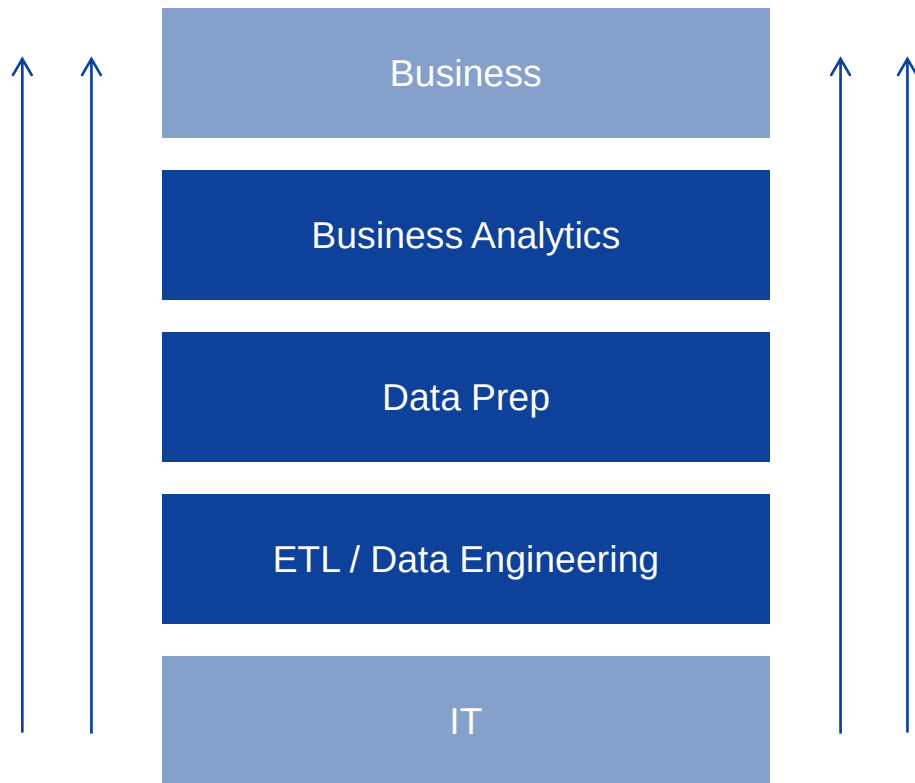
Managing and automating the pipeline



Disparate tools, disparate problems

Today's data landscape is littered with **disparate tools** and **disjointed processes**.

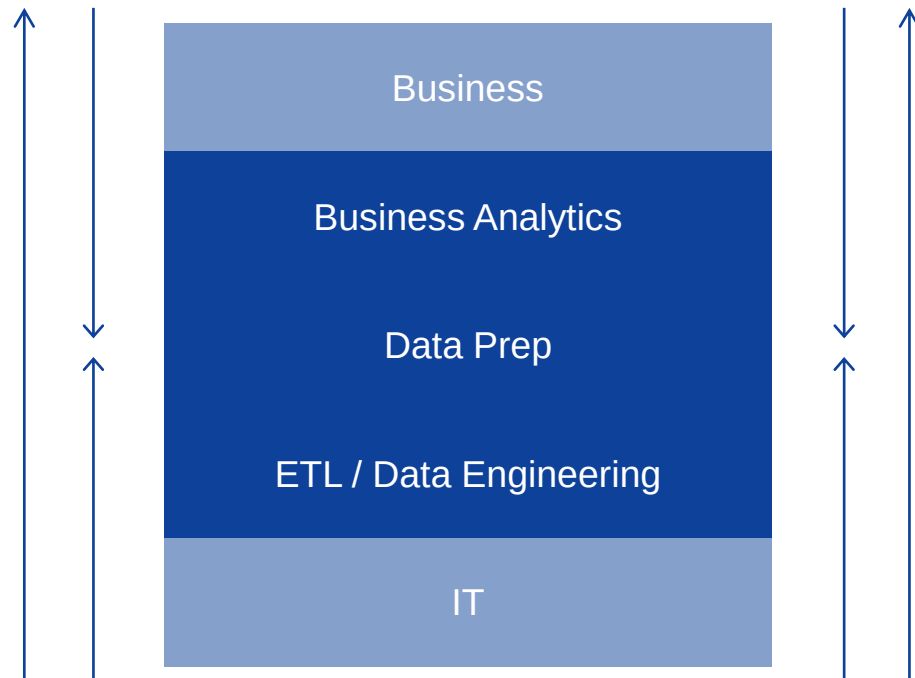
How do you get ahead with that?

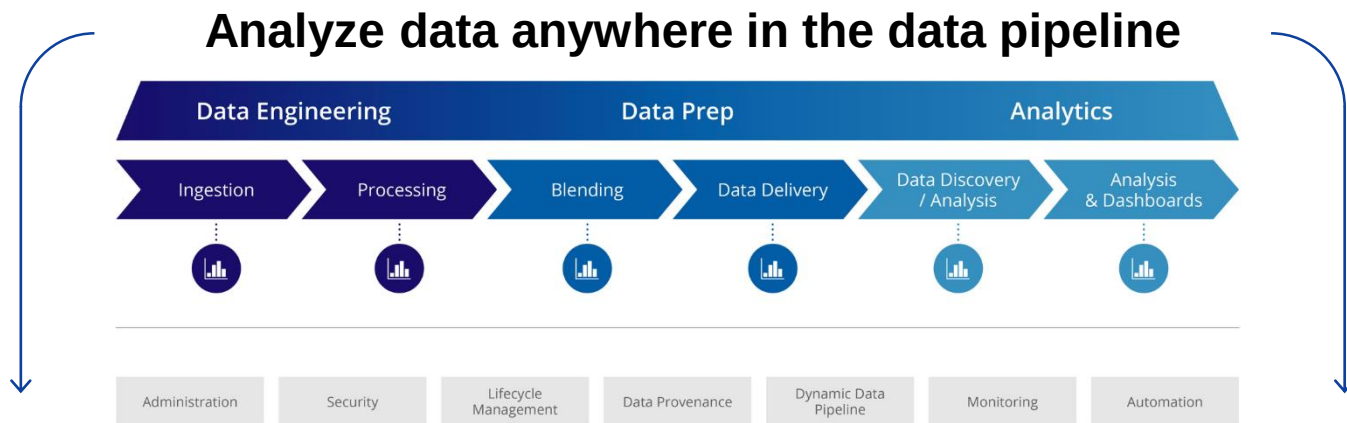


Bridging the gap between data preparation and analytics

Pentaho 7.0 is about **bridging the data divide between business and IT.**

Leveraging the Pentaho platform to drive the business with **usable**, **accurate**, and **accessible** analytics across the organization.





Bridging the gap between data preparation and analytics with **a visual data experience from anywhere in the data pipeline:**

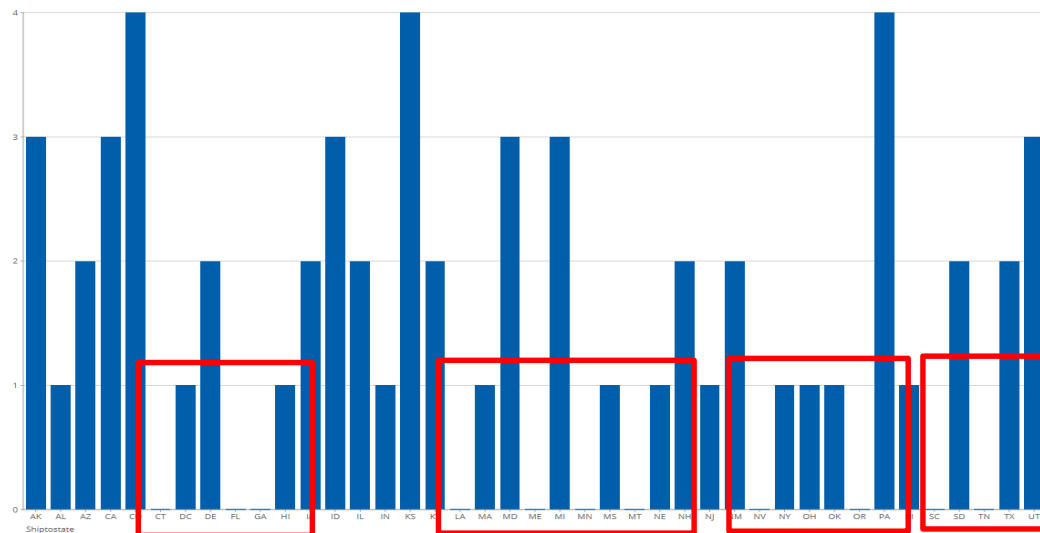
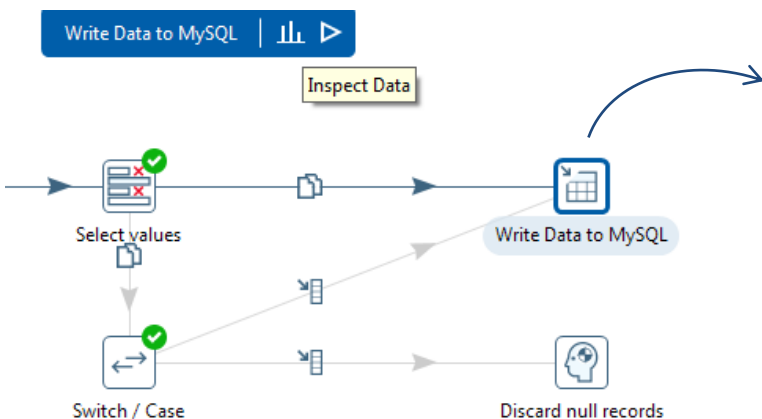
- Data prep analytics;
- Share analytics in PDI.

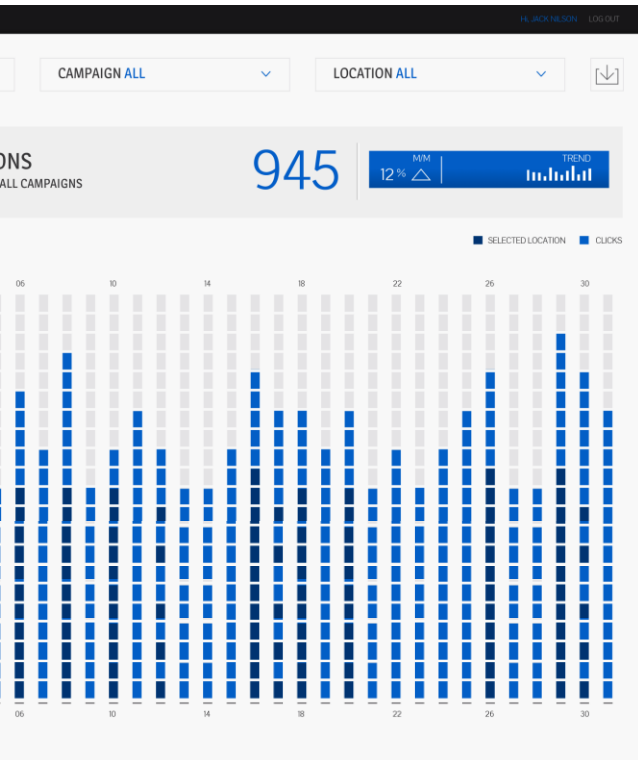
On top of governance, security, and **big data ecosystem support for a blended data world:**

- Spark;
- Metadata Injection;
- Hadoop Security;
- Admin Simplification.

Bringing Analytics into Data Prep

Spot check and visualize data in flight –
without switching applications

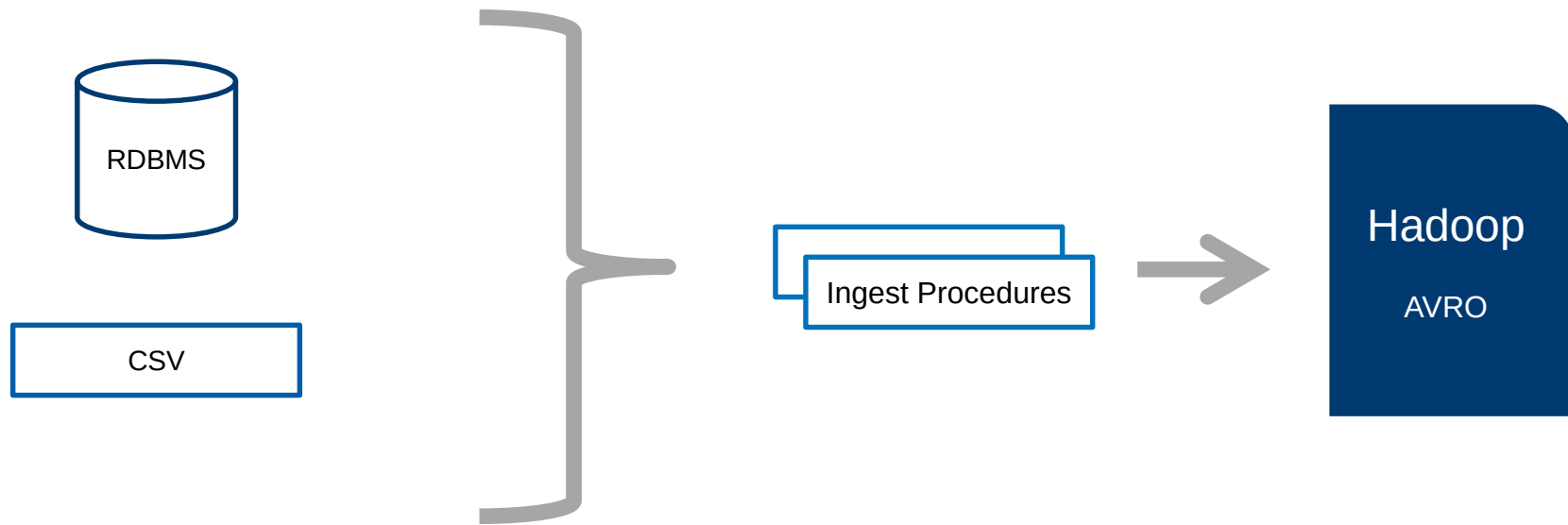




What if data gets big?

Dealing with 100s or 1000s of data sources

Extract, Load, Transform (ELT)

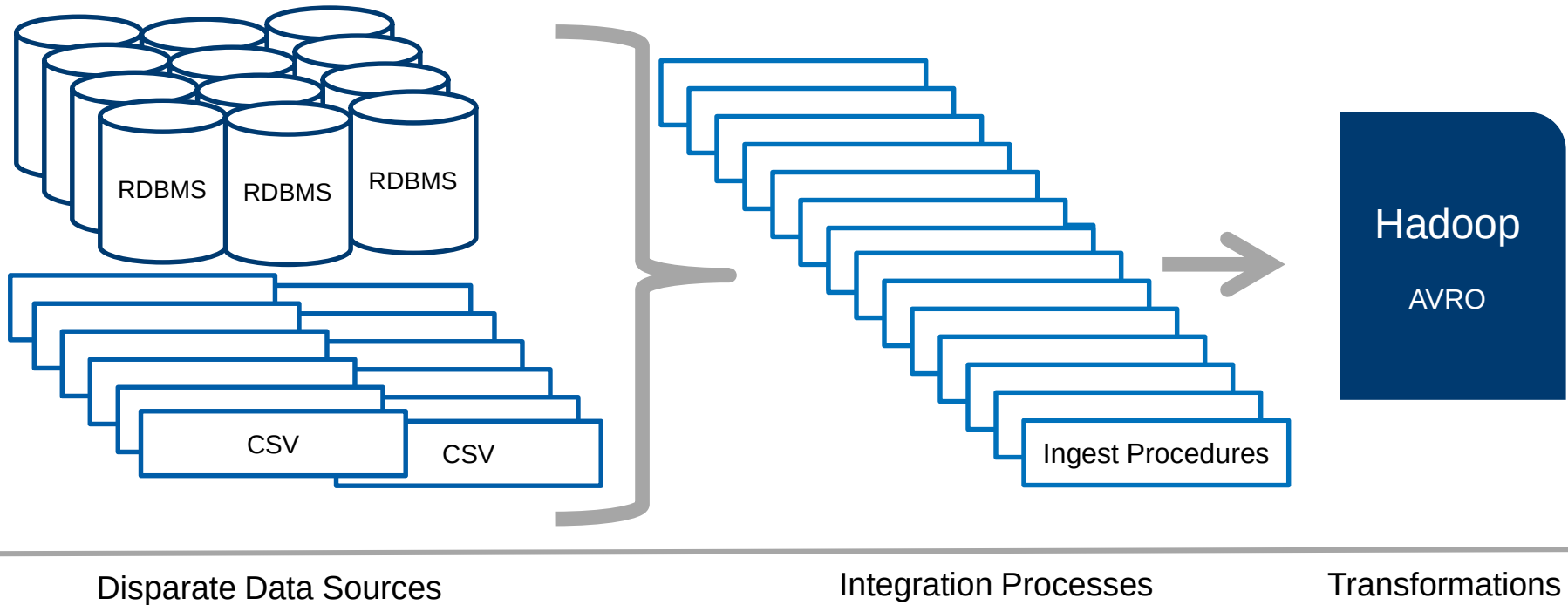


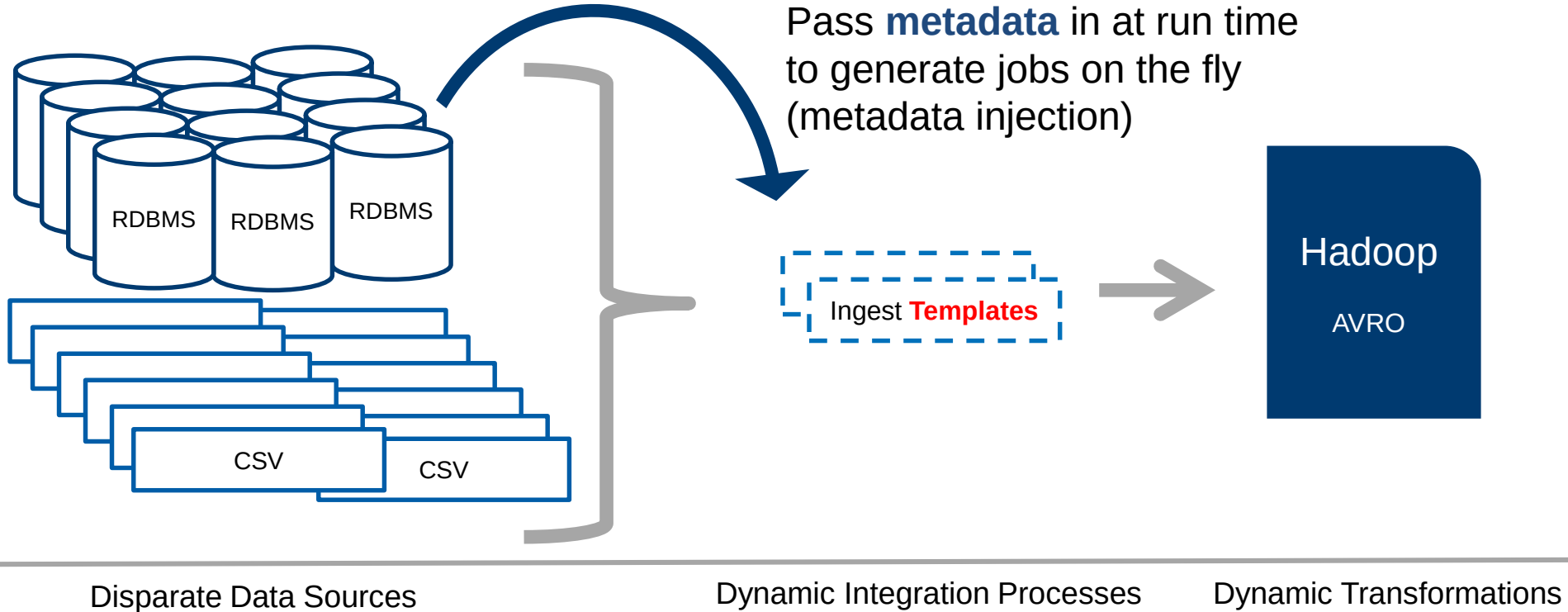
Disparate Data Sources

Integration Processes

Transformations

ELT at Scale



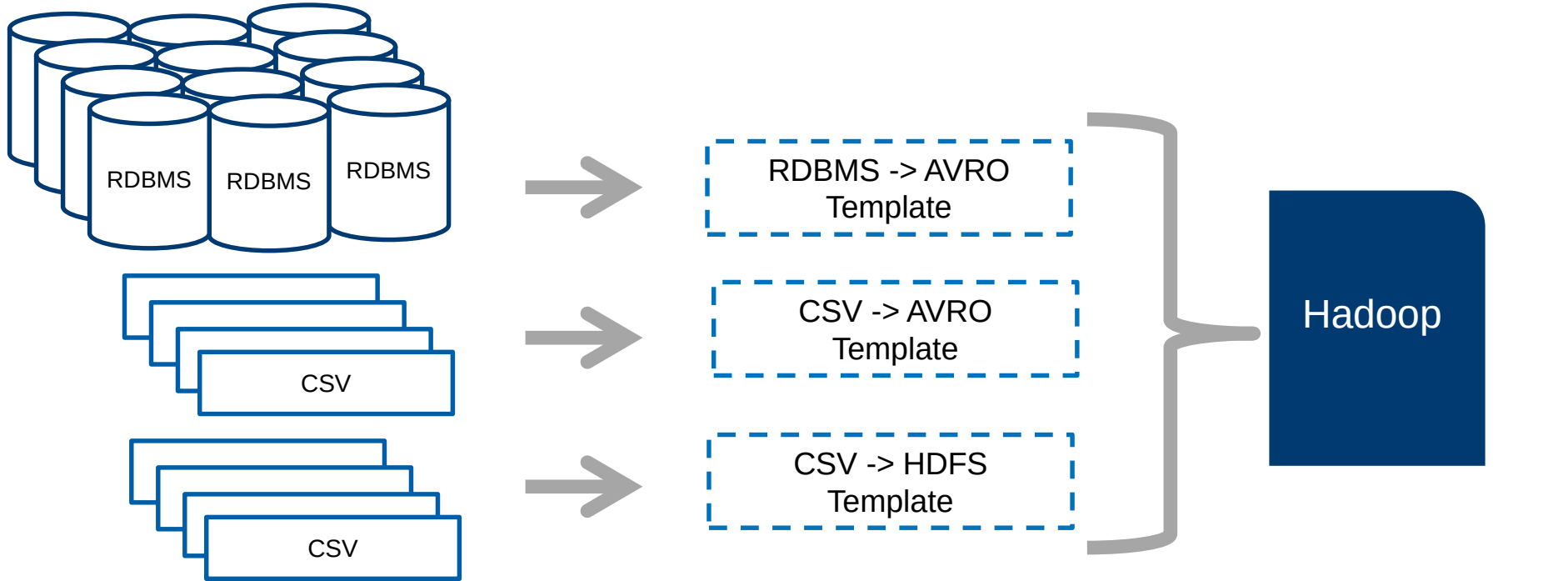


Disparate Data Sources

Dynamic Integration Processes

Dynamic Transformations

Templated workflows



Disparate Data Sources

Dynamic Integration Processes

Dynamic Transformations

Variety – different metadata, one template

Customer	Address	City	State	Zip
Acme	123 Main St	New York	NY	01234
BigCo	22 Spring Ln	San Diego	CA	95123

Order ID	Customer ID	Order Date	Order Total
3425	123	3/15/2016	5000.00
3426	543	3/16/2016	2500.00
3427	432	3/16/2016	700.00

Order ID	Product ID	Price	Qty
3425	4324	25	3
3425	3221	10	50
3426	6554	50	7

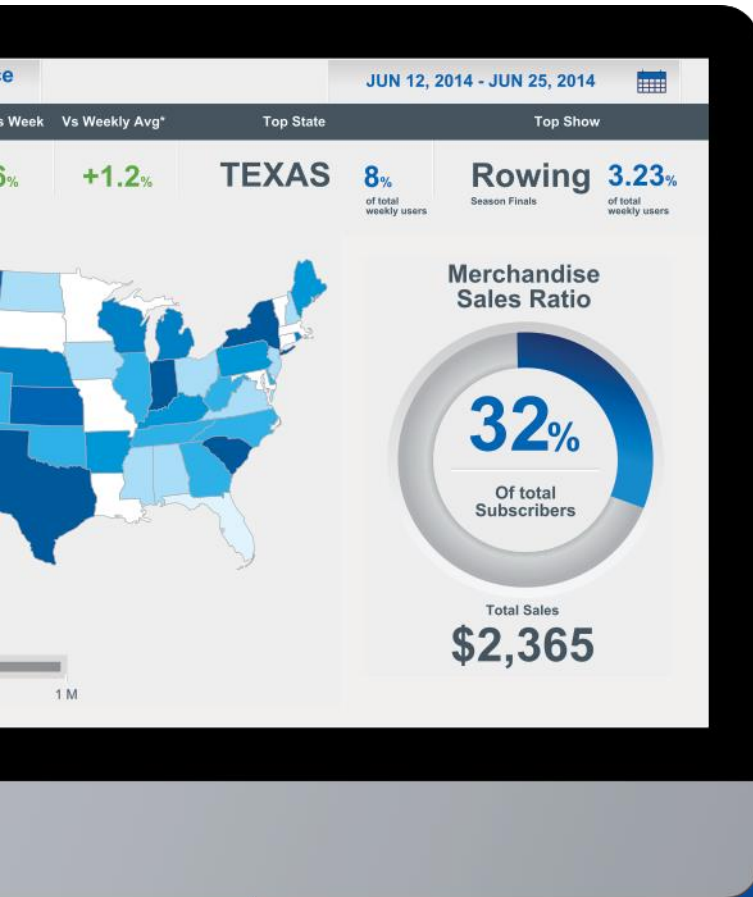
CSV -> AVRO
Template

Hadoop

Disparate Data Sources

Dynamic Integration Processes

Dynamic Transformations



History repeating...

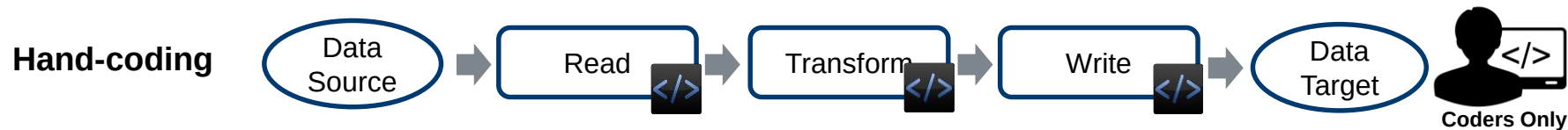
Your toolkit for the journey

BIG DATA IS JUST A FAD



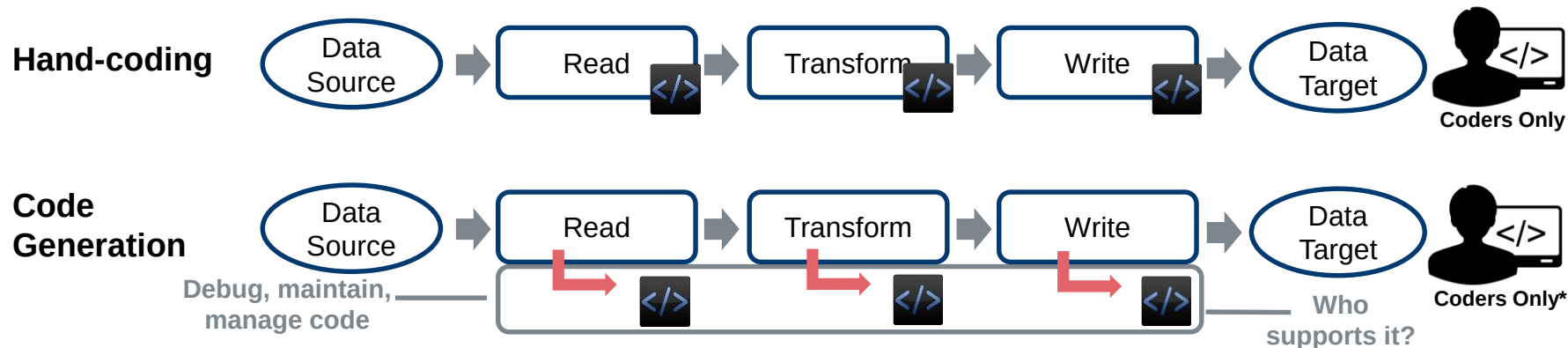
**IT'S JUST
1960's BATCH
PROCESSING,
BUT IN COLOR**

Big Data Integration Approaches



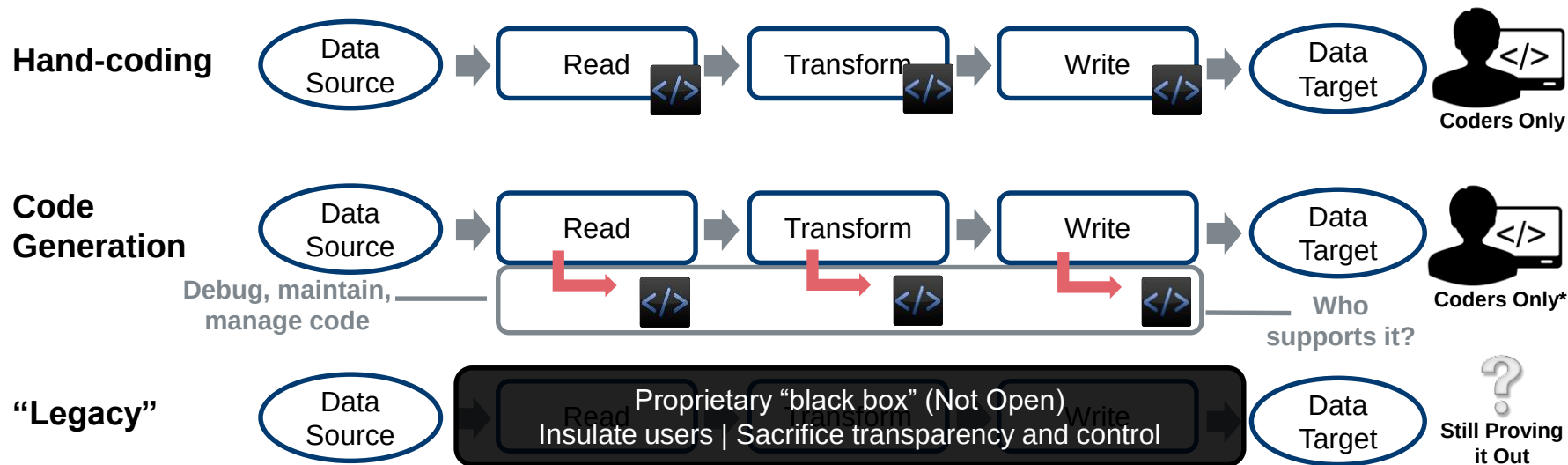
Data integration that is intuitive and powerful

Big Data Integration Approaches



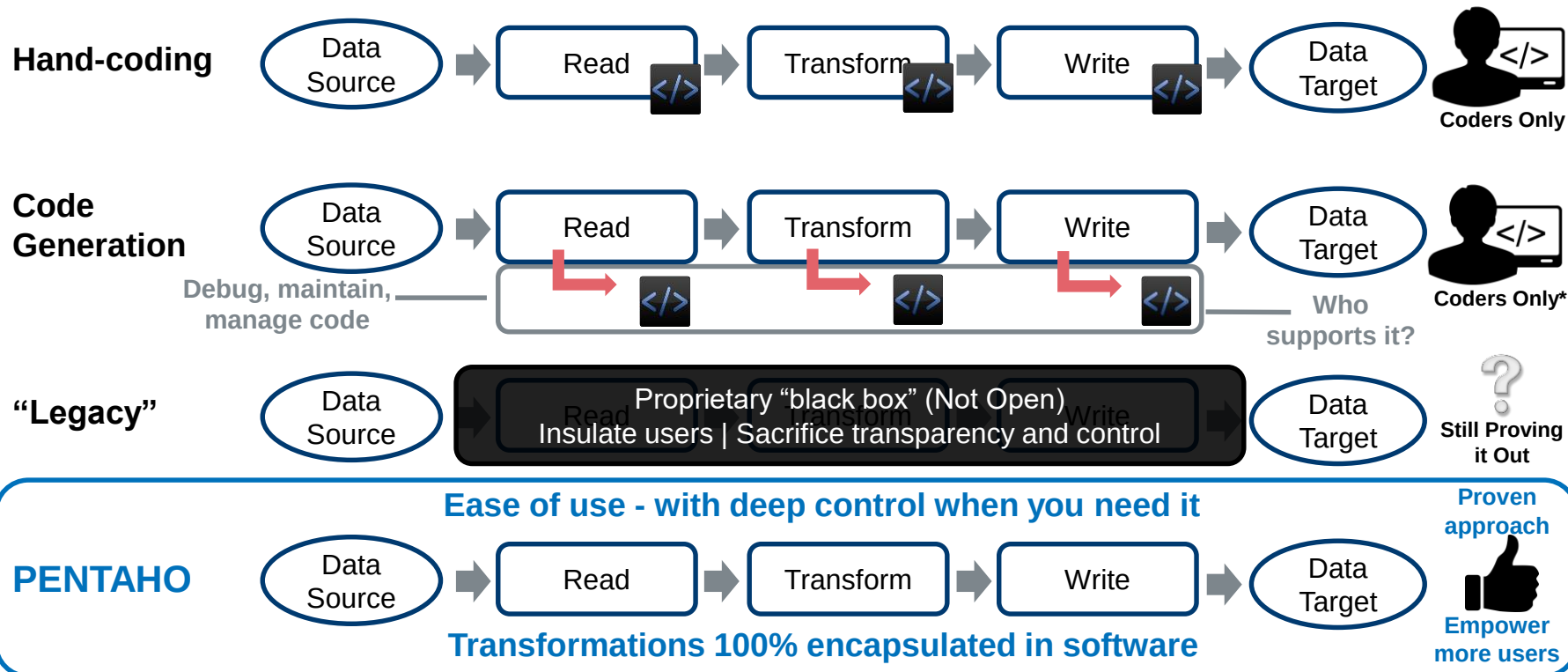
Data integration that is intuitive and powerful

Big Data Integration Approaches



Data integration that is intuitive and powerful

Big Data Integration Approaches



In an immature market, you will *not* find one off-the-shelf tool to meet all requirements

Pliable Technology for the Next Challenge



- Built on open architecture and standards
- Port jobs across Hadoop distributions
- Extend and embed technology

There is No Substitute for Experience



- 6+ years of big data innovation
- Services experts that have been there
- Track record of big data customer success



PAYTRONIX



BORDERFREE



NetApp

Thank You