cloudera®

How Hadoop Changes the Analytics Paradigm

Kunal Taneja, System Engineering Manager A/NZ

Why is Big Data Happening Now?

Instrumentation

Everything that can be measured will be measured.

Consumerization

Employees and customers expect more personal interactions, but not at the cost of their privacy.

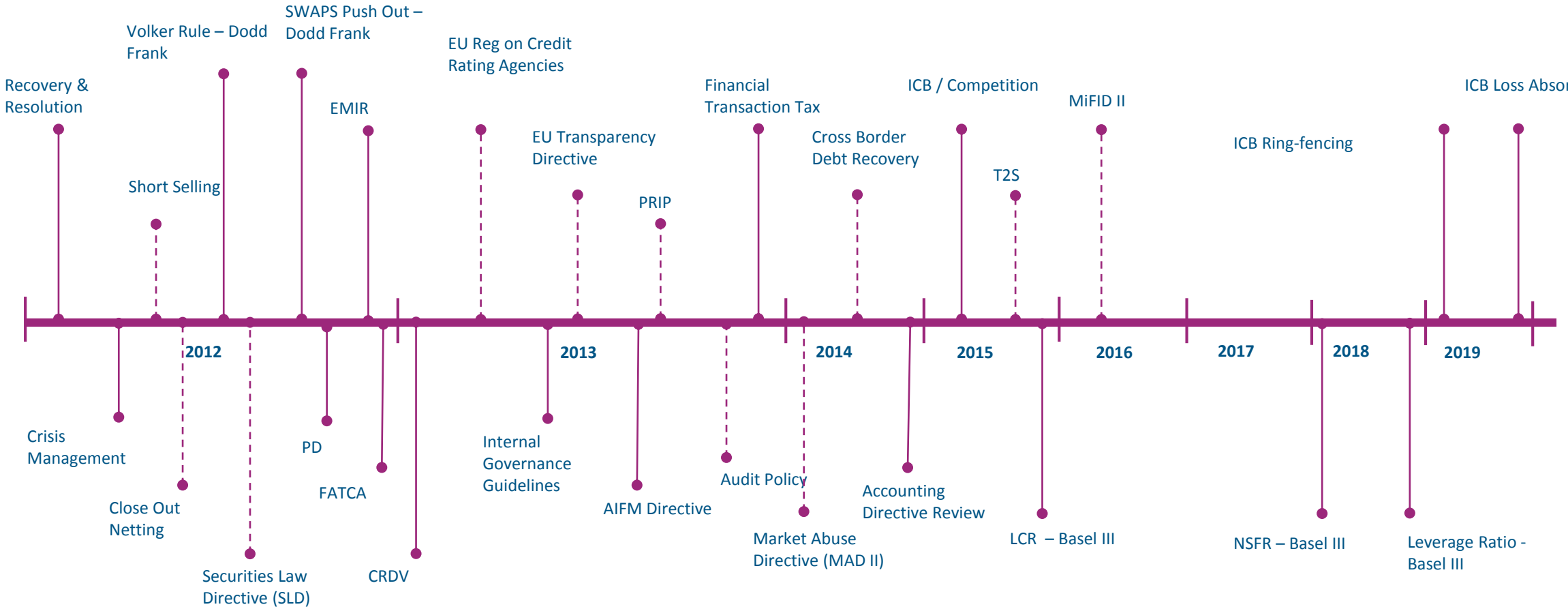
Experimentation

The most innovative companies embrace experimentation and agility.

cloudera

© Cloudera, Inc. All rights reserved.

Example: Instrumentation in Banking

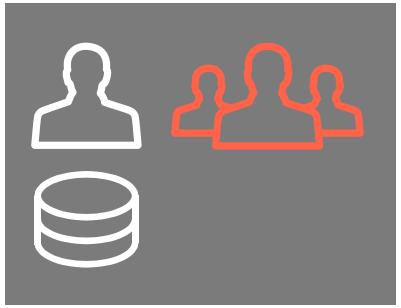


Effective dates yet to be confirmed

cloudera

© Cloudera, Inc. All rights reserved.

Traditional Architectures Under Pressure



Limited Insights

Power users struggle with data.

Many users have no data.

Compliance and Privacy

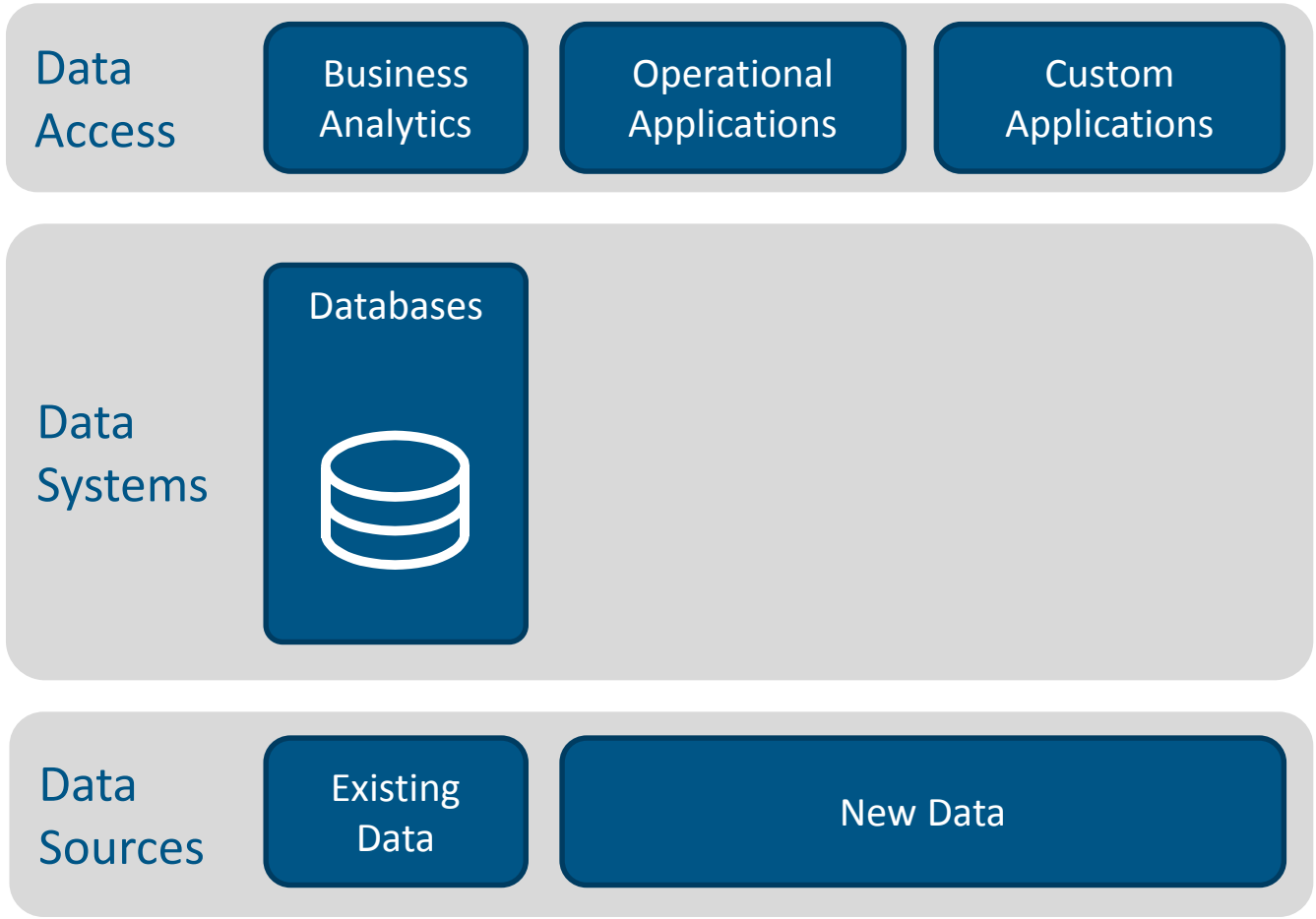
More data, more users, and more tools create complexity.

Need to balance business agility with security and governance.

Limited Data

Not efficient to keep existing data, let alone handle new data sources.

Time consuming to transform data for analysis in existing systems.



cloudera

© Cloudera, Inc. All rights reserved.

Traditional Architectures Under Pressure

Data Modeling

BI Abstraction & Reporting Layer

Dashboard & Reports

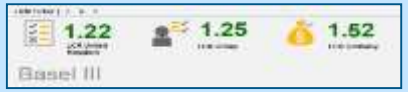
Ad-hoc Analysis

Mobile

Enterprise Data Warehouse

Performance & Access

- Transformation & Calculation
- Performance & Access



ATOMIC Layer

- Normalisation & Storage

Data Acquisition Layer

- Extraction&Staging
- Cleansing

Source Systems

ETL

cloudera

© Cloudera, Inc. All rights reserved.

Example - Limited Data

"iPhone users more likely to Buy?"

DATA MODELLING

BI Abstraction & Reporting Layer

Dashboard & Reports

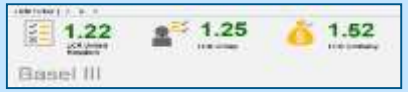
Ad-hoc Analysis

Mobile

Enterprise Data Warehouse

Transformation & Access Layer

- Transformation & Calculation
- Performance & Access



Atomic Layer

- Normalisation & Storage

Data Acquisition Layer

- Extraction&Staging
- Cleansing

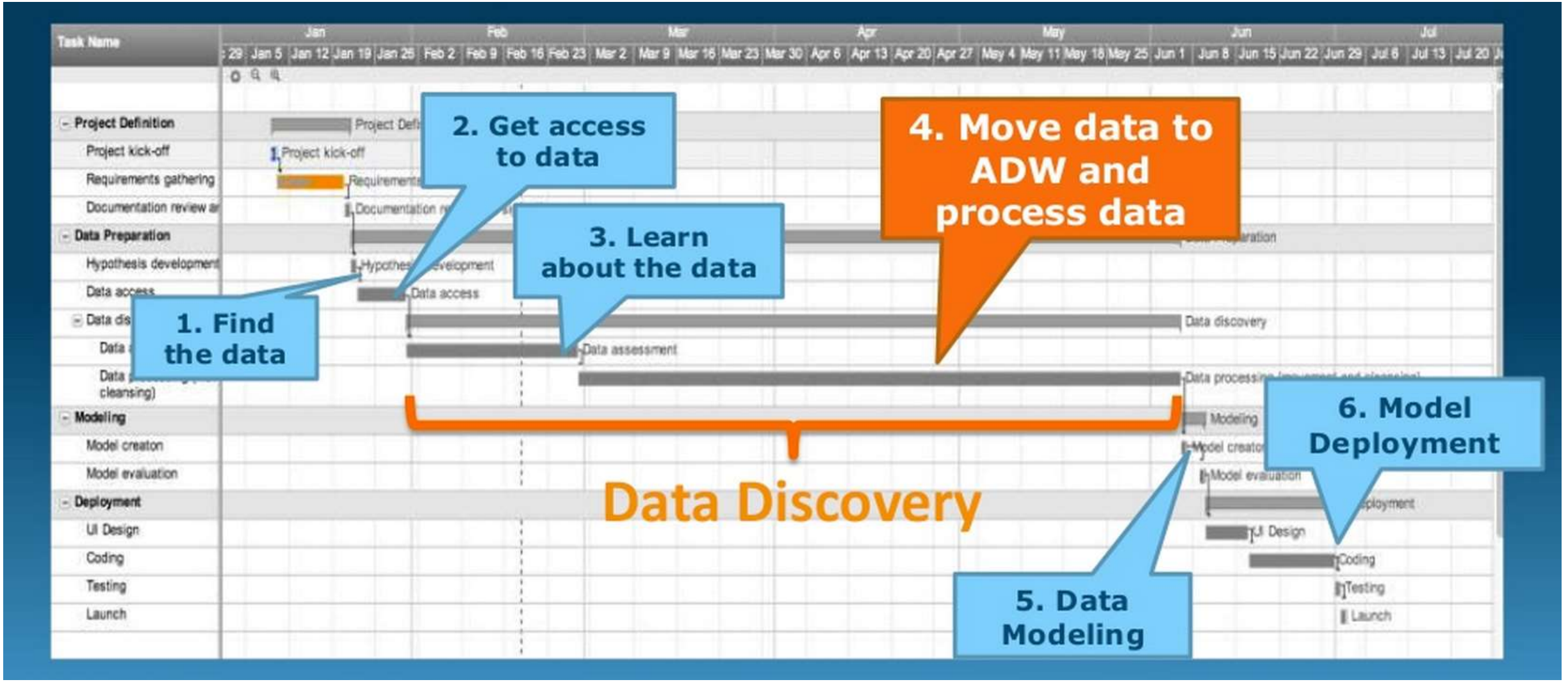
Source Systems

ETL

cloudera

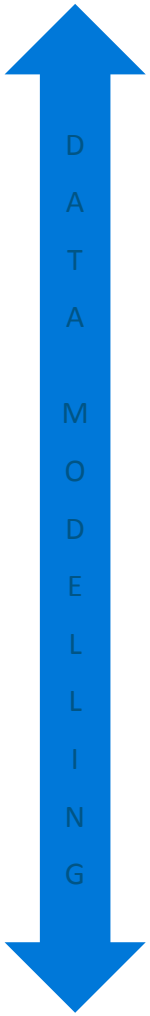
© Cloudera, Inc. All rights reserved.

Example - Limited Data



© Cloudera, Inc. All rights reserved.

Example - Limited Data (slow transforms)



"What is my VaR?"

BI Abstraction & Reporting Layer

Dashboard & Reports

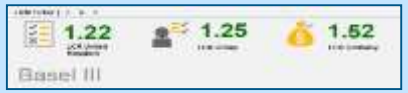
Ad-hoc Analysis

Mobile

2 Hours

Enterprise Data Warehouse

Transformation & Access Layer
- Transformation & Calculation
- Performance & Access



ETL

Atomic Layer
- Normalisation & Storage

12 Hours

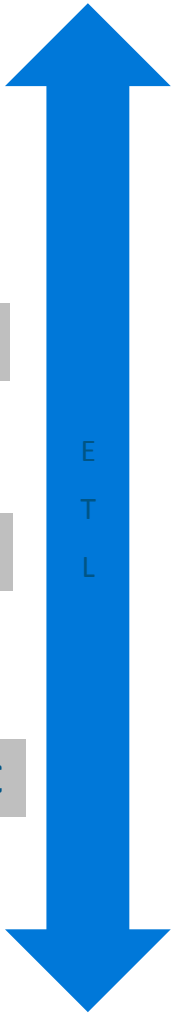
ETL

Data Acquisition Layer
- Extraction&Staging
- Cleansing

8 Hours

CDC

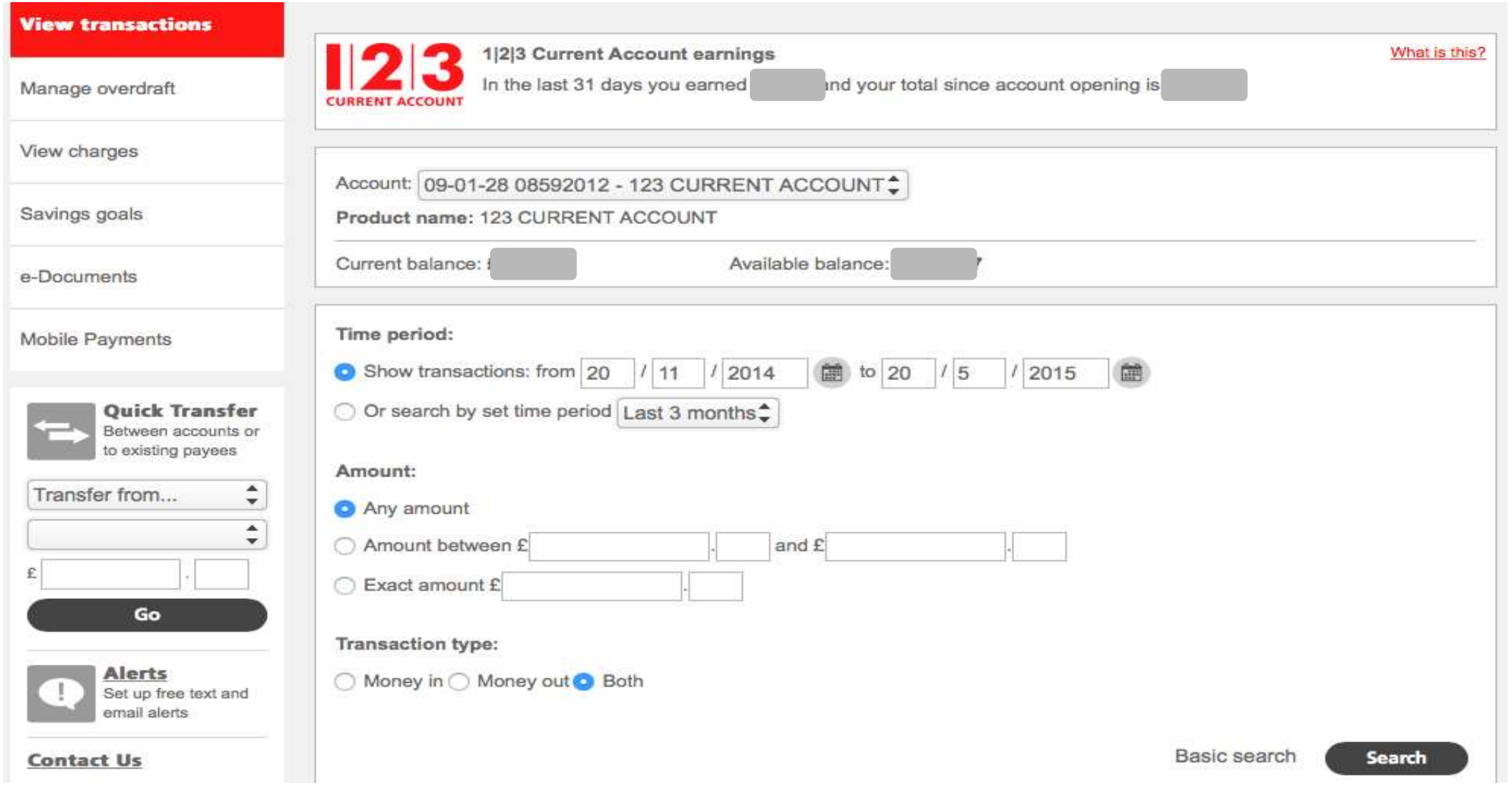
Source Systems



cloudera

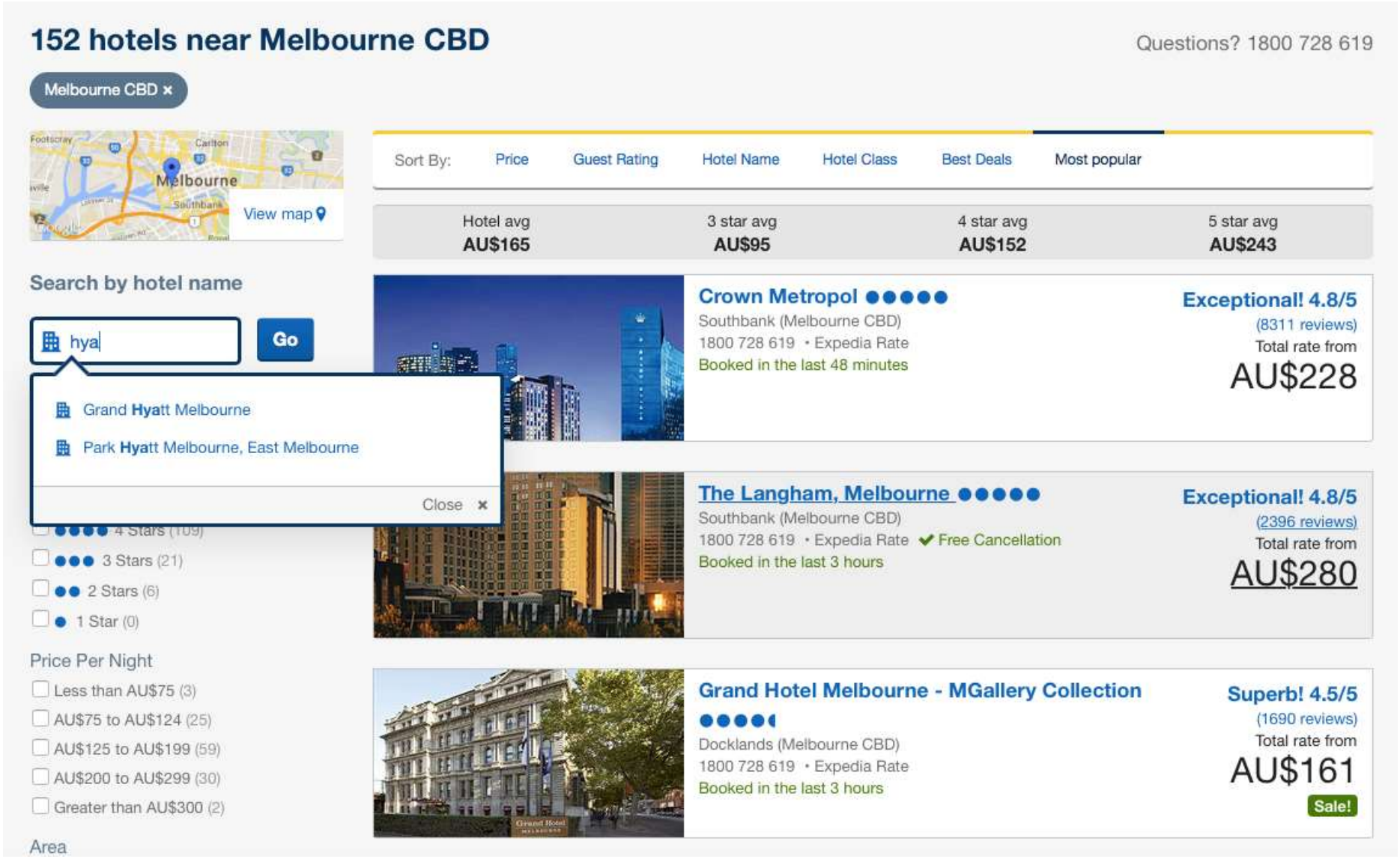
© Cloudera, Inc. All rights reserved.

Example - Limited Insights



© Cloudera, Inc. All rights reserved.

Example - Limited Insights



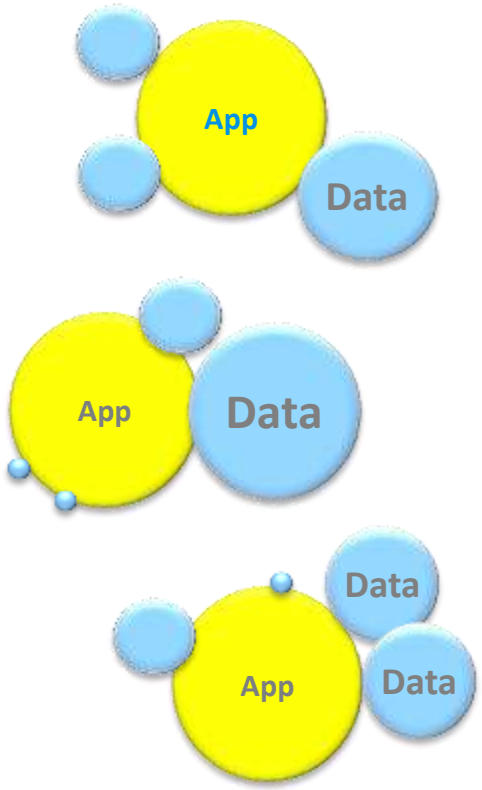
cloudera

© Cloudera, Inc. All rights reserved.

Expanding Data Requires A New Approach

What we do

Copy Data to Applications

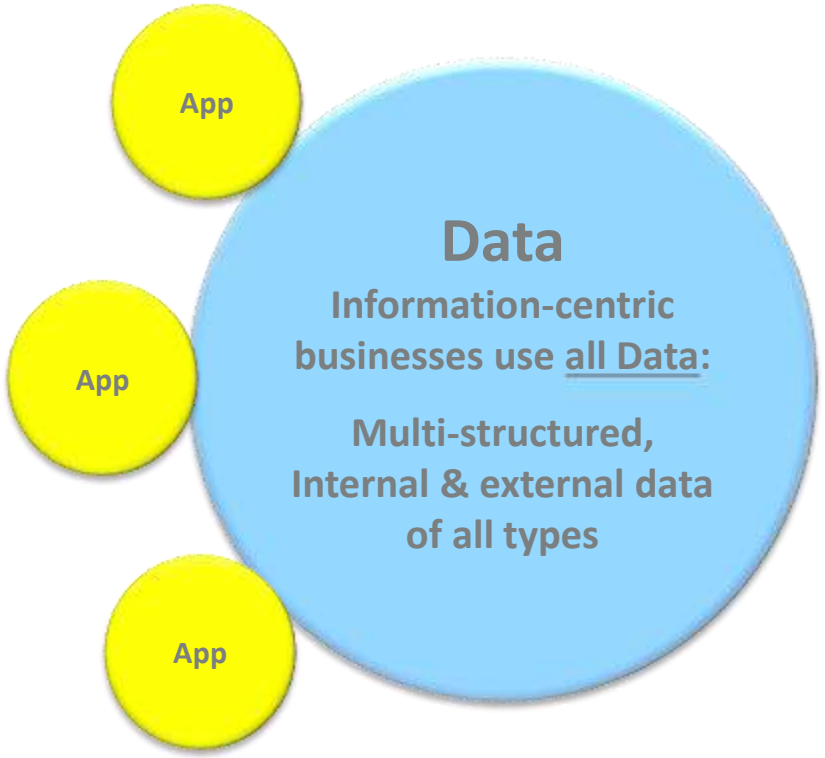


Process-centric businesses use:

- Structured data mainly
- Internal data only
- “Important” data only
- Multiple copies of data

What we should do

Bring Applications to Data



cloudera

© Cloudera, Inc. All rights reserved.

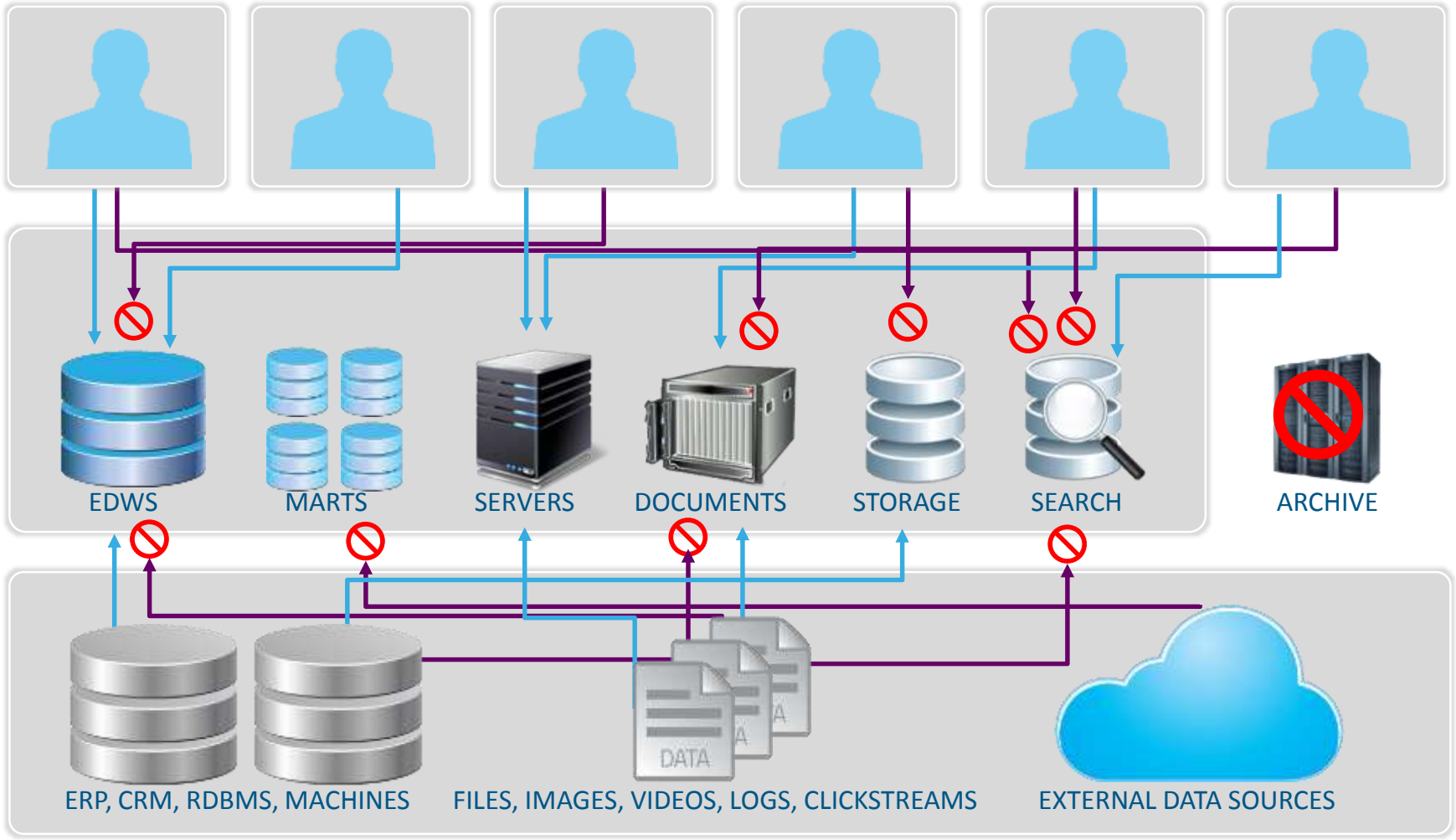
The Old Way: Bringing Data to Applications

4 Can't Get a 360 View
- Many special-purpose systems
- Moving data around
- No complete views

3 Can't Ask New Questions
- Existing systems strained
- No agility
- "BI backlog"

2 Can't Meet ETL SLAs
- Up-front modeling
- Transforms slow
- Transforms lose data

1 Can't Retain Valuable Data
- Leaving data behind
- Risk and compliance
- High cost of storage



cloudera

© Cloudera, Inc. All rights reserved.

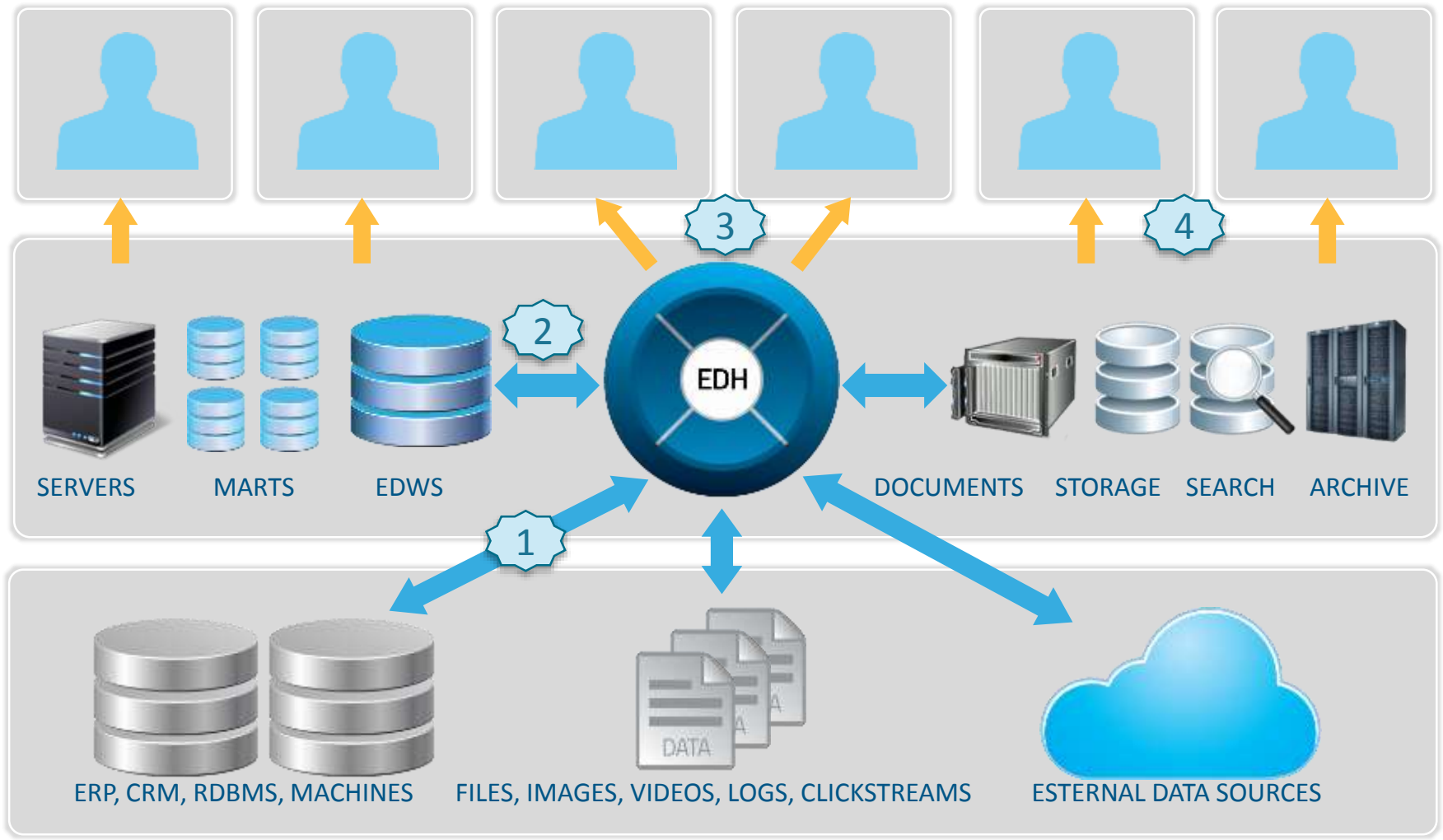
The New Way: Bringing Applications to Data

4 Consolidated Architecture
- Bring applications to data
- Combine different workloads on common data (i.e. SQL + Search)
- ***True analytic agility***

3 Agile Exploration
- Simple search + BI tools
- "Schema on read" agility
- ***Reduce BI user backlog requests***

2 Scalable Transformations
- One source of data for all analytics
- Persist state of transformed data
- ***Significantly faster & cheaper***

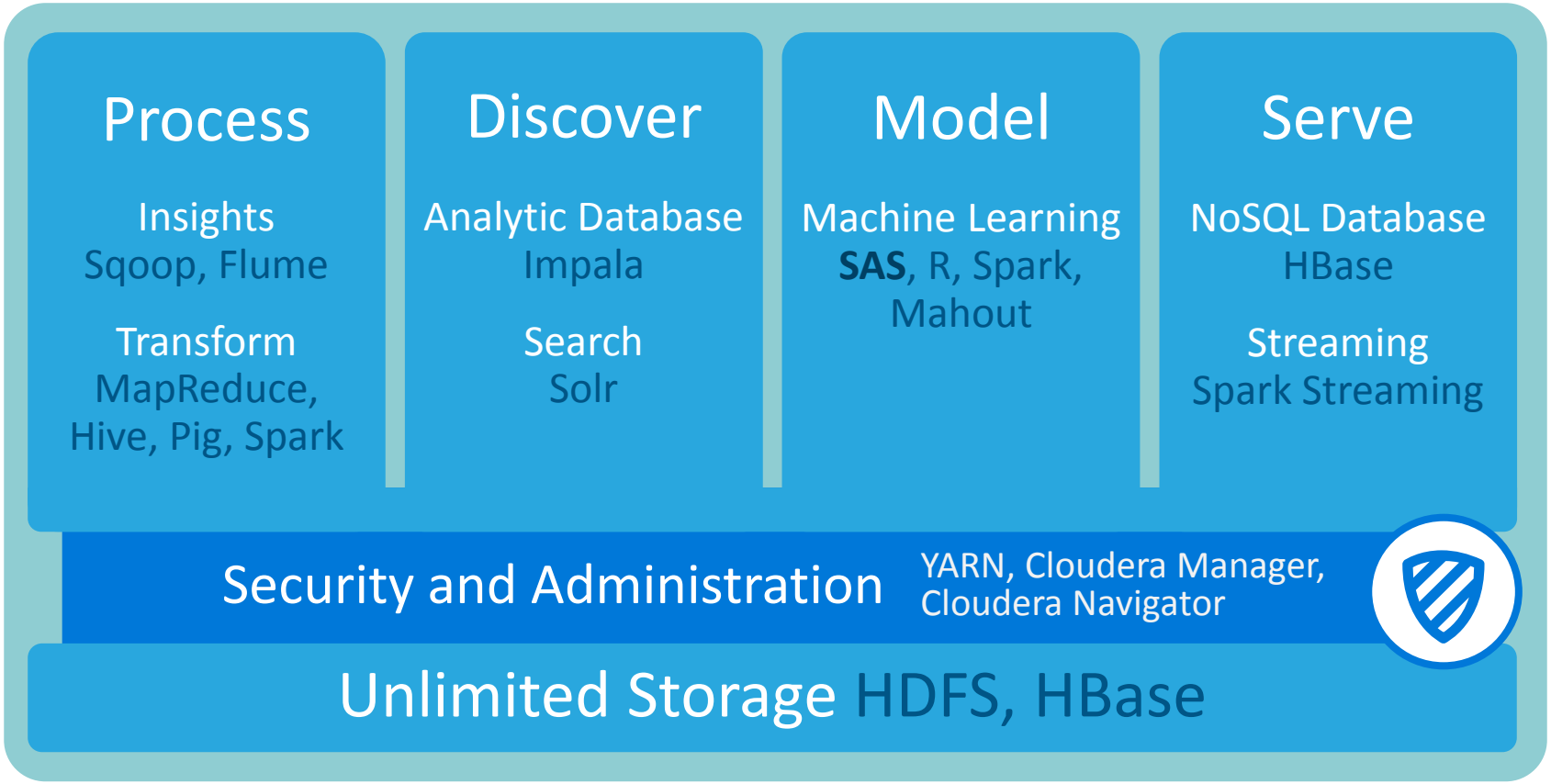
1 Active Archive
- Full fidelity original data
- Indefinite time, any source
- ***Lowest cost storage***



cloudera

© Cloudera, Inc. All rights reserved.

One Platform, Many Workloads



A new kind of data platform

- One place for unlimited data
- Unified, multi-framework data access

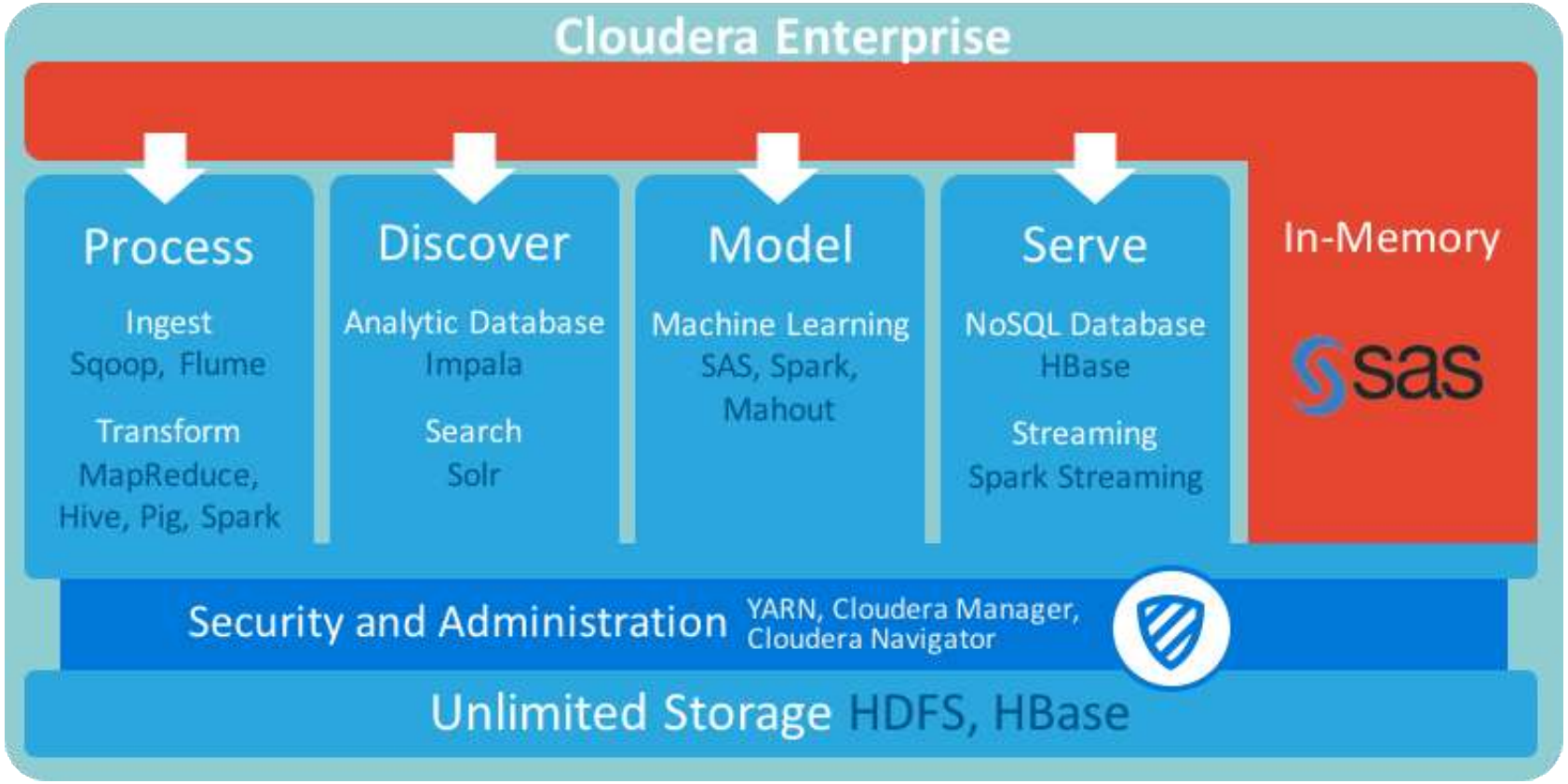
Only with Cloudera:

- Leading performance
- Enterprise system and data management
- Fundamentally secure
- Open source, open standards

cloudera

© Cloudera, Inc. All rights reserved.

SAS on Cloudera Enterprise



- Tight Integration with SAS suite of access engines, big data, and in-memory analytics solutions
- Certified Impala connector to SAS VA delivers the fastest interactive SQL on Hadoop
- Comprehensive data security and governance enable SAS users to innovate with confidence

cloudera

© Cloudera, Inc. All rights reserved.

Joint Customer Successes

Optimize

With SAS® Visual Analytics, business executives at Telecom Italia can compare the performance between all operators for a key indicator – such as accessibility or percentage of dropped calls – on a single screen for a quick overview of pertinent strengths and weaknesses.

Discover

Epsilon built a next-generation marketing application, leveraging Cloudera and taking advantage of SAS® capabilities by our data science/analytics team, that provides its clients with a 360-degree view of their customer

Empower

AMERAN provides 360-degree views into energy usage patterns and similar household comparisons to help consumers save energy.

cloudera

© Cloudera, Inc. All rights reserved.

cloudera

We Are Hiring in NZ

https://jobs.jobvite.com/cloudera/