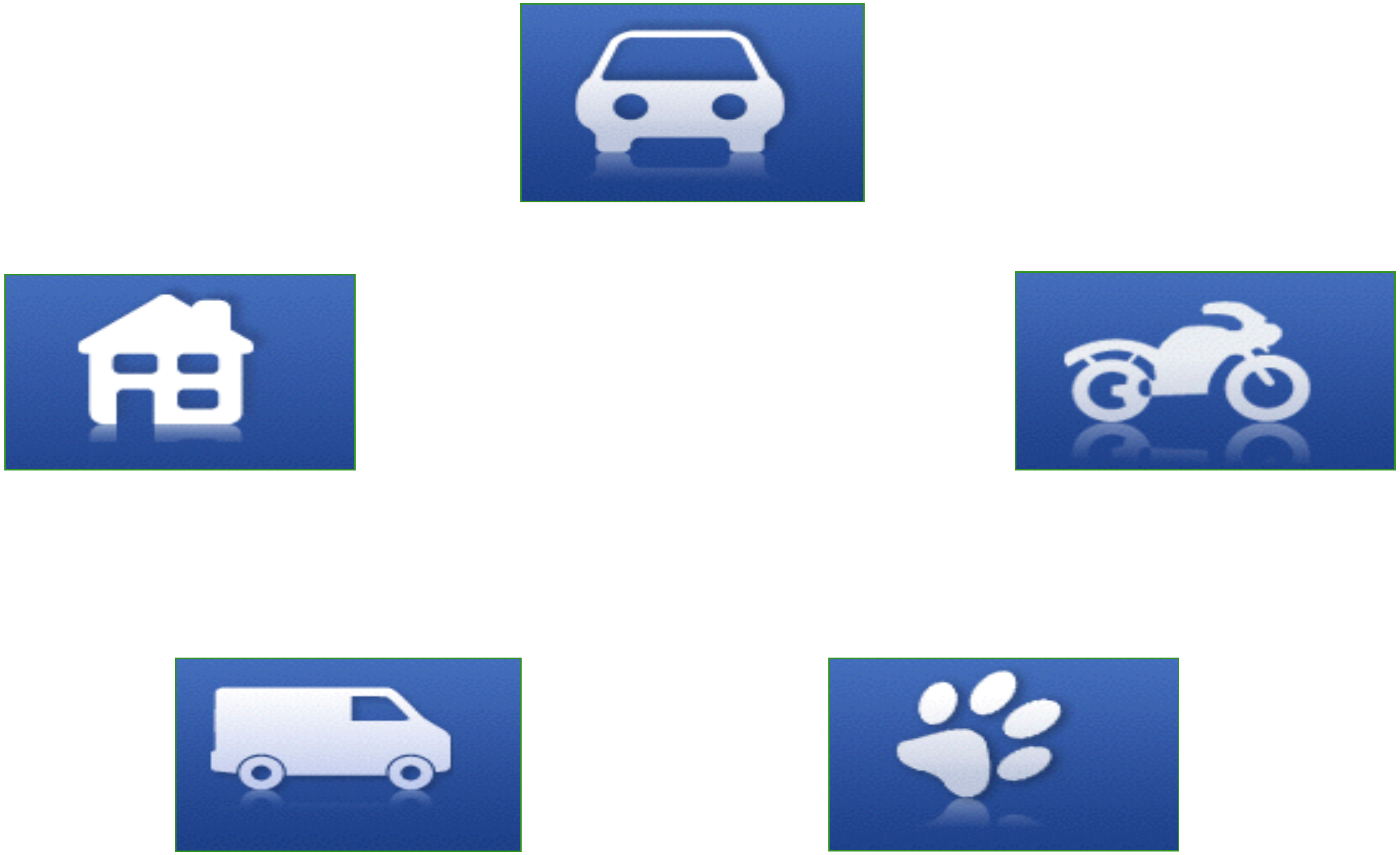
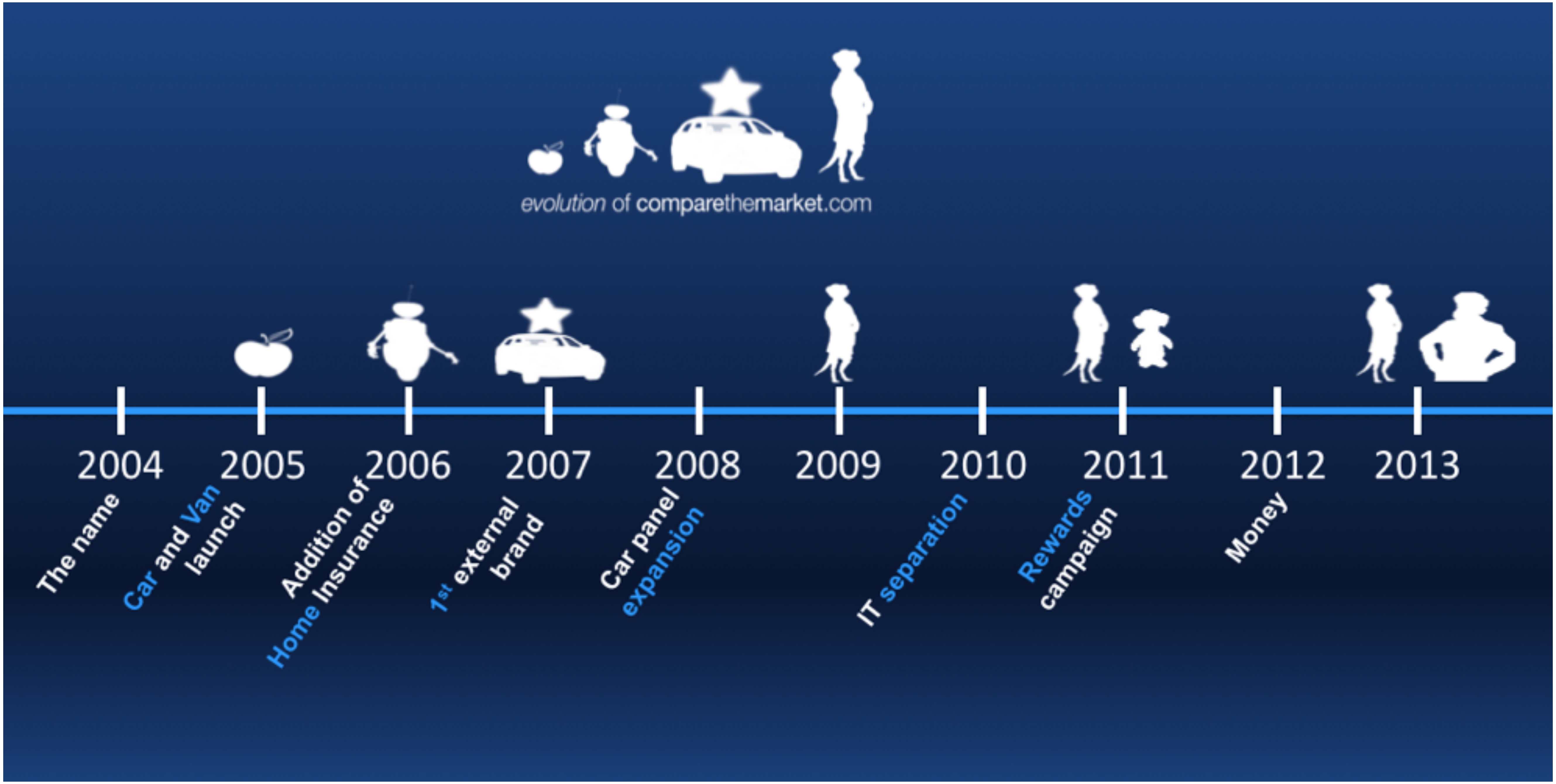
Leading Change in Data Engineering

Rob Siwicki
Neil Martin

Core Products





evolution of comparethemarket.com

2004 The name

2005 Car and Van launch

2006 Addition of Home Insurance

2007 1st external brand

2008 Car panel expansion

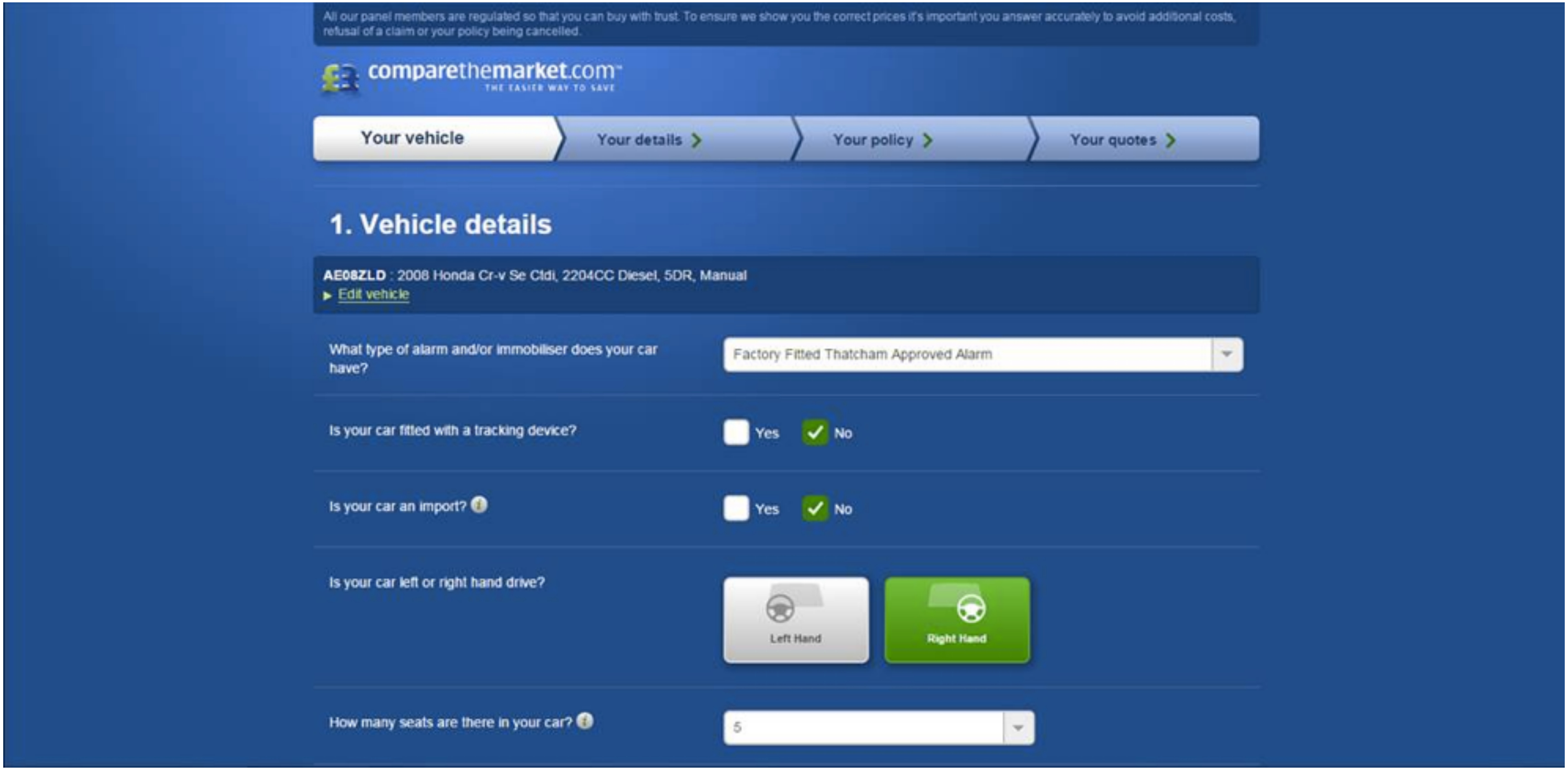
2009

2010 IT separation

2011 Rewards campaign

2012 Money

2013



All our panel members are regulated so that you can buy with trust. To ensure we show you the correct prices it's important you answer accurately to avoid additional costs, refusal of a claim or your policy being cancelled.

comparethemarket.com
THE EASIER WAY TO SAVE
Your vehicle
Your details
Your policy
Your quotes

1. Vehicle details

AE08ZLD : 2008 Honda Cr-v Se Ctdi, 2204CC Diesel, 5DR, Manual

▸ Edit vehicle

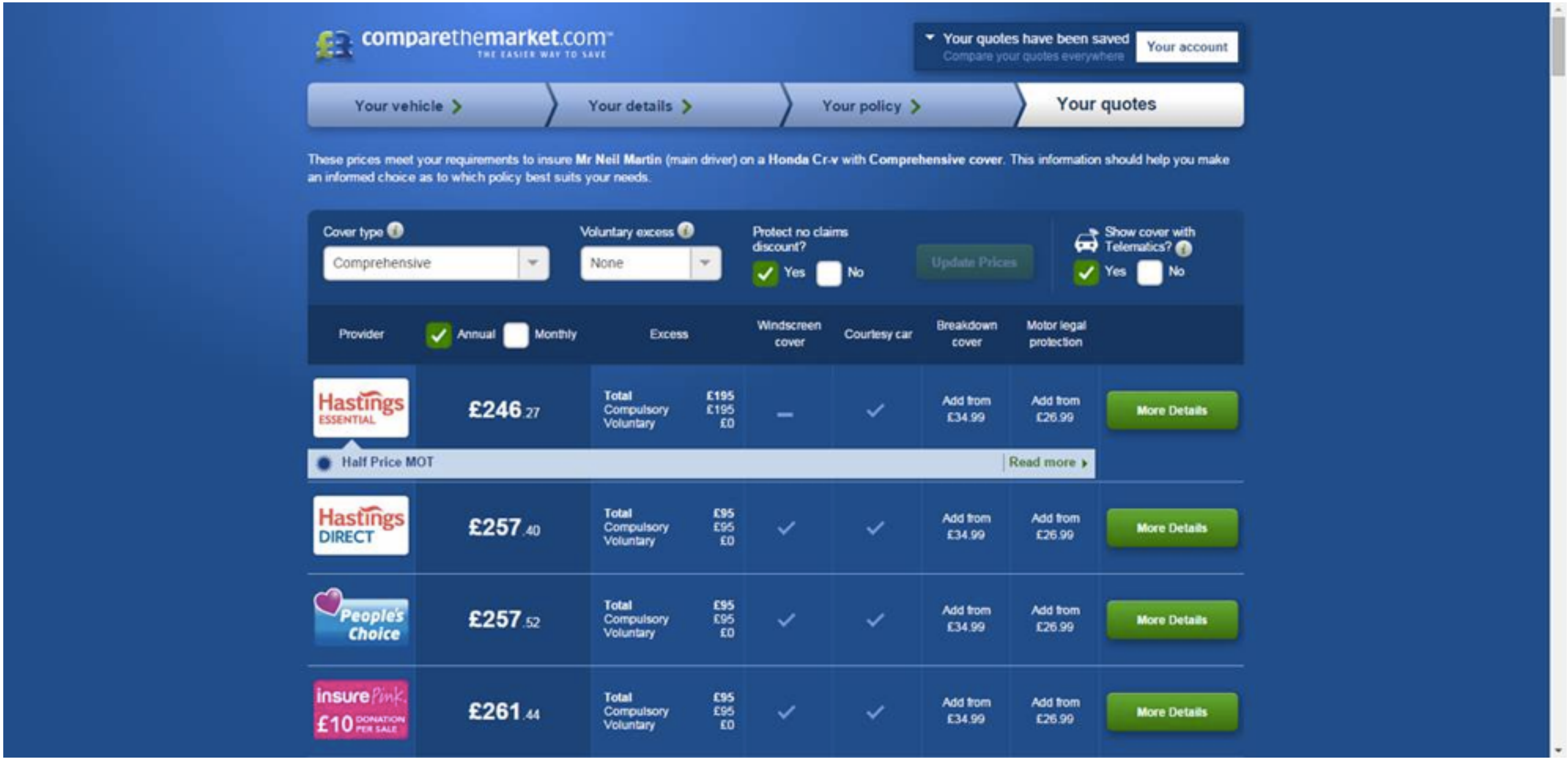
What type of alarm and/or immobiliser does your car have? Factory Fitted Thatcham Approved Alarm

Is your car fitted with a tracking device? ☐ Yes ☑ No

Is your car an import? ☐ Yes ☑ No

Is your car left or right hand drive? Left Hand [✓ Right Hand]

How many seats are there in your car? 5


comparethemarket.com
THE EASIER WAY TO SAVE
Your quotes have been saved
Compare your quotes everywhere
Your account
Your vehicle
Your details
Your policy
Your quotes
Update Prices
More Details
Read more

These prices meet your requirements to insure **Mr Neil Martin** (main driver) on a **Honda Cr-v** with **Comprehensive cover**. This information should help you make an informed choice as to which policy best suits your needs.

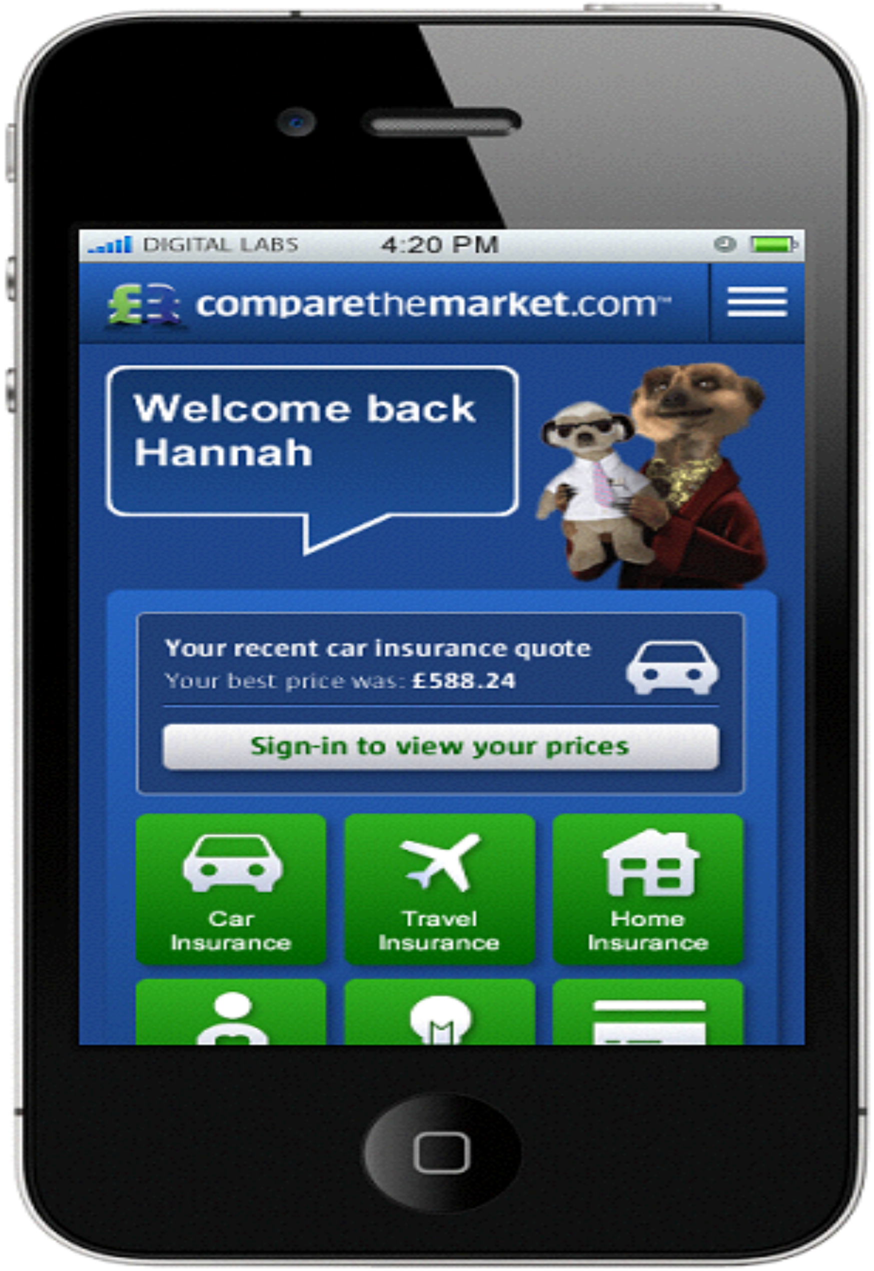
Cover type: Comprehensive

Voluntary excess: None

Protect no claims discount? ☑ Yes ☐ No

Show cover with Telematics? ☑ Yes ☐ No

Provider	☑ Annual ☐ Monthly	Excess		Windscreen cover	Courtesy car	Breakdown cover	Motor legal protection	
Hastings ESSENTIAL	£246.27	Total Compulsory Voluntary	£195 £195 £0	–	✓	Add from £34.99	Add from £26.99	More Details
● Half Price MOT							Read more	
Hastings DIRECT	£257.40	Total Compulsory Voluntary	£95 £95 £0	✓	✓	Add from £34.99	Add from £26.99	More Details
People's Choice	£257.52	Total Compulsory Voluntary	£95 £95 £0	✓	✓	Add from £34.99	Add from £26.99	More Details
insure Pink £10 DONATION PER SALE	£261.44	Total Compulsory Voluntary	£95 £95 £0	✓	✓	Add from £34.99	Add from £26.99	More Details

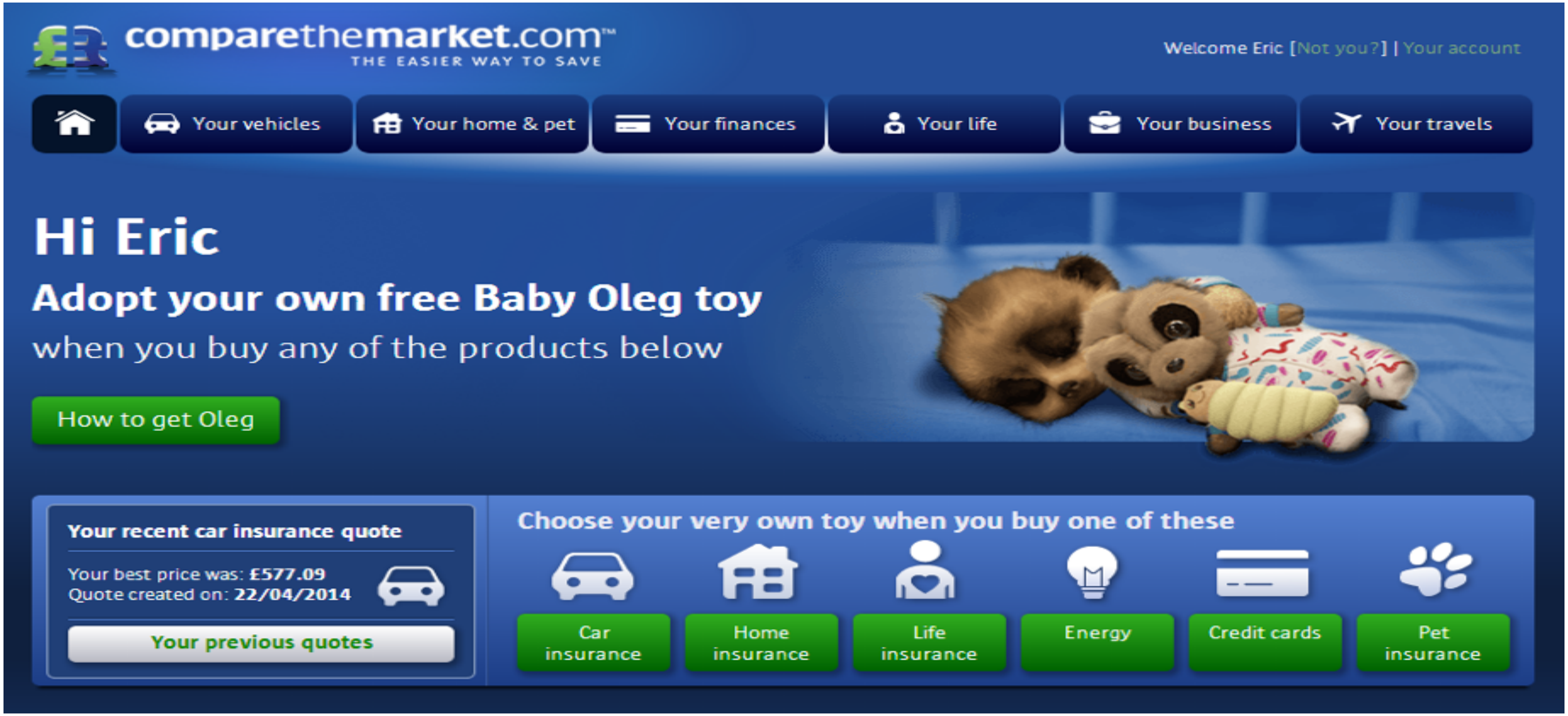

DIGITAL LABS
4:20 PM
comparethemarket.com
Car Insurance
Travel Insurance
Home Insurance

Welcome back Hannah

Your recent car insurance quote

Your best price was: **£588.24**

Sign-in to view your prices


comparethemarket.com™
THE EASIER WAY TO SAVE
Your vehicles
Your home & pet
Your finances
Your life
Your business
Your travels
How to get Oleg
Your previous quotes
Car insurance
Home insurance
Life insurance
Energy
Credit cards
Pet insurance

Welcome Eric [Not you?] | Your account

Hi Eric

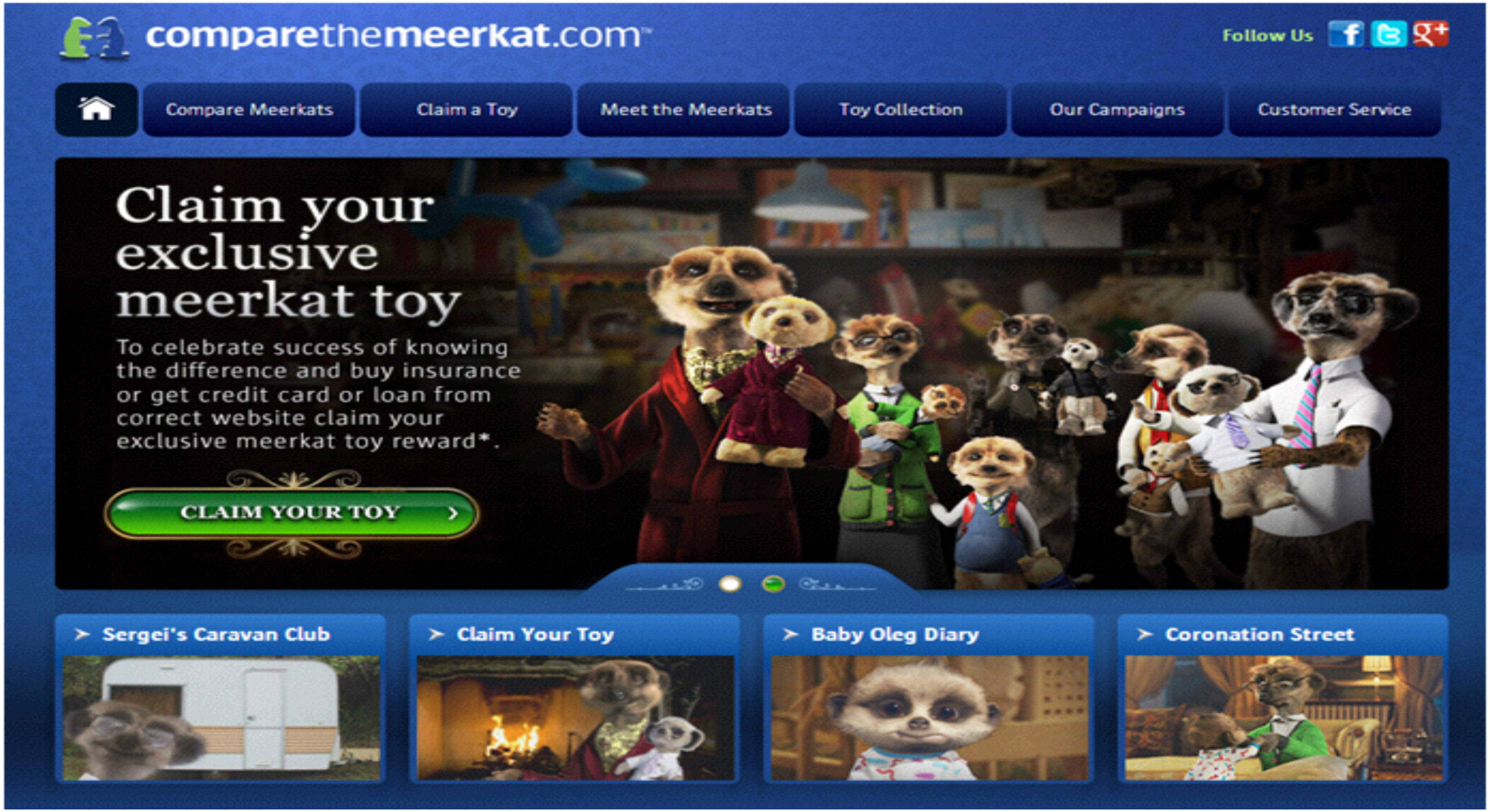
Adopt your own free Baby Oleg toy

when you buy any of the products below

Your recent car insurance quote

Your best price was: **£577.09**
Quote created on: **22/04/2014**

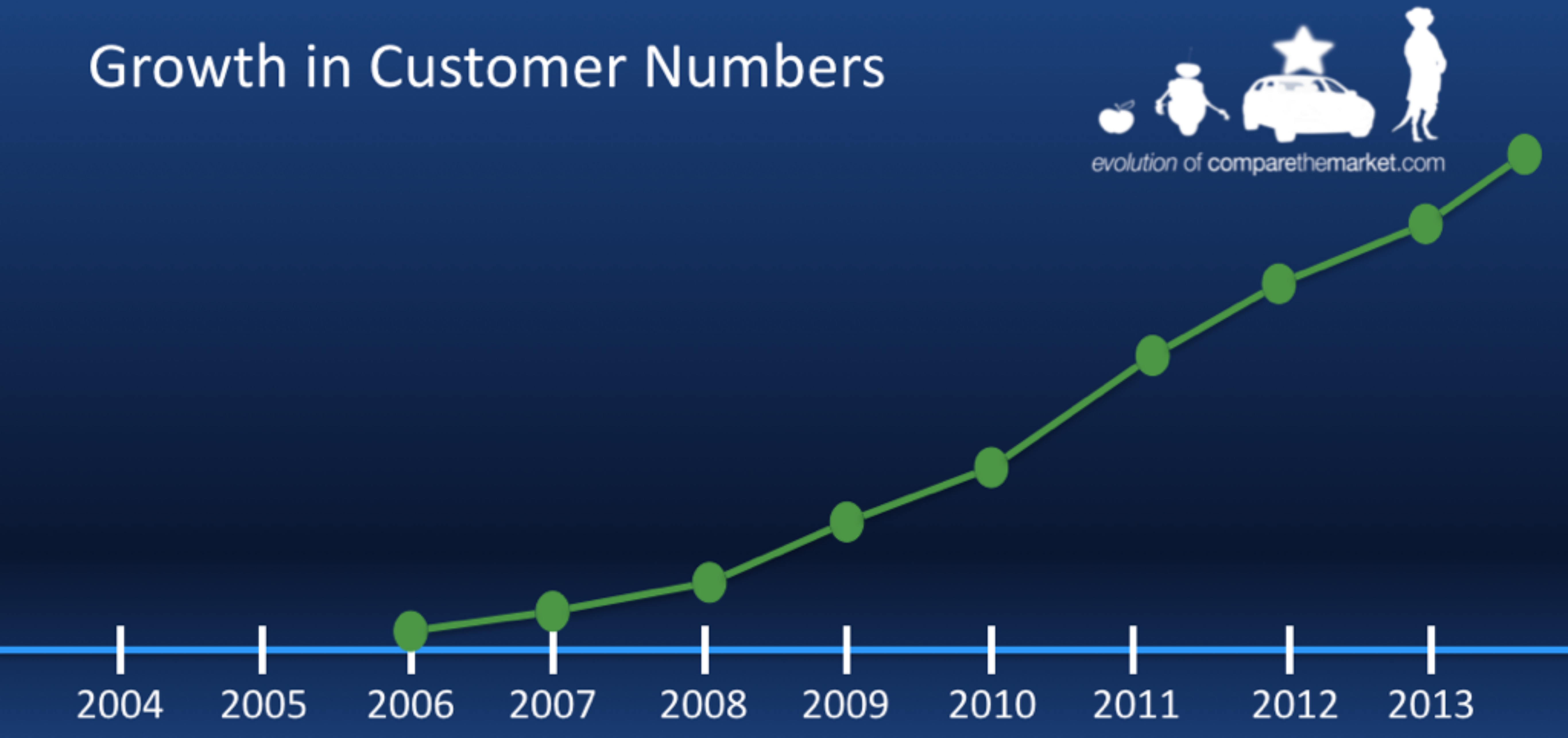
Choose your very own toy when you buy one of these


comparethemeerkat.com™
Follow Us
Compare Meerkats
Claim a Toy
Meet the Meerkats
Toy Collection
Our Campaigns
Customer Service
CLAIM YOUR TOY
Sergei's Caravan Club
Claim Your Toy
Baby Oleg Diary
Coronation Street

Claim your exclusive meerkat toy

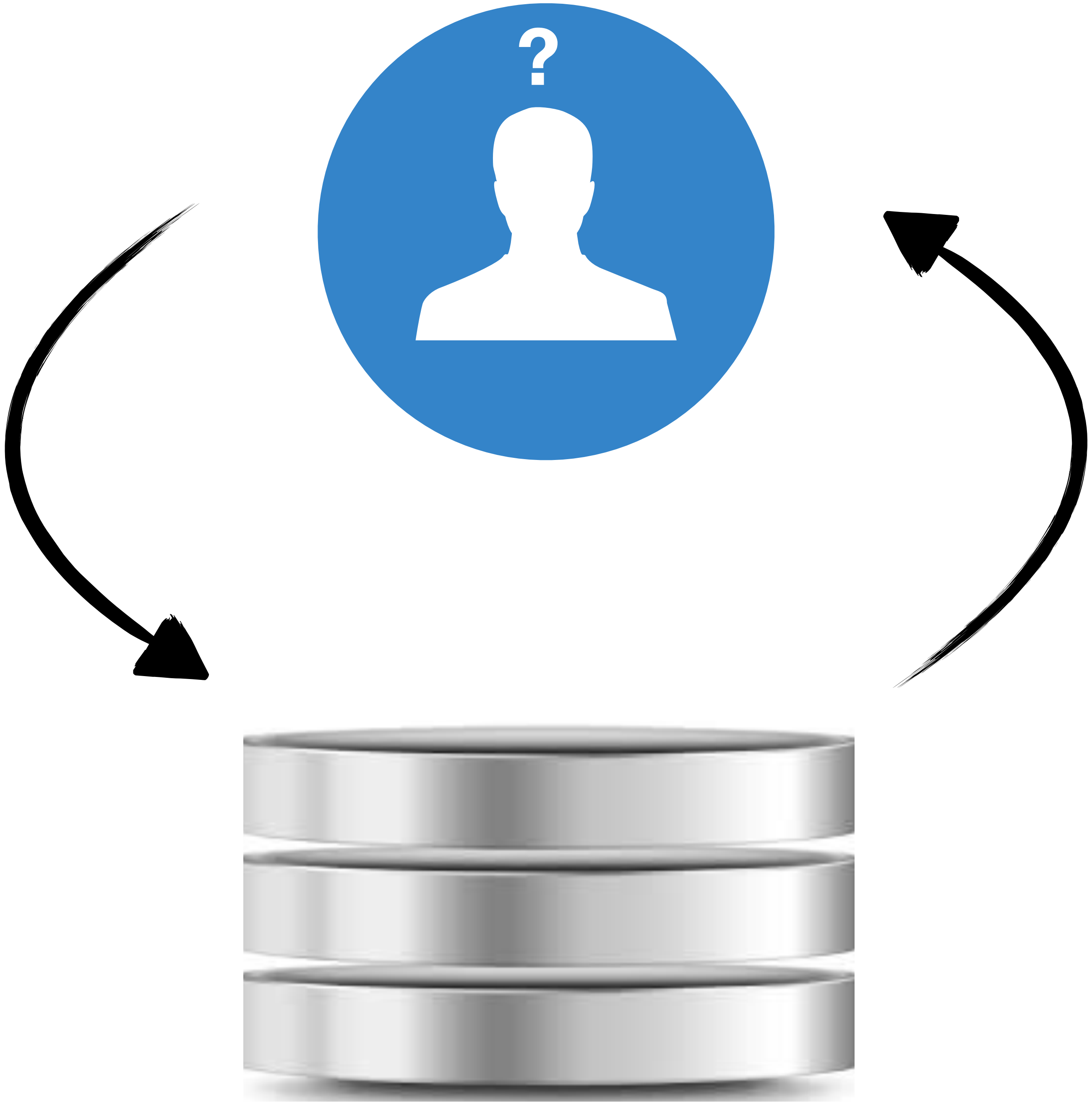
To celebrate success of knowing the difference and buy insurance or get credit card or loan from correct website claim your exclusive meerkat toy reward*.

Growth in Customer Numbers



Video

Inspiring Decisions



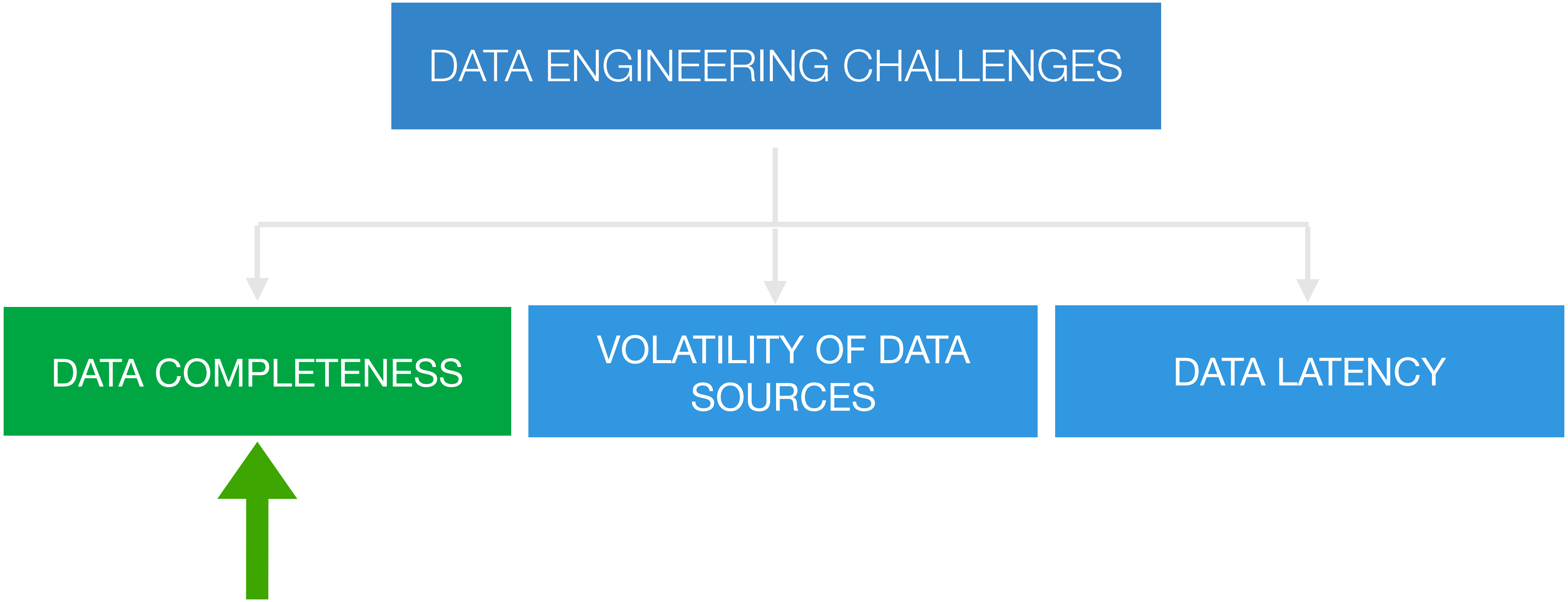
Challenges

Actions

Outcomes

DATA ENGINEERING CHALLENGES

- DATA COMPLETENESS
- VOLATILITY OF DATA SOURCES
- DATA LATENCY

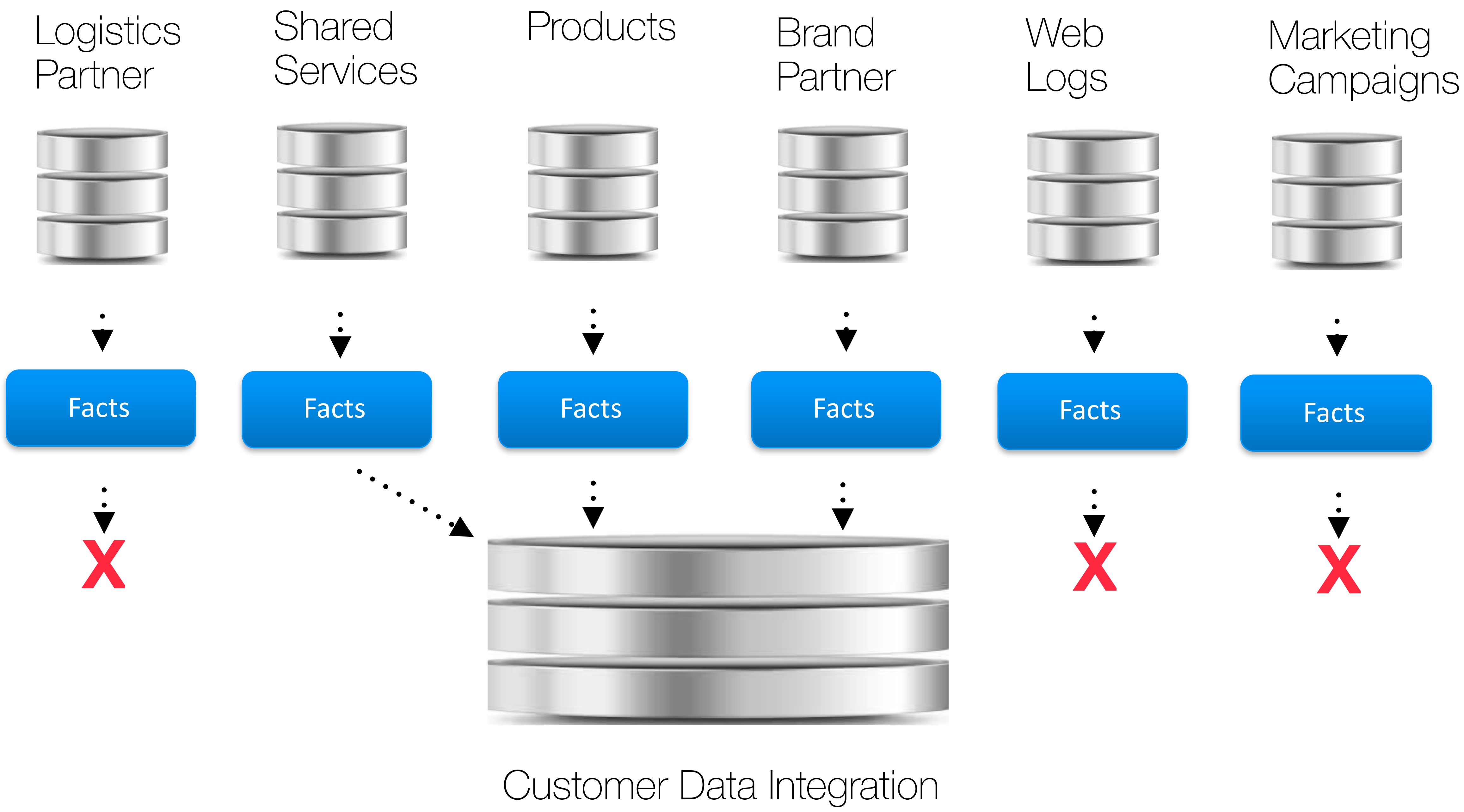


DATA ENGINEERING CHALLENGES

DATA COMPLETENESS

VOLATILITY OF DATA SOURCES

DATA LATENCY



Logistics Partner

Shared Services

Products

Brand Partner

Web Logs

Marketing Campaigns

Facts

Facts

Facts

Facts

Facts

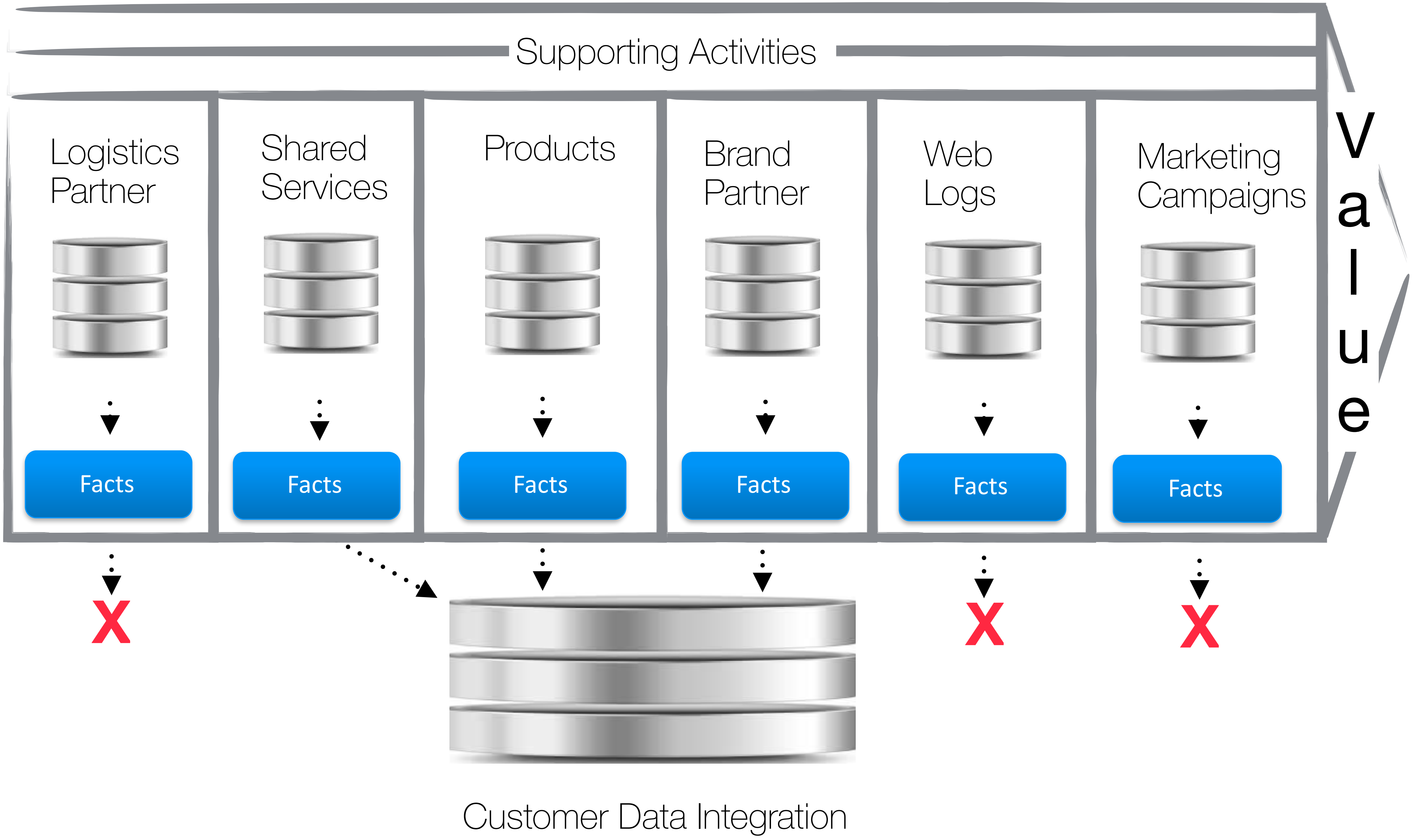
Facts

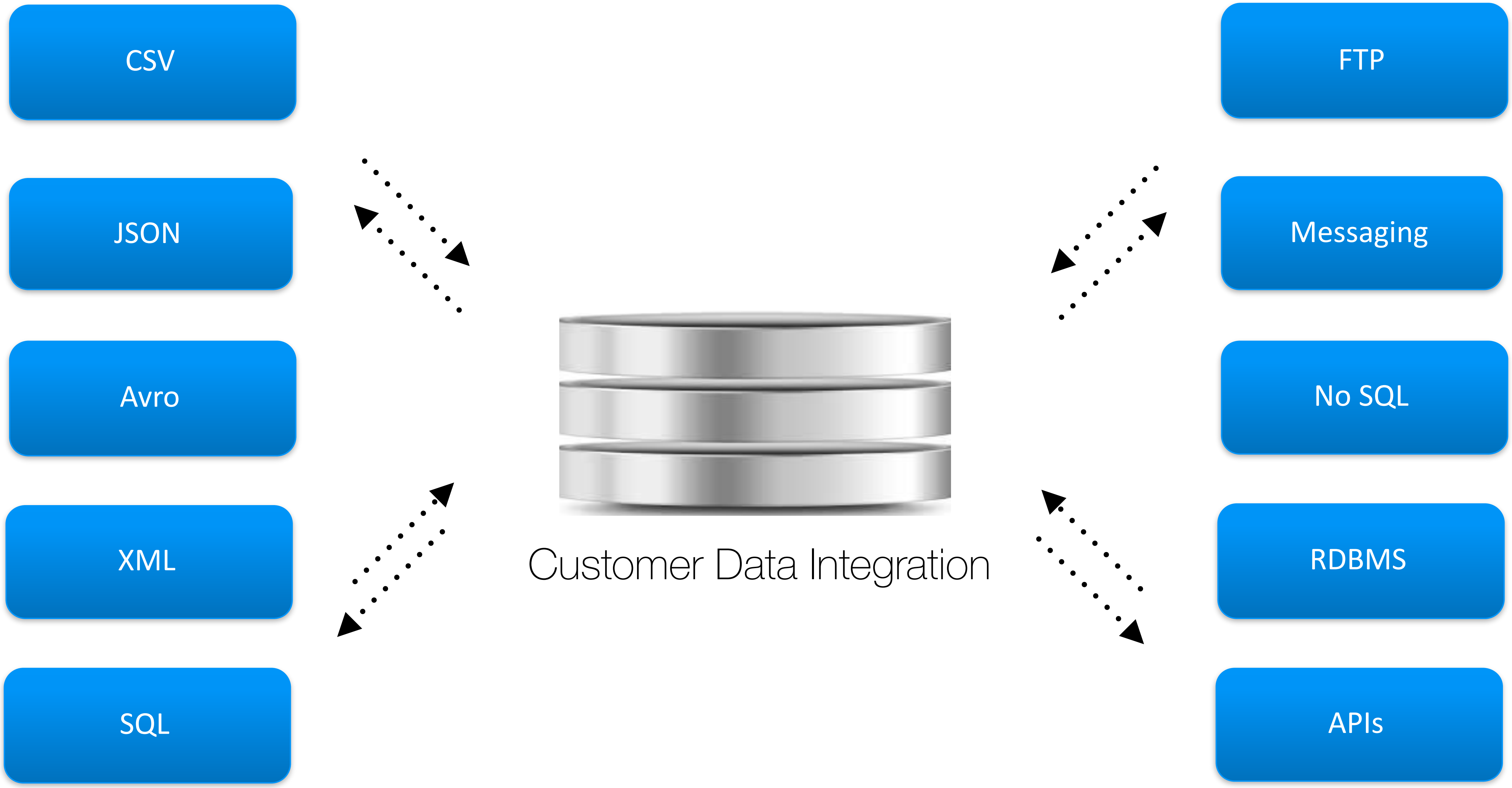
X

X

X

Customer Data Integration


Supporting Activities
Logistics Partner
Shared Services
Products
Brand Partner
Web Logs
Marketing Campaigns
Value
Facts
Facts
Facts
Facts
Facts
Facts
X
X
X
Customer Data Integration



CSV

JSON

Avro

XML

SQL

Customer Data Integration

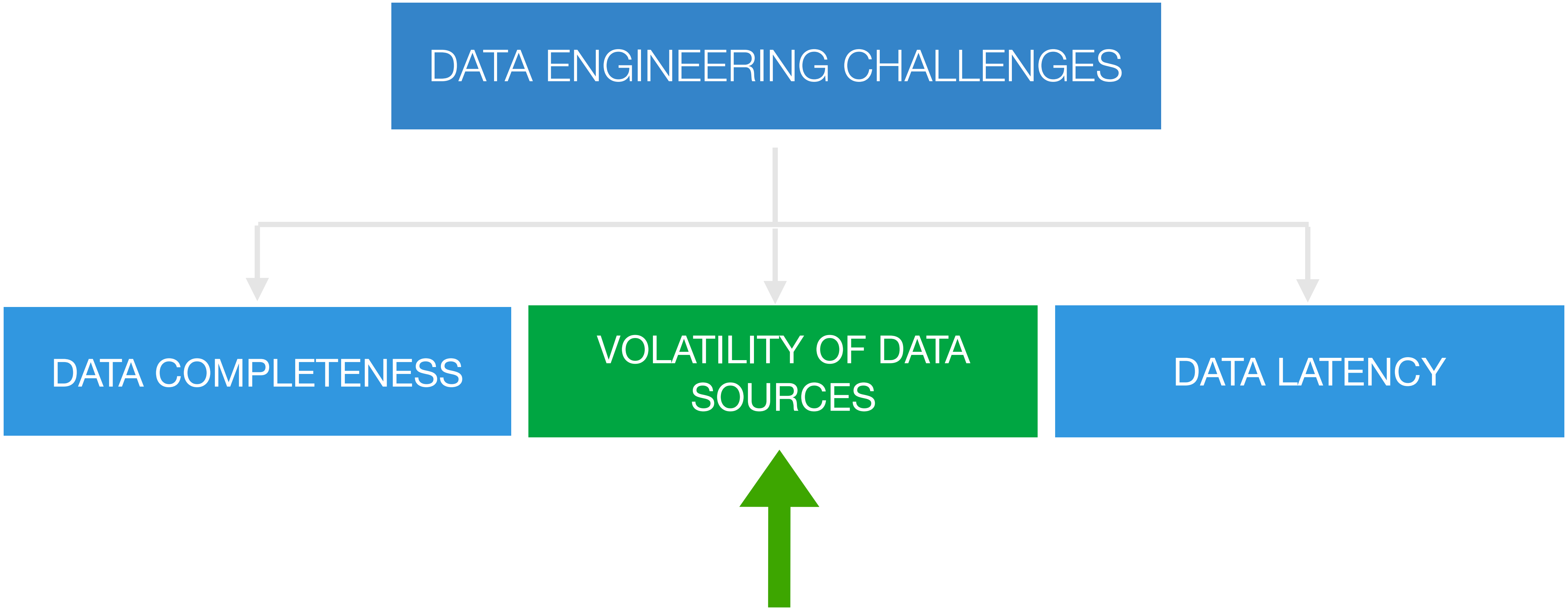
FTP

Messaging

No SQL

RDBMS

APIs



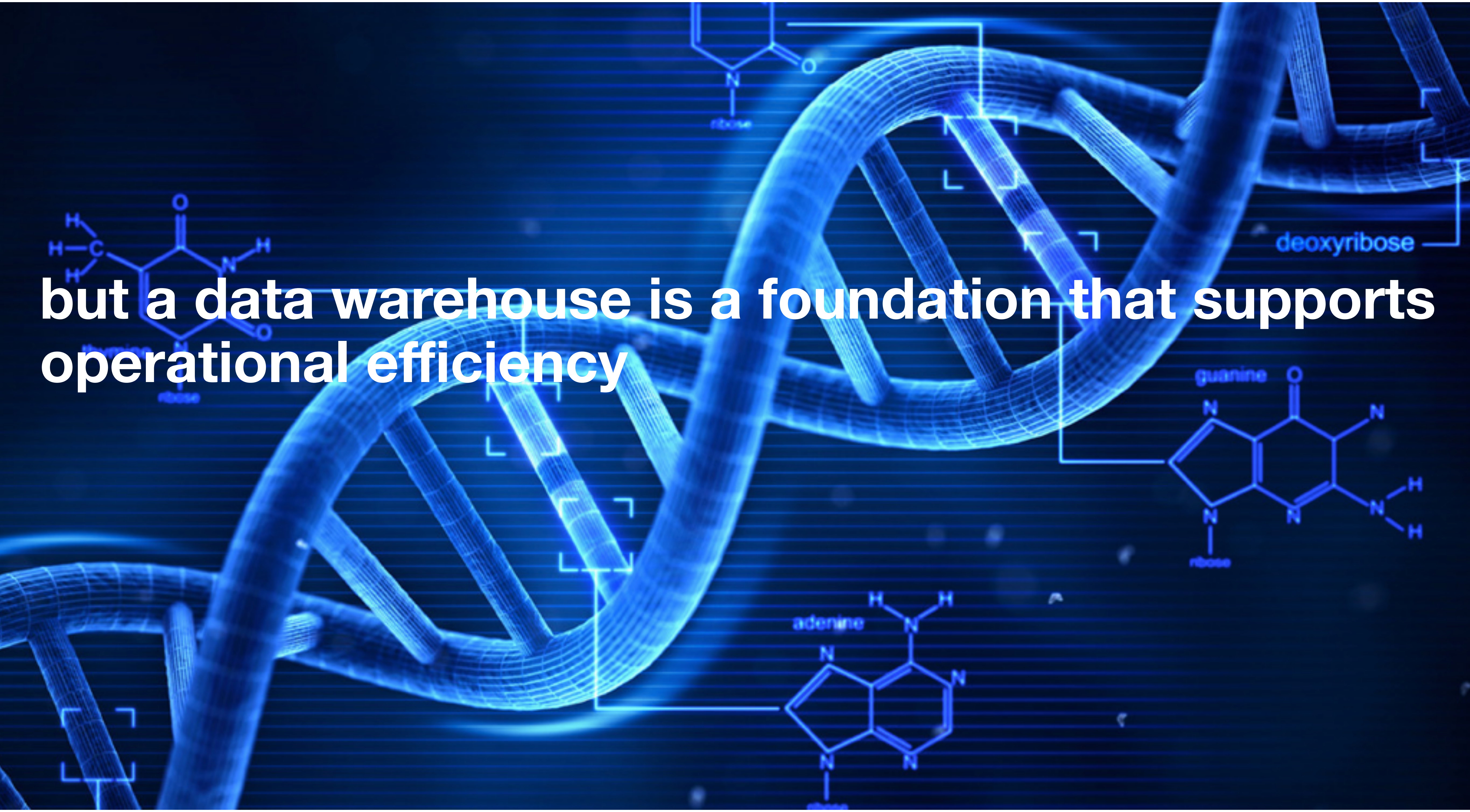
DATA ENGINEERING CHALLENGES

DATA COMPLETENESS

VOLATILITY OF DATA SOURCES

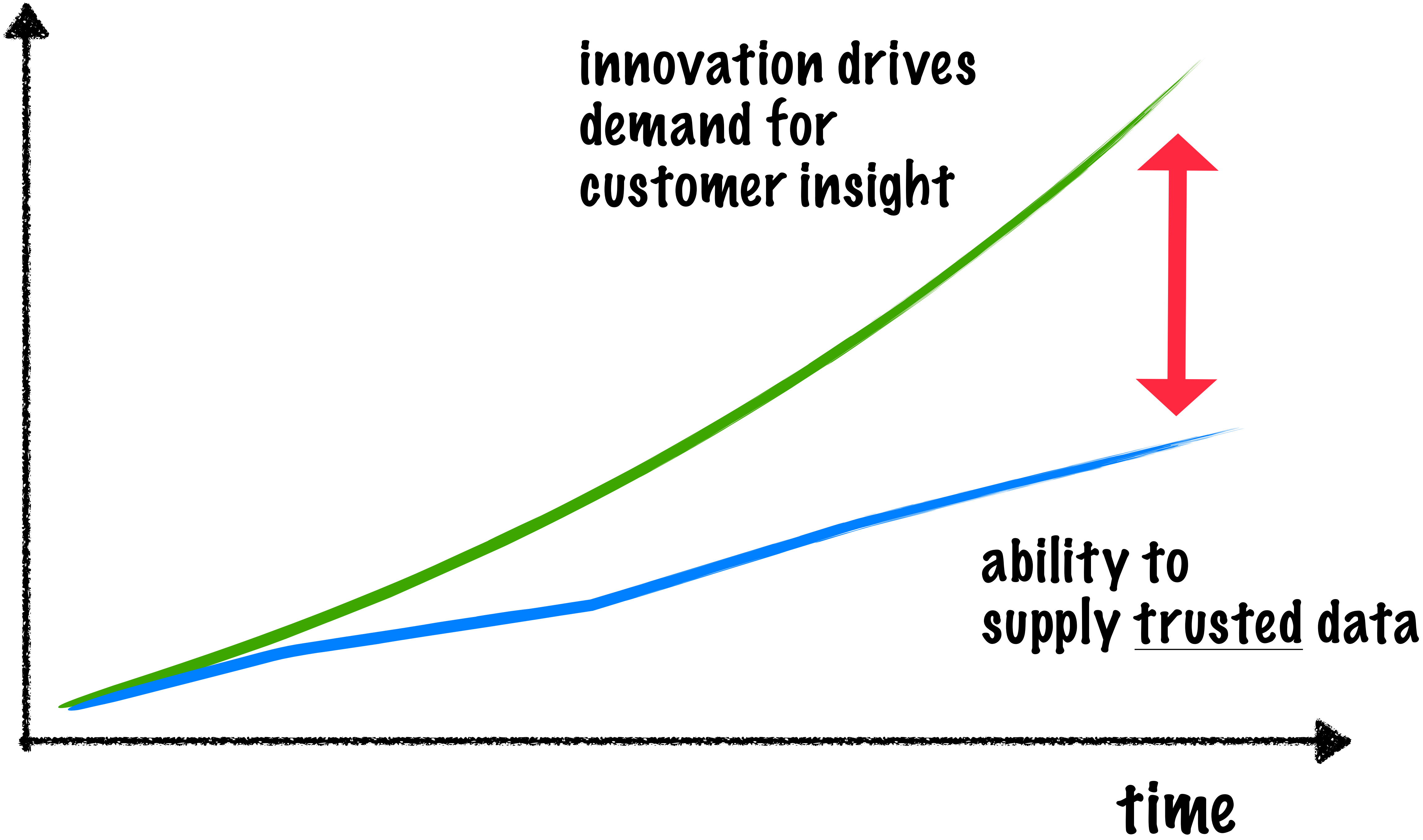
DATA LATENCY

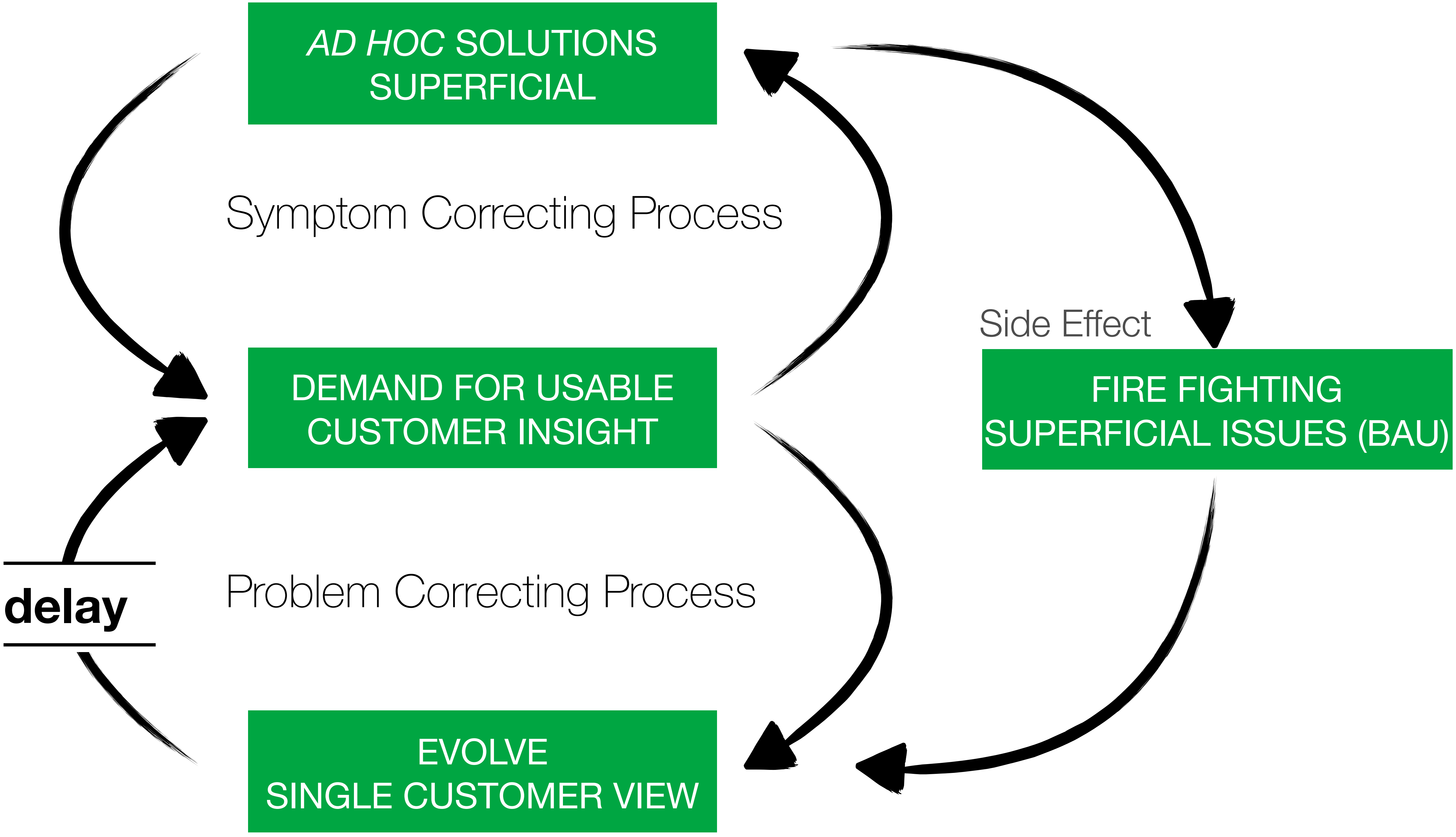
2013/14 Evolution stalling


deoxyribose
guanine
adenine
ribose

but a data warehouse is a foundation that supports operational efficiency

2013/14 Demand overloading supply





AD HOC SOLUTIONS
SUPERFICIAL

Symptom Correcting Process

Side Effect

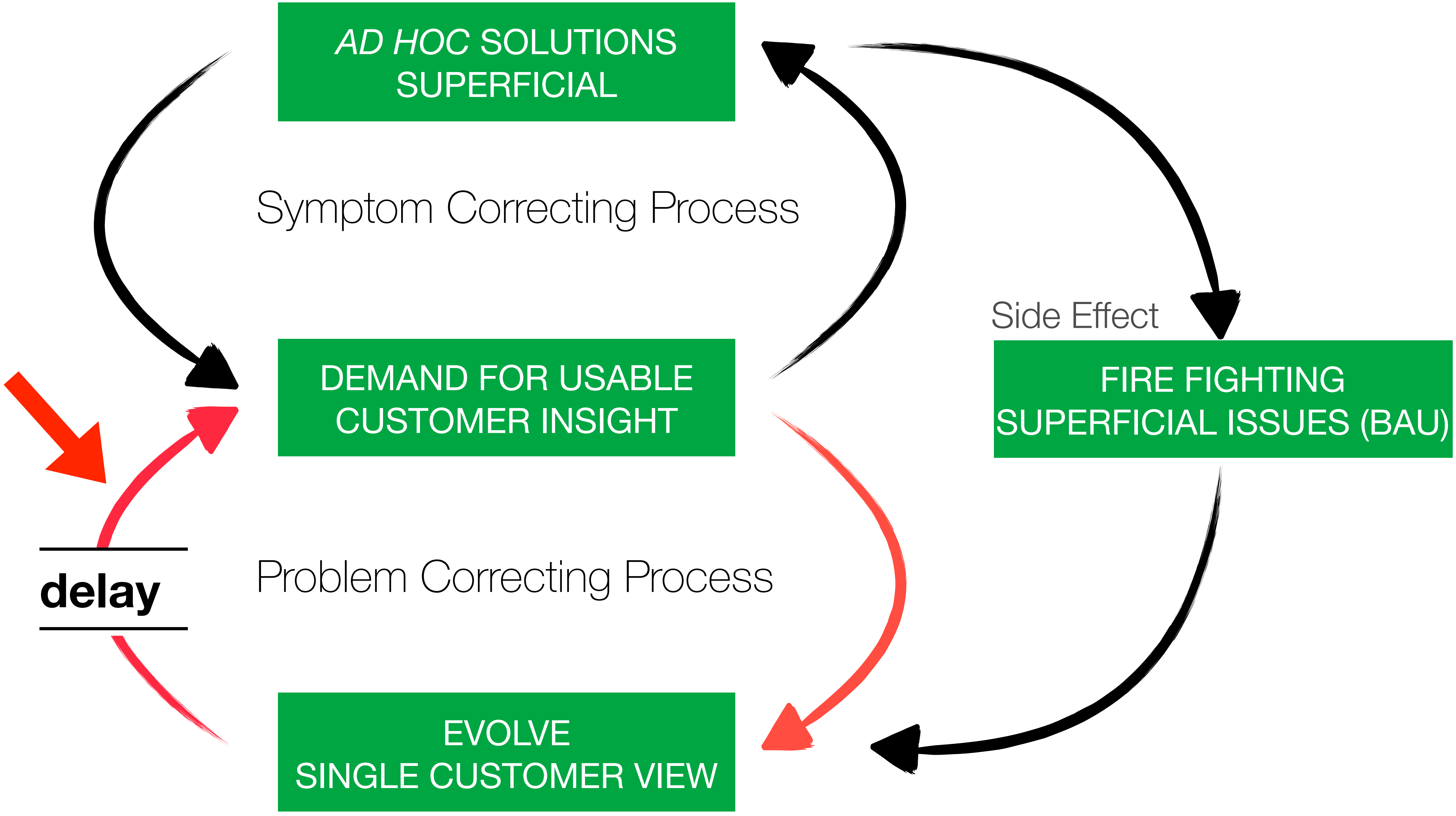
DEMAND FOR USABLE
CUSTOMER INSIGHT

FIRE FIGHTING
SUPERFICIAL ISSUES (BAU)

delay

Problem Correcting Process

EVOLVE
SINGLE CUSTOMER VIEW



AD HOC SOLUTIONS
SUPERFICIAL

Symptom Correcting Process

DEMAND FOR USABLE
CUSTOMER INSIGHT

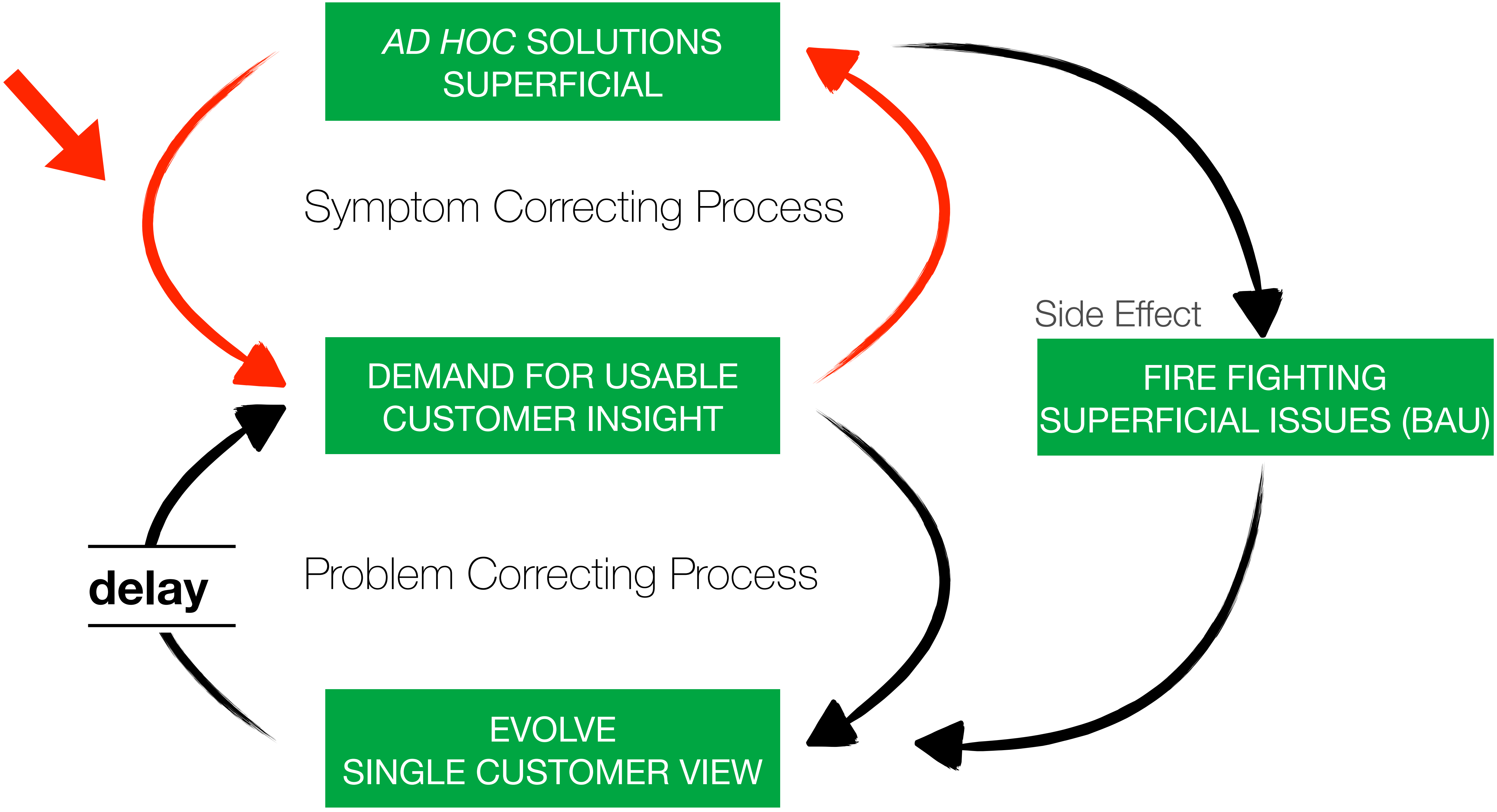
Side Effect

FIRE FIGHTING
SUPERFICIAL ISSUES (BAU)

delay

Problem Correcting Process

EVOLVE
SINGLE CUSTOMER VIEW



AD HOC SOLUTIONS
SUPERFICIAL

Symptom Correcting Process

DEMAND FOR USABLE
CUSTOMER INSIGHT

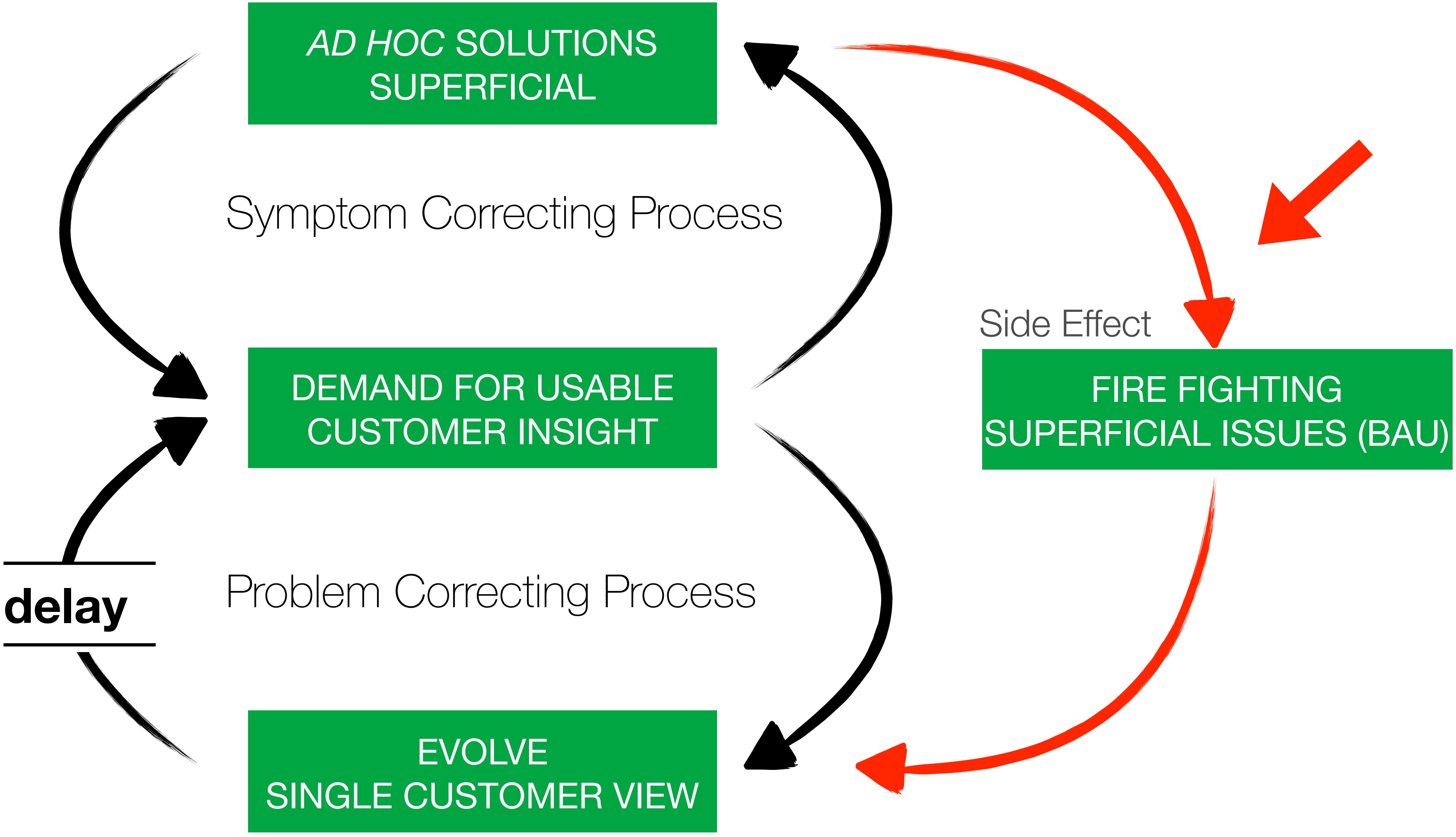
Side Effect

FIRE FIGHTING
SUPERFICIAL ISSUES (BAU)

delay

Problem Correcting Process

EVOLVE
SINGLE CUSTOMER VIEW



AD HOC SOLUTIONS
SUPERFICIAL

Symptom Correcting Process

DEMAND FOR USABLE
CUSTOMER INSIGHT

Side Effect

FIRE FIGHTING
SUPERFICIAL ISSUES (BAU)

delay

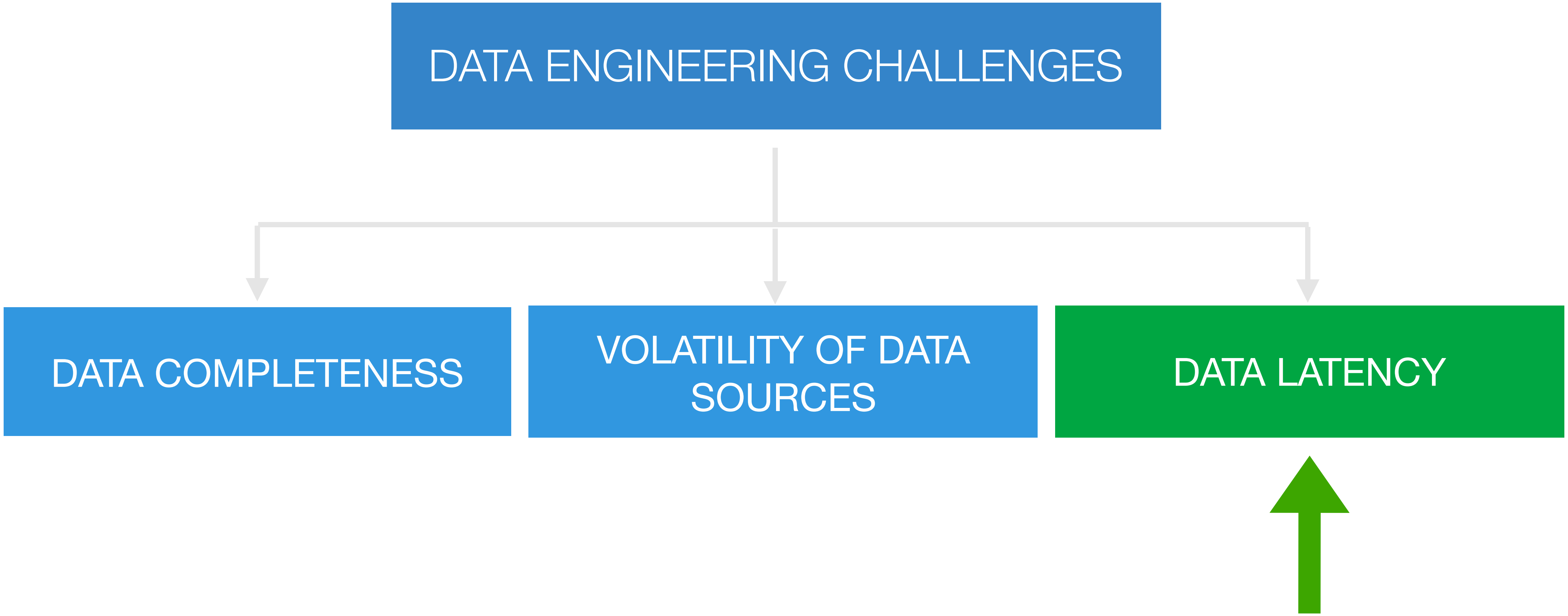
Problem Correcting Process

EVOLVE
SINGLE CUSTOMER VIEW

2013/14 What is causing the **delay** ?

- Accumulating WIP due to cost of batch execution
- Test data not representative of real data
- Cost of human error and reprocessing
- NoSQL No Response

2013/14 Issue Tree



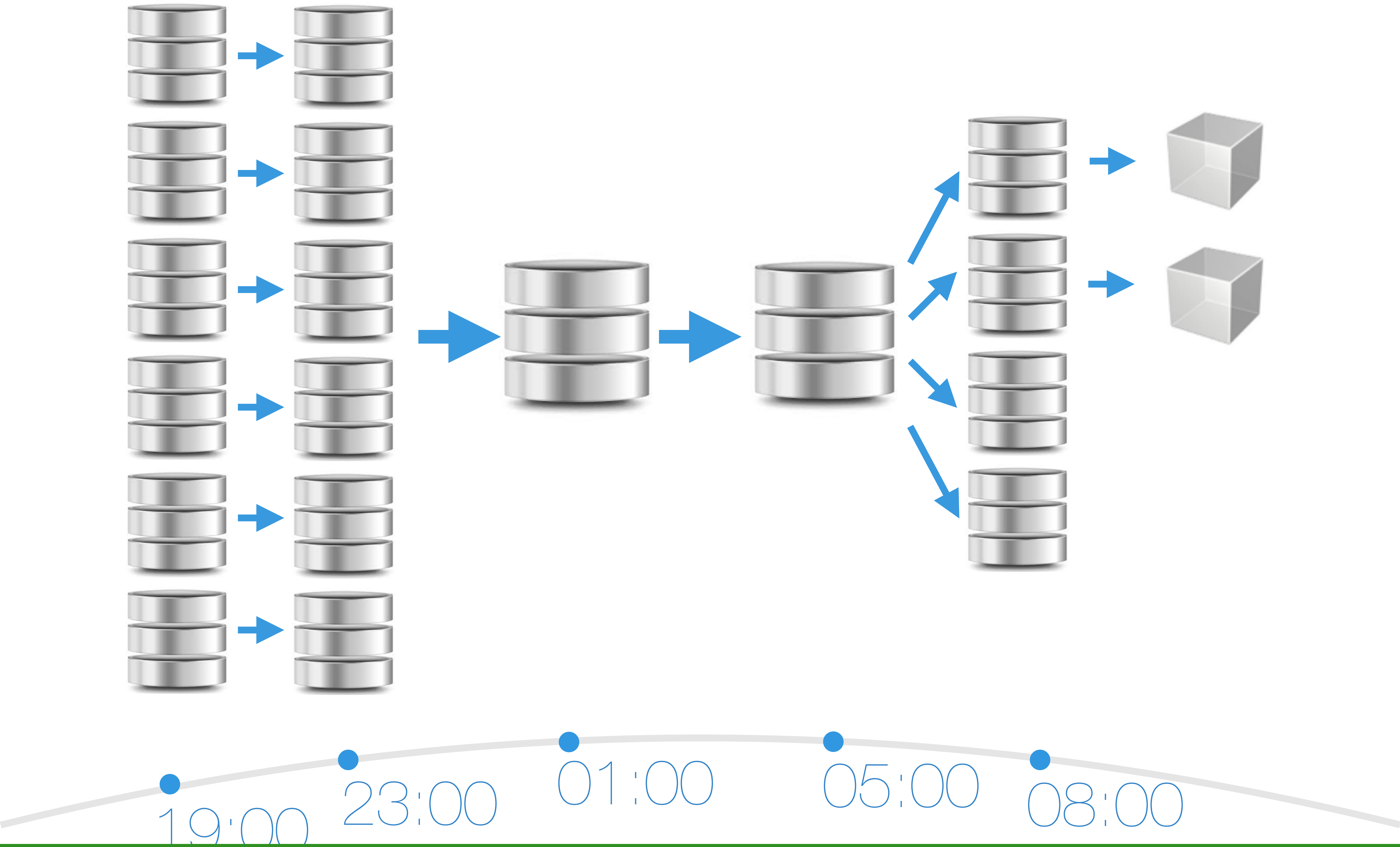
DATA ENGINEERING CHALLENGES

DATA COMPLETENESS

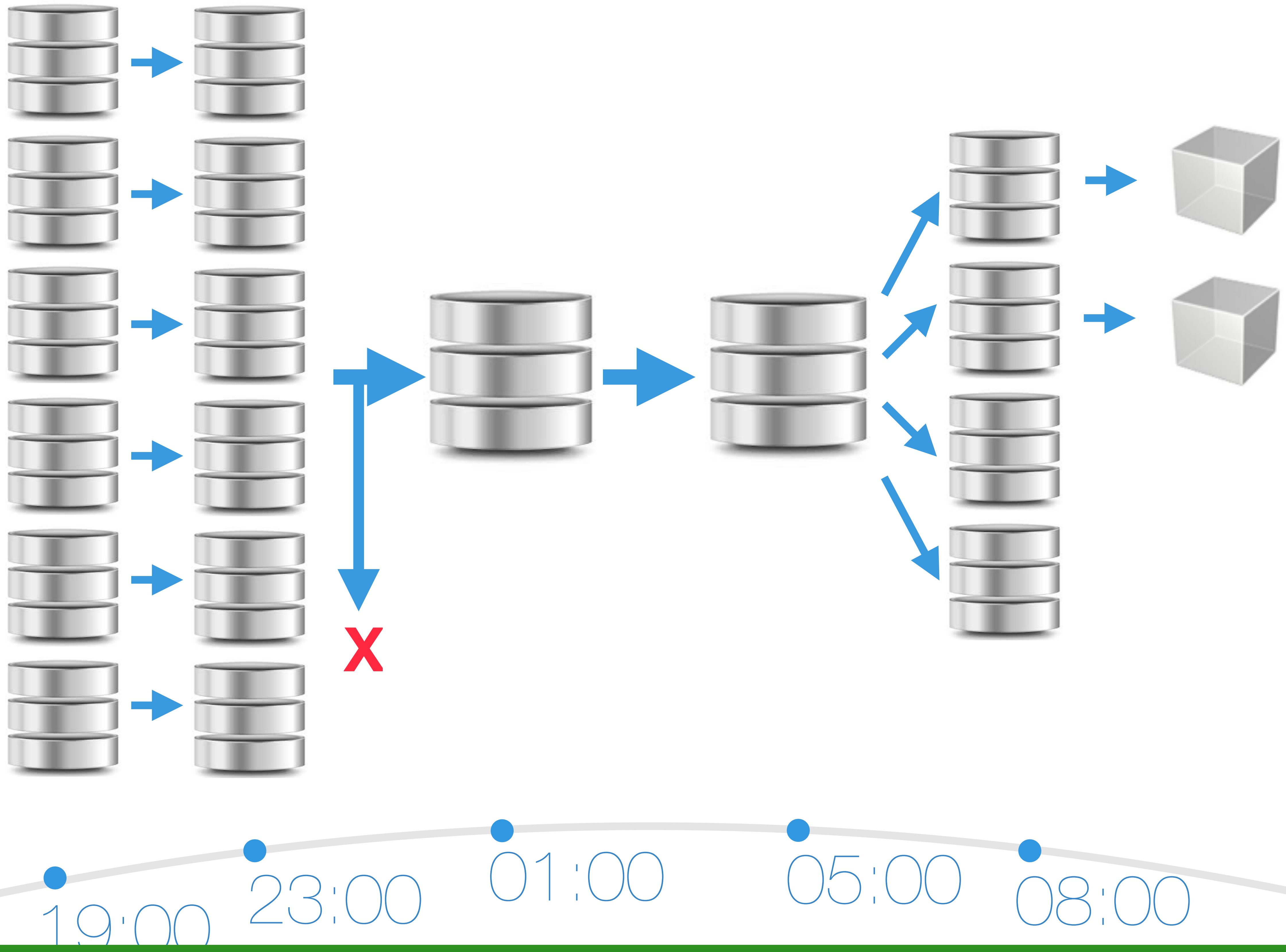
VOLATILITY OF DATA SOURCES

DATA LATENCY

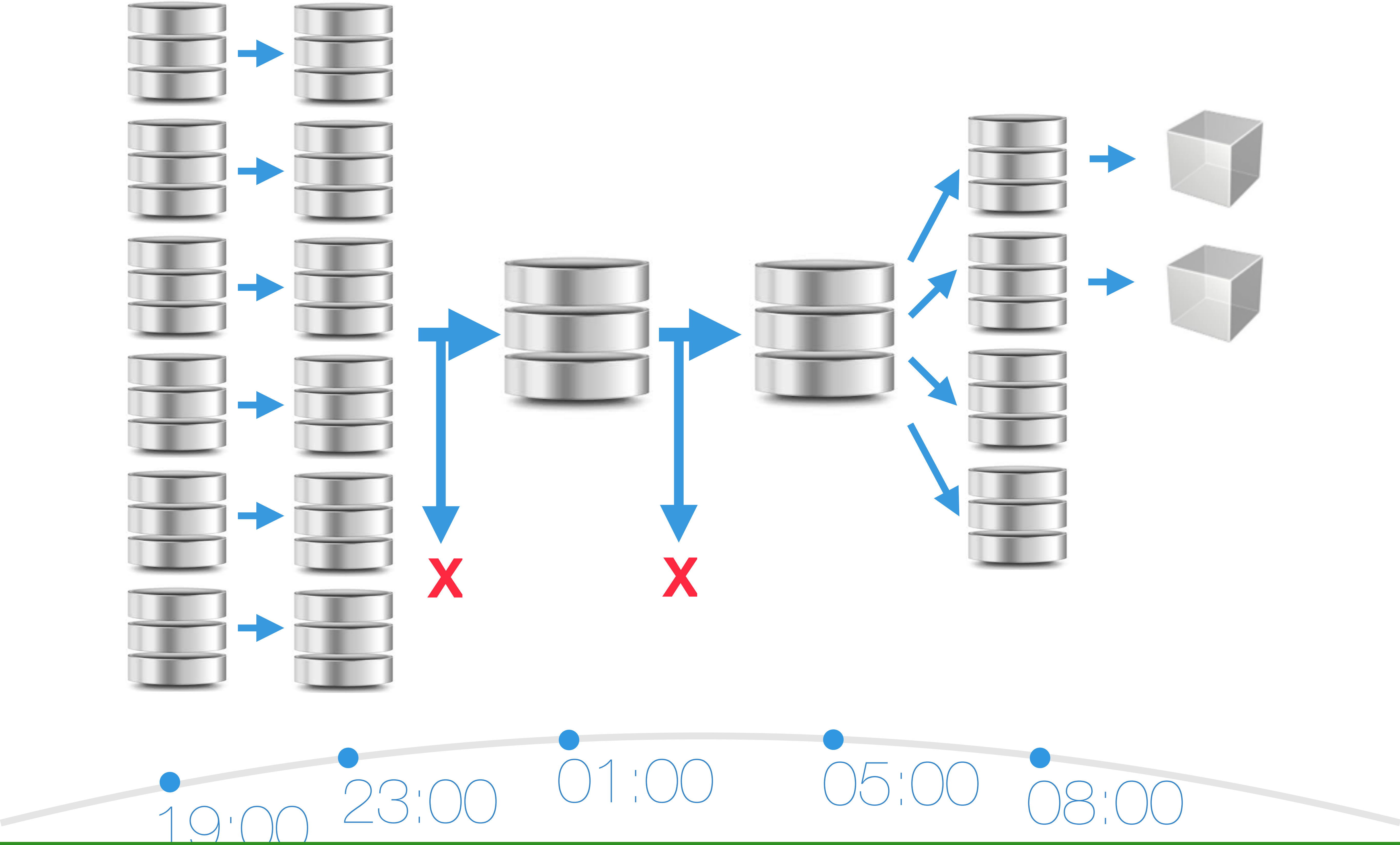
2013/14 Performance, coupling and synchronisation


19:00
23:00
01:00
05:00
08:00

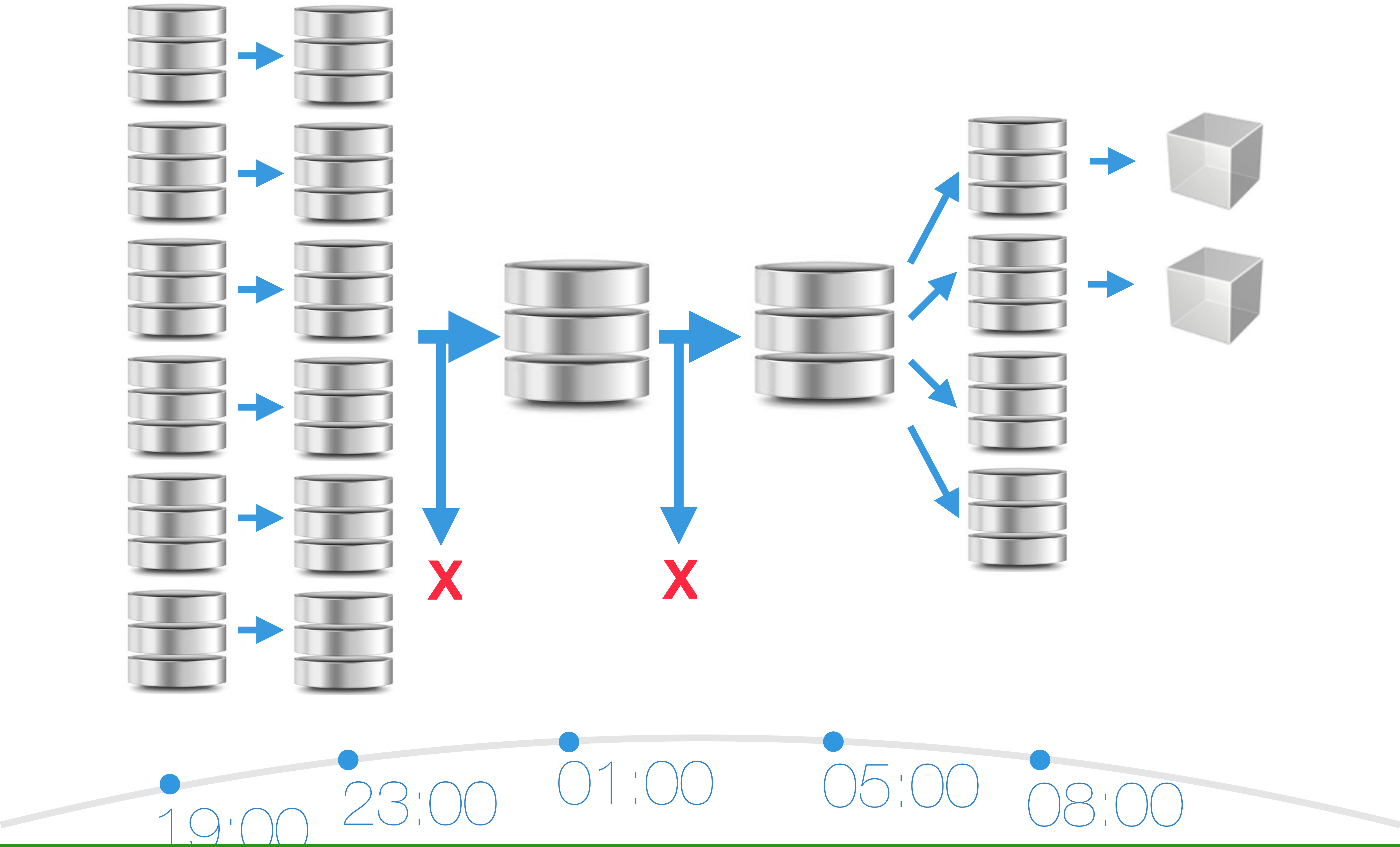
2013/14 Challenges - Performance


X
19:00
23:00
01:00
05:00
08:00

2013/14 Challenges - Performance


X
X
19:00
23:00
01:00
05:00
08:00

2013/14 Challenges - Performance



19:00
23:00
01:00
05:00
08:00

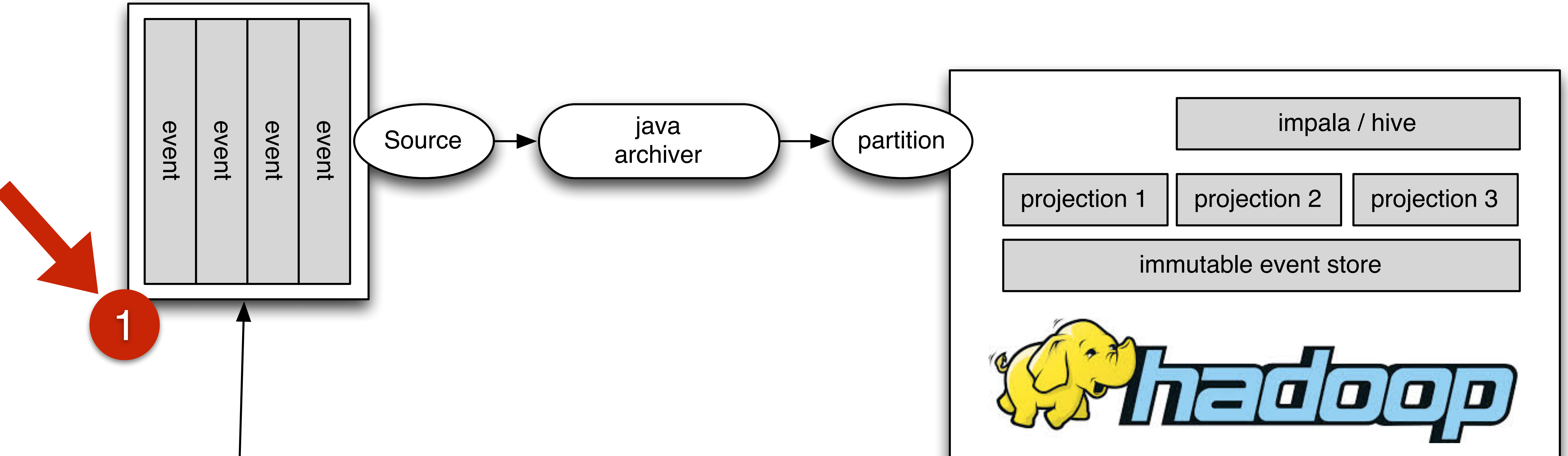
2013/14 Actions - Create a vision

1. Data Definitions, MDM and Governance
2. Operational Analytics
3. Tactical Data Warehouse
4. Visualisation and Tools
5. Next Generation Infrastructure
6. Interfaces to Data
7. Leverage CRM
8. Real-Time Decisions and Personalisation

- Immutable raw data (events) and refine
- Real data where possible
- Schema later, but it is still important
- Loosely coupled pipelines
- Increase data velocity

Actions

PoC 1



1

event

event

event

event

Source

java archiver

partition

impala / hive

projection 1

projection 2

projection 3

immutable event store

hadoop

comparethemarket.com
Welcome | Sign in
Vehicles
Home & pet
Finances
Life
Business
Travel

Hello

Get your own free Limited Editions Safari Oleg toy

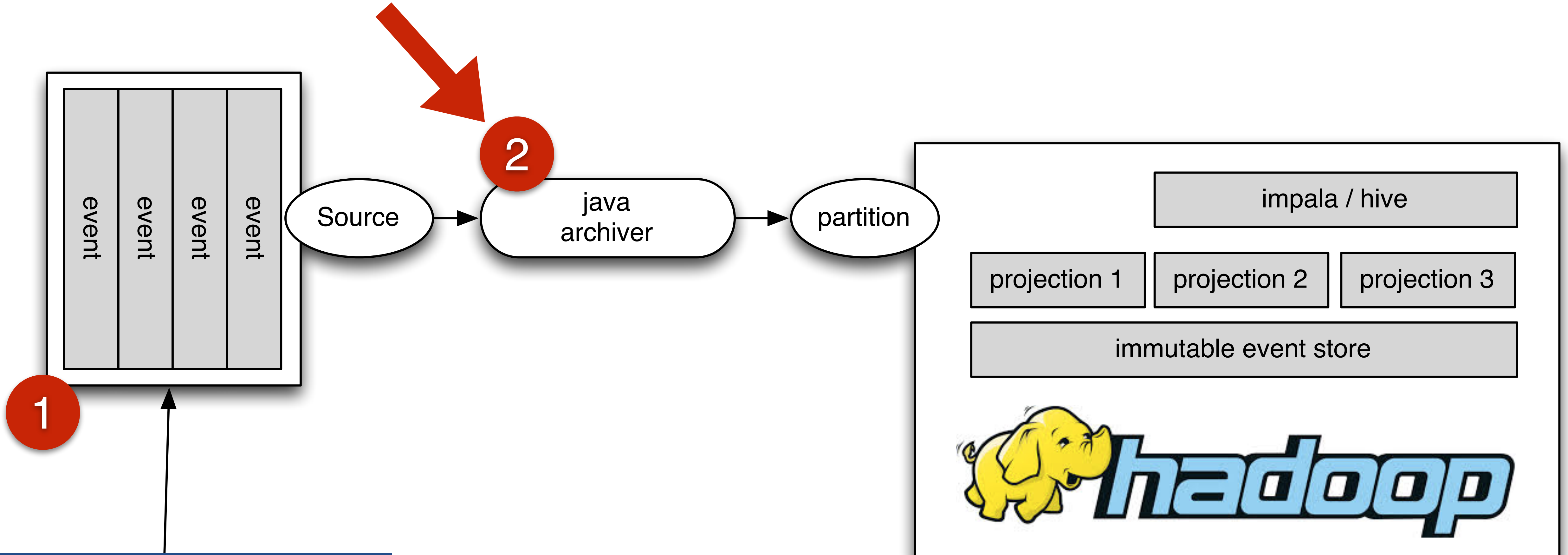
when you buy car or home insurance or switch energy

How to get Oleg

Had a quote from us before? Save time by updating your saved quote.

Your previous quotes
Let's start comparing...
Car insurance
Home insurance
Life insurance
Energy
Credit cards
Travel insurance

PoC 1


2
1
event
event
event
event
Source
java archiver
partition
impala / hive
projection 1
projection 2
projection 3
immutable event store
hadoop

comparethemarket.com
Welcome | Sign in
Vehicles
Home & pet
Finances
Life
Business
Travel
How to get Oleg
Your previous quotes
Car Insurance
Home Insurance
Life Insurance
Energy
Credit cards
Travel Insurance

Hello

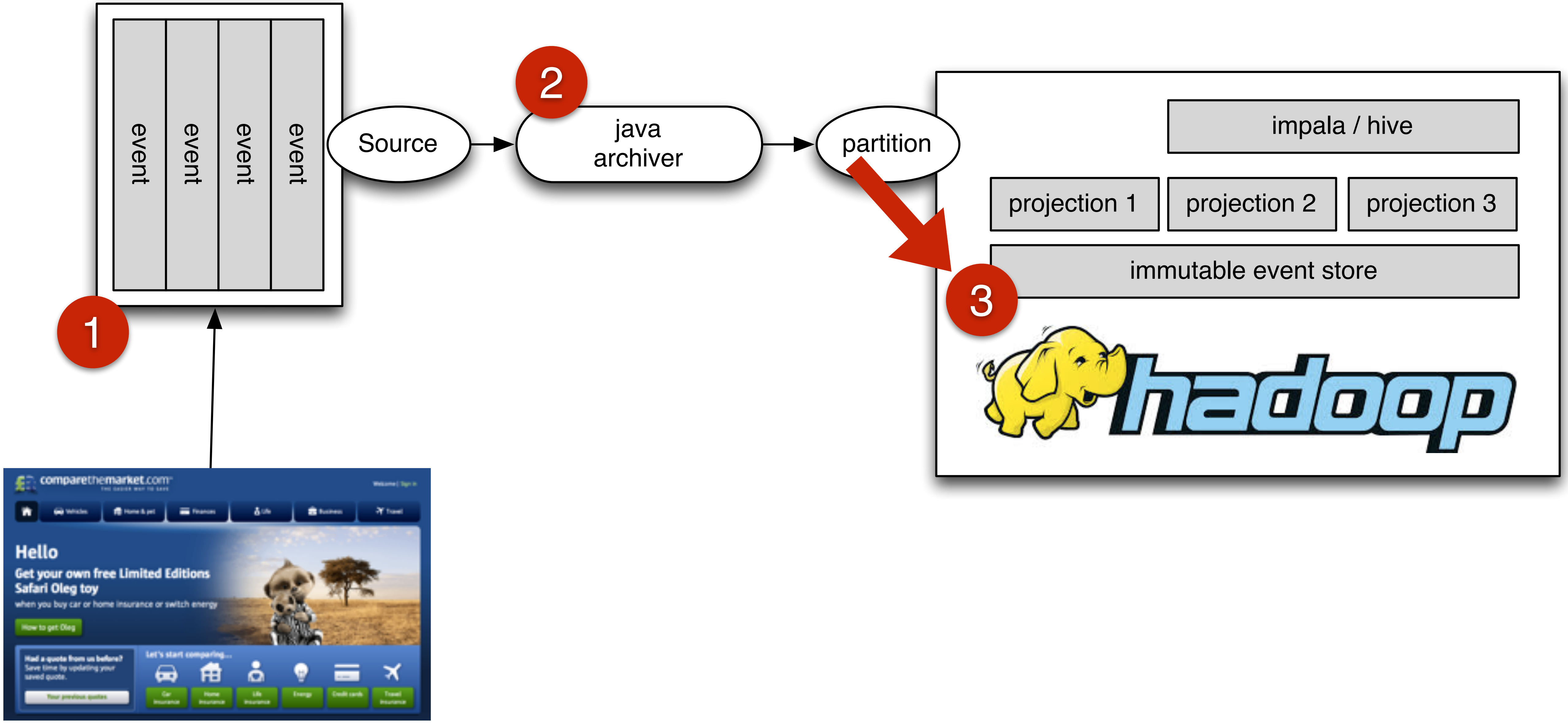
Get your own free Limited Editions Safari Oleg toy

when you buy car or home insurance or switch energy

Had a quote from us before? Save time by updating your saved quote.

Let's start comparing...

PoC 1



event

event

event

event

Source

java archiver

partition

impala / hive

projection 1

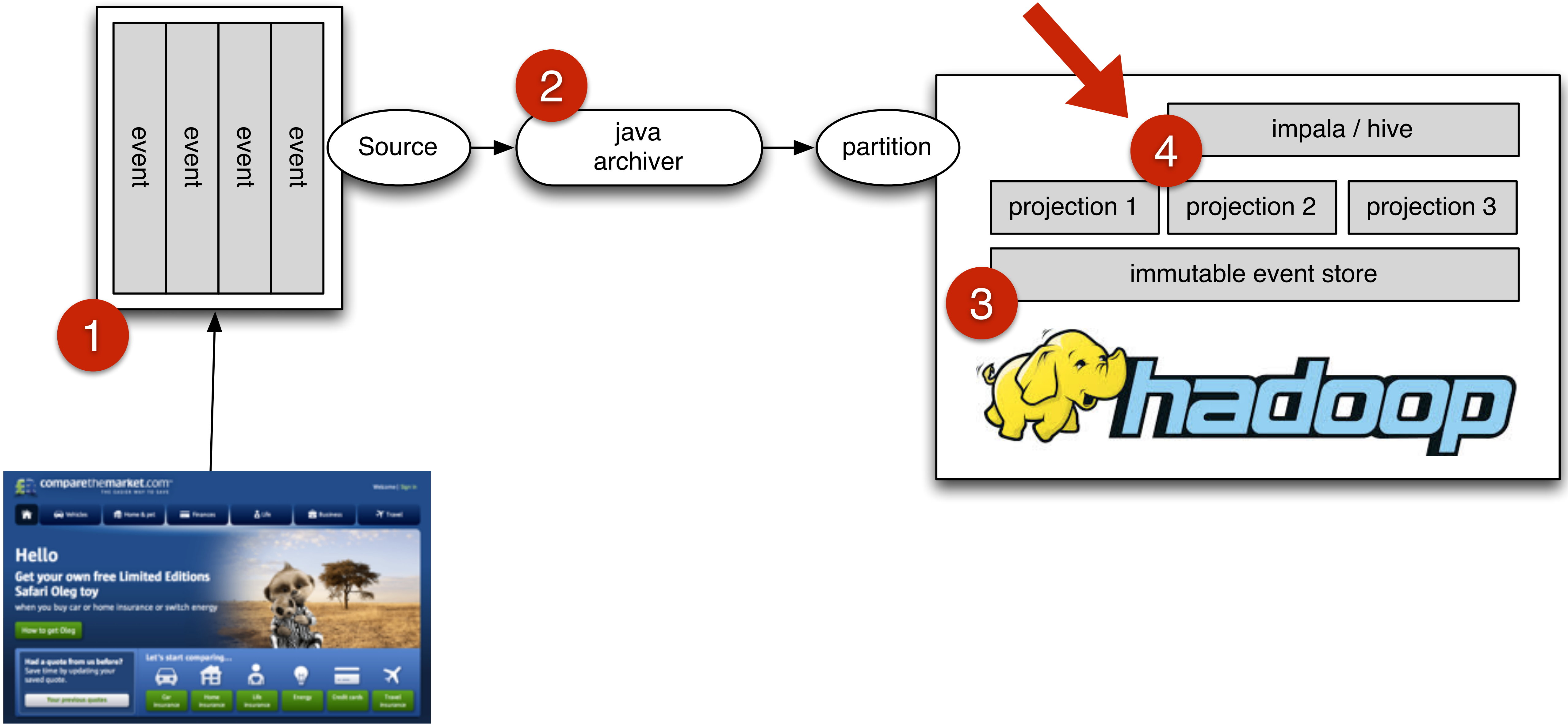
projection 2

projection 3

immutable event store

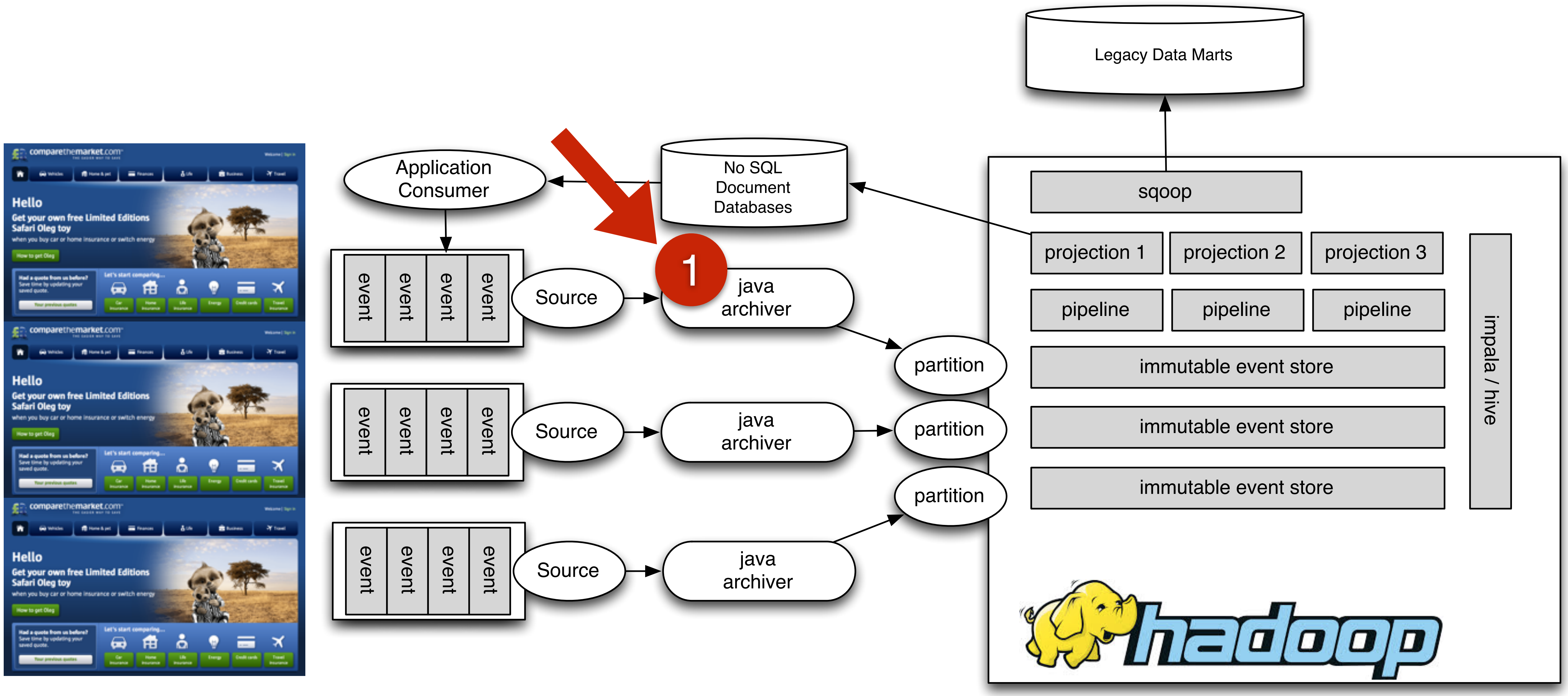
1
2
3
hadoop
comparethemarket.com
Hello
Get your own free Limited Editions Safari Oleg toy
when you buy car or home insurance or switch energy
How to get Oleg
Let's start comparing...

PoC 1



event
event
event
event
1
Source
2
java archiver
partition
4
impala / hive
projection 1
projection 2
projection 3
immutable event store
3
hadoop
comparethemarket.com
Hello
Get your own free Limited Editions Safari Oleg toy
when you buy car or home insurance or switch energy
How to get Oleg
Had a quote from us before?
Let's start comparing...

PoC 2



Legacy Data Marts

Application Consumer

No SQL Document Databases

1

event event event event

Source

java archiver

event event event event

Source

java archiver

event event event event

Source

java archiver

partition

partition

partition

sqoop

projection 1

projection 2

projection 3

pipeline

pipeline

pipeline

immutable event store

immutable event store

immutable event store

impala / hive

hadoop

PoC 2

comparethemarket.com

Hello

Get your own free Limited Editions Safari Oleg toy

when you buy car or home insurance or switch energy

Hello

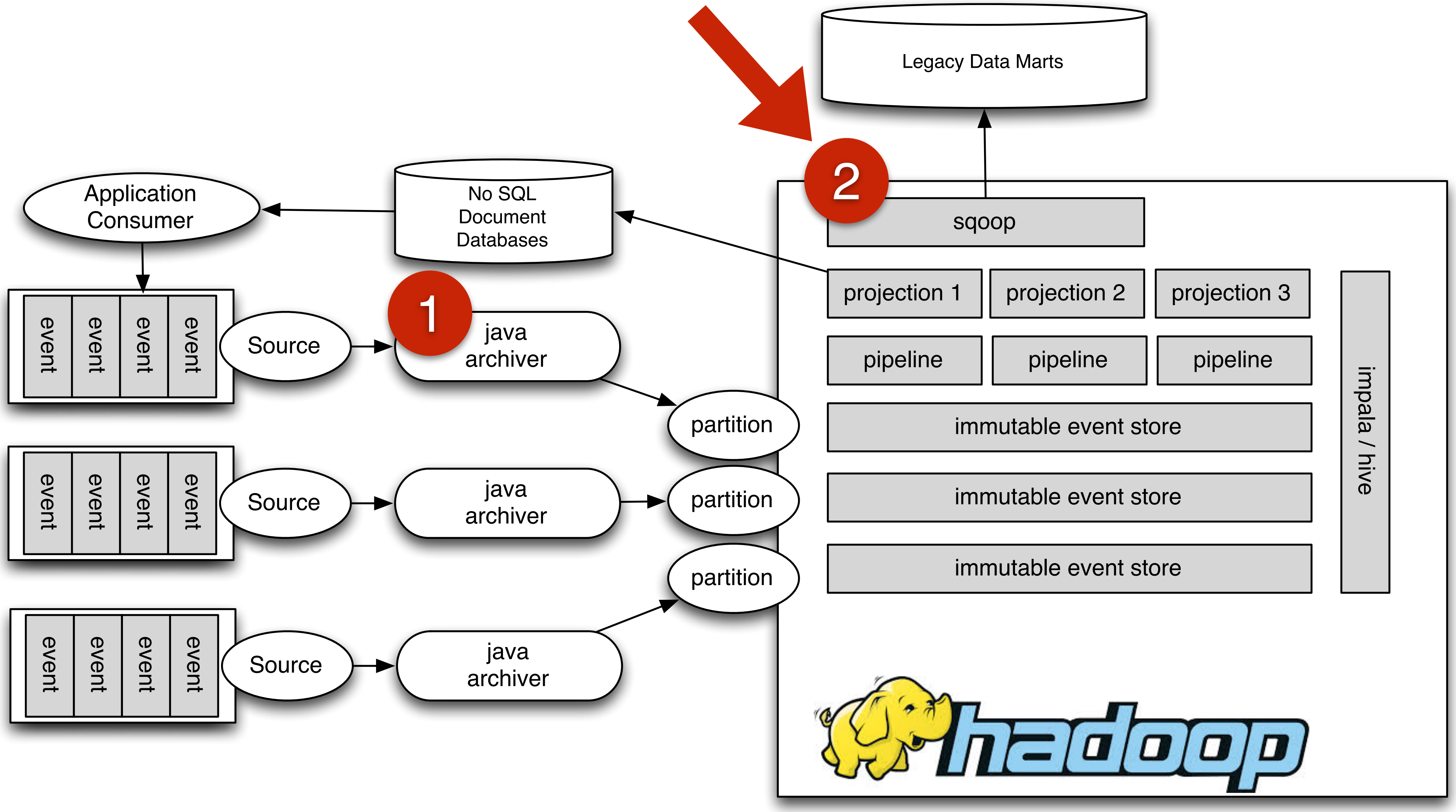
Get your own free Limited Editions Safari Oleg toy

when you buy car or home insurance or switch energy

Hello

Get your own free Limited Editions Safari Oleg toy

when you buy car or home insurance or switch energy



Legacy Data Marts

Application Consumer

No SQL Document Databases

event event event event

Source

① java archiver

partition

event event event event

Source

java archiver

partition

event event event event

Source

java archiver

partition

② sqoop

projection 1

projection 2

projection 3

pipeline

pipeline

pipeline

immutable event store

immutable event store

immutable event store

impala / hive

hadoop

PoC 2

Hello

Get your own free Limited Editions Safari Oleg toy

when you buy car or home insurance or switch energy

Hello

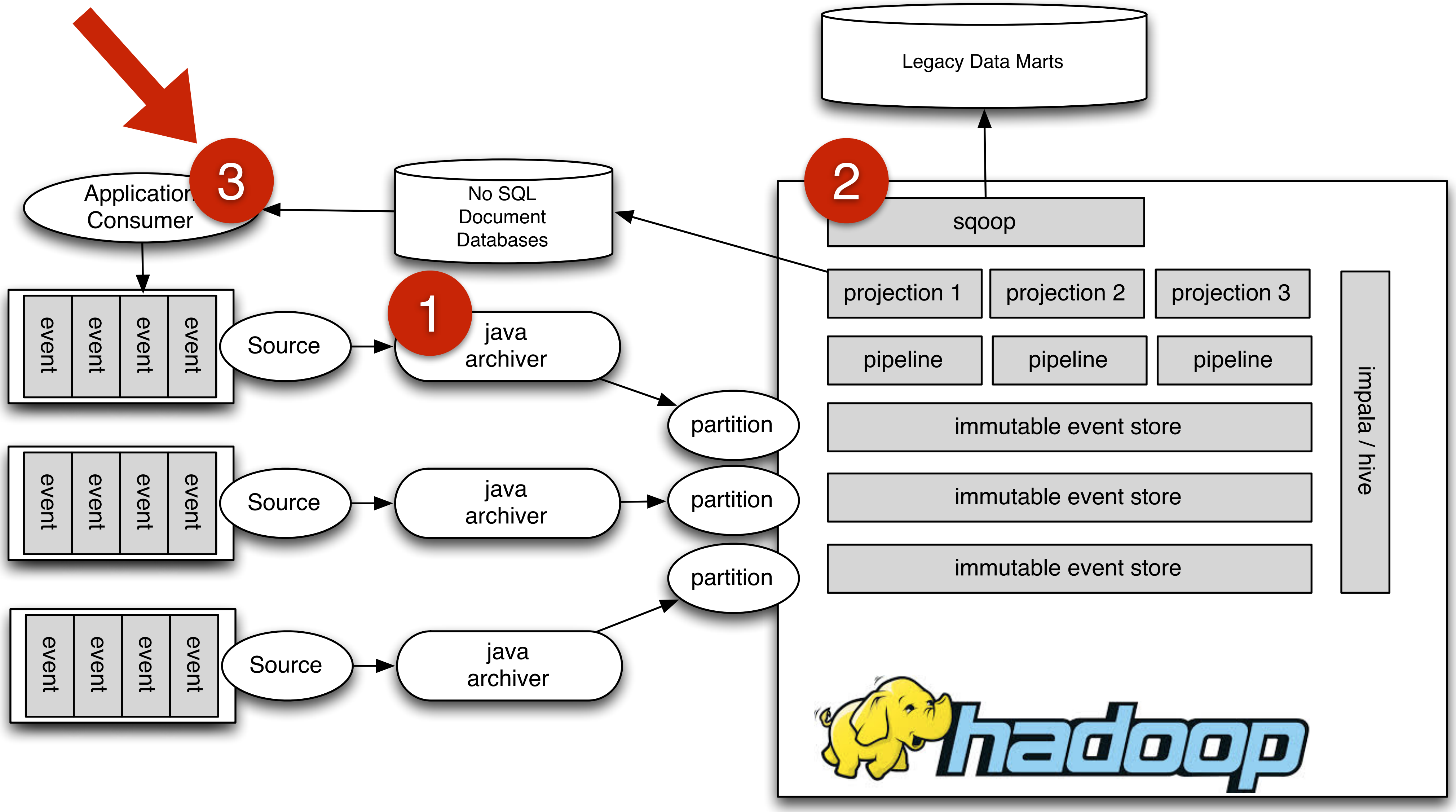
Get your own free Limited Editions Safari Oleg toy

when you buy car or home insurance or switch energy

Hello

Get your own free Limited Editions Safari Oleg toy

when you buy car or home insurance or switch energy



Legacy Data Marts

3

Application Consumer

No SQL Document Databases

2

sqoop

projection 1

projection 2

projection 3

pipeline

pipeline

pipeline

immutable event store

immutable event store

immutable event store

impala / hive

event event event event

Source

1

java archiver

partition

event event event event

Source

java archiver

partition

partition

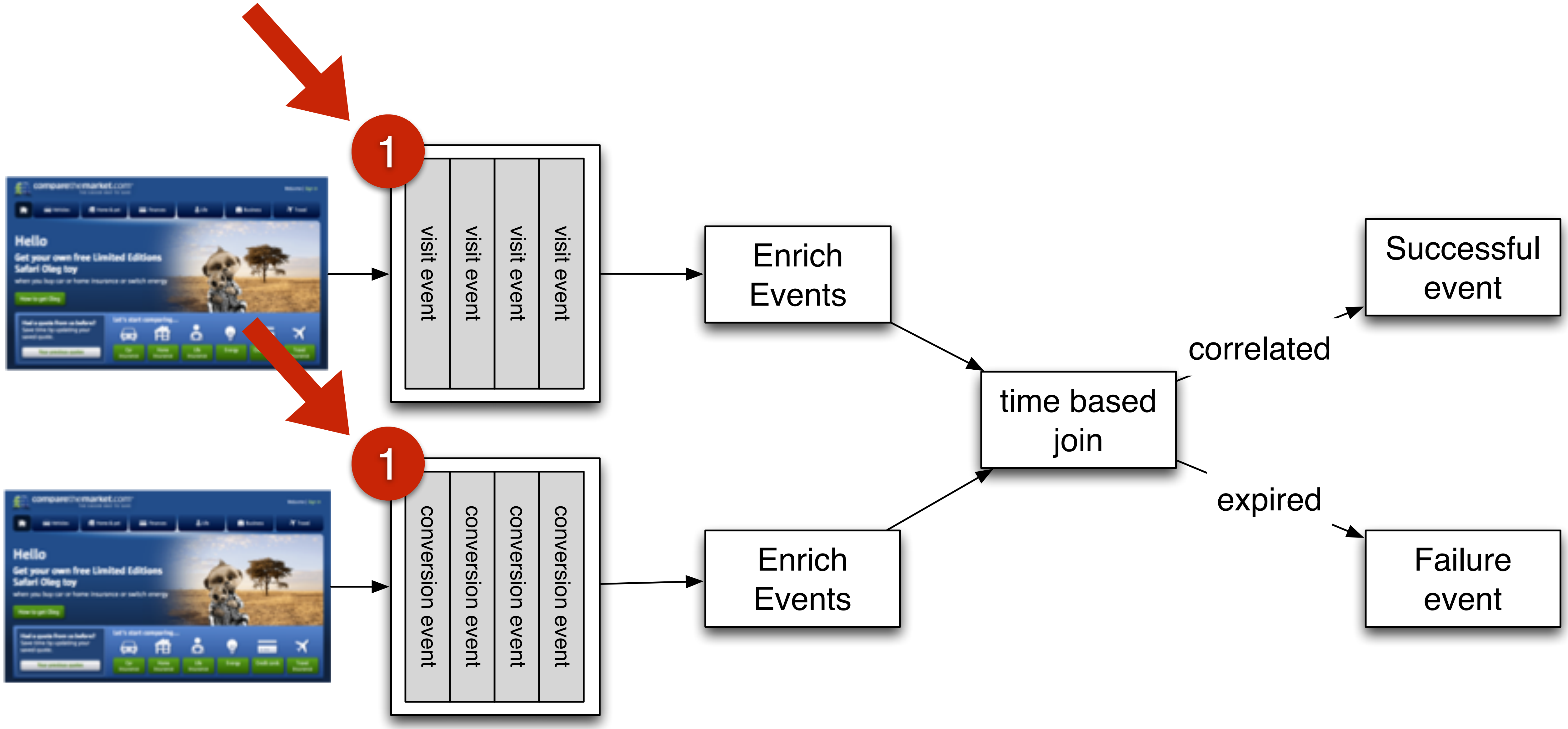
event event event event

Source

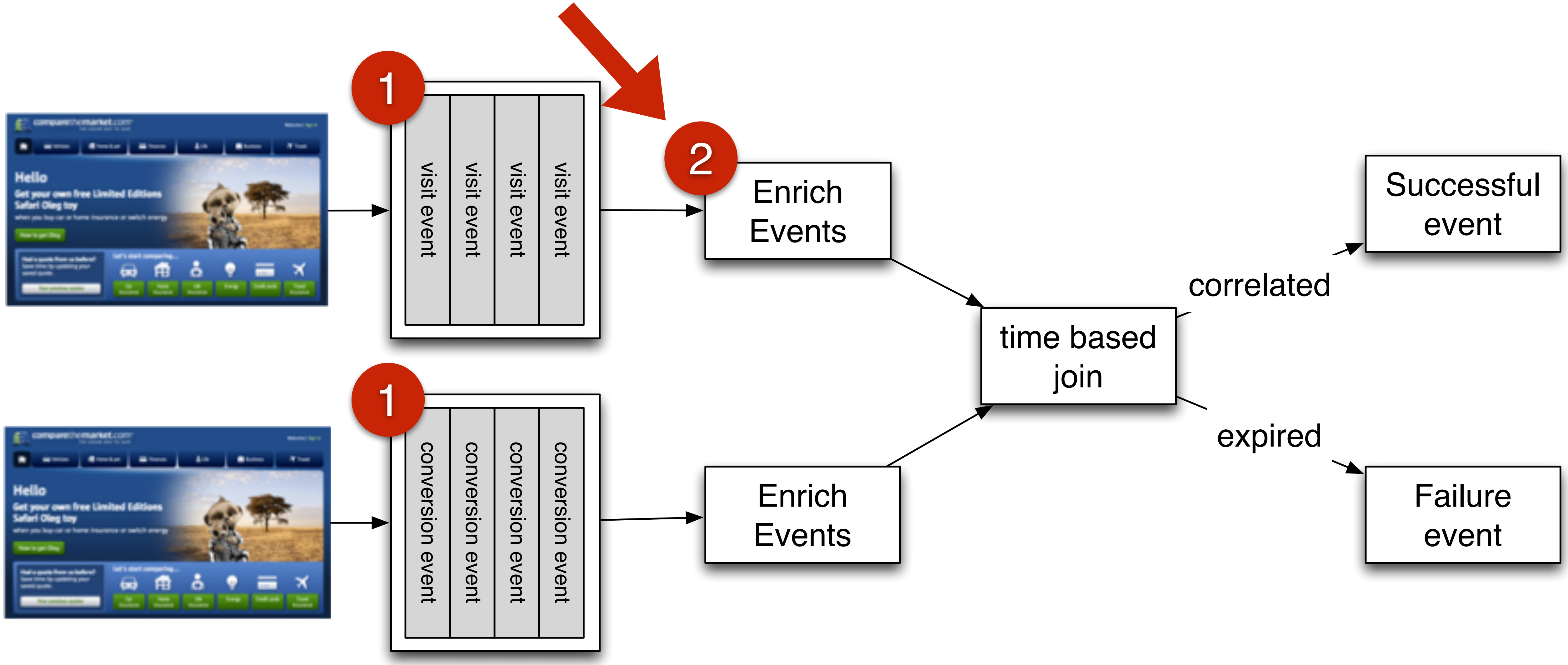
java archiver

hadoop

PoC 3


1
visit event
visit event
visit event
visit event
1
conversion event
conversion event
conversion event
conversion event
Enrich Events
Enrich Events
time based join
correlated
expired
Successful event
Failure event

PoC 3



1

visit event

visit event

visit event

visit event

2

Enrich Events

1

conversion event

conversion event

conversion event

conversion event

Enrich Events

time based join

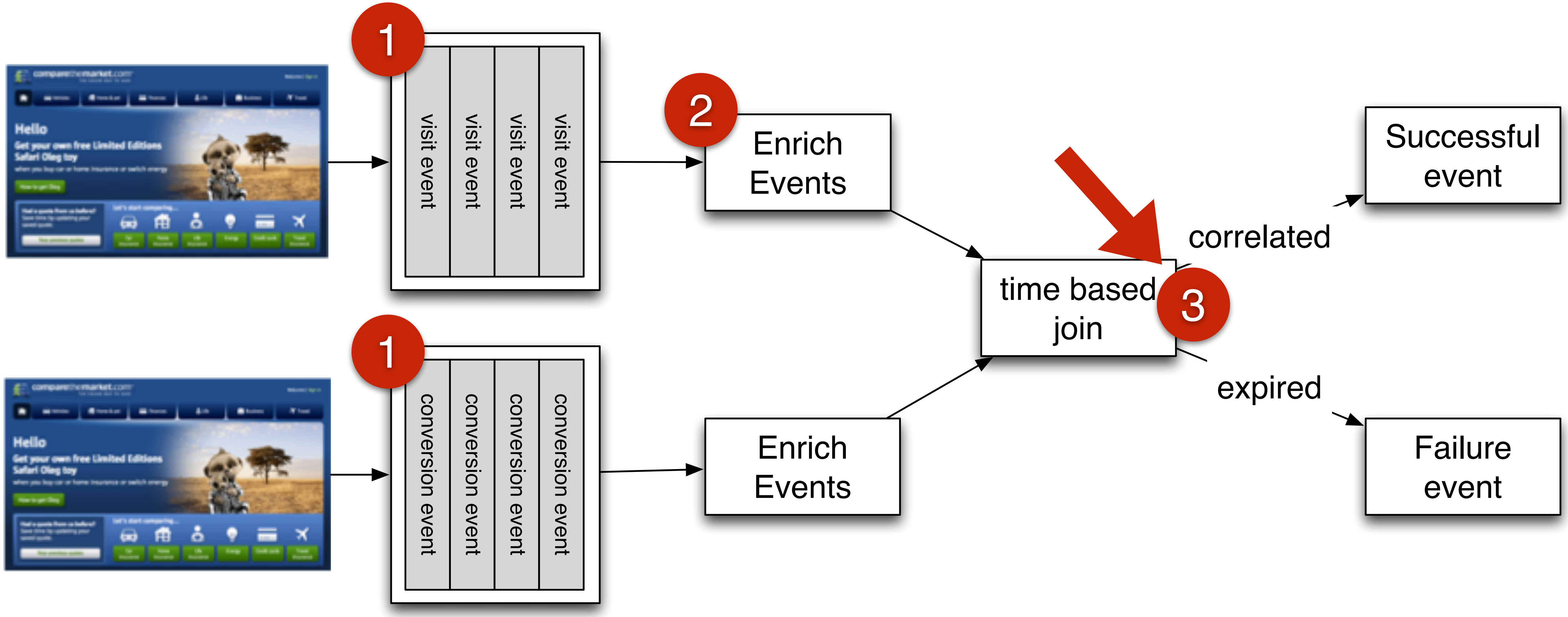
correlated

Successful event

expired

Failure event

PoC 3



1

visit event

visit event

visit event

visit event

1

conversion event

conversion event

conversion event

conversion event

2

Enrich Events

Enrich Events

time based join

3

correlated

expired

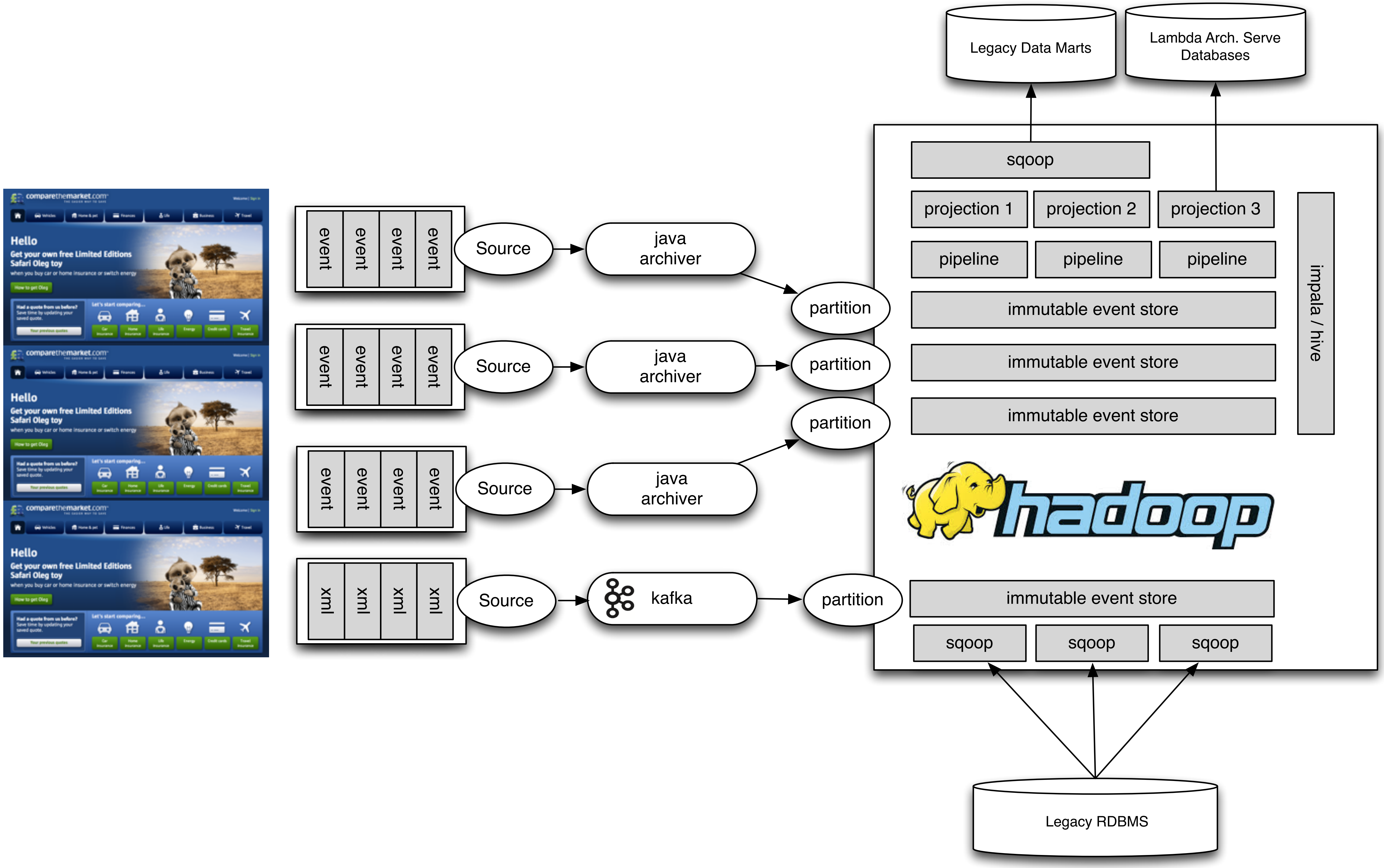
Successful event

Failure event

2013/14 Urgency Increases

- Tipping Point in Demand vs. Supply
- Procurement Process
- Contracted with Cloudera
- Statement of Work
- Onsite Engagement

Production 2



Legacy Data Marts

Lambda Arch. Serve Databases

event event event event

Source

java archiver

partition

event event event event

Source

java archiver

partition

event event event event

Source

java archiver

partition

xml xml xml xml

Source

kafka

partition

sqoop

projection 1

projection 2

projection 3

pipeline

pipeline

pipeline

immutable event store

immutable event store

immutable event store

impala / hive

hadoop

immutable event store

sqoop

sqoop

sqoop

Legacy RDBMS

Outcomes

2014 Outcomes

Hadoop is a key component of our data platform and growing

Customer Data is integrated across CRM and other systems

The Data Vision generates alignment and is making an impact

Hadoop delivers performance and agility at low cost compared to legacy infrastructure

Tactical Data Marts supported by Hadoop

Data Team have moved from being inward looking BAU/CI and engaging with stakeholders

We are gaining new insight from web logs and systems logs

Data Engineering Cycle time has reduced significantly

- Effort into understanding the issues
- Created alignment with a vision for change
- Challenged convention
- Engaged with the problem at range
- Communicated that vision to stakeholders
- Communicated the sense of urgency
- Built a guiding coalition of powerful stakeholders
- Demonstrated quick wins - the solution would work and generated more change

Future Challenges

SUPPLY SIDE CAPABILITY

DEMAND SIDE CAPABILITY

DATA INNOVATION

Future Challenges - Supply Side

- Alignment
- Engineering Capability
- Technology Enablers
- Platform Stability and Data Availability
- Coping with Web Tier Agility

Future Challenges - Demand Side Capability

- Data Team Skills and Structure
- Benefits from Platform Investment
- Reproducible Research
- Leveraging Web Tier Agility
- Tell a story with data and visualisation

Future Challenges - Data Innovation

- Continuing Evolution of Data Strategy
- Open Data Experimentation
- Value from Partnerships and Social
- Trusting Automated Decisions
- Designing Cross Tier Data Products

Thank you