Instant Cloning: Boosting Application Development

Kyle Hailey
kyle@delphix.com
http://kylehailey.com

DELPHIX

Problem in IT

CIO Magazine Survey:

60% Projects Over Schedule

85% delayed waiting for data

Data is the problem

Get the right data
To the right people
At the right time

The Phoenix Project

- IT bottlenecks
- Setting Priorities
- Company Goals
- Defining Metrics
- Fast Iterations

IT version of

"The Goal"

by E. Goldratt

"Any improvement not made at the constraint is an illusion."

What is the constraint ? The DBAs and environments.

"One of the most powerful things that IT can do is get environments to development and QA when they need

Data is the constraint

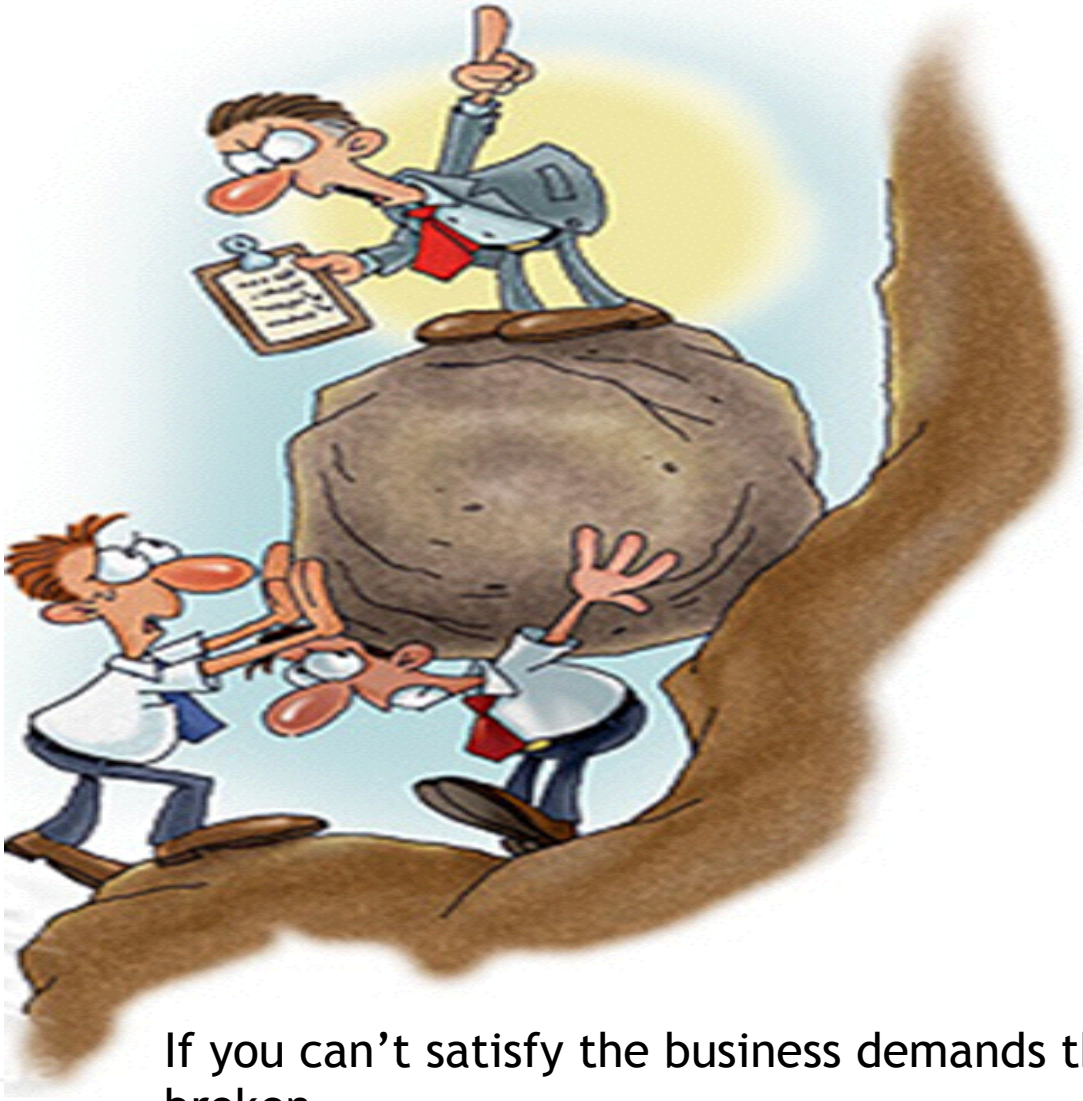
Three points

I. Data Tax strains infrastructure

II. Data Tax price is huge

III. Companies unaware of Data Tax

I. Data Taxes your business



If you can't satisfy the business demands then your process is broken.

II. Data Tax hits the bottom line hard

III. Companies don't think there is a Data Tax problem

I. Data Tax

– Moving data is hard

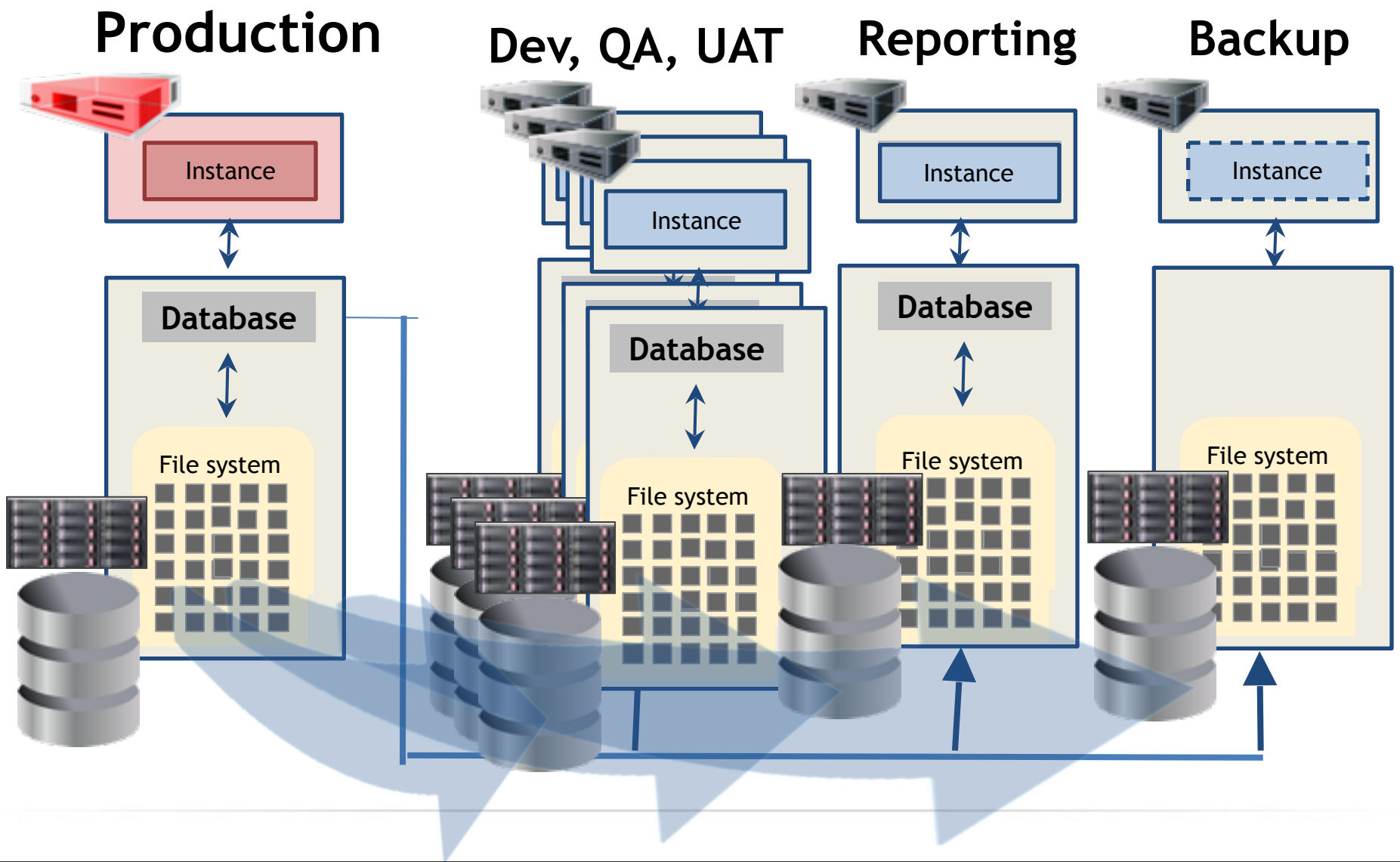
– Triple tax

– Data Floods infrastructure

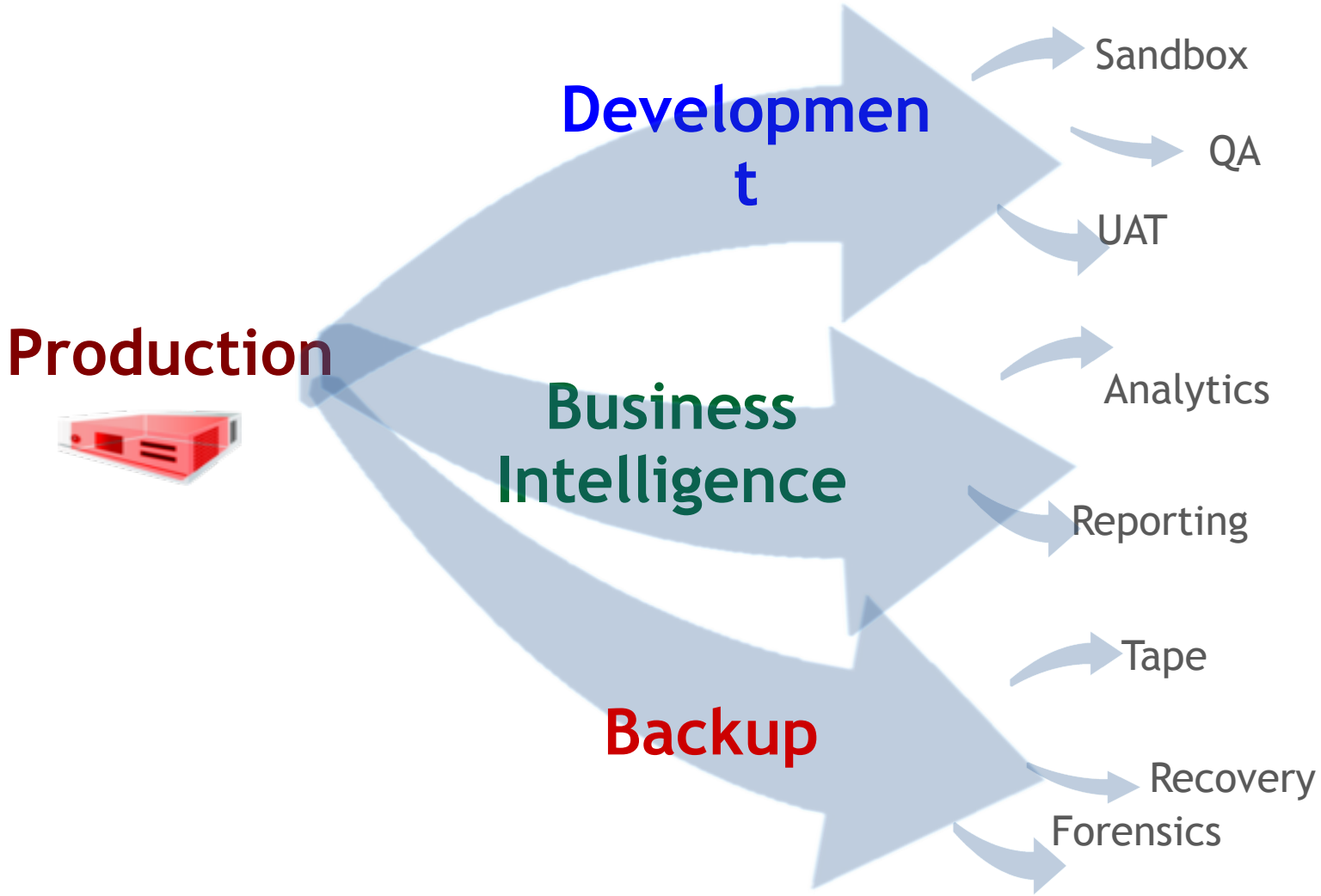
I. Data Tax : moving data is hard

- Storage & Systems : capital resources
- Personnel : operation expenditure
- Time : delayed projects

I. Data Tax : Typical Architecture

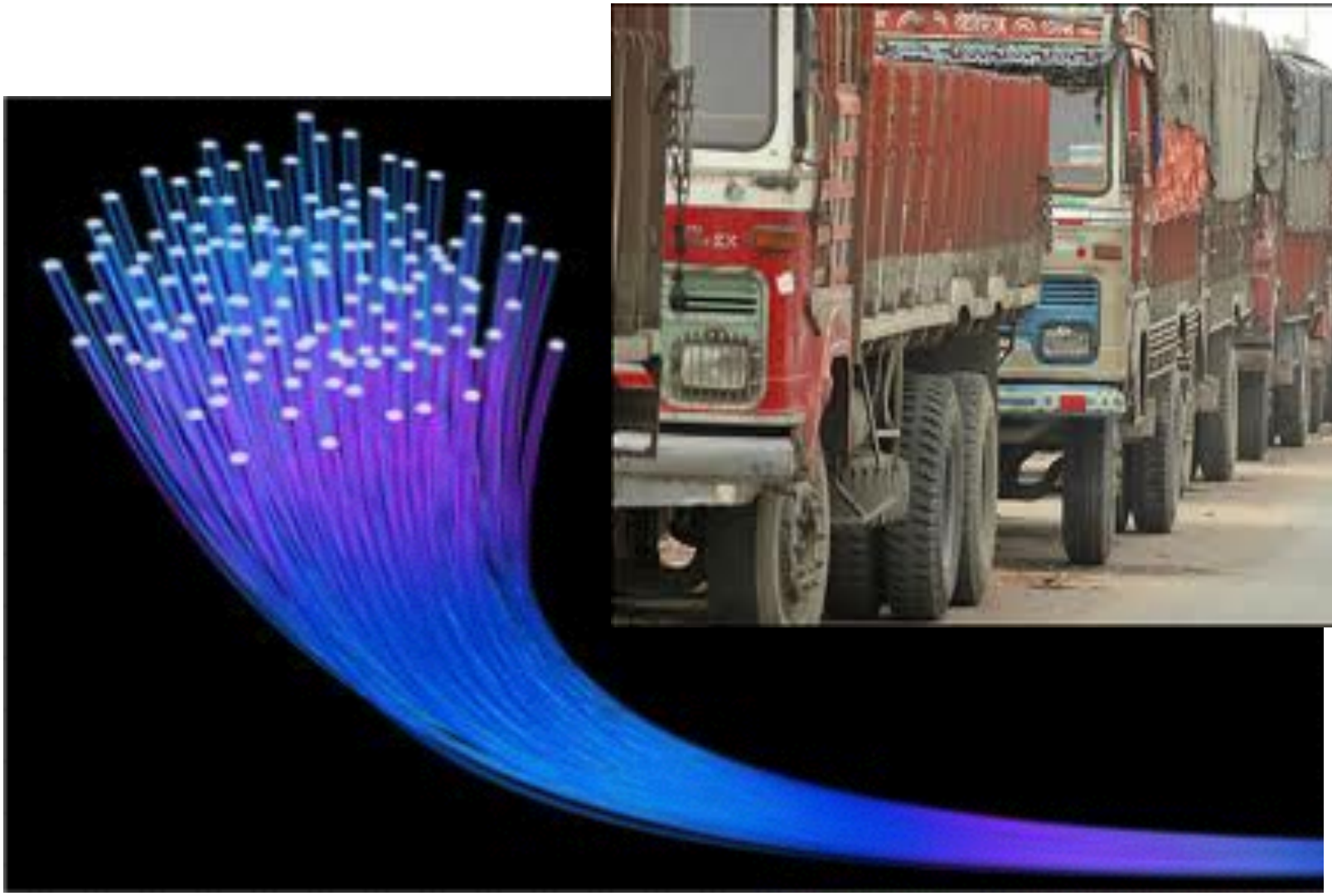


I. Data Tax: triple data tax



DELPHIX

I. Data Tax : Copying data floods infrastructure



I. Data Tax : Data Tax floods company infrastructure

92% of the cost of business
— the financial services business —
is “data”
www.wsta.org/resources/industry-articles

Most companies have
2-9% IT spending
http://uclue.com/?xq=1133

Data management is the largest portion of IT expense

Part II. Data Tax is Huge

Part II. Data Tax is Huge

- Four Areas data tax hits

  1. IT Capital resources
  2. IT Operations personnel
  3. Application Development ***
  4. Business

- How big is the data tax?
  - Measure before and after installing Delphix

II. Data Tax is huge : 1. IT Capital

- Hardware
  - Servers
  - Storage
  - Network
  - Data center floor space, power, cooling

- Example
  - Some customers have over 1 Petabyte duplicate data
  - (1000 TB, ie 1,000,000 GB)

II. Data Tax is huge : 2. IT Operations

- Involves many people
  - DBAs
  - SYS Admin
  - Storage Admin
  - Backup Admin
  - Network Admin
- 1000s of hours annually just for DBAs
  - not including all the other personnel required to supply the infrastructure necessary
- Data center efforts costly and difficult
  - Consolidation
  - Migrations
  - Move to cloud

II. Data Tax is Huge : 3. App Dev

quality and speed

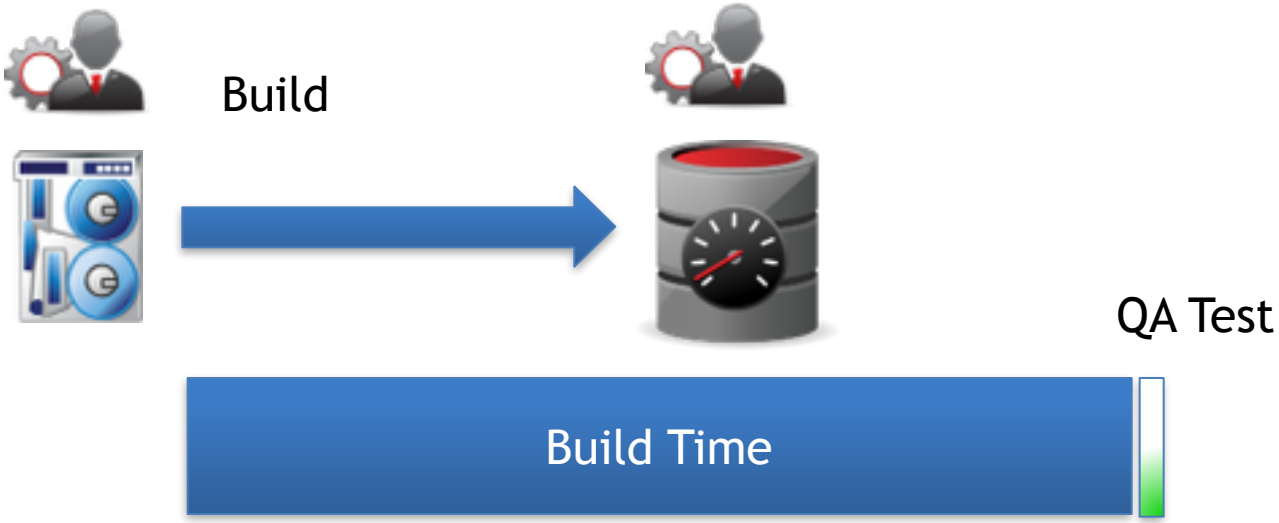
Five examples of application Data Tax impact

- Inefficient QA: Higher costs of QA
- QA Delays : Greater re-work of code
- Sharing DB Environments : Bottlenecks
- Using DB Subsets: More bugs in Prod
- Slow Environment Builds: Delays

“if you can't measure it you can’t manage it”

II. Data Tax is Huge : 3. App Dev

Inefficient QA : Long Build times

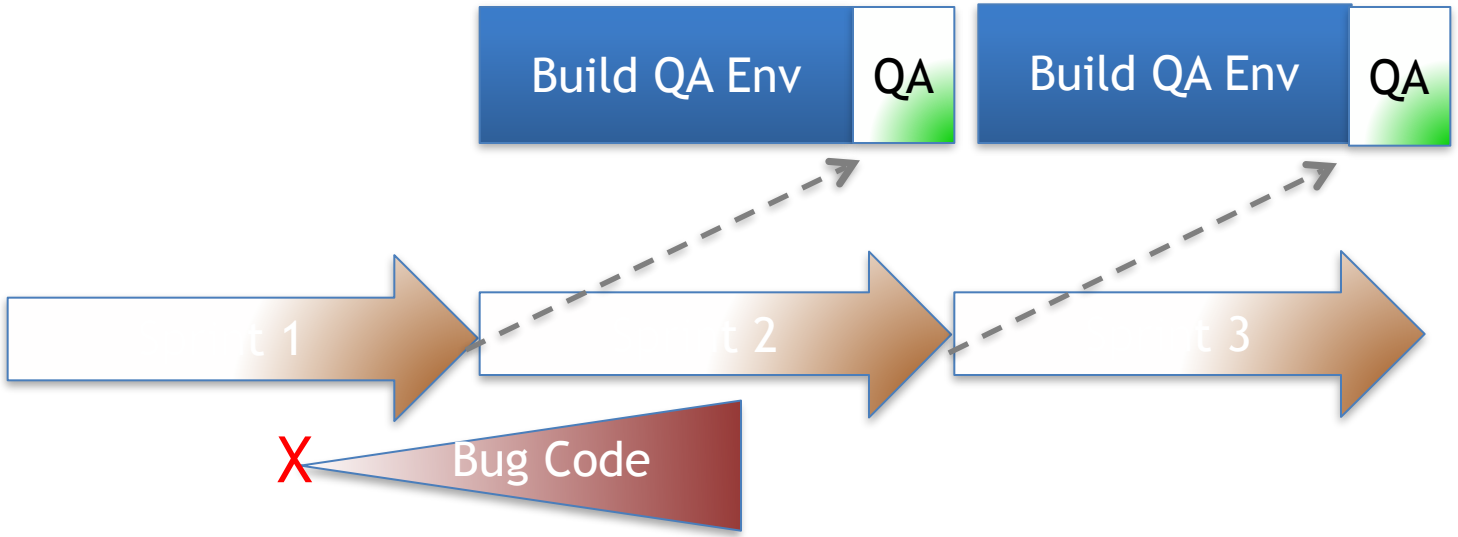


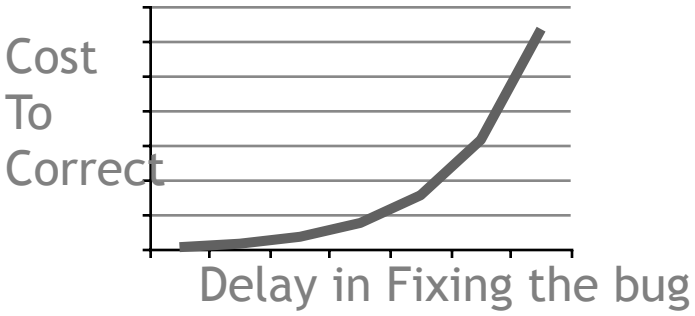
96% of QA time was building environment

$.04/$1.00 actual testing vs. setup

II. Data Tax is Huge : 3. App Dev

QA Delays: bugs found late = more code re-work

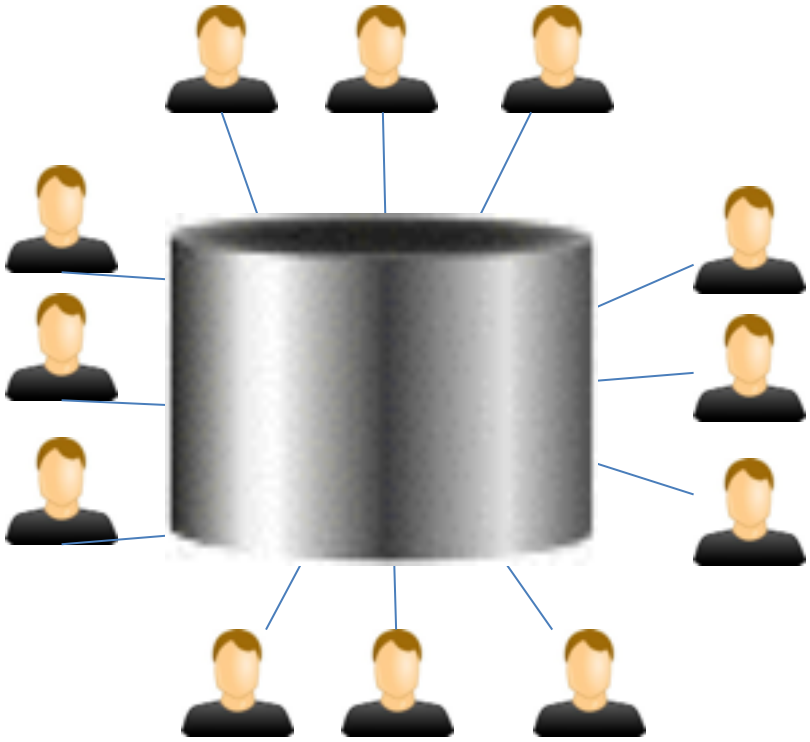




Software Engineering Economics – Barry Boehm (1981)

II. Data Tax is Huge : 3. App Dev

full copies cause bottlenecks

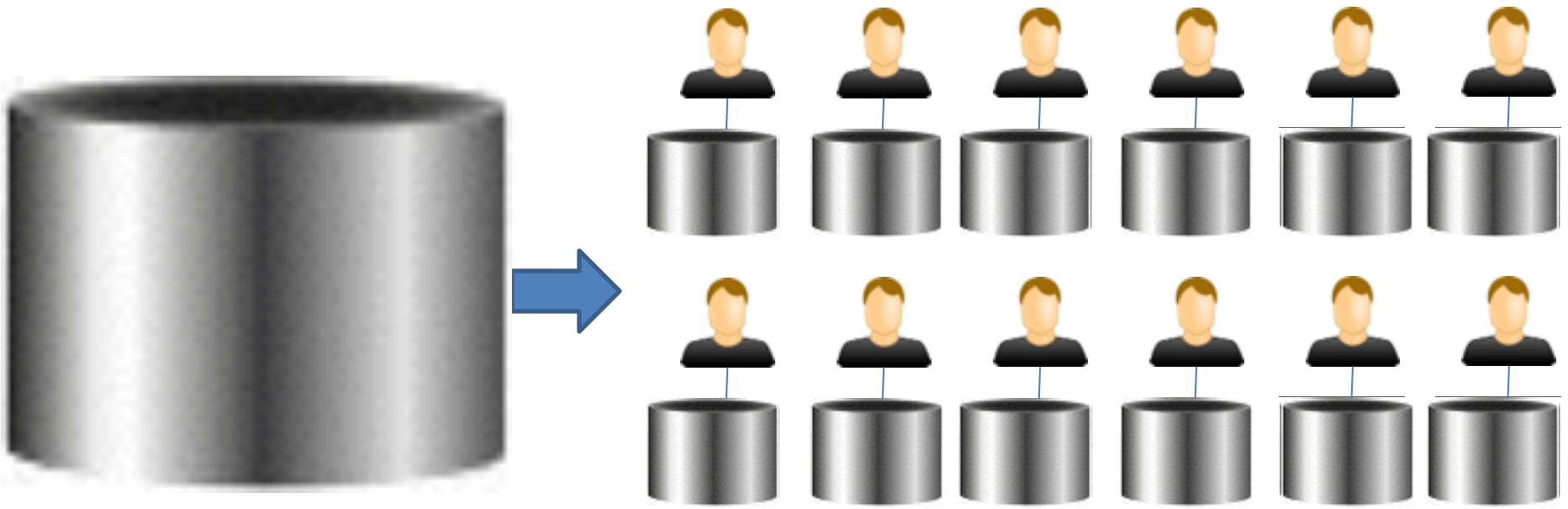


Old Unrepresentative Data

Frustration Waiting

II. Data Tax is Huge : 3. App Dev

subsets cause bugs



II. Data Tax is Huge : 3. App Dev

subsets cause bugs

The Production ‘Wall’

Classic problem is that queries that run fast on subsets hit the wall in production.

Developers are unable to test against all data

II. Data Tax is Huge : 3. App Dev

Slow Environment Builds:

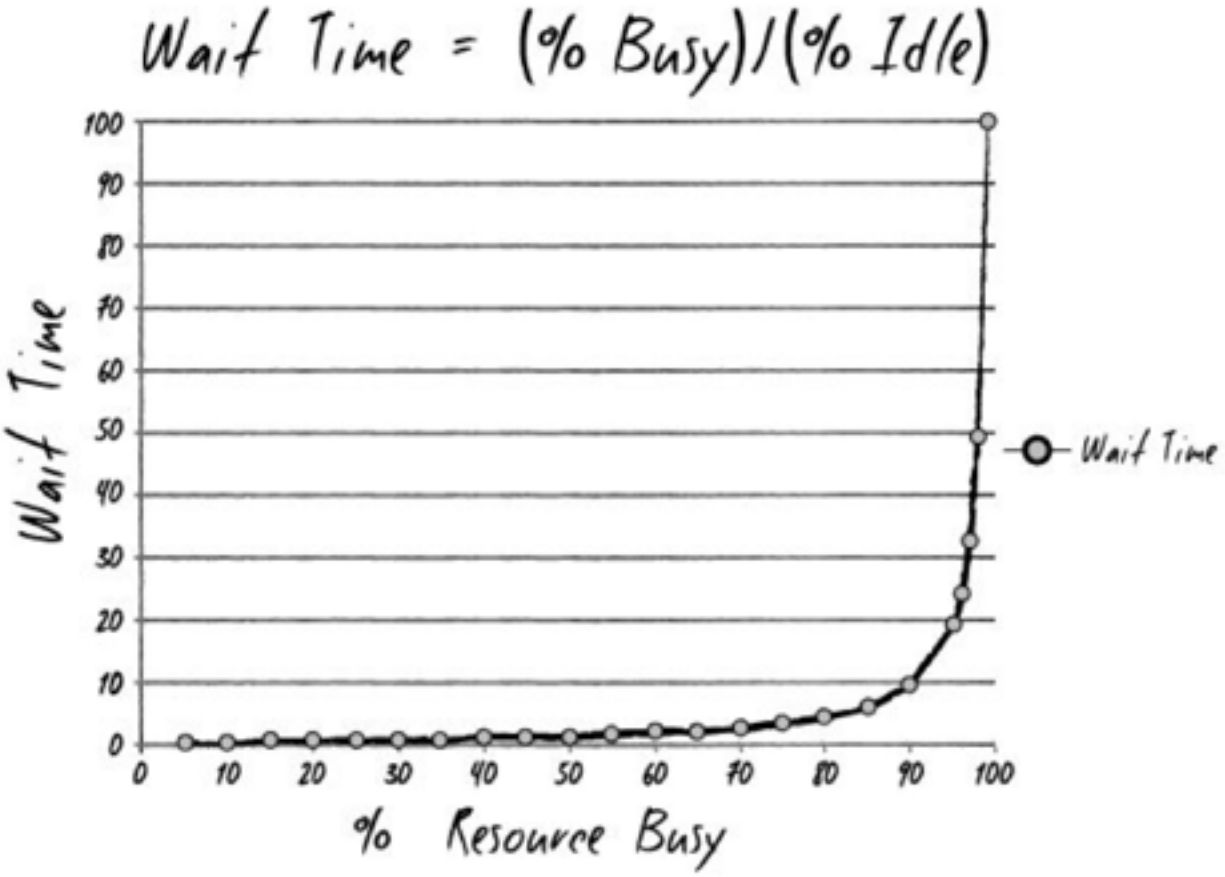
3-6 Months to Deliver Data

Developer	Asks for DB					Get Access
Manager	approves					
DBA		Request system				Setup DB
System Admin			Request storage		Setup machine	
Storage Admin				Allocate storage (take snapshot)		

II. Data Tax is Huge : 3. App Dev

Slow Environment Builds

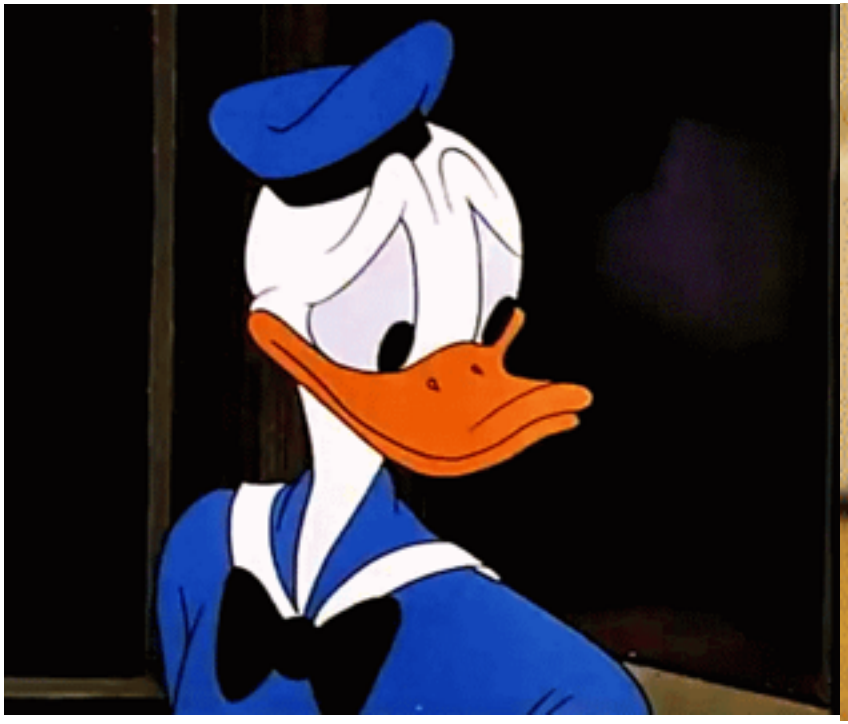
Why are hand offs so expensive?



II. Data Tax is Huge : 3. App Dev

Slow Environment Builds: culture of no

DBA



Developer

II. Data Tax is Huge : 3. App Dev

Slow Environment Builds

Never enough environments

II. Data Tax is Huge : 3. App Dev

What We've Seen

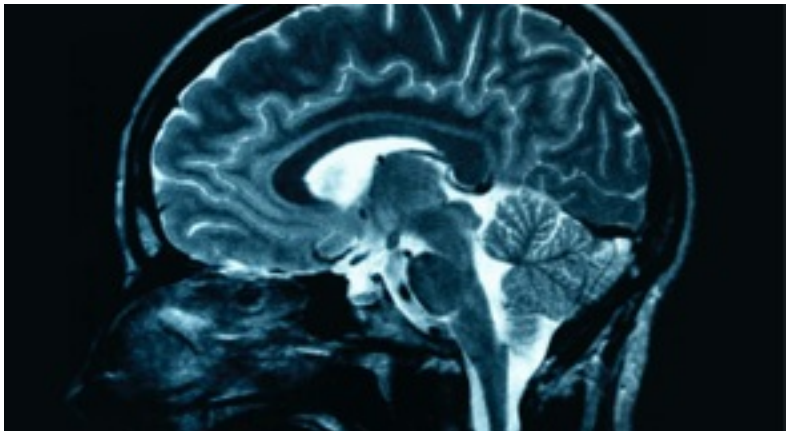
Five examples of application Data Tax impact

- Inefficient QA: Higher costs
- QA Delays : Increased re-work
- Sharing DB : Bottlenecks
- Subset DB : Bugs
- Slow Environment Builds: Delays

II. Data Tax is Huge : 4. Business

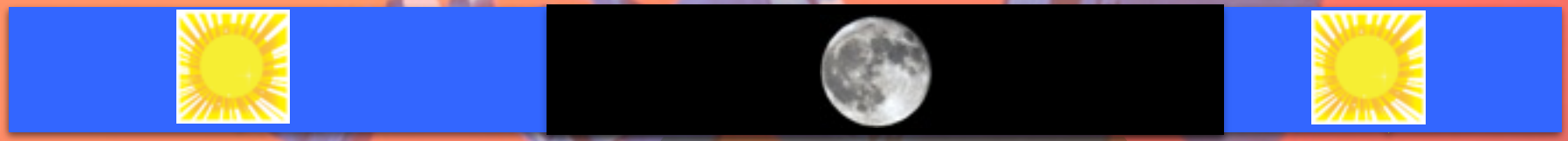
Ability to capture revenue

- Business Intelligence
  - ETLs or data warehouse refreshes
    - Old data = less intelligence
    - Less Intelligence = Missed Revenue



- Business Applications
  - Delays on getting applications that generate the revenue

II. Data Tax is Huge : 4. Business Intelligence



1pm
8am

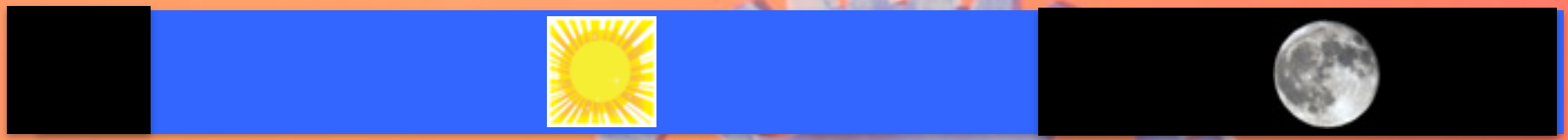
noon

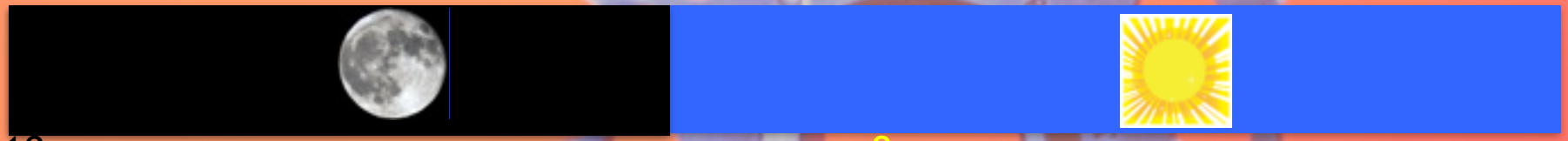
10pm

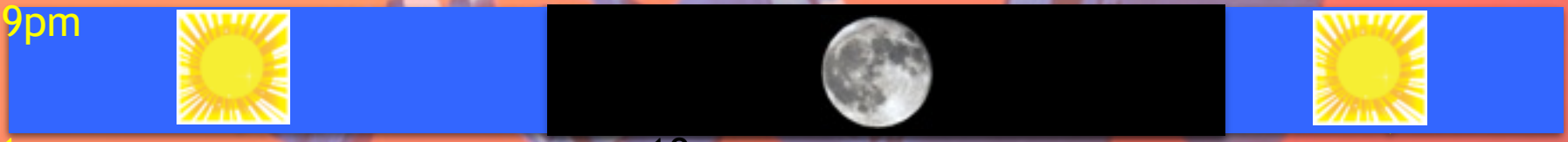
2011
2012
2013
2014
2015

ETL and Refresh Windows

II. Data Tax is Huge : 4. Business Intelligence







1pm
8am
noon

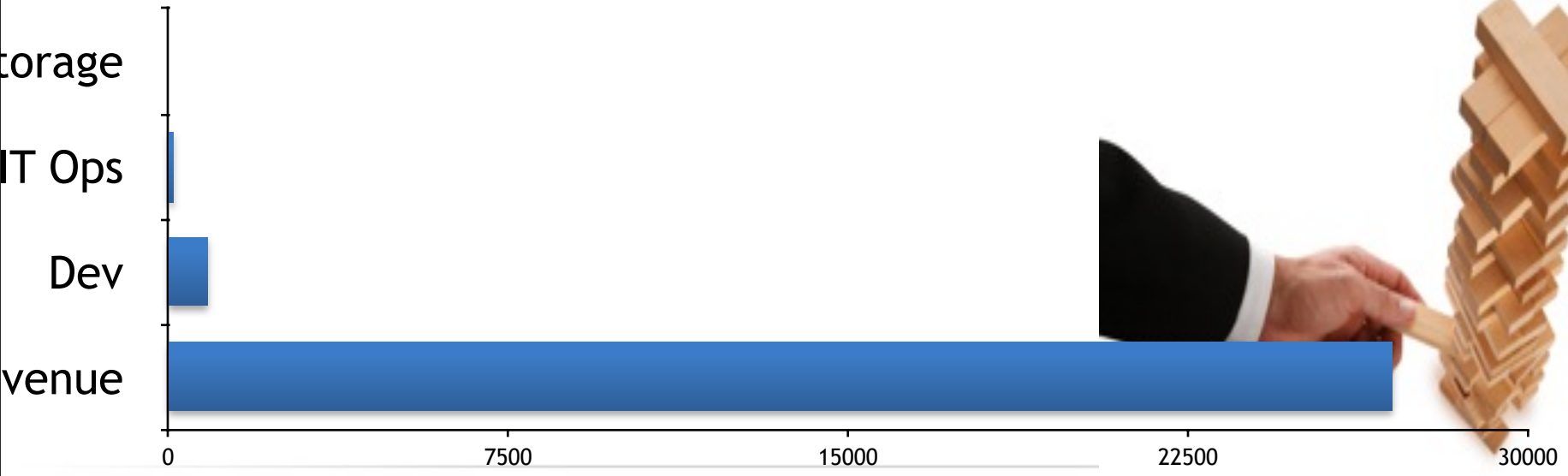
2011
2012
2013
2014
2015

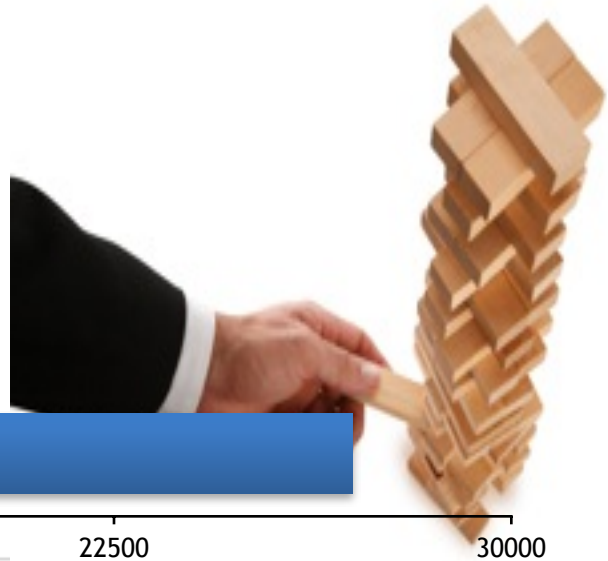
II. Data Tax is Huge : 4. Business

II. Data Tax is Huge : 4. Business

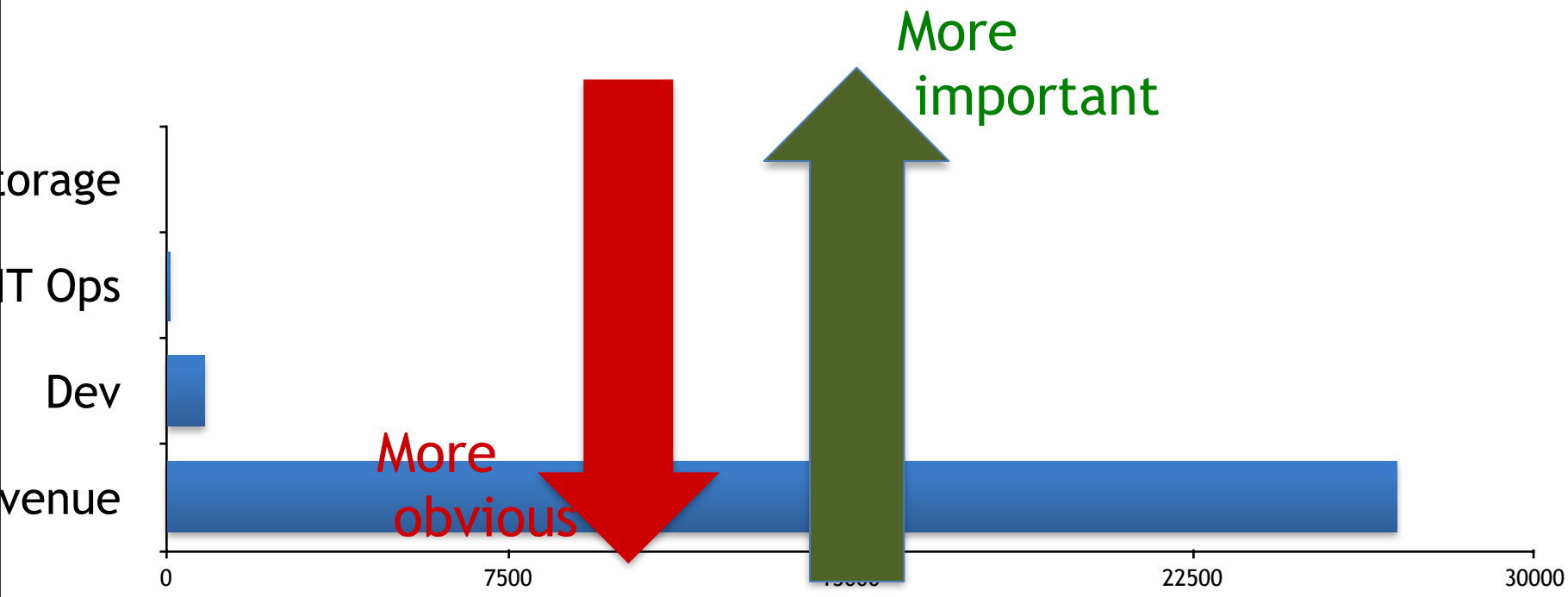
Magnitude of business impact

- $27 Billion Revenue
- $1 Billion IT spend
  - $850M development staff
  - $110M IT Ops
  - $ 40M storage





II. Data Tax is Huge : 4. Business



Part III. Companies unaware of the Data Tax

III. Companies unaware of the Data Tax

Isn't there technology that does that?
Why do I need Delphix?

Why do I need an iPhone ?

III. Companies unaware of the Data Tax

Nobody does what we do

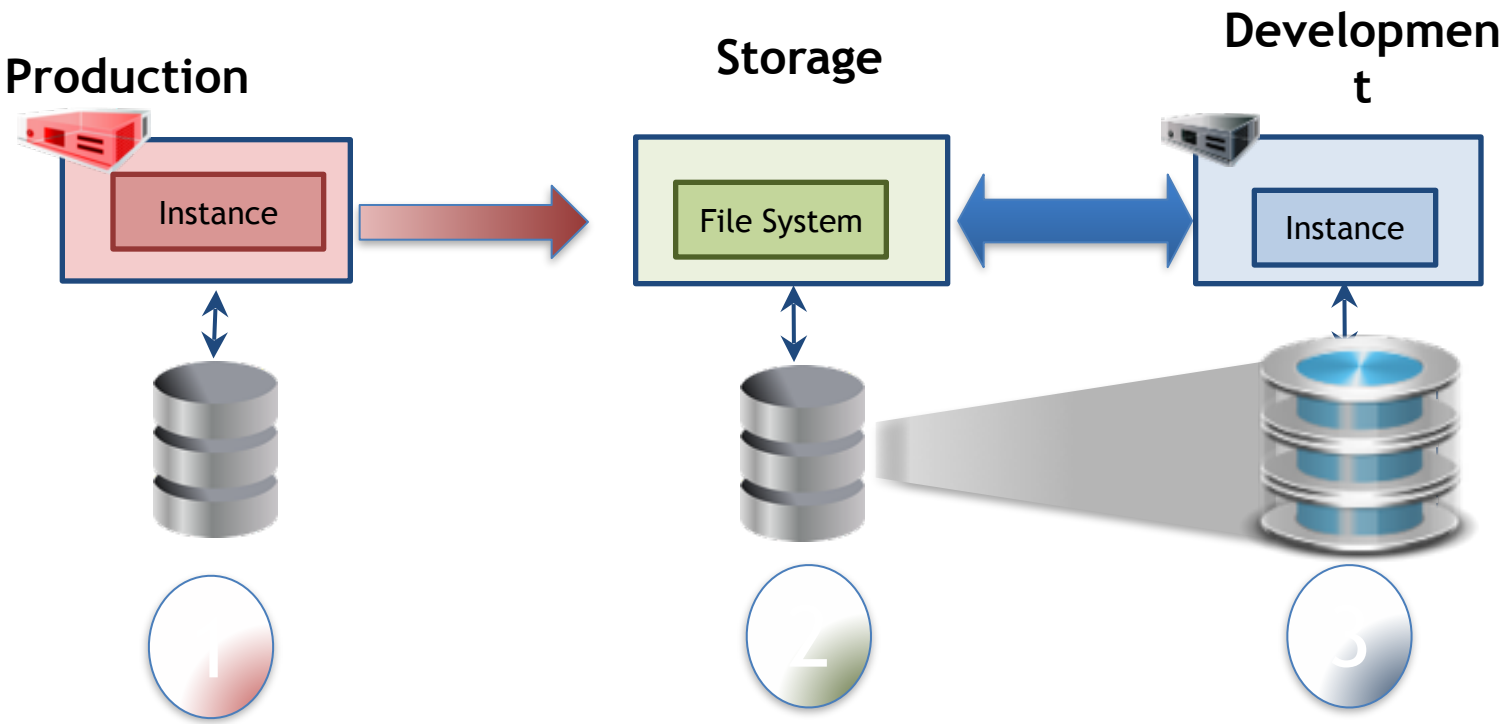
- NetBackup
- Flashback
- DataGuard
- GoldenGate
- SnapClone
- FlexClone

≠

DELPHIX®

≠

Three Core Parts



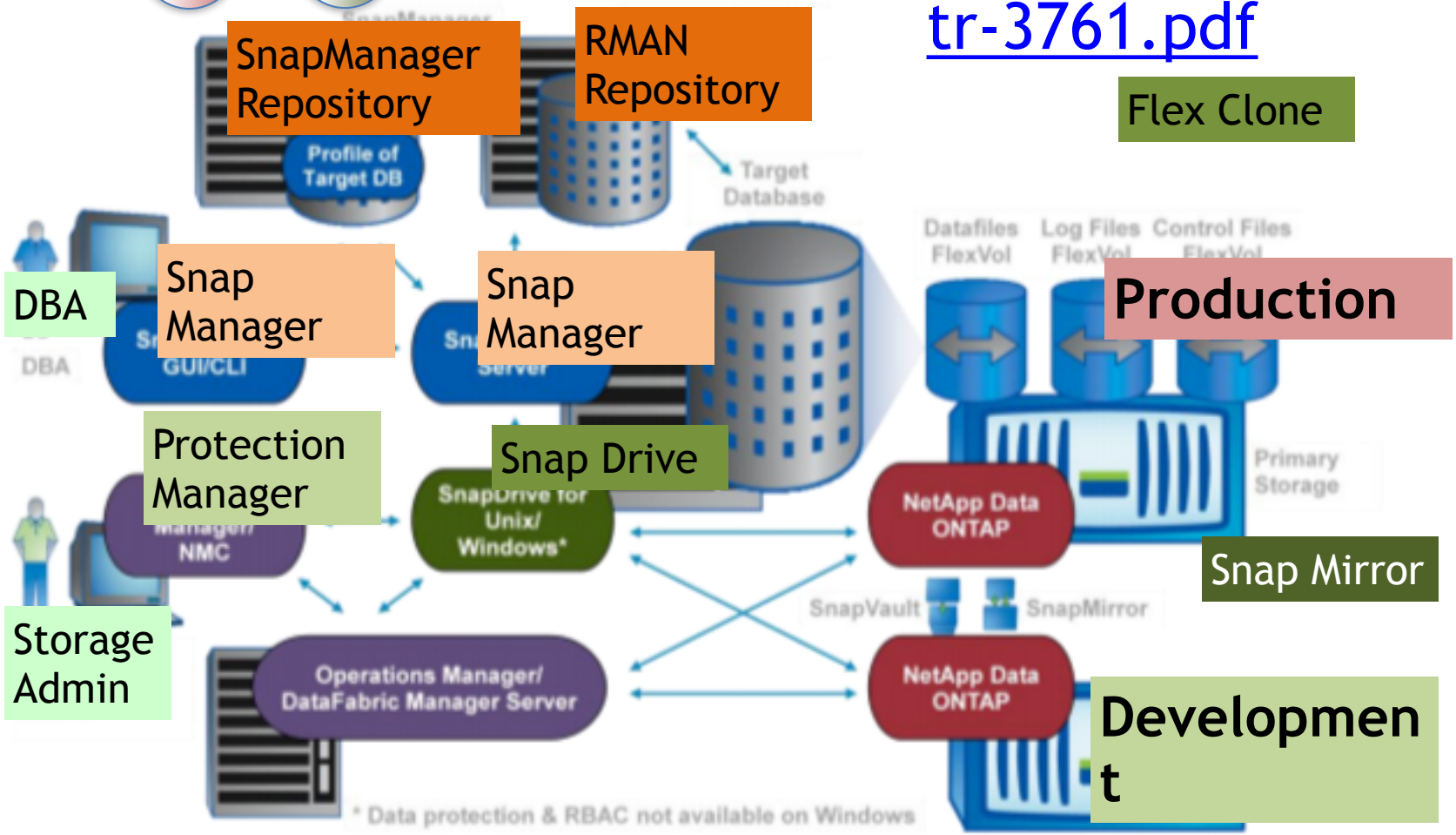
Copy
Sync Snapshots
Time Flow
Purge

Clone (snapshot)
Compress
Share Cache
Storage Agnostic

Mount, recover, rename
Self Service, Roles & Security
Rollback & Refresh
Branch & Tag

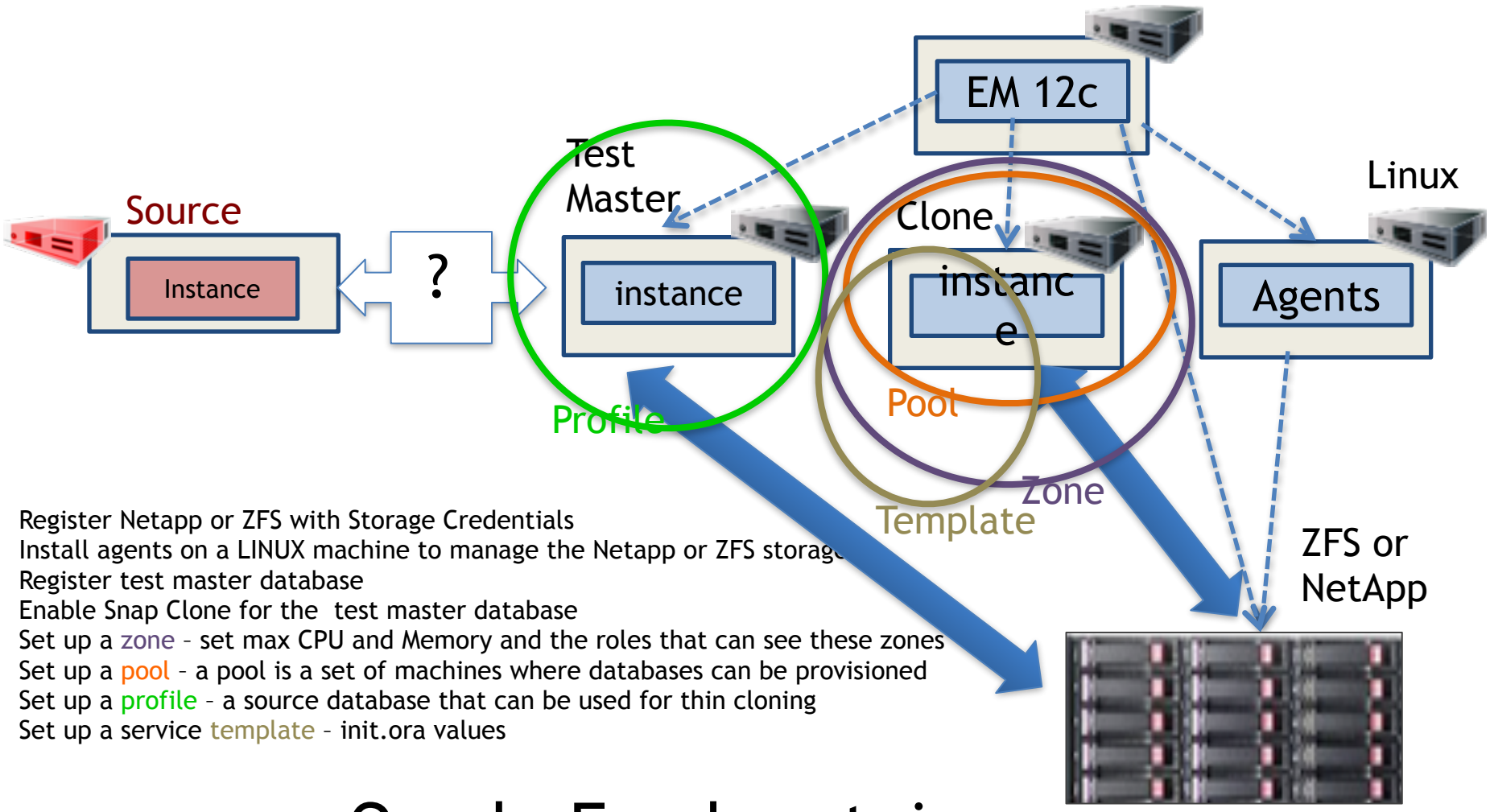
Netapp

tr-3761.pdf



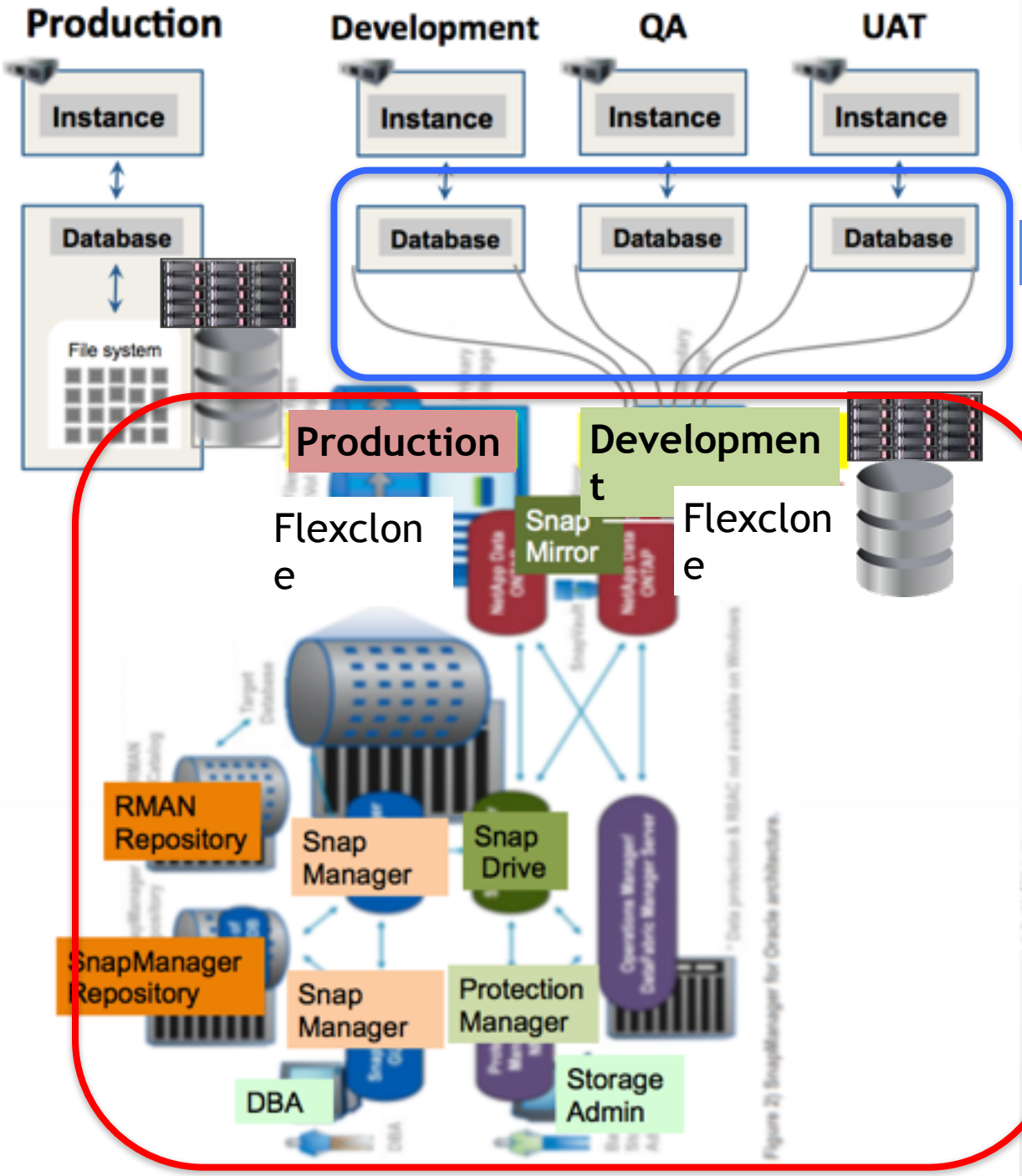
Netapp Frankenstein

Oracle EM 12c Snap Clone



- Register Netapp or ZFS with Storage Credentials
- Install agents on a LINUX machine to manage the Netapp or ZFS storage
- Register test master database
- Enable Snap Clone for the test master database
- Set up a zone - set max CPU and Memory and the roles that can see these zones
- Set up a pool - a pool is a set of machines where databases can be provisioned
- Set up a profile - a source database that can be used for thin cloning
- Set up a service template - init.ora values

Oracle Frankenstein



EM 12c: Snap Clone

Netapp
Snap Manager for Oracle

Other technology?

- Prove it
- Bake off
- Customer refs

III. Companies unaware of the Data Tax

#1 Biggest Enemy IT departments believe

- have best processes that exist
- have latest and greatest technology
- nothing better exists
- Just the way it is
- Data management is a drop in the bucket of overall issues

"The status quo is pre-ordained failure"

III. Companies unaware of the Data Tax

- Ask Questions
  - **Delphix**: we provision environments in minutes for almost not extra storage.
  - Customer: We already do that
  - Delphix: How long does it take a developer to get an environment after they ask ?
  - Customer: 2-3 weeks
  - Delphix: we do it in 2-3 minutes

No other product does what we do

III. Companies unaware of the Data Tax

- How to enlighten companies ? Ask for metrics
  - Batch window size for ETL
  - How new (old) is their BI data?
  - Number: app projects per year
  - How long does it take a developer to get a DB copy?
  - How long does it take QA to setup an environment
    - How long to rollback
    - How long to refresh
    - How many time do they run a QA cycle
  - How old is data in QA and DEV

Summary

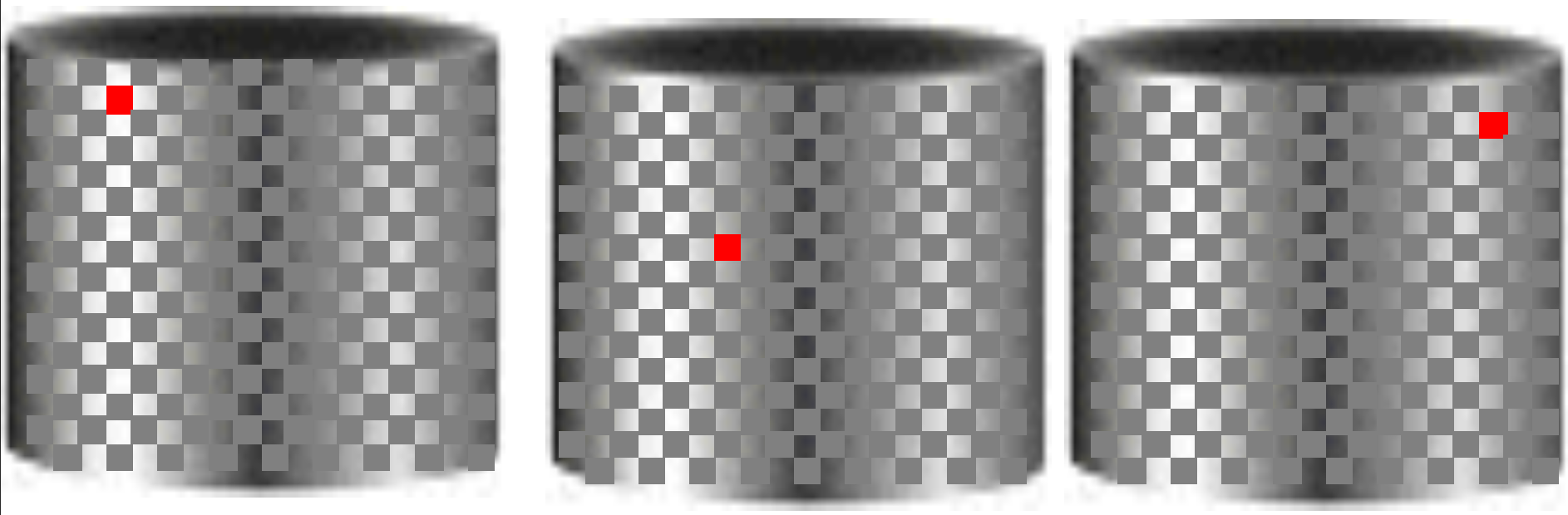
I. Companies pay a Data Tax
II. Data Tax is huge
III. Companies unaware of the Data Tax

99% of blocks are identical

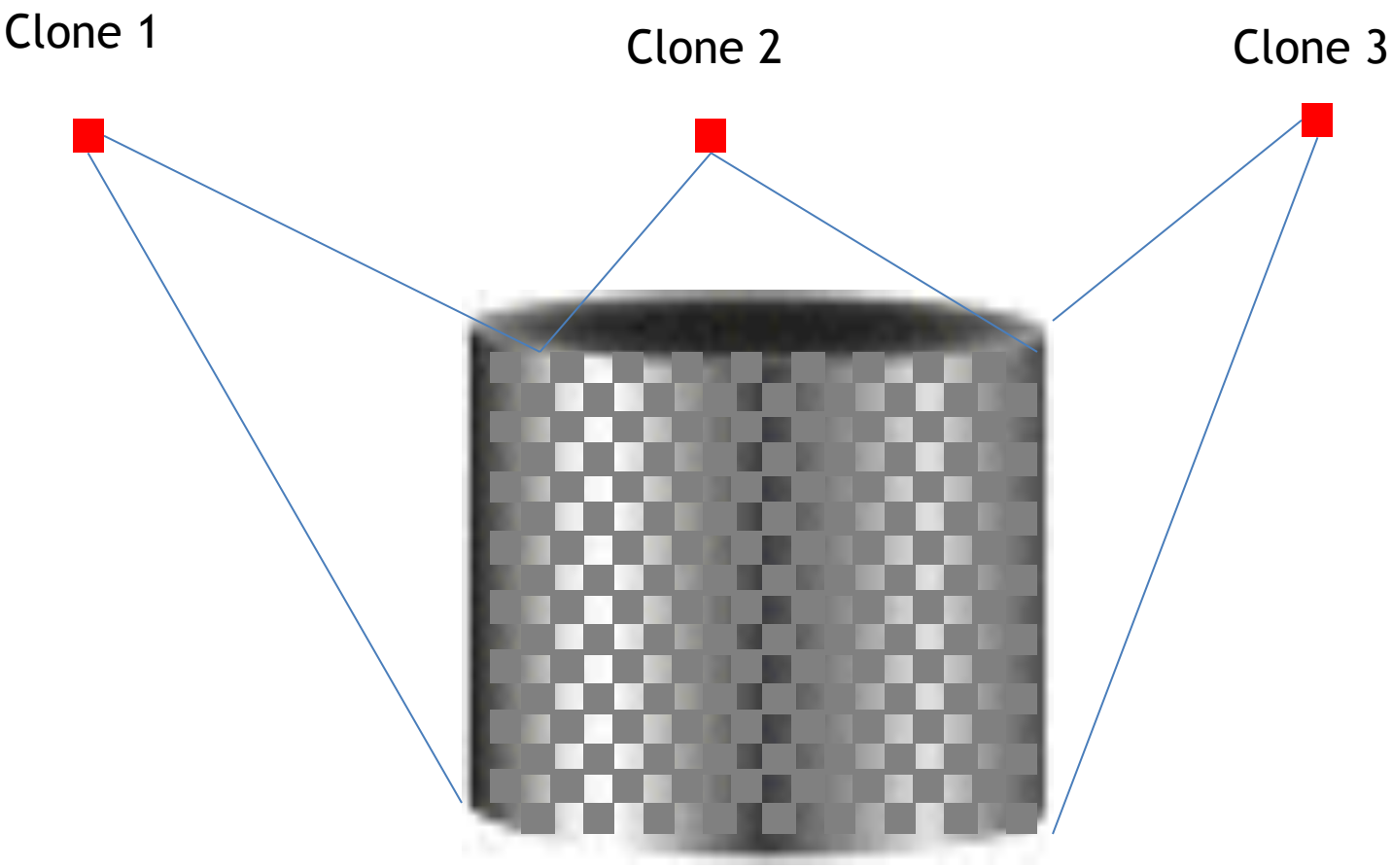
Clone 1

Clone 2

Clone 3

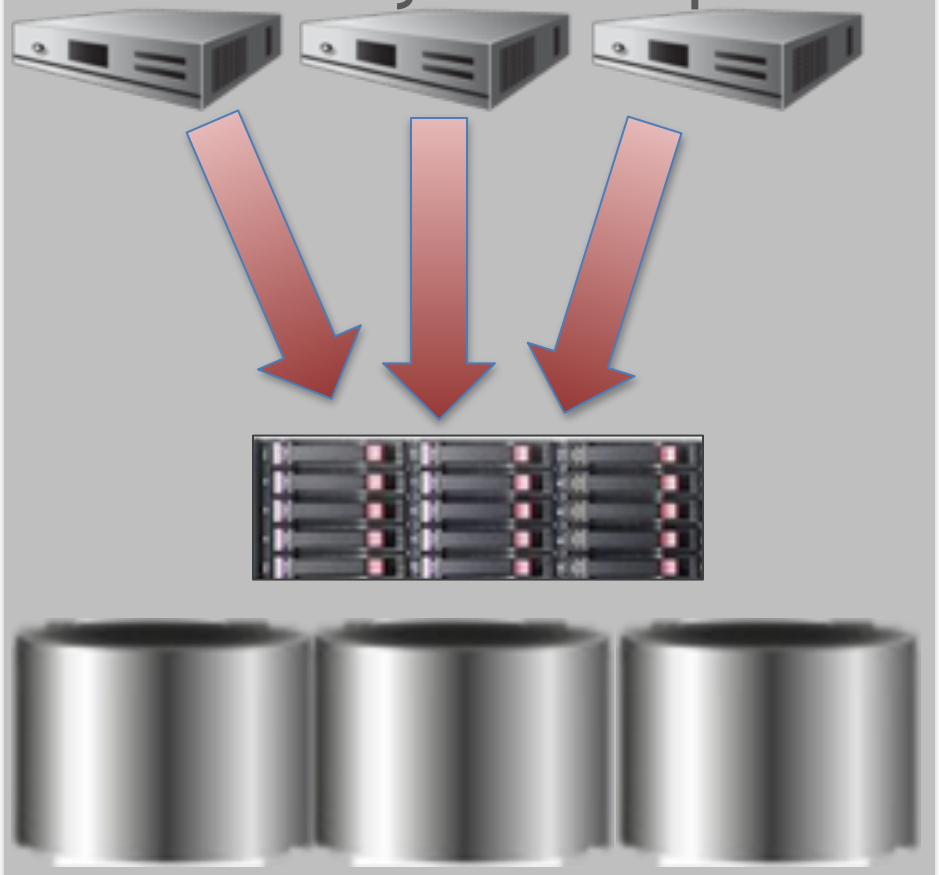


Thin Clone

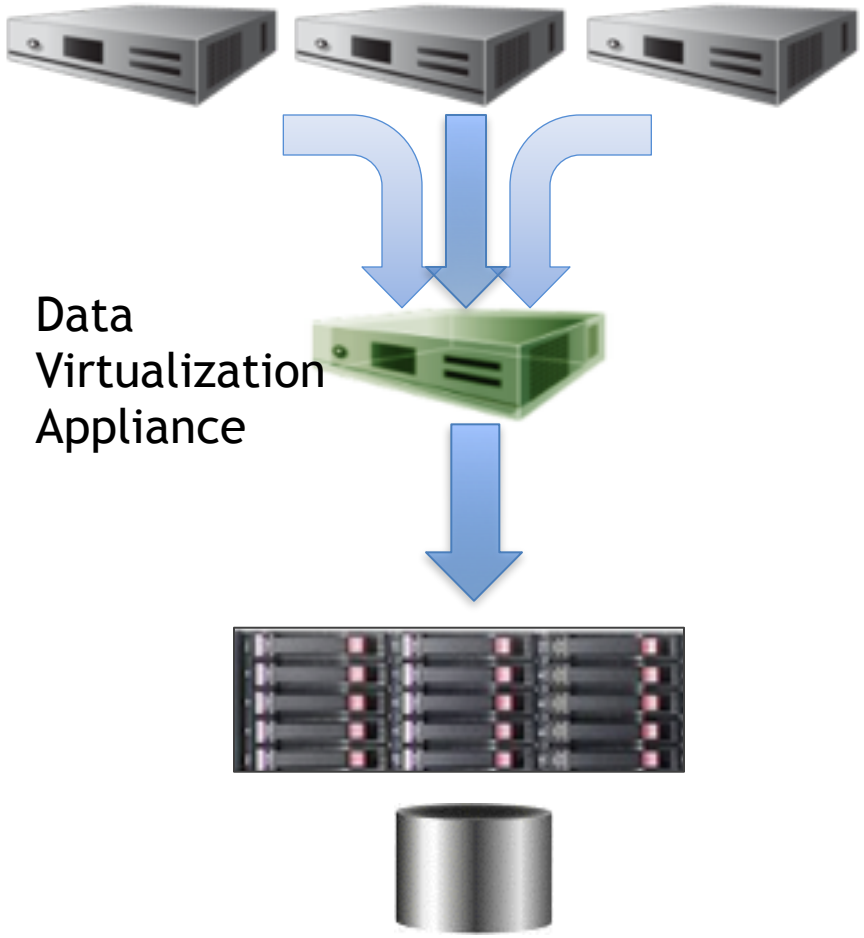


3. Database Virtualization

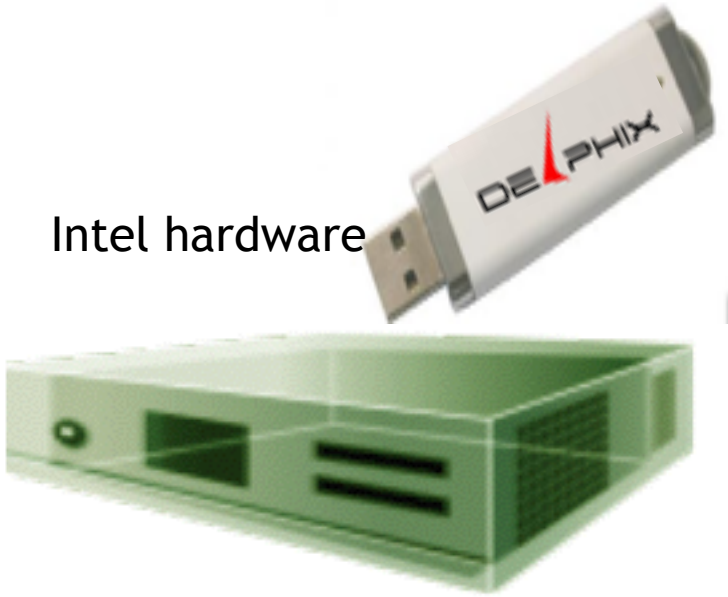
Three Physical Copies



Three Virtual Copies



Install Delphix on x86 hardware

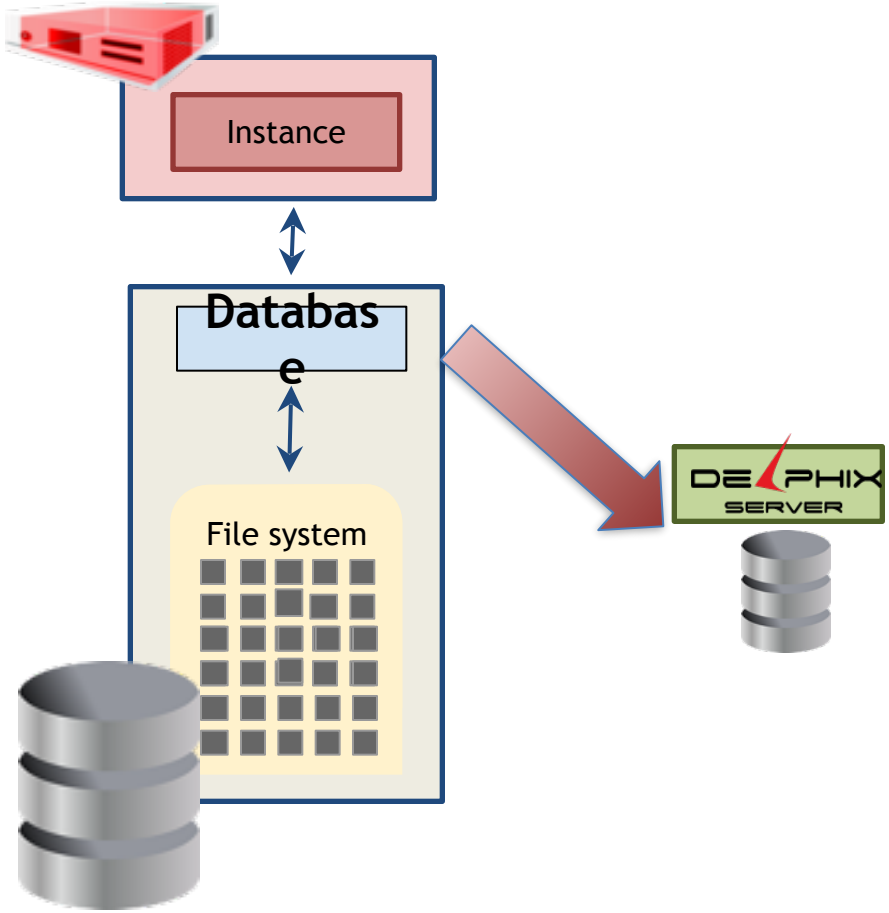


Allocate Any Storage to Delphix

Allocate Storage
Any type

Pure Storage + Delphix
Better Performance for
1/10 the cost

One time backup of source database



Supports

Microsoft SQL Server

ORACLE

PostgreSQL

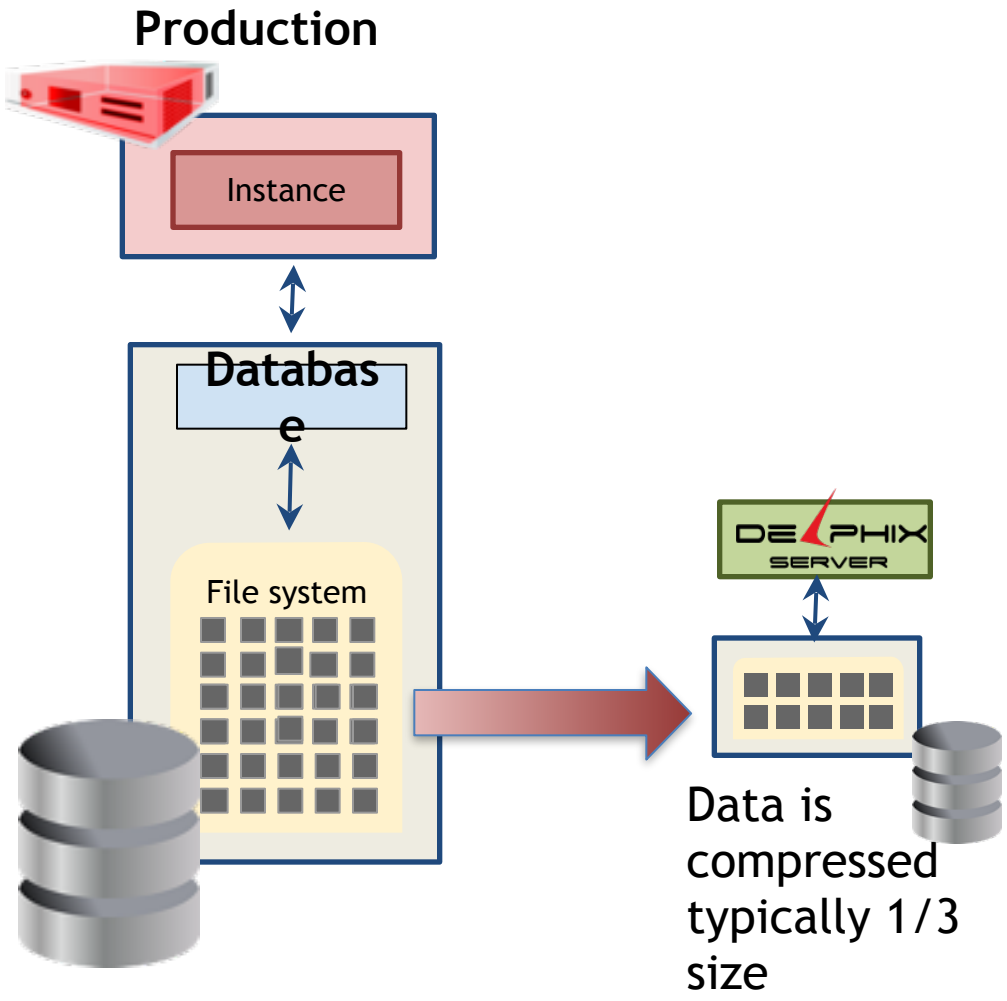
Application Stack Data

Upcoming

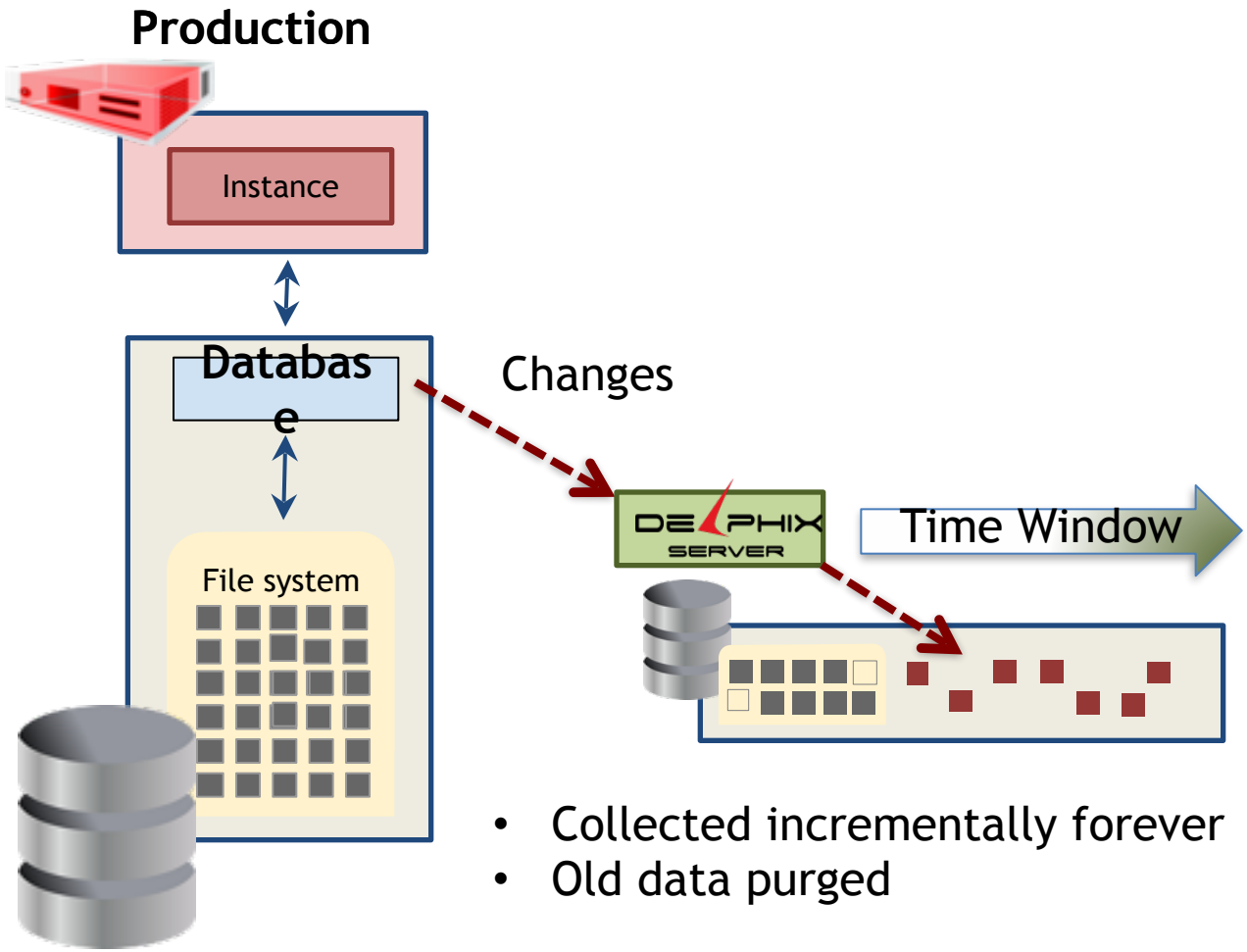
SYBASE

MySQL

DxFS (Delphix) Compress Data



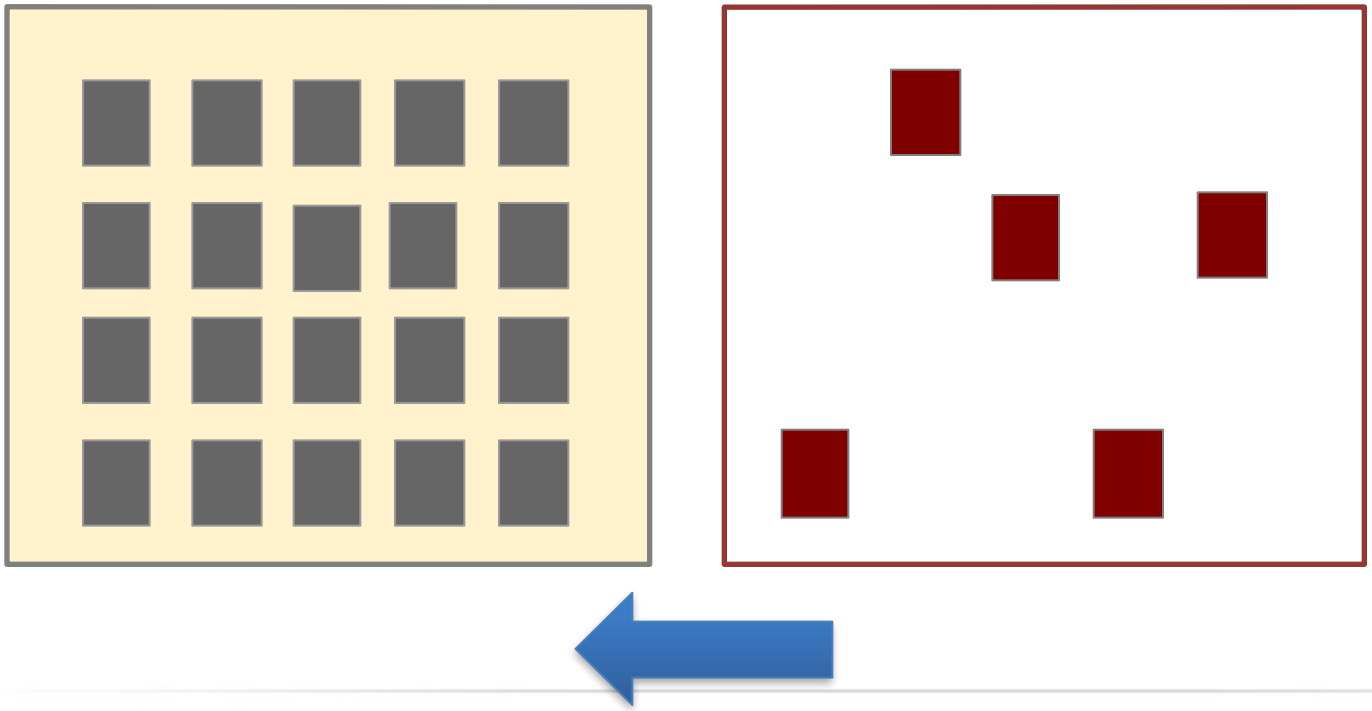
Incremental forever change collection

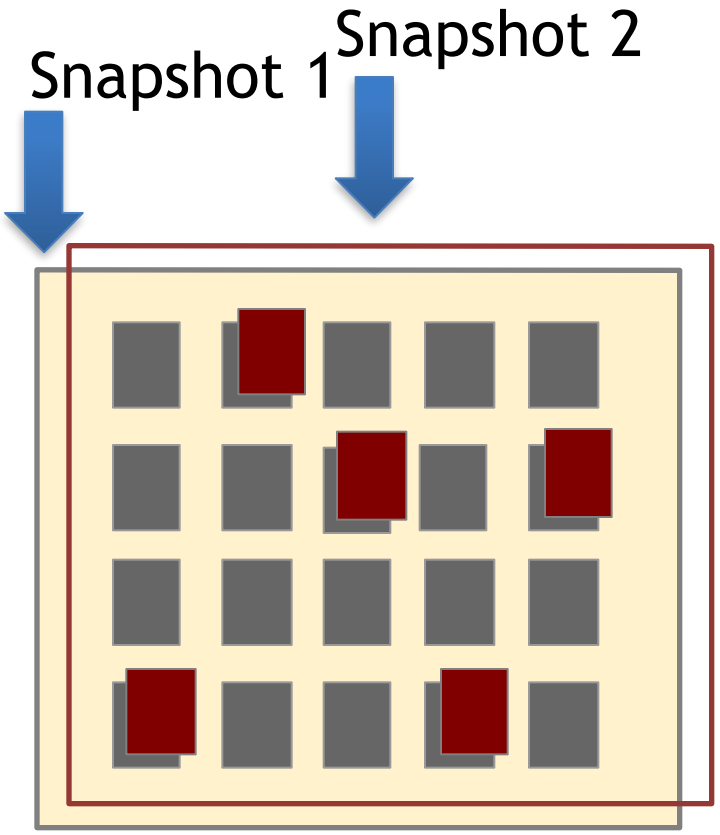


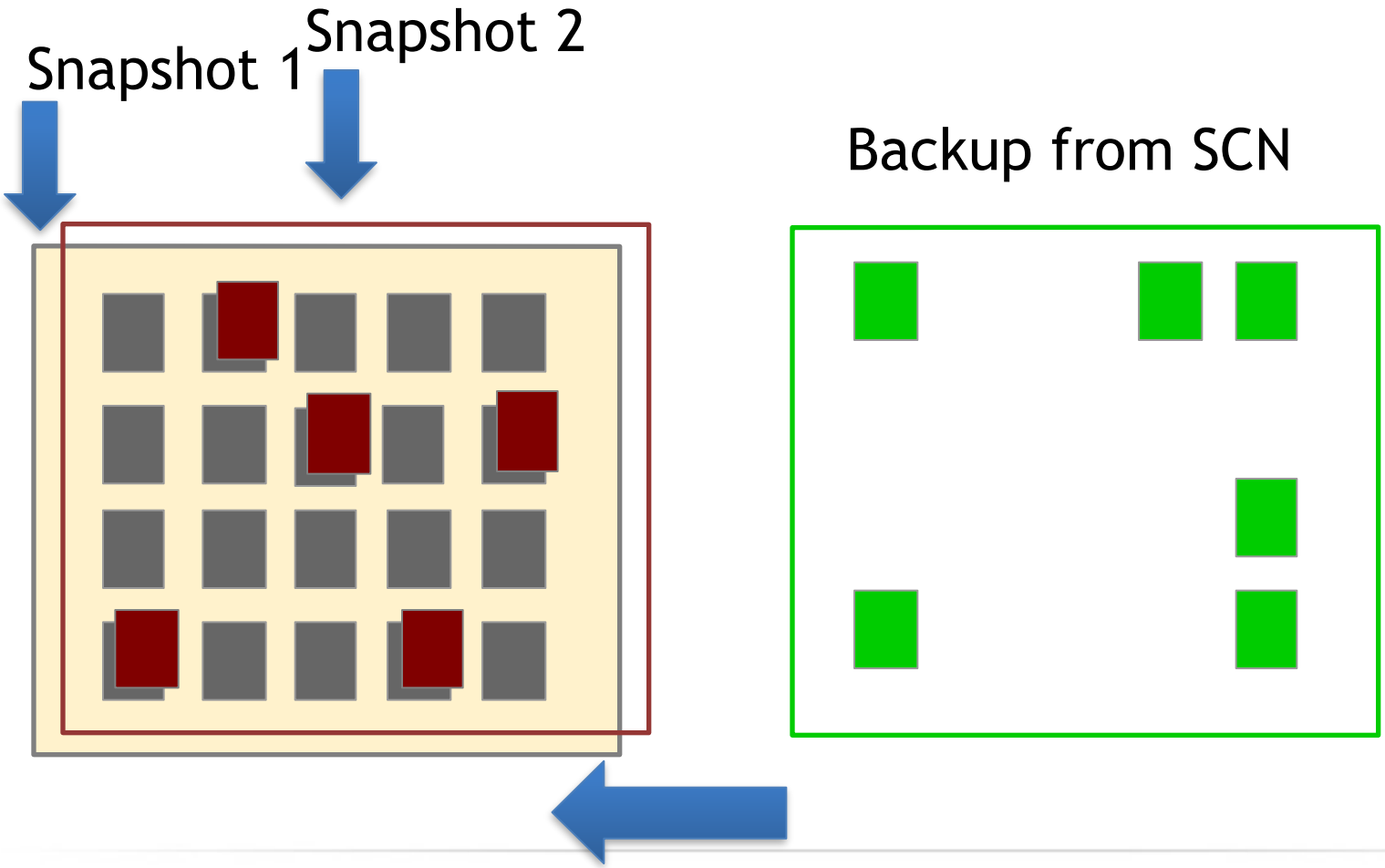
- Collected incrementally forever
- Old data purged

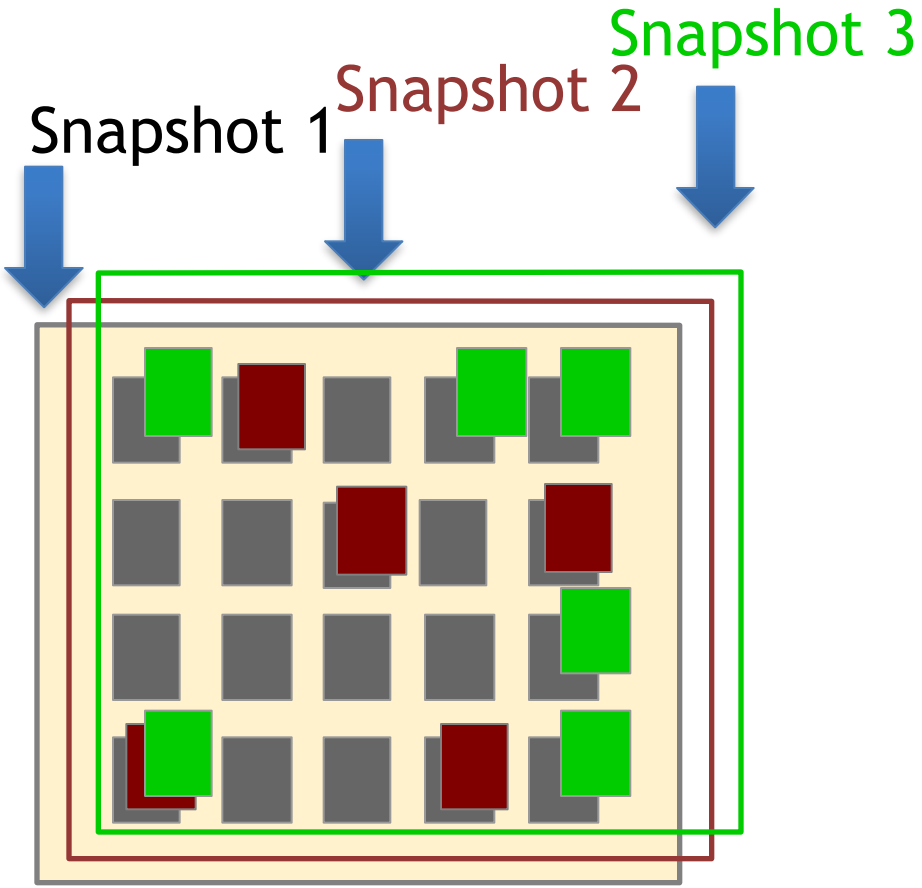
Source Full Copy

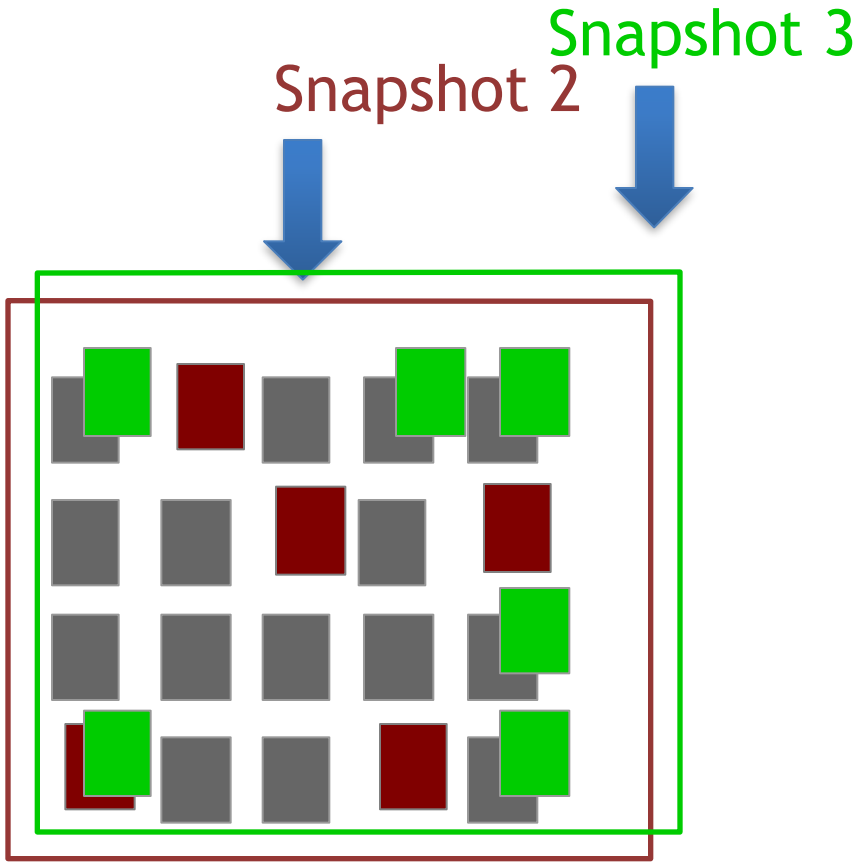
Source backup
from SCN 1




Snapshot 1
Snapshot 2

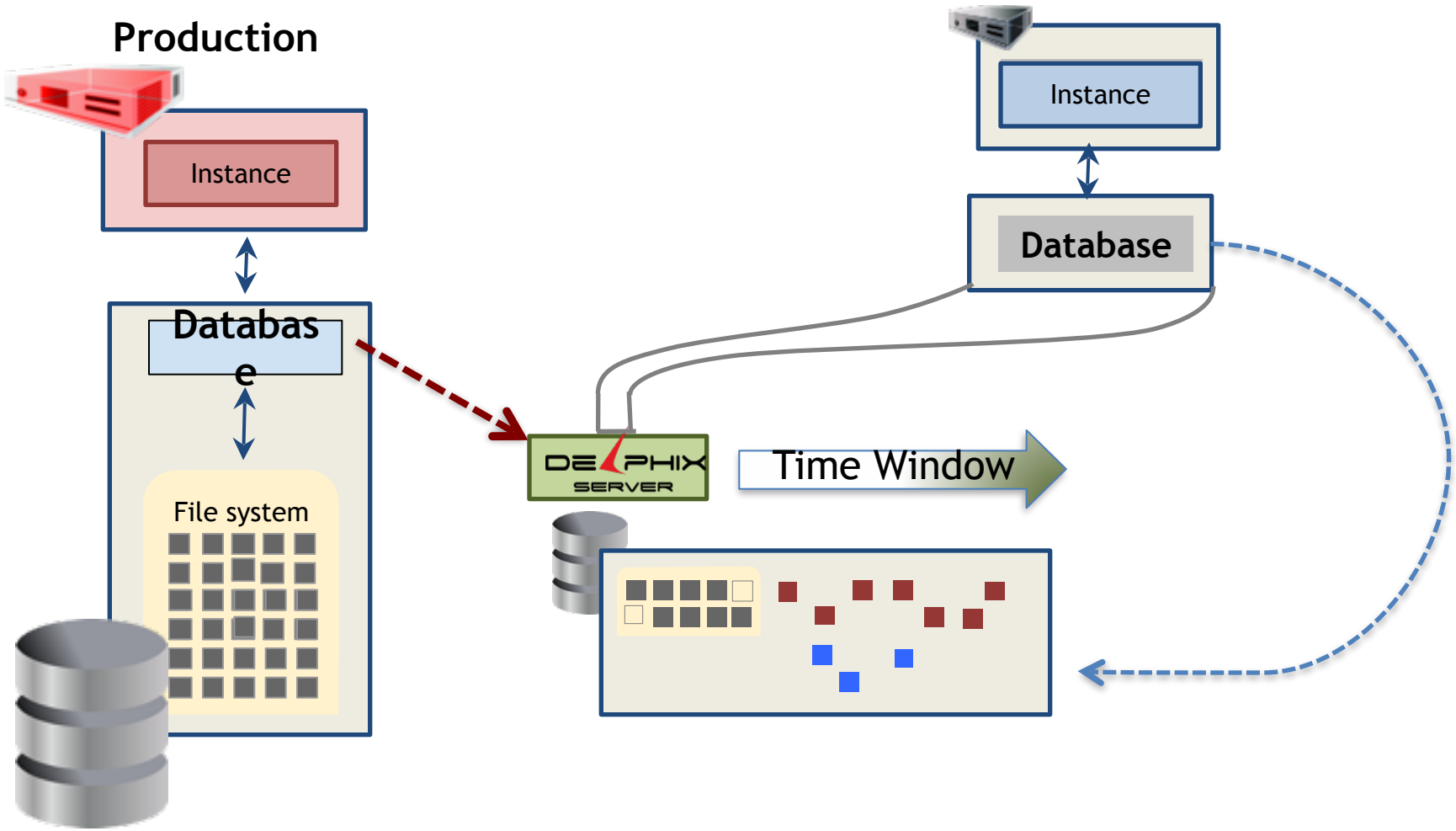

Snapshot 1
Snapshot 2
Backup from SCN


Snapshot 3
Snapshot 2
Snapshot 1



Drop
Snapshot 1

Cloning


Production
Instance
Database
File system
Instance
Database
DELPHIX SERVER
Time Window

DELPHIX

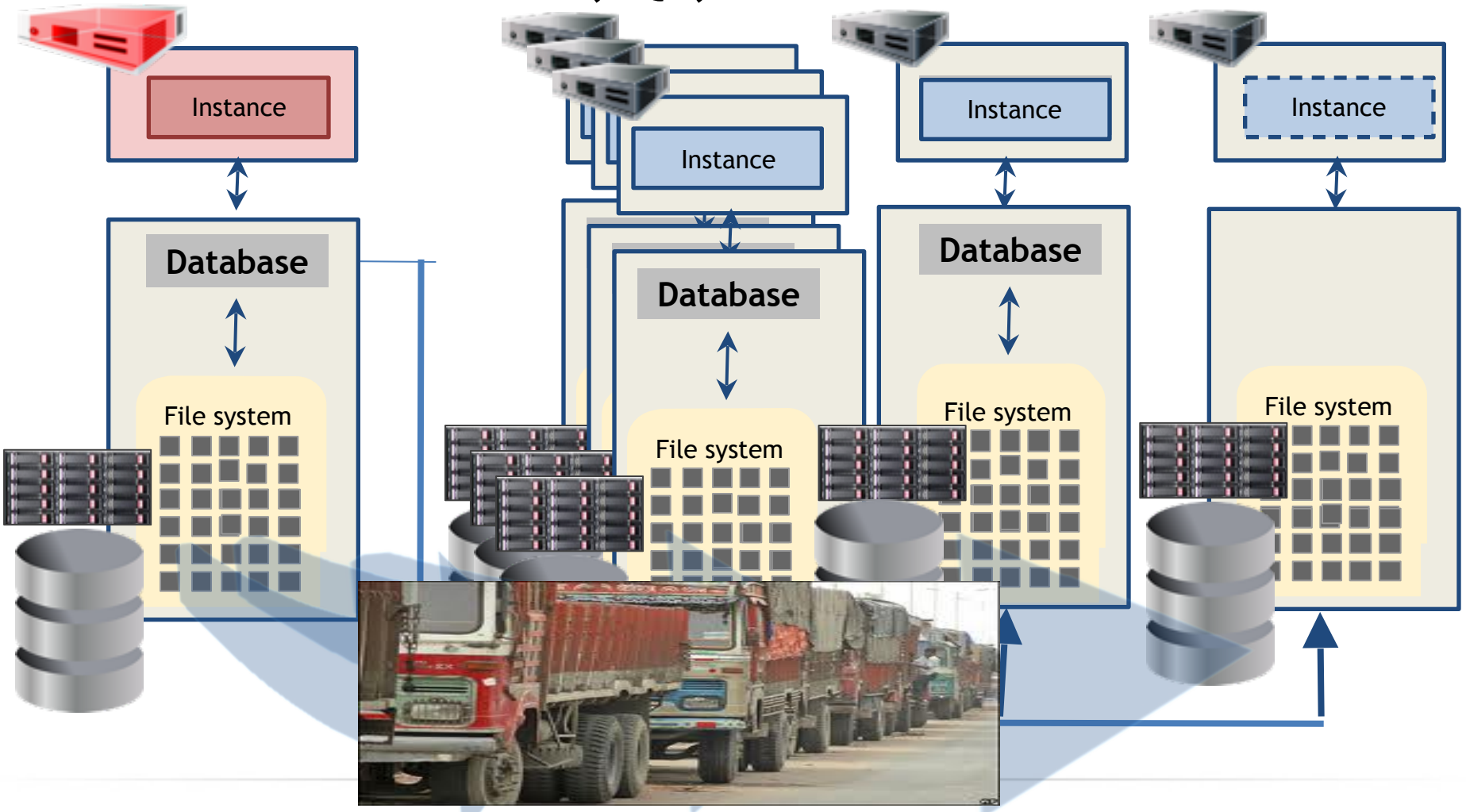
Typical Architecture

Production

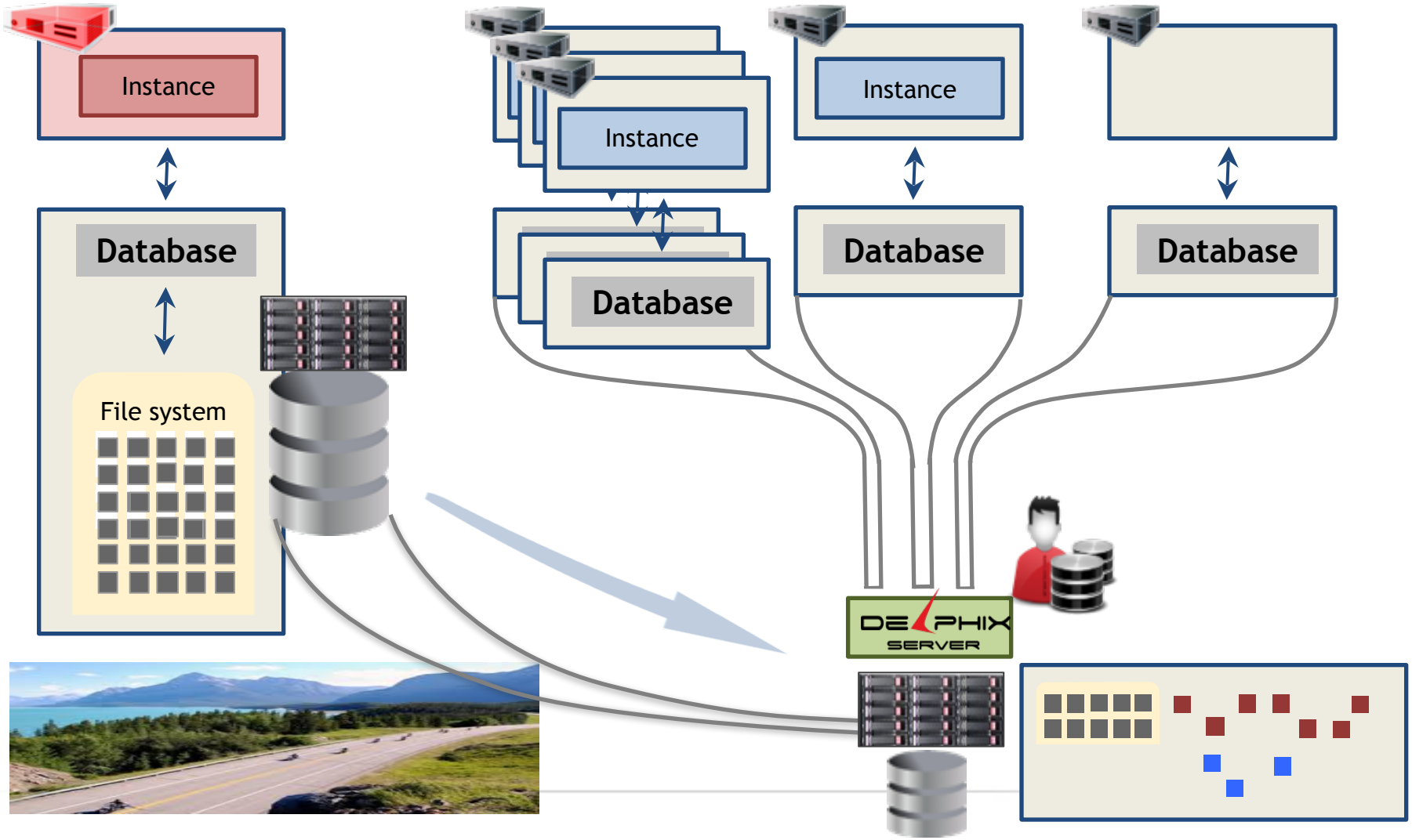
Dev, QA, UAT

Reporting

Backup

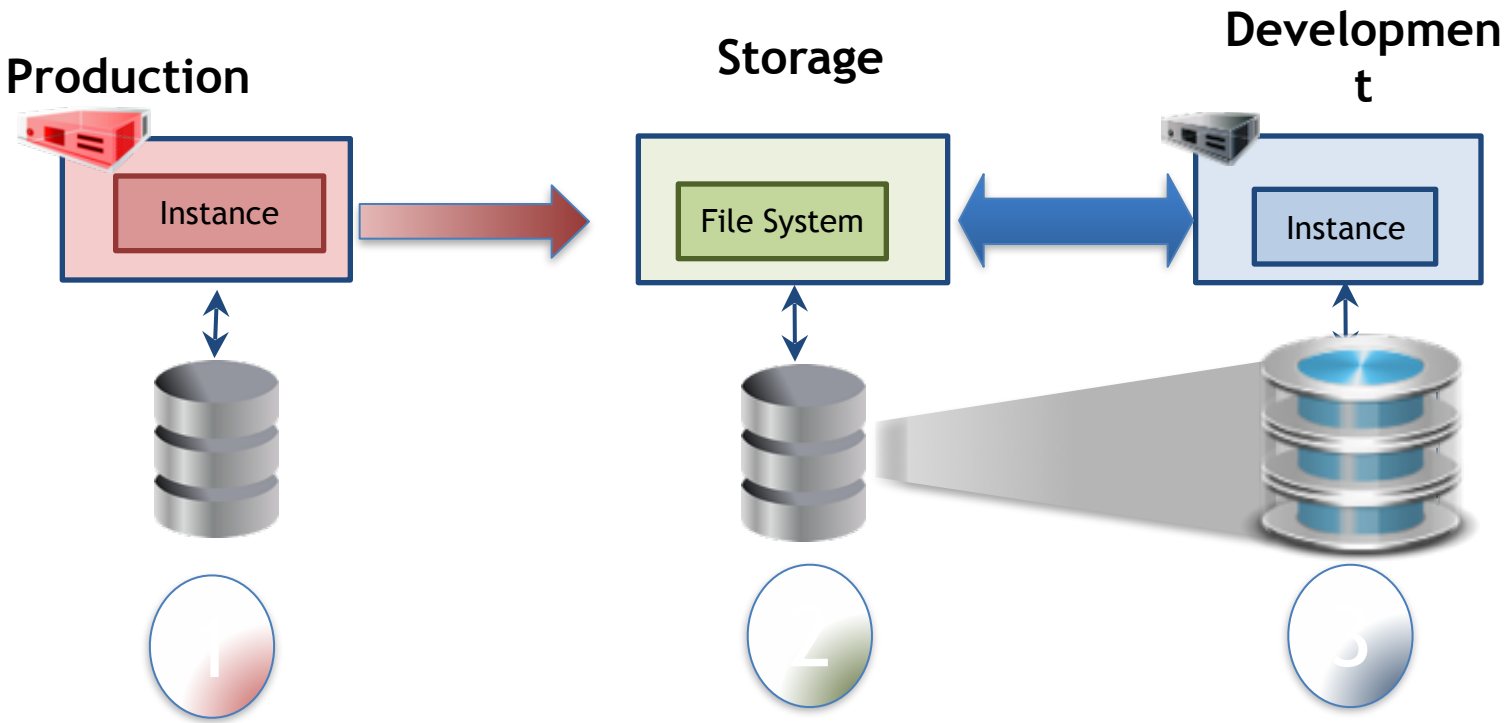


With Delphix



Databases
Host Servers
Snapshot
Sep 09 2010 5:53:10 PM EDT
Input AccurateTime:
Provision
No LogSync Delay

Three Core Parts



Copy
Sync Snapshots
Time Flow
Purge

Clone (snapshot)
Compress
Share Cache
Storage Agnostic

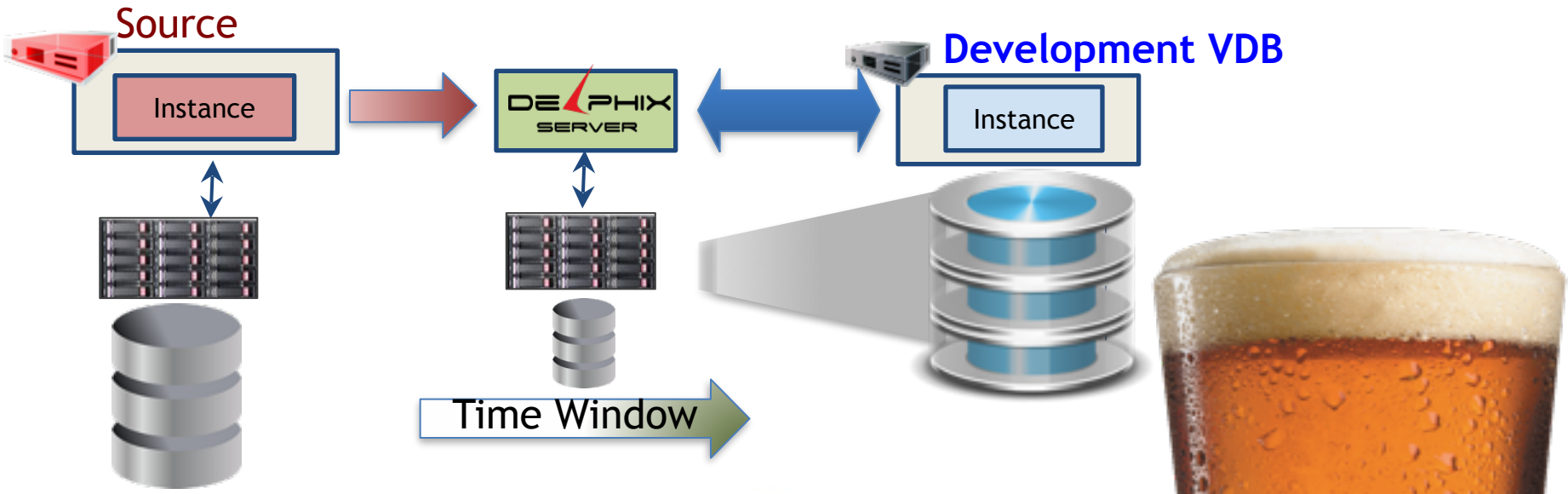
Mount, recover, rename
Self Service, Roles & Security
Rollback & Refresh
Branch & Tag

DELPHIX

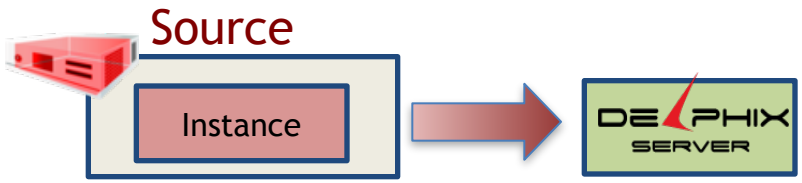
Overview of Database Virtualization

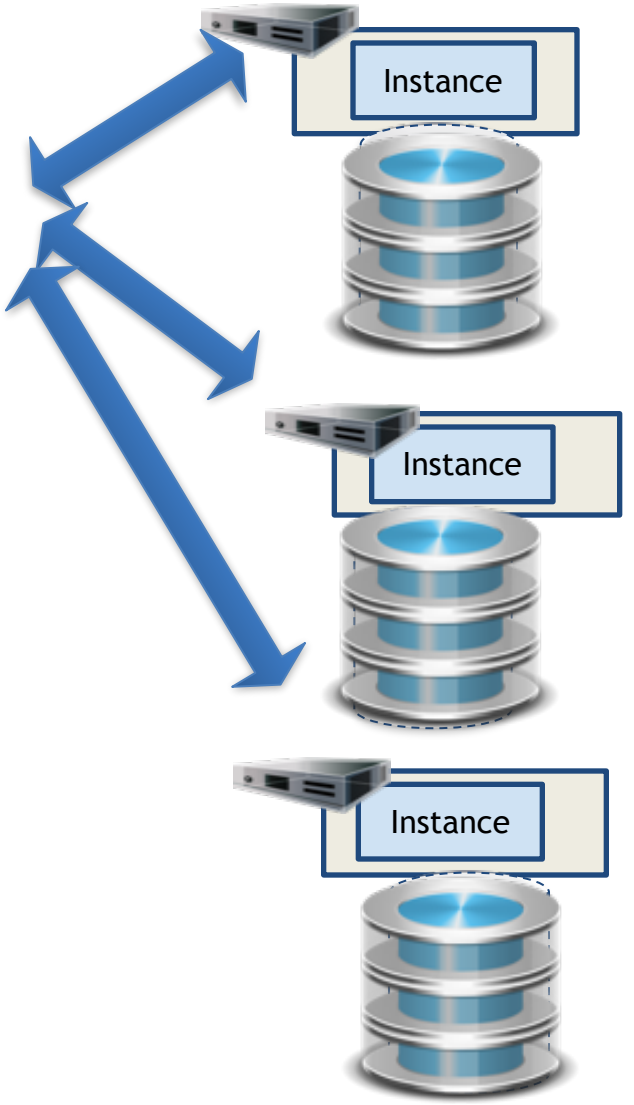
1. Fast, Fresh, Full
2. Free
3. Branching
4. Federated
5. Self Serve

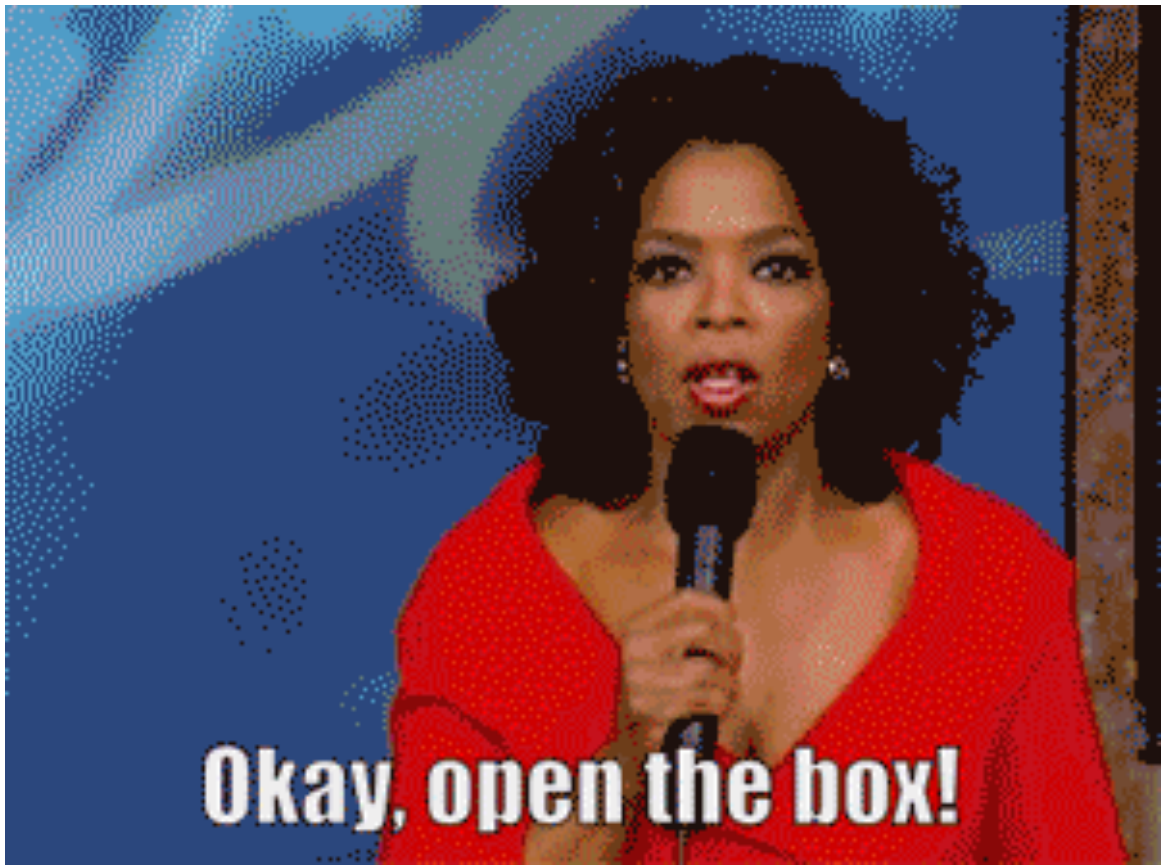
Fast, Fresh, Full



Free

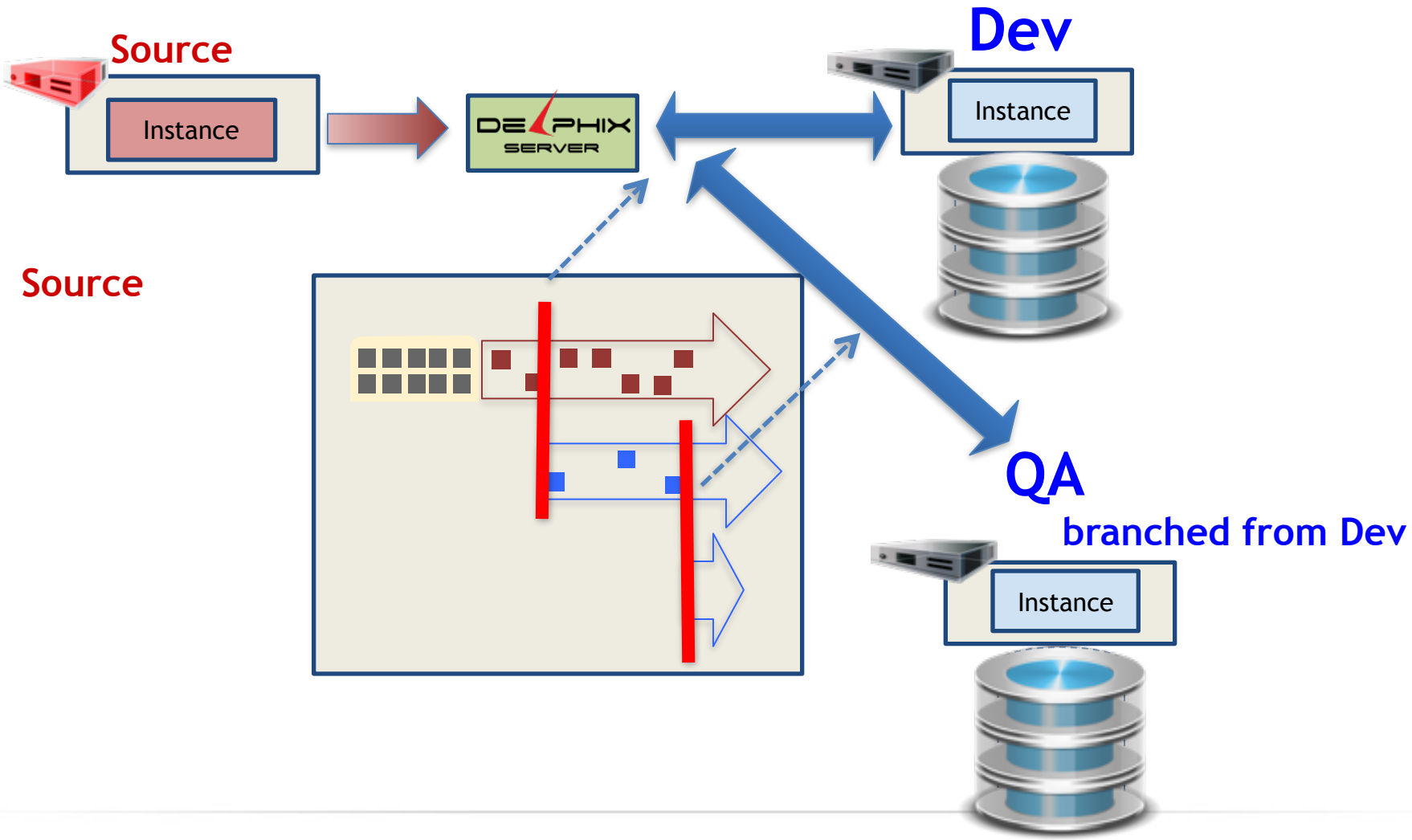






gif by Steve Karam

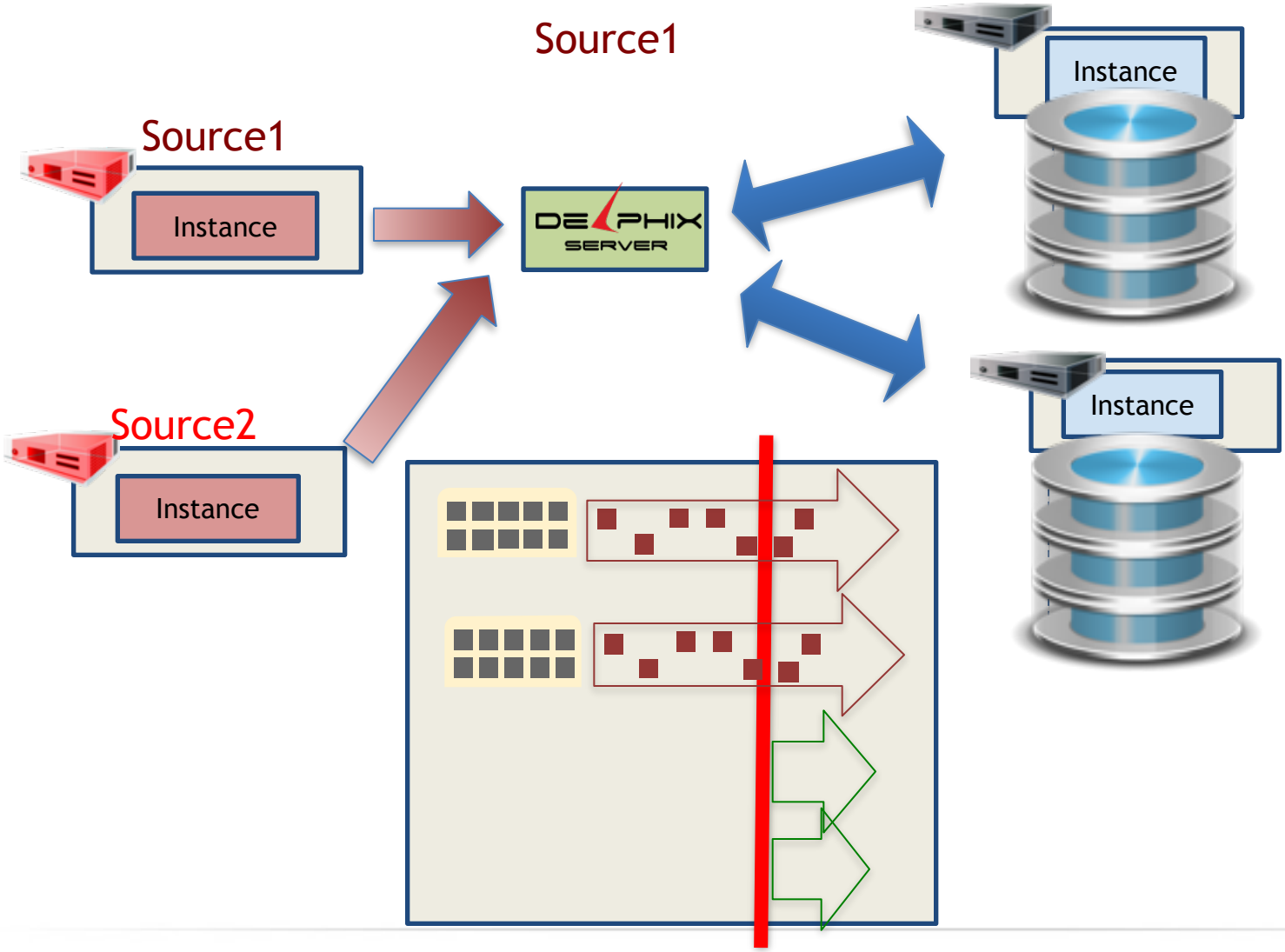
Branching


Source
Instance
DELPHIX SERVER
Dev
Instance
Source
QA
branched from Dev
Instance

DELPHIX

Federated Cloning

Federated


Source1
Source1
Instance
Source2
Instance
DELPHIX
SERVER
Instance
Instance

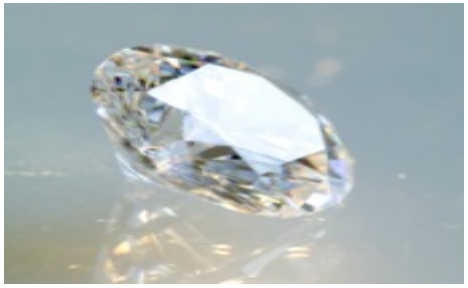
“I looked like a hero”
Tony Young, CIO
Informatica

Self Service

Use Cases

1. Development Acceleration

2. Quality



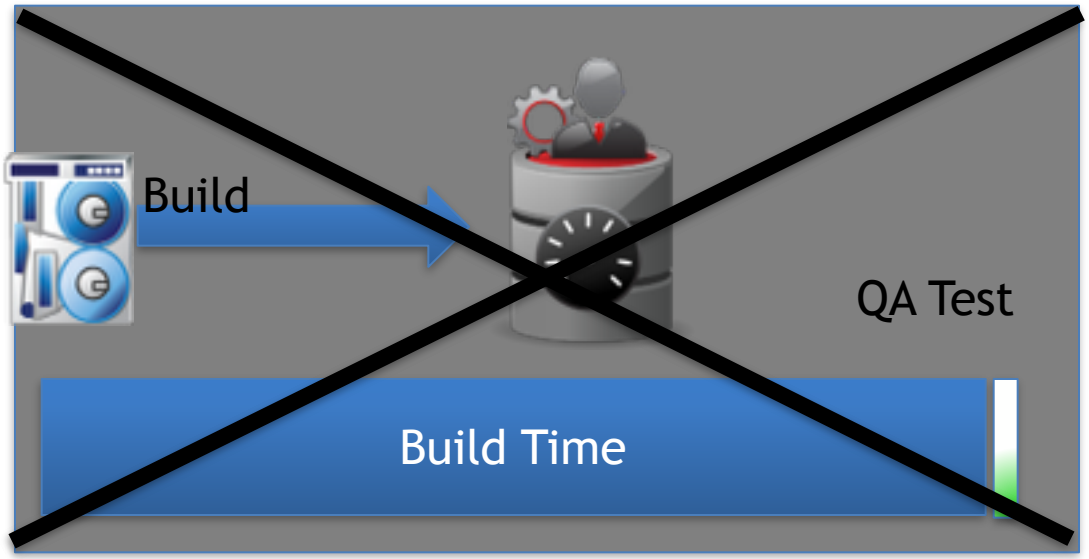
3. BI

DevOps

DevOps With Delphix

1. Efficient QA: Low cost, high utilization
2. Quick QA : Fast Bug Fix
3. Every Dev gets DB: Parallelized Dev
4. Full DB : Less Bugs
5. Fast Builds: Culture of Yes

1. Efficient QA: Lower cost

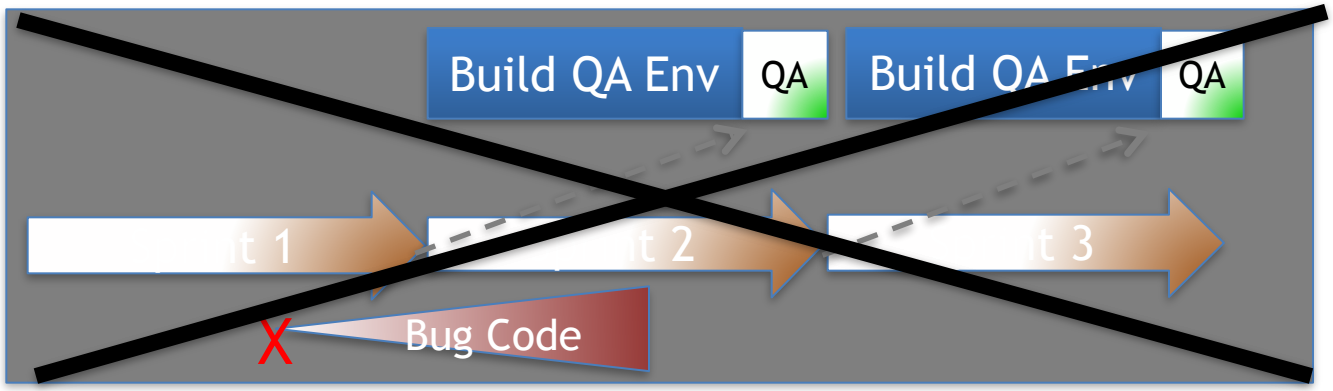


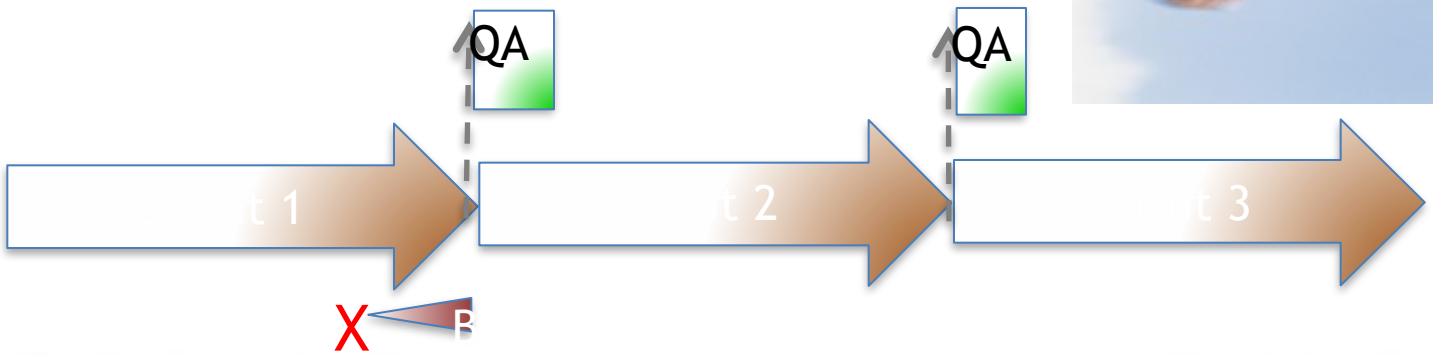
1% of QA time was building environment

$.99/$1.00 actual testing vs. setup

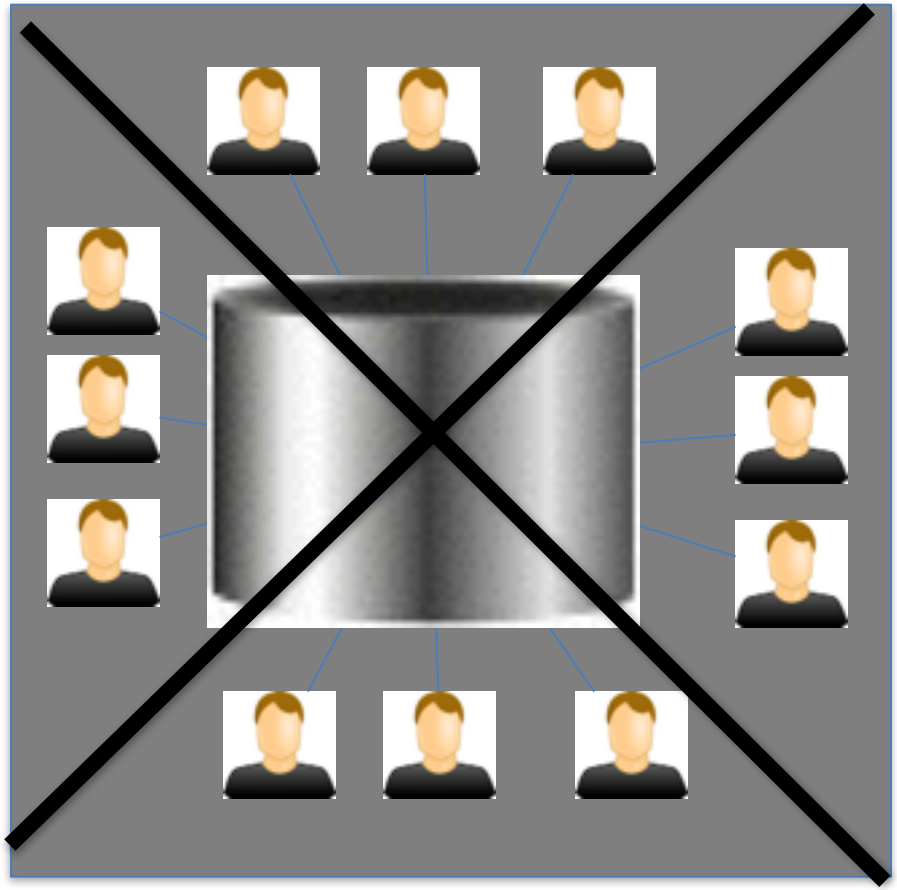
Rapid QA via Branching

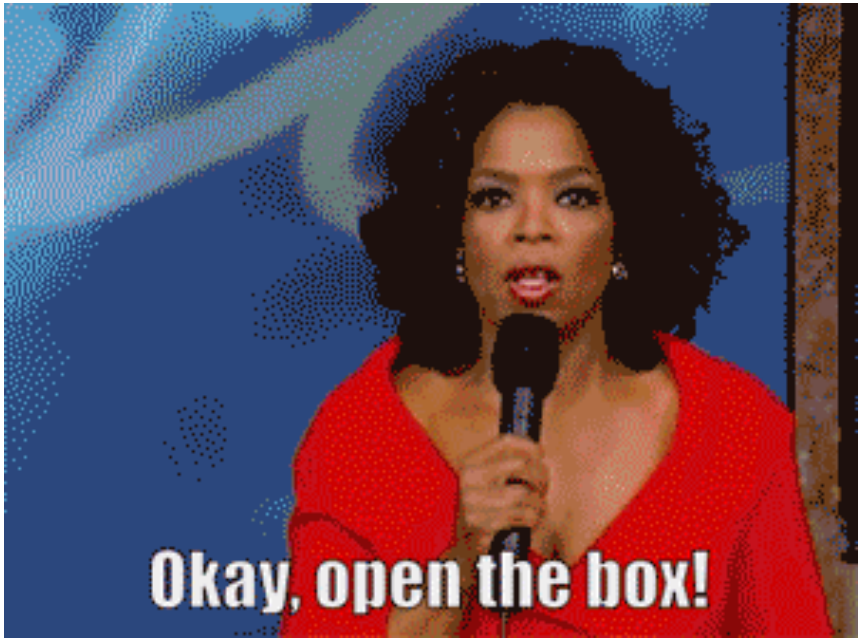
2. QA Immediate: bugs found fast and fixed





3. Private Copies: Parallelize

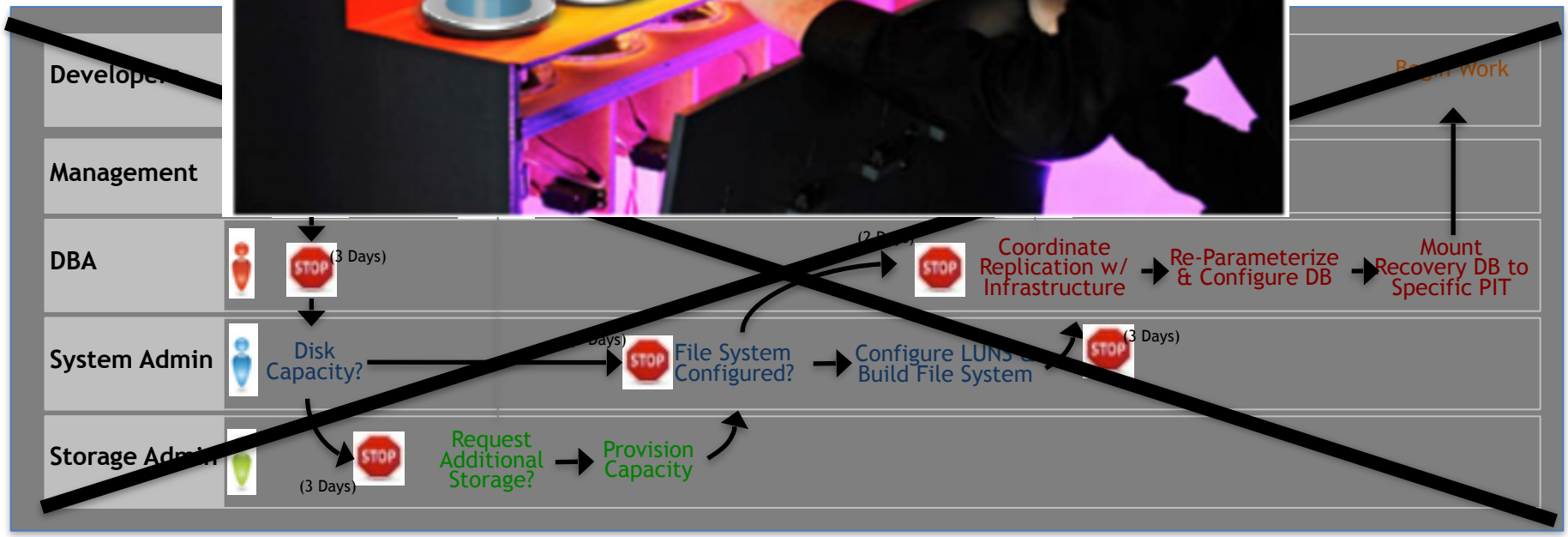




gif by Steve Karam

4. Full Size DB : Eliminate bugs

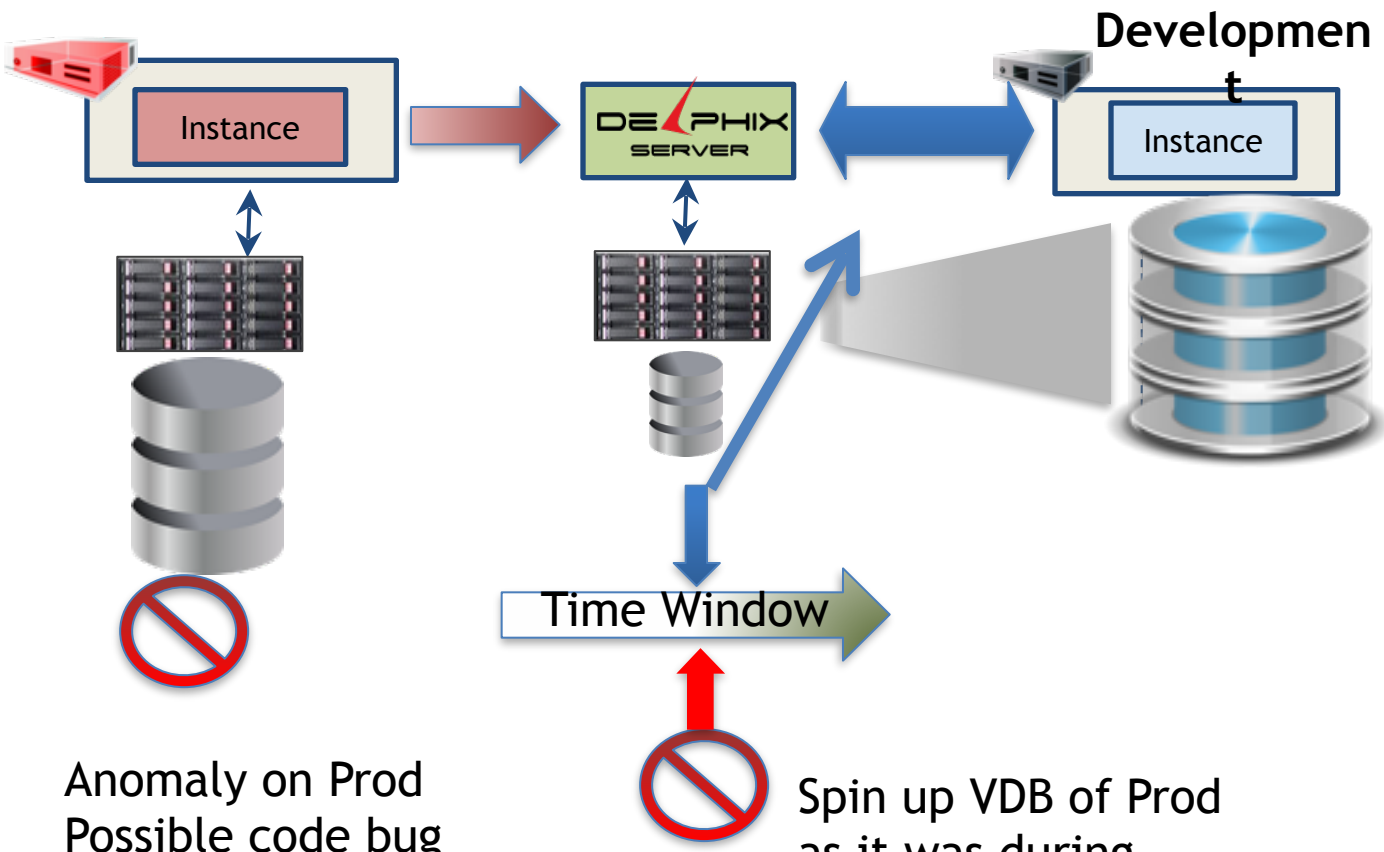
5. Self Service: Fast, Efficient. Culture of Yes!



Quality

1. Forensics
2. Testing
3. Recovery

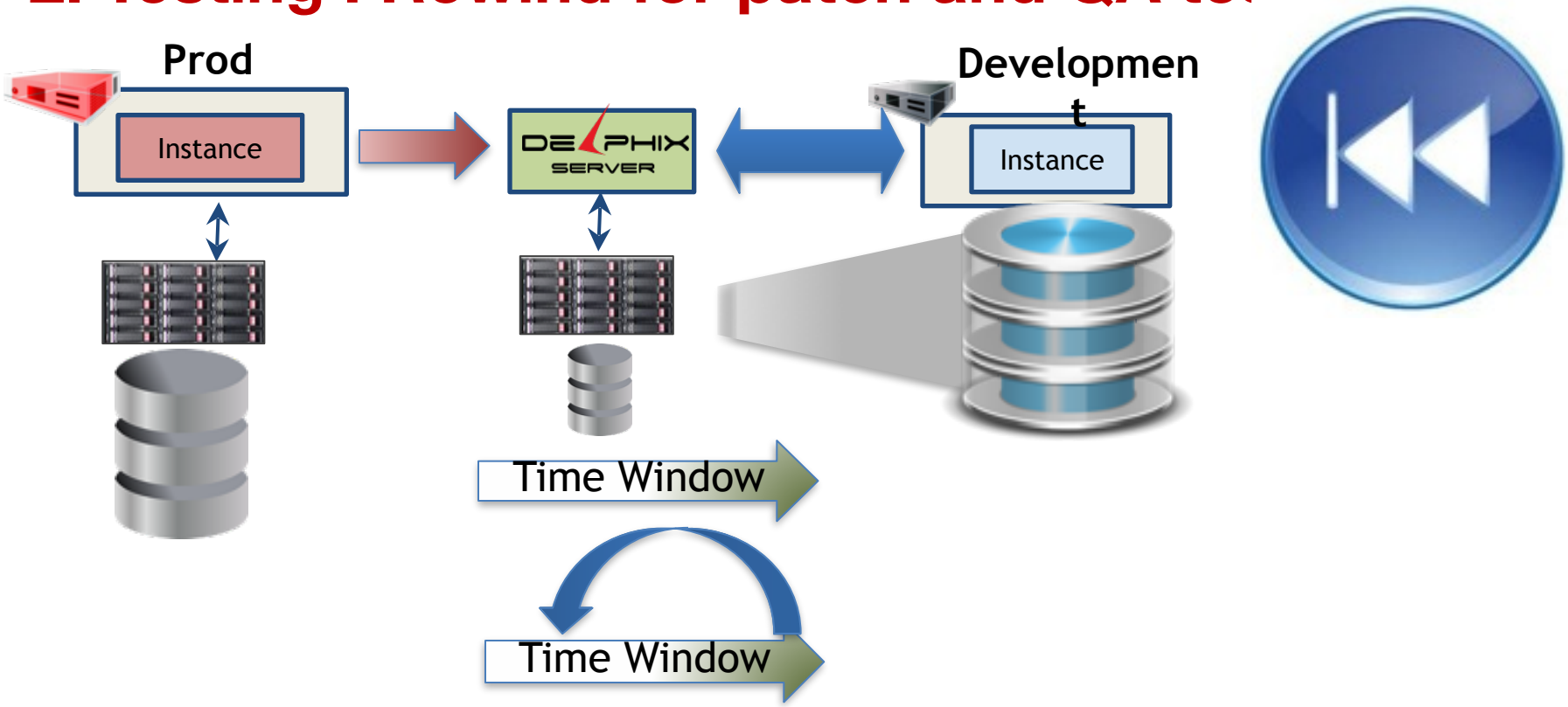
1. Forensics: Investigate Production Bugs



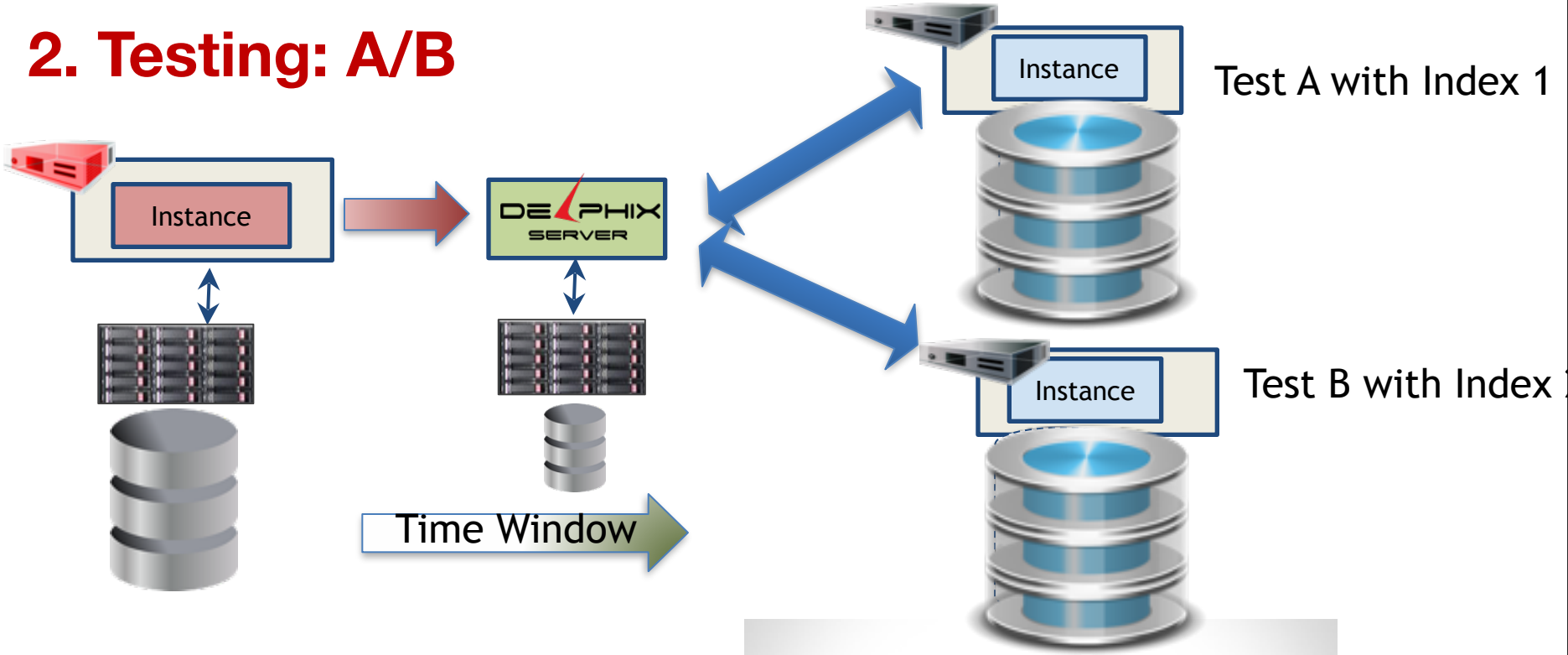
Anomaly on Prod
Possible code bug
At noon yesterday

Spin up VDB of Prod
as it was during
anomaly

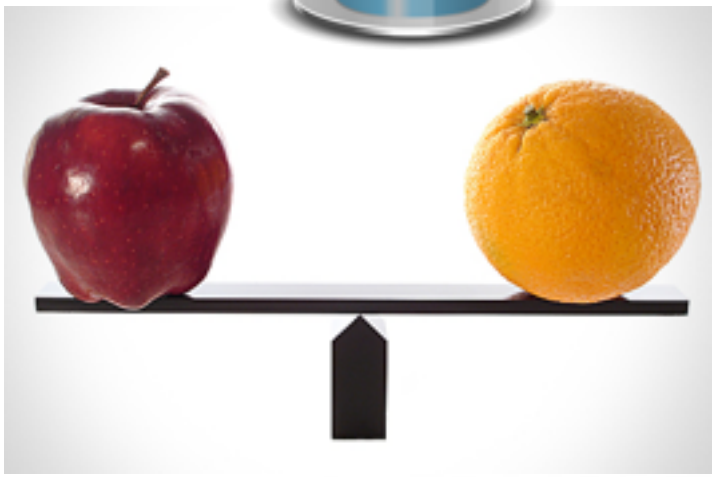
2. Testing : Rewind for patch and QA testing



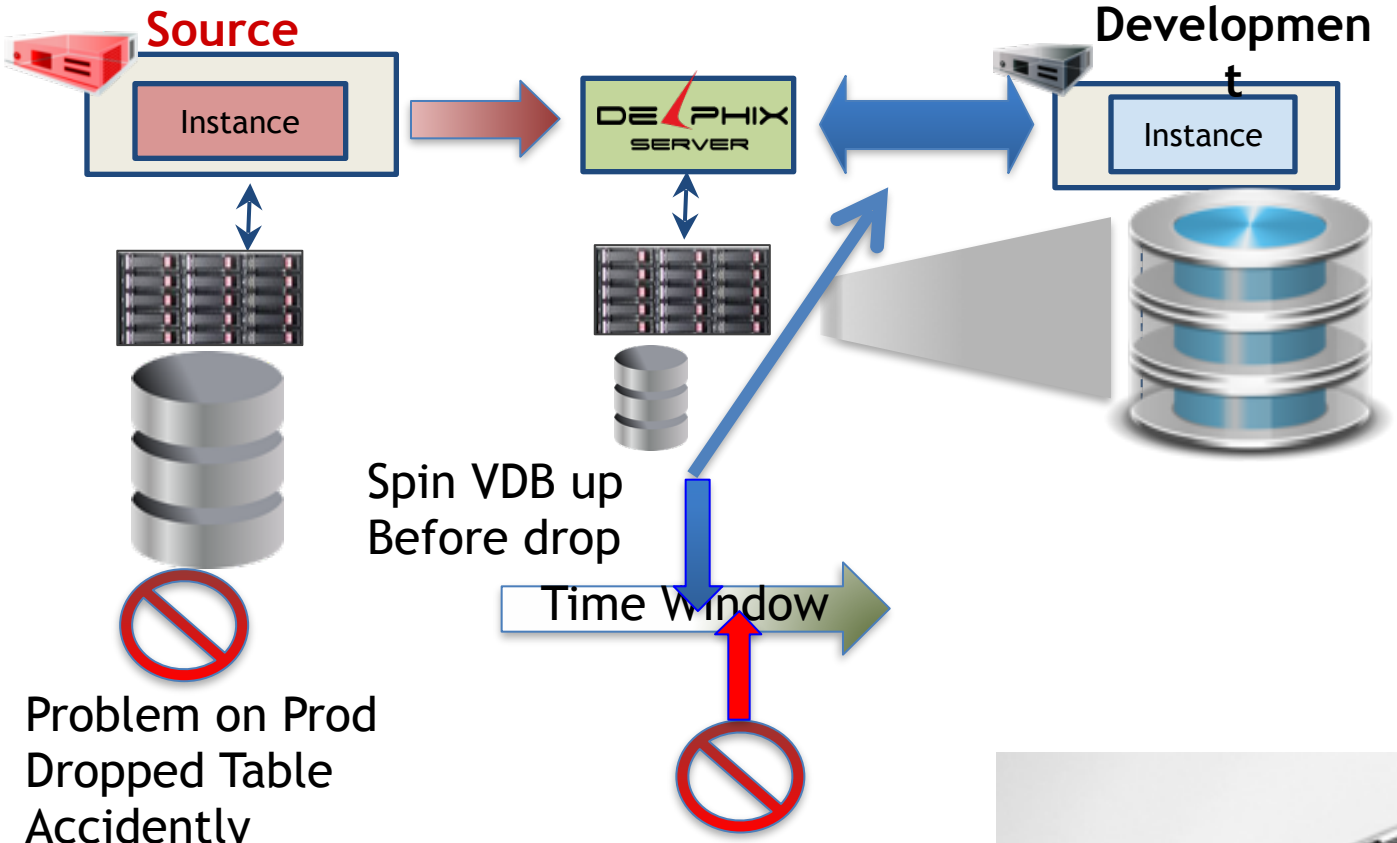
2. Testing: A/B



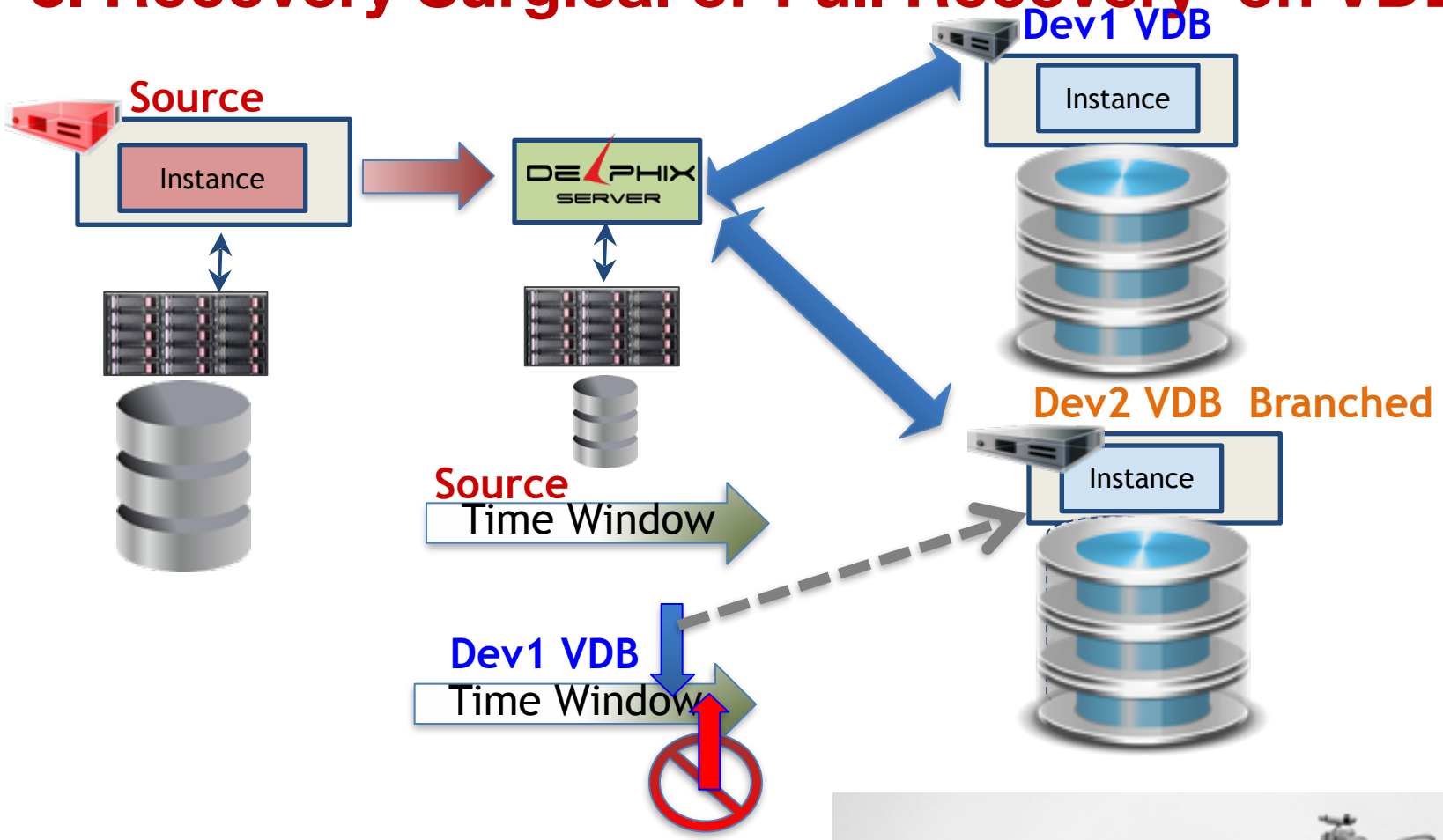
- Keep tests for compare
- Production vs Virtual
  - invisible index on Prod
  - Creating index on virtual
- Flashback vs Virtual



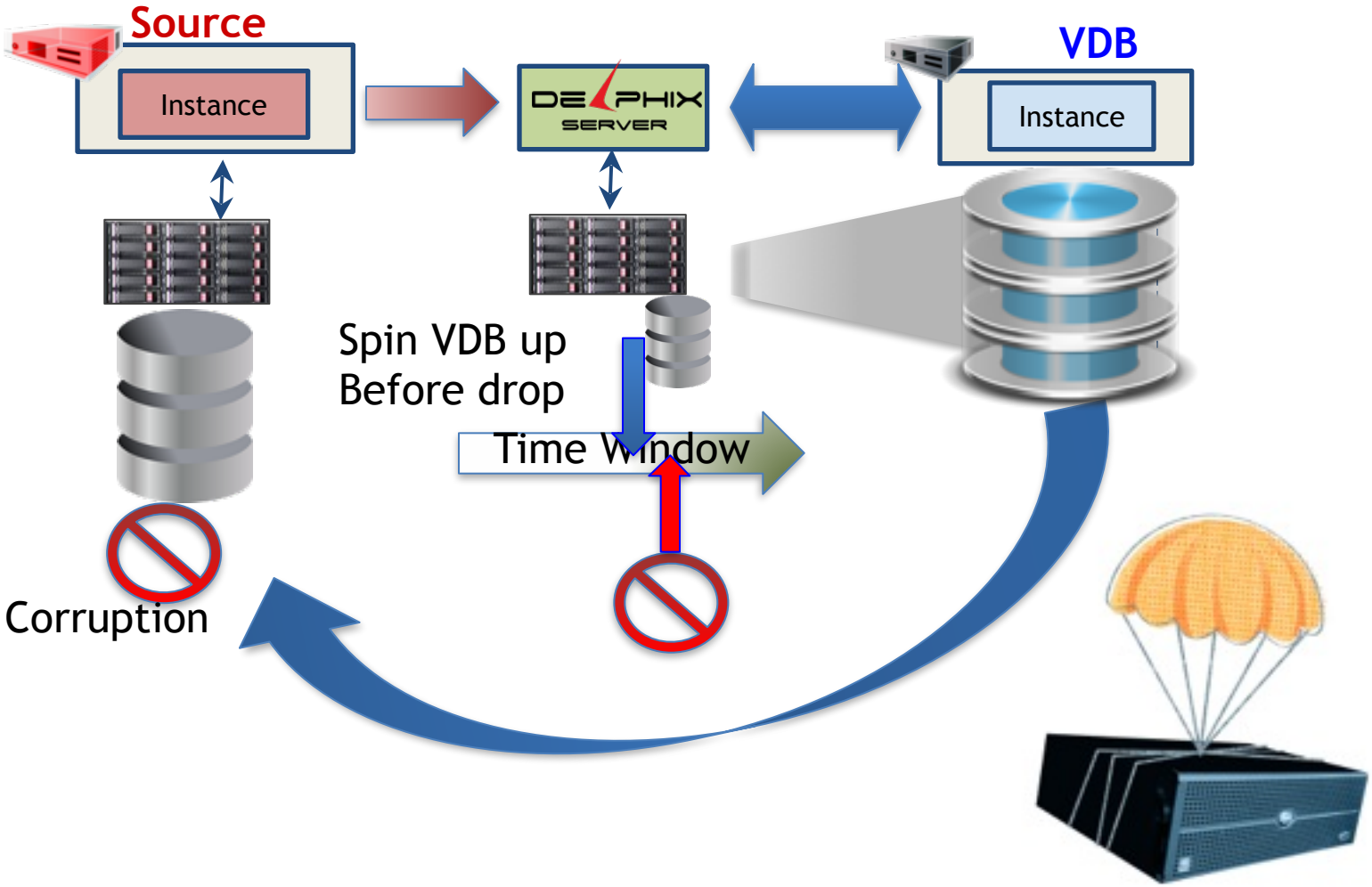
3. Recovery: Surgical recover of Production

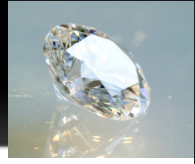


3. Recovery Surgical or Full Recovery on VDB



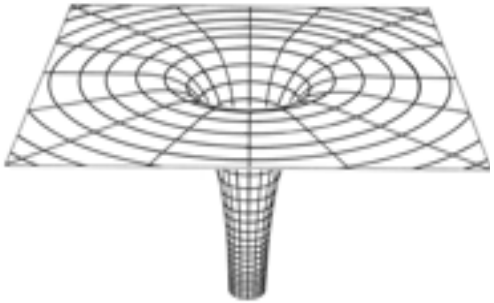
3. Recovery: Virtual to Physical

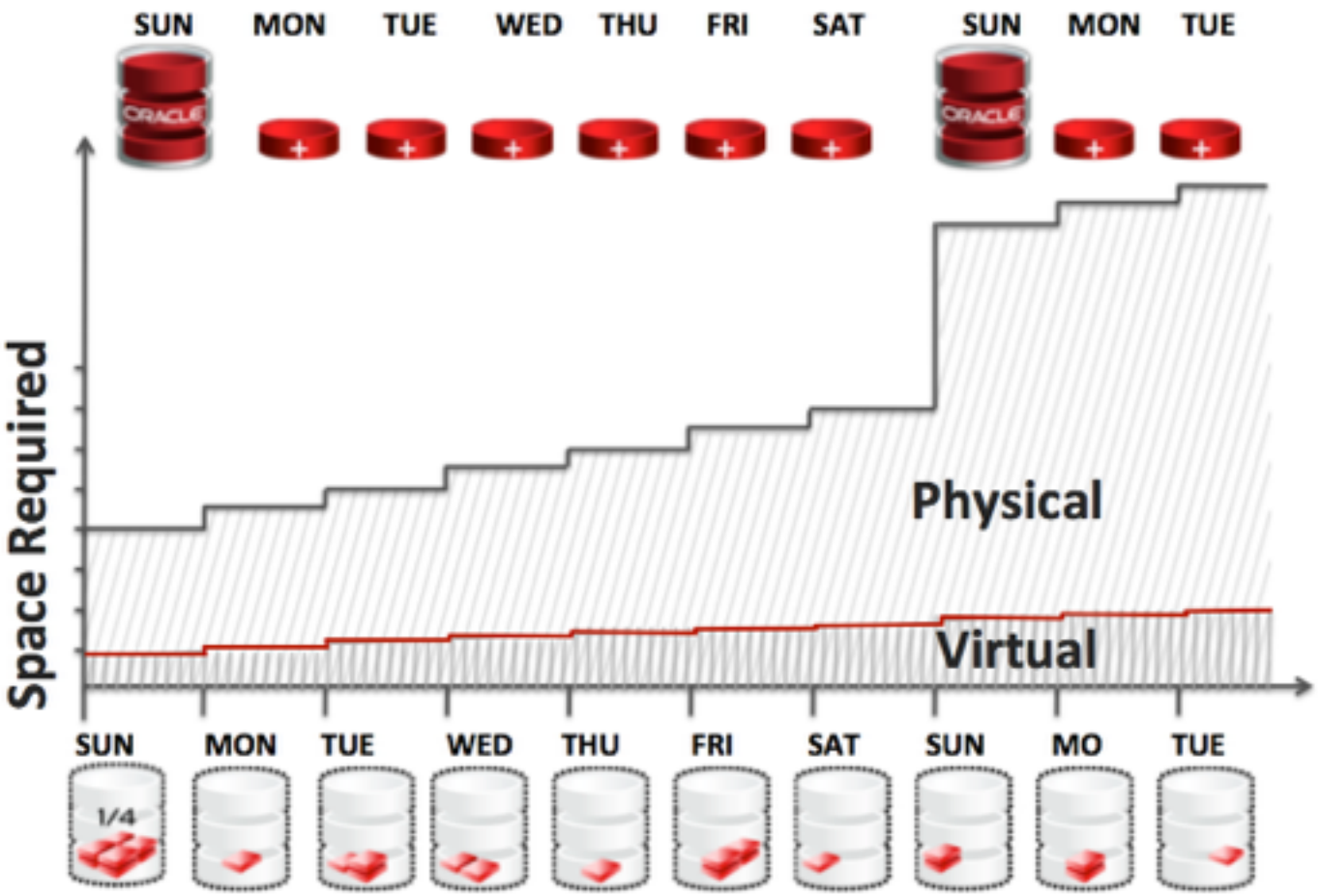




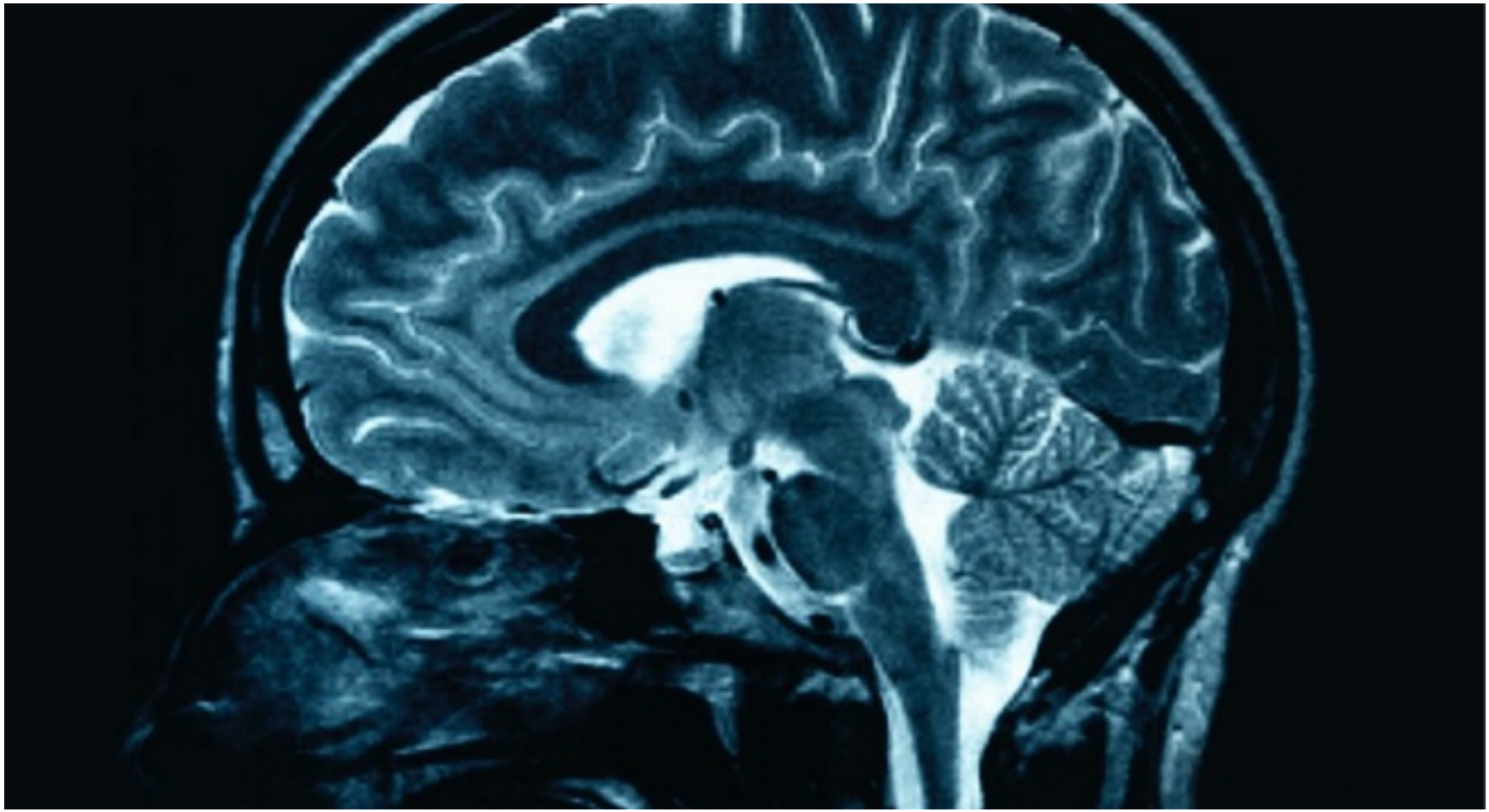
3. Recovery

BACKUP



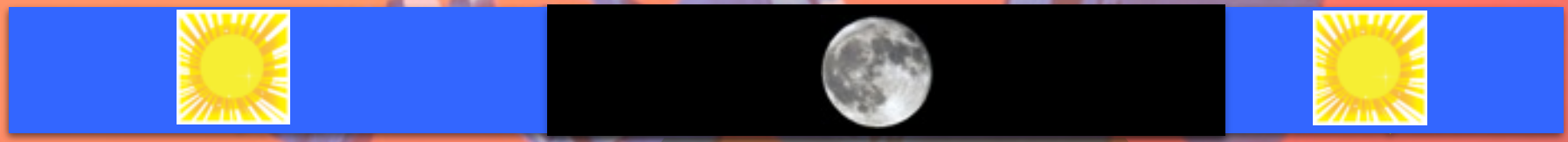

SUN
MON
TUE
WED
THU
FRI
SAT
SUN
MON
TUE
ORACLE
Space Required
Physical
Virtual
SUN
MON
TUE
WED
THU
FRI
SAT
SUN
MO
TUE
1/4

Business Intelligence



ETL and Refresh Windows

ETL and DW refreshes taking longer



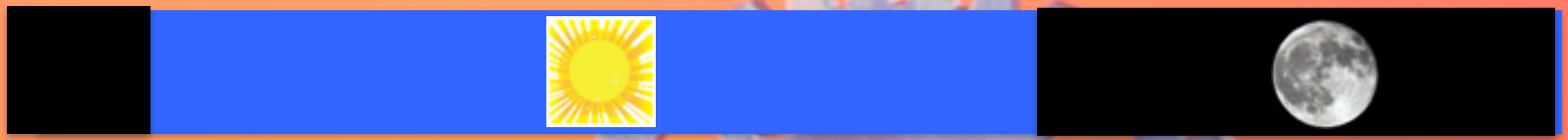
1pm
8am

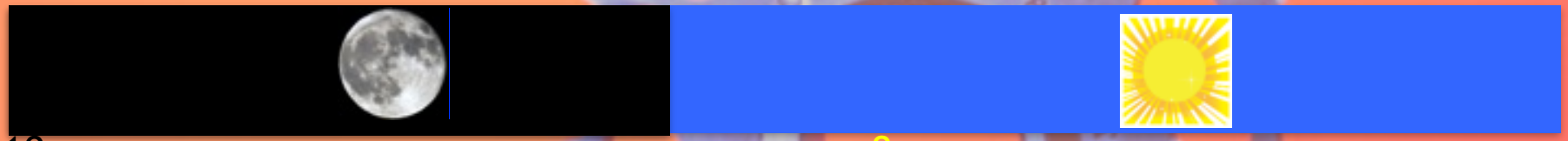
noon

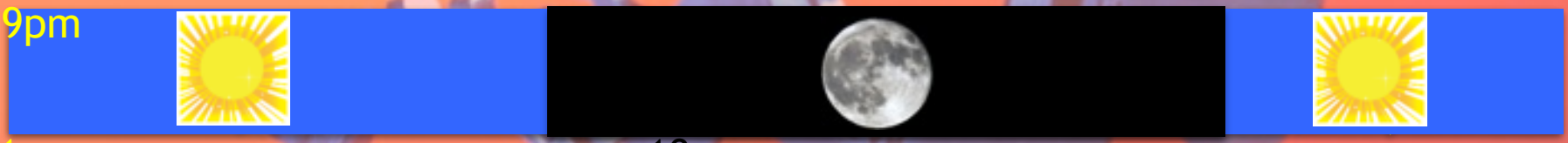
10pm

2011
2012
2013
2014
2015

ETL and Refresh Windows







8am
noon

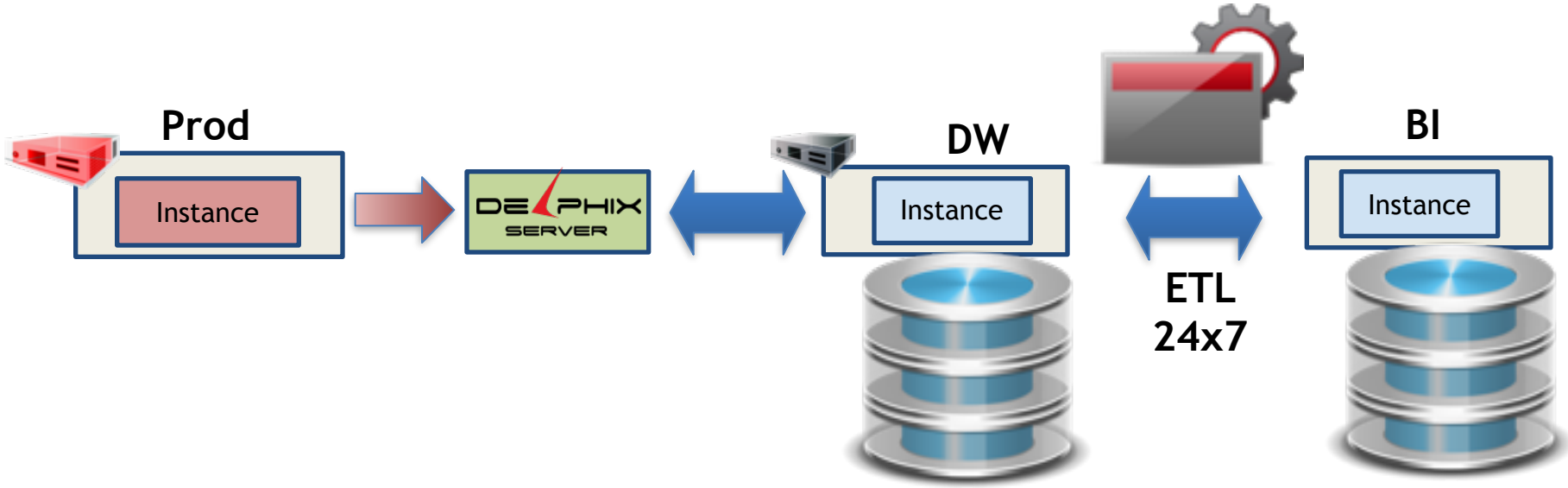
2011
2012
2013
2014
2015

ETL and DW Refreshes

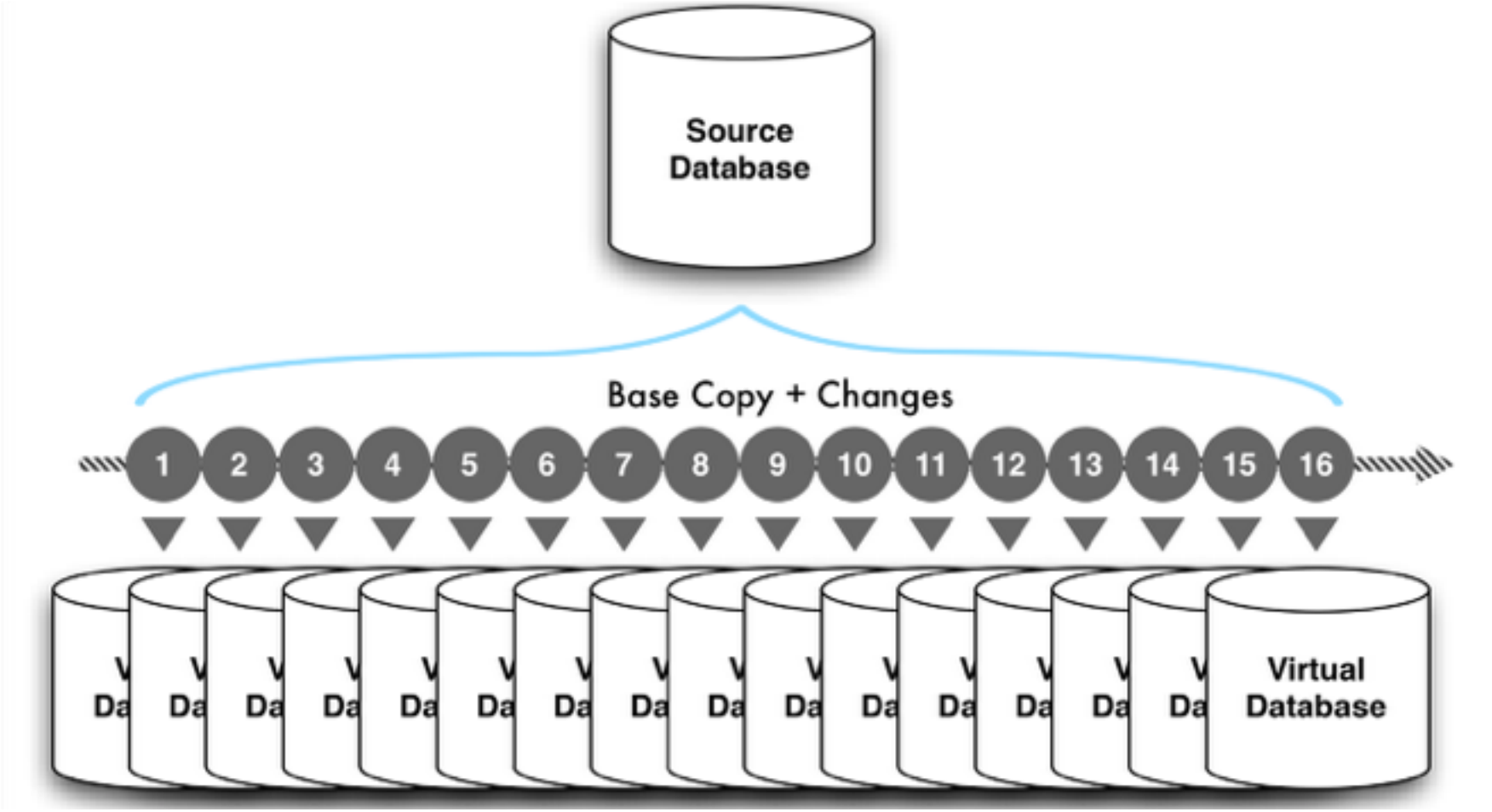
Data Guard – requires full refresh if used
Active Data Guard – read only, most reports don't work

Fast Refreshes

- Collect only Changes
- Refresh in minutes



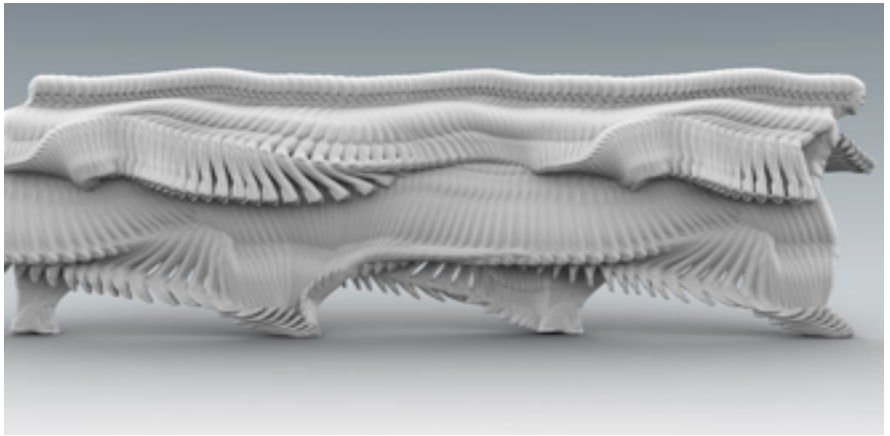
Temporal Data



BI

a) Fast refreshes

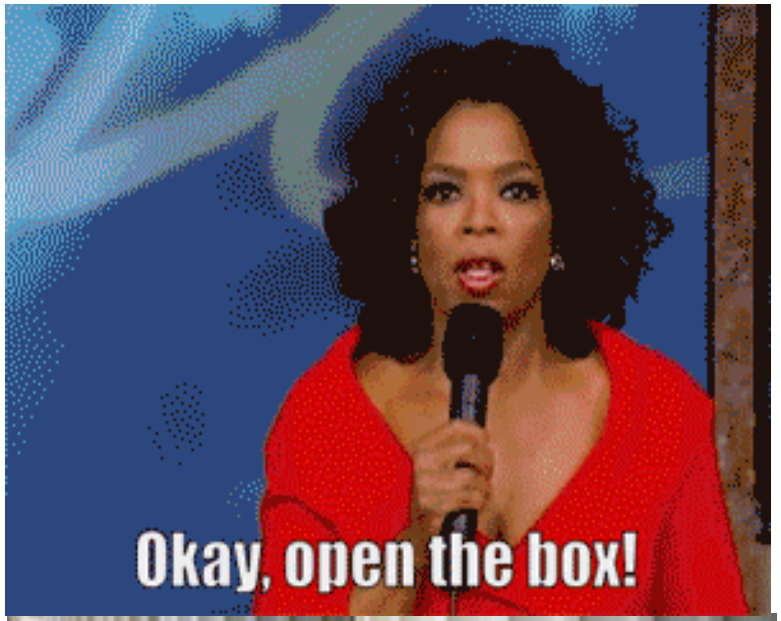
a) Temporal queries



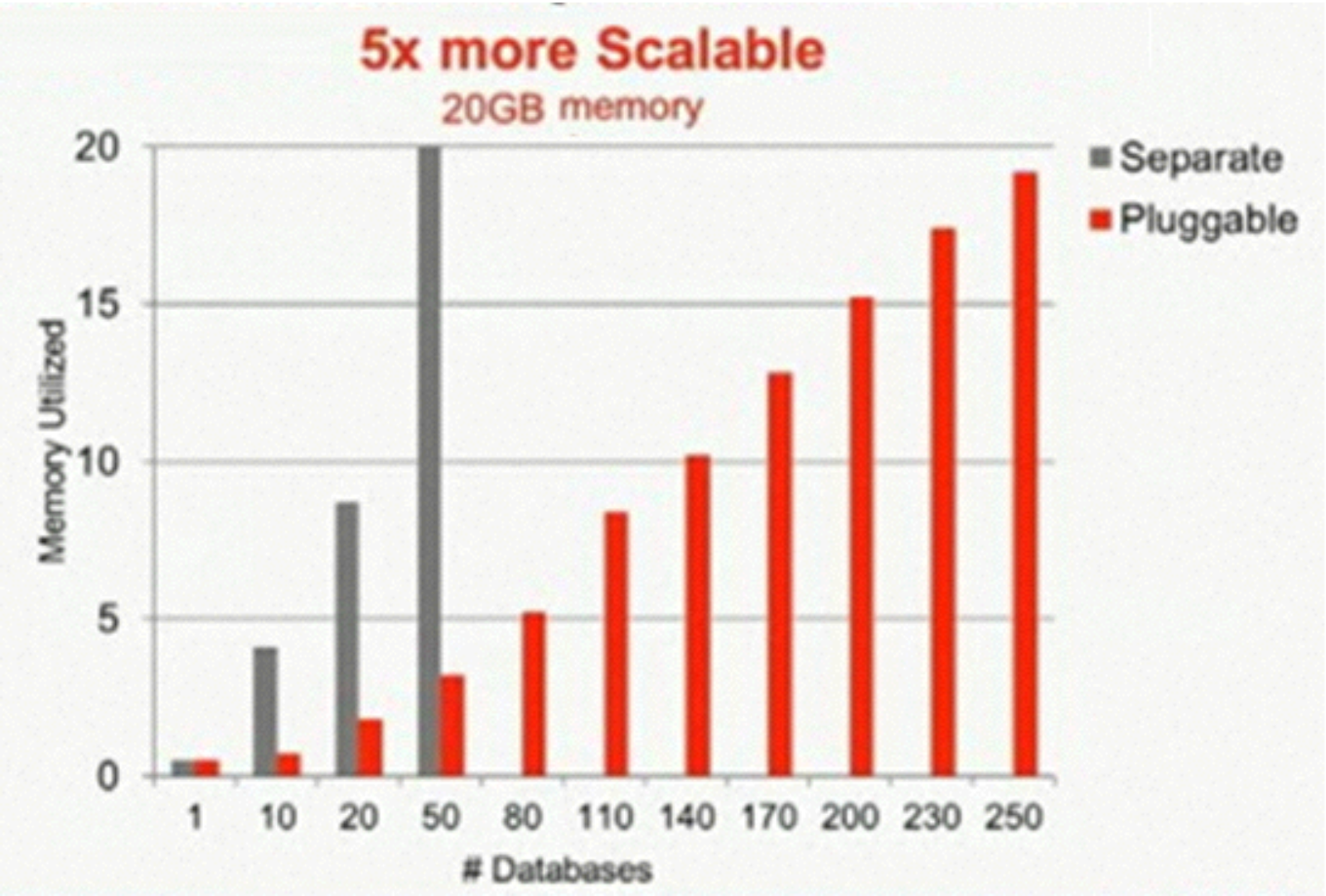
b) Confidence testing

Review: Use Cases

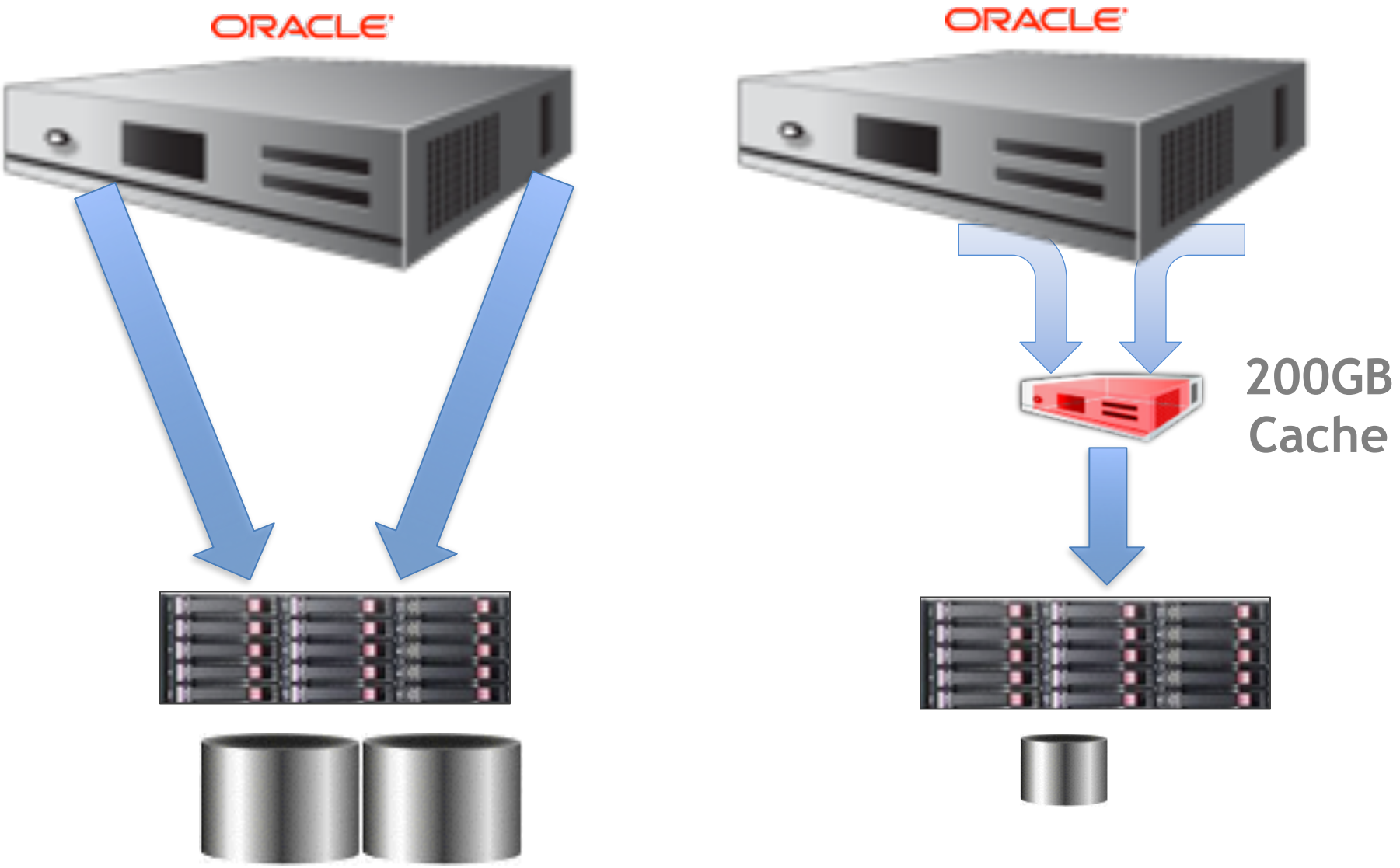
1. Development Acceleration
   a) Full, Fresh, Fast , Self Serve
   b) QA Branching
   c) Federated
2. Quality
   a) Forensics
   b) Testing : A/B, upgrade, patch
   c) Recovery: logical, physical
3. BI
   a) Fast refresh
   b) Temporal Data
   c) Confidence testing

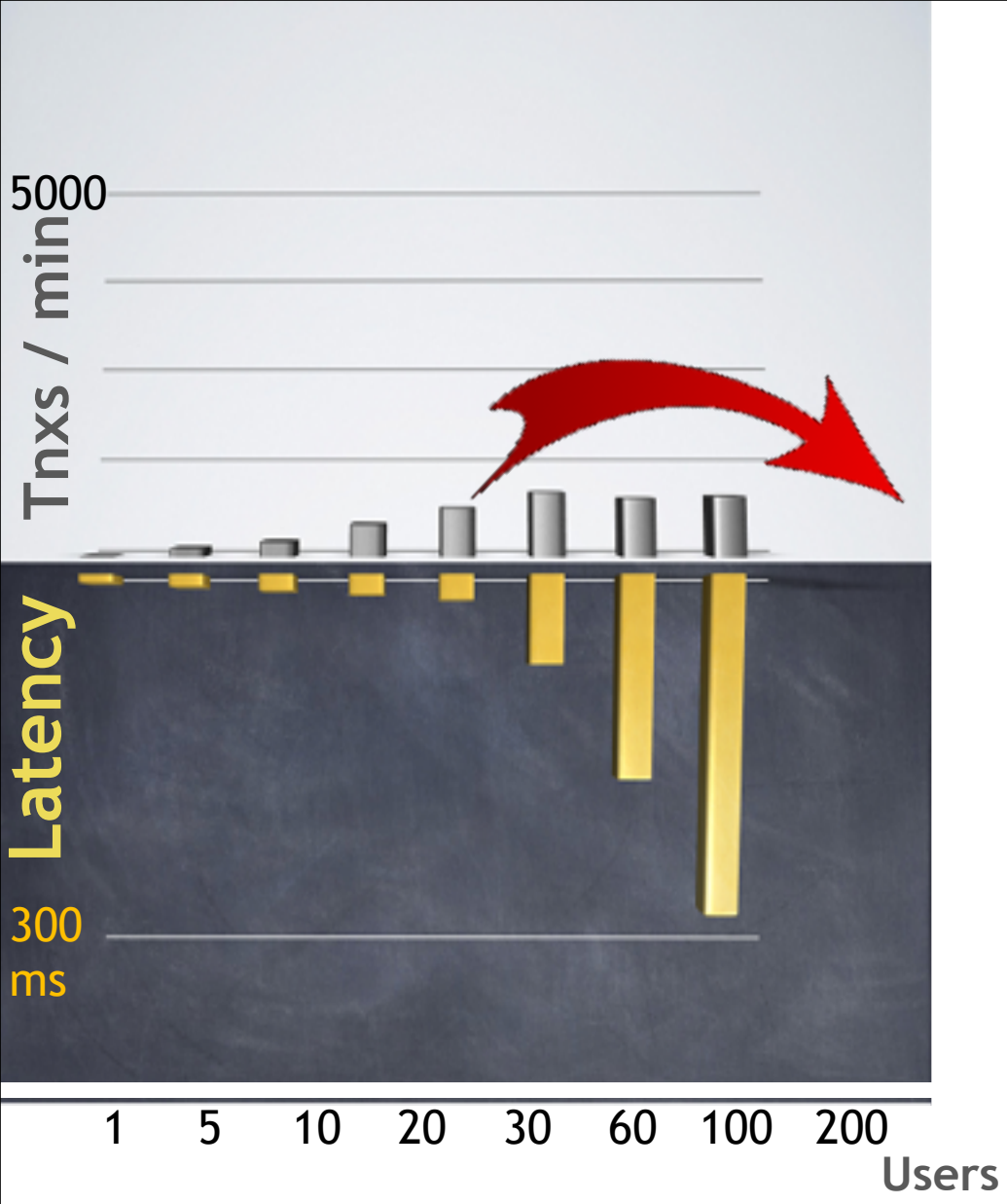

Okay, open the box!

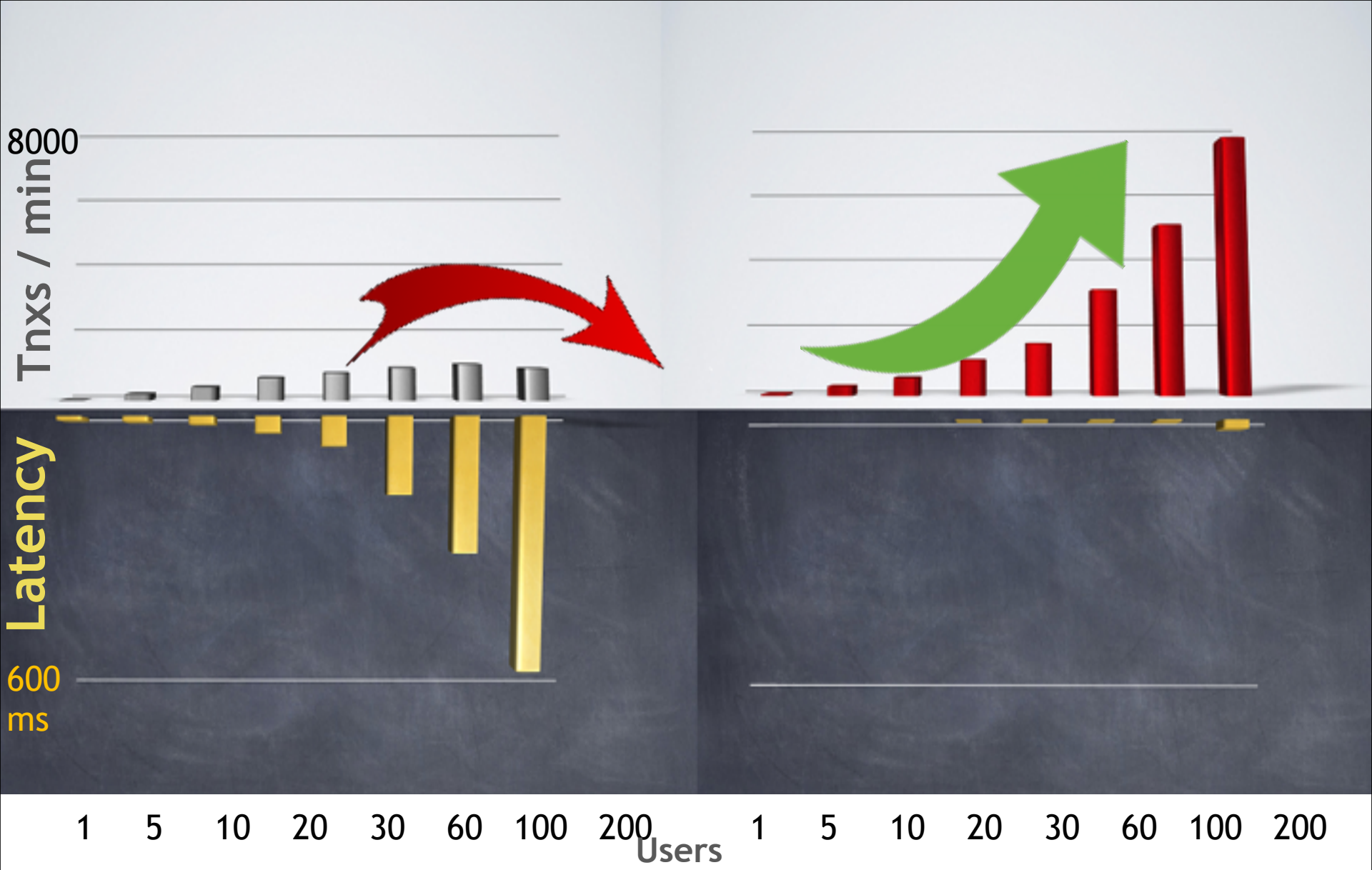
Oracle 12c


5x more Scalable
20GB memory
Memory Utilized
20
15
10
5
0
1
10
20
50
80
110
140
170
200
230
250
Databases
Separate
Pluggable

80MB buffer cache ?


ORACLE
ORACLE
200GB
Cache

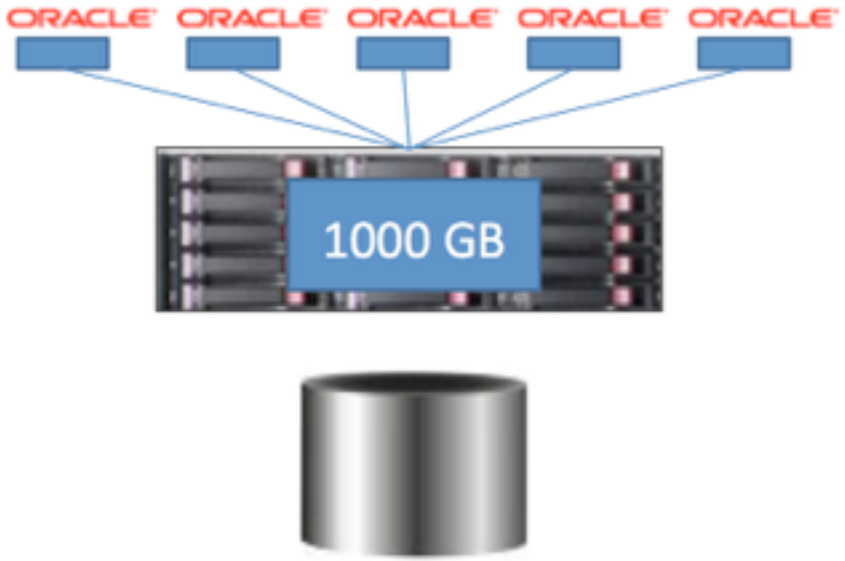

5000
Tnxs / min
Latency
300
ms
1
5
10
20
30
60
100
200
Users


8000
Tnxs / min
Latency
600
ms
1
5
10
20
30
60
100
200
Users
1
5
10
20
30
60
100
200

Five 200GB database copies are cached with :

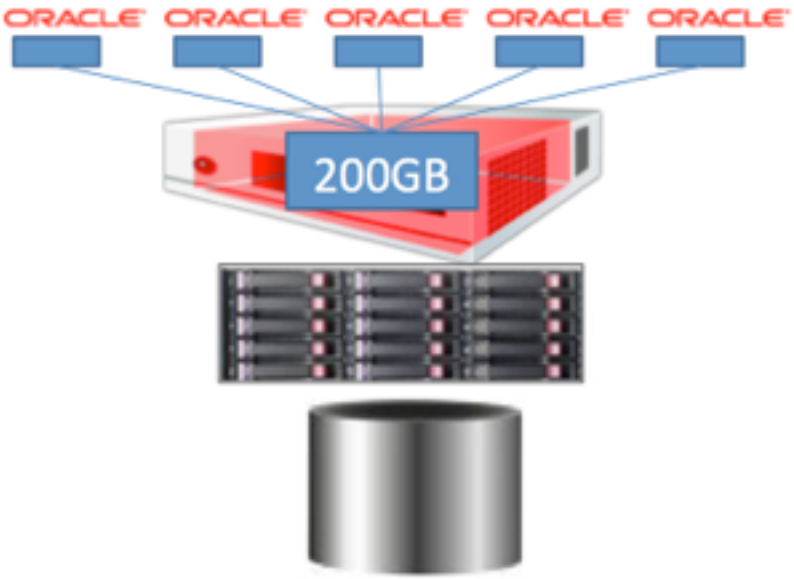
$1,000,000

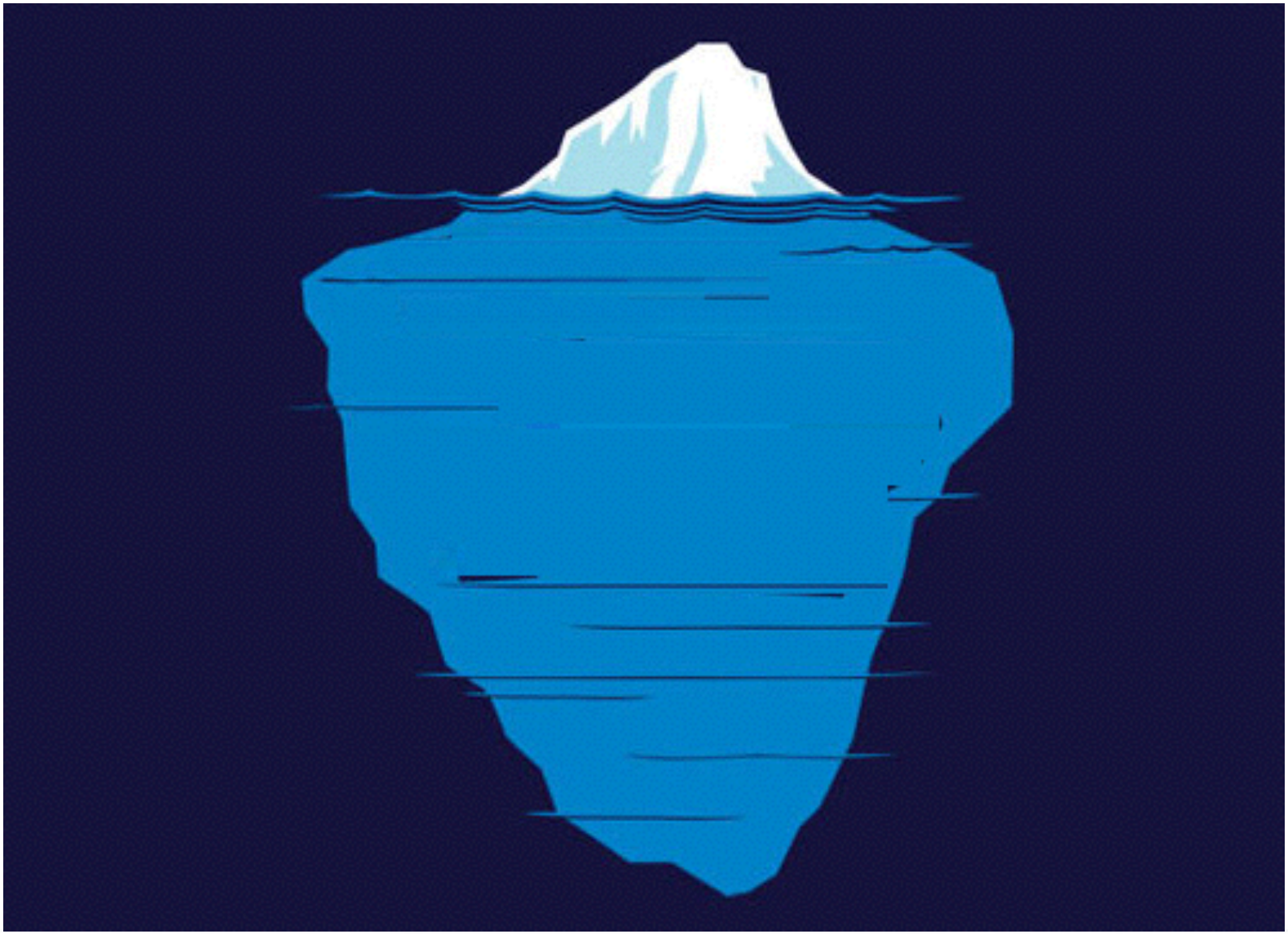
1TB cache on SAN



$6,000

200GB shared cache on Delphix



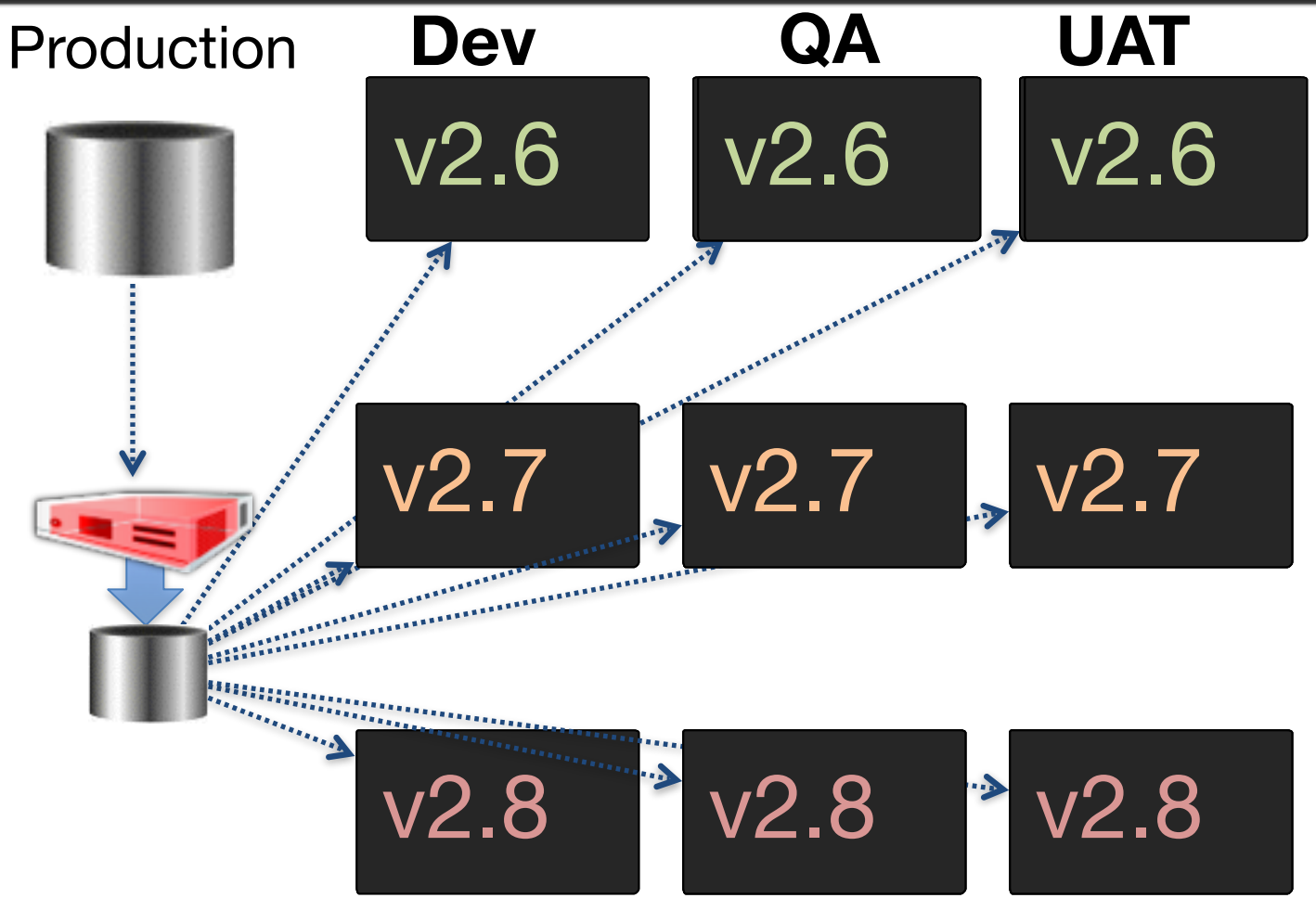


About Delphix

- Founded in 2008, launched in 2010
- CEO Jedidiah Yueh (founder of Avamar: >$1B revenue))
- Based in Silicon Valley, Global Operations

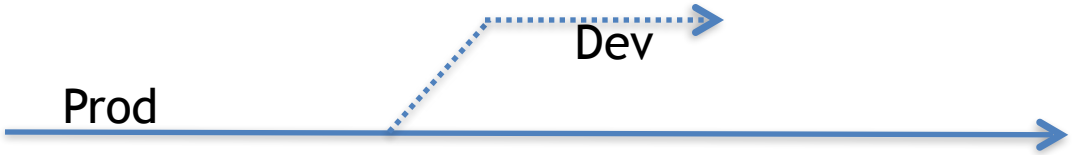
Fortune #1 Walmart
#1 social: Facebook
#1 US Bank: Wells Fargo
#1 Networking : Cisco
#1 cable provider: Comcast
#1 auction site: Ebay
#1 insurance: New York Life
#1 chip manufacture: Intel

Dev
QA
UAT
v2.6
v2.6
v2.6
v2.7
v2.7
v2.7
v2.8
v2.8
v2.8



Source Control for the database data

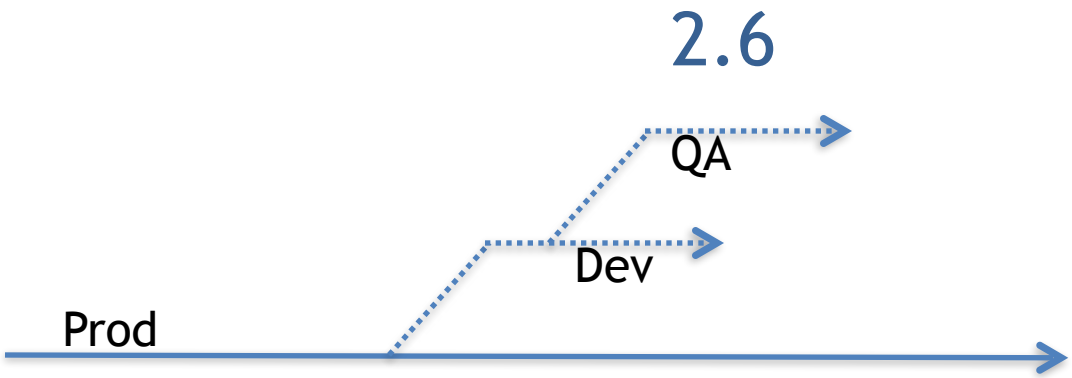
2.6

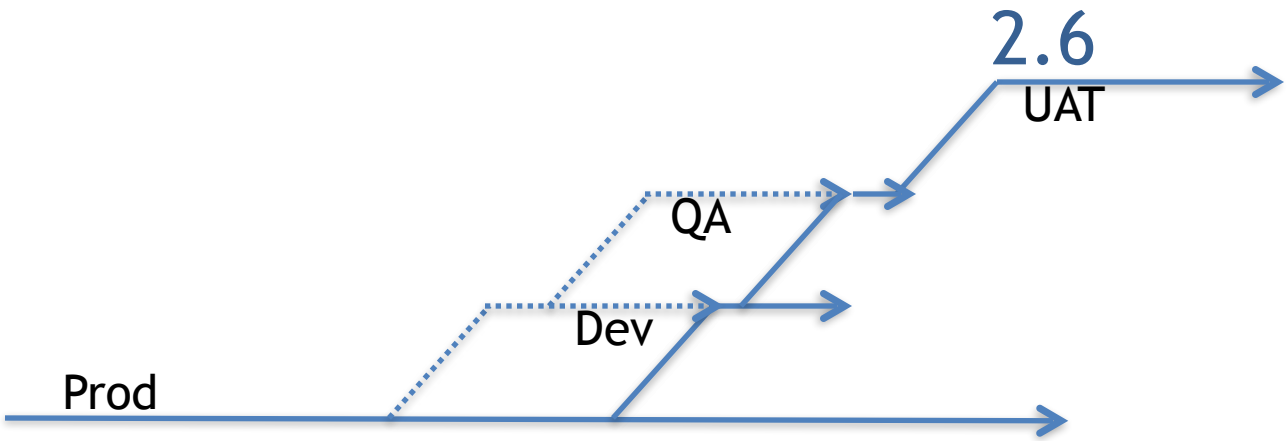


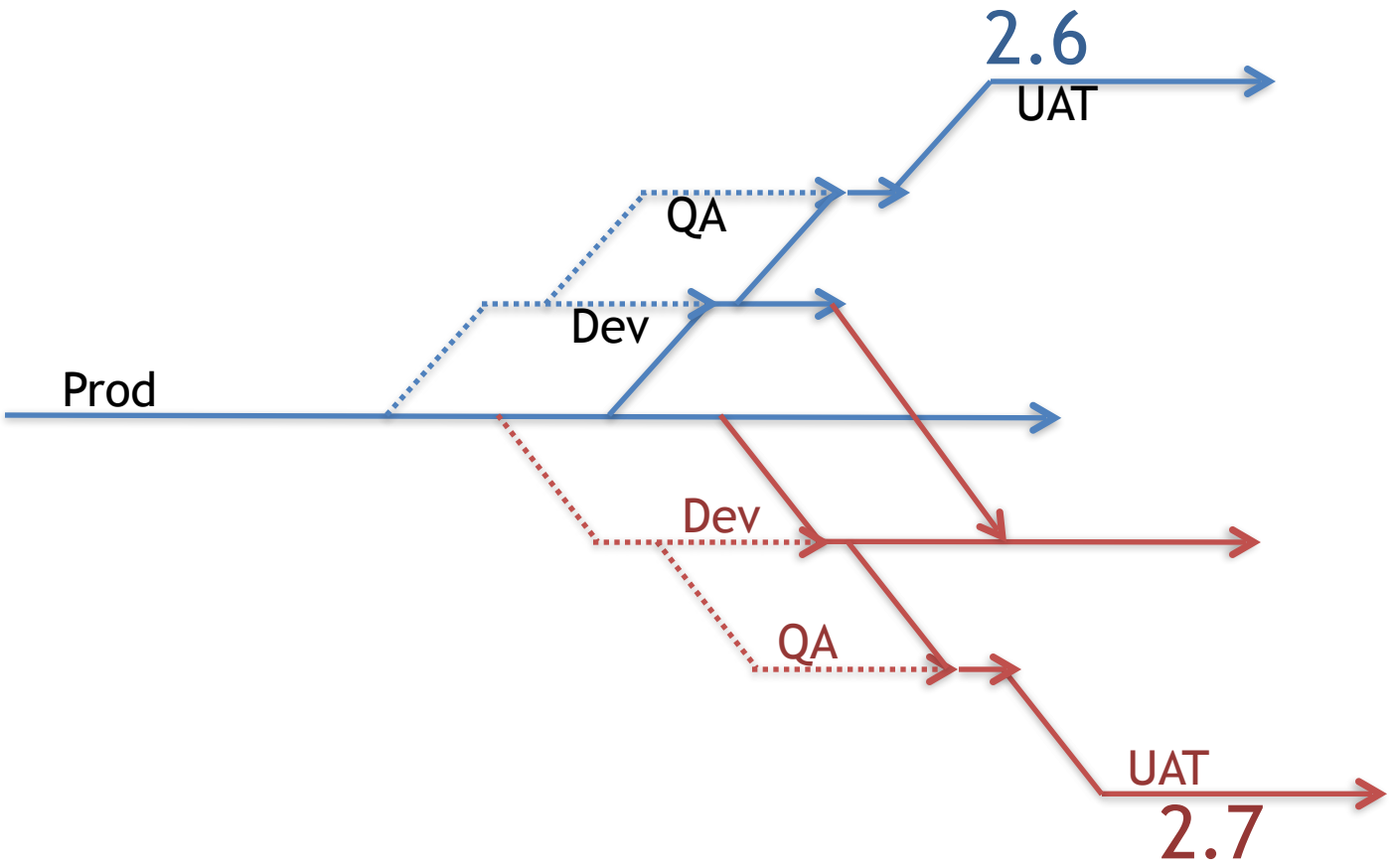
branching

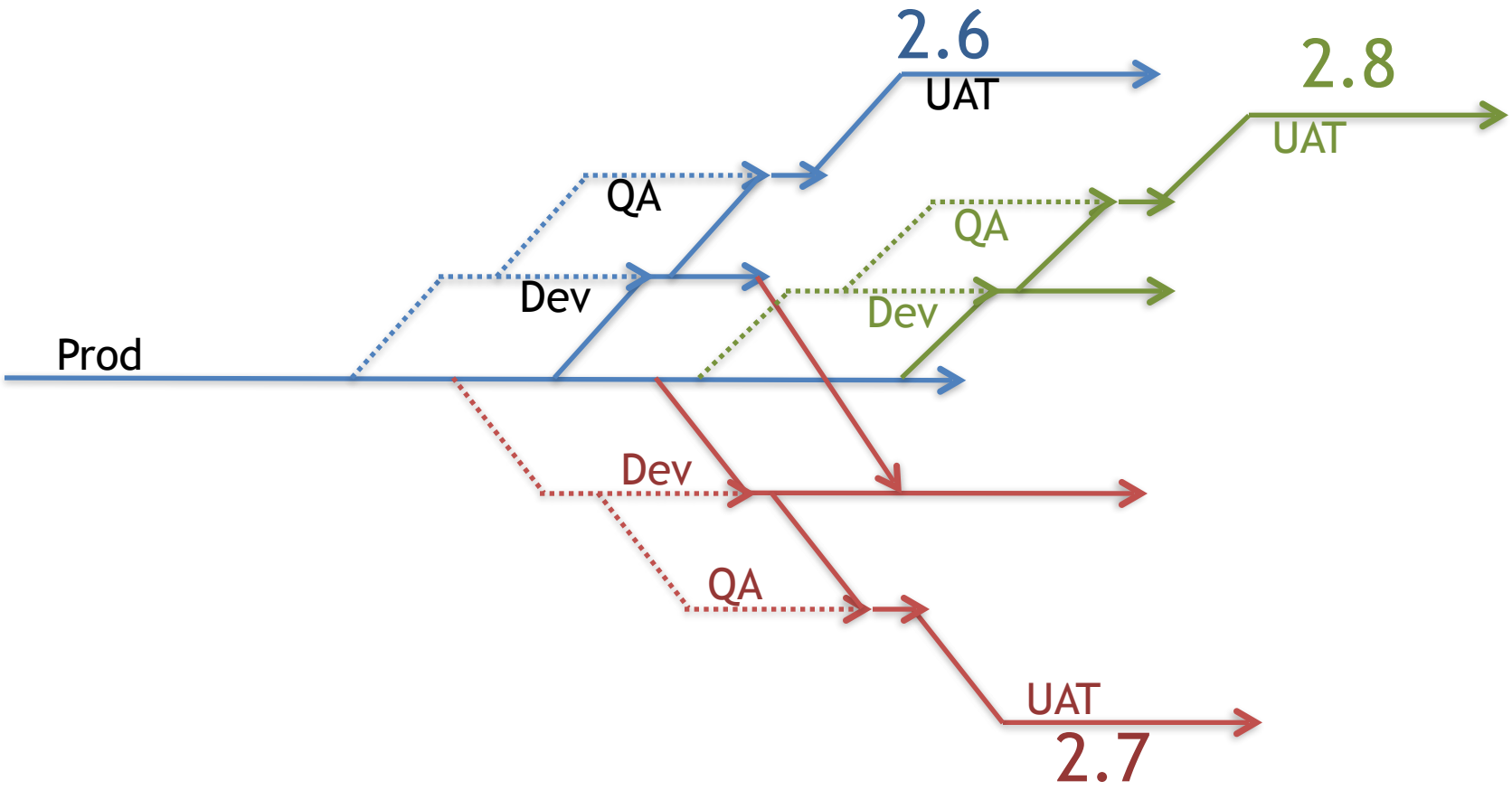
checkout

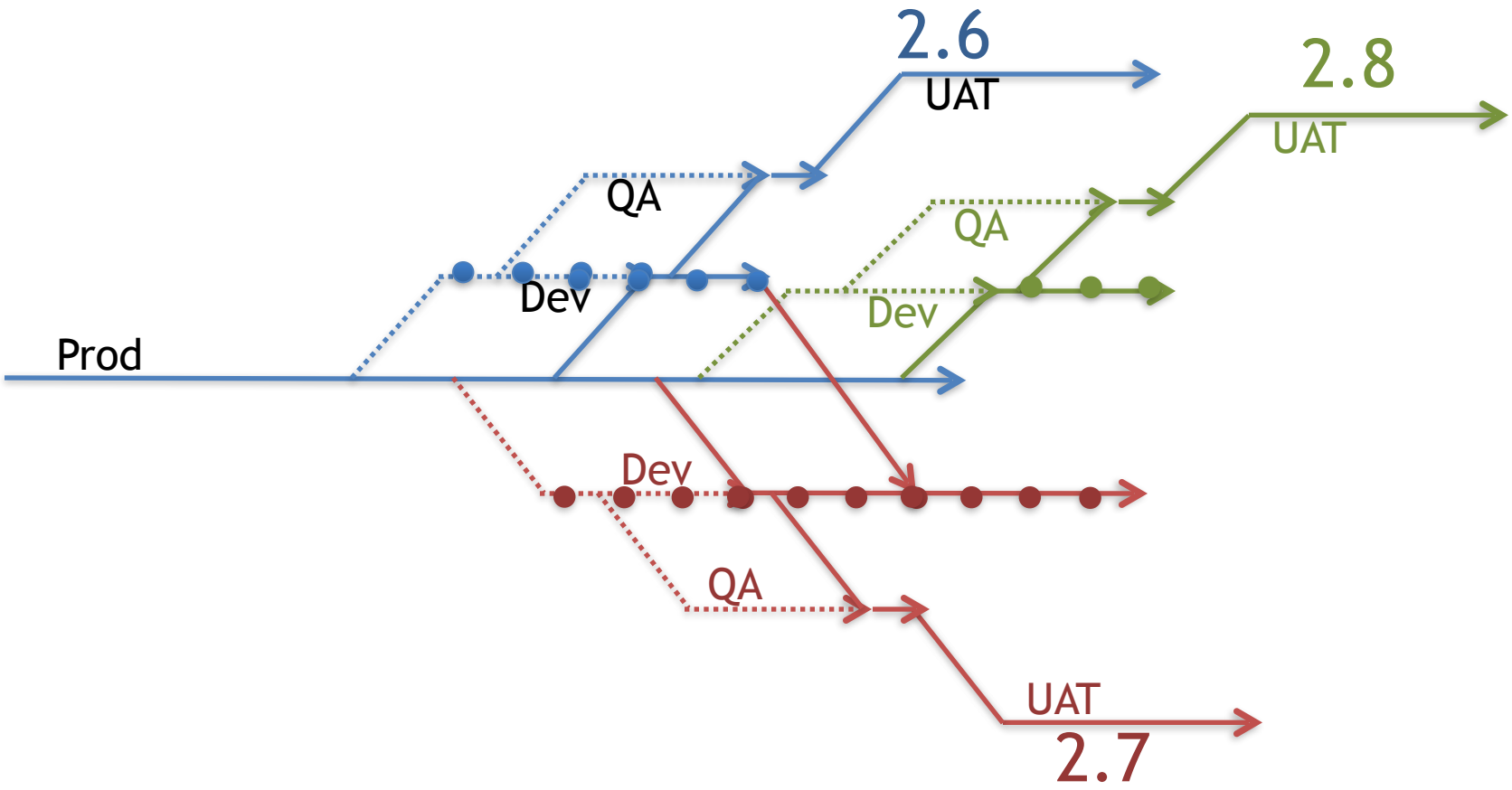
bookmark


2.6
QA
Dev
Prod

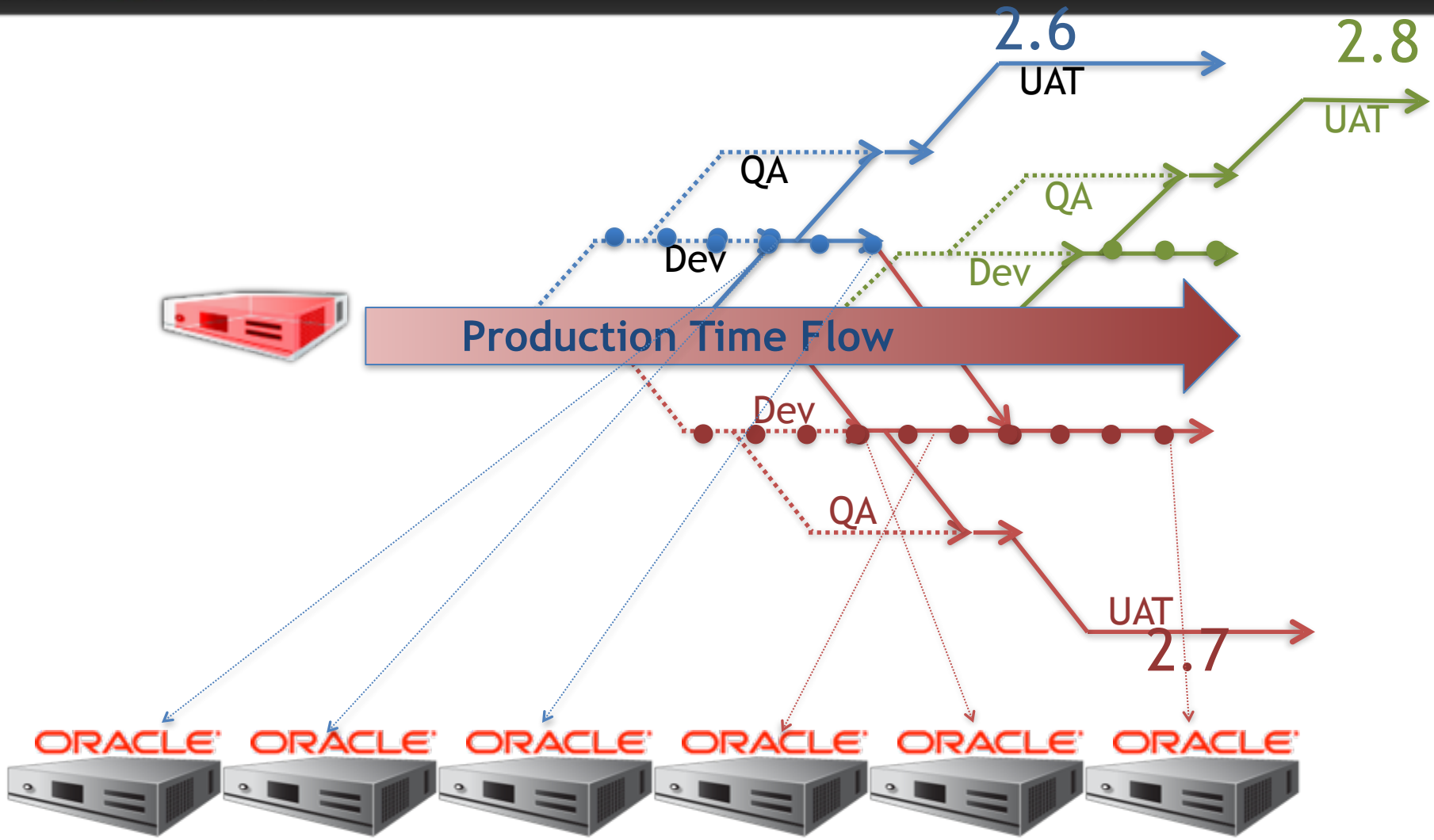

2.6
UAT
QA
Dev
Prod


2.6
UAT
QA
Dev
Prod
Dev
QA
UAT
2.7


2.6
UAT
2.8
UAT
QA
QA
Dev
Dev
Prod
Dev
QA
UAT
2.7

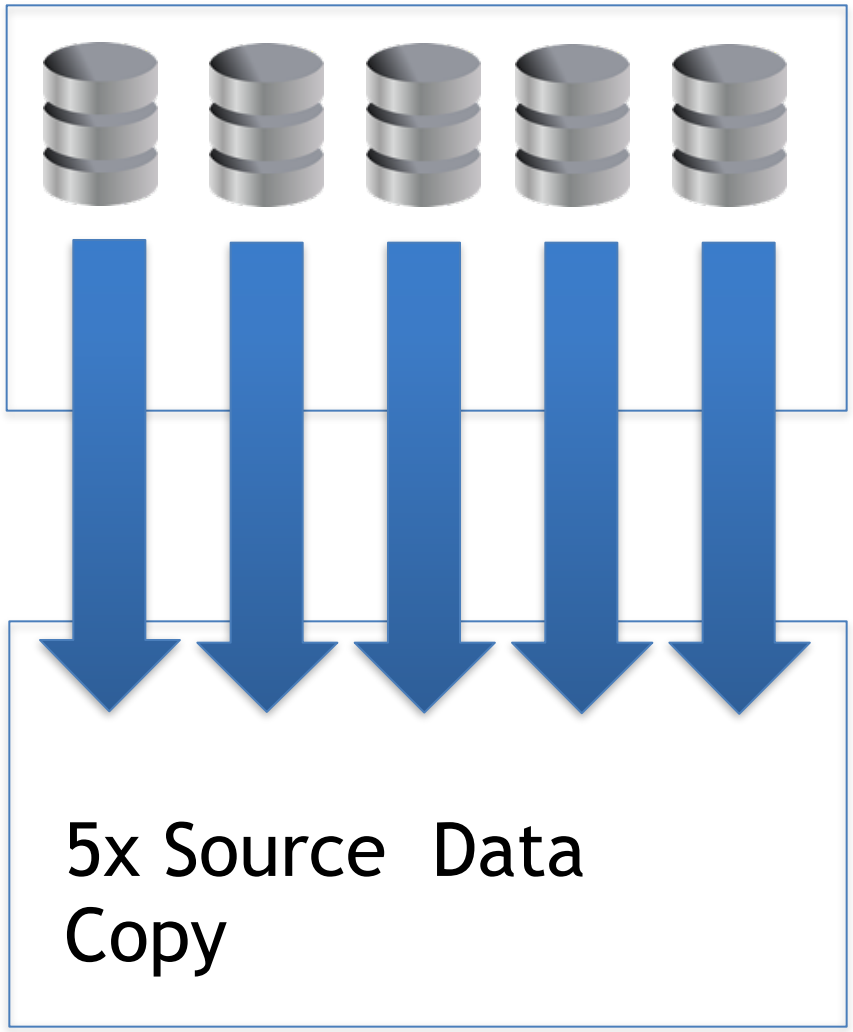


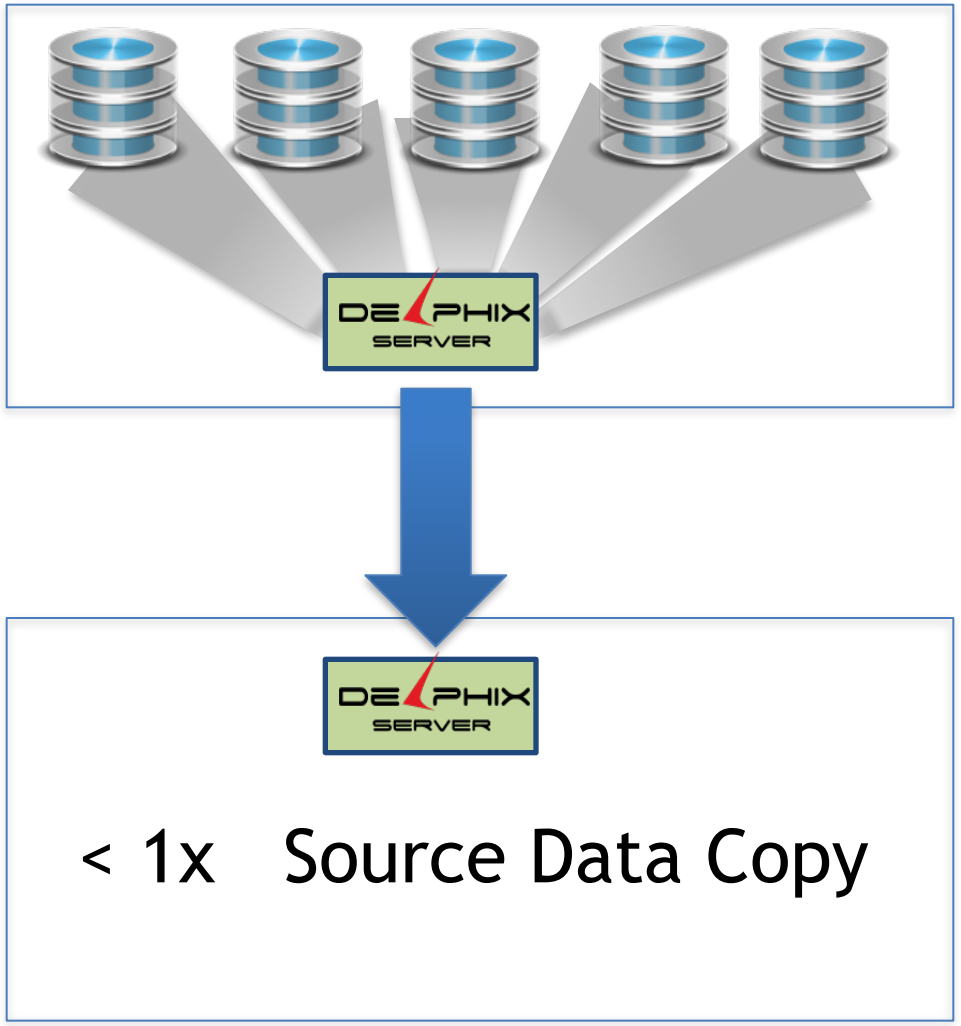
Data Control = Source Control for the Database



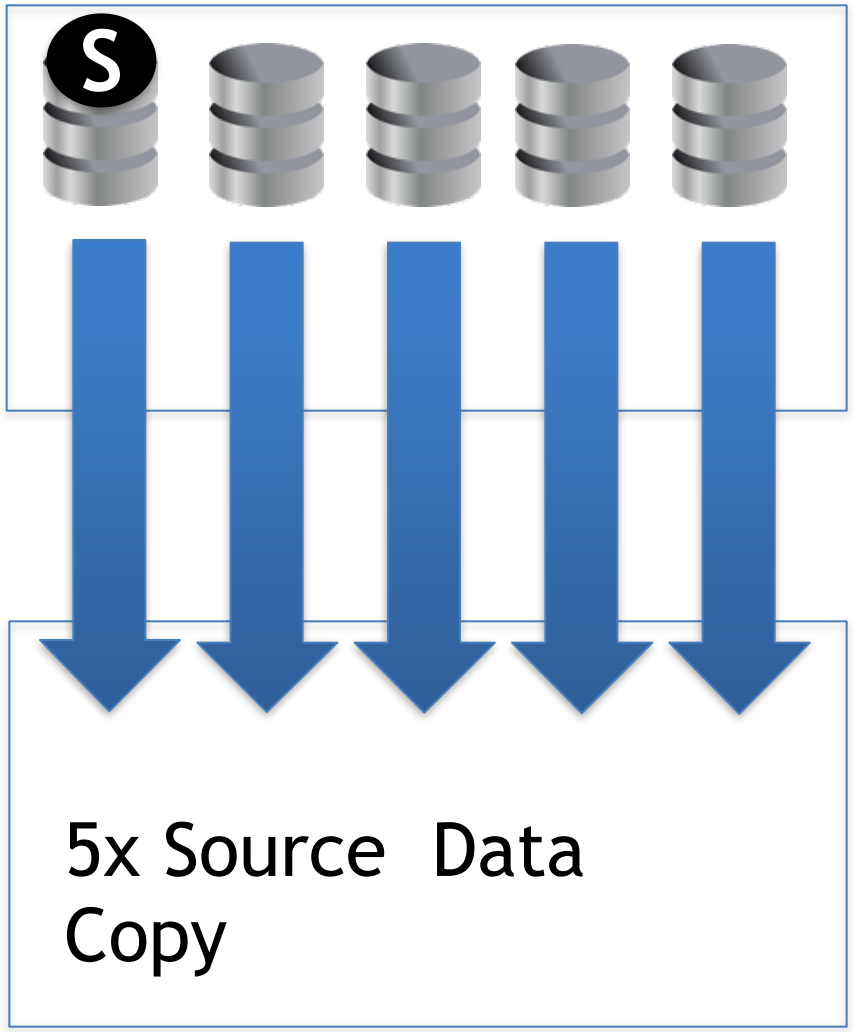
Data Control = Source Control for the Database

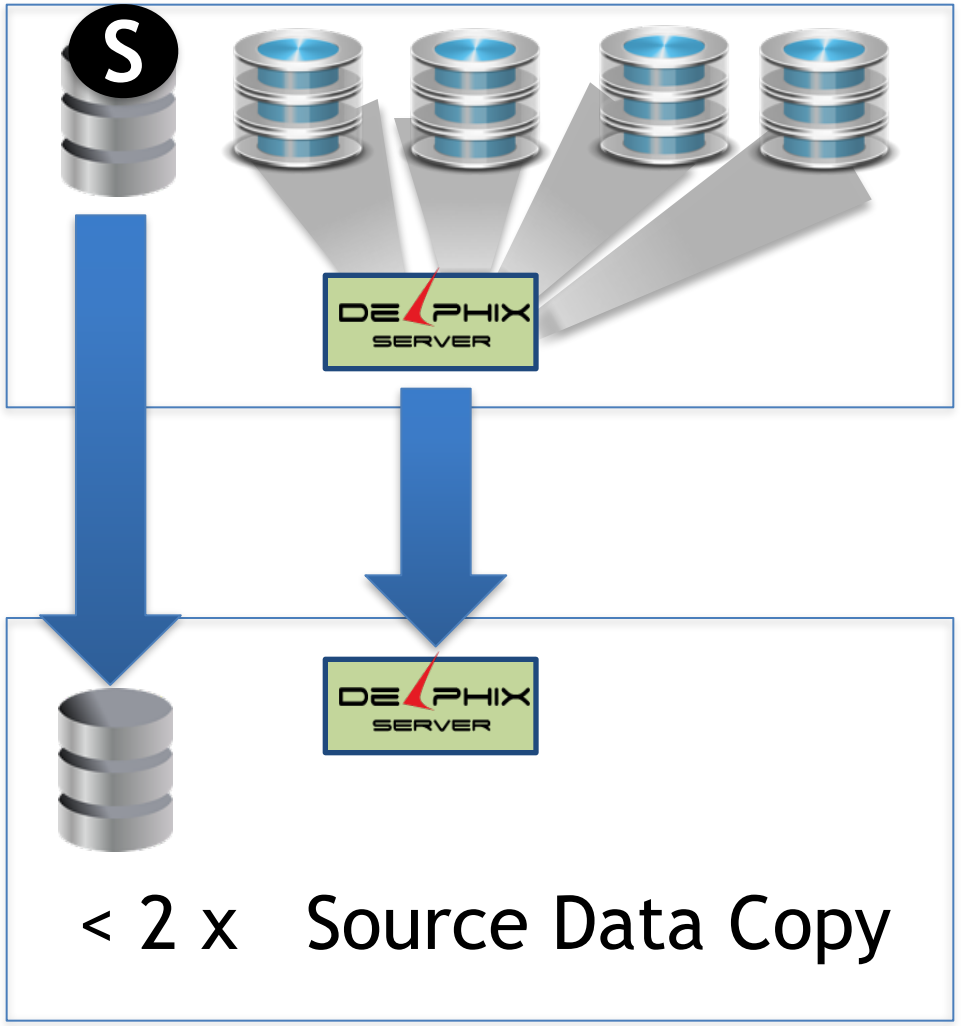
Data Center Migration : clone migration



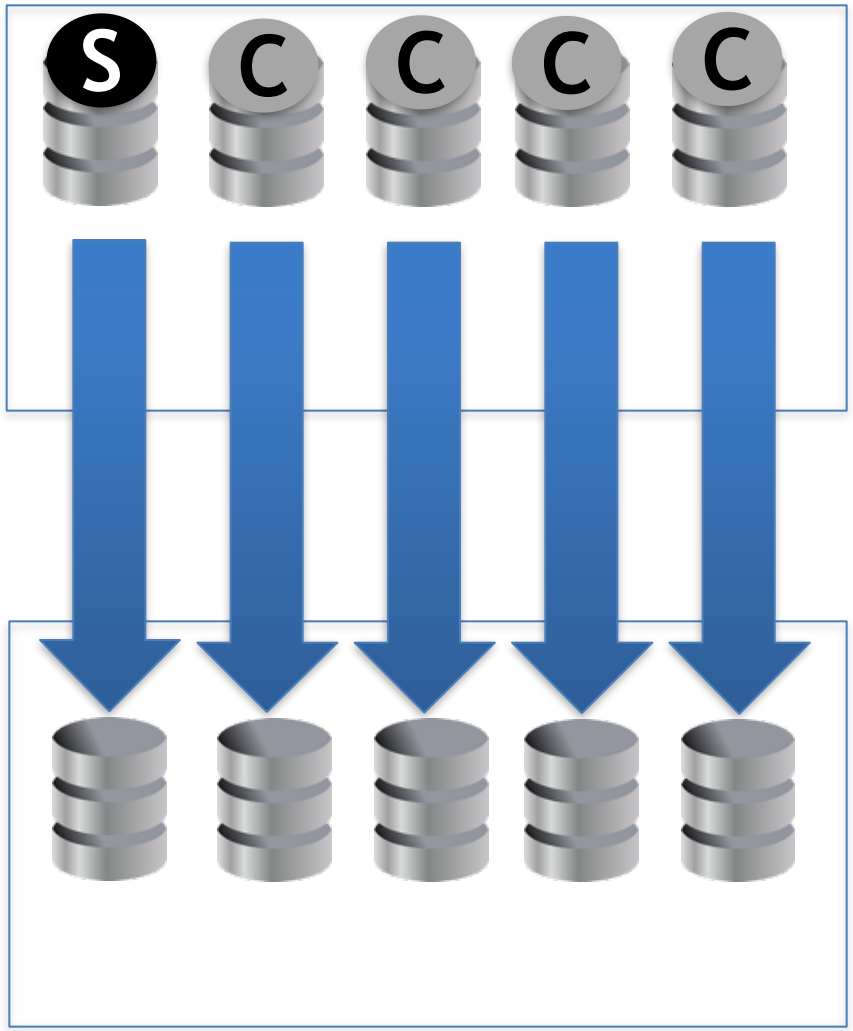


Data Center Migration : clone migration + source

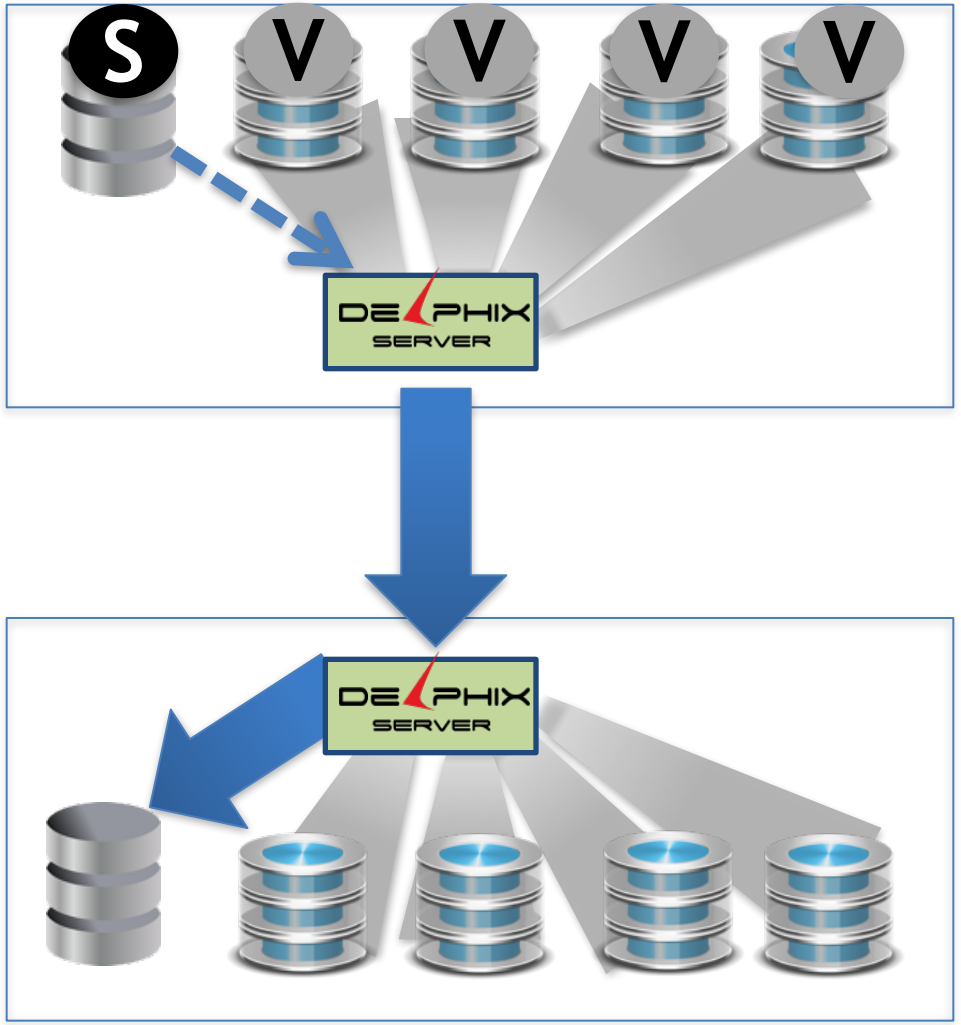




Data Center Migration : clone migration + source

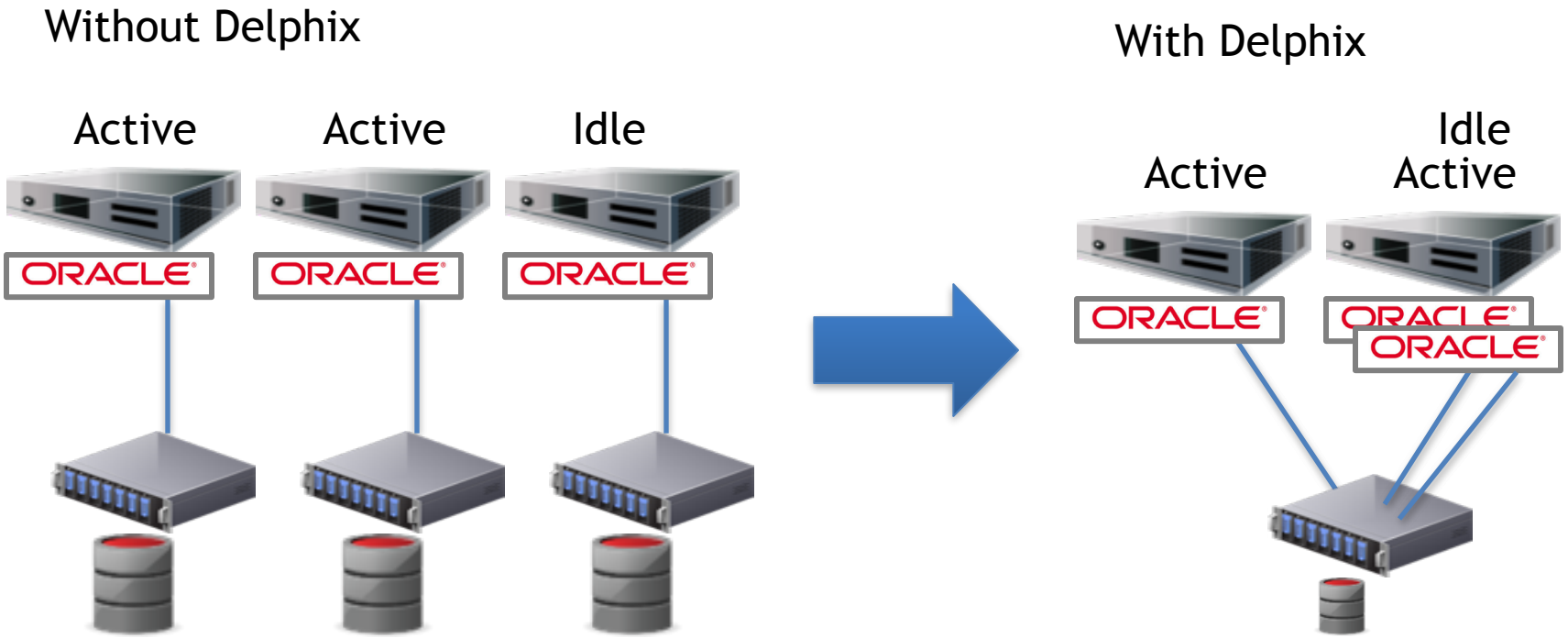


5x Source Data

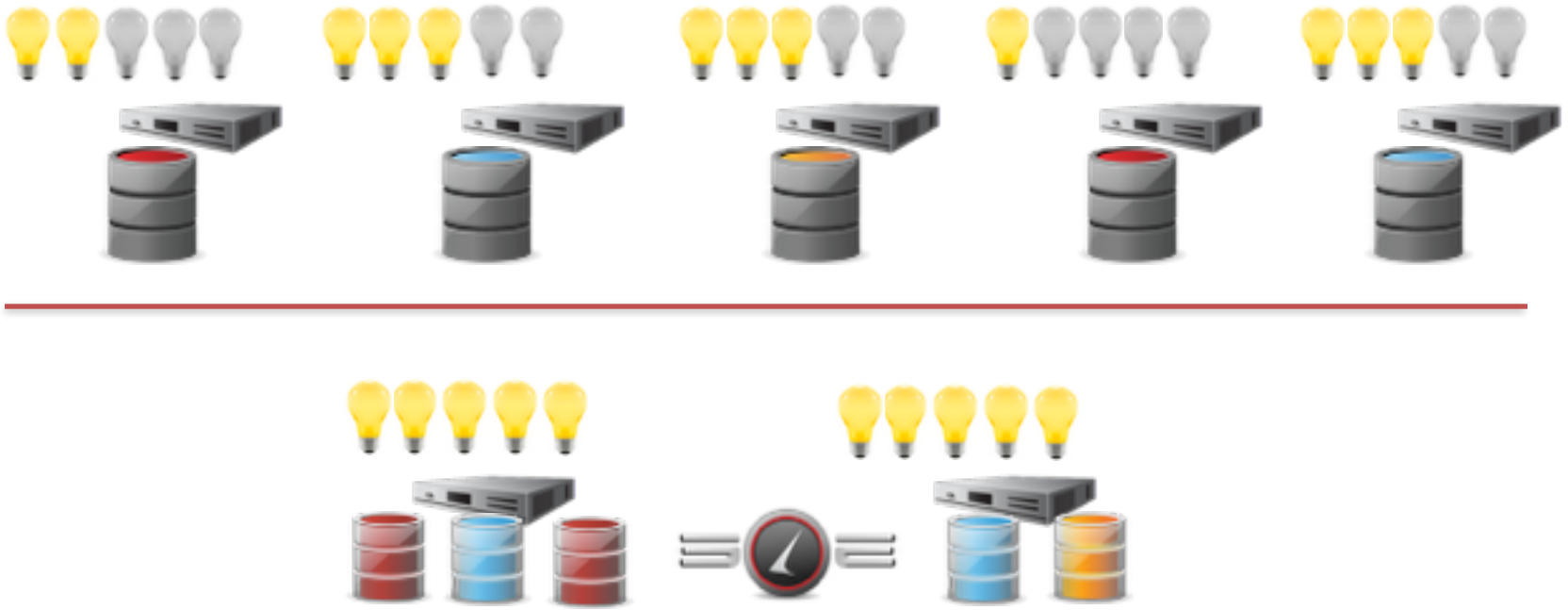


< 1 x Source Data Copy

Consolidation



DBMS License Utilization



- Most DBMS time spent waiting on environment setup, data loads, backups, etc.
- Coupling of data and DB servers forces over-provisioning, limits utilization
- Delphix can attach / detach data to server tier on-demand to drive consolidation
- Potential 2X reduction in DB server footprint cuts HW and DBMS license costs

© 2014 DELPHIX. ALL RIGHTS RESERVED. PRIVATE & CONFIDENTIAL.