

Agile Data : Virtual Data Revolution



Kyle@delphix.com

kylehailey.com

slideshare.com/khailey



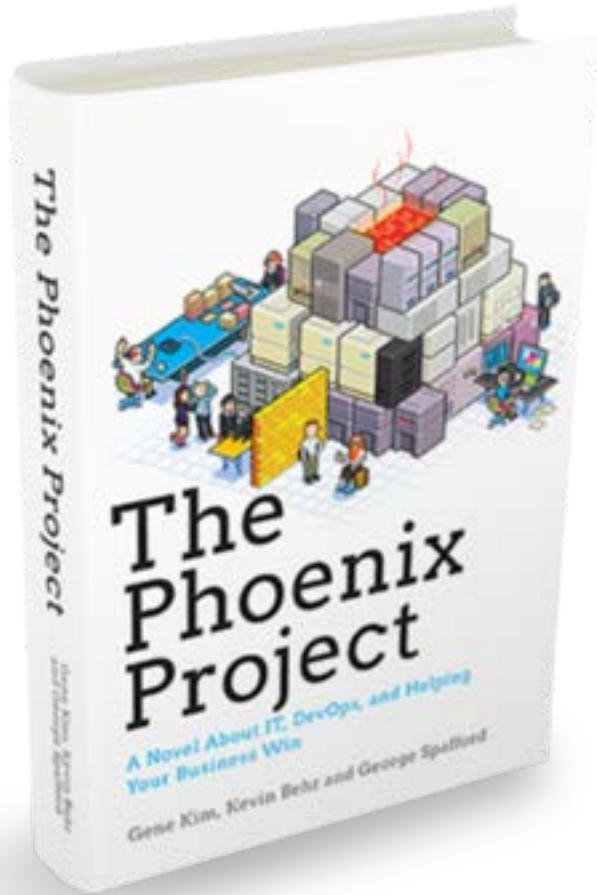
In this presentation :

- Problem in IT
- Solution
- Use Cases

In this presentation :

- Problem in IT
- Solution
- Use Cases

The Phoenix Project

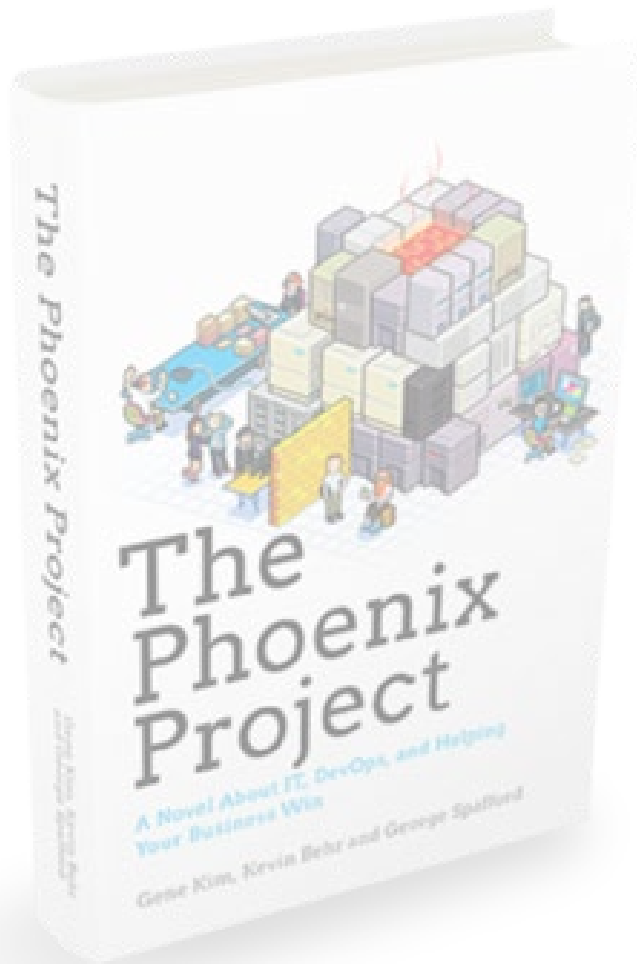


- Bottlenecks
- Metrics
- Priorities
- Goals
- Iterations

“The Goal”

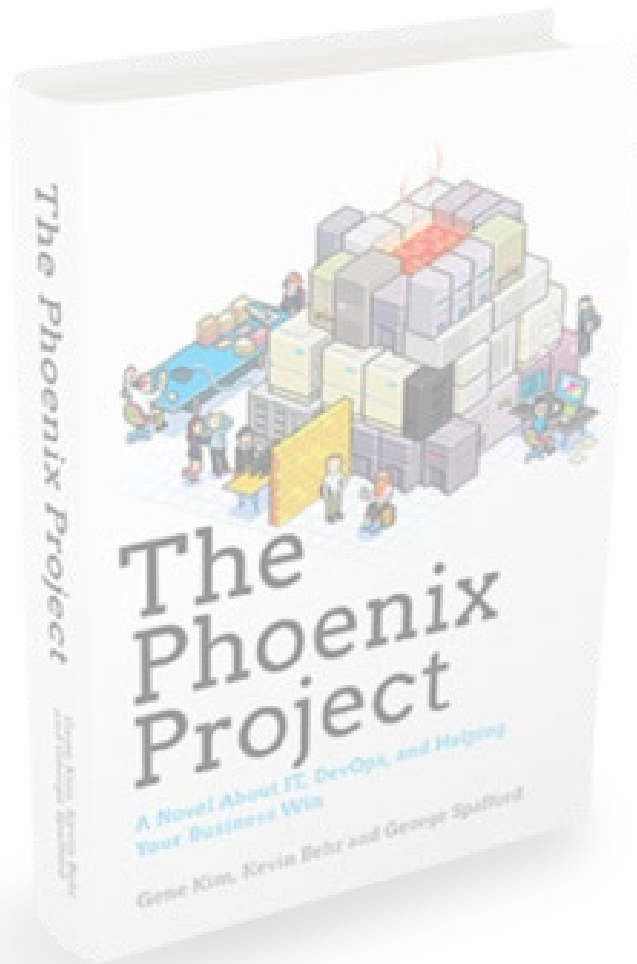
by E. Goldratt

The Phoenix Project



“Any improvement not made at the **constraint** is an illusion.”

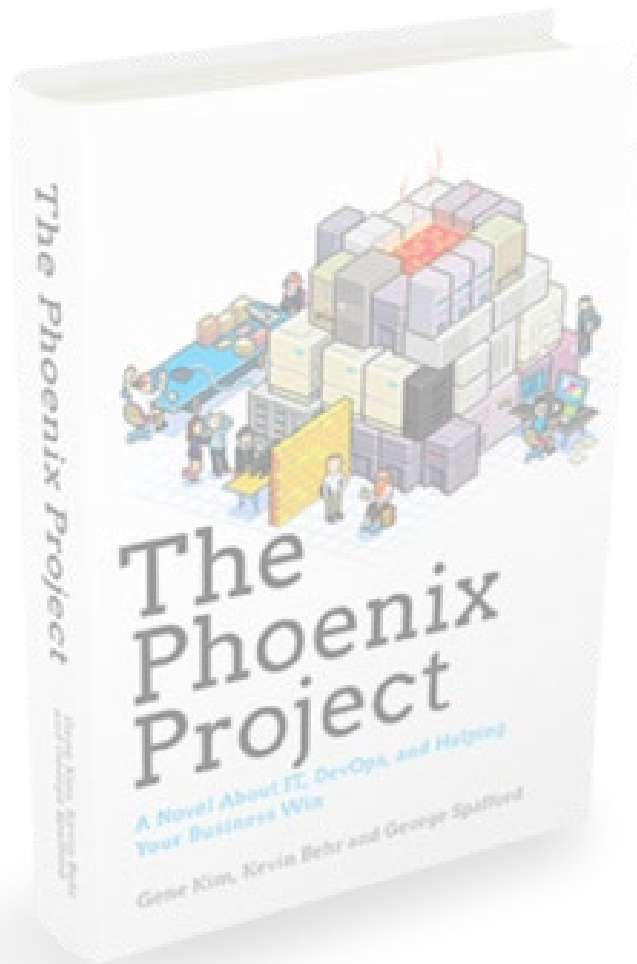
The Phoenix Project



“Any improvement not made at the **constraint** is an illusion.”

What is the **constraint**?

The Phoenix Project



“Any improvement not made at the **constraint** is an illusion.”

What is the **constraint**?

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

Problem in IT

- I. **Data Constraint** strains IT
- II. **Data Constraint** price is huge
- III. **Data Constraint** companies unaware

Problem in IT

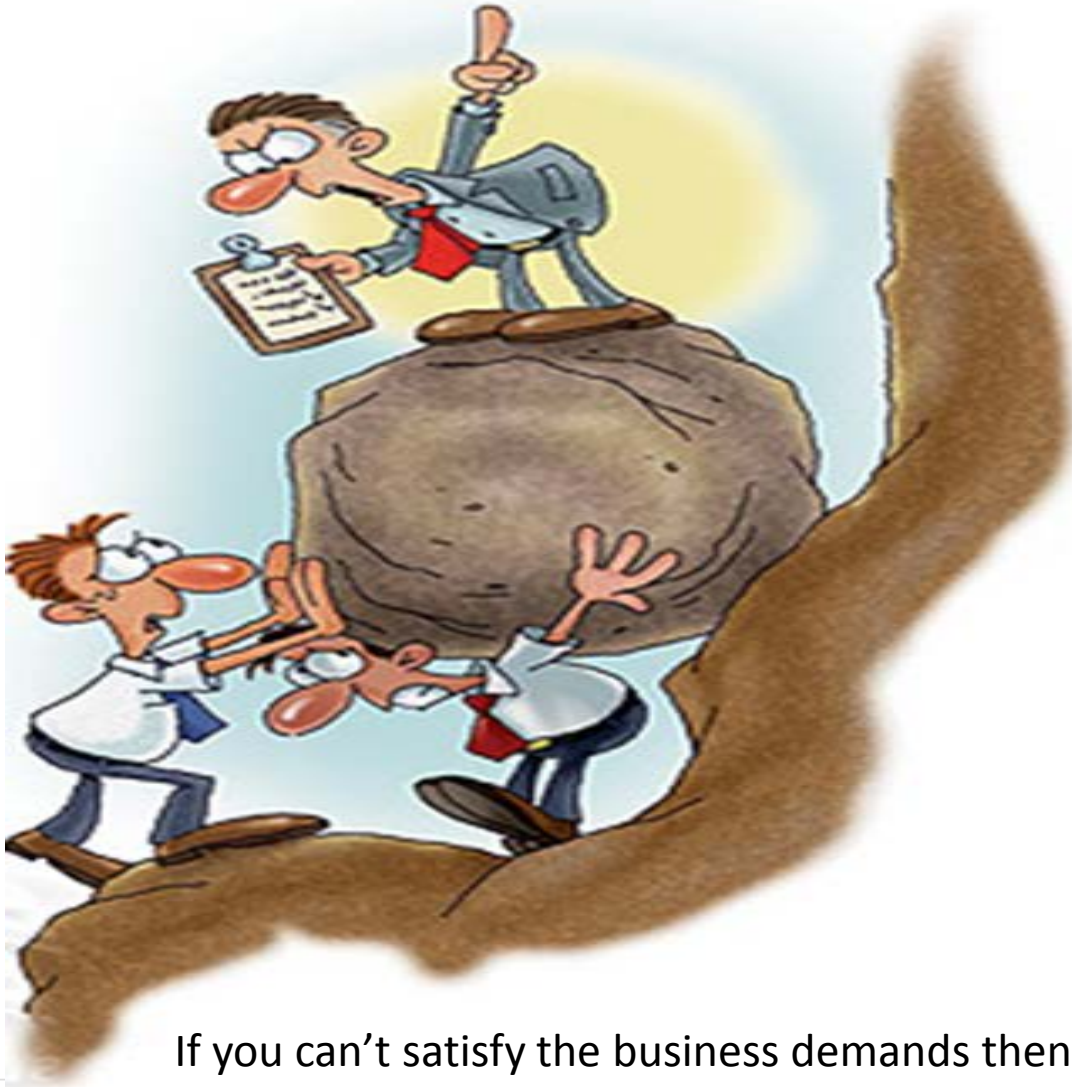
CIO Magazine Survey:

60% Projects Over Schedule

85% delayed waiting for data

Data is the **Constraint**

I. Data Constraint strains IT



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

II. Data Constraint price is huge



III. Data Constraint : companies unaware



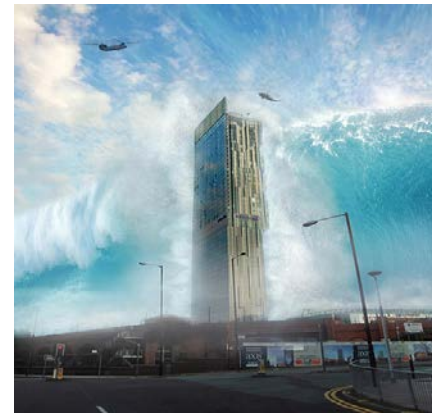
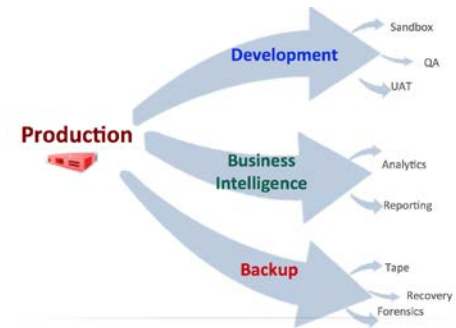
Data is the constraint

- I. **Data Constraint** strains IT
- II. Data Constraint price is huge
- III. Data Constraint companies unaware



I. Data Constraint companies unaware

- Moving data is hard
- Triple tax
- Data Floods infrastructure





I. Data Constraint : moving data is hard

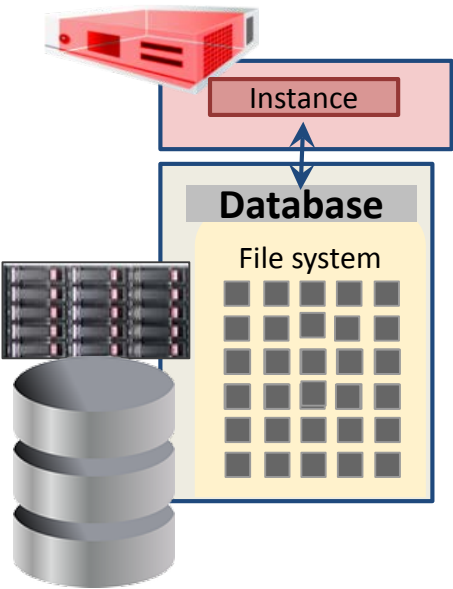
- Storage & Systems
- Personnel
- Time





Typical Architecture

Production





Typical Architecture

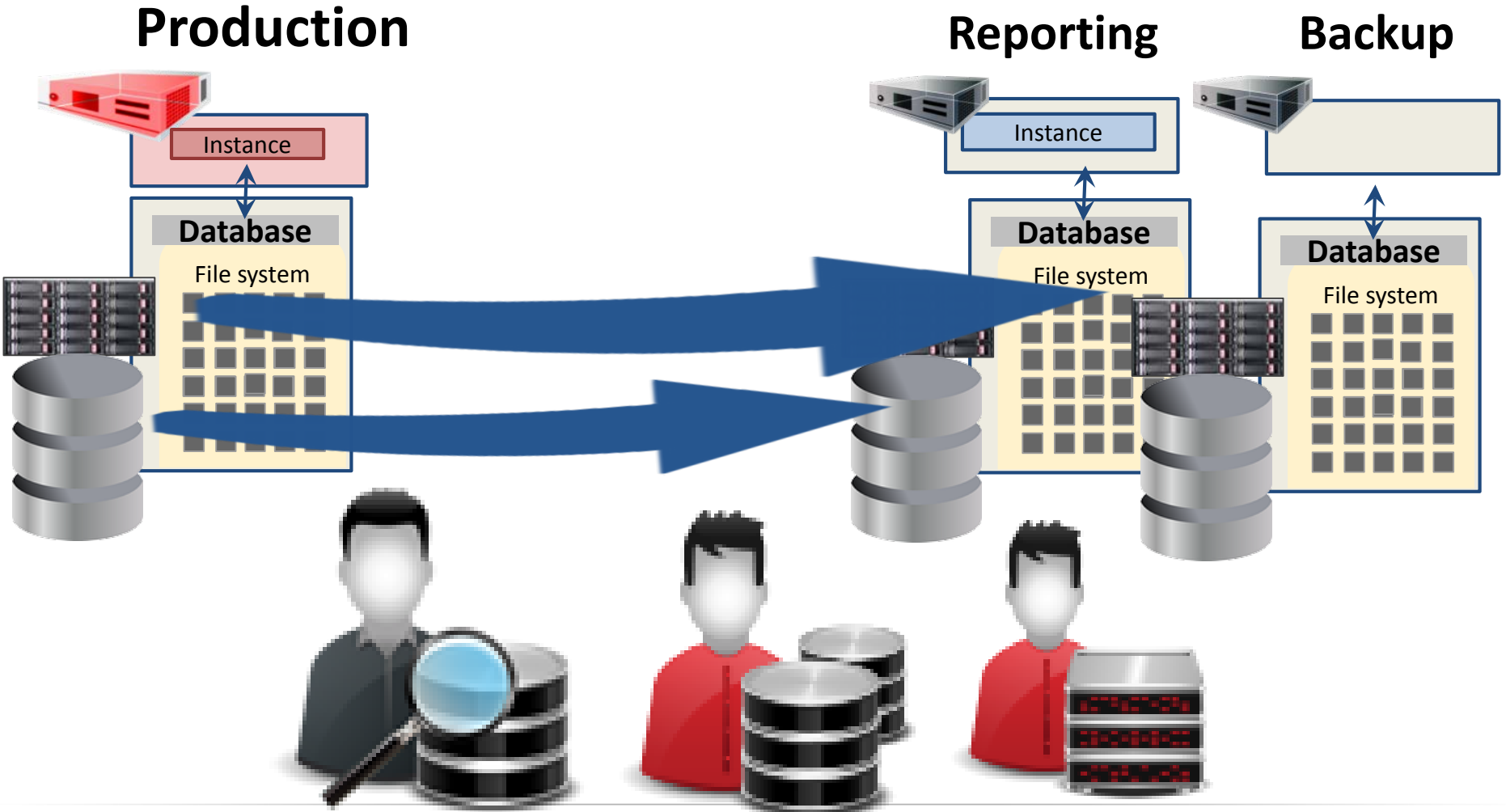
Production

Backup





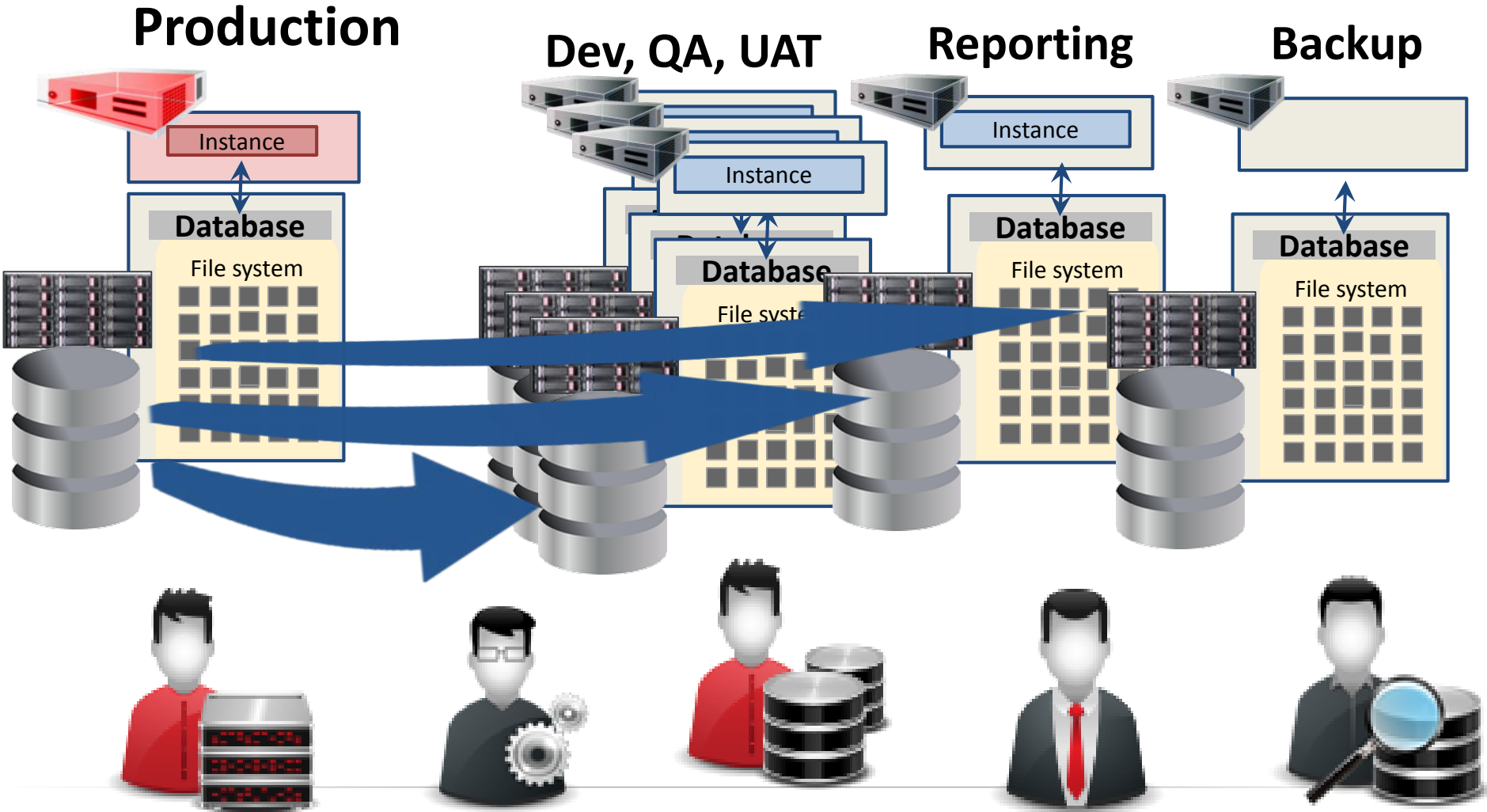
Typical Architecture





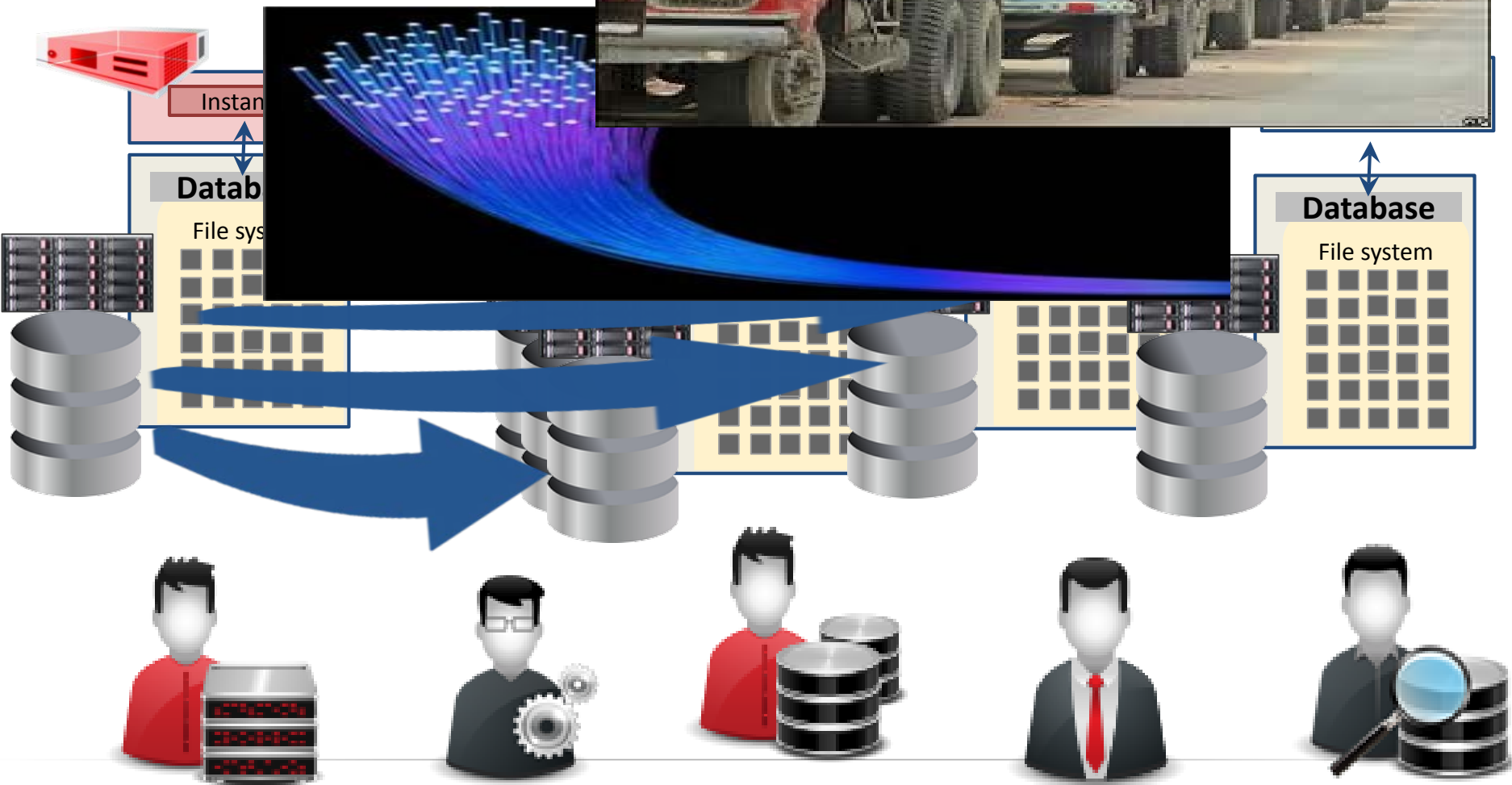
Typical Architecture

Triple Tax



Typical Architecture

Production





I. Data constraint: Data floods company infrastructure

92% of the cost of business

, in 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
Part of IT expense



Data is the constraint

- I. Data Constraint strains IT
- II. **Data Constraint** price is huge
- III. Data Constraint companies unaware



Part II. Data constraint price is Huge





Part II. Data constraint price is Huge

- Four Areas data tax hits
 1. IT Capital resources
 2. IT Operations personnel
 3. Application Development
 4. Business



Part II. Data constraint price is Huge

- Four Areas data tax hits
 1. IT Capital resources
 2. IT Operations personnel
 3. Application Development
 4. Business



II. Data constraint price is huge : 1. IT Capital

- Hardware

- Servers



- Storage



- Network



- Data center floor space, power, cooling





Part II. Data constraint price is Huge

- Four Areas data tax hits
 1. IT Capital resources
 2. IT Operations personnel
 3. Application Development
 4. Business



II. Data constraint price is huge : 2. IT Operations

- People

- DBAs
- SYS Admin
- Storage Admin
- Backup Admin
- Network Admin



- Hours : 1000s just for DBAs

- \$100s Millions for data center modernizations





Part II. Data constraint price is Huge

- Four Areas data tax hits
 1. IT Capital resources
 2. IT Operations personnel
 3. Application Development
 4. Business



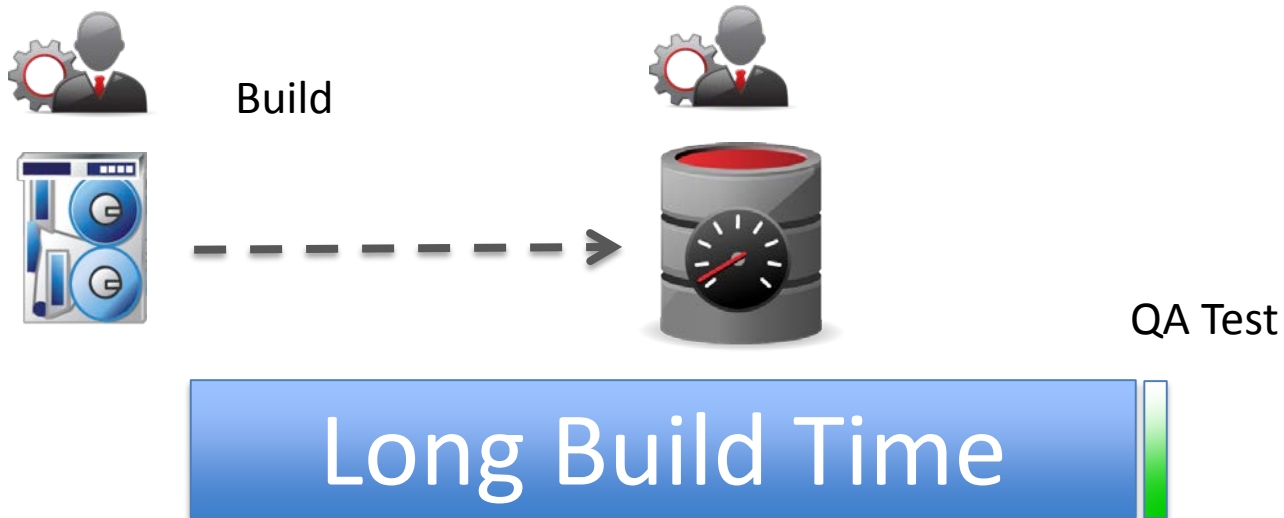
II. Data constraint price is Huge : 3. App Dev

- 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”



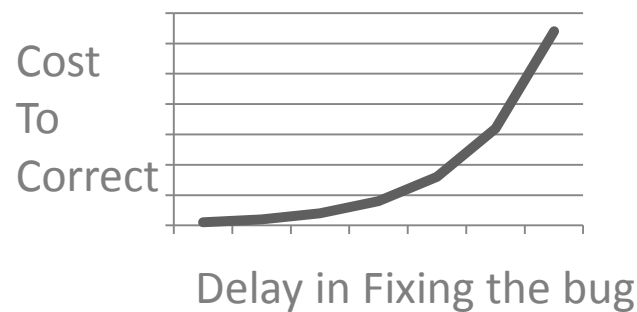
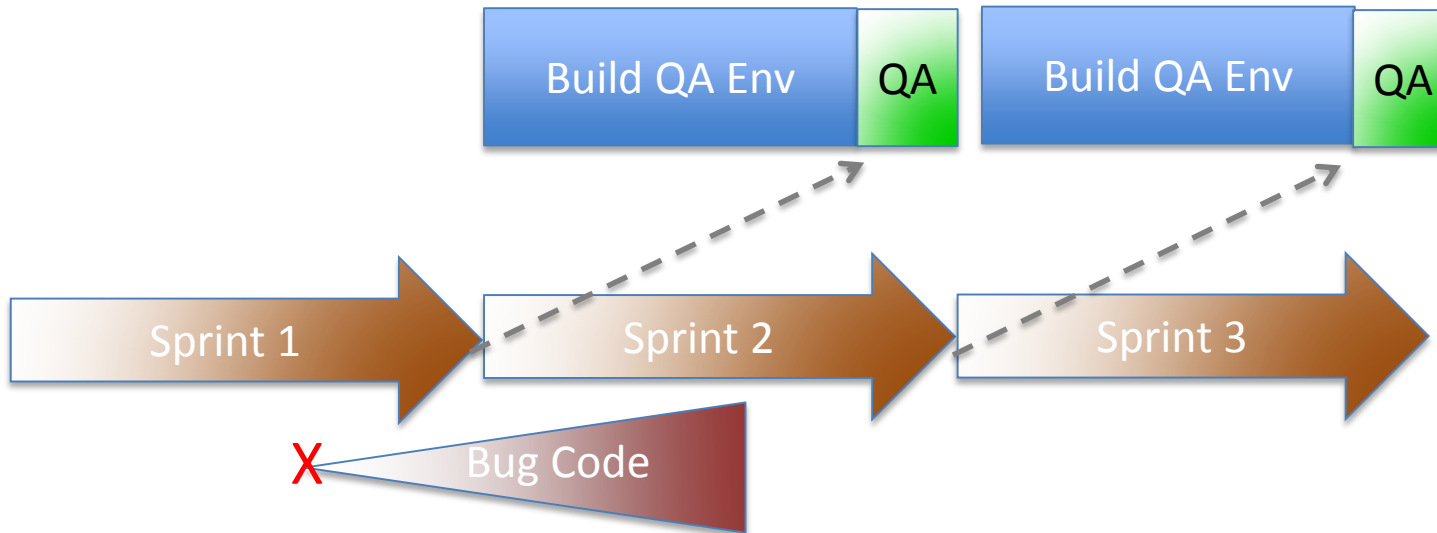
II. Data Tax is Huge : 3. App Dev



96% of QA time was building environment
\$.04/\$1.00 actual testing vs. setup

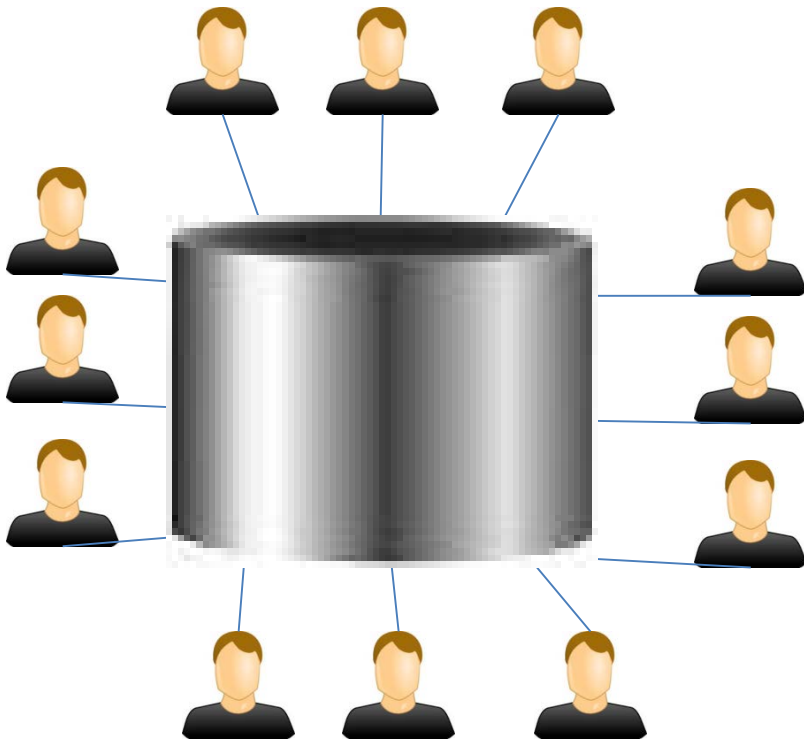


II. Data Tax is Huge : 3. App Dev





II. Data Tax is Huge : 3. App Dev full copies cause bottlenecks



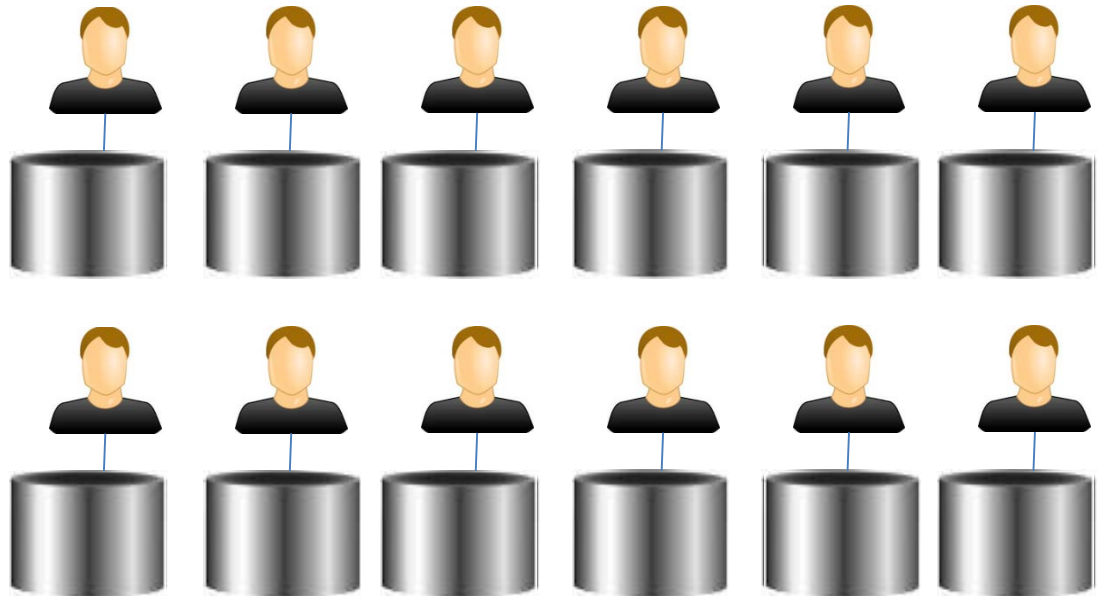
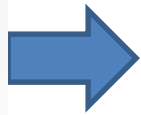
Old Unrepresentative Data



Frustration Waiting



II. Data Tax is Huge : 3. App Dev subsets cause bugs





II. Data Tax is Huge : 3. App Dev subsets cause bugs

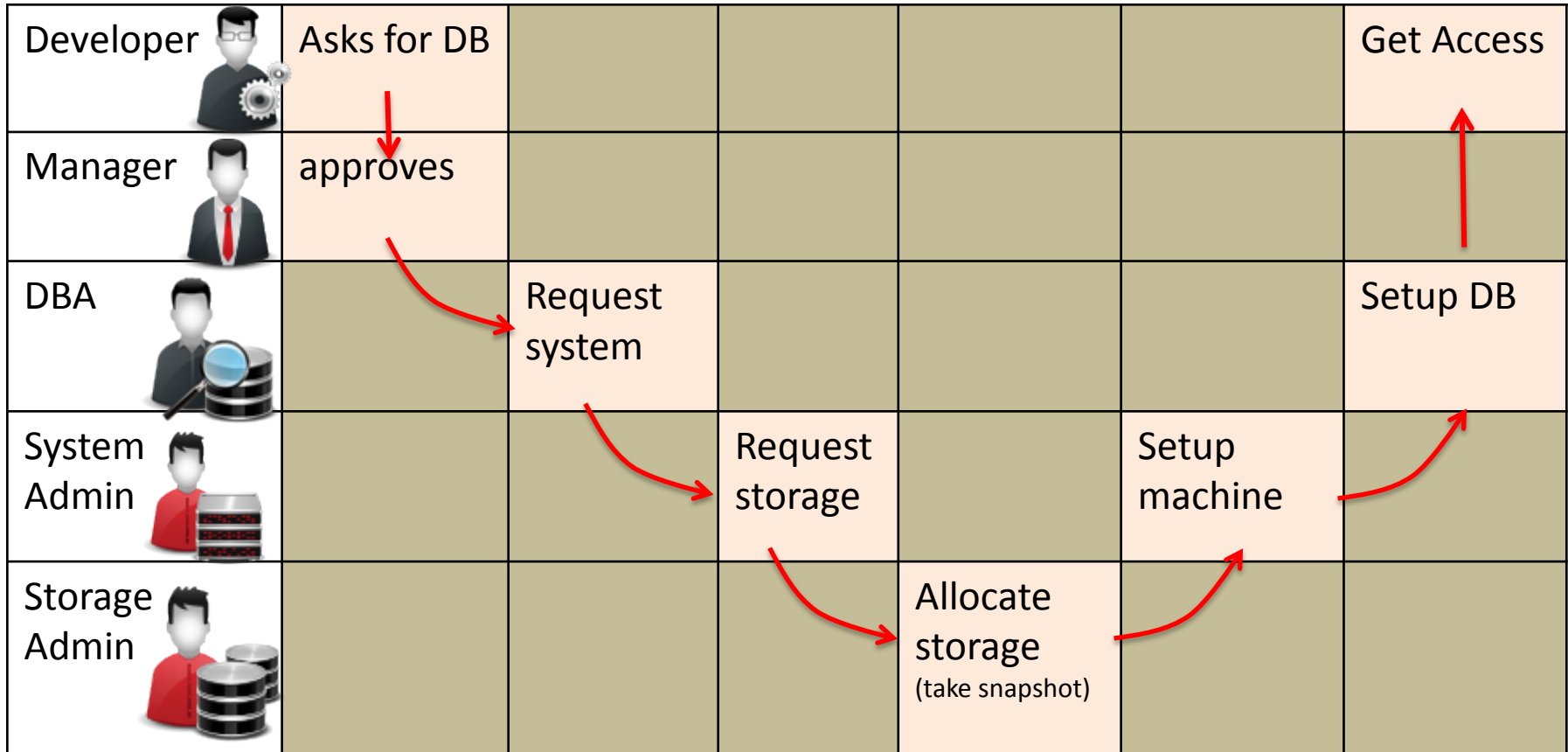
The Production 'Wall'





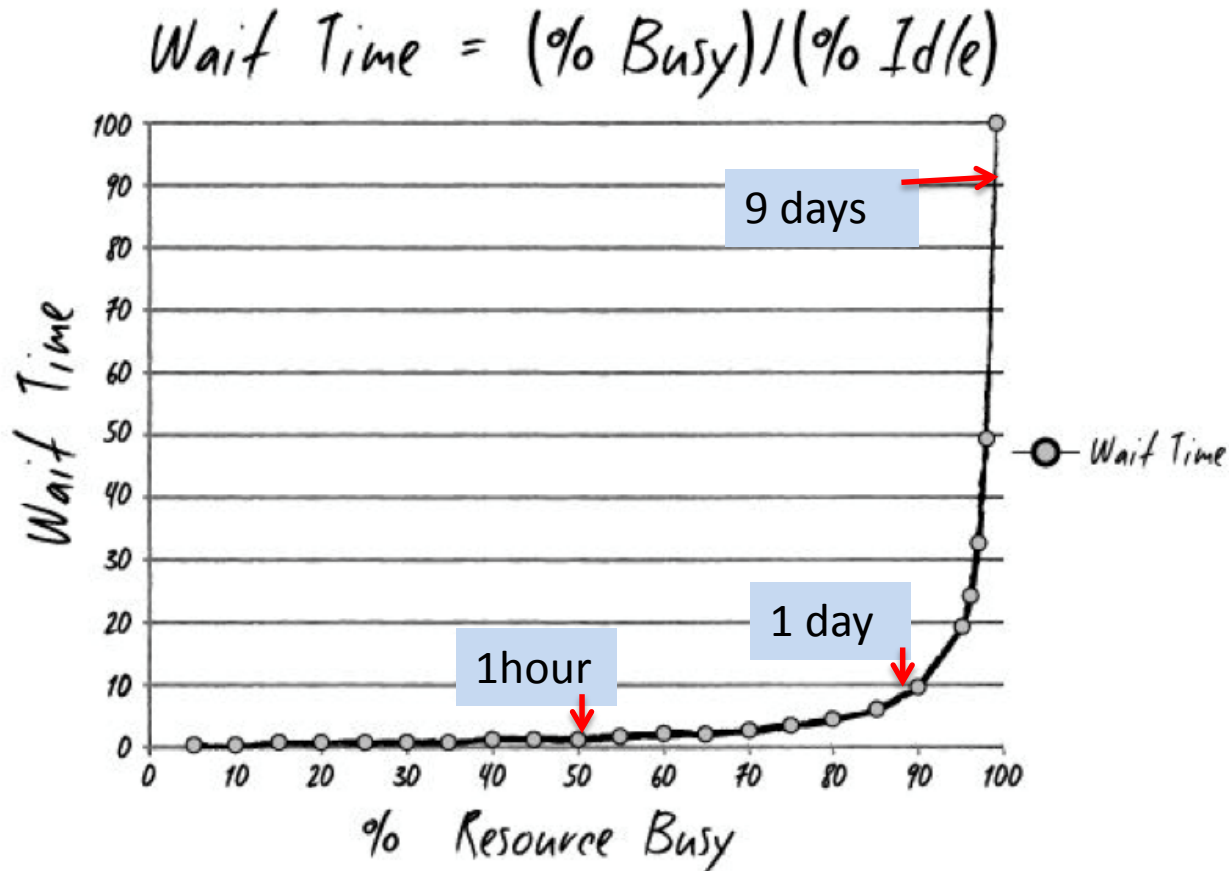
II. Data Tax is Huge : 3. App Dev

3-6 Months to Deliver Data



II. Data Tax is Huge : 3. App Dev

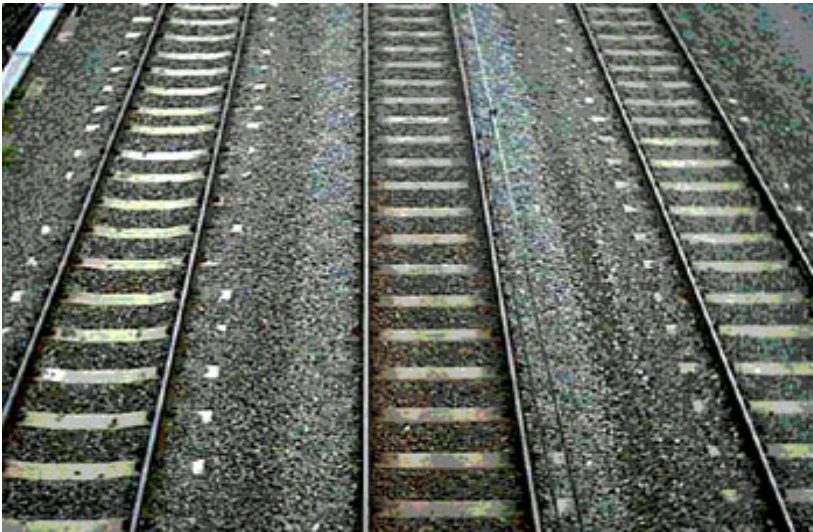
Why are hand offs so expensive?





II. Data Tax is Huge : 3. App Dev Slow Environment Builds

Never enough environments





Part II. Data constraint price is Huge

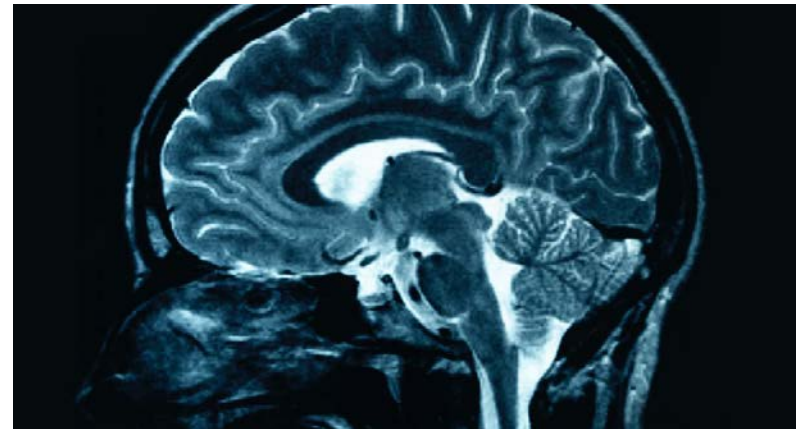
- Four Areas data tax hits
 1. IT Capital resources
 2. IT Operations personnel
 3. Application Development
 4. Business



II. Data constraint price is Huge : 4. Business

Ability to capture revenue

- Business Intelligence
 - Old data = less intelligence
- Business Applications
 - Delays cause



=> Lost Revenue





II. Data constraint price is Huge : 4. Business

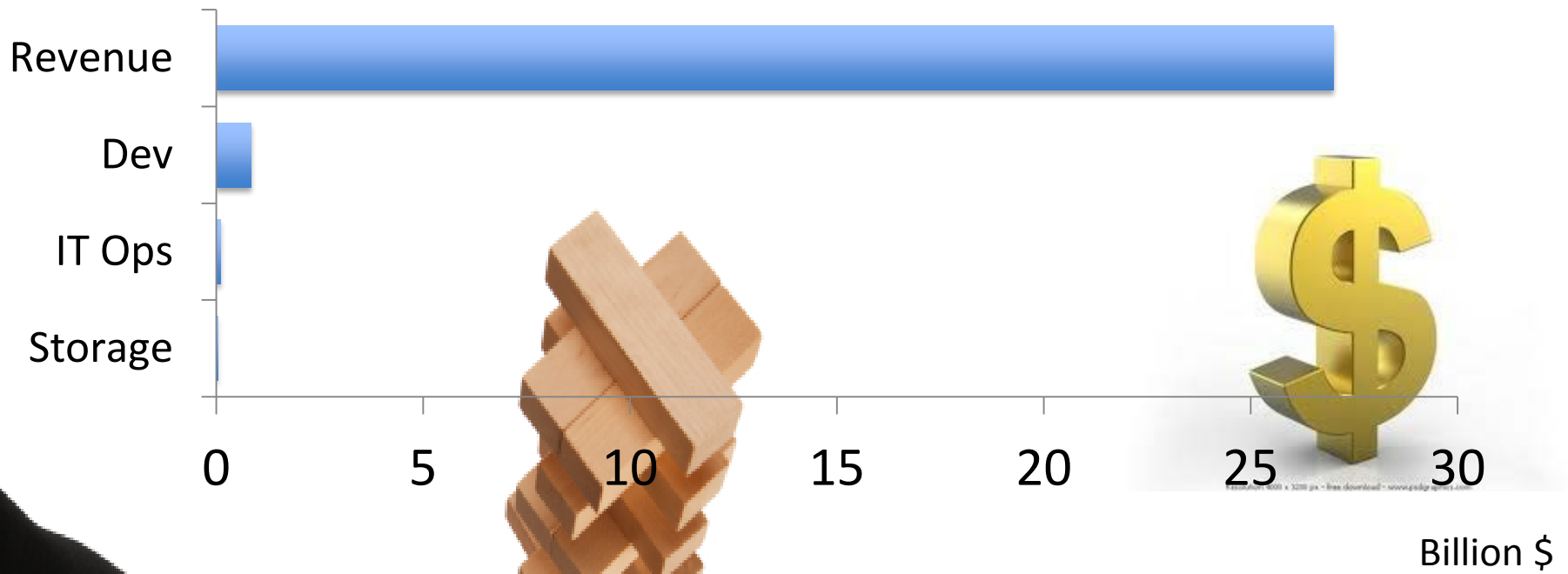


2. Delays





II. Data constraint price is Huge : 4. Business



Billion \$

Data is the constraint

- I. Data Constraint strains IT
- II. Data Constraint price is huge
- III. **Data Constraint** companies unaware



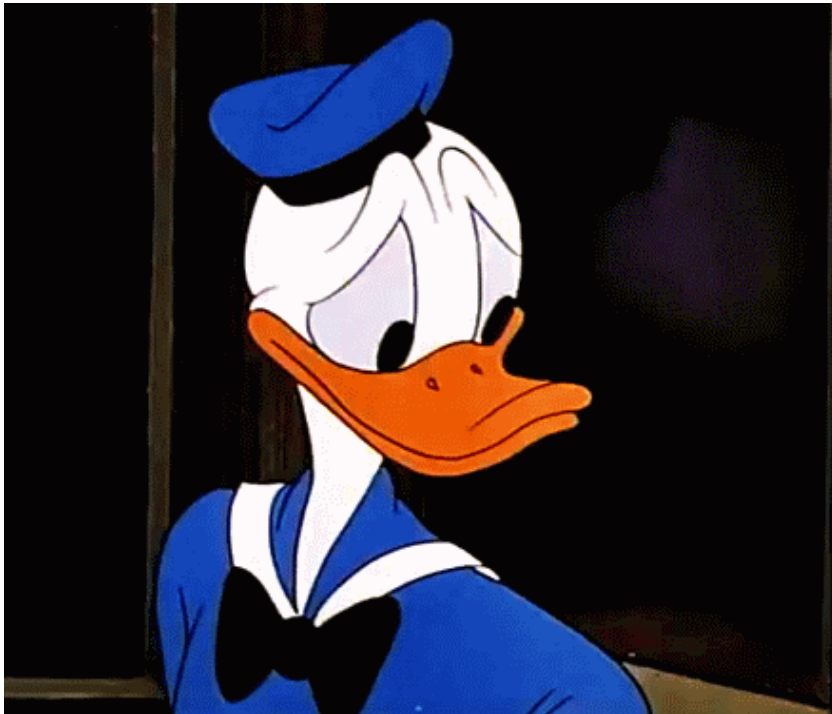
Part III. Data Constraint companies unaware





III. Data Constraint companies unaware

DBA



Developer





III. Data Constraint companies unaware

#1 Biggest Enemy :

IT departments believe

- best processes
- greatest technology
- Just the way it is





III. Data Constraint companies unaware

Don't we already do that ?



Why do I need an iPhone ?



III. Data Constraint companies unaware

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

III. Data Constraint companies unaware

How to enlighten? Ask for metrics

- How old is data in
 - BI and DW : ETL windows
 - QA and Dev : how often refreshed
- How long does it take a developer to get a DB copy?
- How long does it take QA to setup an environment

Data is the constraint

- I. Data Constraint strains IT
- II. Data Constraint price is huge
- III. Data Constraint companies unaware

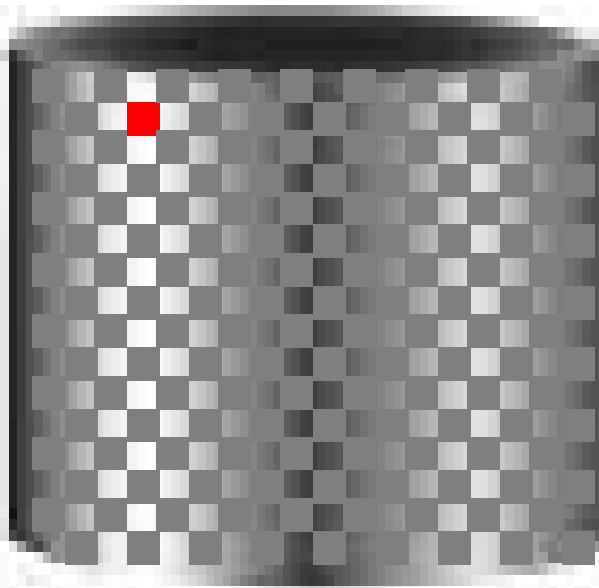
In this presentation :

- Problem in the Industry
- **Solution**
- Use Cases

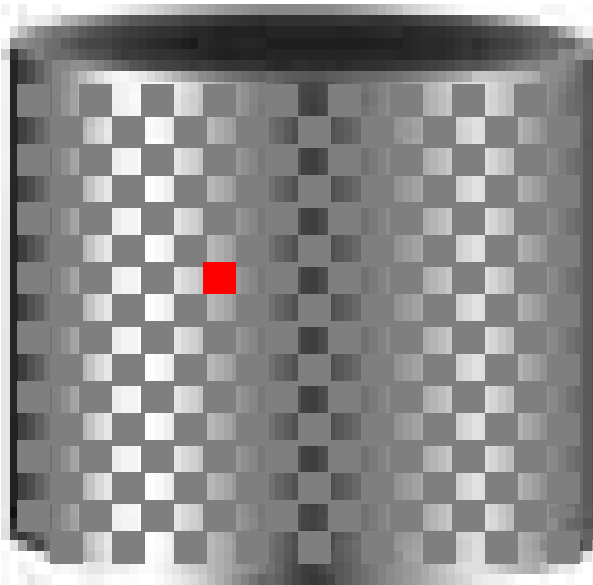


99% of blocks are identical

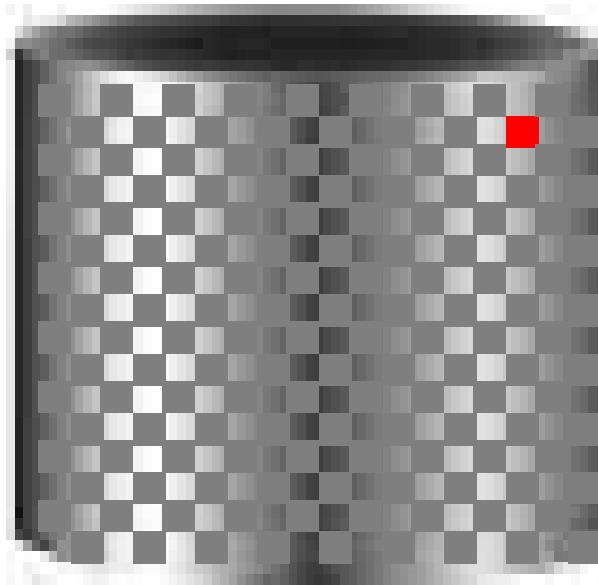
Clone 1



Clone 2



Clone 3



Solution



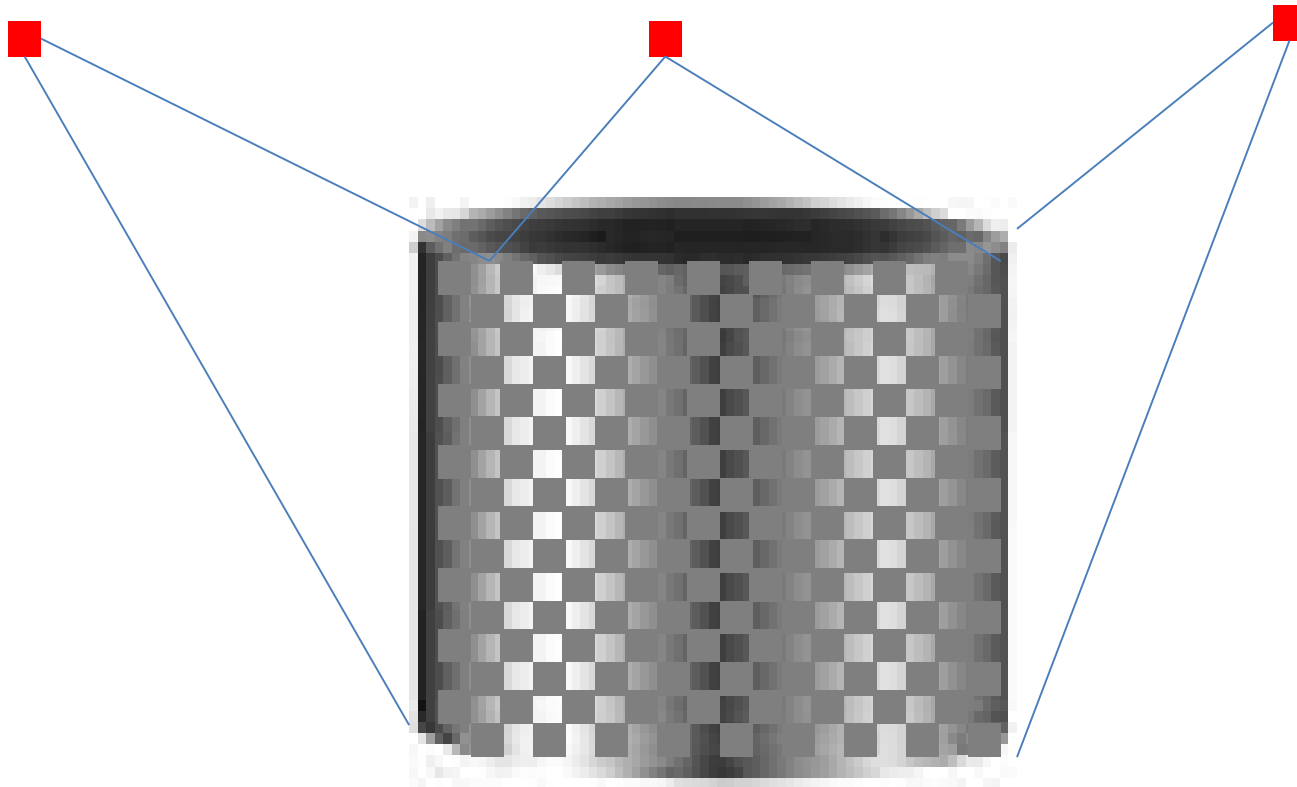


Thin Clone

Clone 1

Clone 2

Clone 3

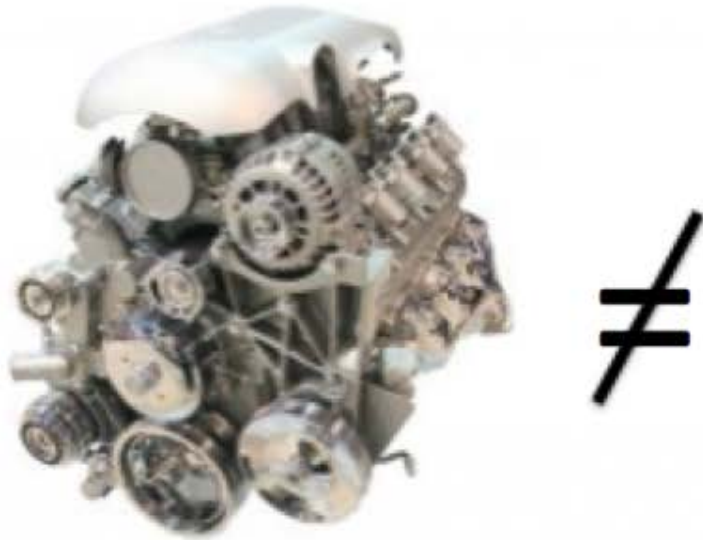


Technology Core : file system snapshots

- VMware Linked Clones
 - Not supported for Oracle
- EMC
 - 16 snapshots
 - Write performance impact
- Netapp
 - 255 snapshots
- ZFS
 - Unlimited snapshots

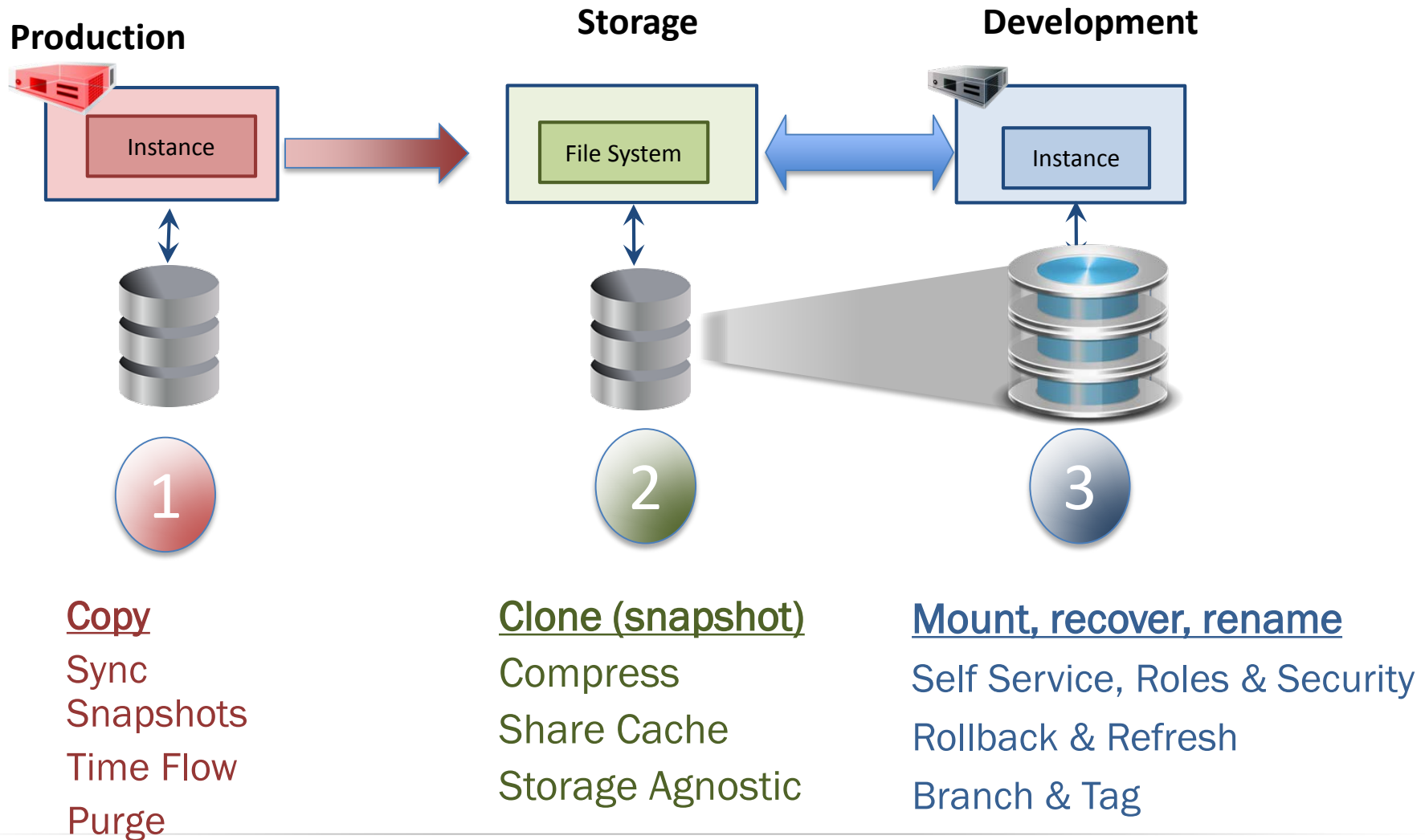


III. Companies unaware of the Data Tax



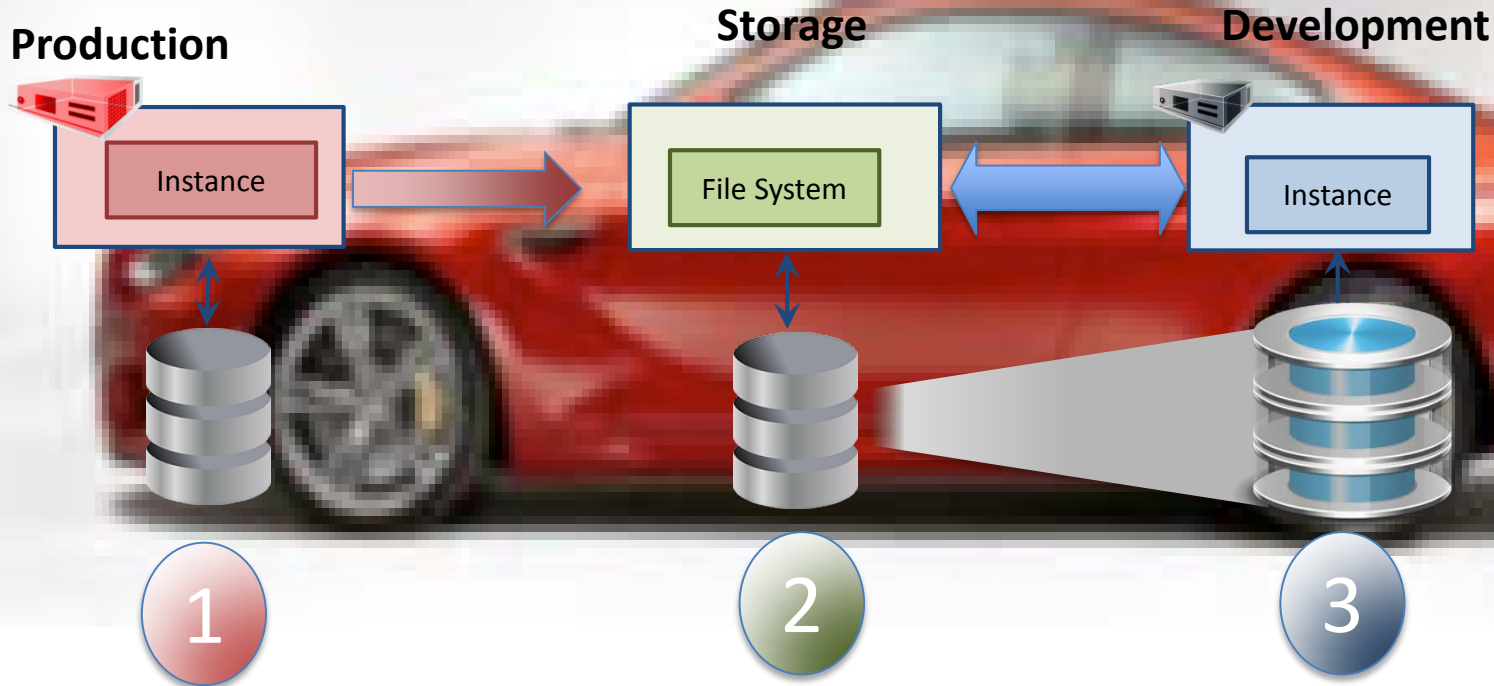


Three Core Parts





Three Core Parts



Copy
Sync
Snapshots
Time Flow
Purge



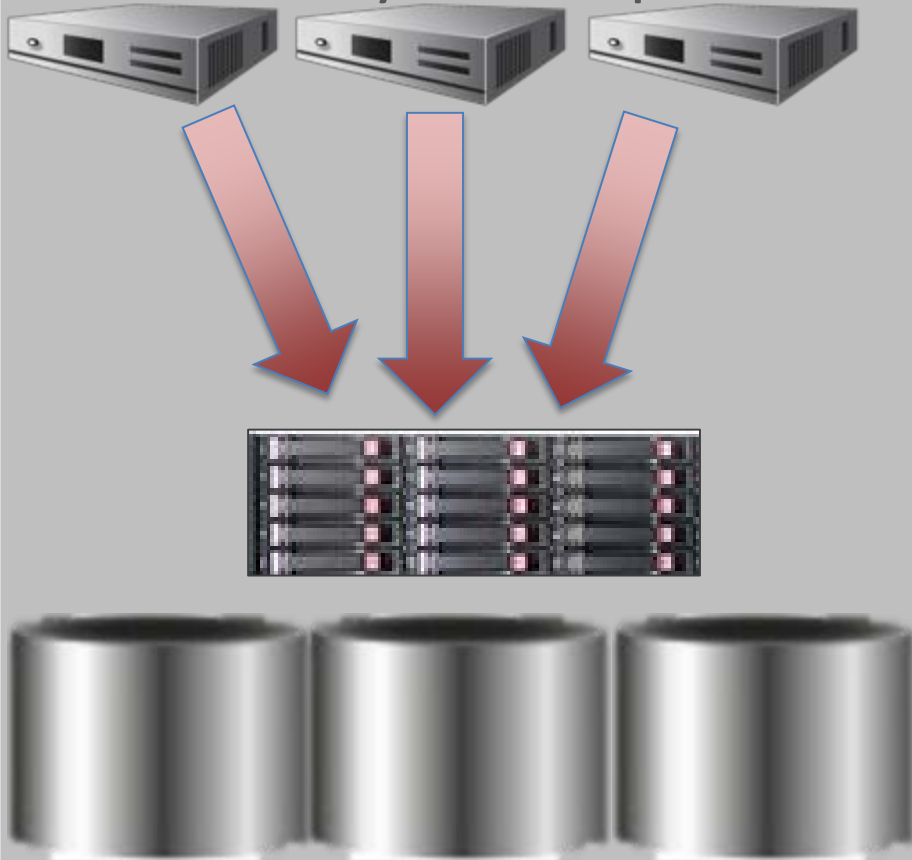
Clone (snapshot)
Compress
Share Cache
Storage Agnostic

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

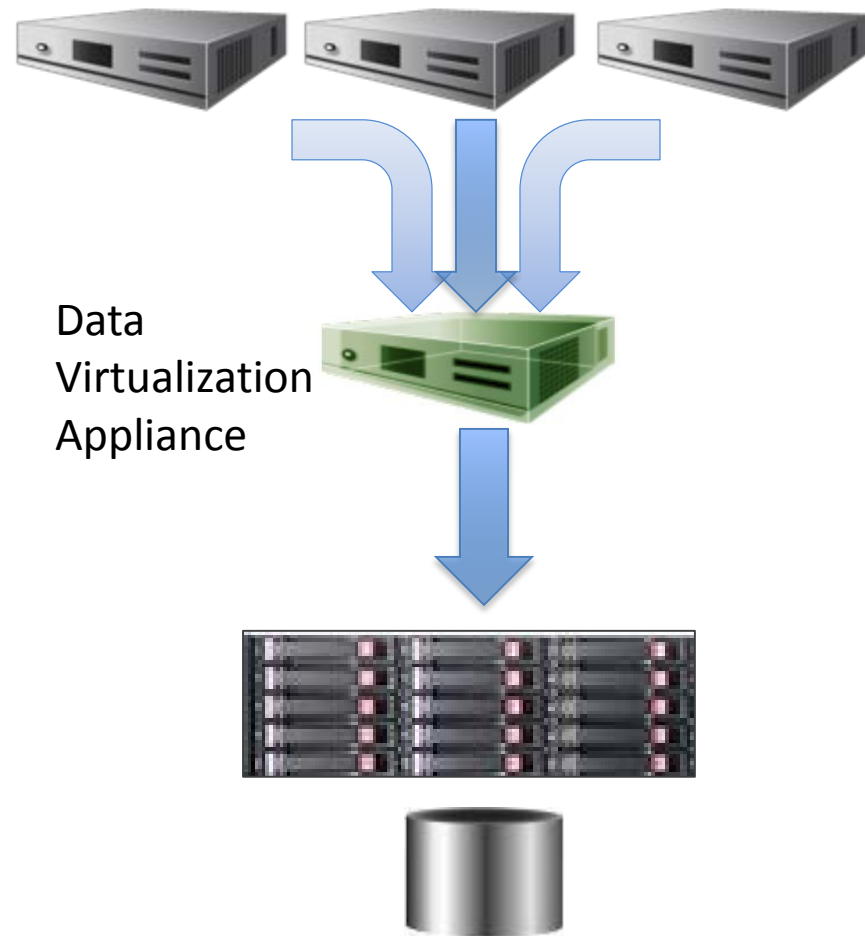
3. Database Virtualization



Three Physical Copies



Three Virtual Copies



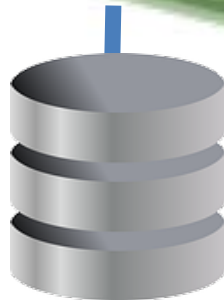
Install Delphix on x86 hardware





Allocate Any Storage to Delphix

DELPHIX
SERVER



Allocate Storage
Any type

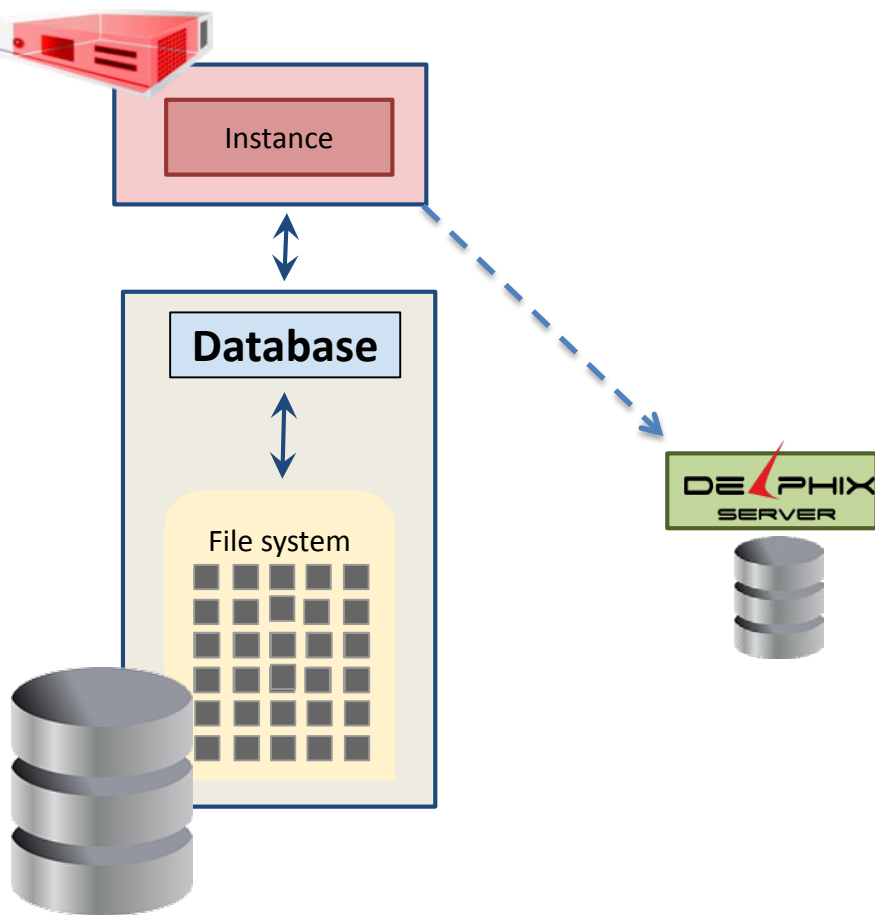


Pure Storage + Delphix
Better Performance for
1/10 the cost



One time backup of source database

Production



Supports

Microsoft
SQL Server

ORACLE[®]

PostgreSQL

Application Stack Data

Upcoming

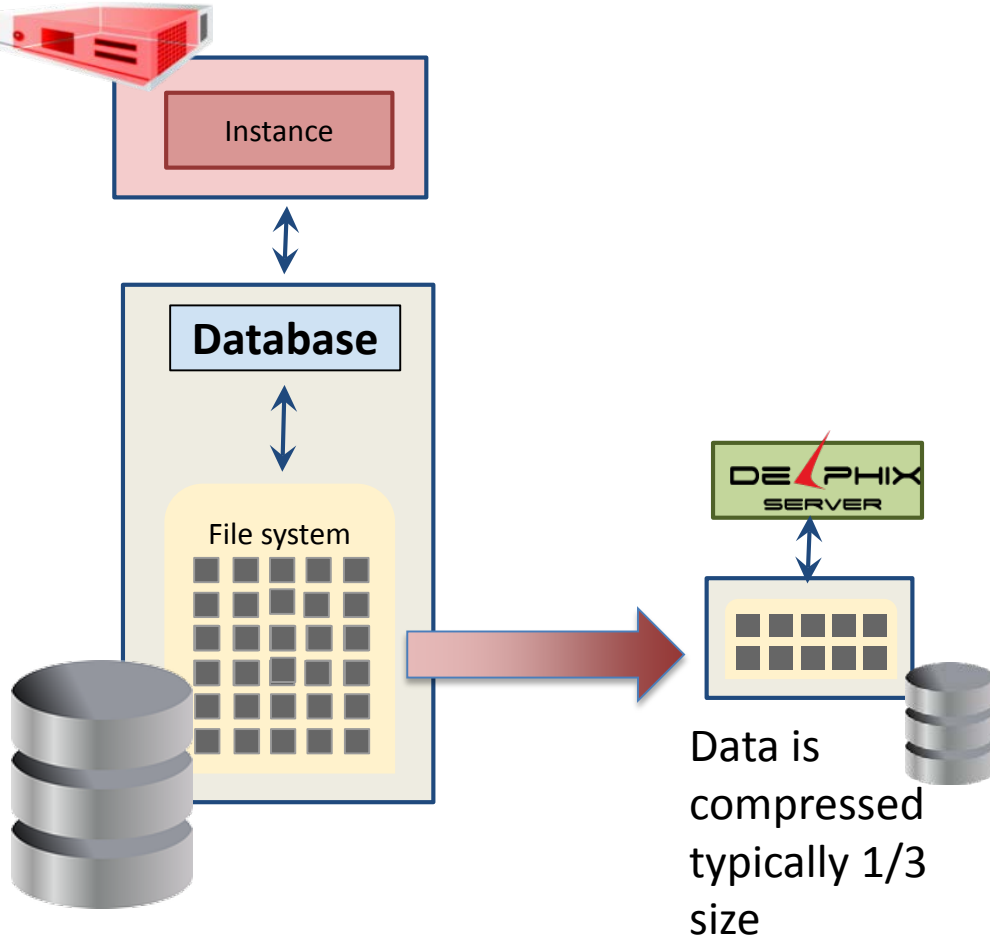
SYBASE[®]

MySQL[®]



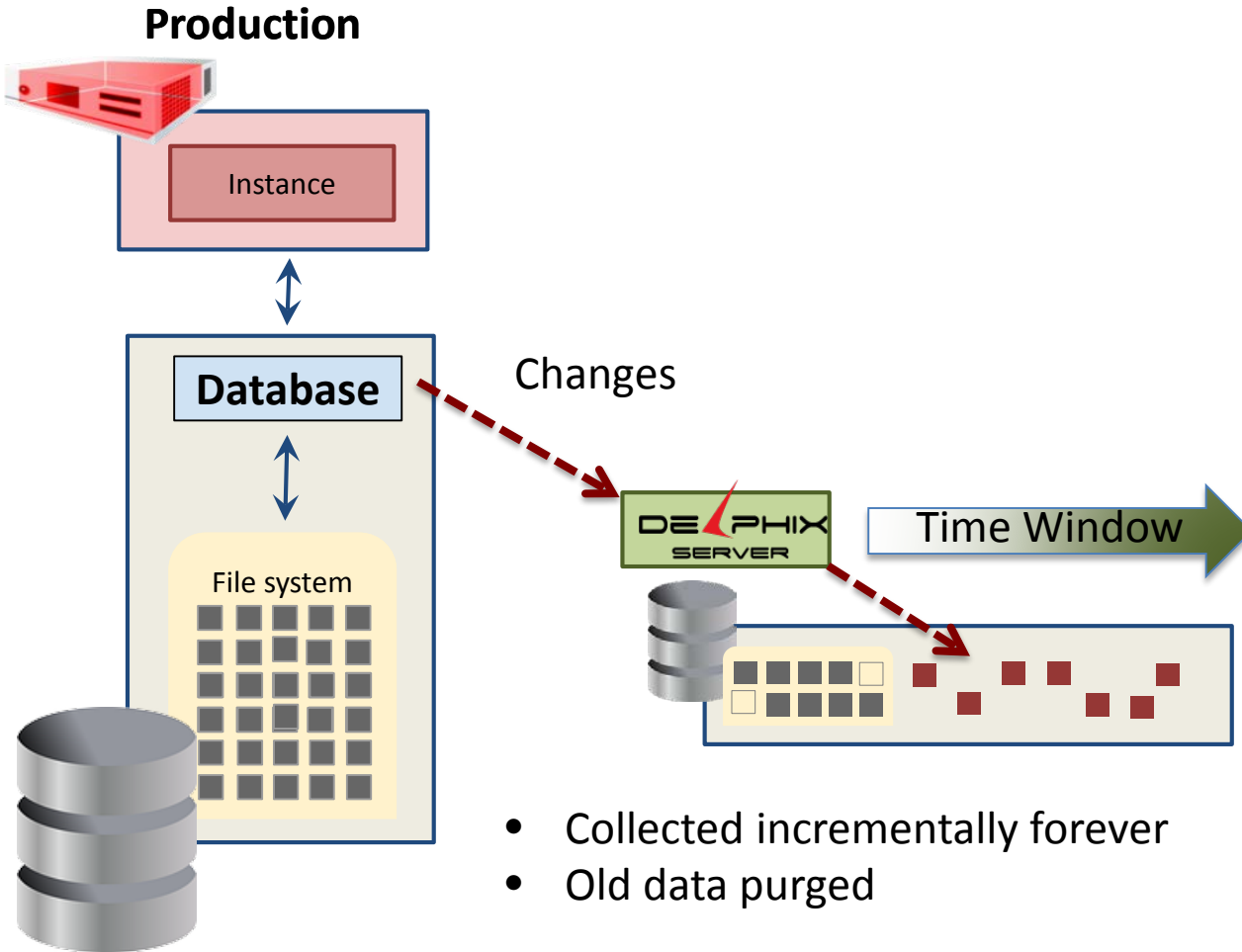
DxFS (Delphix) Compress Data

Production

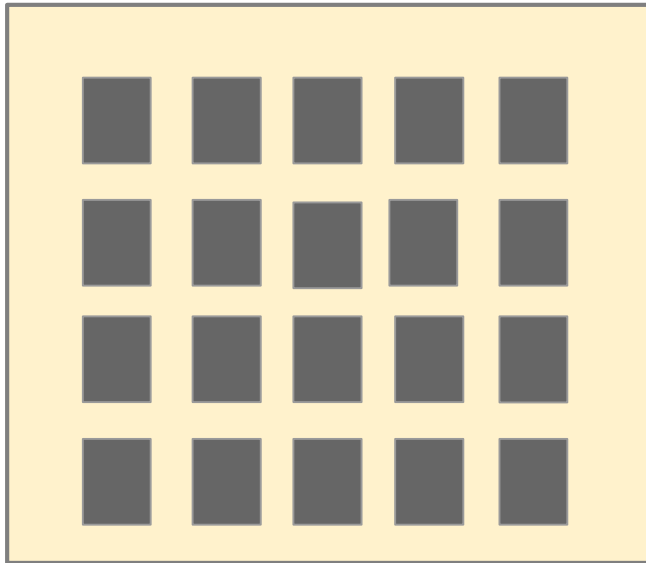




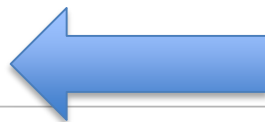
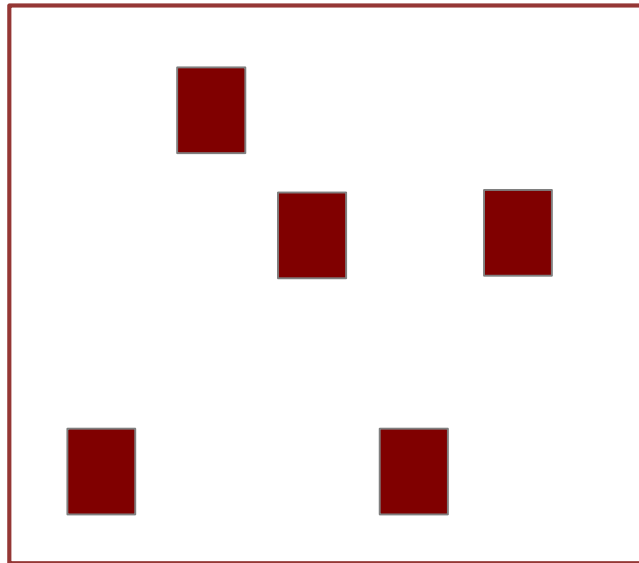
Incremental forever change collection



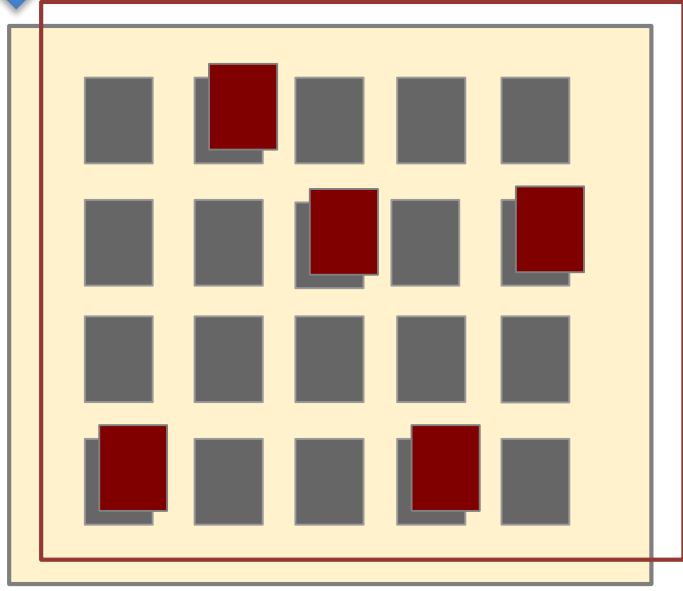
Source Full Copy



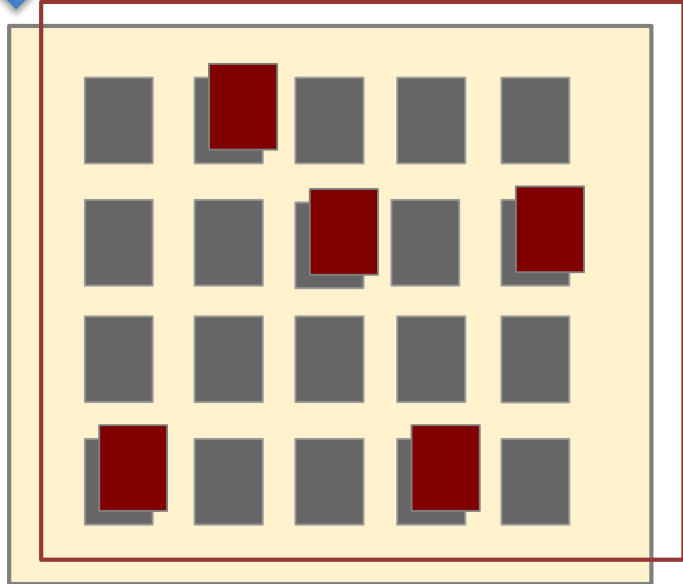
Source backup
from SCN 1



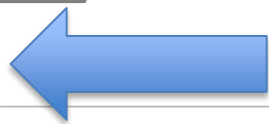
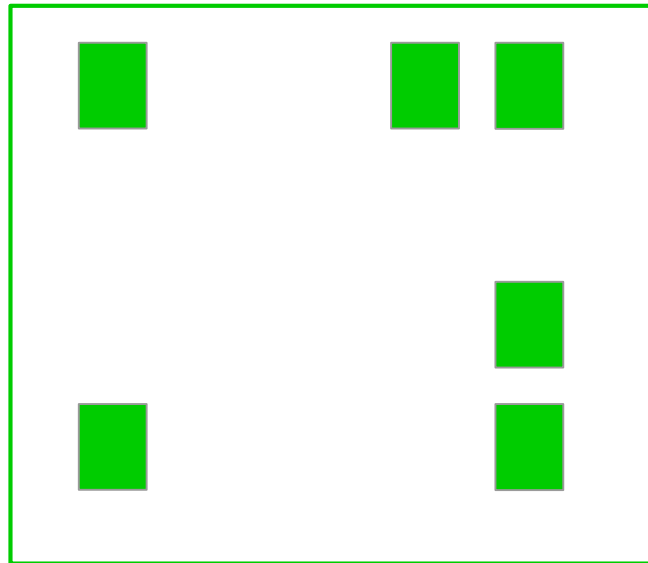
Snapshot 1 Snapshot 2

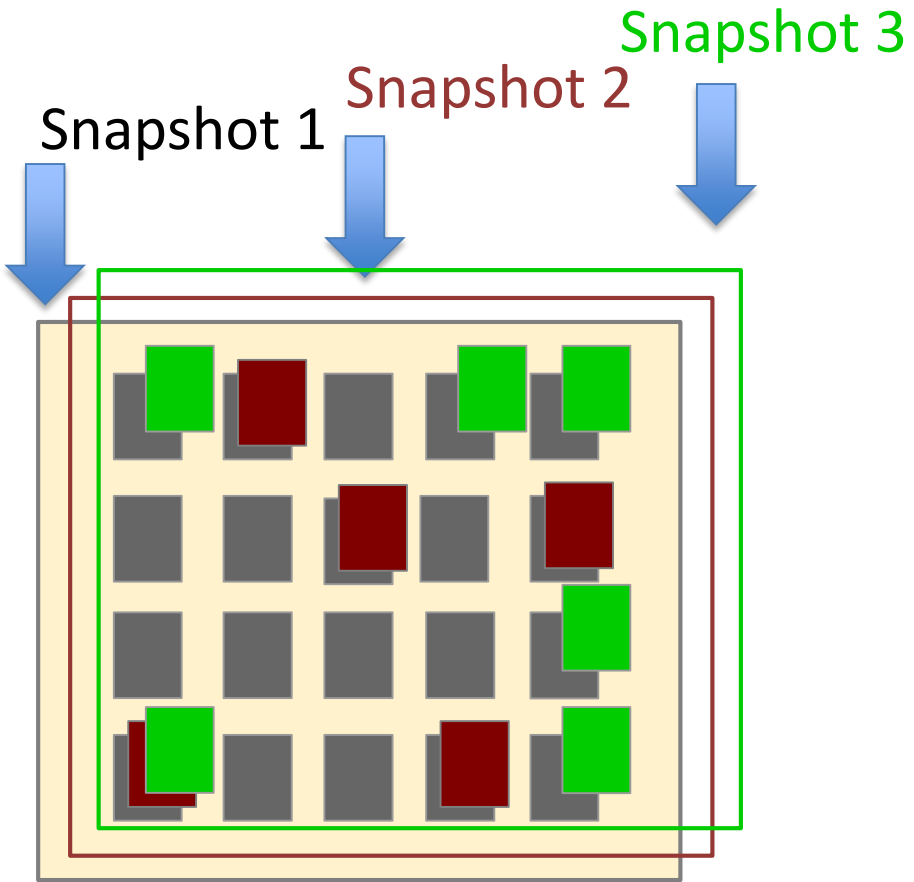


Snapshot 1 Snapshot 2

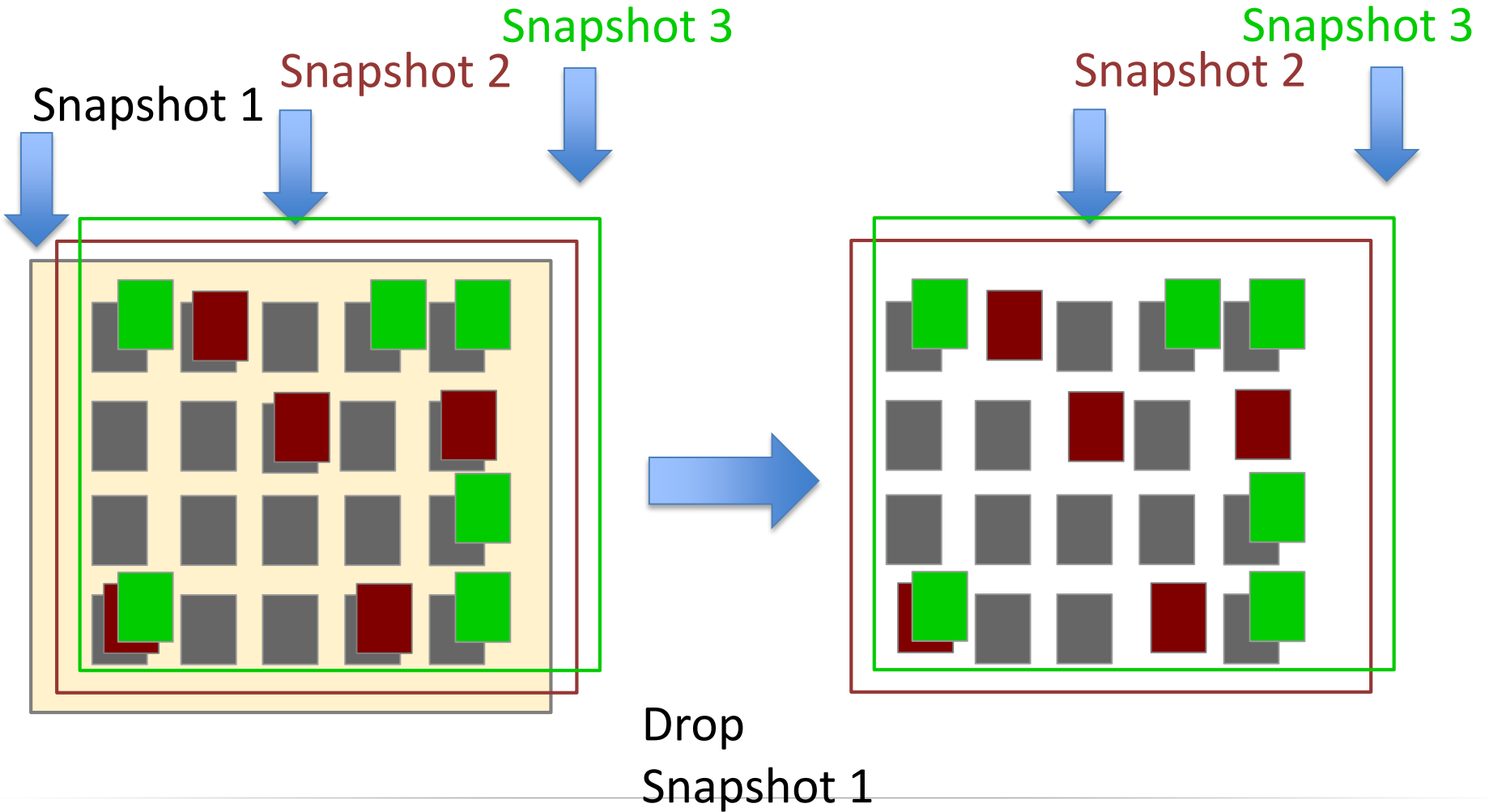


Backup from SCN

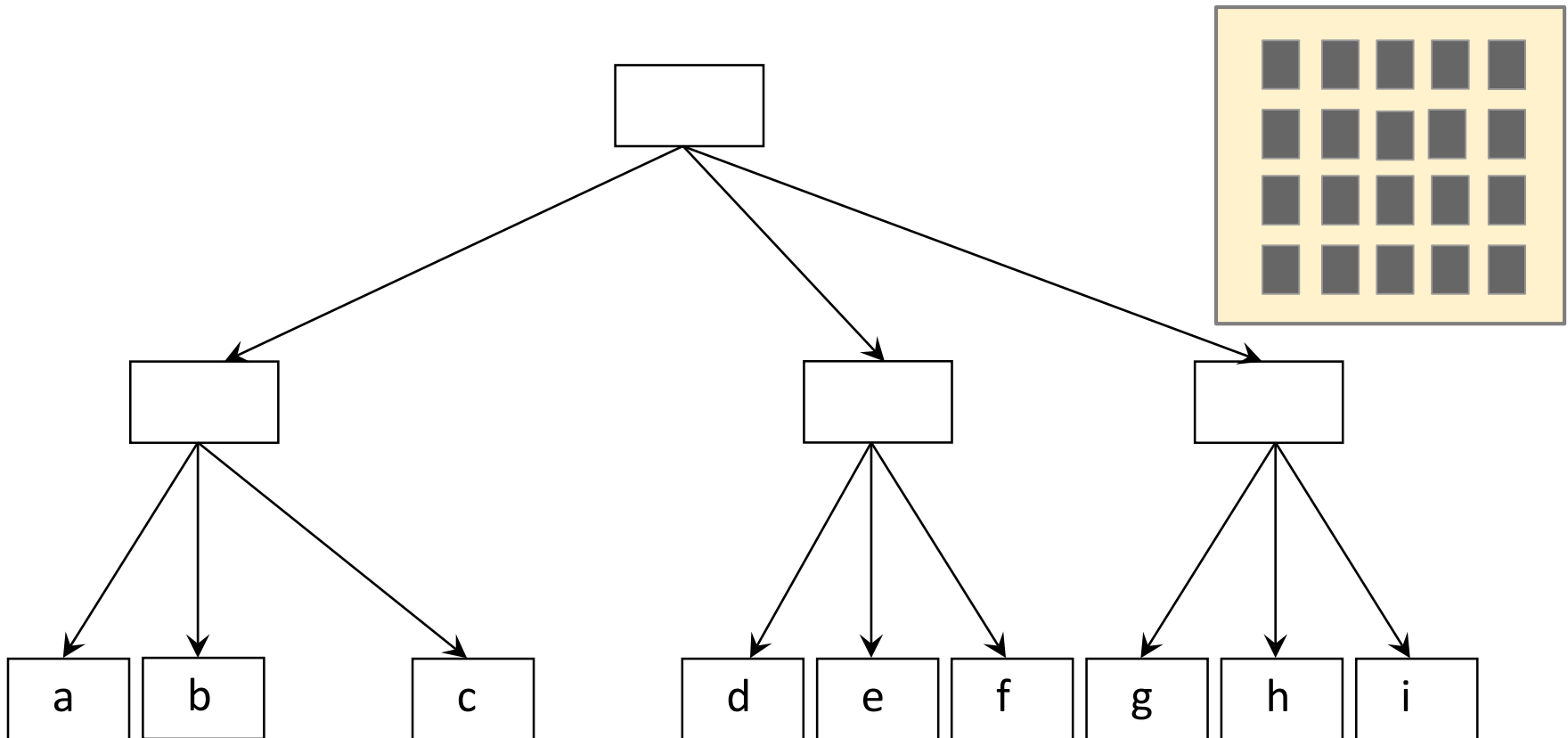




Drop Snapshot



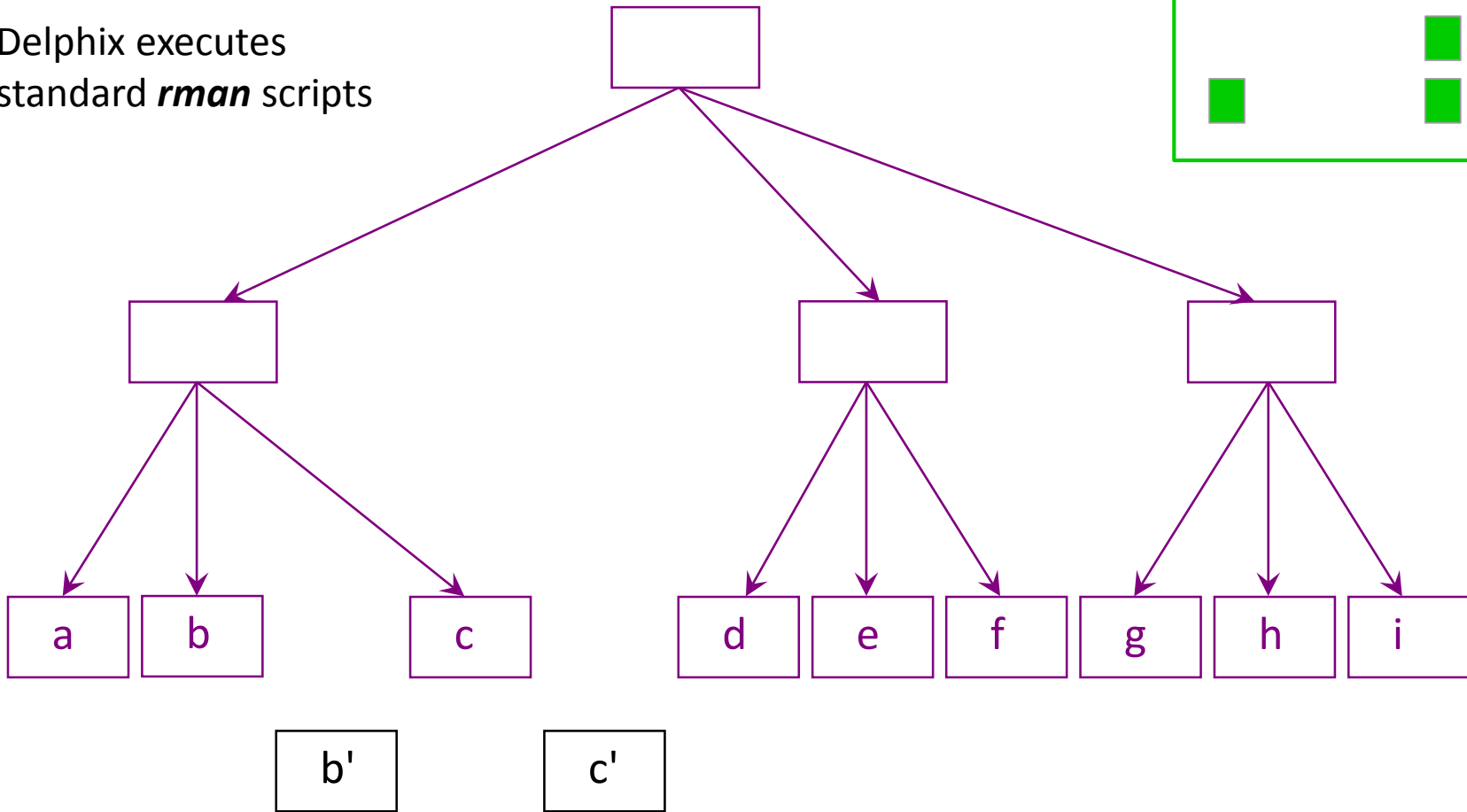
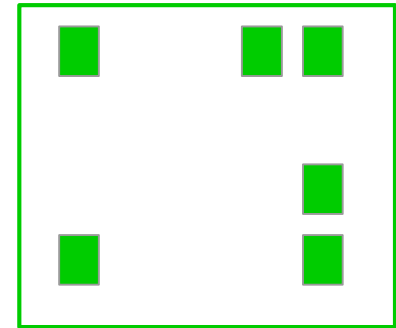
Snapshot 1 – full backup once only at link time



We start with a full backup - analogous to a level 0 *rman* backup. Includes the archived redo log files needed for recovery. Run in archive log mode.

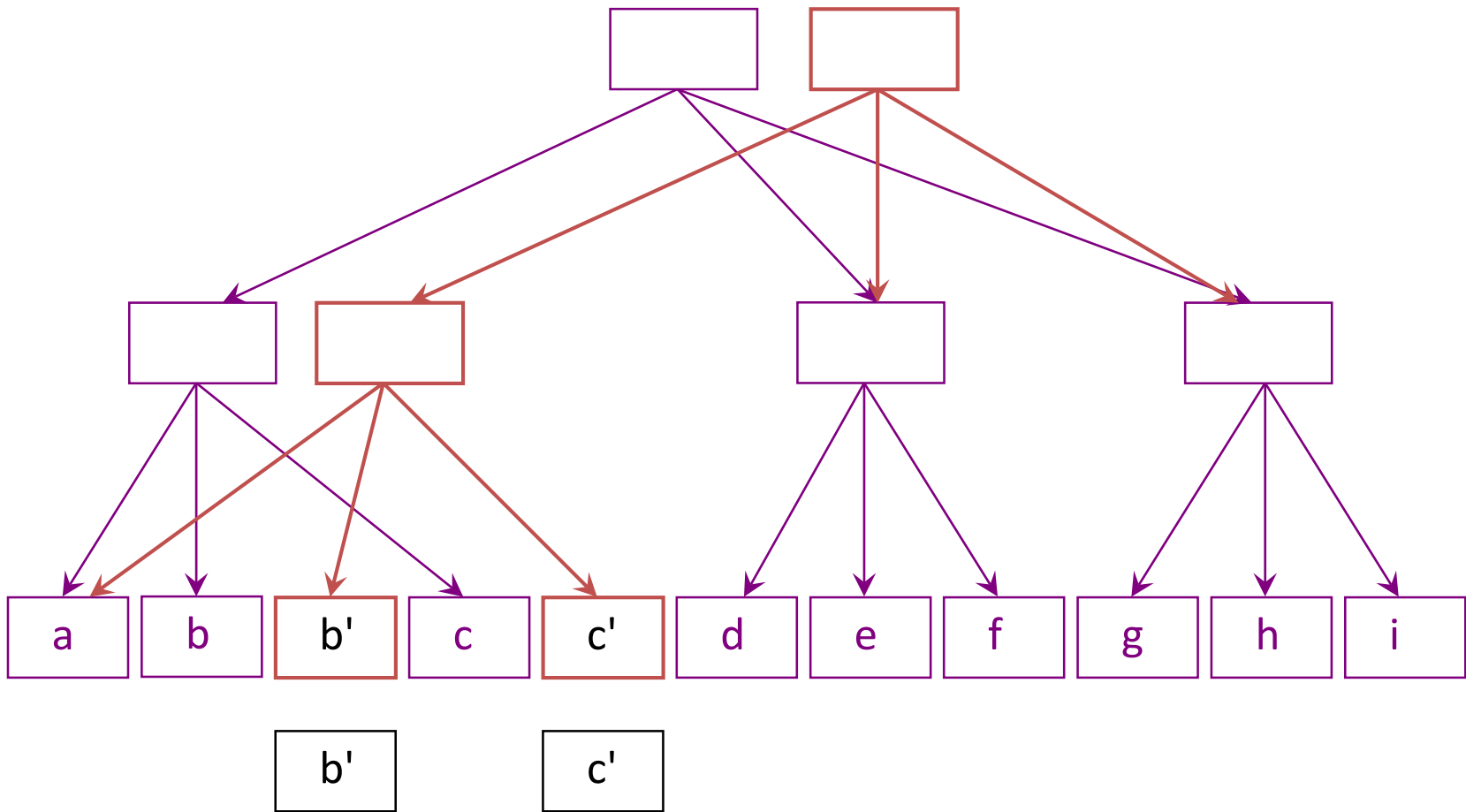
Snapshot 2 (from SCN)

Delphix executes standard *rman* scripts



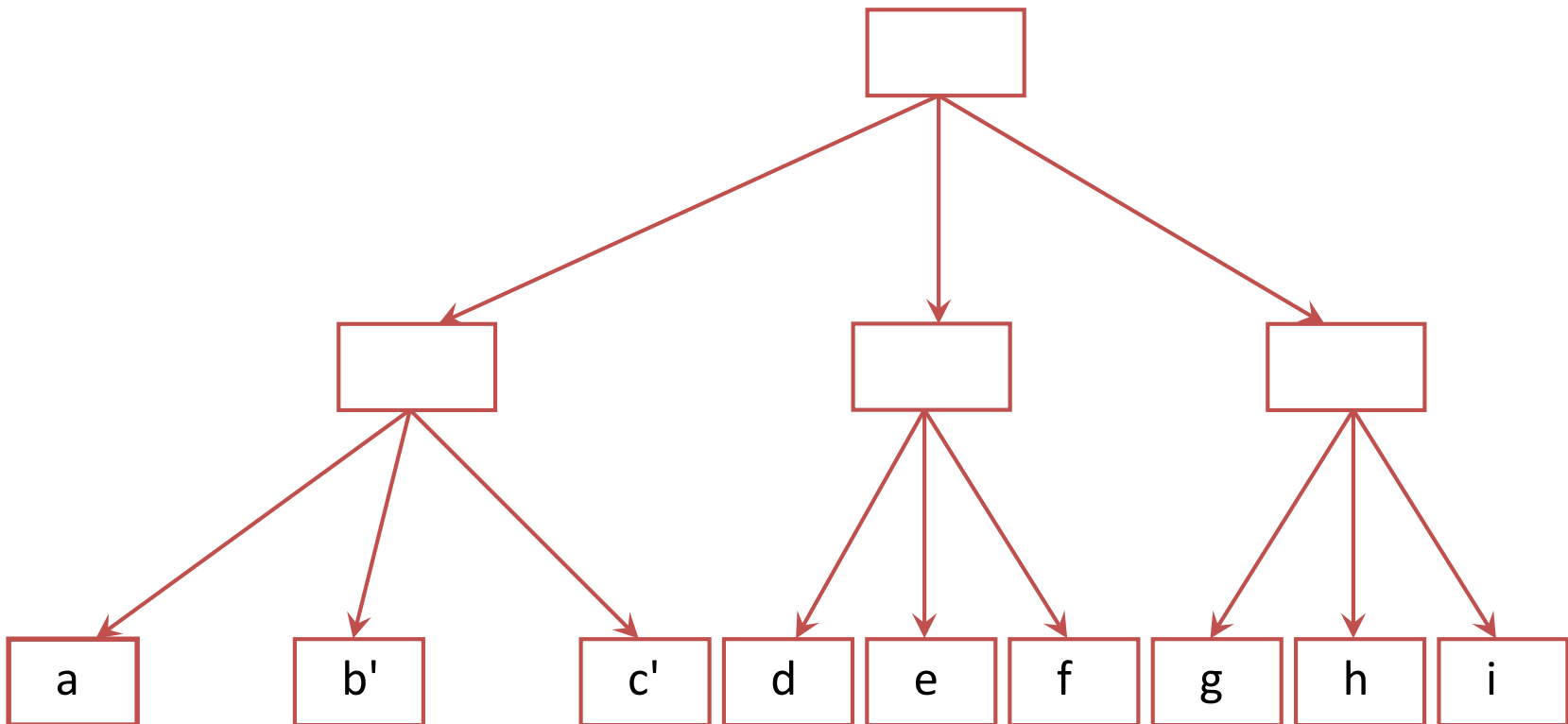
The "backup from SCN" is analogous to a level 1 incremental backup (which includes the relevant archived redo logs). Sensible to enable BCT.

Apply Snapshot 2



The Delphix appliance unpacks the *rman* backup and "overwrites" the initial backup with the changed blocks - but DxFs makes new copies of the blocks

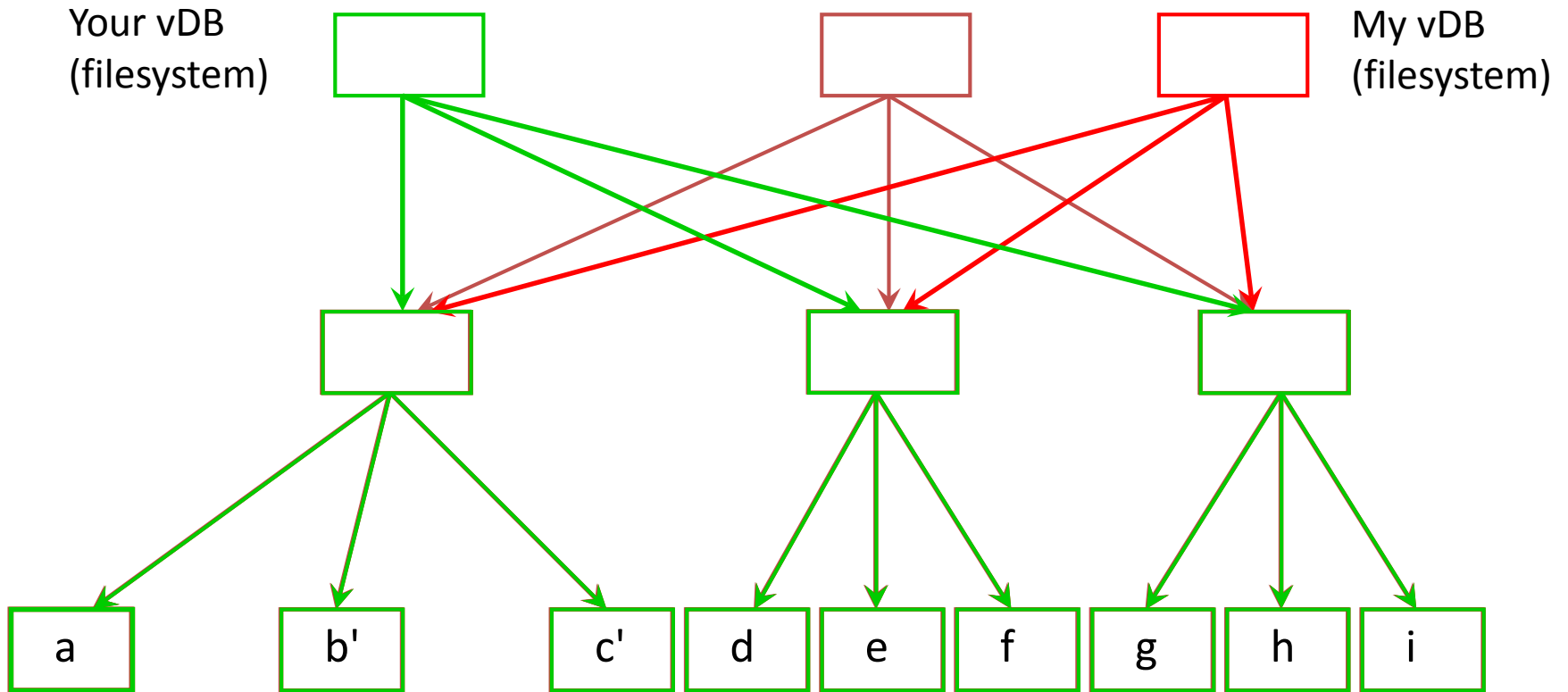
Derived Full Backup at Snapshot 2



The call to *rman* leaves us with a new level 0 backup, waiting for recovery.

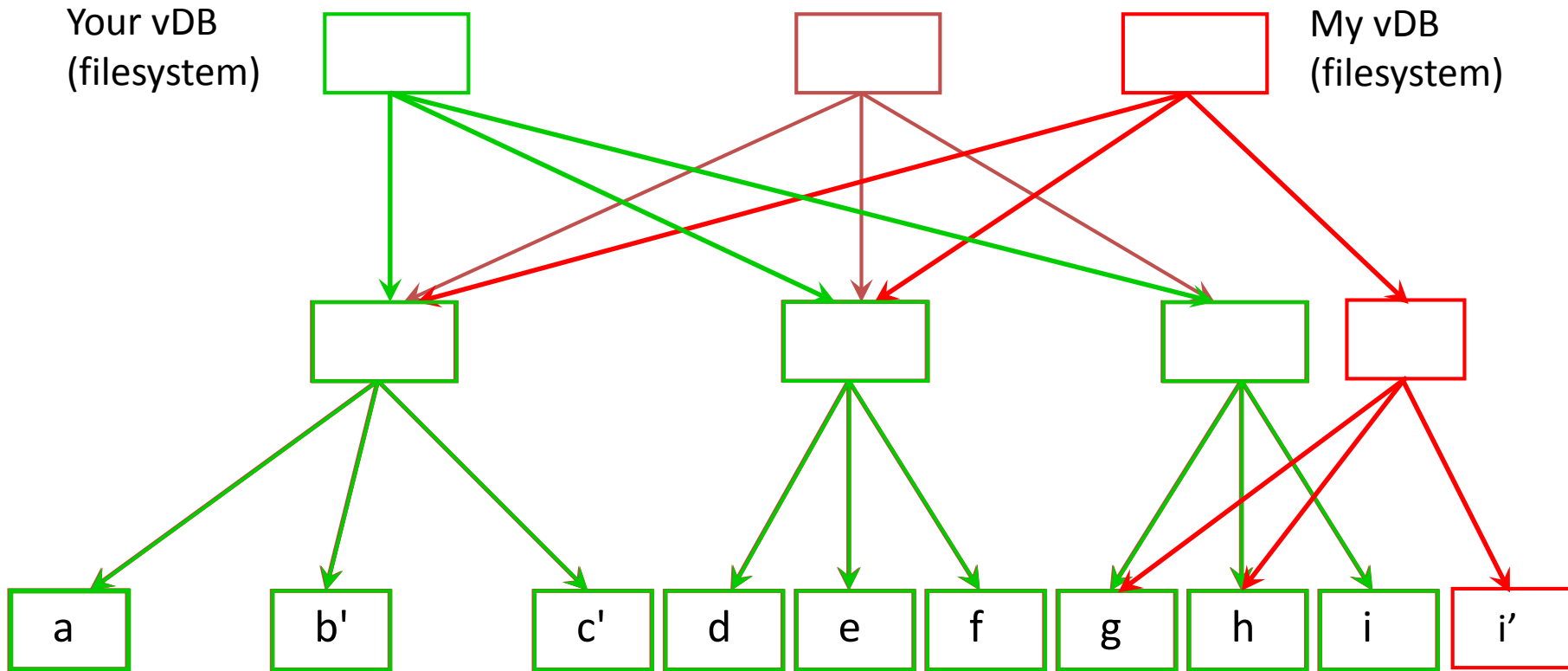
But we can pick the snapshot root block. We have EVERY level 0 backup

Creating a vDB



The first step in creating a vDB is to take a snapshot of the filesystem as at the backup you want (then roll it forward)

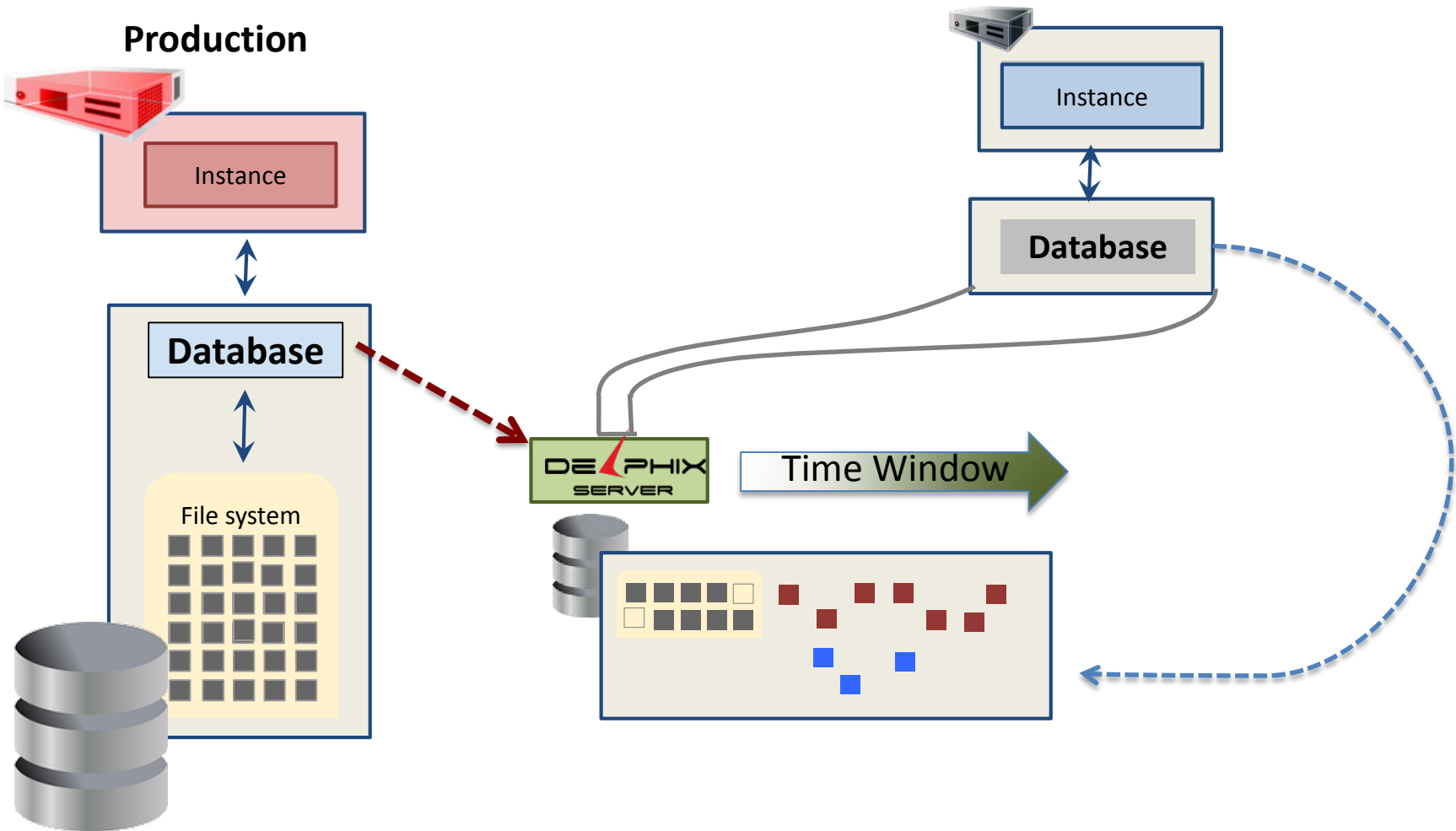
Creating a vDB



The first step in creating a vDB is to take a snapshot of the filesystem as at the backup you want (then roll it forward)



Cloning



In this presentation :

- Problem in the Industry
- Solution
- **Use Cases**

Use Cases

1. Development
2. QA
3. Recovery
4. Business Intelligence
5. Modernization



Use Cases

1. Development
2. QA
3. Recovery
4. Business Intelligence
5. Modernization



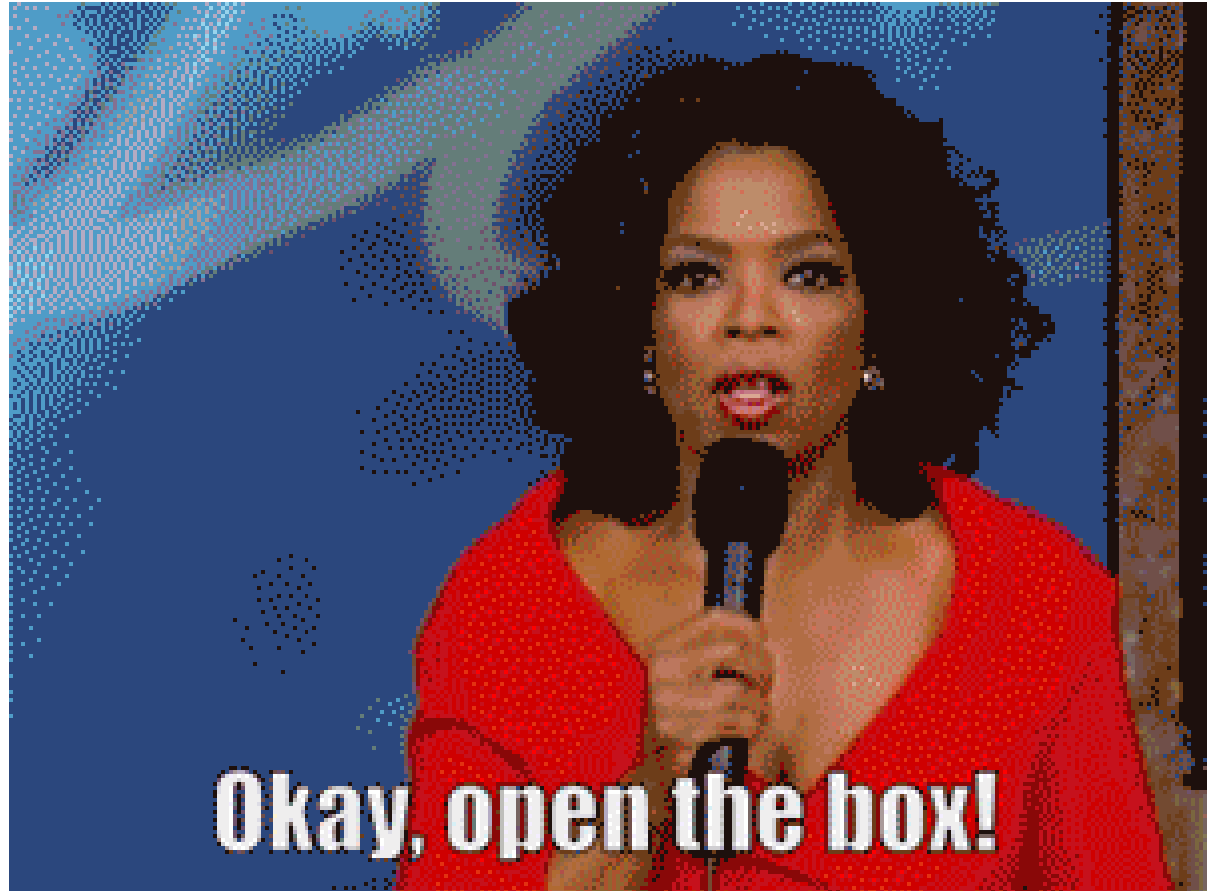
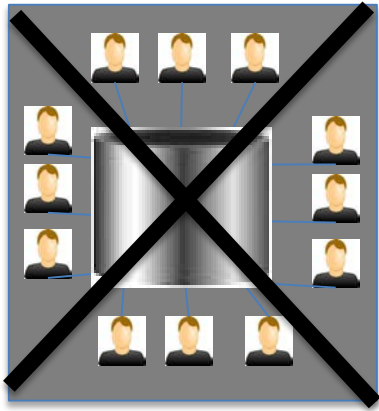
Development

- Parallelized Environments
- Full size environments
- Self Service





Development: Parallelize Environments



gif by Steve Karam

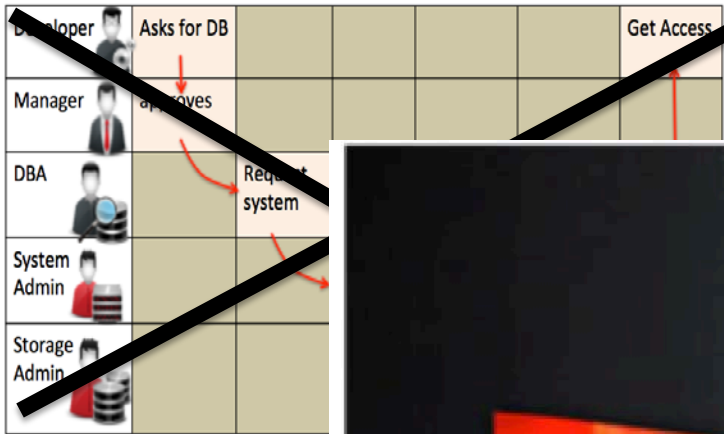


Development: Full size copies





Development: Self Service



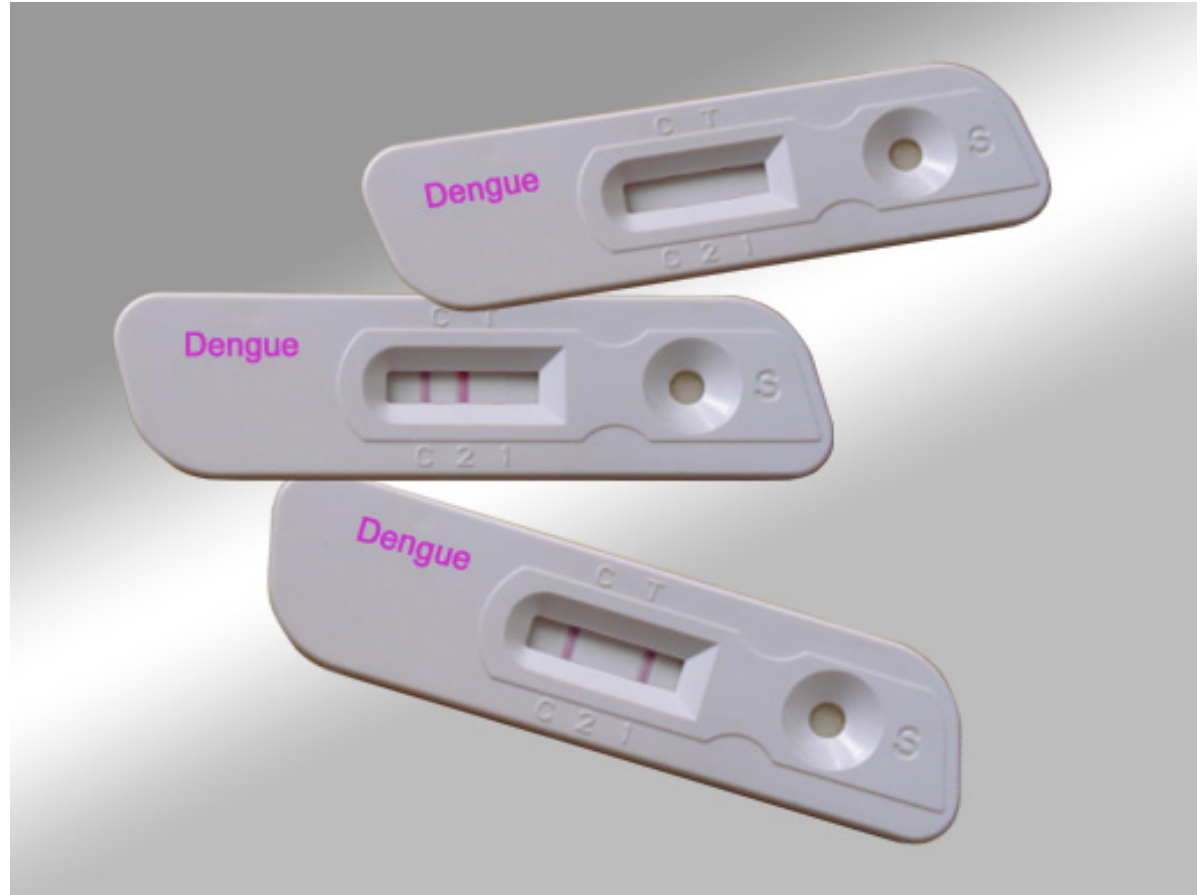
Use Cases

1. Development
2. QA
3. Recovery
4. Business Intelligence
5. Modernization



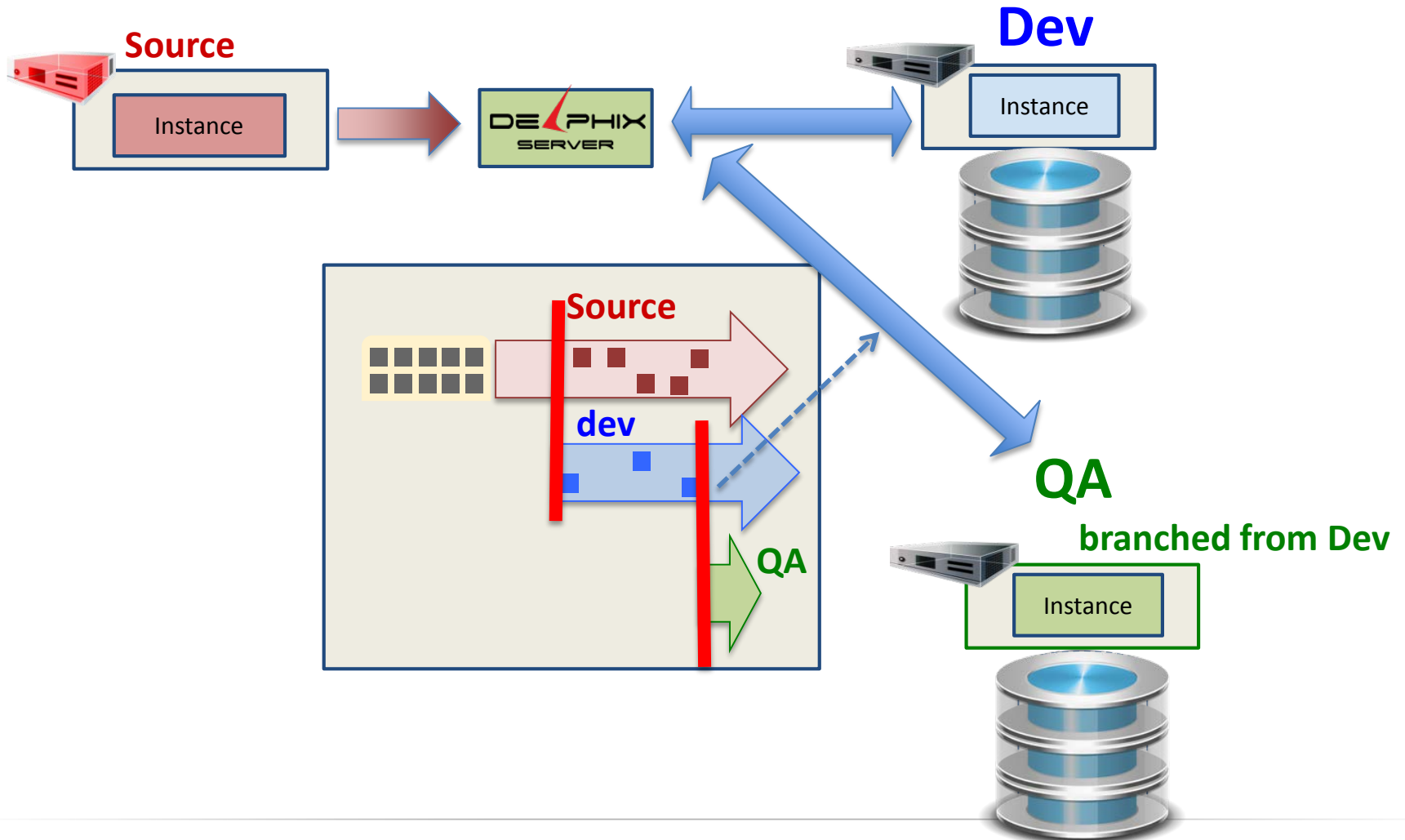
QA

- Fast
- Parallel
- Rollback
- A/B testing



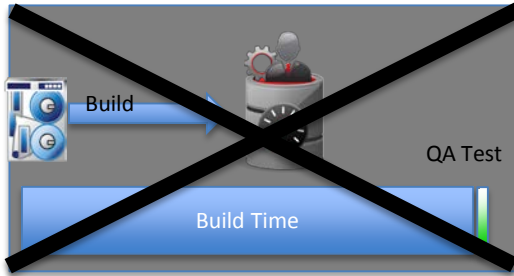


QA : Fast environments with Branching





QA : Fast environments with Branching



B
u
i
l
d
T
i
m
e

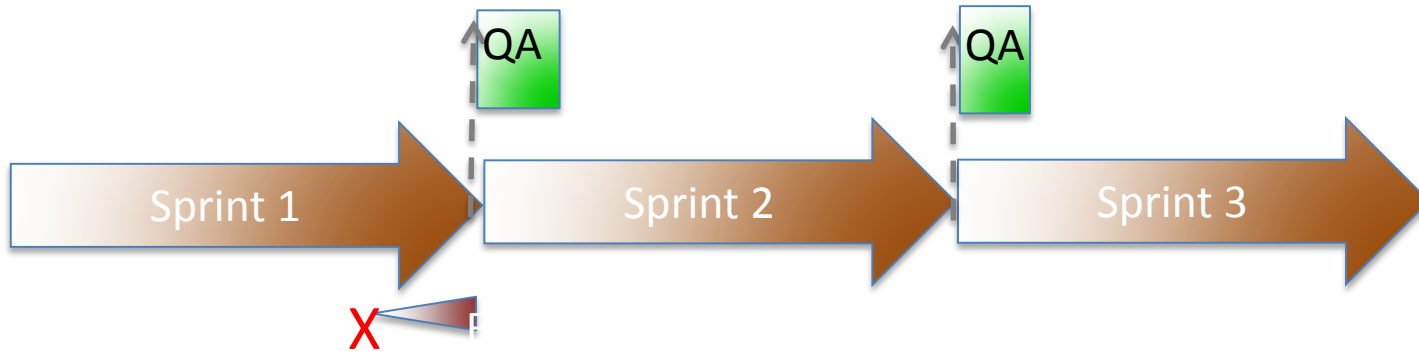
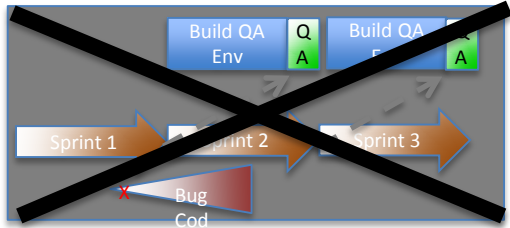
QA Test



1% of QA time was building environment
\$.99/\$1.00 actual testing vs. setup

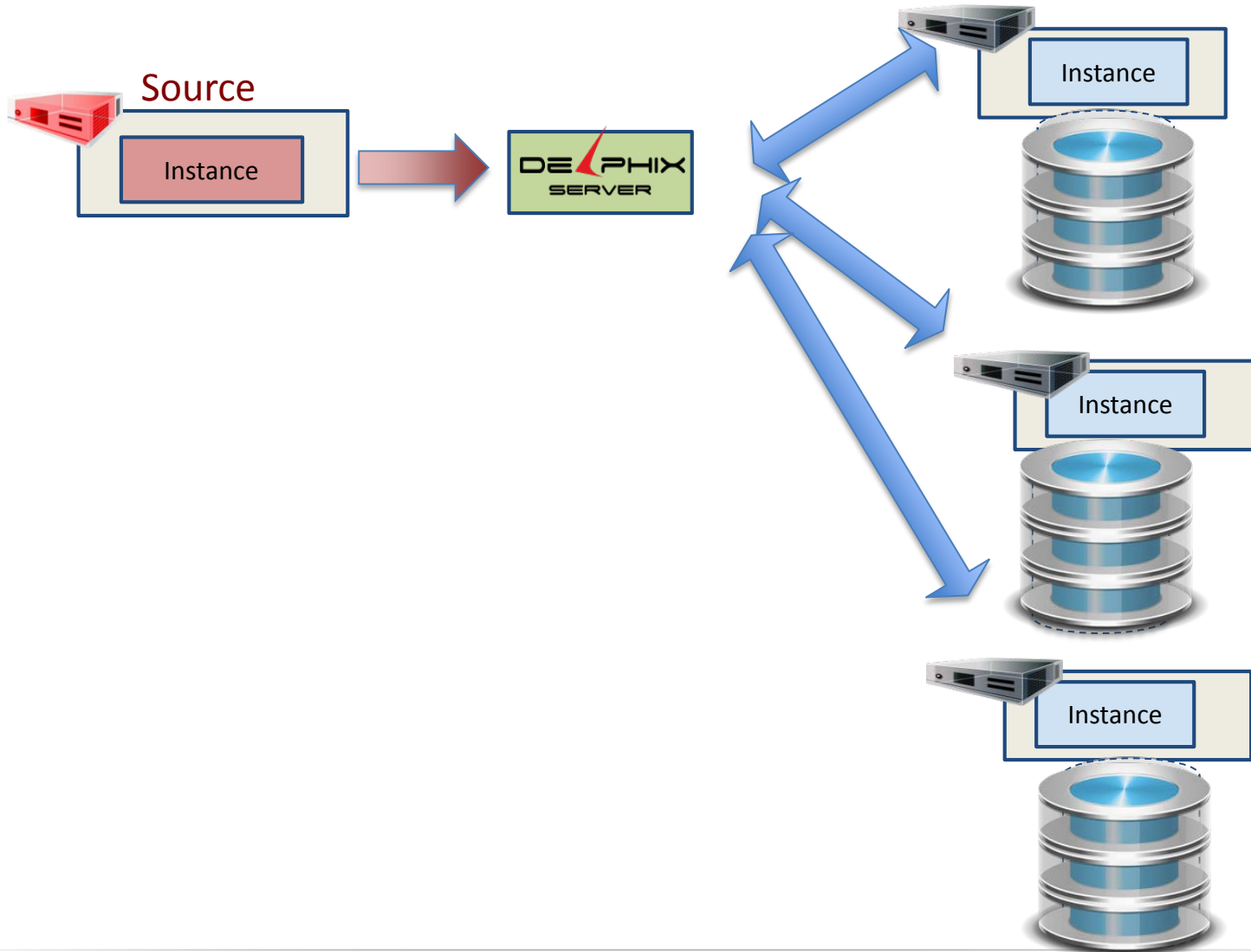


QA : bugs found fast



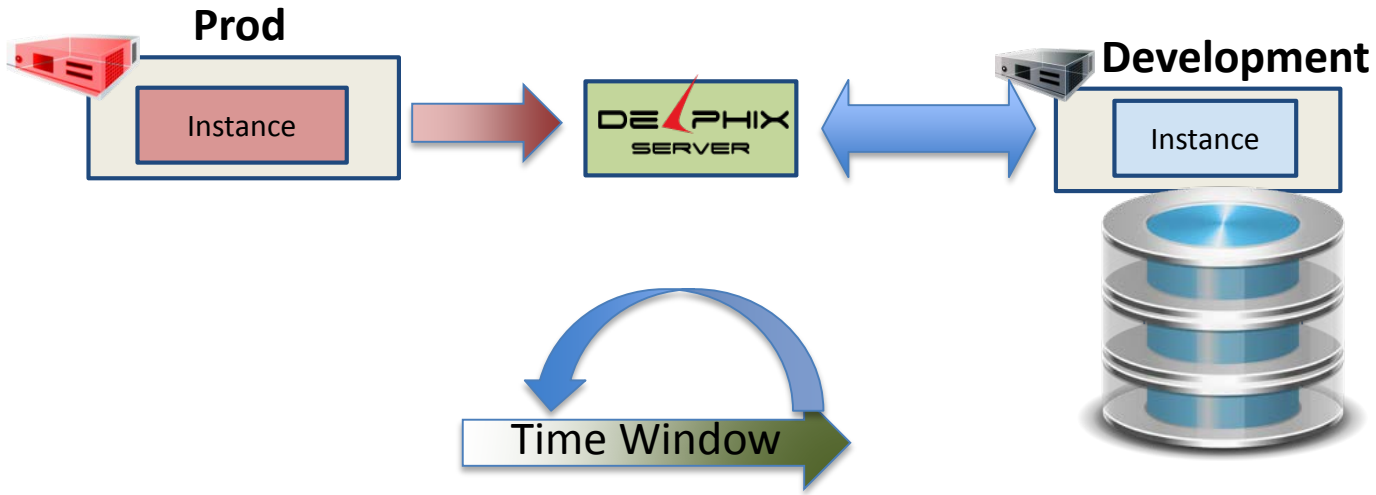


QA : Parallel environments



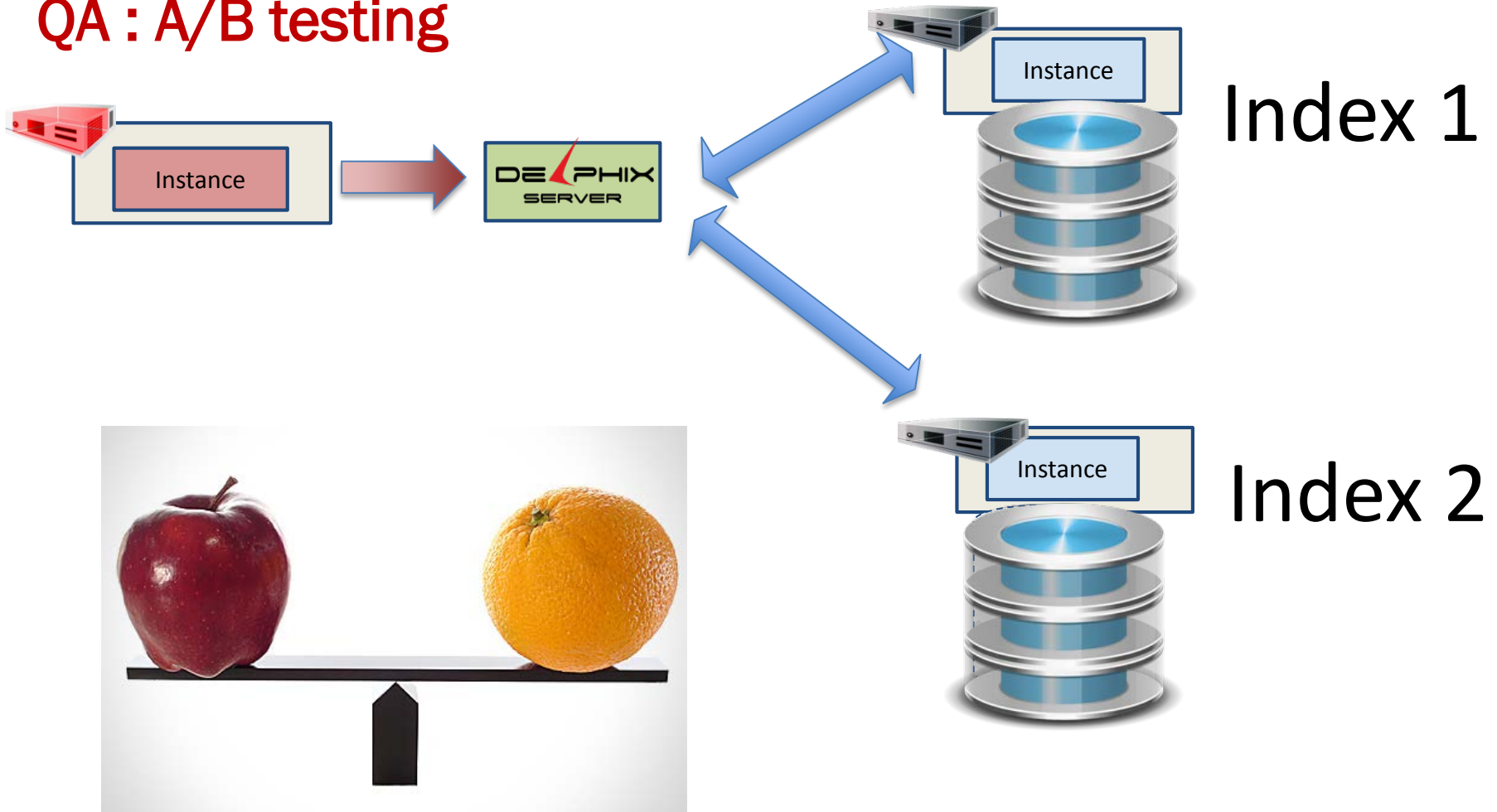


QA : Rewind for patch and QA testing





QA : A/B testing



Use Cases

1. Development
2. QA
3. Quality
4. Business Intelligence
5. Modernization



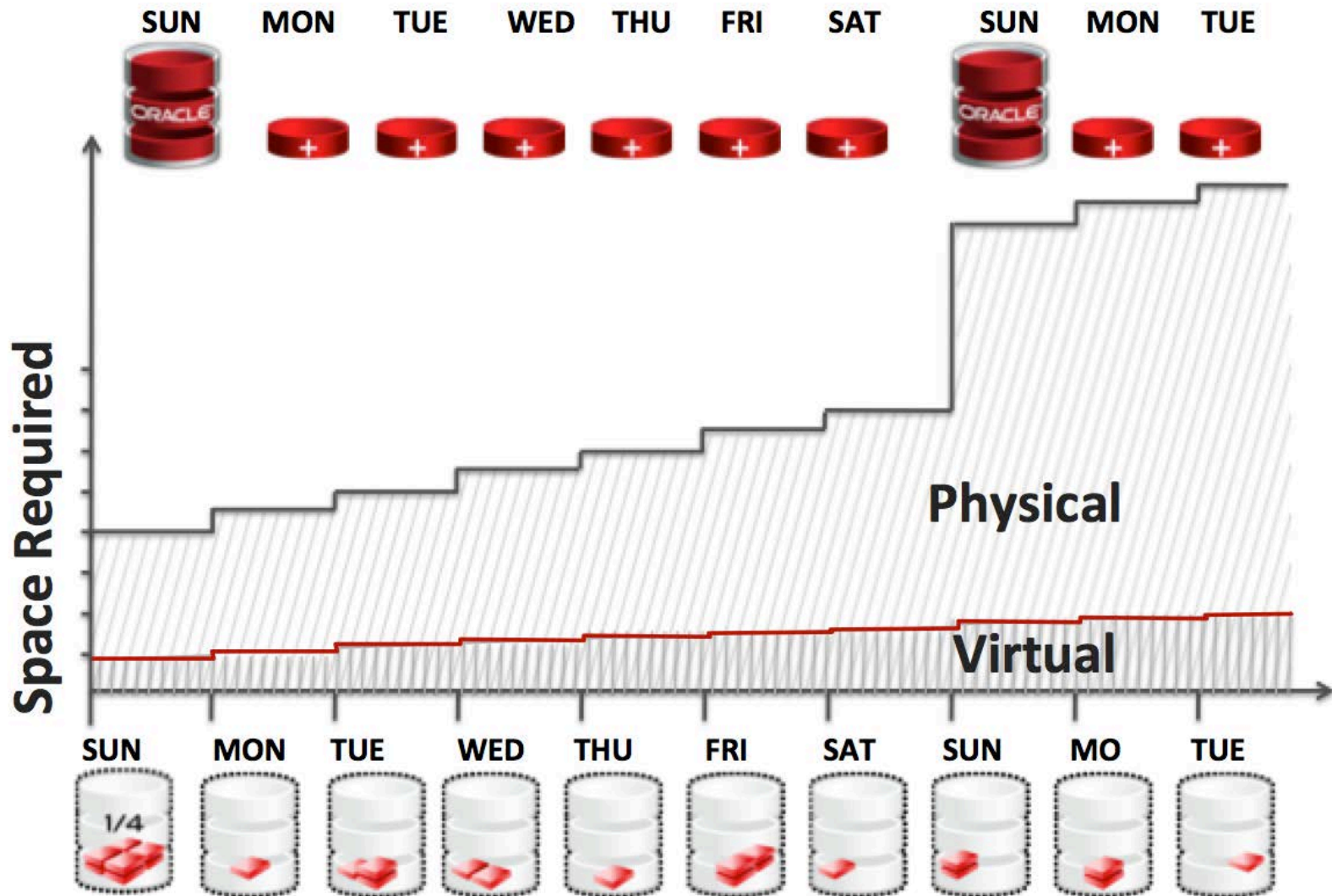
Quality

1. Prod & Dev Backups
2. Surgical recovery
3. Recovery of Production
4. Recovery of Development
5. Bug Forensics



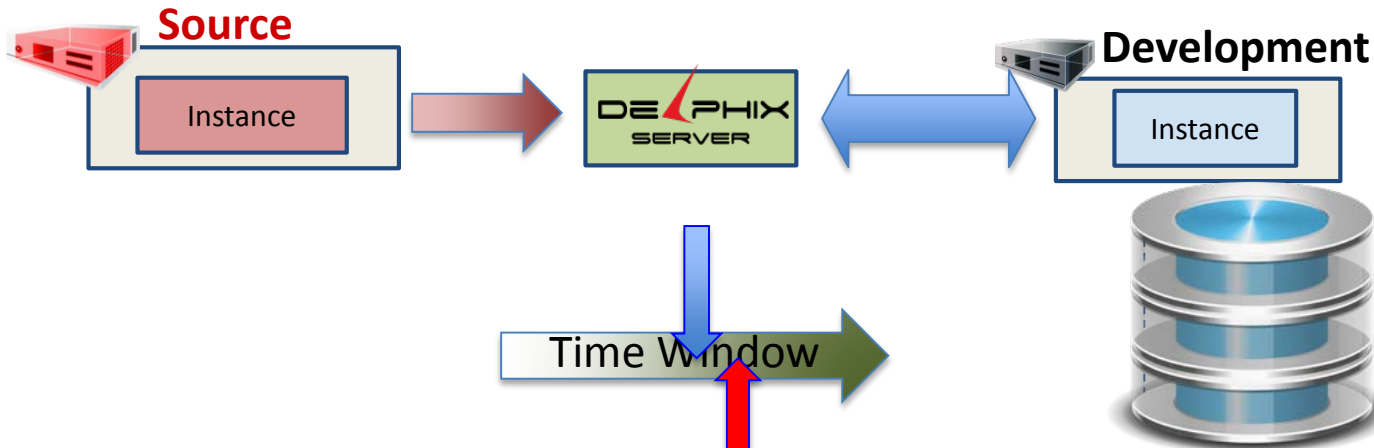


Quality : 50 days of backup in size of production



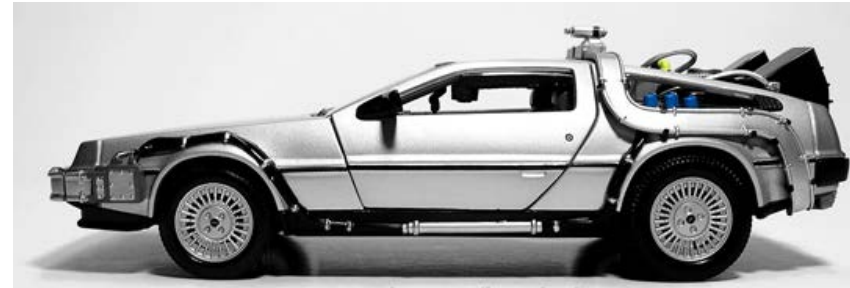


Quality : Surgical recovery



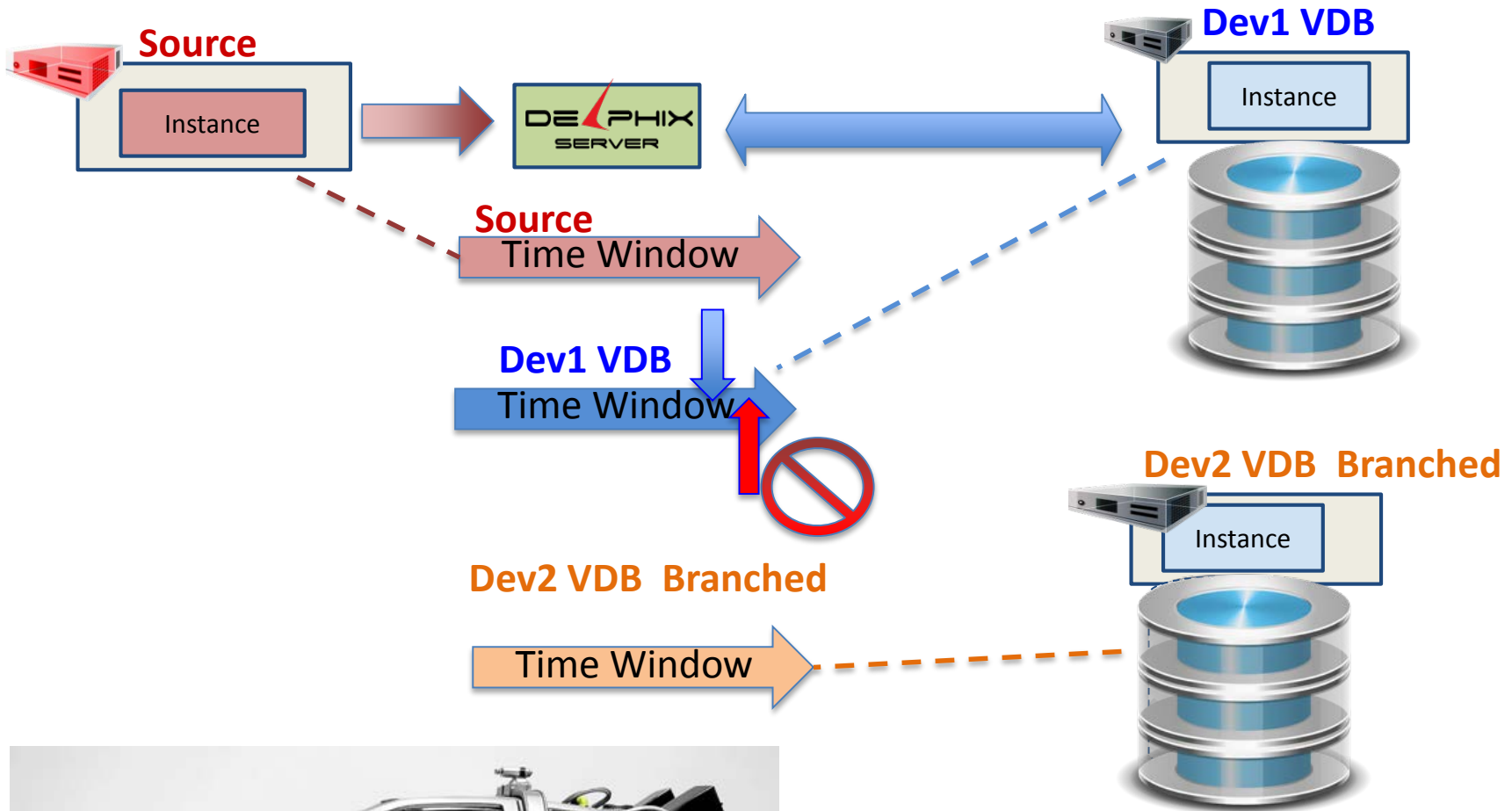

Drop


Before drop



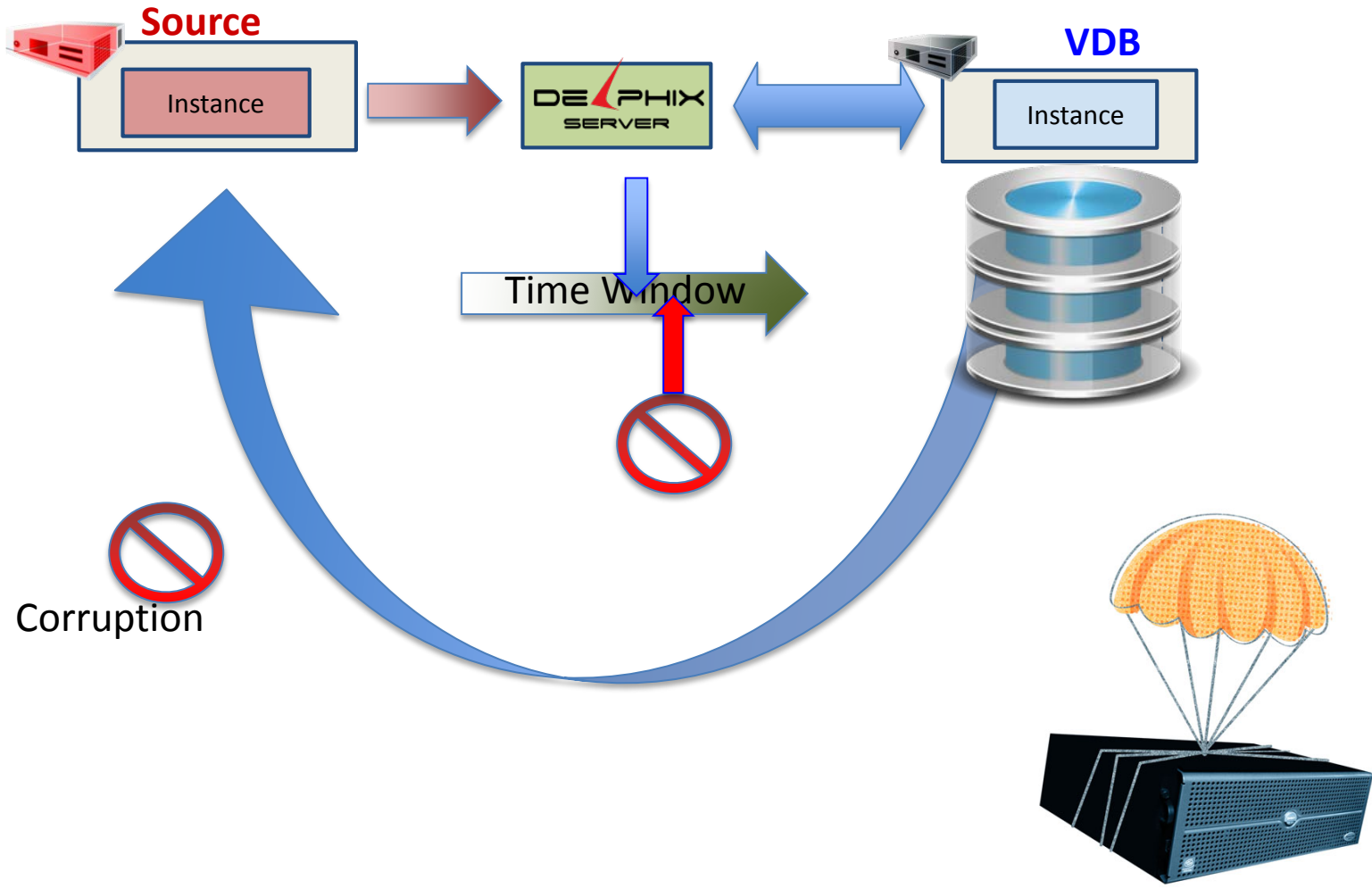


Quality: recovery of development



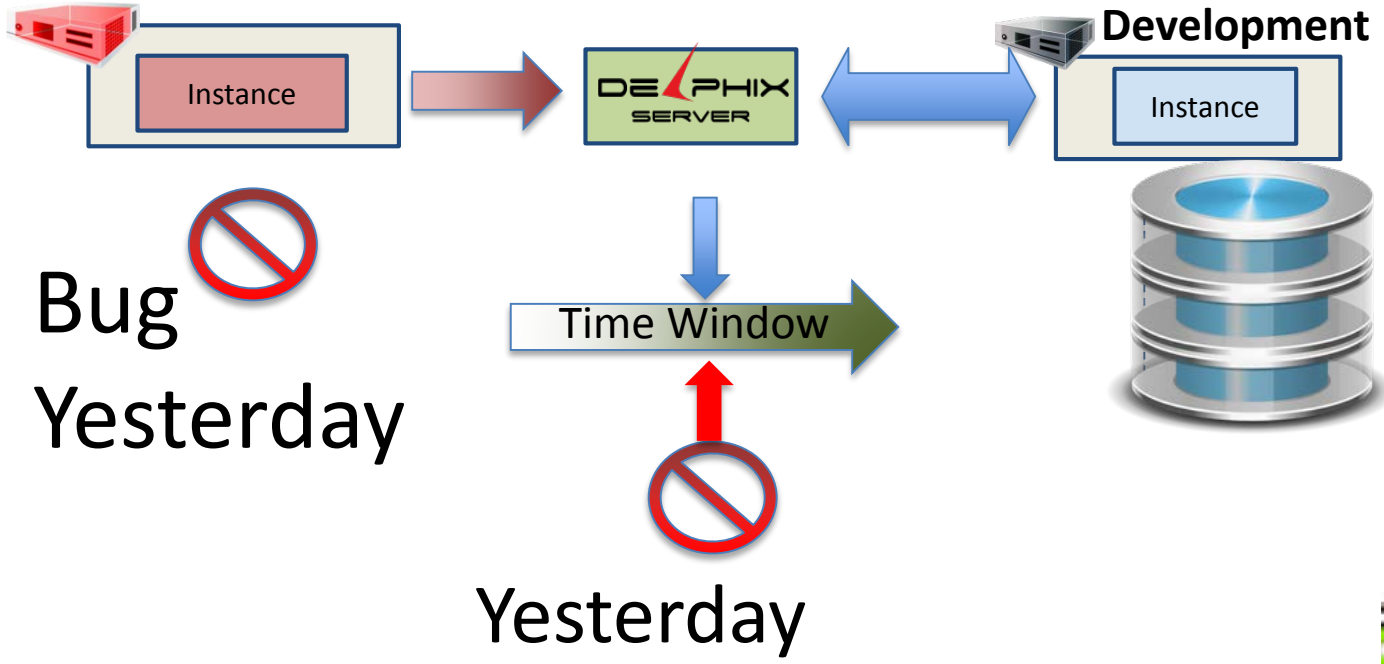


Quality : recovery of production





1. Forensics: Investigate Production Bugs



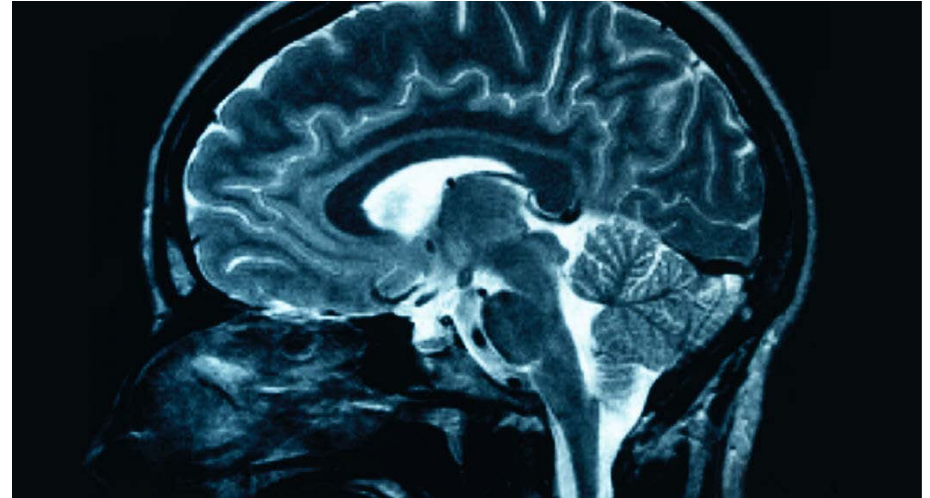
Use Cases

1. Development
2. QA
3. Quality
4. Business Intelligence
5. Modernization



Business Intelligence

- 24x7 Batches
- Low Bandwidth
- Temporal Data
- Confidence Testing



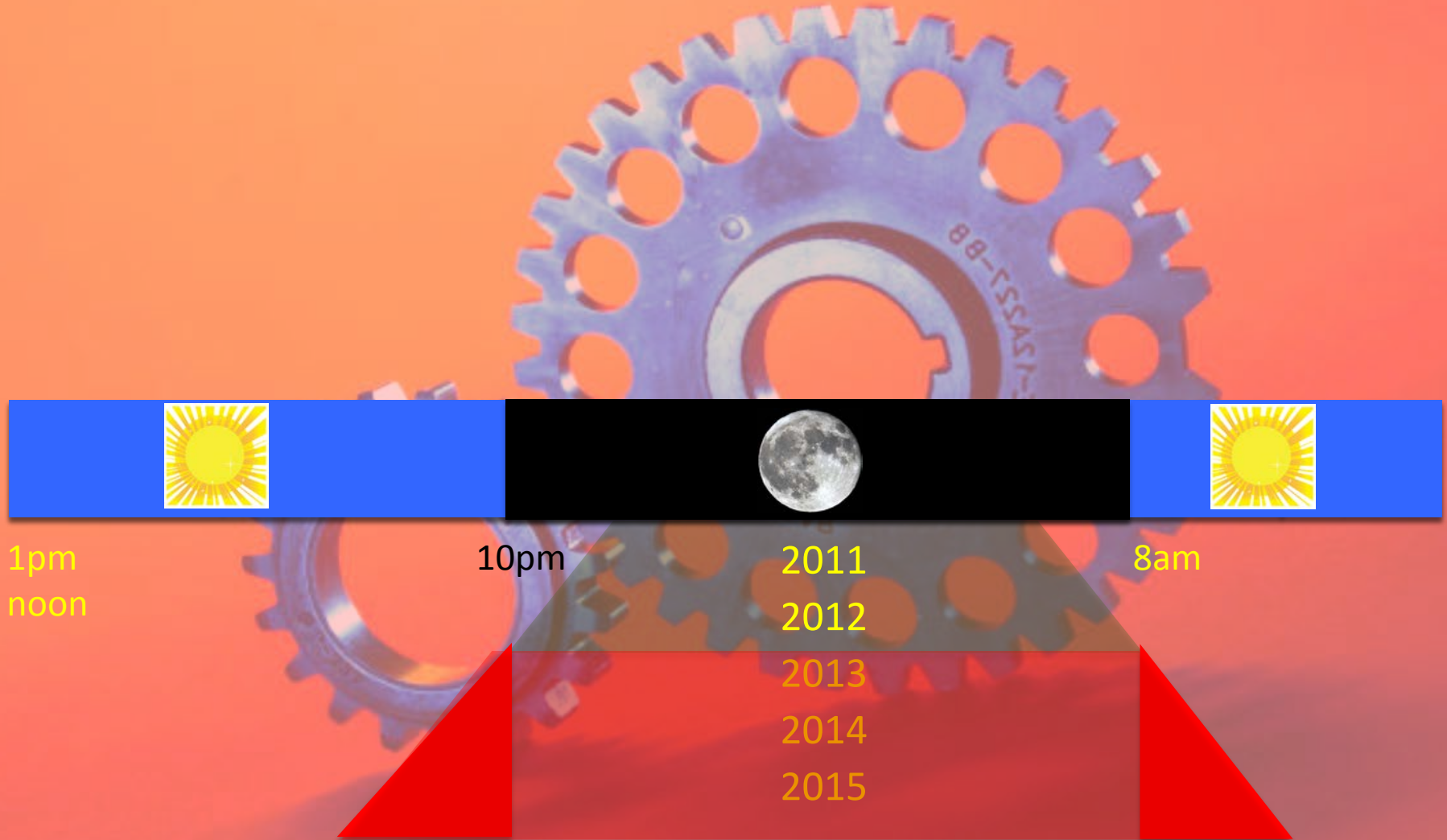


Business Intelligence: ETL and Refresh Windows



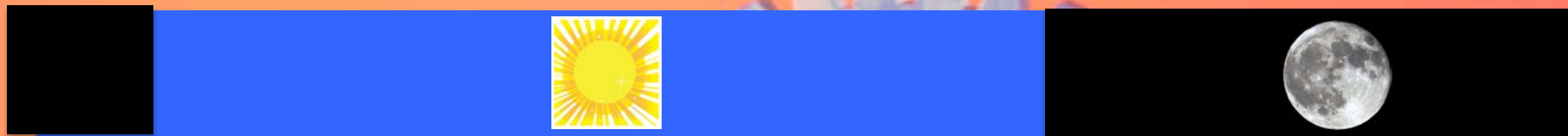


Business Intelligence: ETL and DW refreshes taking longer



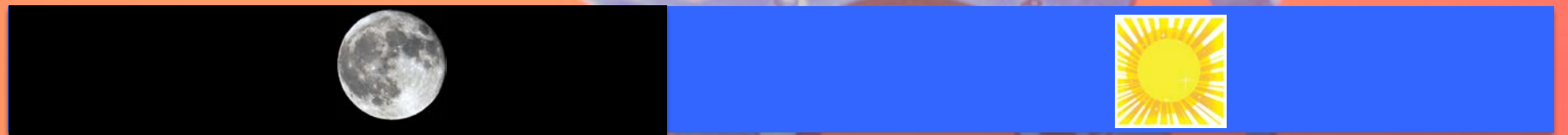


Business Intelligence ETL and Refresh Windows



6am 8am

10pm



10pm

8am

noon

9pm

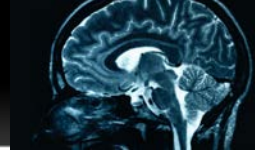


1pm
noon

10pm

8am

- 2011
- 2012
- 2013
- 2014
- 2015

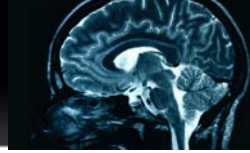


Business Intelligence: ETL and DW Refreshes



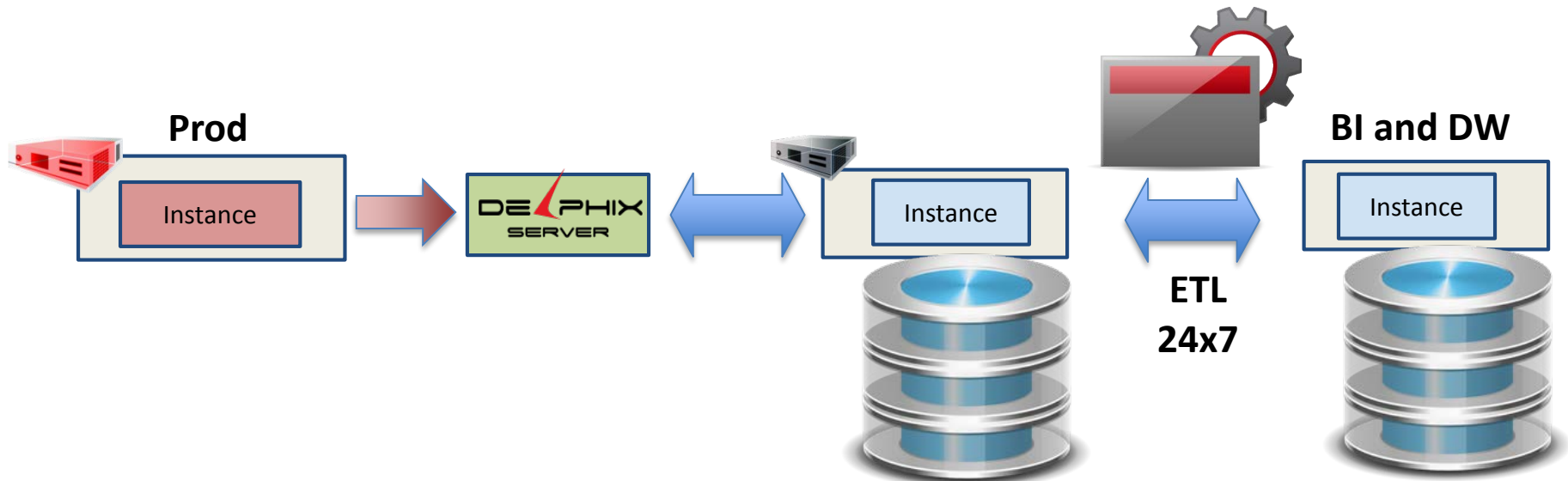
Data Guard – requires full refresh if used

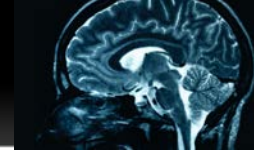
Active Data Guard – read only, most reports don't work



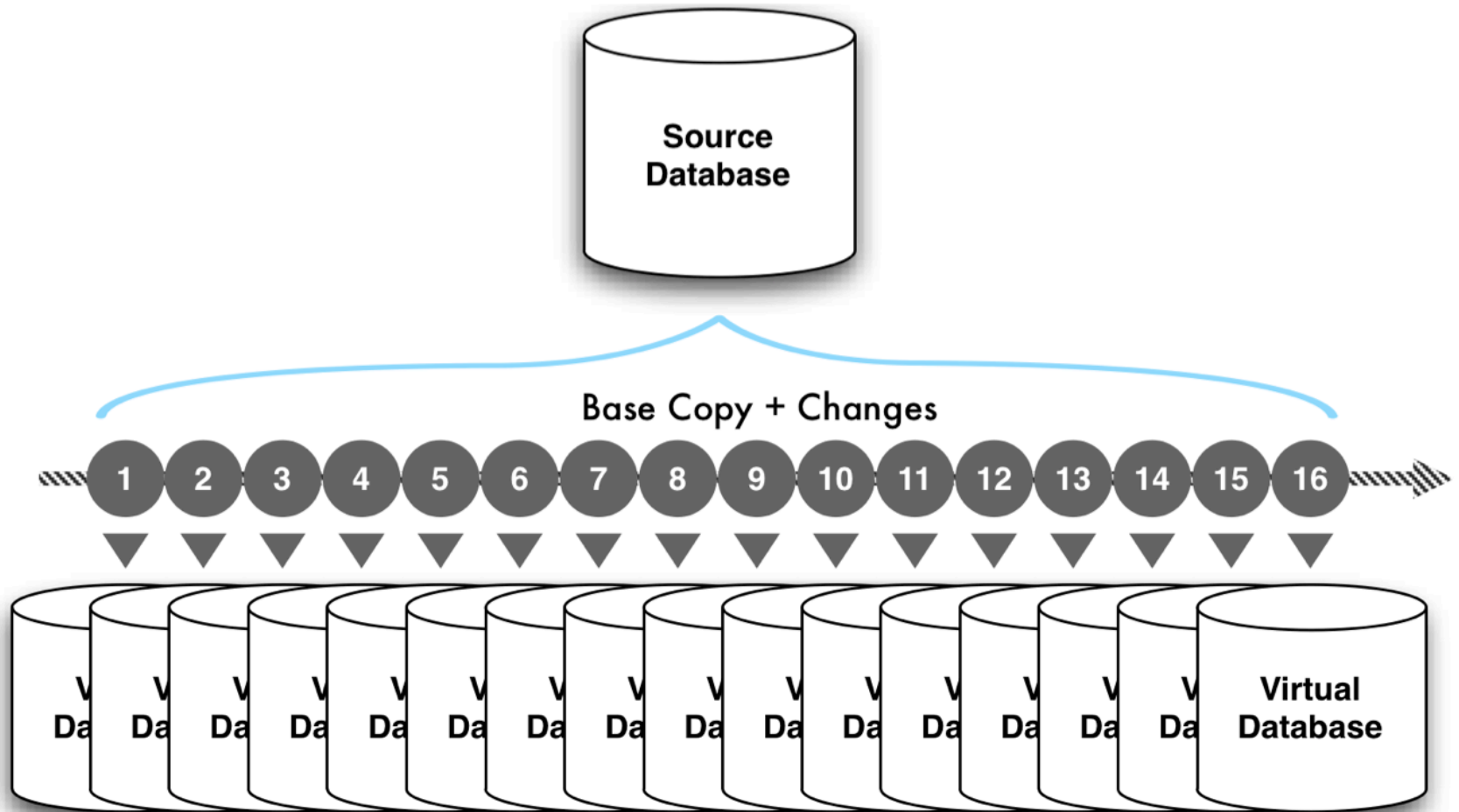
Business Intelligence: Fast Refreshes

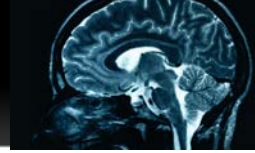
- Collect only Changes
- Refresh in minutes





Business Intelligence: Temporal Data



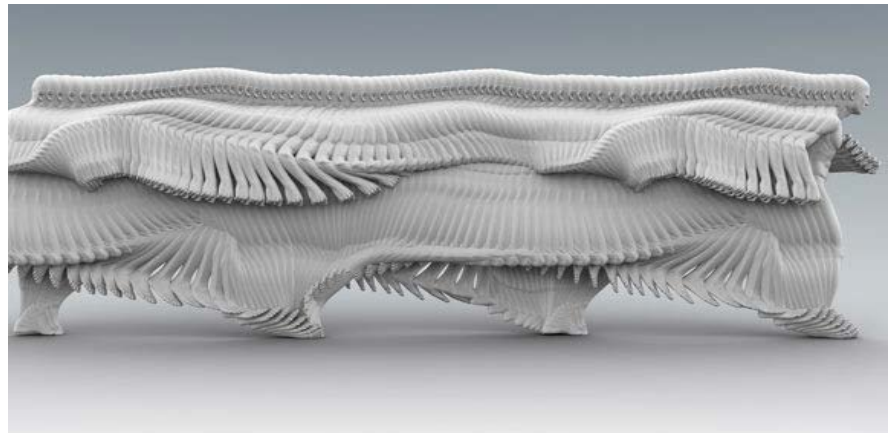


Business Intelligence

a) 24x7 Batches & Refreshes



a) Temporal queries



b) Confidence testing



Use Cases

1. Development
2. QA
3. Quality
4. Business Intelligence
5. Modernization



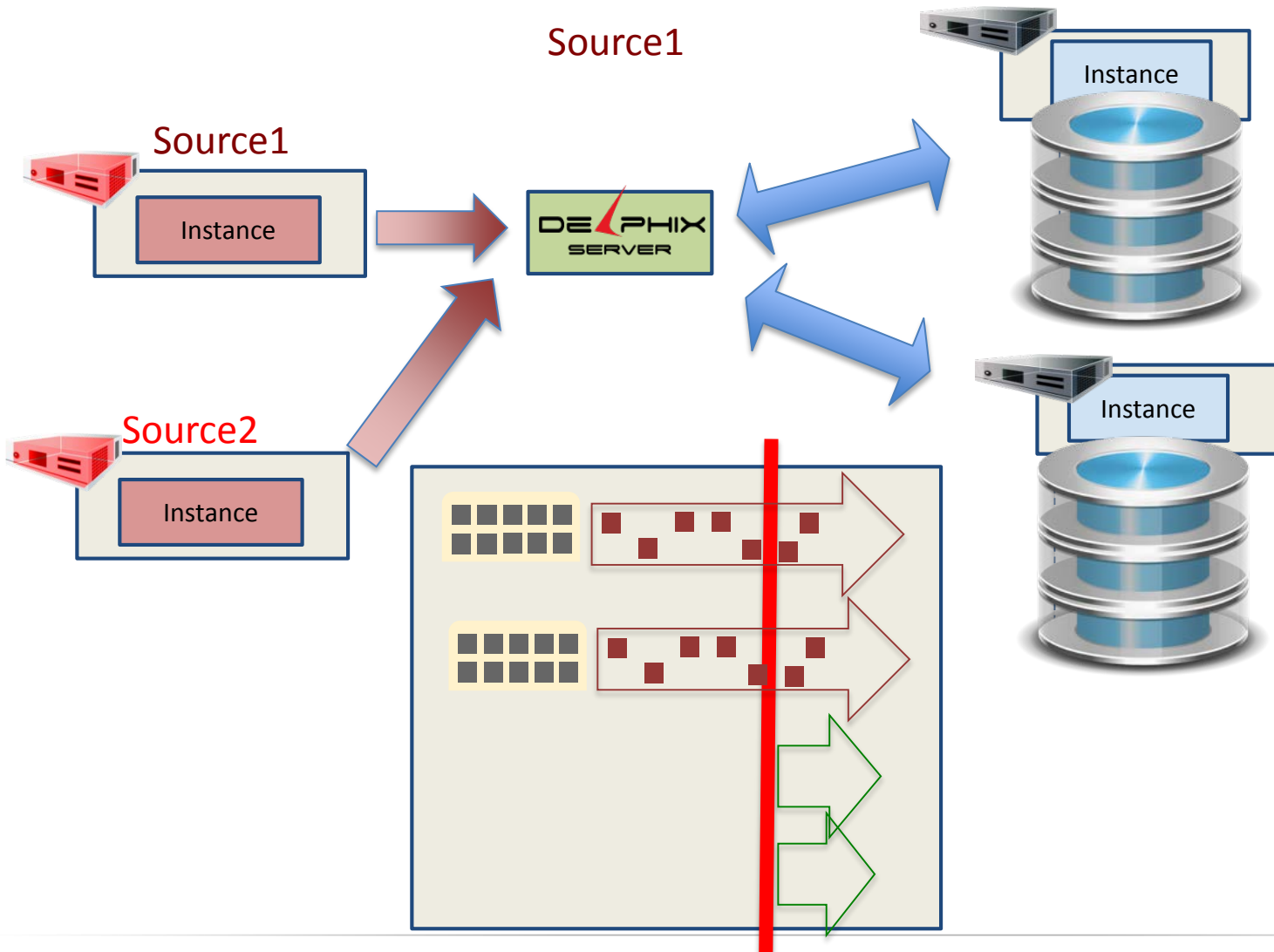
Modernization

1. Federated
2. Consolidation
3. Migration
4. Auditing





Modernization: Federated





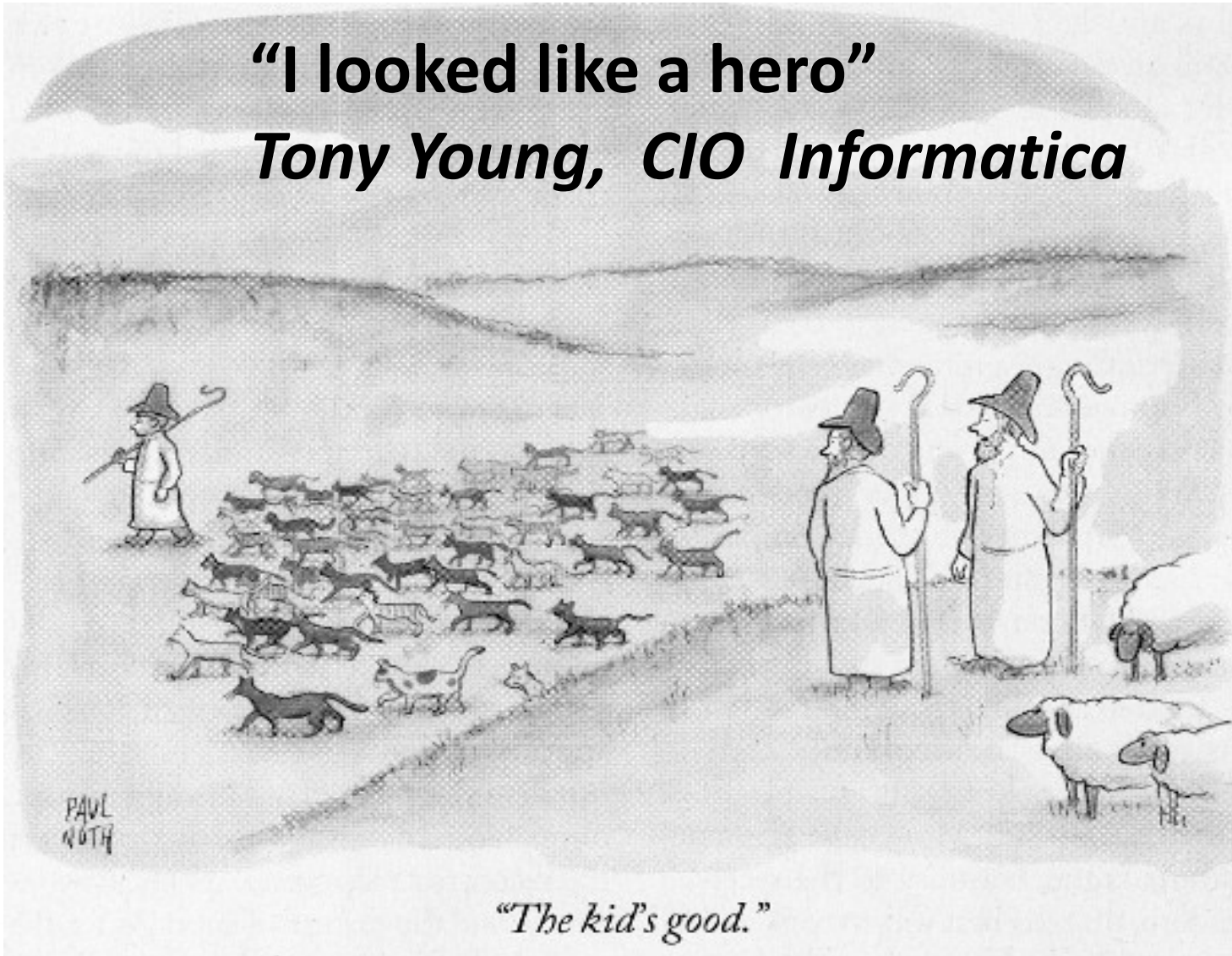
Modernization: Federated





Modernization: Federated

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

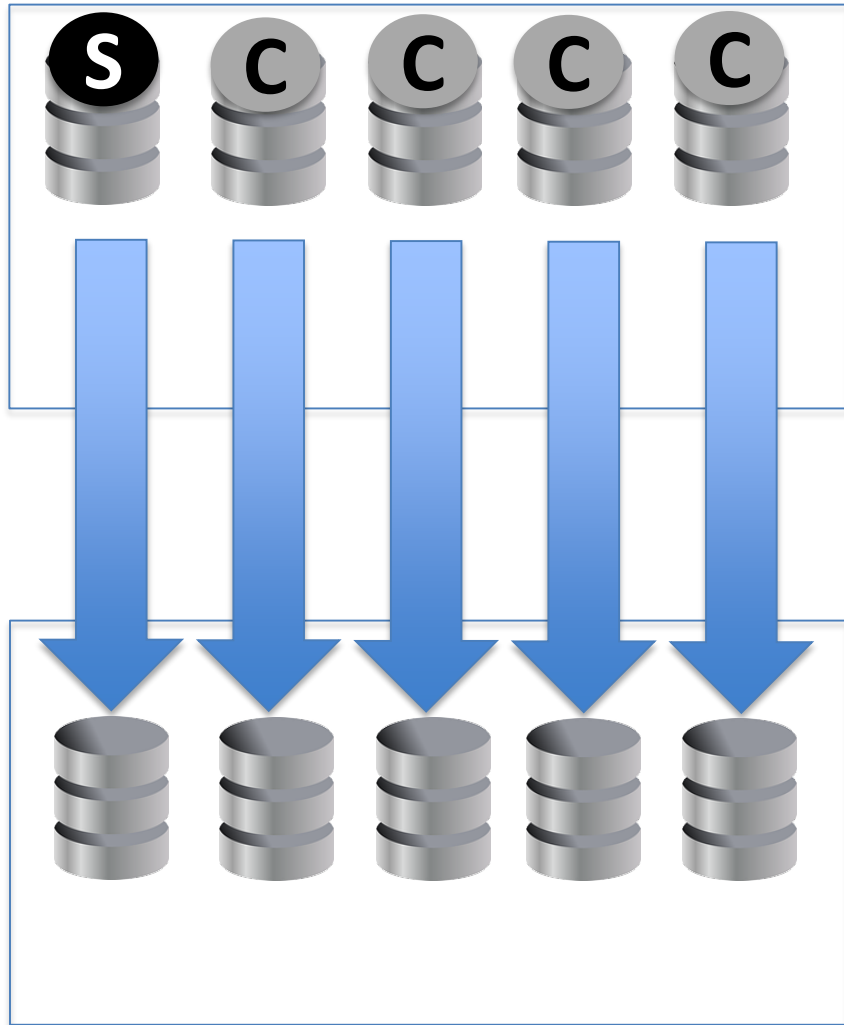


“The kid’s good.”

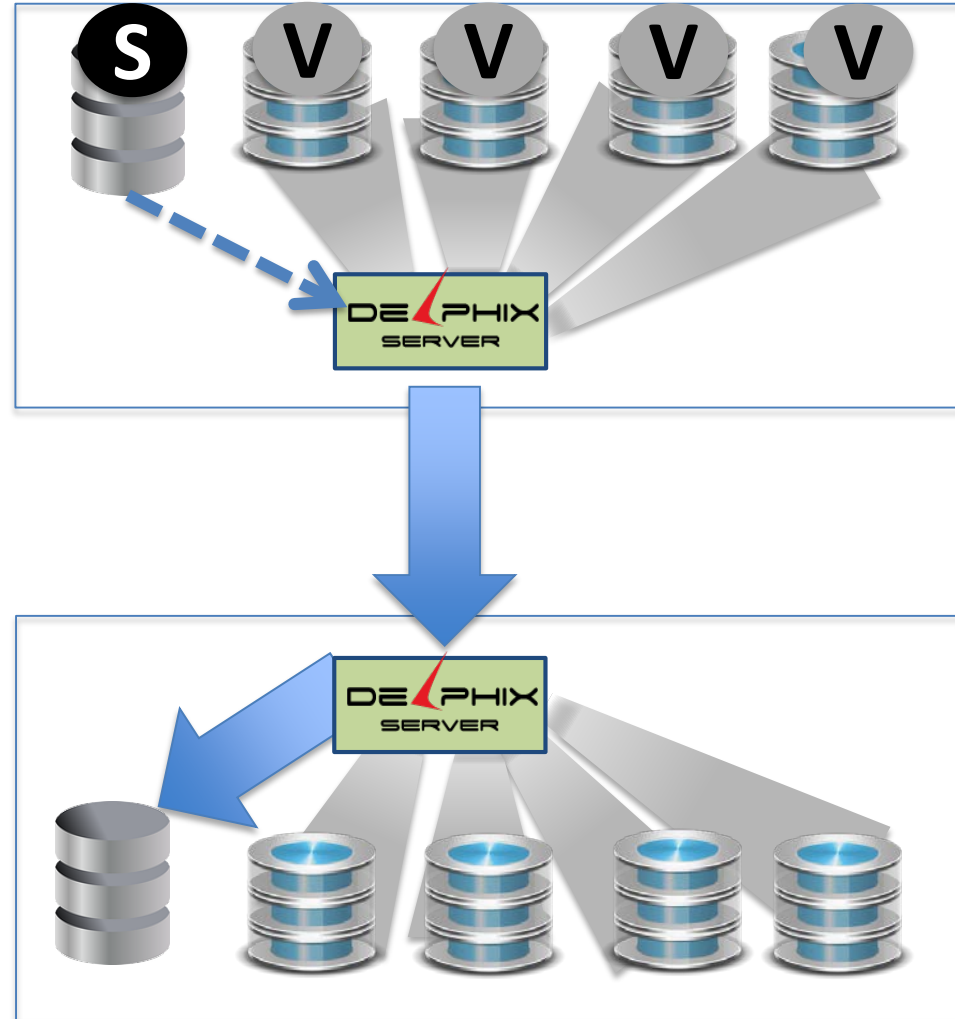




Modernization: Data Center Migration



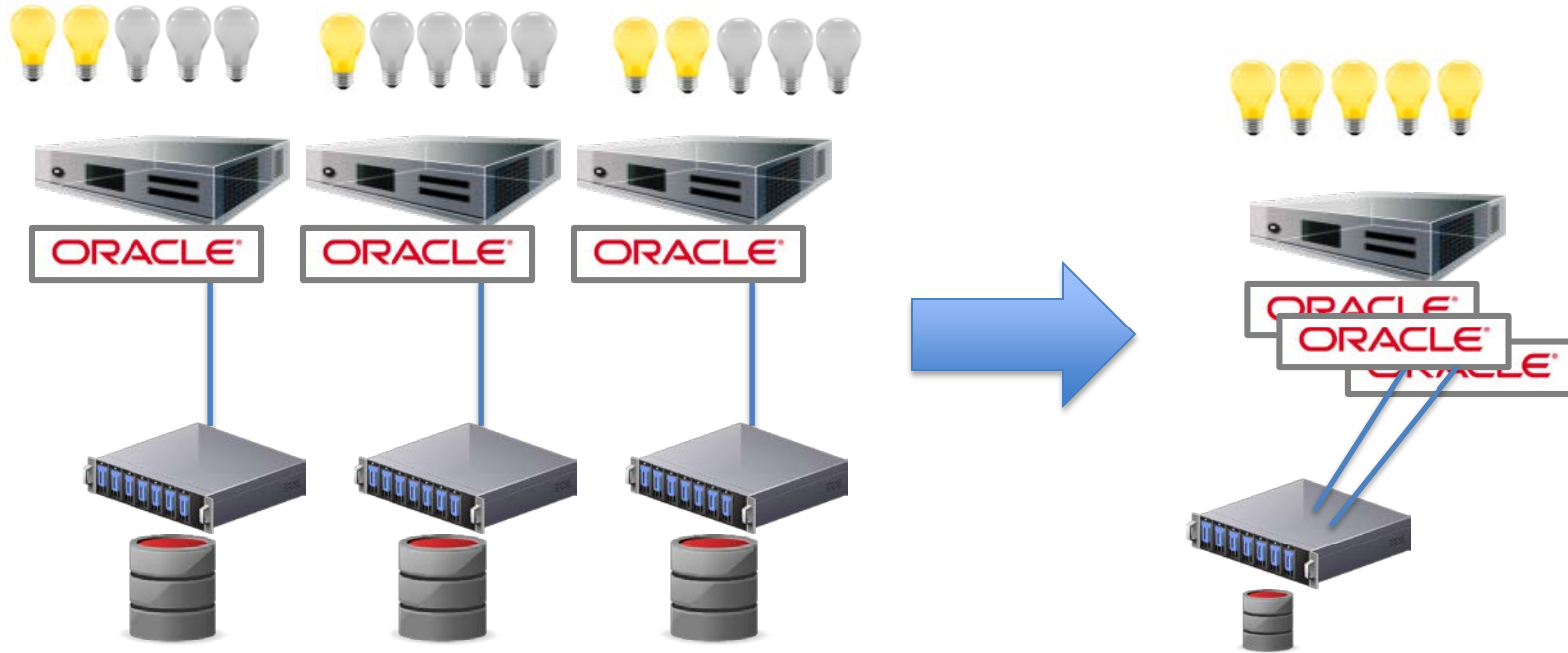
5x Source Data Copy



< 1 x Source Data Copy



Modernization: Consolidation

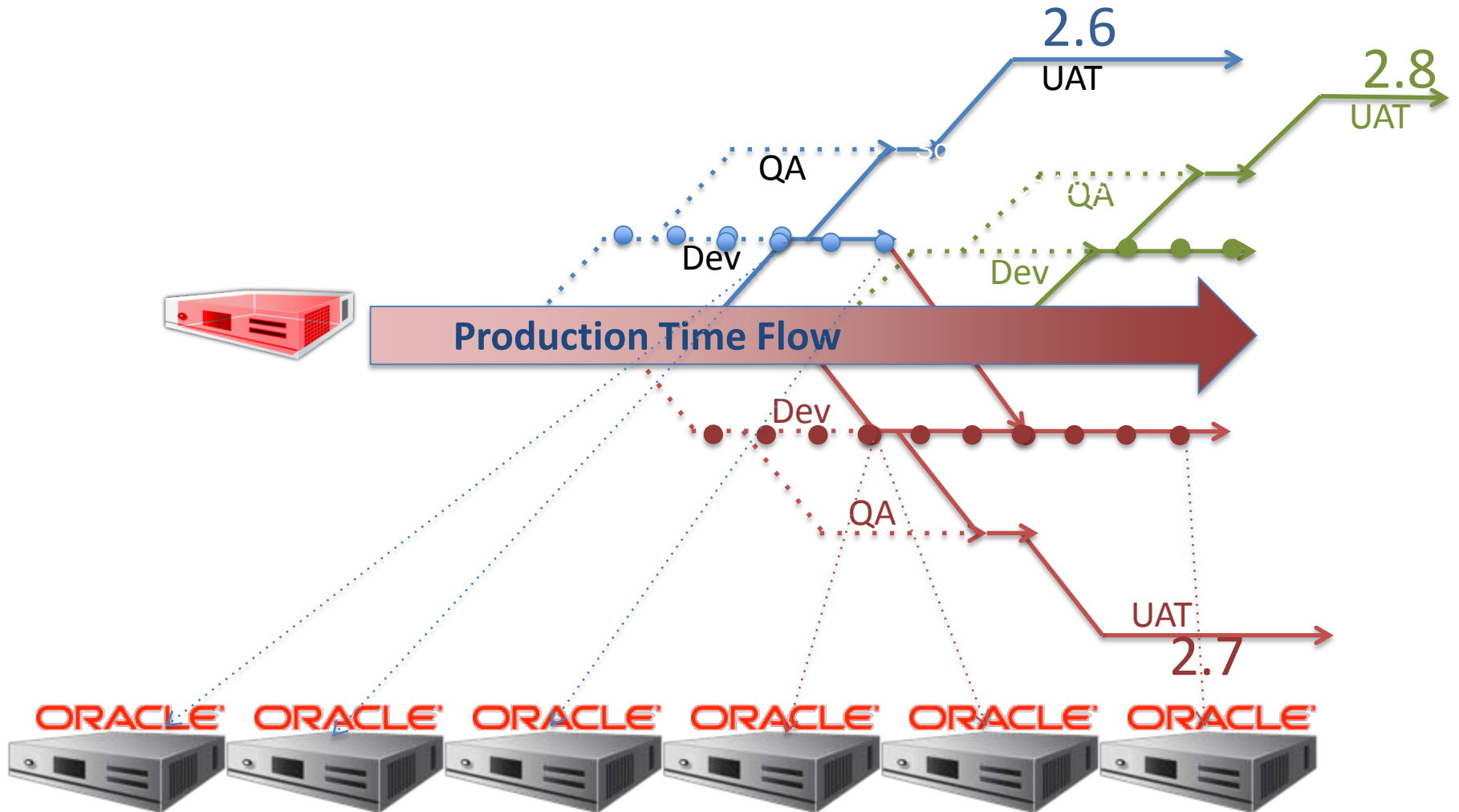


Without Delphix

With Delphix



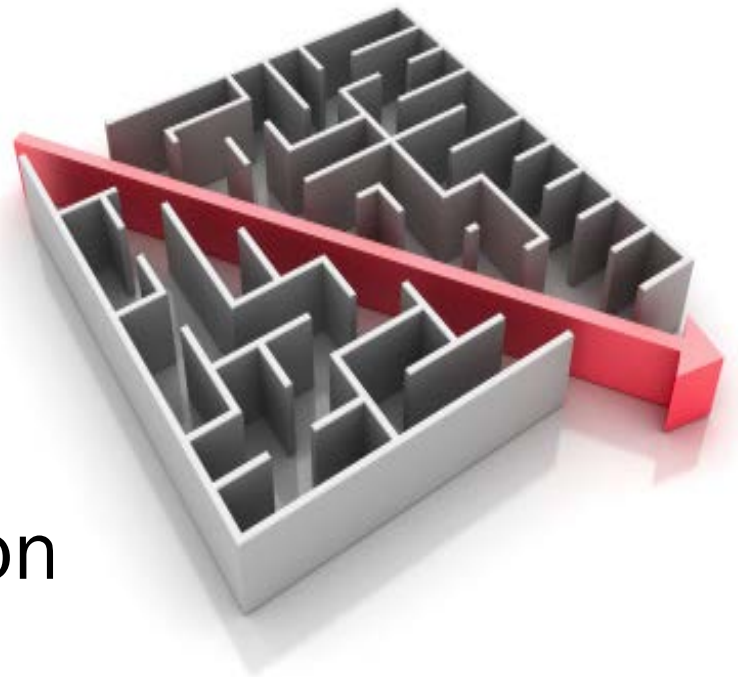
Modernization: Auditing & Version Control



Data Control = Source Control for the Database

Use Case Summary

1. Development
2. QA
3. Quality
4. Business Intelligence
5. Performance Acceleration





How expensive is the Data Constraint?

Measure before and after Delphix w/ Fortune 500 :

Median App Dev throughput increase by 2x



How expensive is the Data Constraint?

- 10 x Faster Financial Close
- 9x Faster BI refreshes
- 2x faster Projects
- 20 % less bugs



Agile Data Quotes

- “Allowed us to shrink our project schedule from **12 months to 6 months.**”
 - BA Scott, NYL VP App Dev
- "It used to **take 50-some-odd days** to develop an insurance product, ... Now we can get a product to the customer in **about 23 days.**”
 - Presbyterian Health
- “**Can't imagine working without it**”
 - Ramesh Shrinivasan CA Department of General Services



Summary

- Problem: Data is the constraint
- Solution: Agile data is small & fast
- Results: Deliver projects
 - Half the Time
 - Higher Quality
 - Increase Revenue

Kyle@delphix.com

kylehailey.com

slideshare.net/khailey

Future

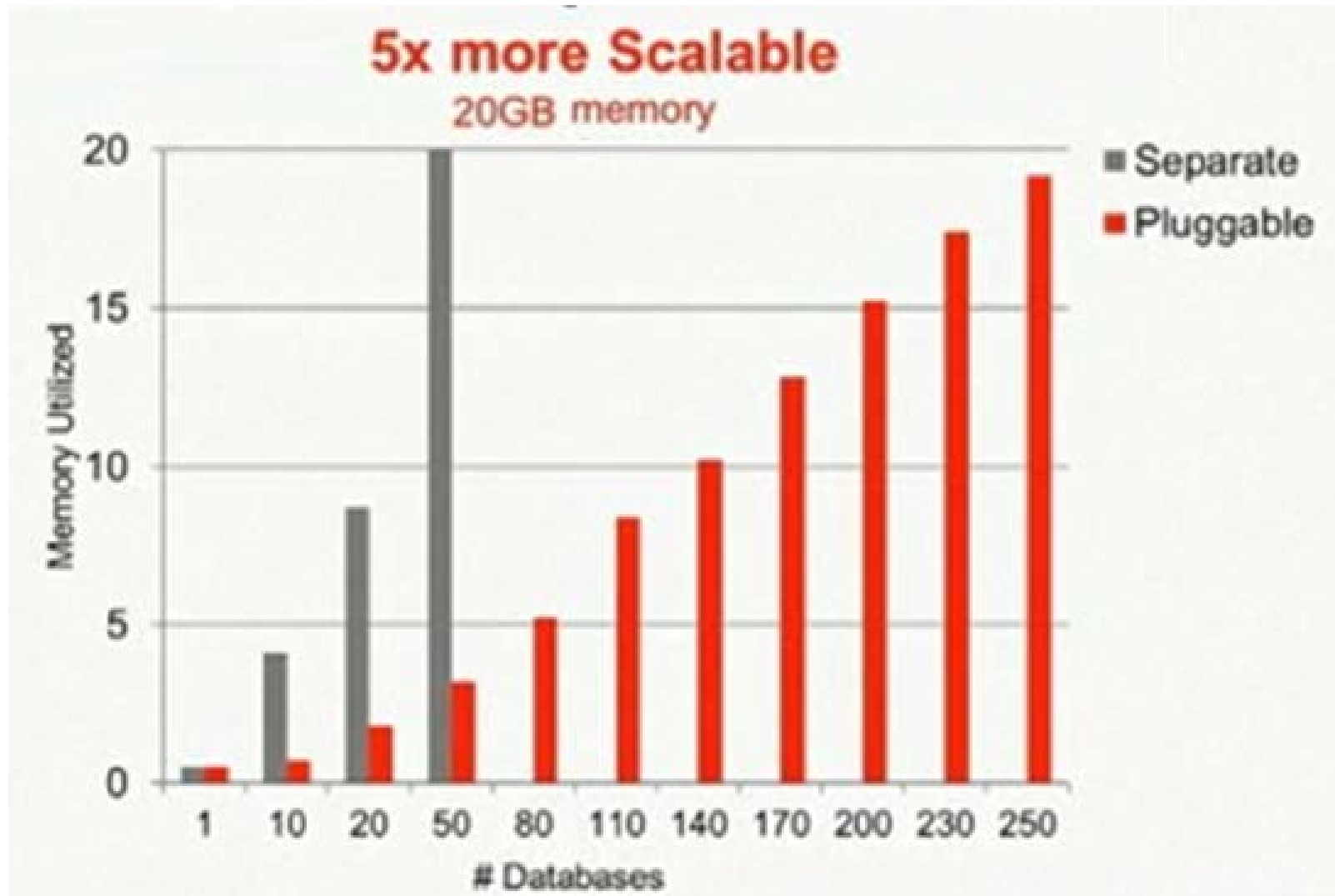
Now

- Application Stack Cloning
- Cross Platform Cloning : UNIX -> Linux
- Postgres

Coming

- VM cloning
- Workflows
 - Chef, Puppet, etc workflows for virtual data provisioning
- Developer workspaces
 - Check out, check in, bookmark, tagging, rollback, refresh
- Secure Data
 - Masking
- More Databases
 - MySQL, Sybase, DB2, Hadoop, Mongo, Cassandra
- DR and HA

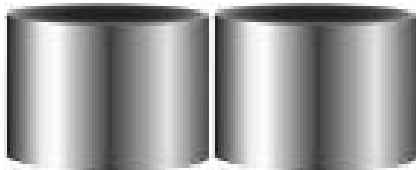
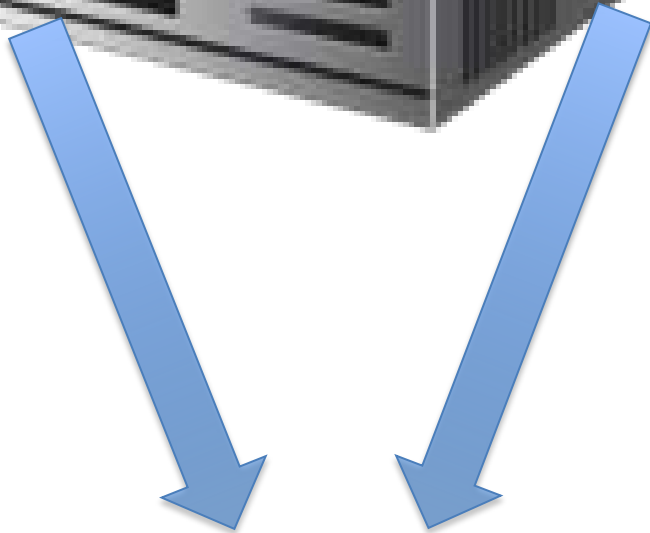
Oracle 12c



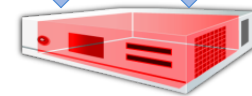
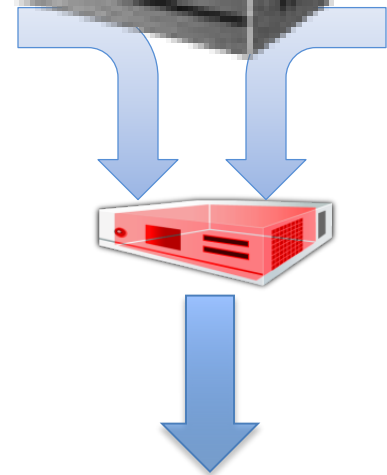
80MB buffer cache ?



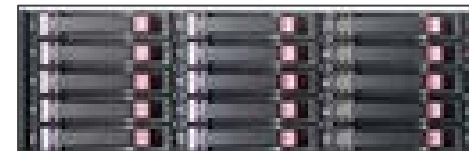
ORACLE®

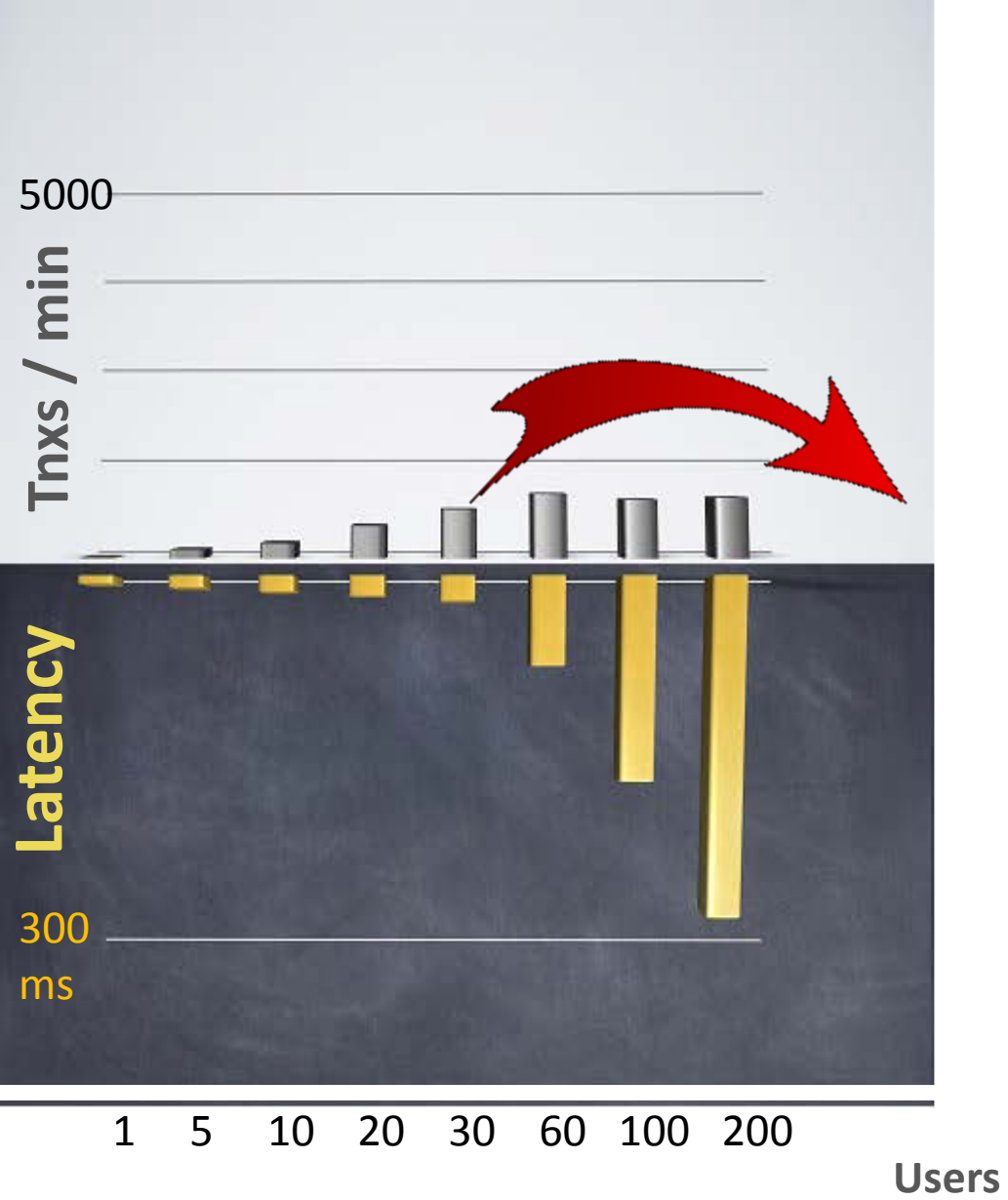


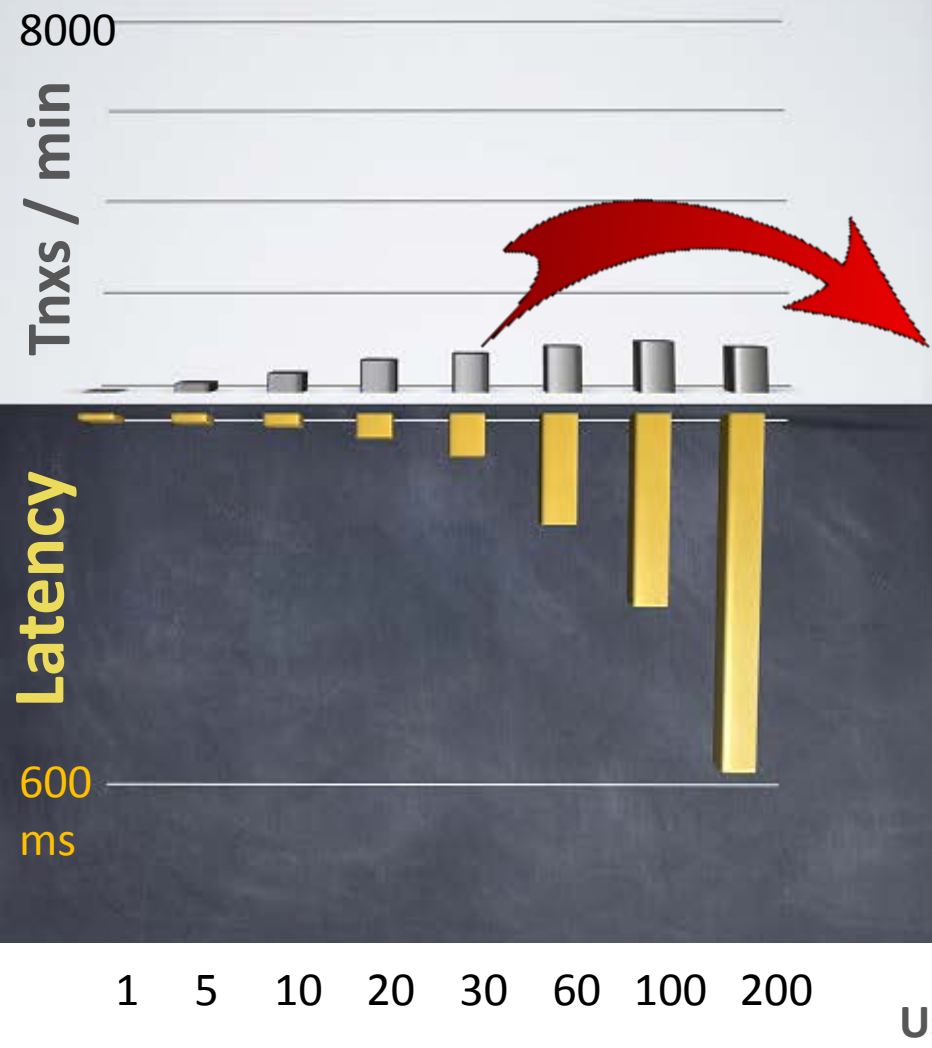
ORACLE®



200GB
Cache







Five 200GB database copies are cached with :

\$1,000,000
1TB cache on SAN

\$6,000
200GB shared cache on Delphix

