DELPHIX

Screw DevOps,
Let's Talk DataOps

Kellyn Pot'Vin-Gorman | Technical Intelligence Manager| February, 2018

Kellyn Pot'Vin-Gorman

Technical Intelligence Manager, Delphix

- Multi-platform DBA, (Oracle, MSSQL, MySQL, Sybase, PostgreSQL, Informix…)

- Oracle ACE Director, (Alumni)

- Oak Table Network Member

- Idera ACE 2018

- APEX Women in Technology Award, CTA

- STEM education with Raspberry Pi and Python, including DevOxx4Kids, Oracle Education Foundation and TechGirls

- President, Rocky Mtn Oracle User Group

- President, Denver SQL Server User Group

- DevOps author, instructor and presenter.

- Author, blogger, (http://dbakevlar.com)

DevOps + DBA = DataOps

- Database Administrators heavily influence decisions.

- Adverse to bleeding edge, high risk.

- Introduction of the cloud, especially and SaaS, (Software as a Service).

- Demand for decreased development cycles introduce risk, DBAs are seen as a roadblock to this.

Still, DBA skills are in high demand towards DevOps initiatives…

DataOps- Origin

DataOps takes DevOps to the next level, recognizing that many DevOps projects have data integrated into them and requires that data to move at the same speed the rest of development and testing.

Let's Talk Data Gravity
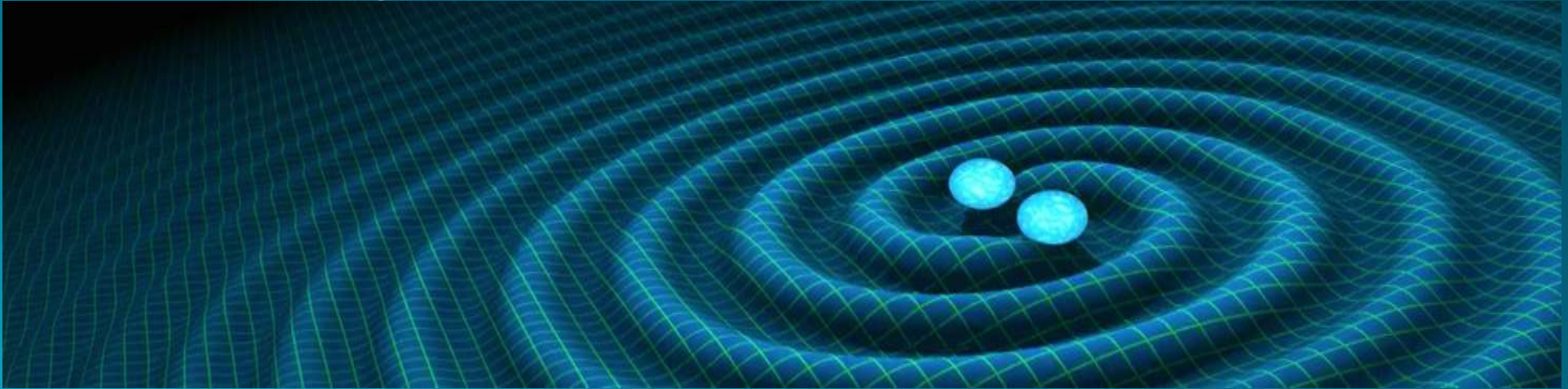
What is Data Gravity?

Data Gravity is the ability for data to attract applications, services, etc. As with the laws of physical gravity, data, due to its mass, will attract all else that has less mass.
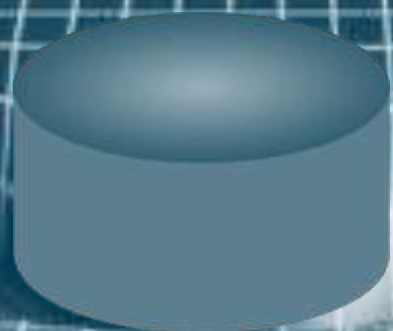
This is Data Gravity

How much data are we talking about?

- More data has been created in just the last two years than the previous history of humanity
- 1.7MB of new information created every second per human on the planet.
- Introduction of big data often has same development pain points.

From DevOps to DataOps

Dave is a Developer New to Agile

His development team is about to start Scrum as their Agile Methodology

It will be their first Scrum "sprint"
- Looking to increase productivity 20%
- 25% increase in collaboration.
- Gains in revenue are predicted.

And they're looking forward to succeeding…

The Company has hired George as the new Scrum Master

The Periodic Table of DevOps Tools

George creates the burn down list for the sprint…

Scrum

Product Backlog

Sprint Planning

Sprint Backlog

Daily Scrum

Sprint

Burndown Chart

Task Board
To Do | In Progress | Done

Sprint Review & Retro

Working Software

Team   George   Dave
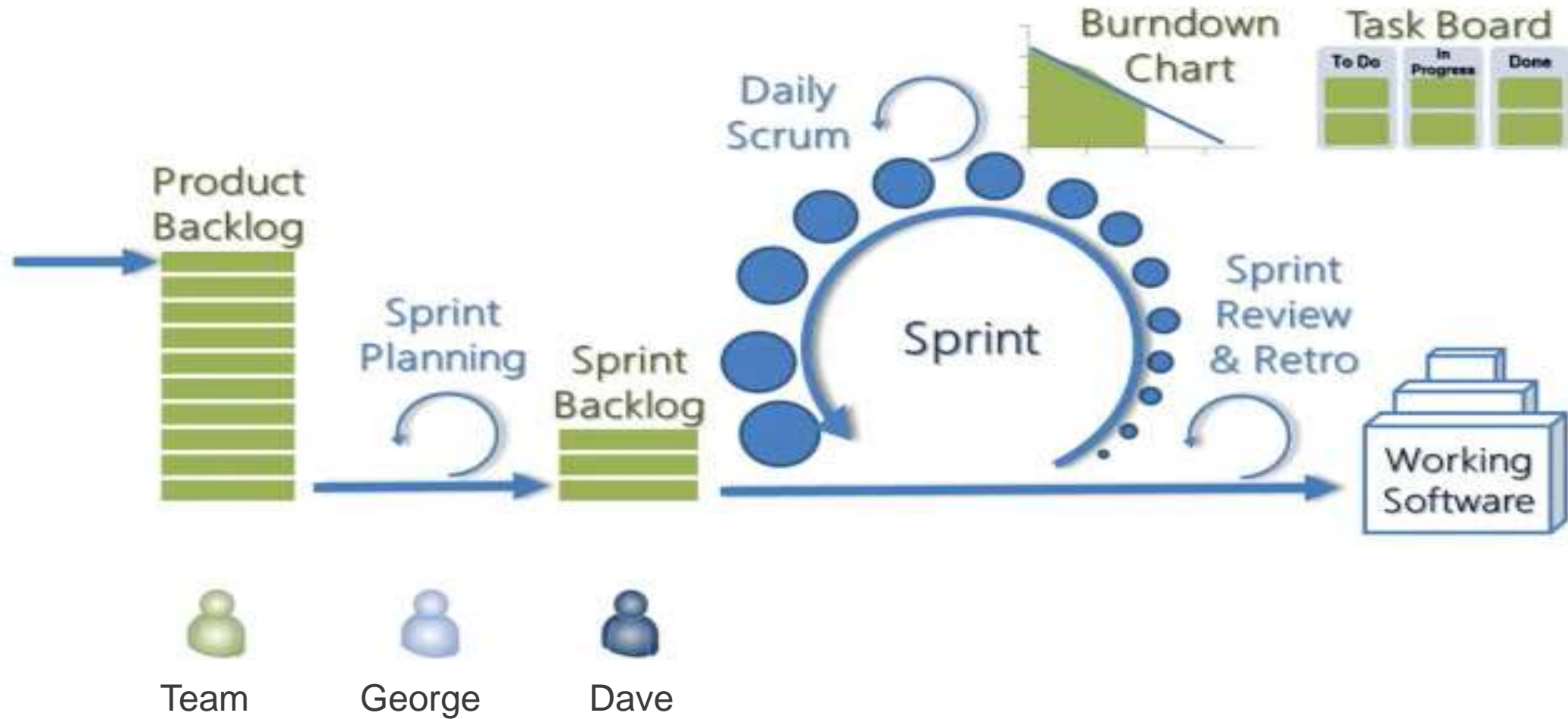
George is heading up the daily scrum meetings with the team…
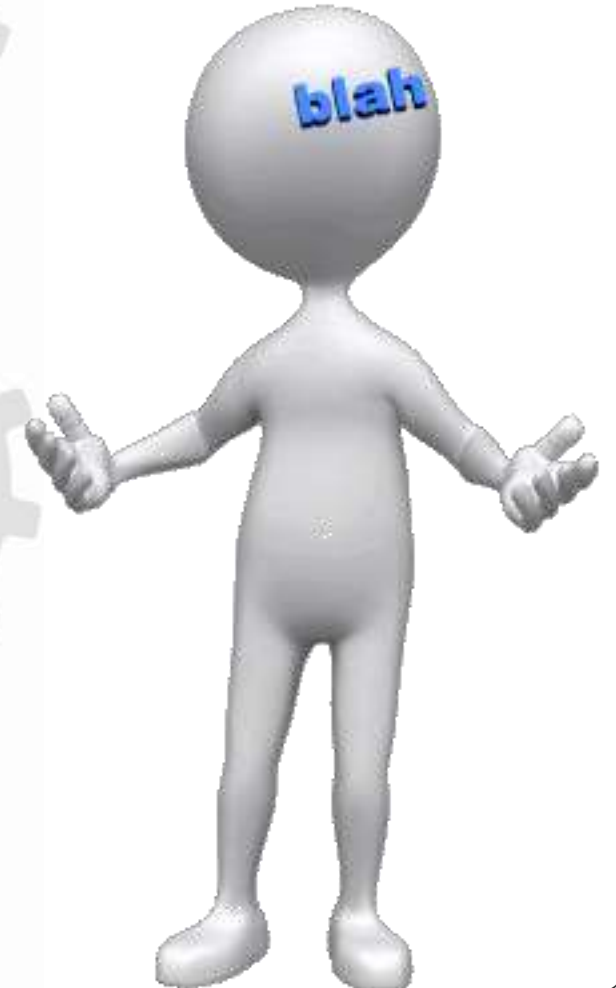
Problem is, the team all needs their own development and testing environments.

And Dave Needs to Test His New Script...

```
- name: Transfer and execute a script.
  hosts: server
  remote_user: test_user
  sudo: yes
  tasks:
      - name: Transfer the script
        copy: src=test.sh dest=/home/test_user
mode=0777

      - name: Execute the script
        command: sh /home/test_user/test.sh
```

Data is Holding Them Back
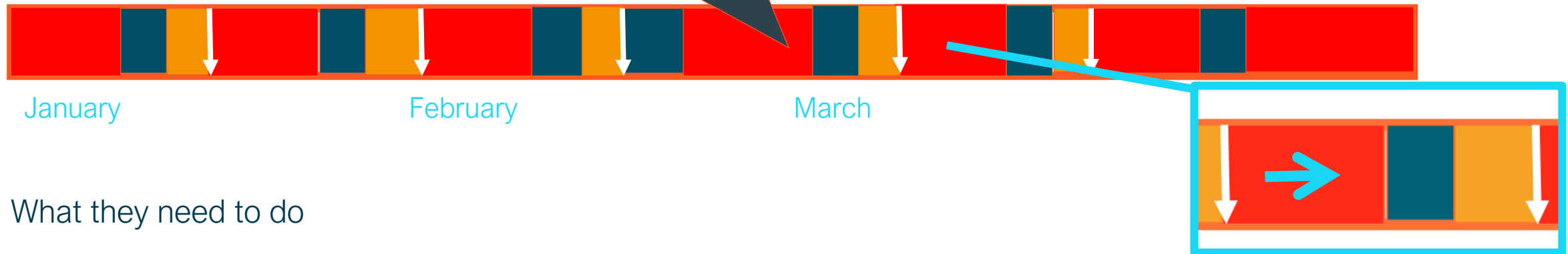
Without a self-service portal, Dave and the team is also chained to the DBAs…

The Natural Life of Data is Growth

- DBAs are busier
- Development cycle shortens
- Demands increases
- DBA is less available
- Previous way of performing refreshes, provisioning and supporting are no longer sustainable.

Where is George, Dave and the rest of the team going to get what they need?

Segway: the Cloud Backlash

Remember that One Outlier project?

Evolution of the Cloud in the DBAs Eyes
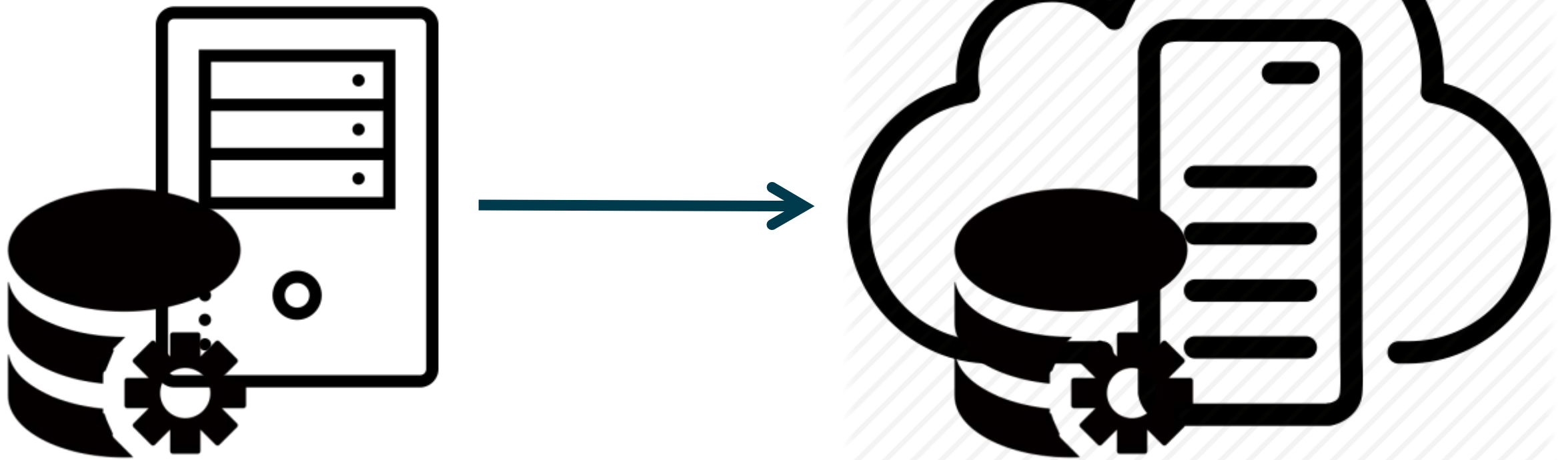
Data Causes Friction

- If those managing the data sources aren't included and working well with those needing that data to produce features and products, friction is the result.
- DBAs are working with robust, but archaic utilities that guarantee outcome, rarely speed of results.

The pain for friction is felt by everyone.

The Result are the Developers and Testers Spending Upwards of 80% of Their Time **NOT** Performing their Primary Function…
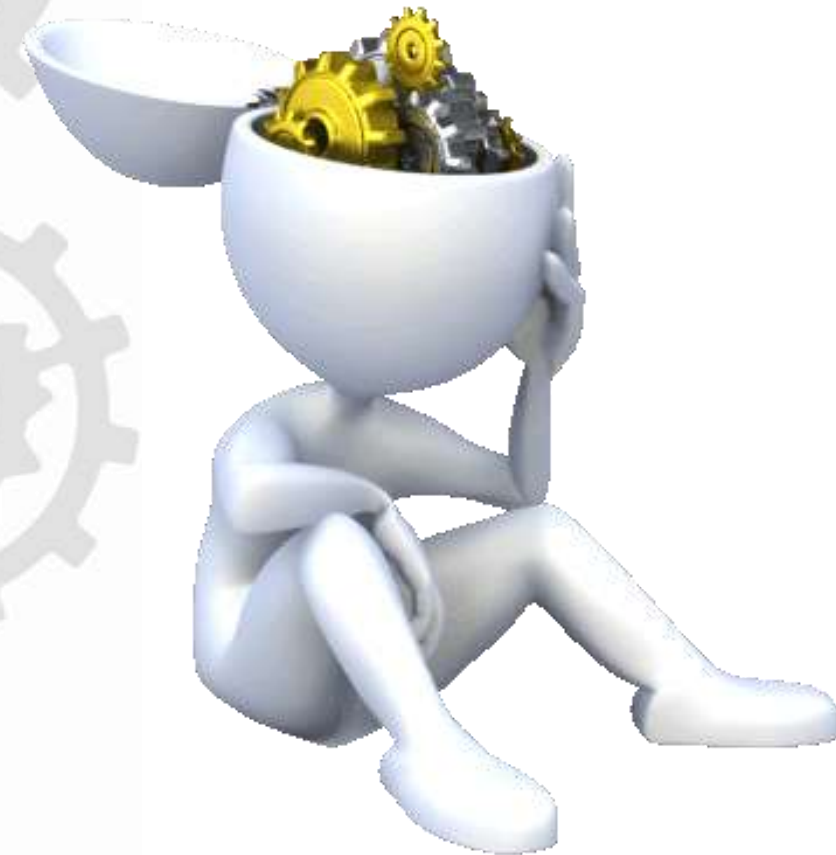
Luckily, their DBA, Marc, is keen on Database Virtualization and Containers…

1ˢᵗ- Marc Embraces Virtualization…

A technical approach in which users and applications do not use physical machines, but simulated systems running on actual, "real" hardware. Virtualization can be used to eliminate resource usage and enable savings for databases, network, file and application management, along with server infrastructure.

Older Cloning Methods
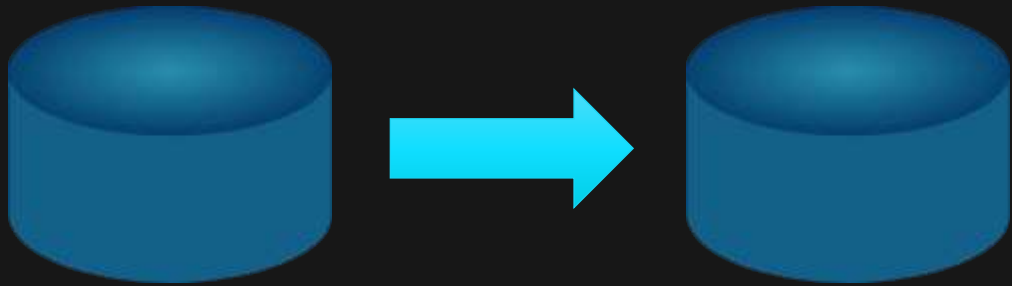
Time Consuming

Virtualizes the Data Sources

Compress and Deduplicate
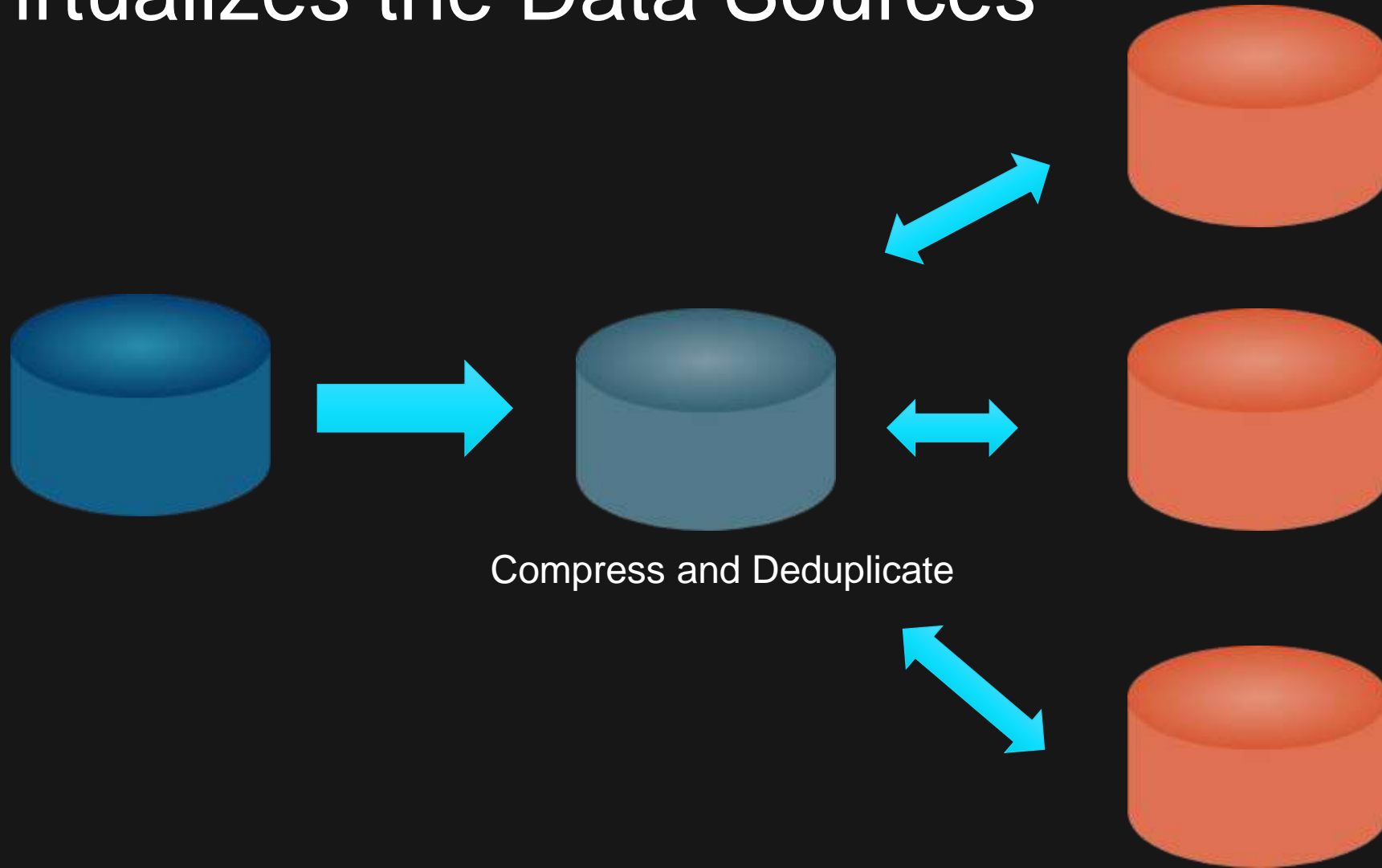
Database Virtualization Tools

- Windocks
- Redgate SQL Clone
- Rubrik SQL Mount
- Actifio
- Delphix

Why Database Virtualization is Different

- DBA is "only as good as their last backup".

- Many database tools take considerable time to recover.

- DevOps is often about automation- automating an "undo" for development and testing that includes data rewind.

- Include a self-service tool that will allow for rewind without DBA intervention.

- Allows for data version control and DataOps, the next step in DevOps
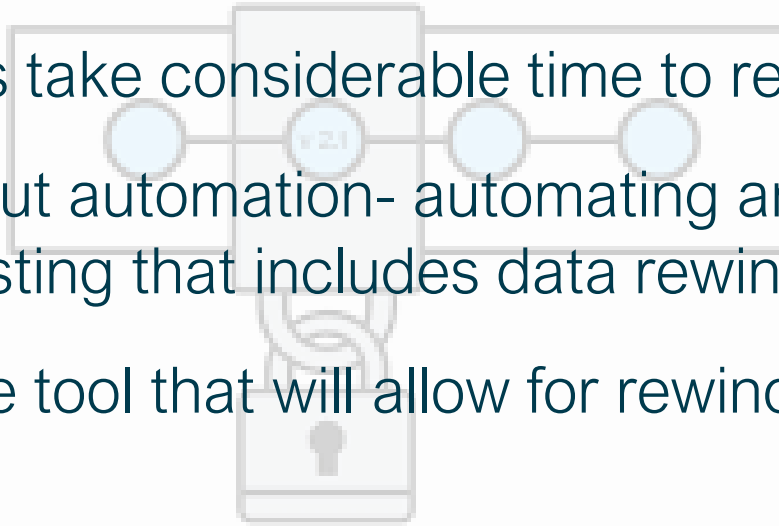
Example of Virtualized Environment- Delphix

PRODUCTION

Database/App Tier

1 TB

1 TB

Each Virtual Database takes up around 5-10Gb upon creation, (dependent upon parameters)

Storage Pool for Delphix

QA

TEST

DEV

PATCH TEST

Read From Production

◀ ◀ Virtualize and Deployed ▶

▶ ▶

600GB

Read AND Write
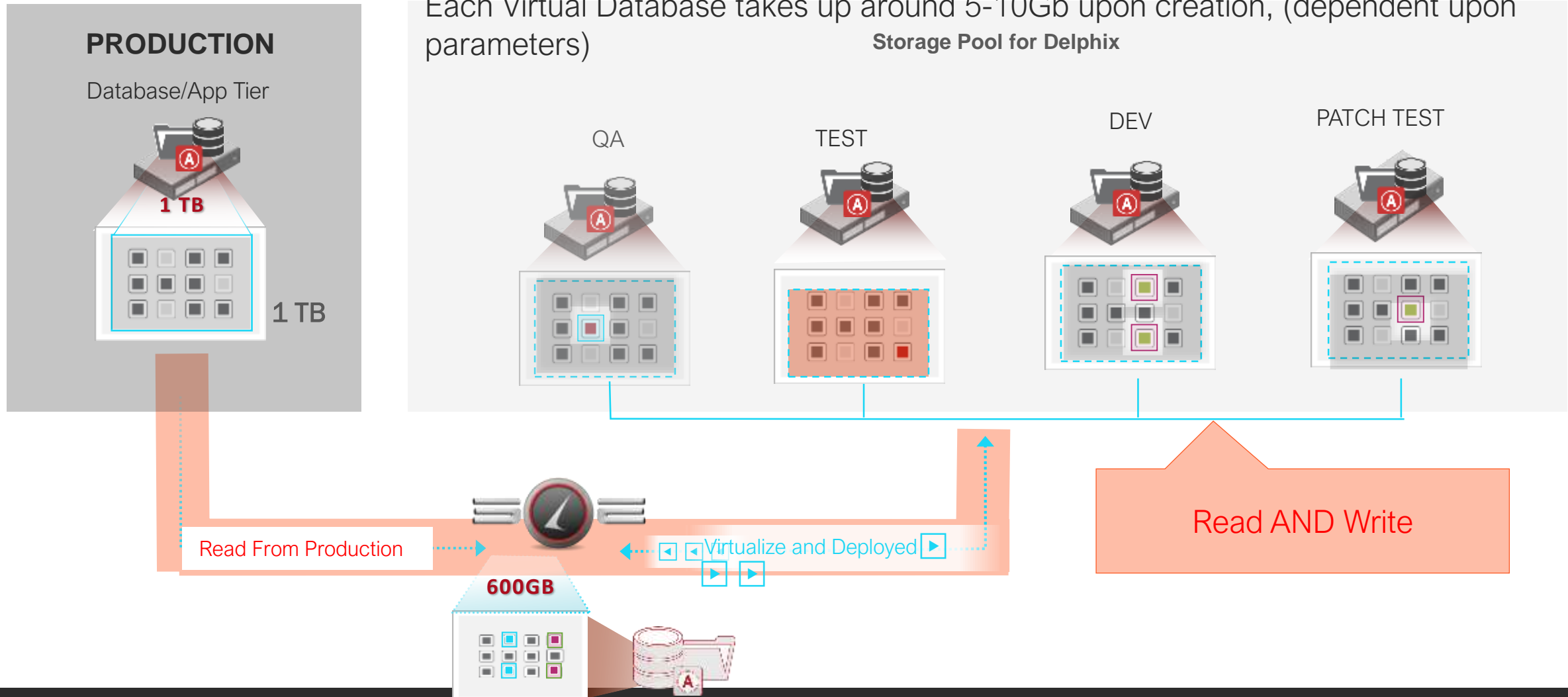
Virtualized Database, Independence

SGA

PGA

Library Cache

Buffer Cache

Java Pool

Large Pool

Redo Logs
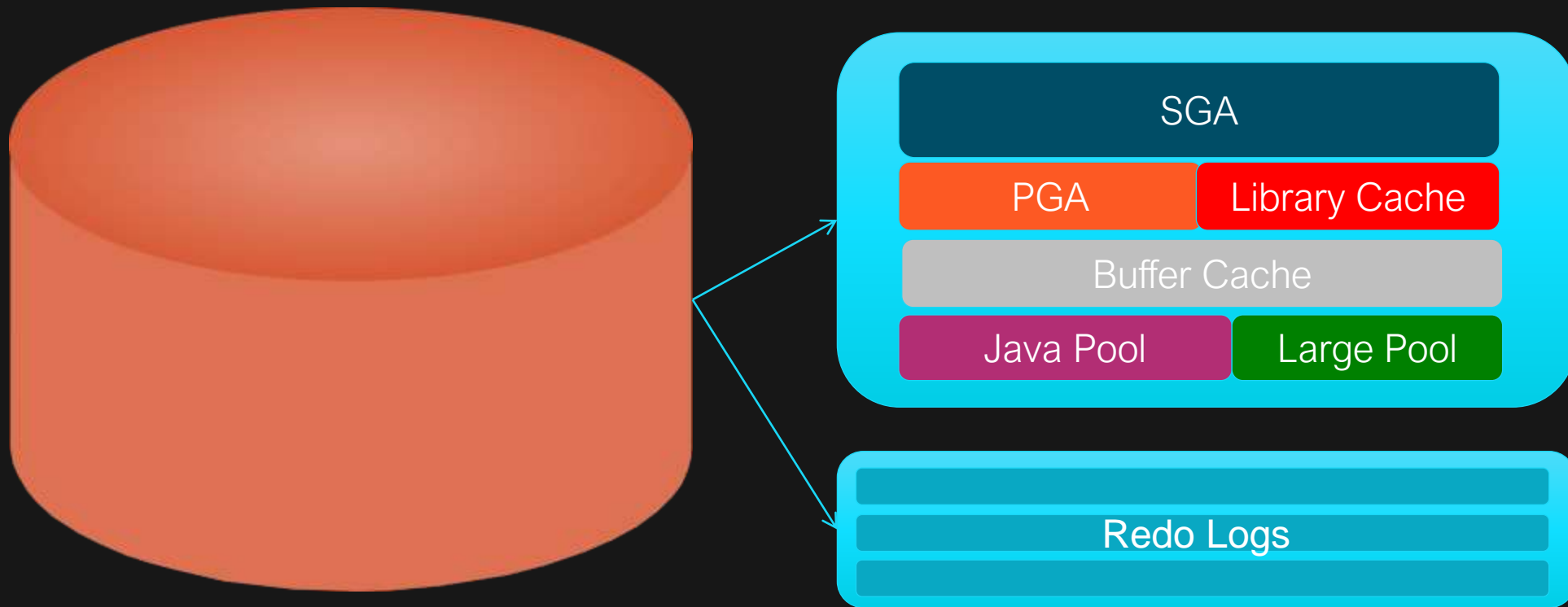
Virtualized Database, Datafiles

/Pointers to Dsource/datafile1

/Pointers to Dsource/datafile2

/Pointers to Dsource/datafile3

/Pointers to Dsource/datafile4

DSource
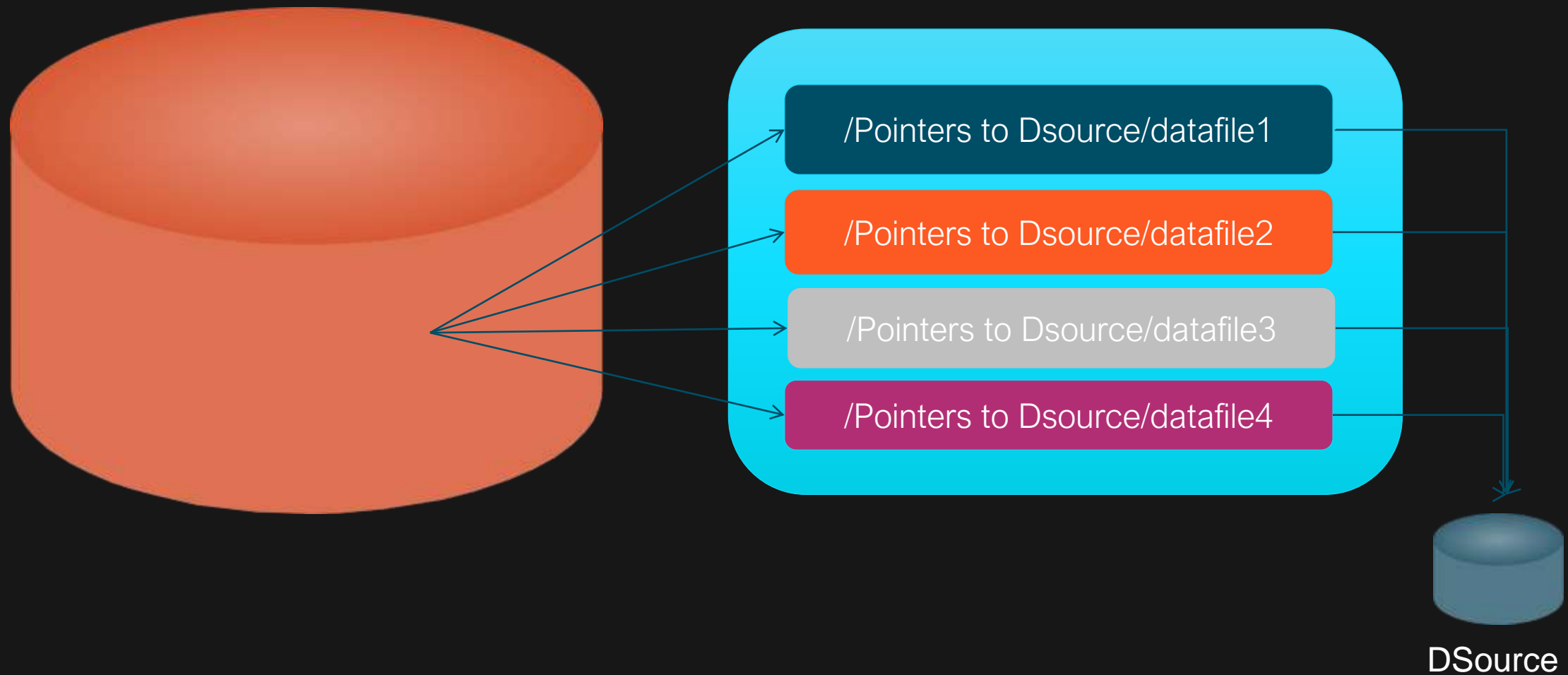
Virtualized Database, (VDB)
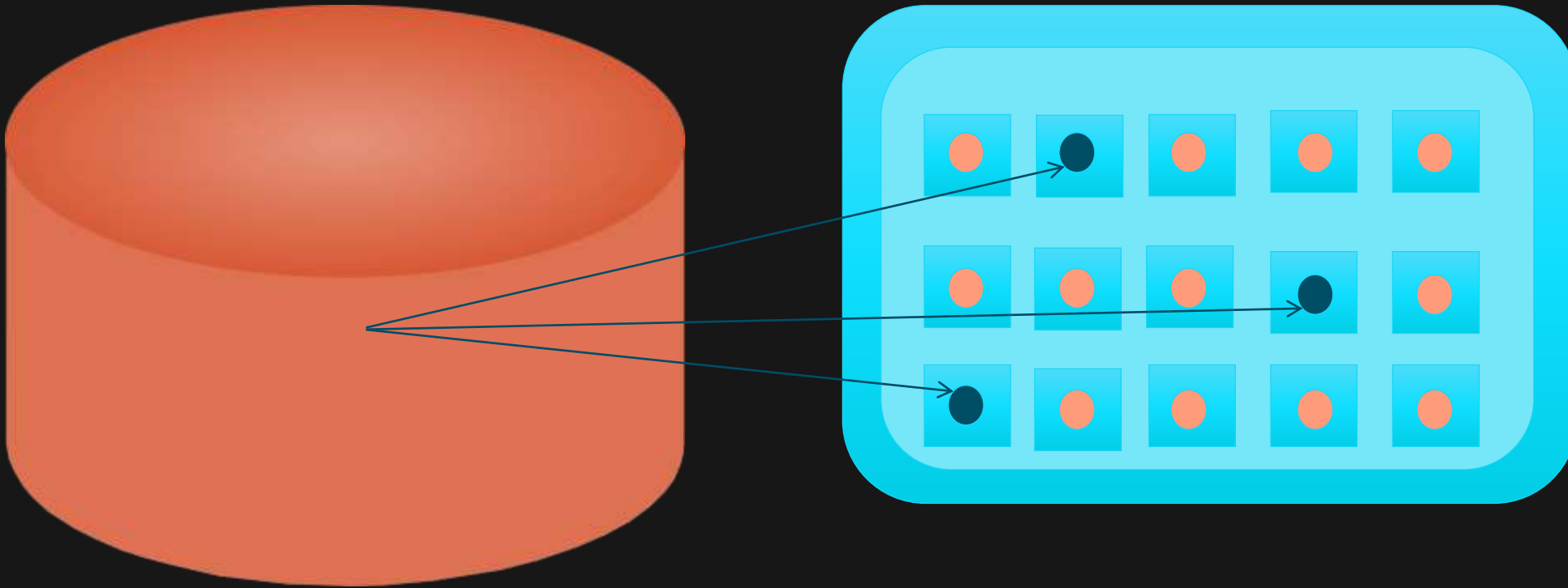
Marc Automates it all with Jenkins CLI

```
$ curl -d "script=./dlphx_dply_rend.groovy"
http://jenkins/script


$ curl --data-urlencode
"script=$(./dlphx_dply_rend.groovy)"
http://jenkins/scriptText

$ curl --user 'username:password' --data-urlencode
"script=$(<./dlphx_dply_rend.groovy)" http://jenkins/scriptText
```

Marc Automates Database Deployments

- Creates Python and Shell Scripts to perform tasks he and his team used to perform manually.
- If any pieces are missing in his logic/scripting, able to use github repository/community examples to build out what is needed.
- As Virtual environments take little resources and almost no storage, creating a DBA specific environment for development, testing and maintenance makes the DBA team less invasive to the Development/Testing Team's scrum deadlines.

Plugin for Delphix in Jenkins

Marc Adds Maven for Builds

a=Jenkins.instance.getExtensionList(hudson.tasks.Maven.DescriptorImpl.class)[0];
b=(a.installations as List);
b.add(new hudson.tasks.Maven.MavenInstallation("MAVEN3", "/home/apache-maven-3", []));
a.installations=b
a.save()

2nd- Marc Creates Containers/Data Pods

Containers offer the ability to isolate application code and/or the whole infrastructure stack into a **package able entity to ease deployment**, even from the same kernel. This is a powerful tool for DevOps to ease deployment for complex tiers, applications and multiple data stores. Data Pods is the next step, where a pod is created from virtualized environments.

Docker != Container

- Virtualization is important, this is OS level

- Resource level isolation at the kernel level

- Technology such as OverlayFS, allows for containers, which removes overhead of other virtualization technologies.

Database Containers:
- Open Container Initiative, (OCI)
- Kubernetes, (Google)
- CoreOS and rkt
- Apache Mesos
- Canonical and LXD
- Windocks
- Delphix Data Pods
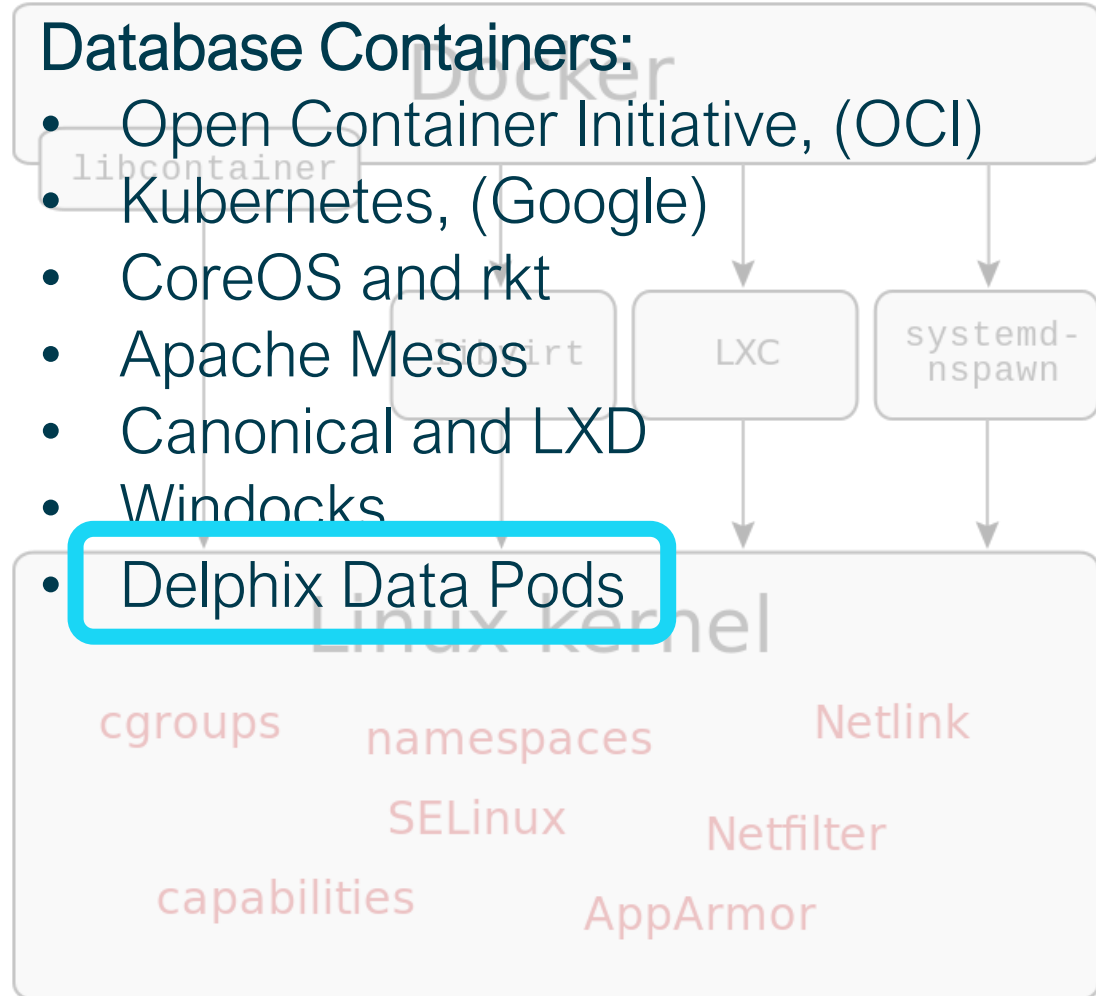
Group Environments into Containers
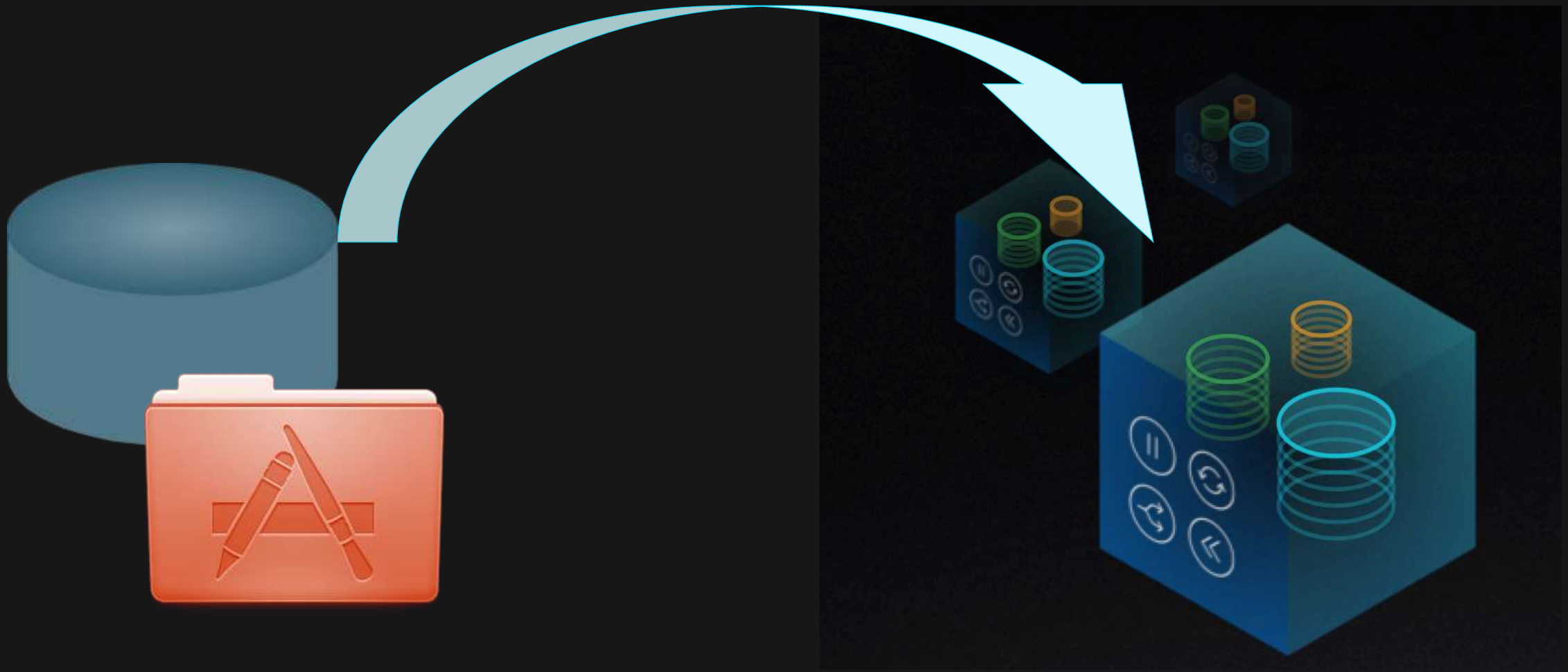
Why Create Containers/Data Pods

- Development is done by project or feature vs. tier or product.
- With the introduction to the cloud, the business requires those that can do more at a global scale.
- Ease of management and maintenance
- Provide more value to the business

Data Pods

Database

Unstructured Data

Application
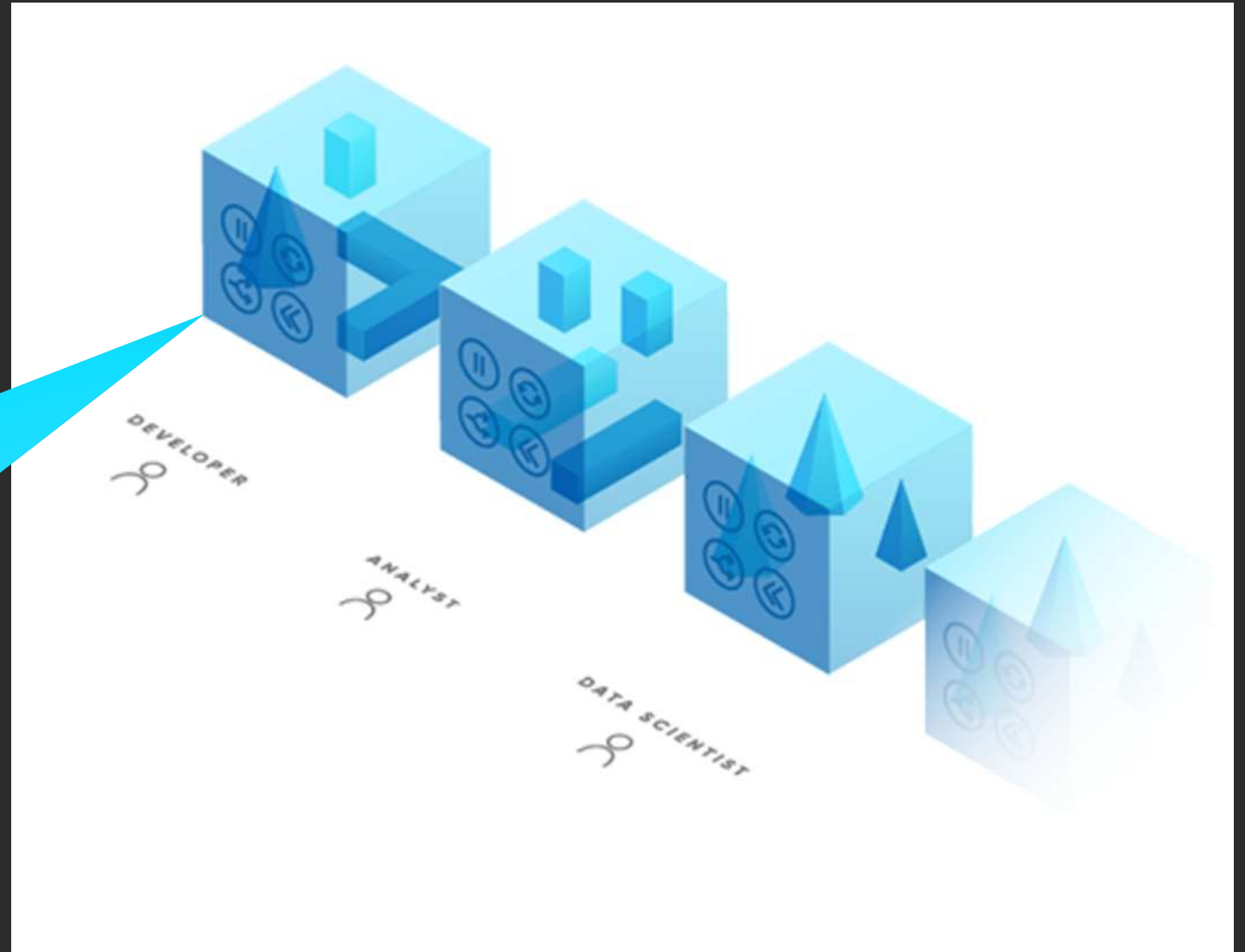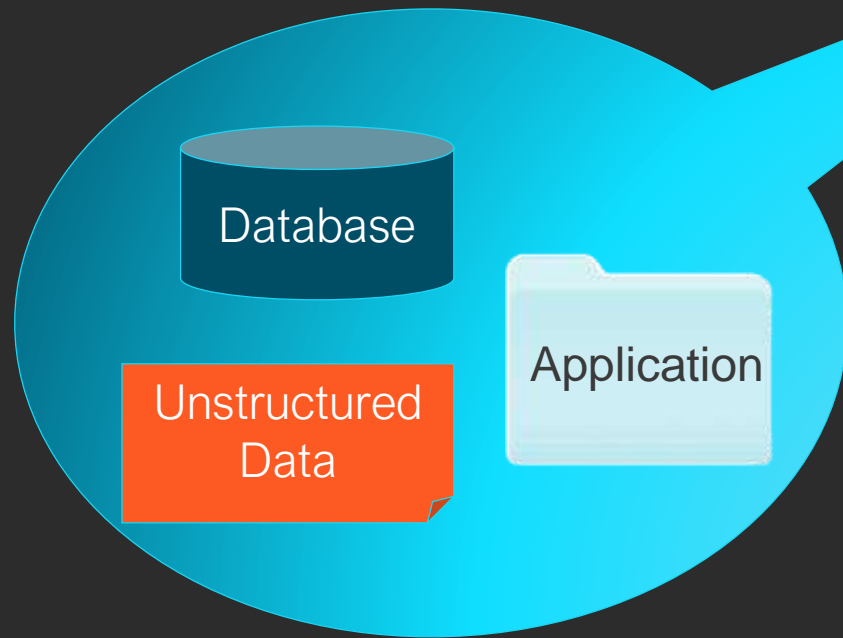
Data Pods for Ease of Delivery

Delphix Virtualization Engine
Same Storage Size as Production

Deployed using Jenkins

Rsync (UNIX/Linux)
Robocopy (Windows)

NFS
iSCSI

NFS
iSCSI

NFS
iSCSI

Development

Testing

Reporting
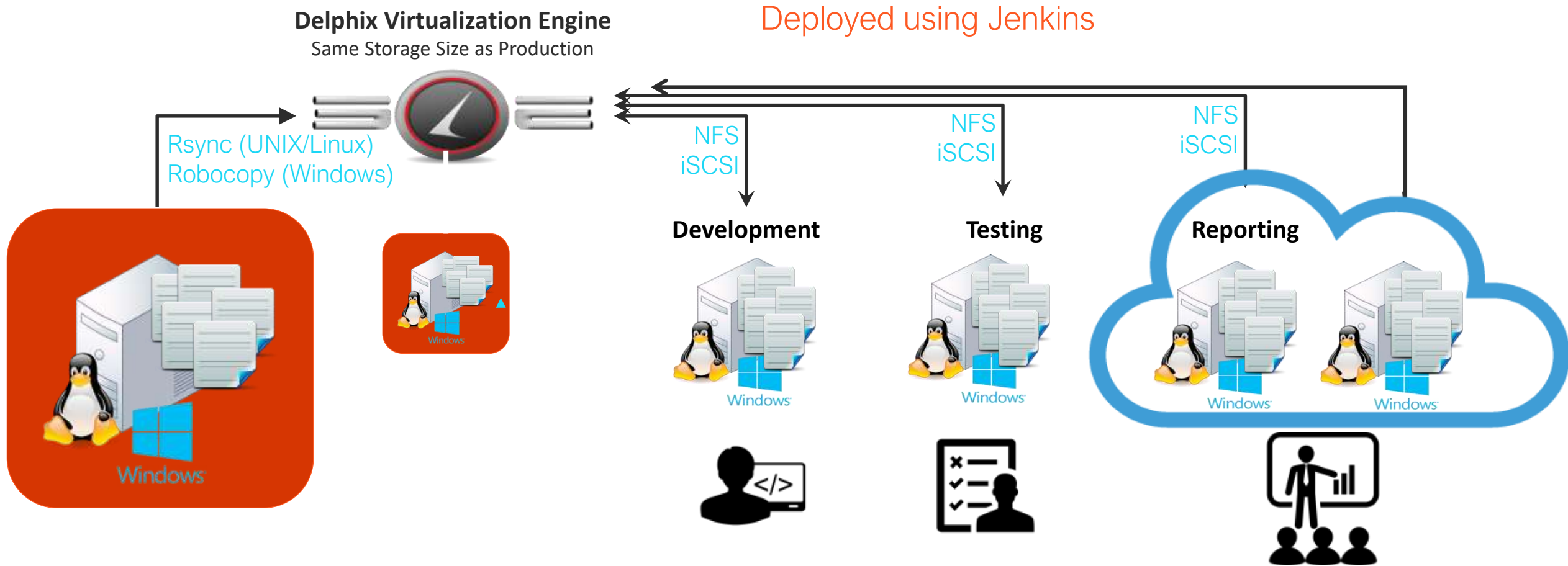
Now the Developers and Testers can work at the speed the business needs them to…

Thank you

DataOps and Source Control
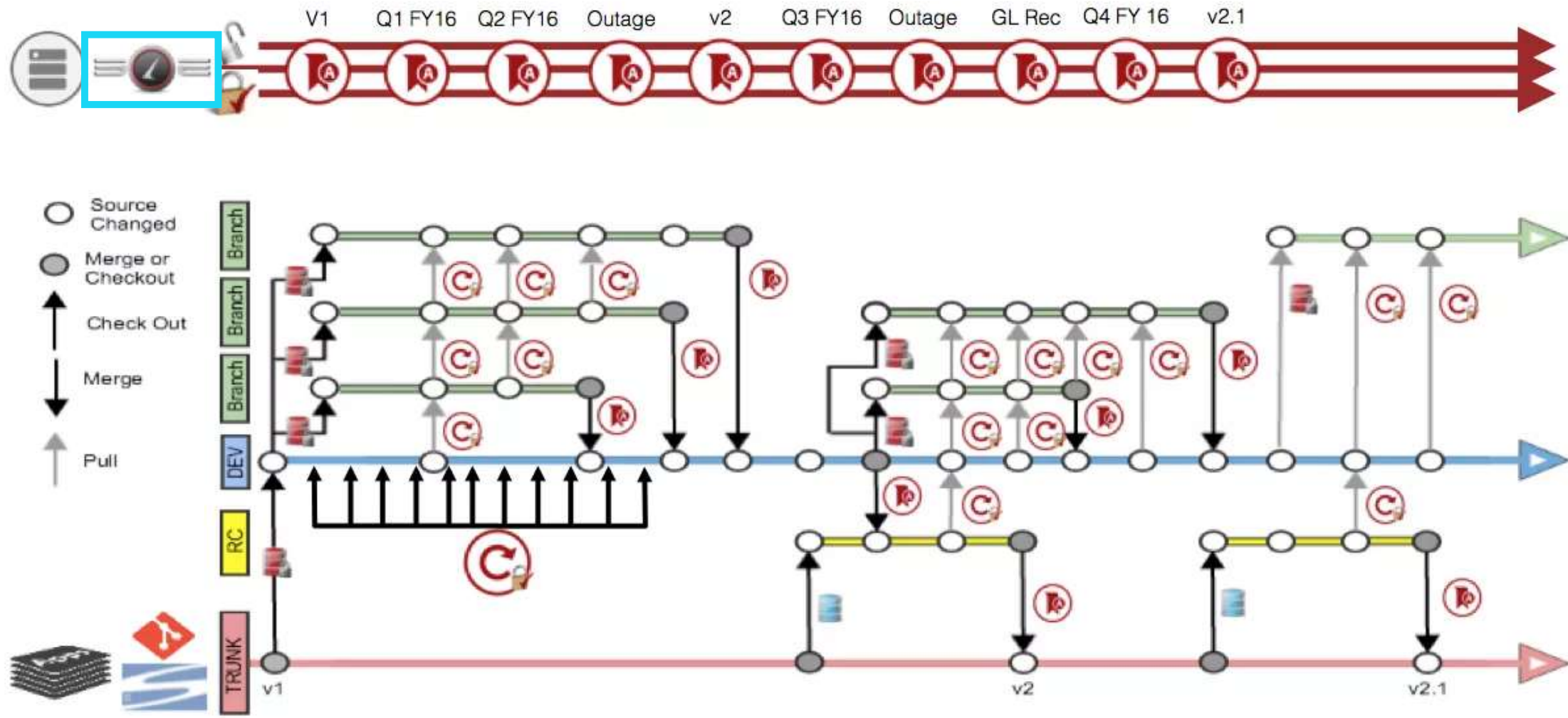
Branching and Bookmarking

- The ability to mark each iteration of development with a bookmark

- Simplify to lock and deliver while testing a consistent image via a virtual database, (VDB)

- If a something goes wrong, the ability to "bookmark", (and subsequent snapshot) to deliver to development to address.

Self Service Portal
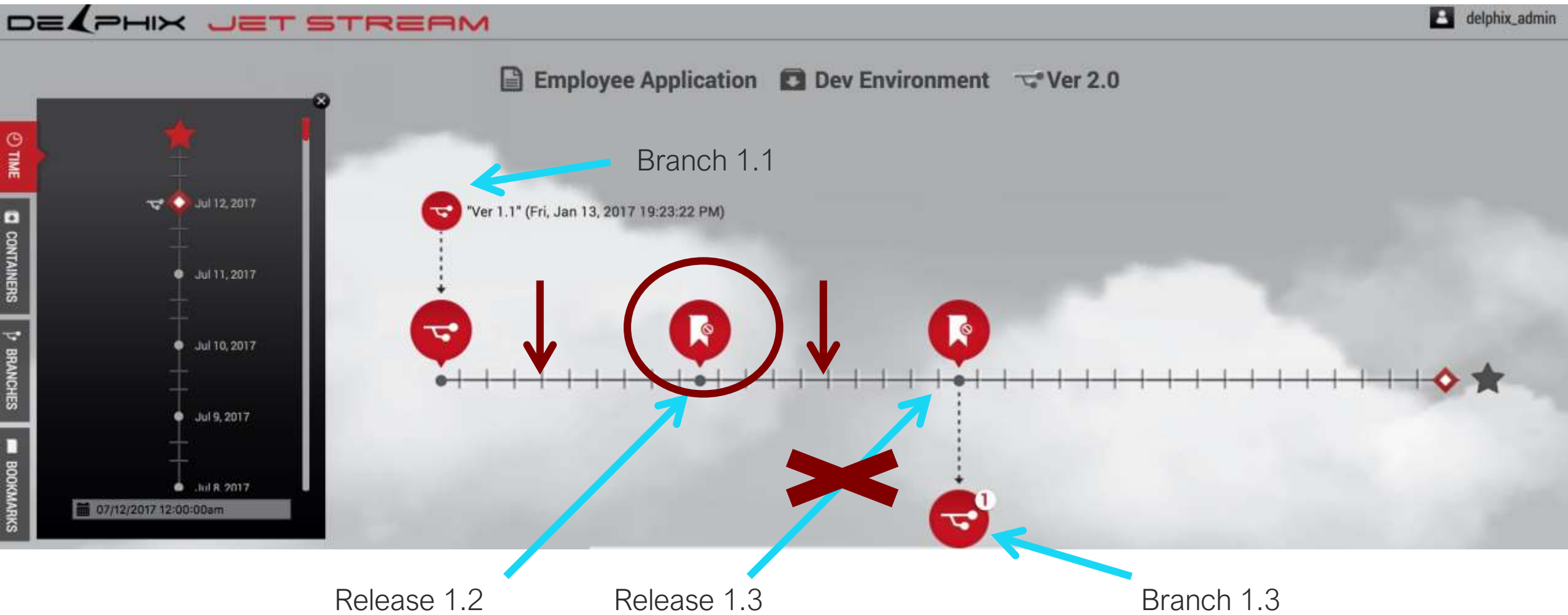
Sprint Goals are Now Possible to Meet!

Secure It

Confidential Data

- GDPR, (General Data Protection Regulation States):
  - Must have lawful basis to post process data
  - Subject must have consented
  - Must have contract that needs data
  - Necessary and in order:
    - Compliance
    - Protect vital interests
    - In public interest
    - In legitimate interest of the owner/other party and doesn't violate the freedom and rights of the subject.

What is GDPR Data?

- Standard Personal Data: Name, Address, Phone Number

- Identification Numbers: Social Security Numbers, IP Address

- Medical and Financial Information

- Cookies, Geo Tracking Info

Confidential data

Exposure

Production
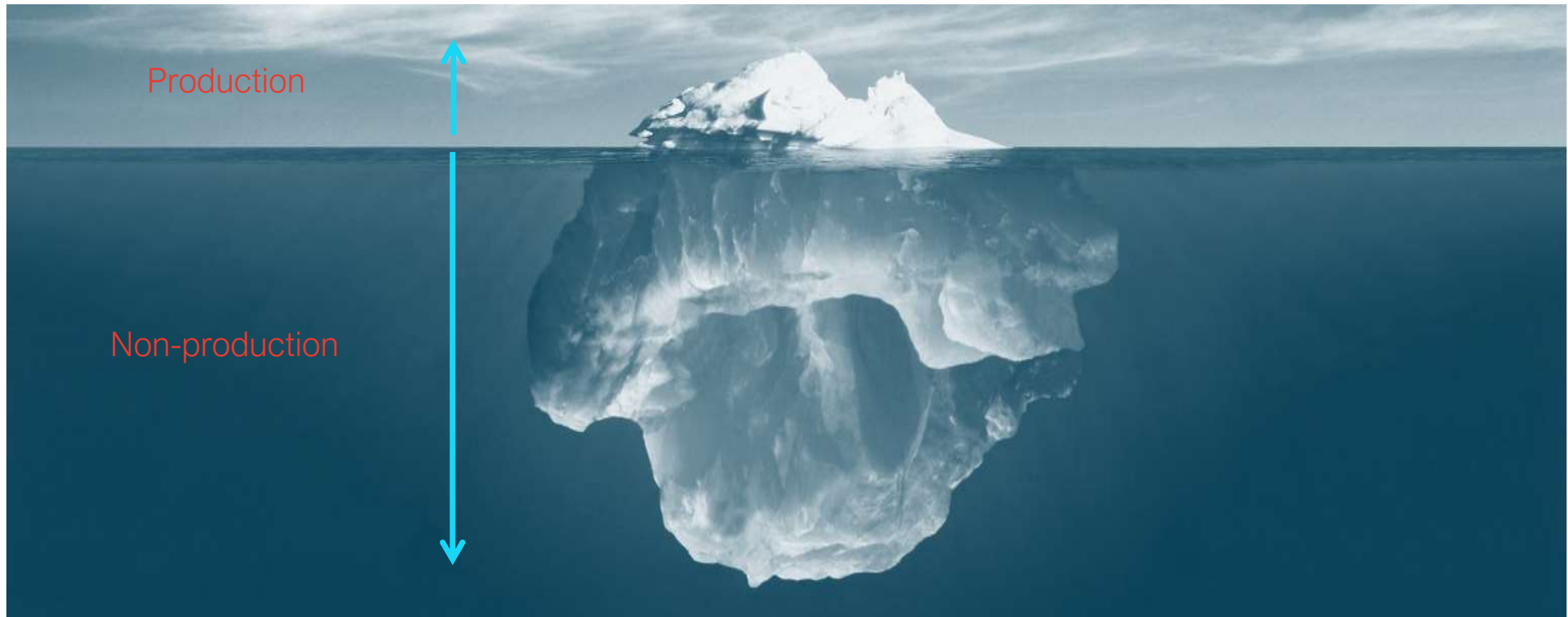
Non-production

Confidential data

- *Encryption* is **reversible** data obfuscation, which is essential for production access and MUST be DONE.

- Data masking is ***non-reversible.***

  - It solves the issue at the **data level**.

- Is authentication and authorization in non-production in compliance with security goals?

  - All organizations will soon need to review if critical data in non-production environments be accessible to developers, testers and users.

Why Masking is Important-

- Masking personally-identifiable, (PII, HIPPA, PCI, etc.) information renders it useless from a security standpoint
- Resolves both the *technical* and personal responsibility issue.
- The data can be masked before it moves to non-production, removing unnecessary risk.
- As we discussed, on average, 80% of data is non-production.
- GDPR is on the horizon and US must be compliant by May, 2018

Data Virtualization, Masking, to the Cloud

Delphix Virtualization Engine
8 TB storage

MSSQL Source
8 TB database

MSSQL Target
Oracle Database

Delphix Masking Engine

Application Server

File Server
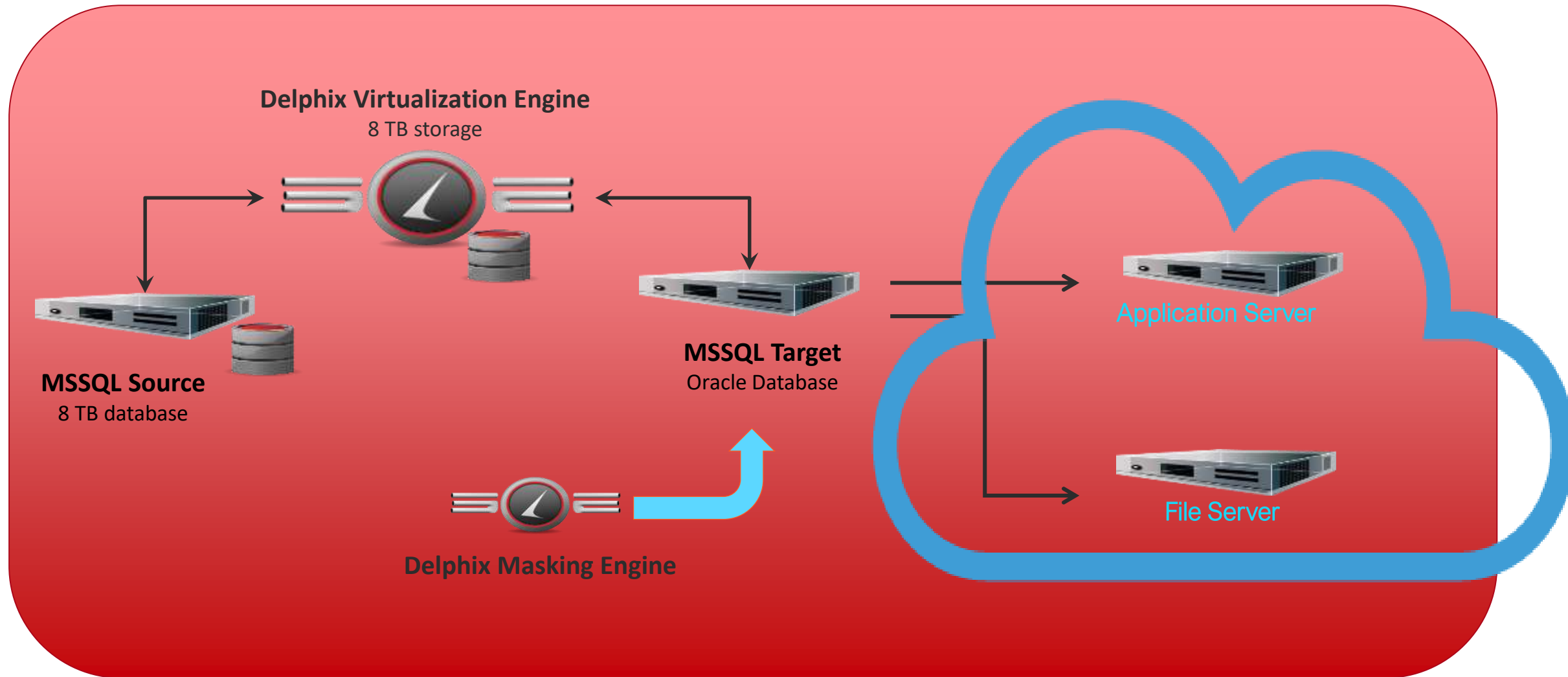
Confidential Data with Masking

Exposure

Solution

Production

Encryption

Non-production

Mask
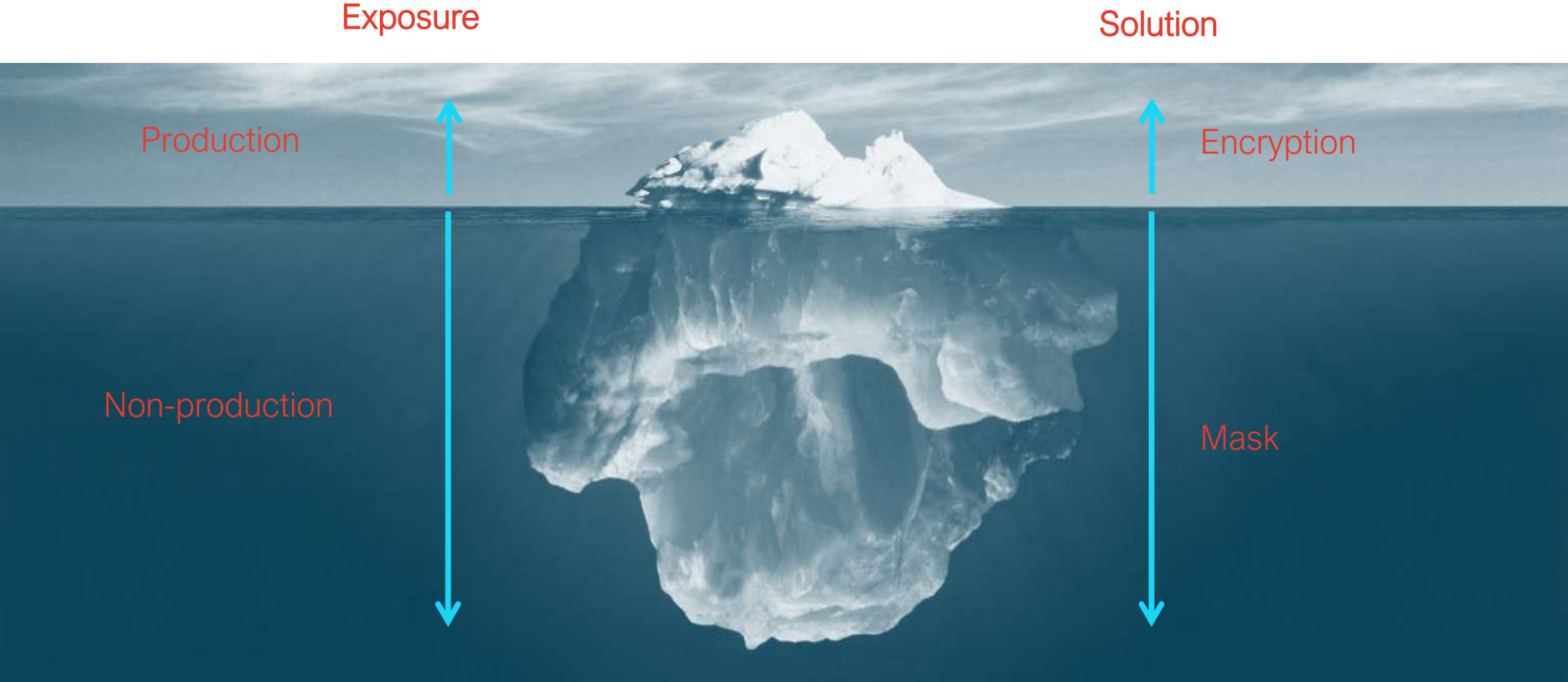
DataOps Summary

The Goals Achieved?

- DevOps skills, (scripting, release builds, security, advanced optimization.

- Automate or phase out tedious tasks with platform automation, (monitoring, backups, maintenance jobs, etc.)

- Escape data gravity through features such as advanced self-service portals, virtualization and container technology.

- Which encompasses advance features of DataOps, removing the data friction and allowing data to move as fast as the rest of development and testing.

And George and the DBAs are no longer viewed as a roadblock, but heroes…

DevOps != DataOps

- DevOps is automation, culture and collaboration, WITH tools

- DBAs are viewed as the roadblock and often are the solution!

- Data, although gold, is viewed as the cause of friction.

- DevOps and Agile can't be successful if they can't get the data they need.

- Just because you've gone DevOps doesn't mean anything with GDPR is on the horizon.

This is the difference between DataOps and DevOps in the World of Data…

With the Enhancement to DataOps

Dave's company is experiencing an increase of 10% in data accessibility which should result in more than $65 million in additional net income. You can see why DataOps was important.

Embracing DataOps

- Removes the last piece holding developers and testers back.
- Uses virtualization and containers to simplify.
- Self-service portals for developers and testers to refresh and develop the way they do in an agile environment.
- Tools with an interface made for Agile and Development data at its focus.

How DevOps Becomes DataOps

Where do you go from here?

- Automating repeatable and complex processes

- Locating heterogeneous tools/scripting languages that can support multiple tiers of technology.

- Use tools to speed up the development cycle.

- Secure ALL the data- GDPR is coming!

- Understanding that automation does not lessen anyone's value.

- The tools are important, but that cultural changes are the biggest hurdle.
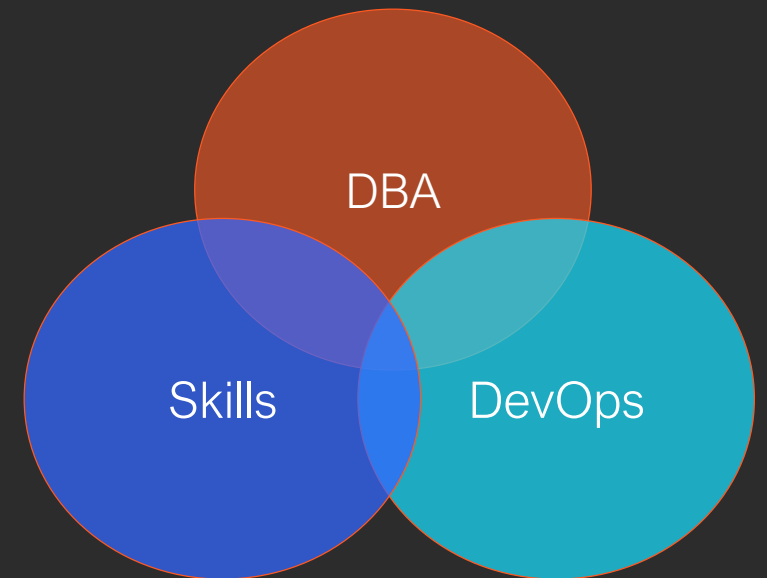
References

Blog Posts-
- [The DBA and DevOps: The Last Frontier](#)
- [The DBA and DevOps: Automation and Configuration Management](#)
- [The DBA and DevOps - Orchestration and Monitoring](#)
- [DBA and DevOps: The Ghost in the Machine](#)

Webinar Recordings
- [The DBA and DevOps- the Last Frontier](#)

- Read- [The Phoenix Project](#)and the [DevOps Handbook](#)!

Q&A

Kellyn Pot'Vin-Gorman
Technical Intelligence Manager
kellyn@delphix.com
http://dbakevlar.com