



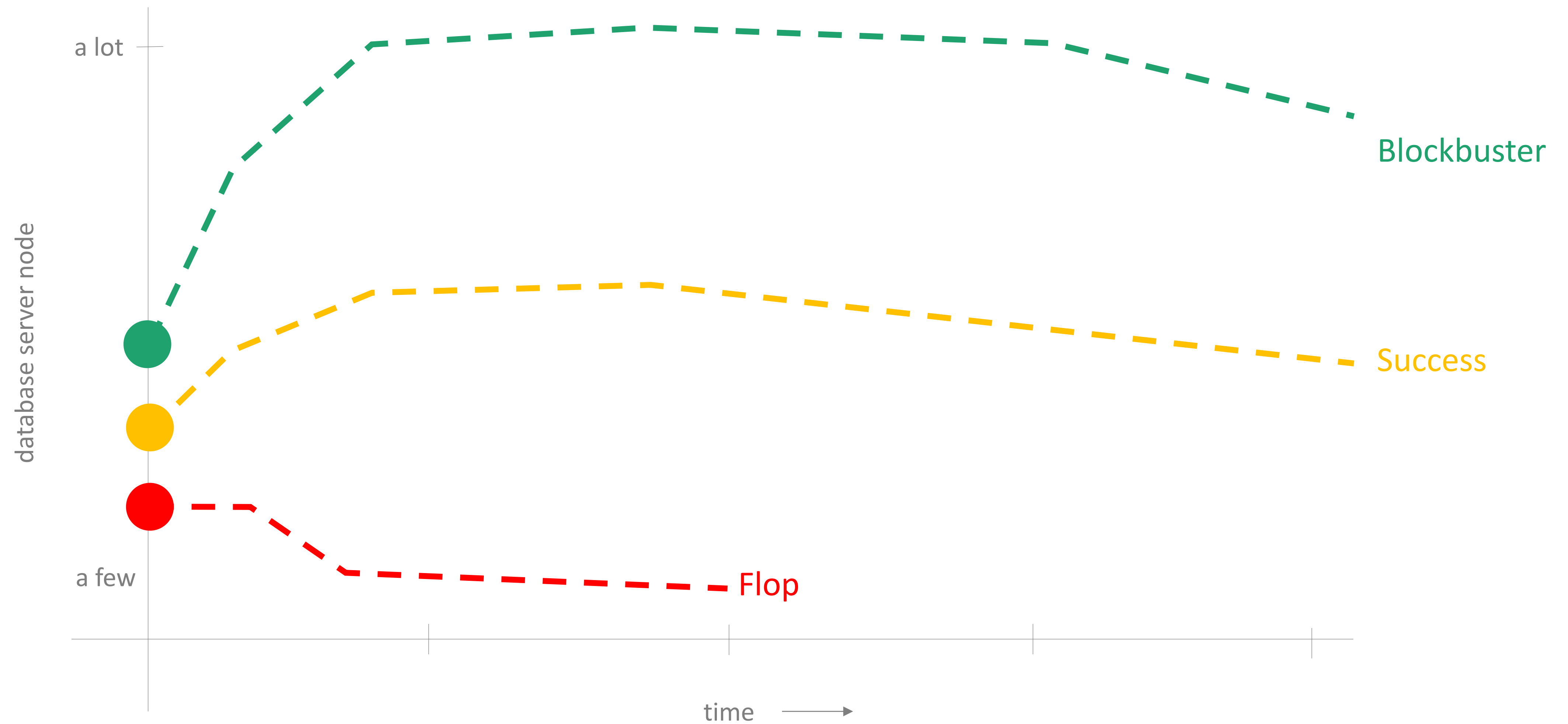
# Vertica's Eon architecture

*Separation of compute from storage*

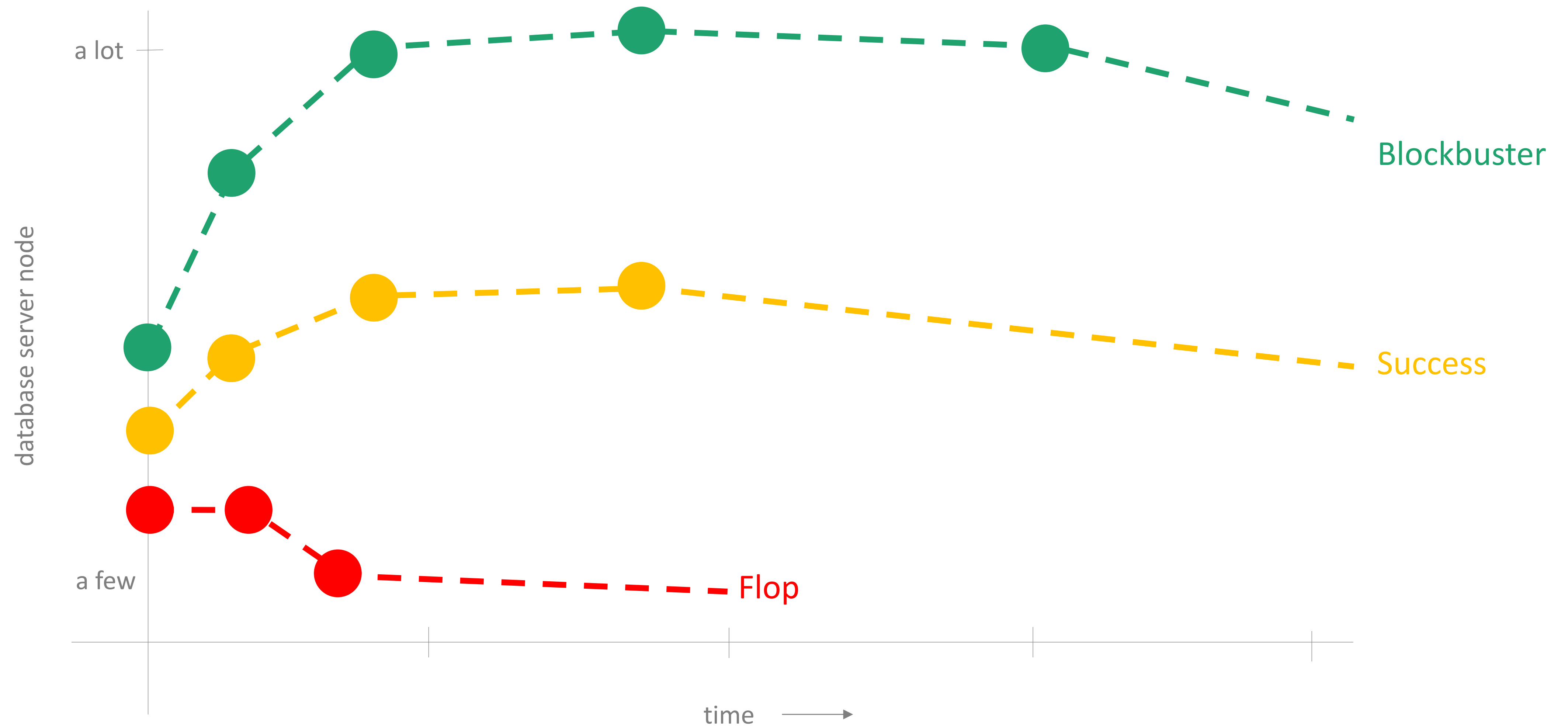
David Sprogis



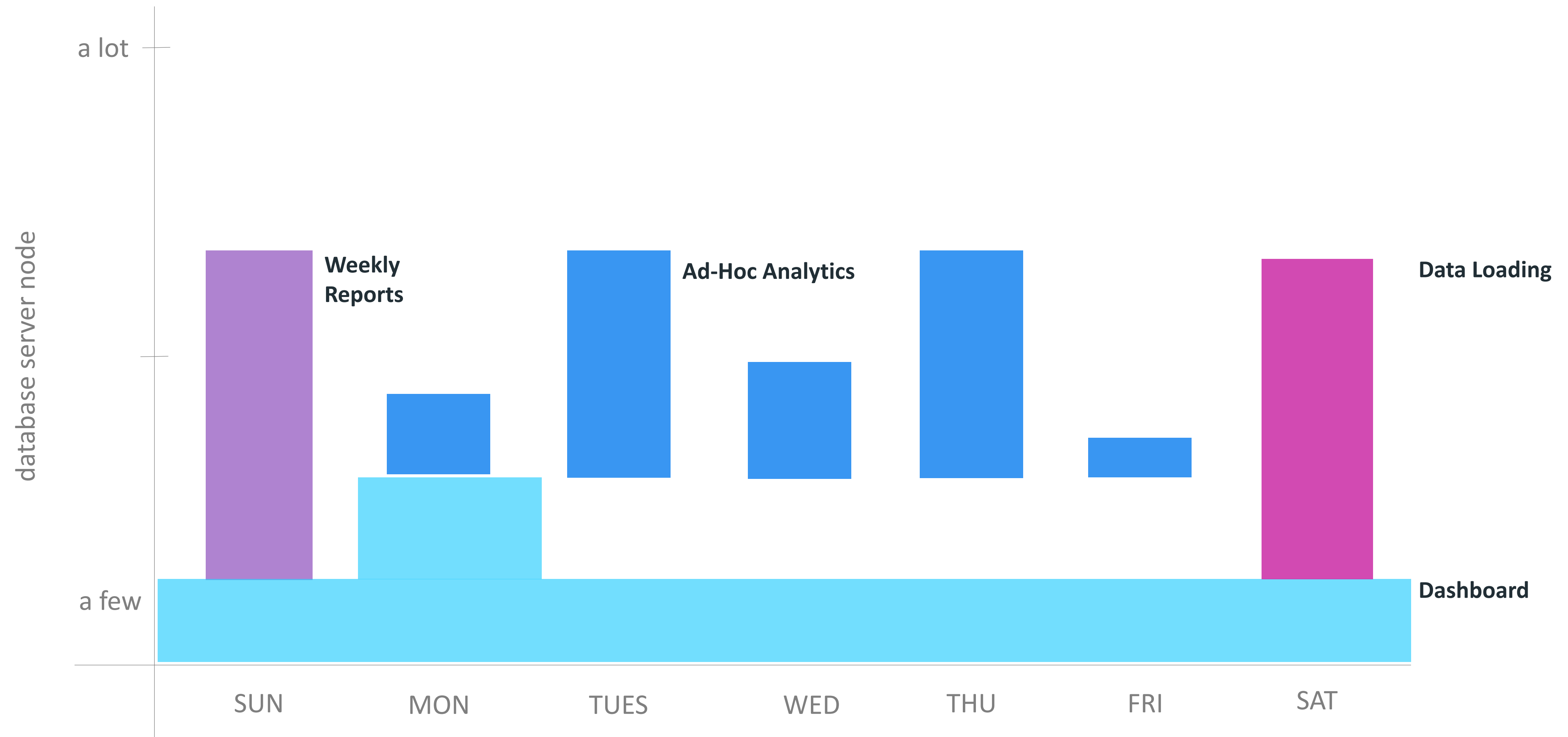
# The struggle to size database infrastructure



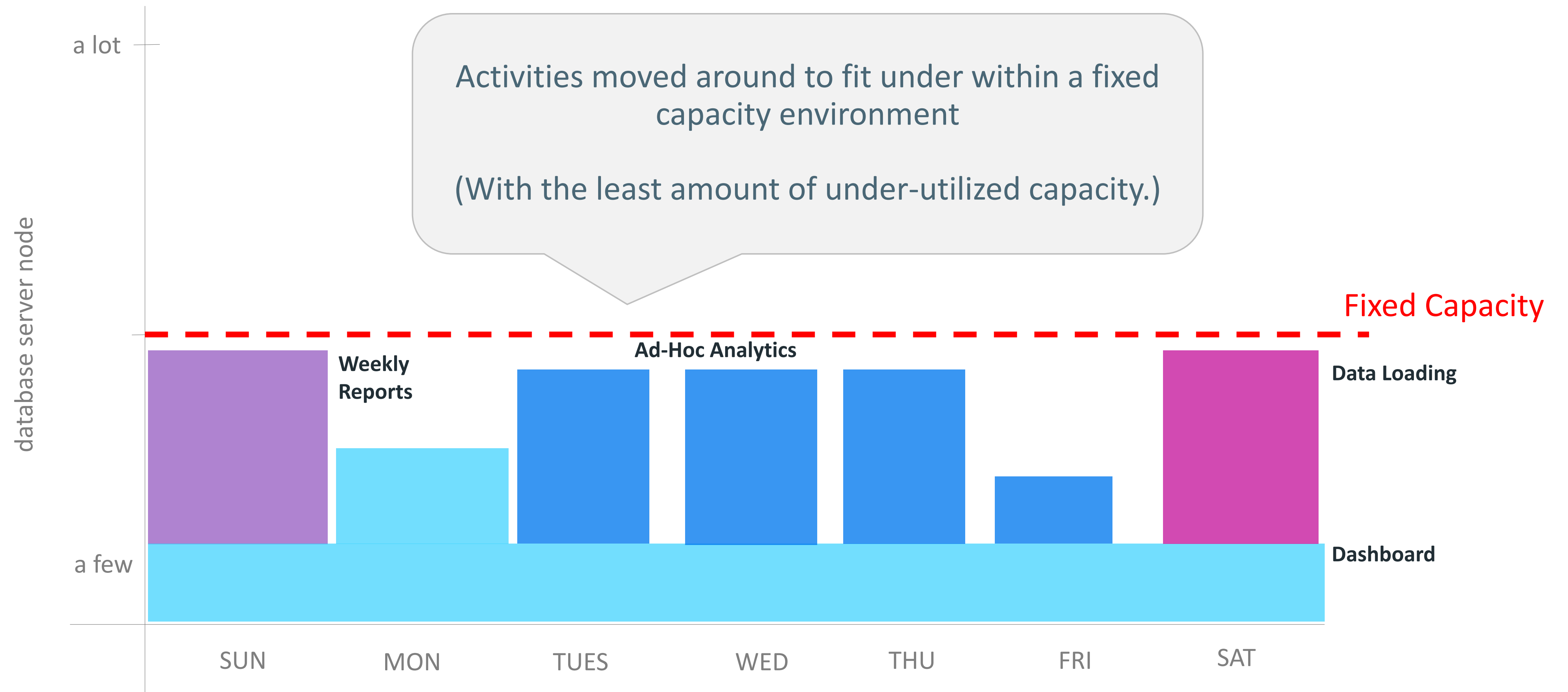
... then resize across a product lifecycle



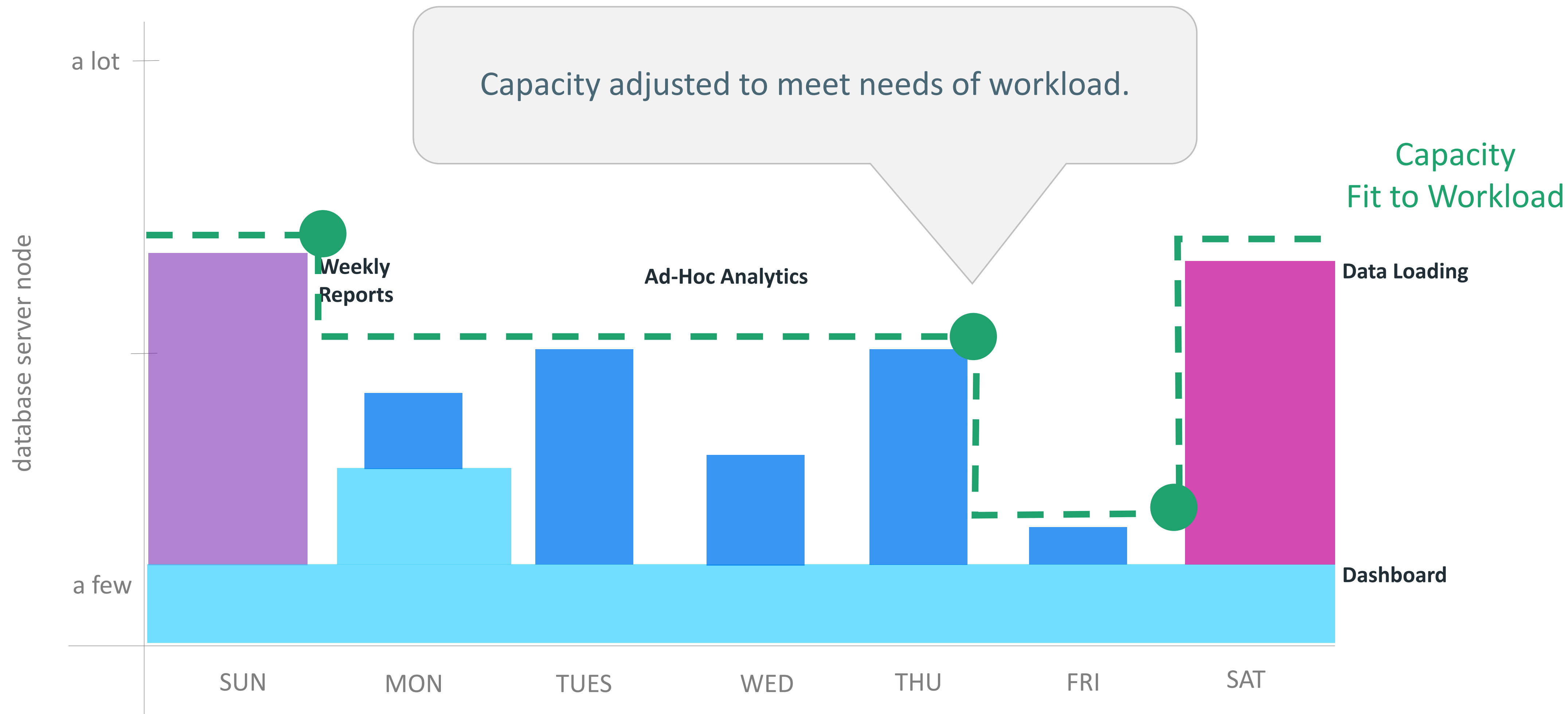
## Not just gaming companies



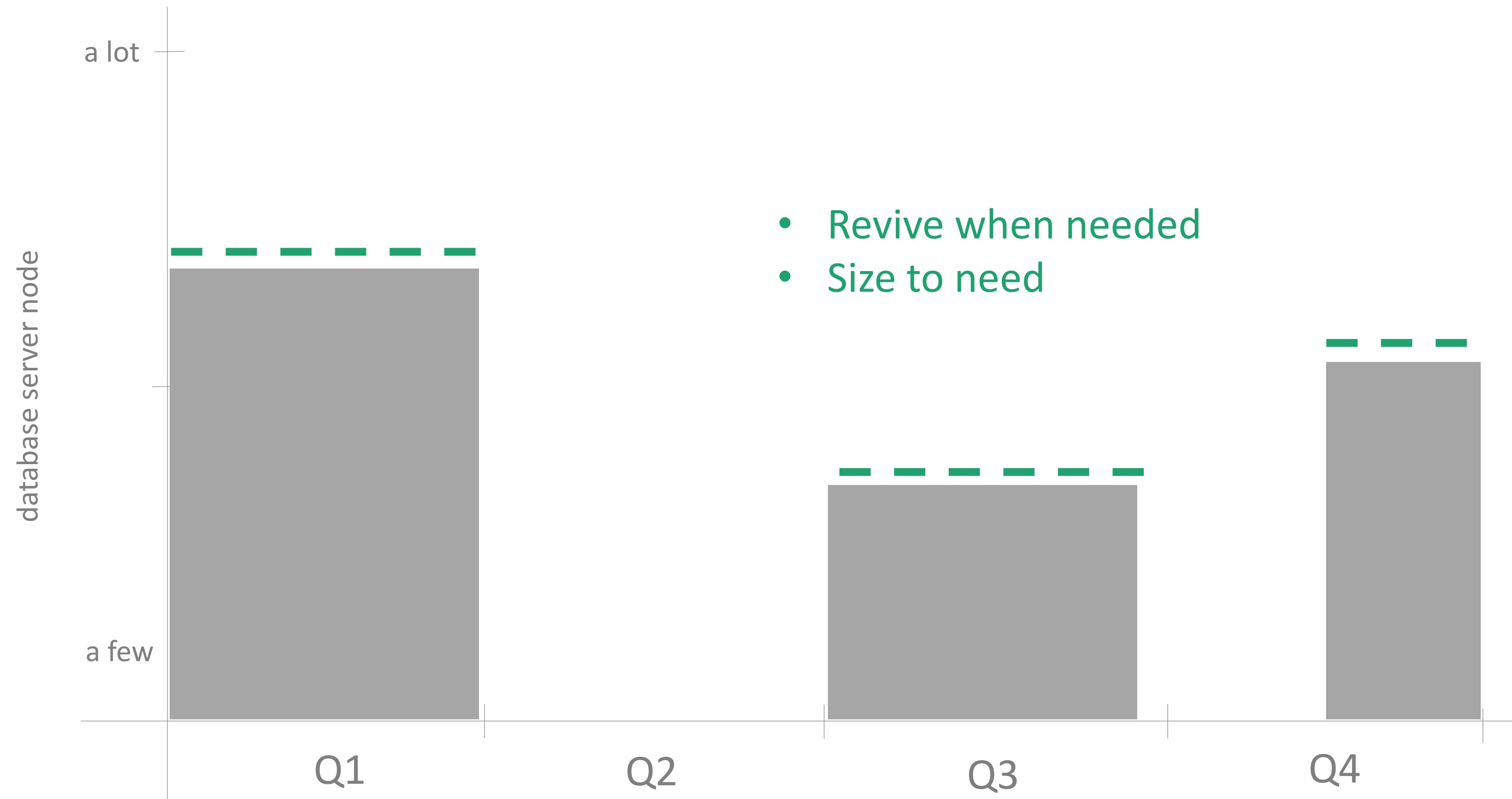
# Adjusting workload to fit capacity



# Eon mode lets you fit capacity to workload, scaling as needed



# Hibernate between project phases



- Revive when needed
- Size to need



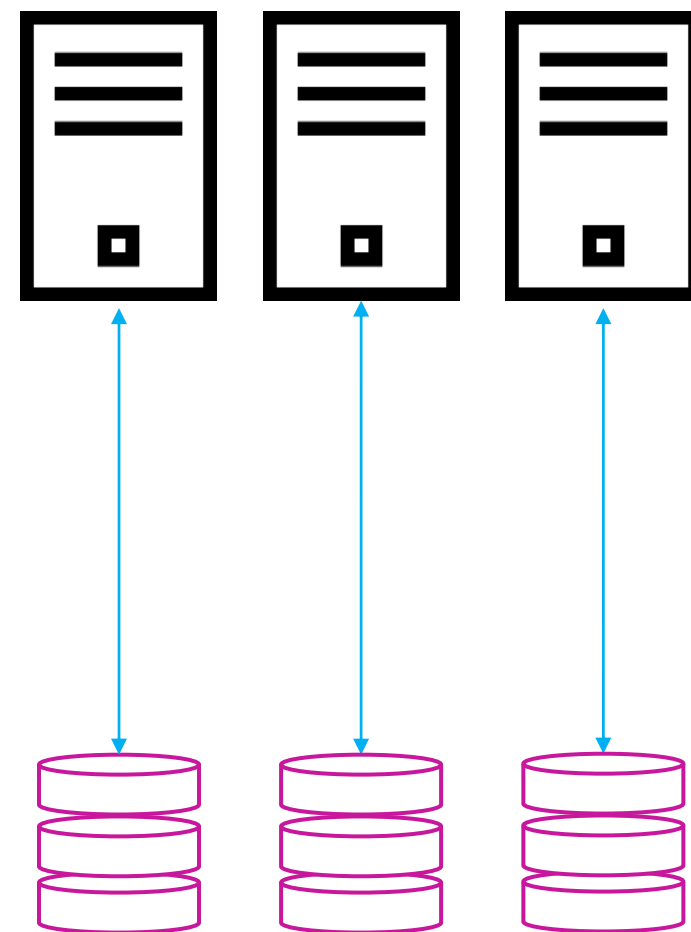


**How does Eon work?**



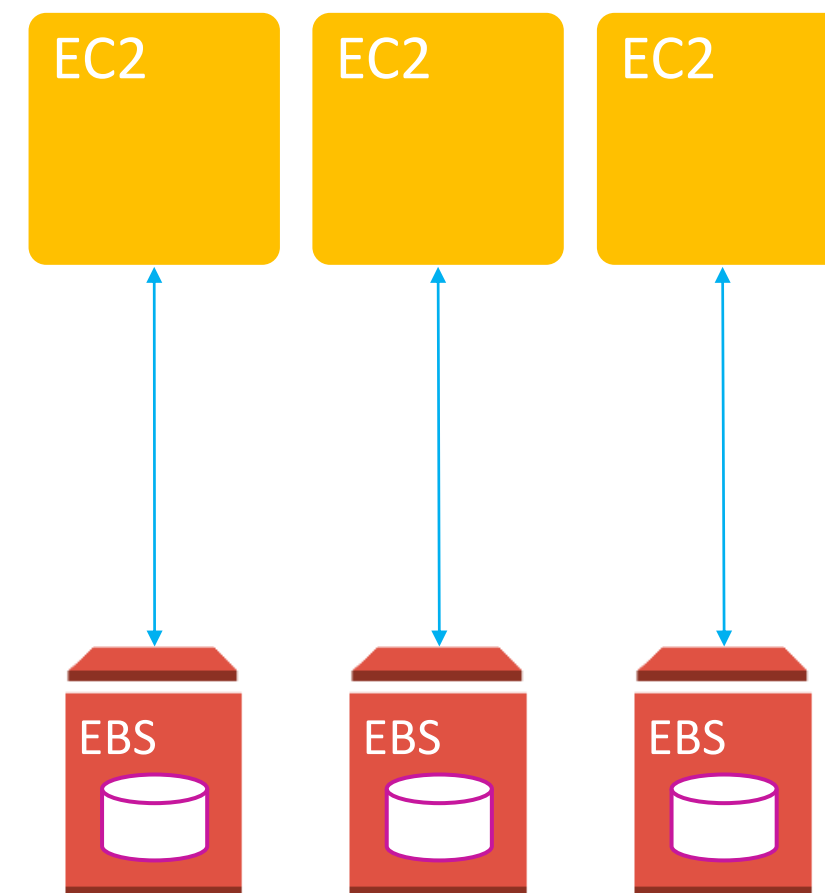
# Growth and Extension of Vertica

## Enterprise Mode Vertica On-Premises



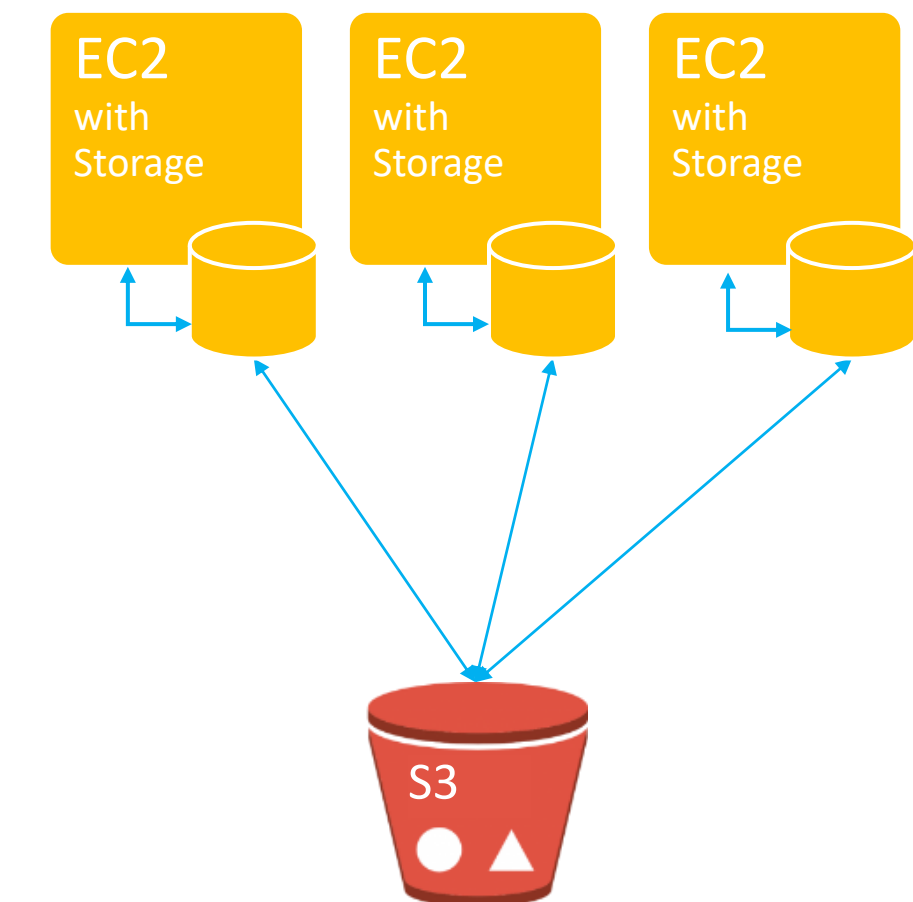
*Fastest, open architecture,  
advanced analytics SQL  
database on commodity  
hardware.*

## Enterprise Mode Vertica AWS



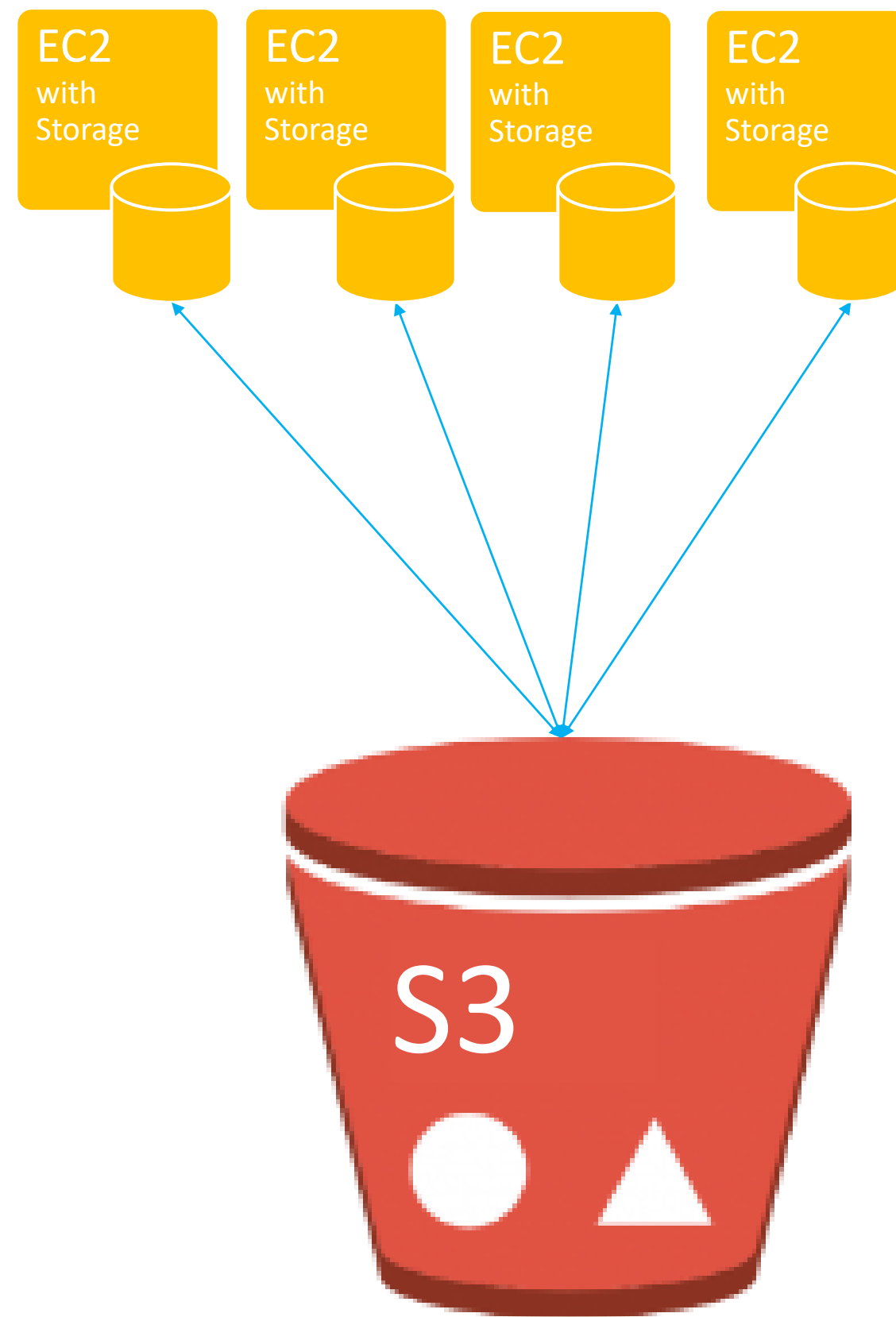
*Fastest, open architecture,  
advanced analytics SQL  
database in the cloud.*

## Eon Mode Vertica AWS



*Fastest, open architecture,  
advanced analytics SQL  
database capable of scaling  
quickly to keep pace with your  
changing workload.*

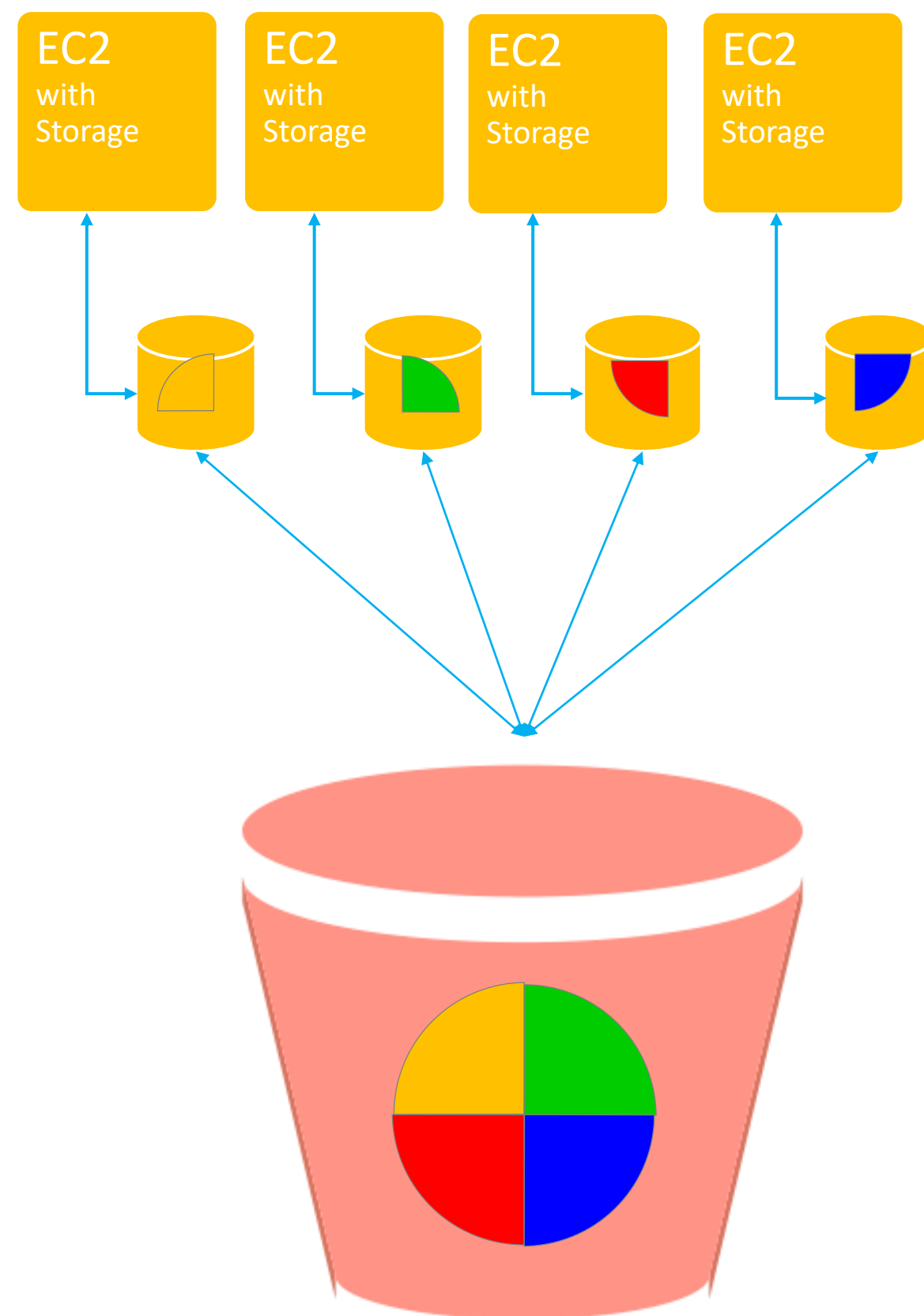
# Simpler, Faster Provisioning



Getting started just got easier!

All you need are EC2 instances with storage and an S3 Bucket.

# How does Eon Mode work?



Nodes with instance storage are provisioned.

Each node caches a shard of the database from S3 into instance storage from which it can service queries with the blazing performance you expect from Vertica.

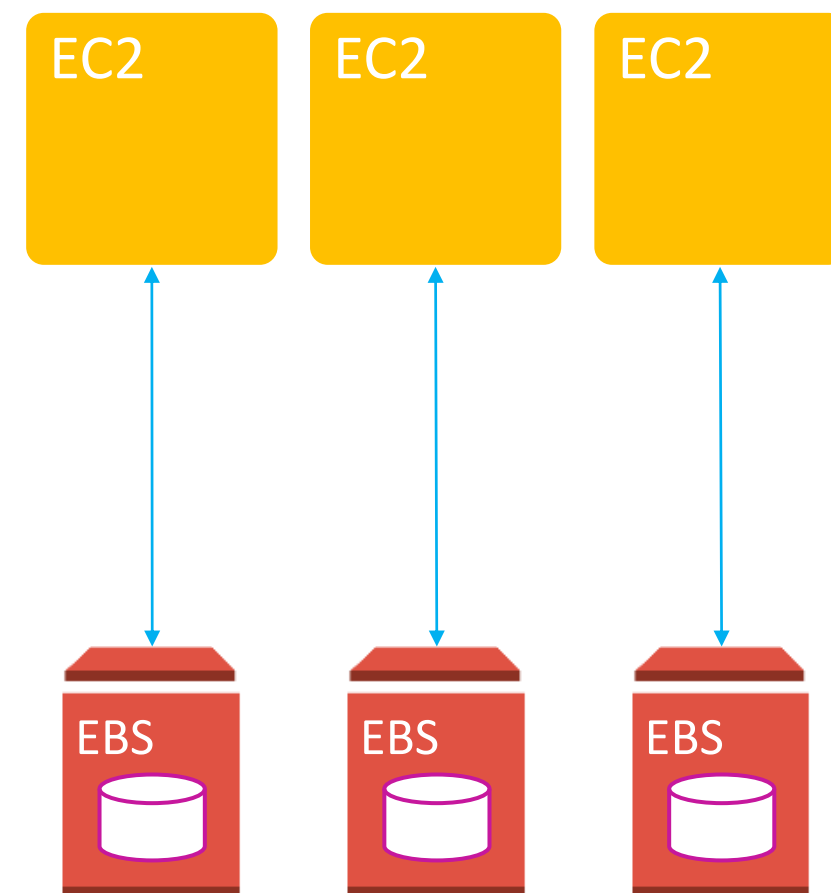
Collectively, this ephemeral instance storage layer is called the “depot” and remains consistent with the full database maintained in S3.

Your database is stored in S3 with 11 9’s of reliability.

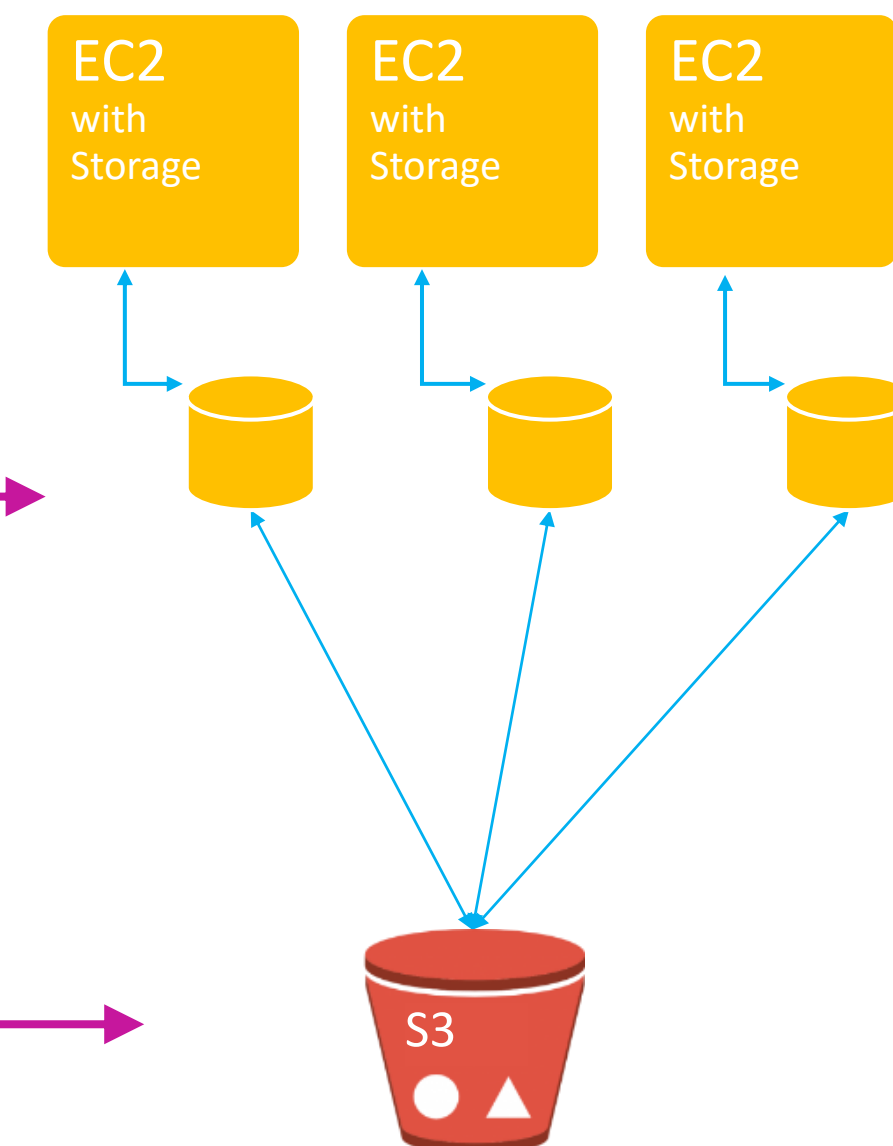


# Growth and Extension of Vertica

## Enterprise Mode Vertica AWS



## Eon Mode Vertica AWS

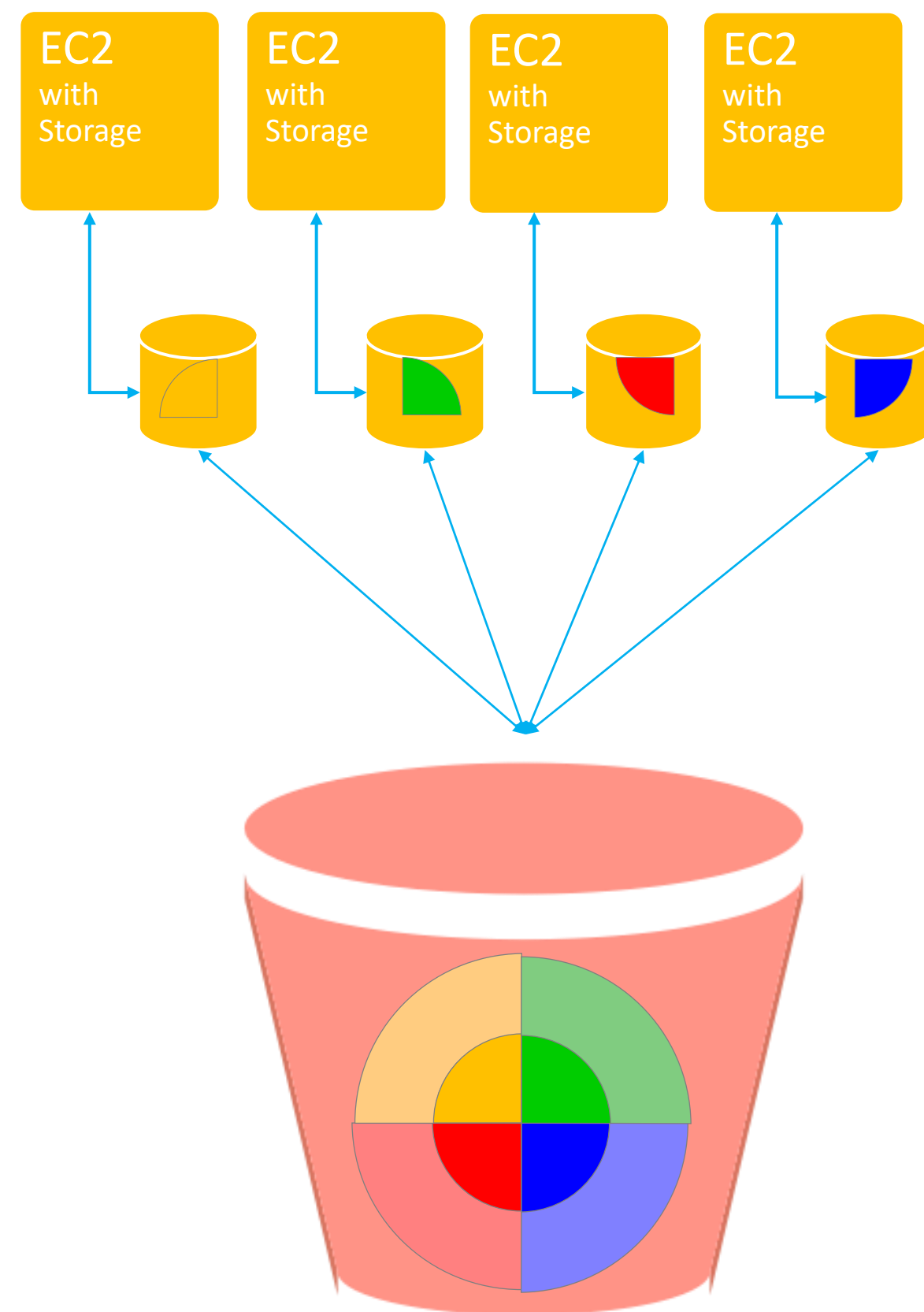


Performance

Durability

\* Enterprise Mode or Eon Mode decided at the time the database is created

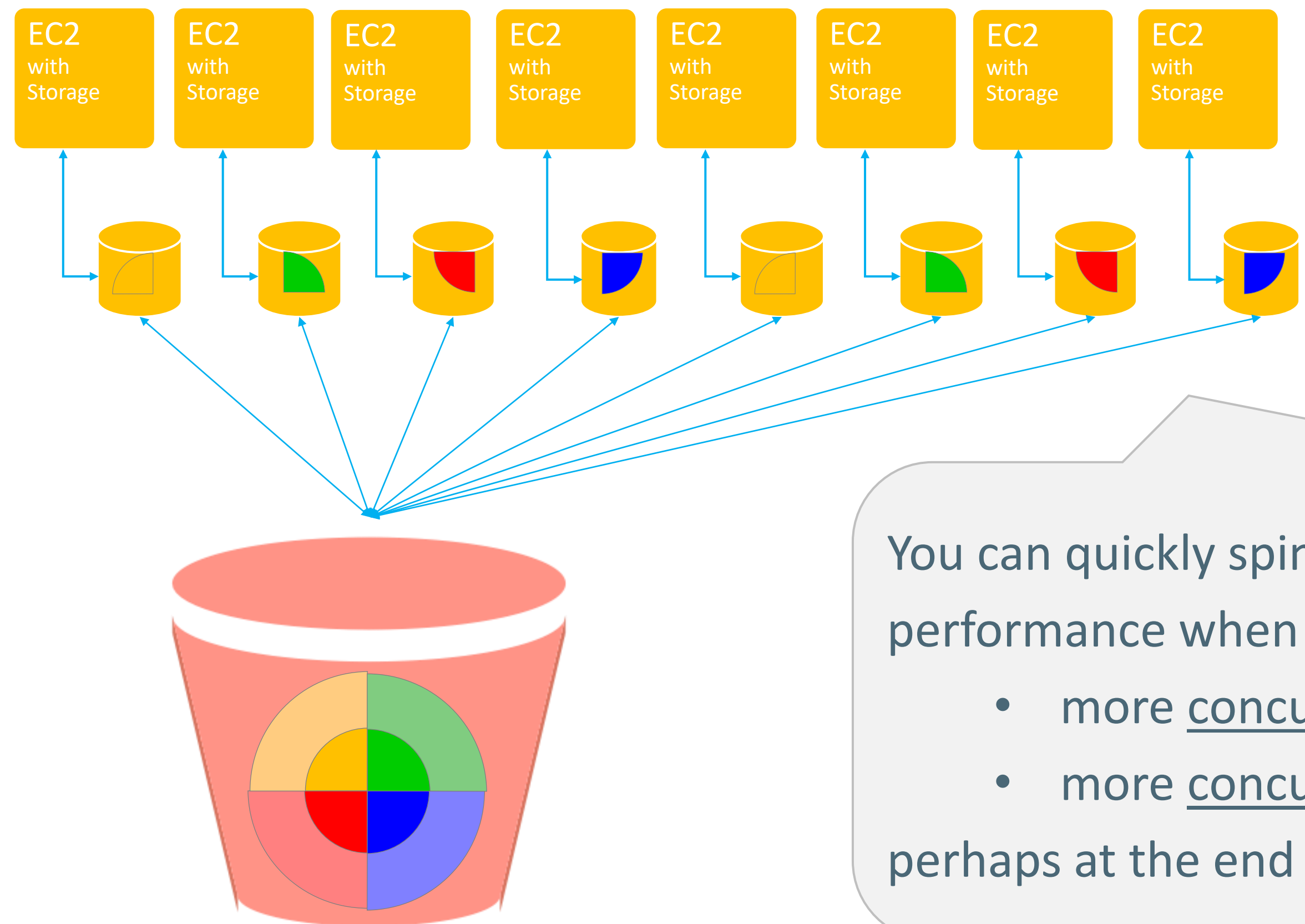
# How does Eon Mode work?



A node may not store the entire shard in the depot.

When a new query can't be satisfied by the depot, the node will run the query directly against S3 while updating the depot.

# Rapid Scaling



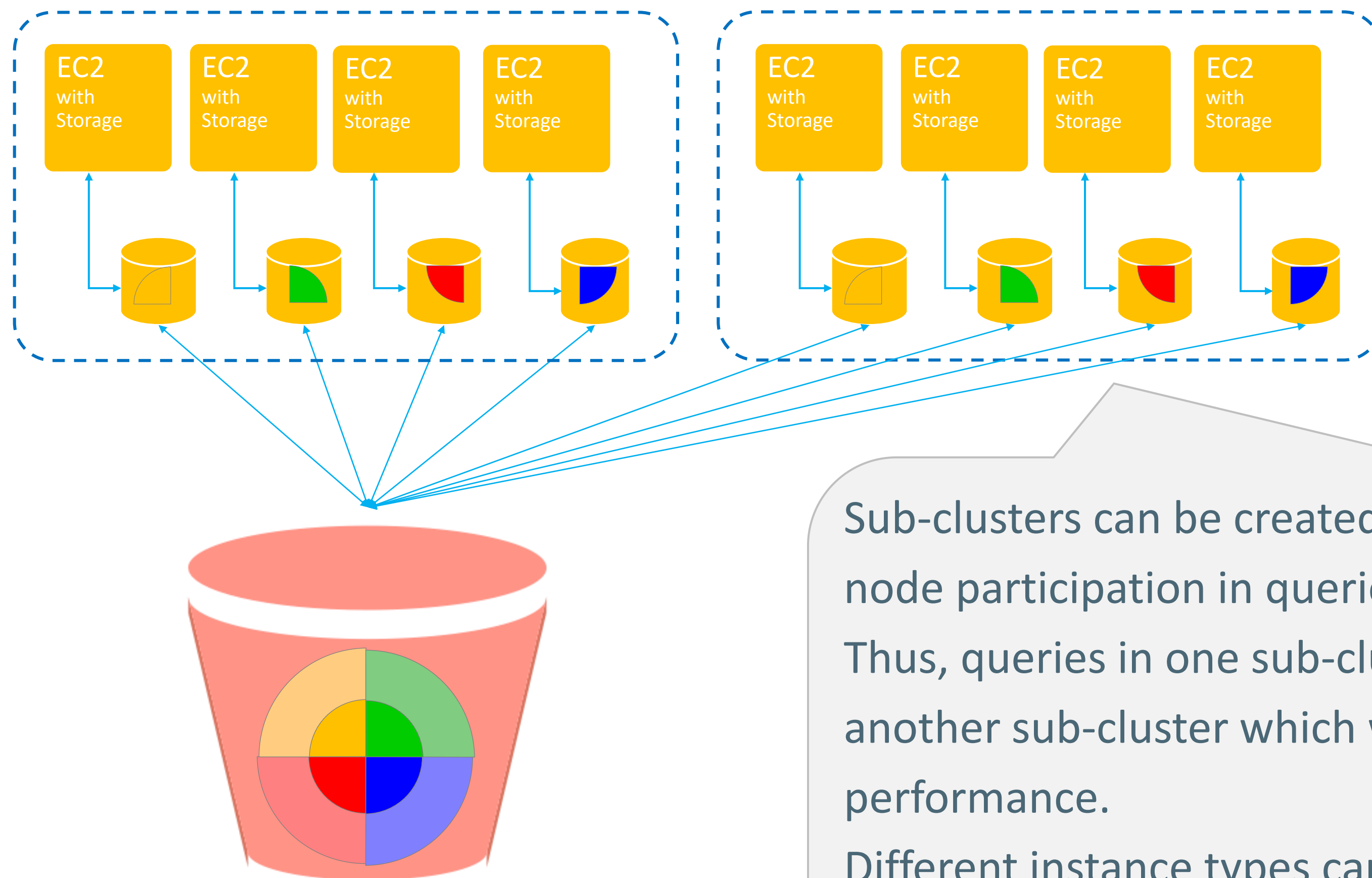
You can quickly spin-up additional nodes to maintain blazing fast performance when you need to run:

- more concurrent queries or
- more concurrent loads

perhaps at the end of a month or simply at the end of the day.



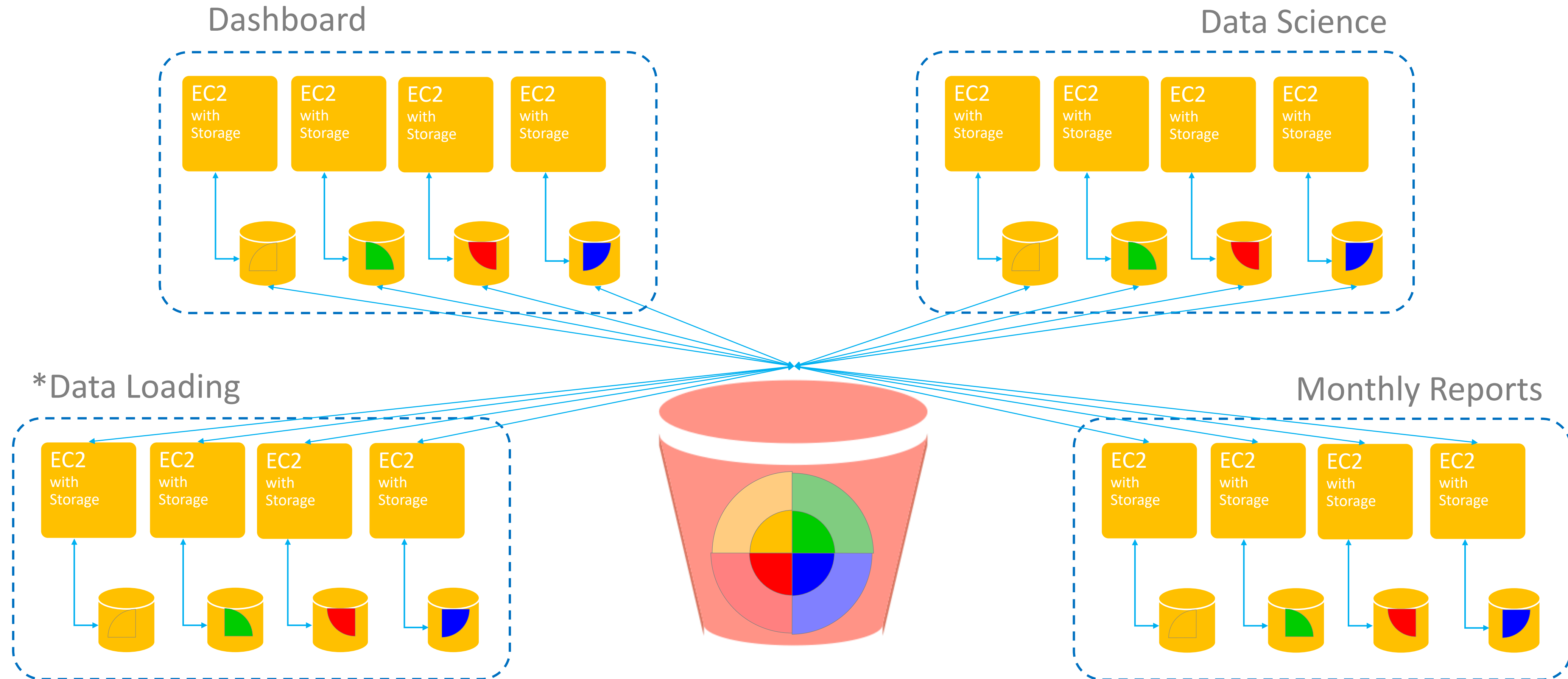
# Isolate Workloads using Sub-clusters



Sub-clusters can be created using Fault Groups. Fault Groups isolate node participation in queries which, in turn, results in Depot isolation. Thus, queries in one sub-cluster will never evict files from the depot of another sub-cluster which would force an S3 fetch and reduce overall performance.

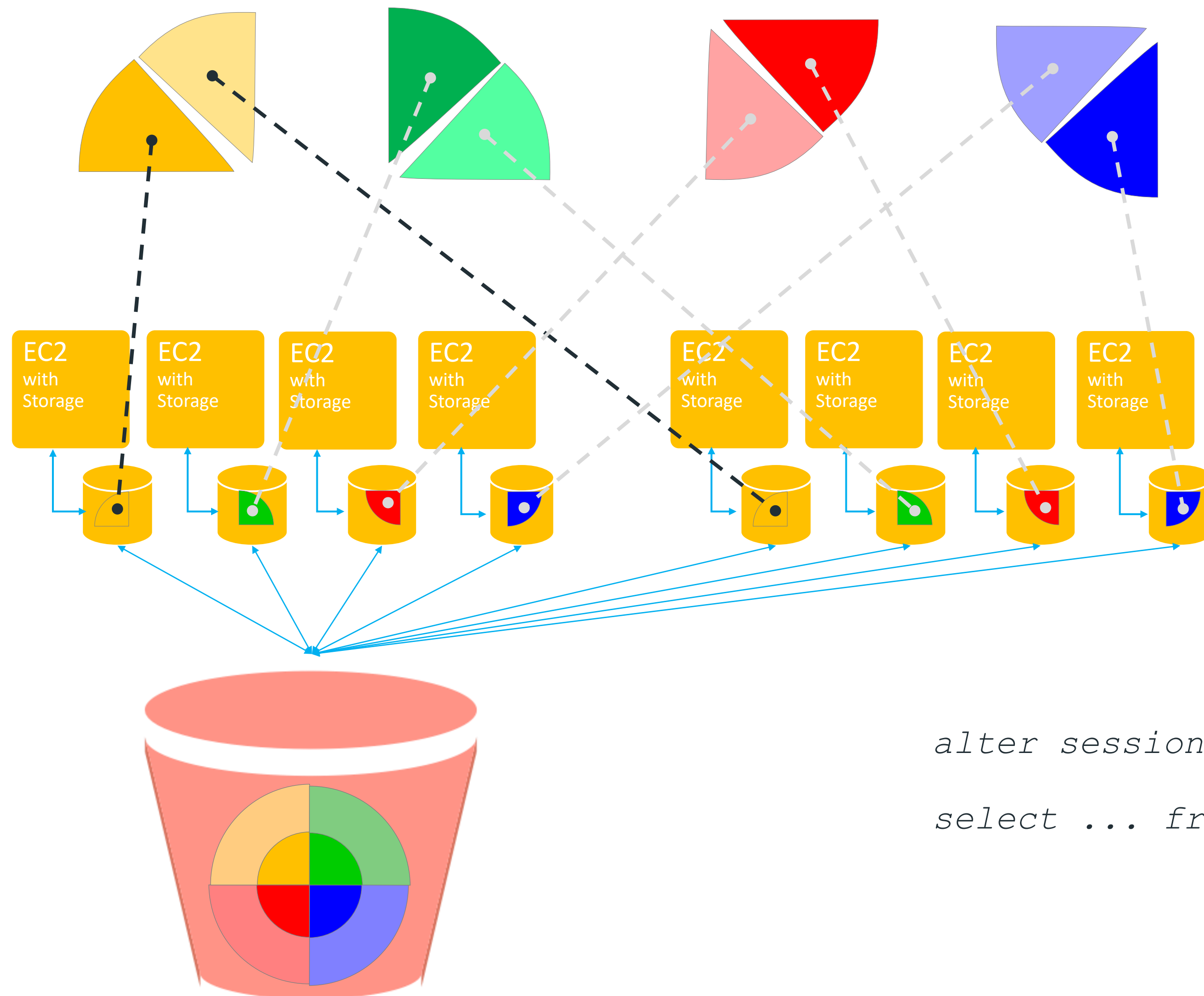
Different instance types can be used in different sub-clusters to better address specific activities of the sub-cluster.

# Isolate Workloads using Sub-clusters



\*Data Loading is an exception to workload isolation. All depots will be updated with most recent data files. We are planning a way to suppress peer-filling in this case.

# Query Crunching for Increase Query Performance



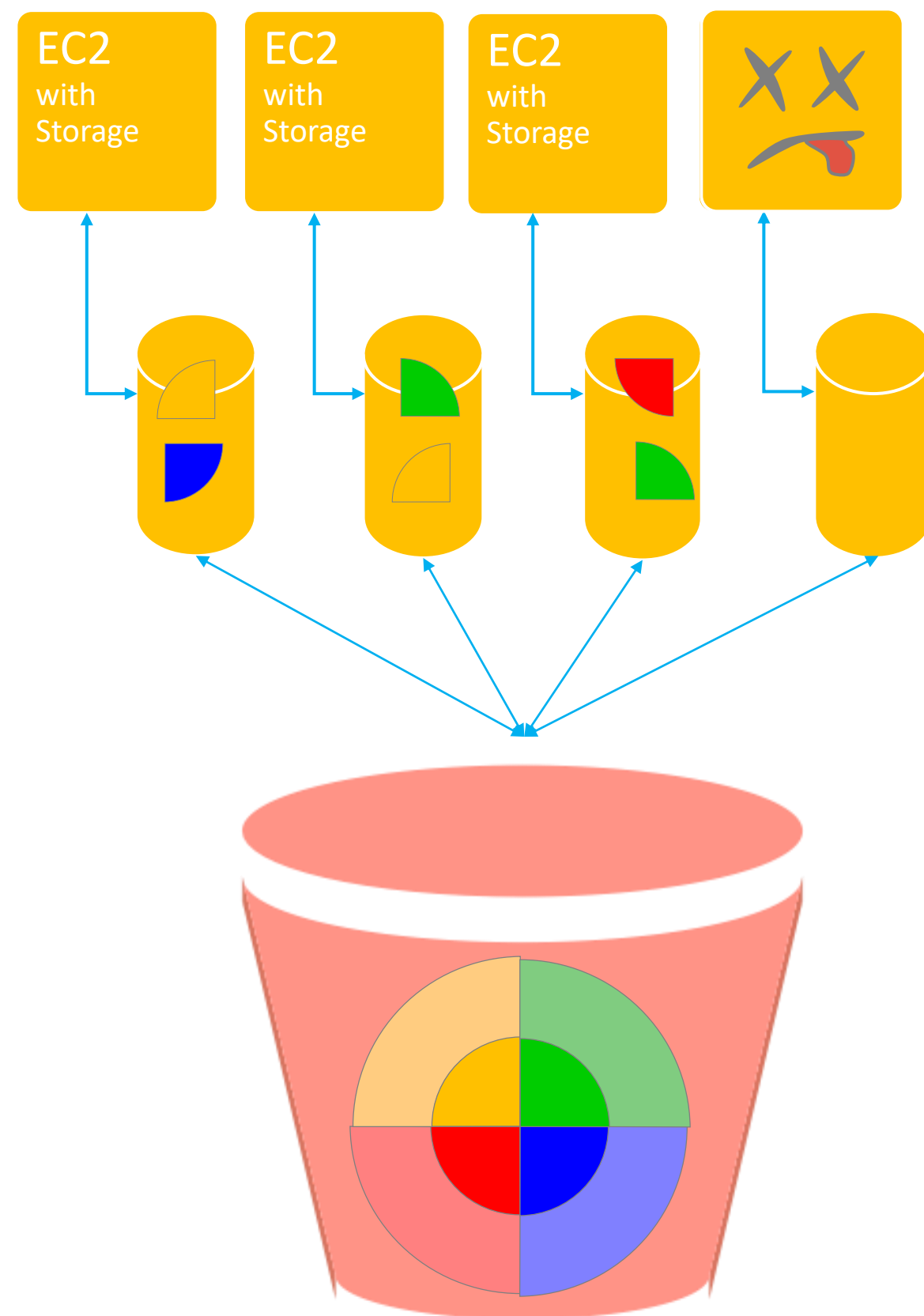
Divide compute across more nodes (shards) in order to increase query performance.

## NOTATION

```
alter session set EnableElasticThroughputScaling = 0;  
select ... from ... where ...
```



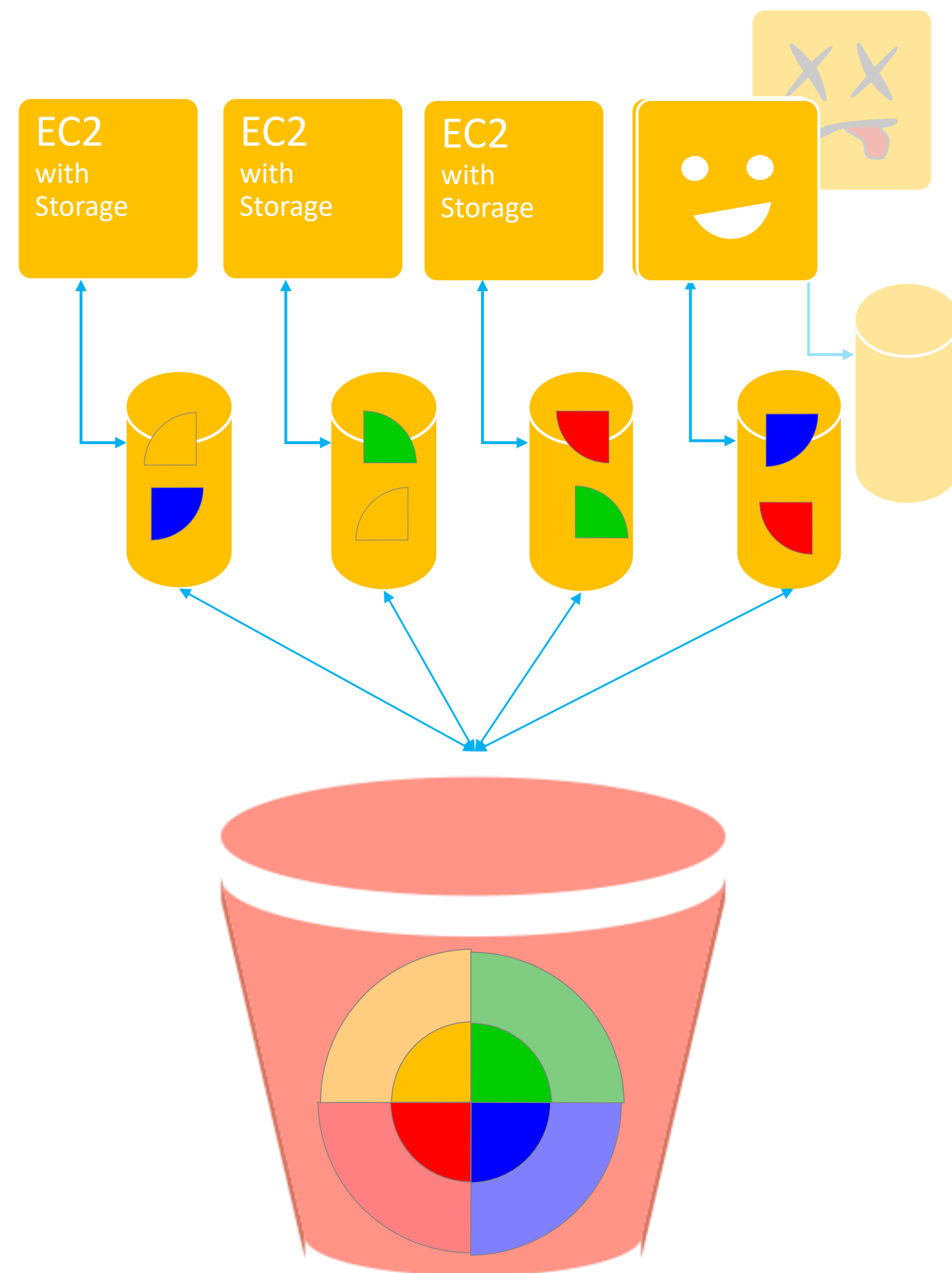
# High Availability through Multiple Shards



Each node is responsible for a multiple shards so that when a node goes down, queries continue to be satisfied by alternate nodes responsible for the shard.

This is similar to K-Safety.

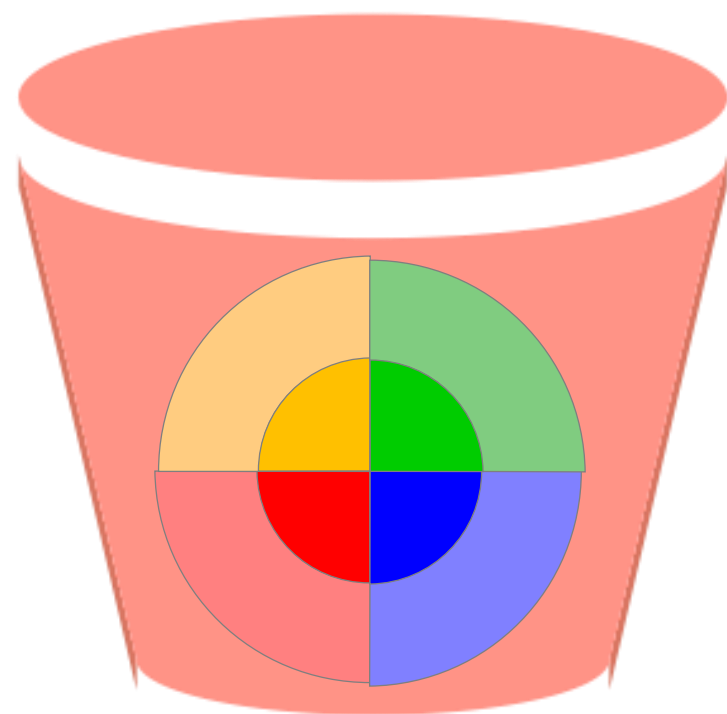
# Rapid Node Recovery



A failed node can be restored or quickly replaced with a new node.

The new node starts up quickly by filling its cache from peer nodes or directly from S3. Performance is maintained because table locks are not required.

# Hibernating

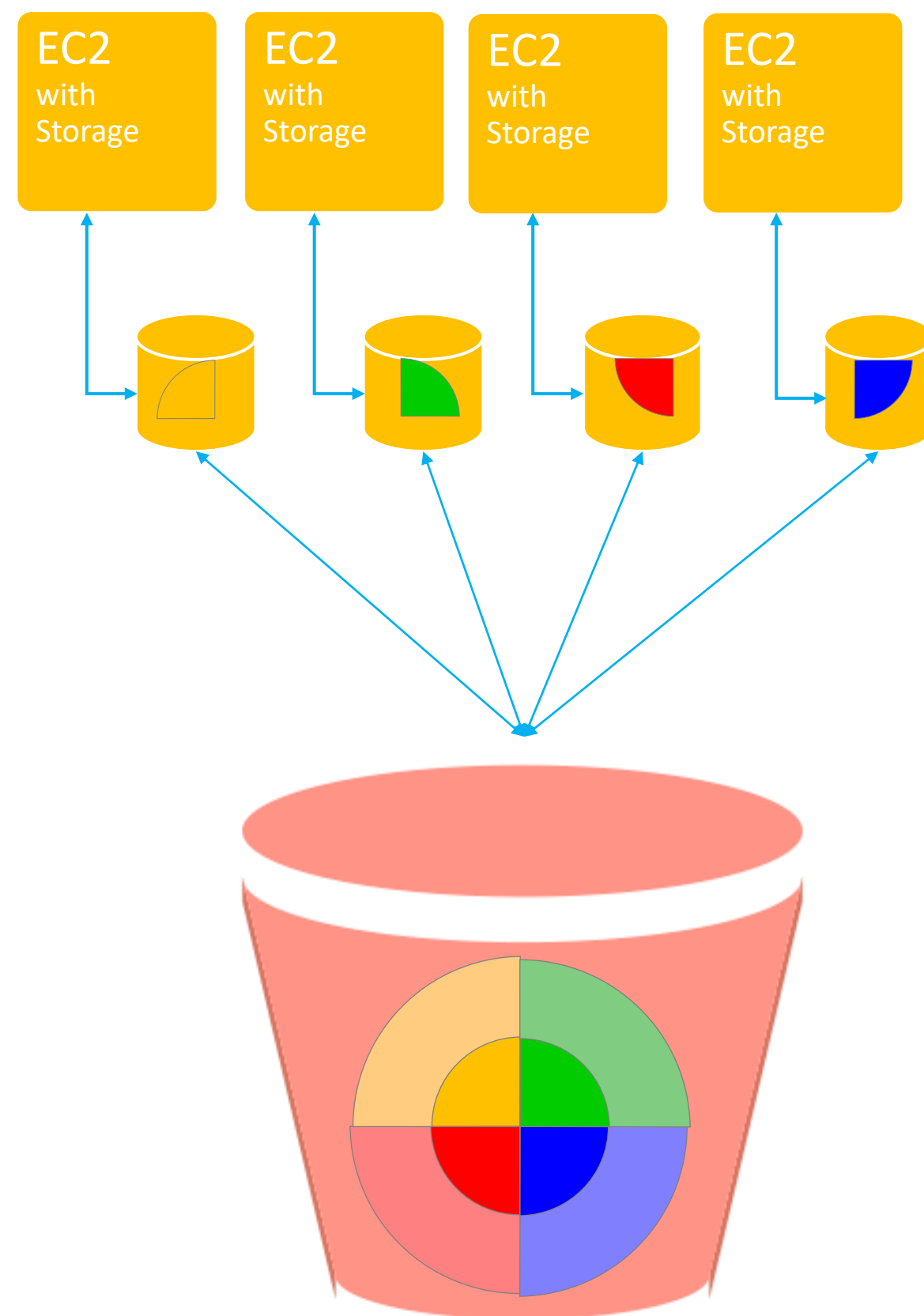


Spin down your entire cluster to save money if your system or project goes dormant or you want to use your Vertica license on other data.

Simply spin it back up when you need it again.



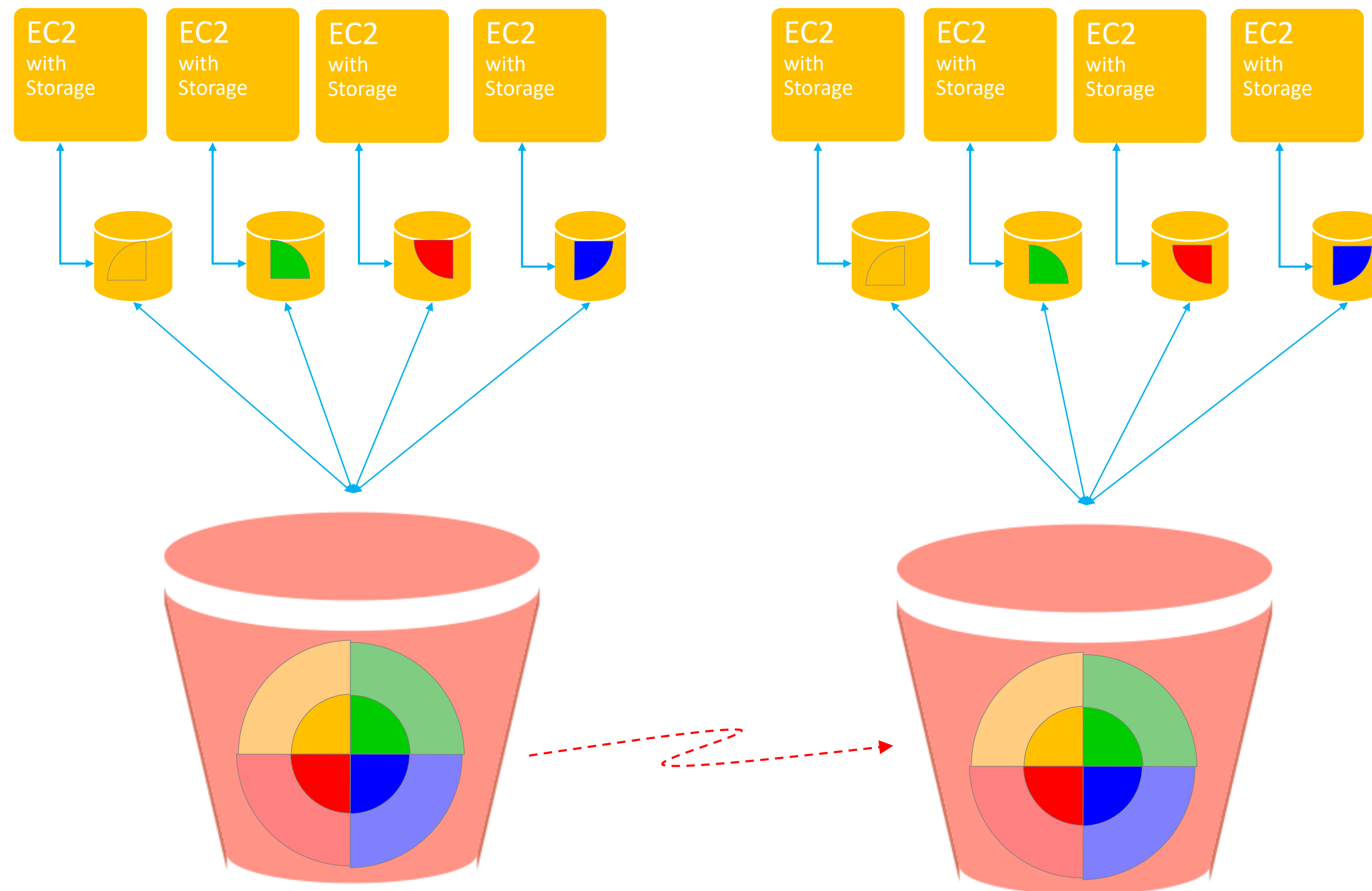
# Revive Database



When you are ready to restore your database, simply:

1. Provision a new cluster and
2. Use `revive_db` to bring the database out of hibernation.

# Replication through S3 Snapshots

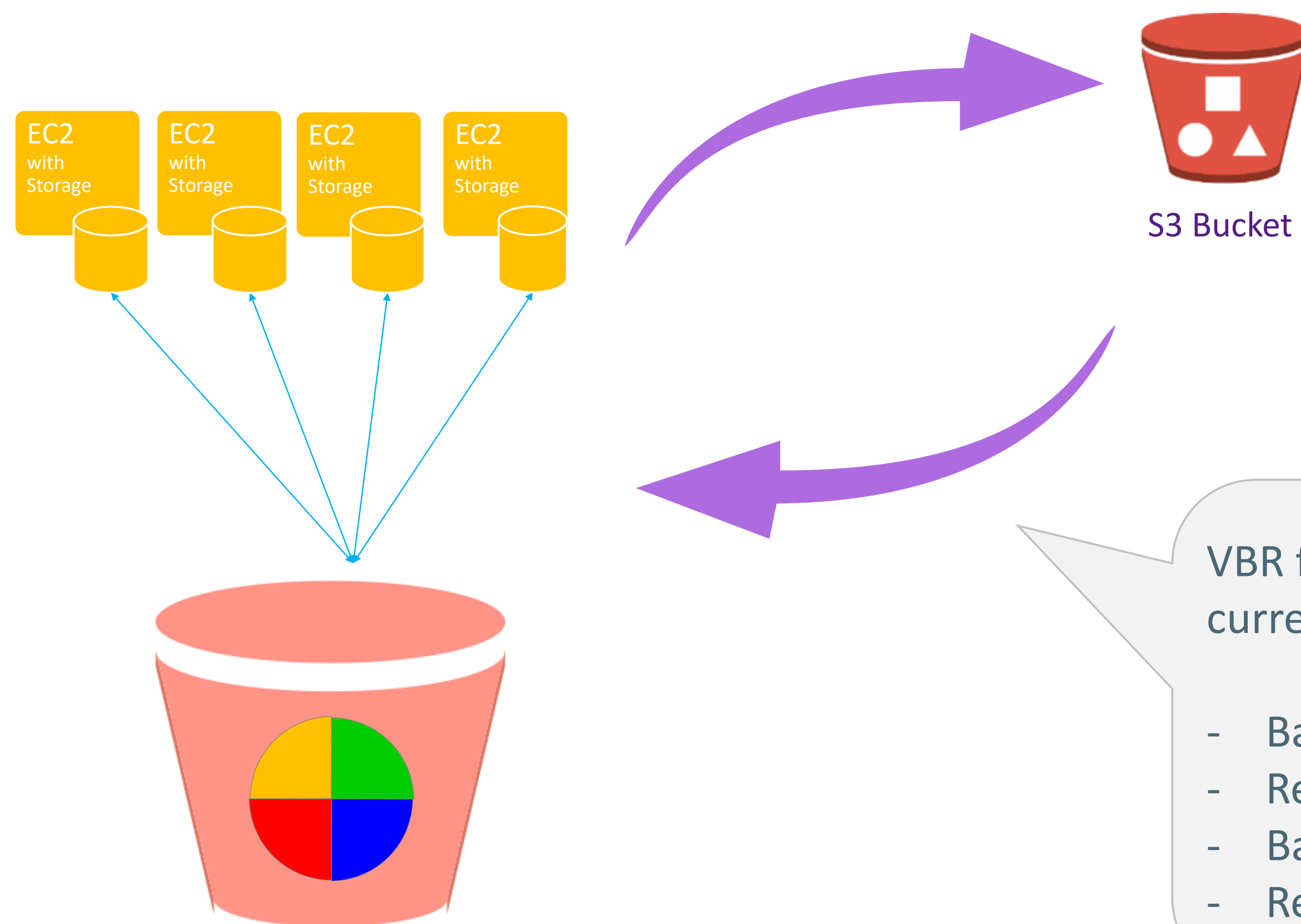


Snapshot your S3 database for replicas on-demand.

For example, dedicate a cluster to service another geographical region or cost center.

“Revive” a replica by simply adding nodes.

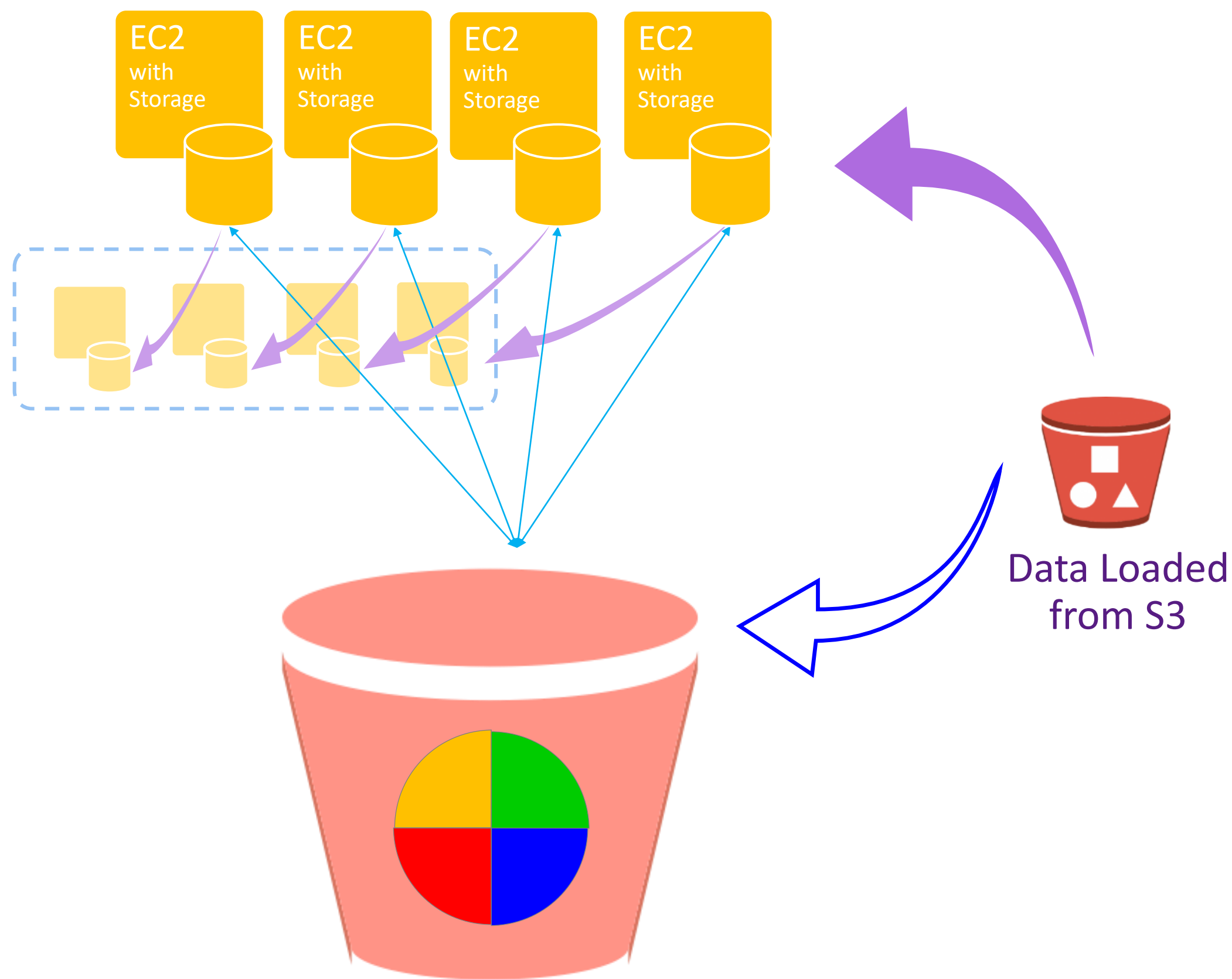
## Vertica Backup and Restore (“VBR”) to S3



VBR for EON will operate much like “Back-up to S3” currently available in Enterprise Mode

- Backup entire DB (Full & Incremental)
- Restore entire DB
- Backup selected Objects
- Restore selected Objects

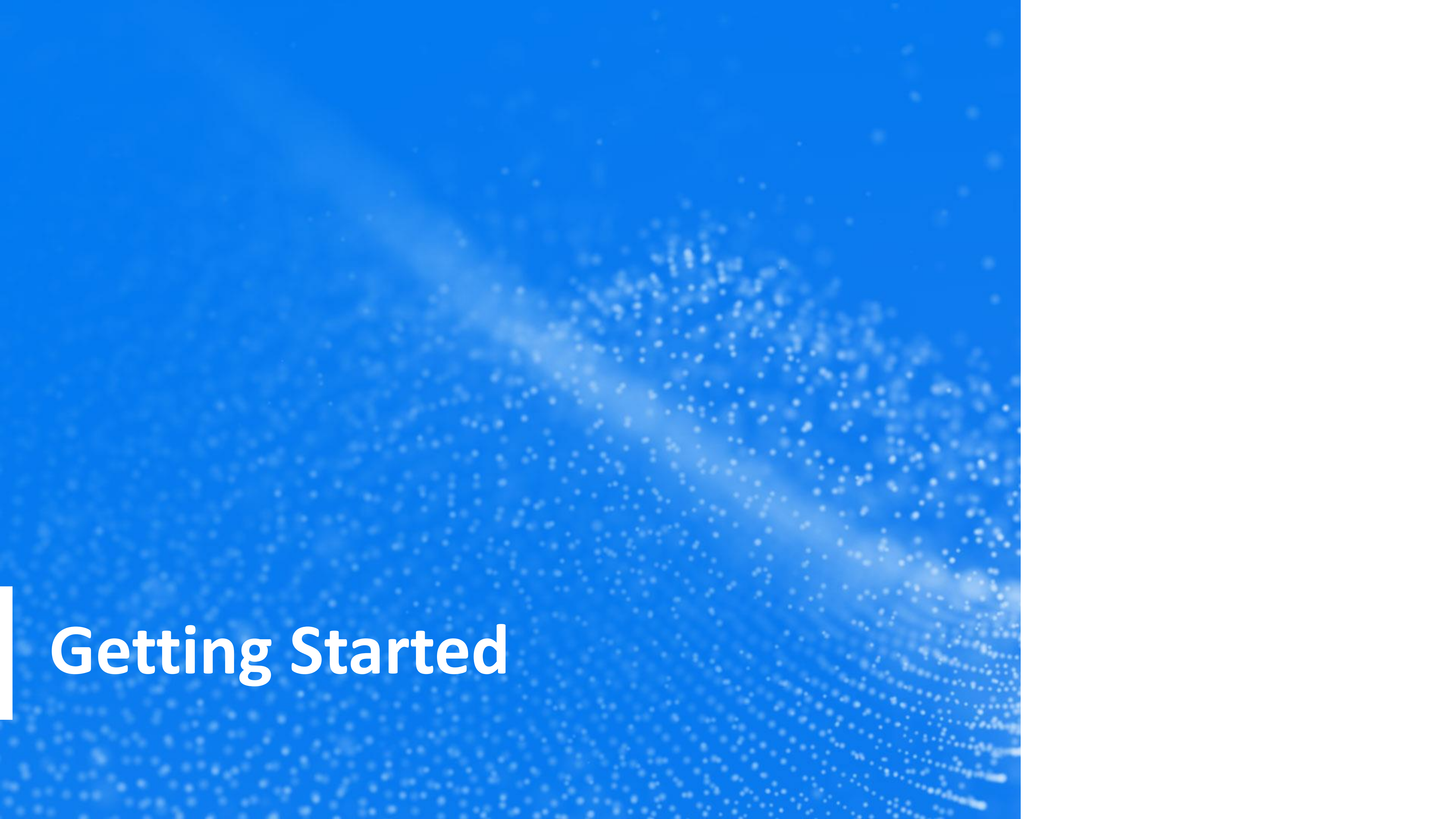
# Loading Data, through Depot and Direct Writes to S3



Data can be loaded into Vertica Eon through the Depot or directly to S3

Depot	Direct to S3
Immediately available for queries	Not constrained by Depot size
Option to share with peer nodes in same subcluster and/or other subclusters	No contention for Depot space

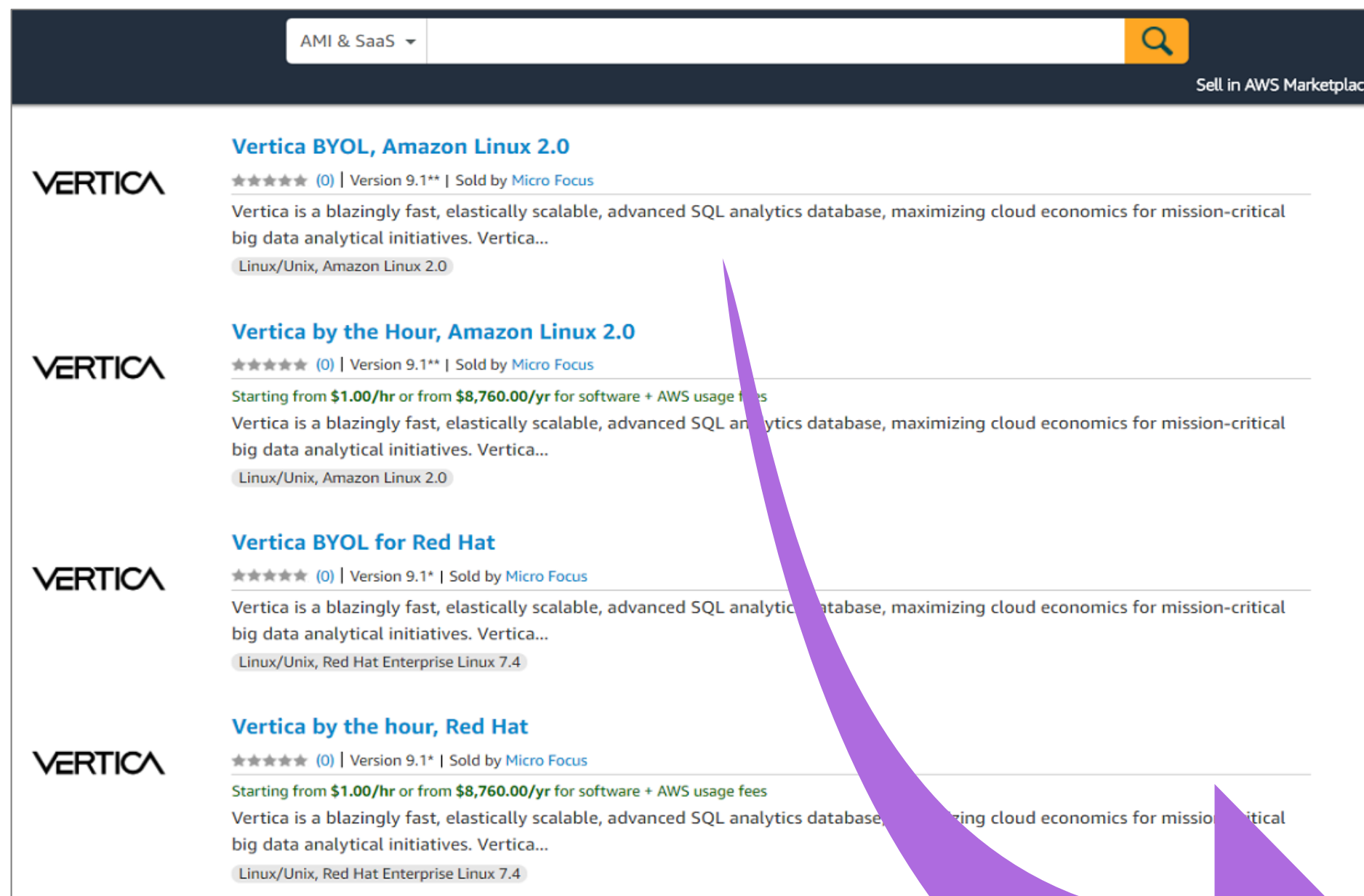




# Getting Started



# Search for “Vertica” in AWS Marketplace (http://aws.amazon.com/marketplace)



AMI & SaaS

Sell in AWS Marketplace

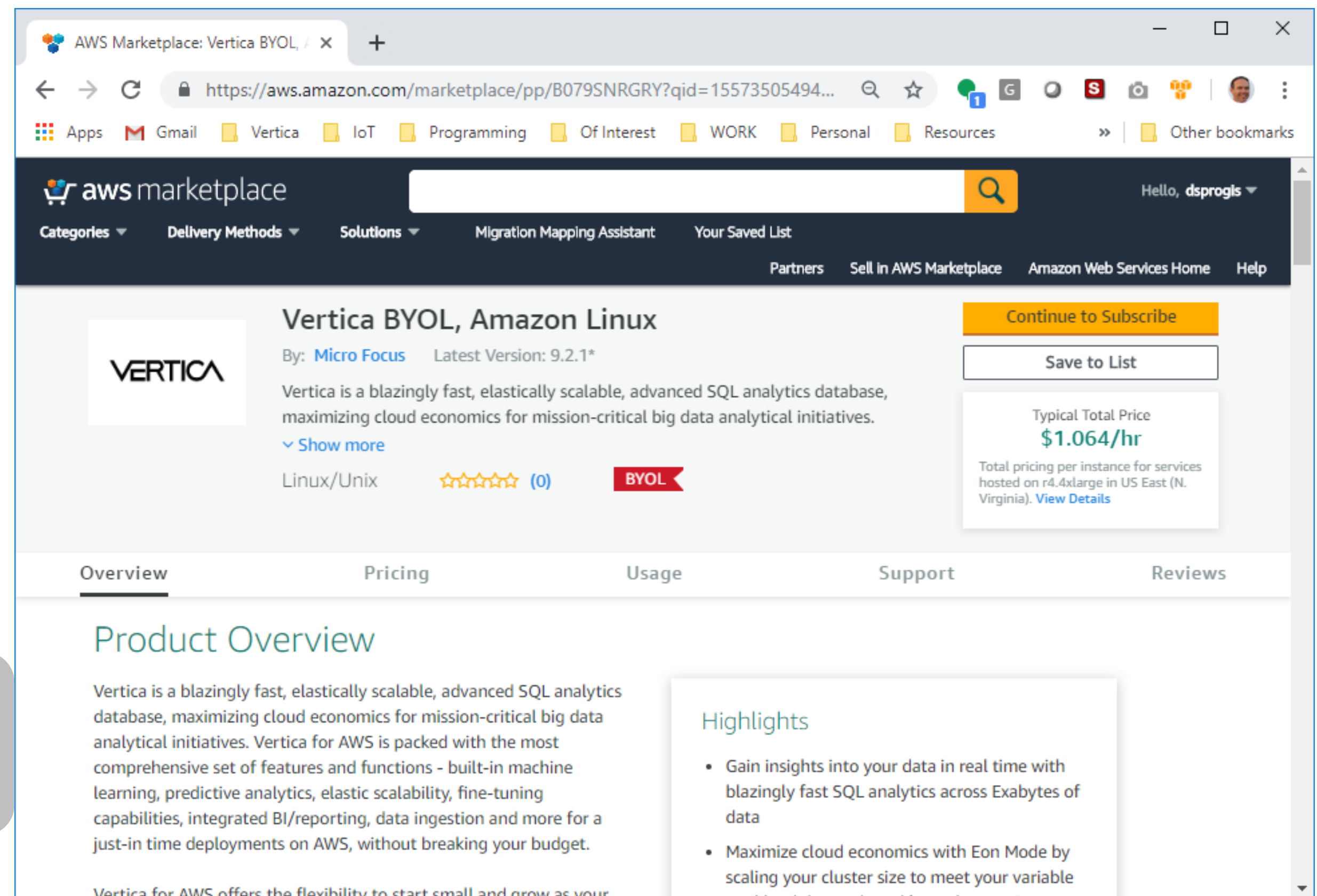
**VERTICA**  
**Vertica BYOL, Amazon Linux 2.0**  
★★★★★ (0) | Version 9.1\*\* | Sold by Micro Focus  
Vertica is a blazingly fast, elastically scalable, advanced SQL analytics database, maximizing cloud economics for mission-critical big data analytical initiatives. Vertica...  
Linux/Unix, Amazon Linux 2.0

**VERTICA**  
**Vertica by the Hour, Amazon Linux 2.0**  
★★★★★ (0) | Version 9.1\*\* | Sold by Micro Focus  
Starting from \$1.00/hr or from \$8,760.00/yr for software + AWS usage fees  
Vertica is a blazingly fast, elastically scalable, advanced SQL analytics database, maximizing cloud economics for mission-critical big data analytical initiatives. Vertica...  
Linux/Unix, Amazon Linux 2.0

**VERTICA**  
**Vertica BYOL for Red Hat**  
★★★★★ (0) | Version 9.1\* | Sold by Micro Focus  
Vertica is a blazingly fast, elastically scalable, advanced SQL analytics database, maximizing cloud economics for mission-critical big data analytical initiatives. Vertica...  
Linux/Unix, Red Hat Enterprise Linux 7.4

**VERTICA**  
**Vertica by the hour, Red Hat**  
★★★★★ (0) | Version 9.1\* | Sold by Micro Focus  
Starting from \$1.00/hr or from \$8,760.00/yr for software + AWS usage fees  
Vertica is a blazingly fast, elastically scalable, advanced SQL analytics database, maximizing cloud economics for mission-critical big data analytical initiatives. Vertica...  
Linux/Unix, Red Hat Enterprise Linux 7.4

Click “Vertica BYOL, Amazon Linux”  
Which provides Community Edition Vertica free



AWS Marketplace: Vertica BYOL, / x +

https://aws.amazon.com/marketplace/pp/B079SNRGRY?qid=15573505494...

Apps Gmail Vertica IoT Programming Of Interest WORK Personal Resources Other bookmarks

aws marketplace Hello, dsprogis

Categories Delivery Methods Solutions Migration Mapping Assistant Your Saved List Partners Sell in AWS Marketplace Amazon Web Services Home Help

**Vertica BYOL, Amazon Linux**  
By: Micro Focus Latest Version: 9.2.1\*  
Vertica is a blazingly fast, elastically scalable, advanced SQL analytics database, maximizing cloud economics for mission-critical big data analytical initiatives.  
Show more  
Linux/Unix ★★★★★ (0) BYOL

Continue to Subscribe  
Save to List

Typical Total Price  
**\$1.064/hr**  
Total pricing per instance for services hosted on r4.xlarge in US East (N. Virginia). View Details

Overview Pricing Usage Support Reviews

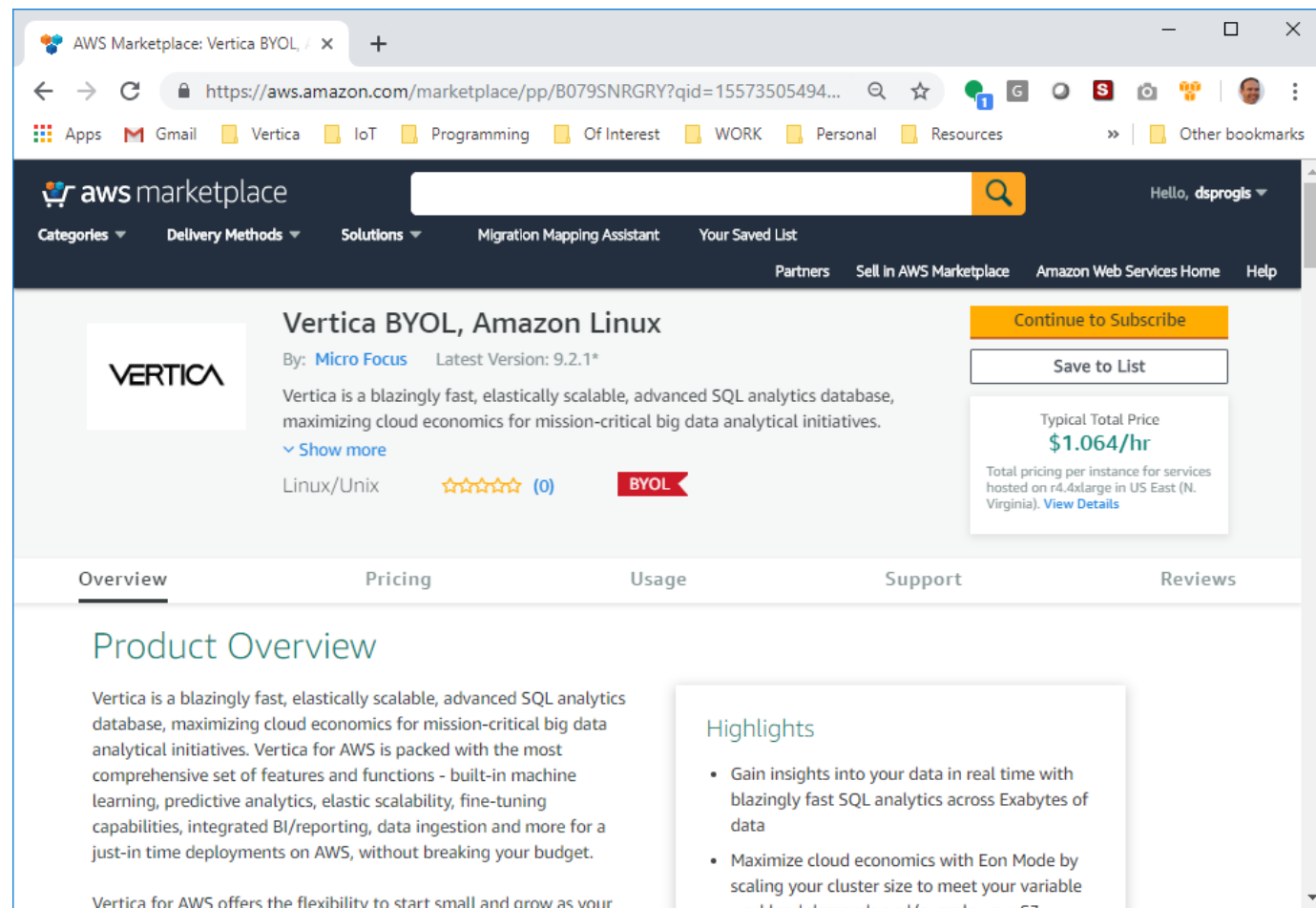
**Product Overview**

Vertica is a blazingly fast, elastically scalable, advanced SQL analytics database, maximizing cloud economics for mission-critical big data analytical initiatives. Vertica for AWS is packed with the most comprehensive set of features and functions - built-in machine learning, predictive analytics, elastic scalability, fine-tuning capabilities, integrated BI/reporting, data ingestion and more for a just-in time deployments on AWS, without breaking your budget.

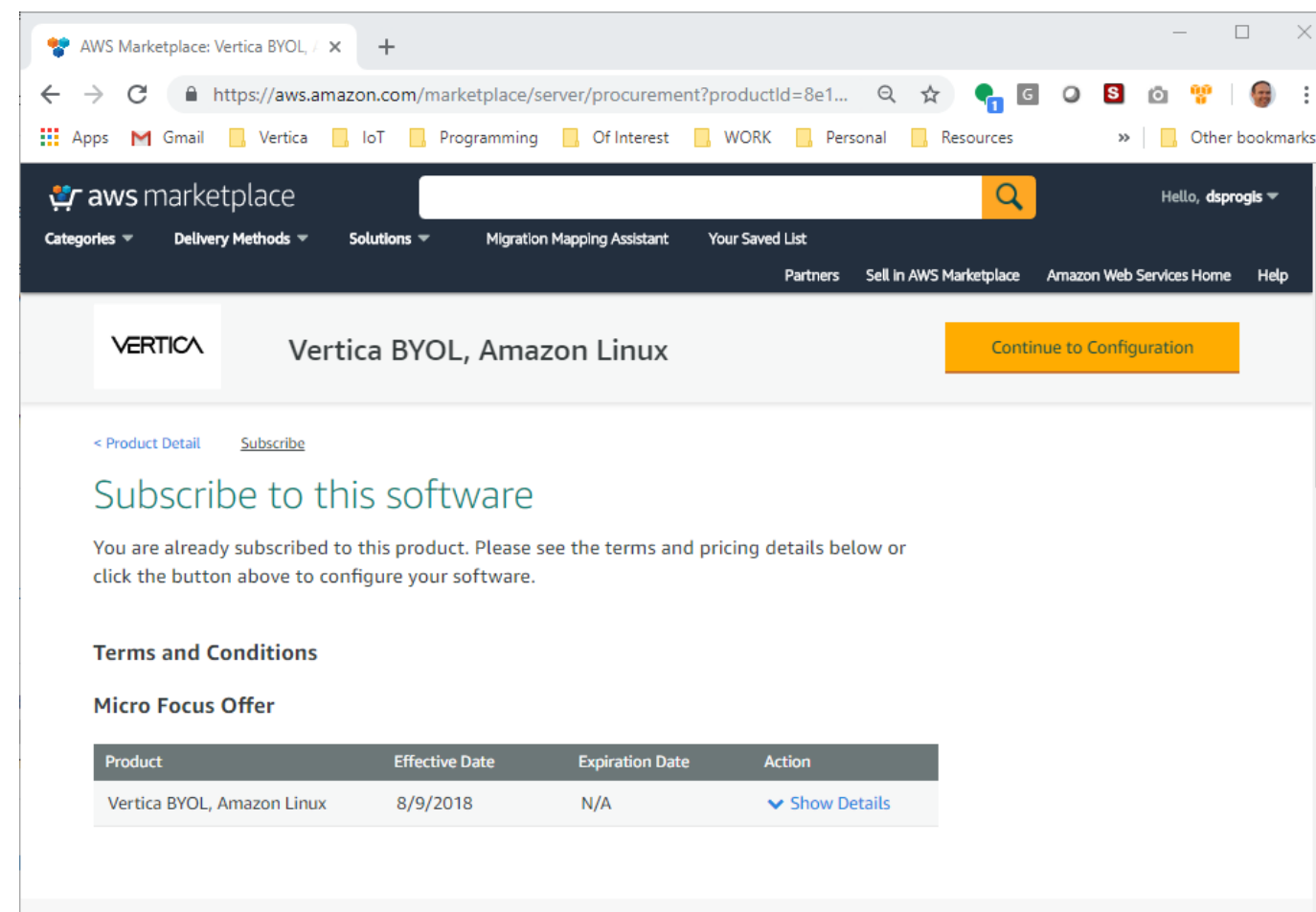
Vertica for AWS offers the flexibility to start small and grow as your

**Highlights**

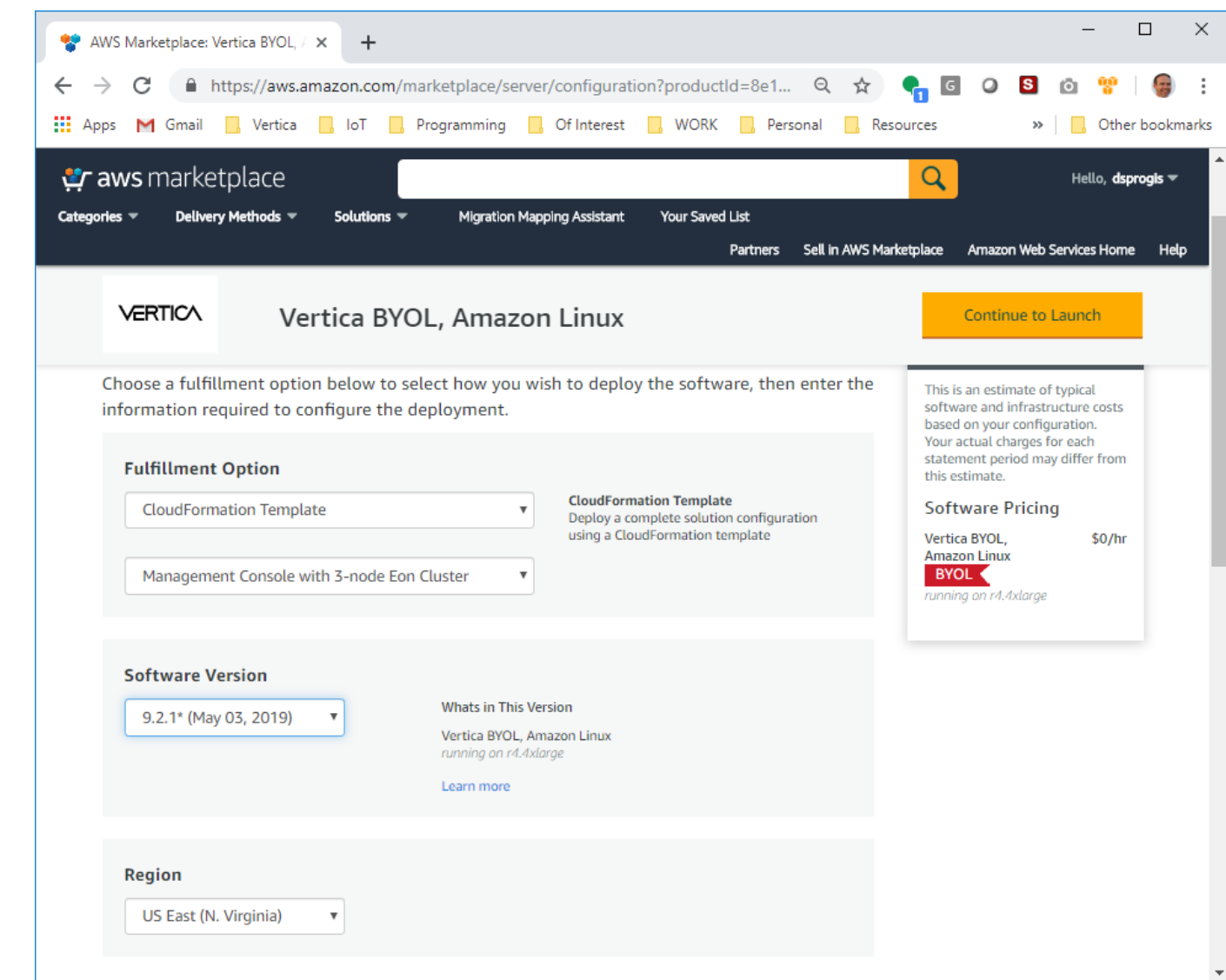
- Gain insights into your data in real time with blazingly fast SQL analytics across Exabytes of data
- Maximize cloud economics with Eon Mode by scaling your cluster size to meet your variable workload demands and requirements



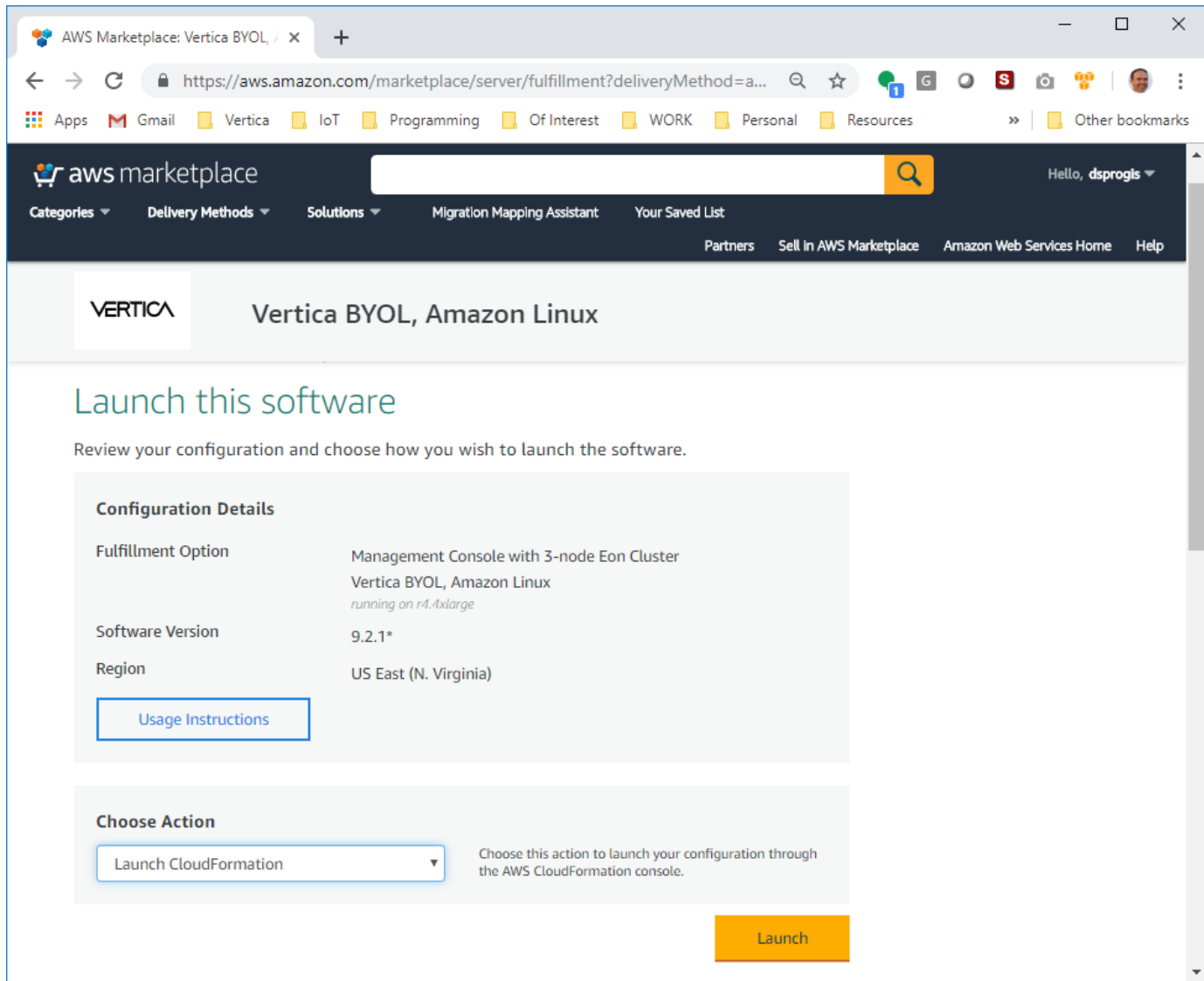
Click “Continue to Subscribe”



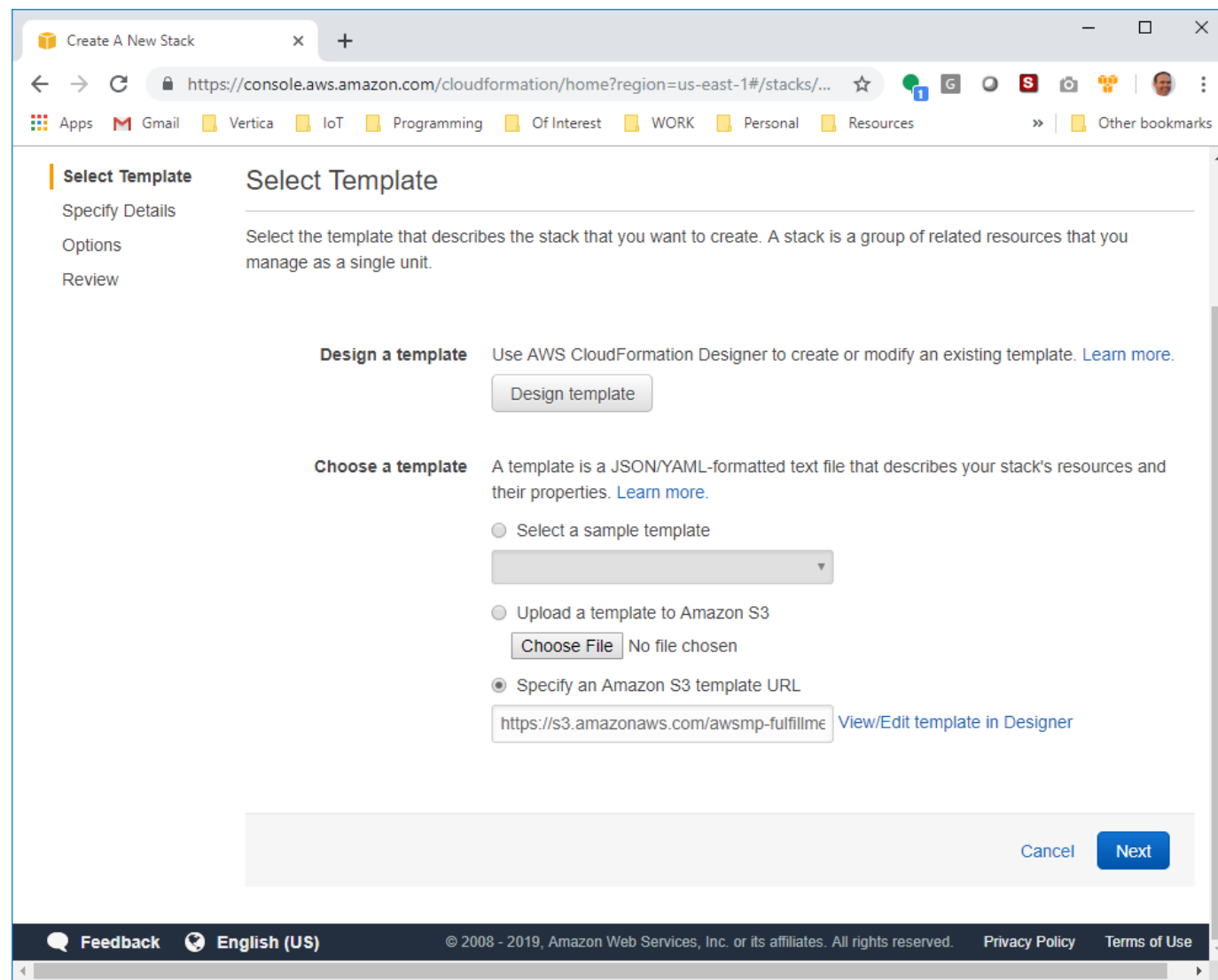
Click “Continue to Configuration”



- Select “CloudFormation Template”
- Select “Management Console with 3-node Eon Cluster”
- Select desired Version
- Select desired Region
- Click “Continue to Launch”



- Select “Launch CloudFormation”
- Click “Launch”



- Click “Next”



aws

Services

Resource Groups

CloudFormation

Stacks

Create Stack

dsprogis @ vertica

N. Virginia

Support

Create stack

Select Template

Specify Details

Options

Review

Specify Details

Specify a stack name and parameter values. You can use or change the default parameter values, which are defined in the AWS CloudFormation template. [Learn more](#).

Stack name

Parameters

Vertica Eon Mode Database Information

Database Name

verticadb

Alphanumeric, must start with alphabet, may also include underscore

Communal Storage

Location in AWS S3 to store database data. Enter S3 bucket followed with a path, for example, s3://bucketname/foldername/dname. S3 bucket must exist and owned by user in the same region as Vertica Cluster instance(s). The communal storage path must not exist

Login

dbadmin

Vertica administrator login

Password

Vertica administrator password

Cluster EC2 type

r4.xlarge

AWS EC2 Instance Type for Vertica Cluster Nodes. Please note, depot data on volumes from c5d/m5d/r5d/i3/i2/i2 instance store will be lost after instance stop.

EBS Volume Type

gp2

For instances other than c5d/m5d/r5d/i3/i2/i2, select from EBS General Purpose SSD (gp2), EBS Provisioned IOPS SSD (io1), or Throughput Optimized HDD (st1).

EBS Volume Size

50

EBS Volume (GB) per volume (Total 8 such volumes for each Vertica node)

io1 Provisioned IOPS

0

Provisioned IOPs for EBS, applicable for io1 Volume Type only

Vertica Management Console Information

Login

uidbadmin

Vertica MC administrator login

Password

Vertica MC administrator password

EC2 type

c4.xlarge

AWS EC2 Instance Type for Vertica Management Console

Authentication method

IAM Role Instance Profile

Authentication method for Vertica AWS instance provisioning and cluster management

Access Information

Key Pair

Search

Name of an existing EC2 key pair for SSH access to the EC2 instances

CIDR Range

The range of IP addresses to allow access to Vertica Management Console, JDBC client connection to Vertica nodes, and SSH access to the all nodes. Use CIDR notation, for example: 192.0.2.0/24 or 0.0.0.0/0

Notification

Enable Notifications?

No

Enable Cloudwatch Alarm Notifications

Email

Needed if you enable notifications. You must confirm to subscription email notice from 'AWS Notification' to receive alerts for this cluster

Other

Availability Zone

Search

Availability Zone of the Subnet

Accept Terms and Conditions

You must accept End User License Agreement (EULA), see <https://software.microfocus.com/en-us/about/end-user-agreement-terms>

Cancel

Previous

Next

Feedback

English (US)

© 2008 - 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy

Terms of Use

- Enter new name for CloudFormation Stack

- Enter new name storage path in existing bucket

- Enter new database password

- Enter new Management Console password

- Select Key Pair
- Enter CIDR Range

- Select Availability Zone
- Accept Terms and Conditions

- Click “Next”



aws

Services

Resource Groups

CloudFormation

Stacks

Create Stack

dsprogis @ vertica

N. Virginia

Support

Create stack

Select Template

Specify Details

Options

Review

Options

Tags

You can specify tags (key-value pairs) for resources in your stack. You can add up to 50 unique key-value pairs for each stack. [Learn more.](#)

	Key (127 characters maximum)	Value (255 characters maximum)	
1	<input type="text"/>	<input type="text"/>	<a href="#">+</a>

Permissions

You can choose an IAM role that CloudFormation uses to create, modify, or delete resources in the stack. If you don't choose a role, CloudFormation uses the permissions defined in your account. [Learn more.](#)

IAM Role

Choose a role (optional)

Enter role am

Rollback Triggers

Rollback triggers enable you to have AWS CloudFormation monitor the state of your application during stack creation and updating, and to rollback that operation if the application breaches the threshold of any of the alarms you've specified. [Learn more](#)

Monitoring Time ⓘ

0-180

Minutes

Minimum value of 0. Maximum value of 180.

Available triggers remaining: 5

	Type	ARN (Amazon Resource Name)	
1	AWS::CloudWatch::Alarm	<input type="text"/>	<a href="#">+</a>

Advanced

You can set additional options for your stack, like notification options and a stack policy. [Learn more.](#)

Cancel

Previous

Next

Feedback

English (US)

© 2008 - 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy

Terms of Use

- Click “Next”

aws

Services

Resource Groups

CloudFormation

Stacks

Create Stack

dsprogis @ vertica

N. Virginia

Support

Create stack

Select Template

Specify Details

Options

Review

Review

Template URL

https://s3.amazonaws.com/awsmp-fulfillment-cf-templates-prod/8e10071d-8a2f-4eb3-8bc8-d4d771a89011.dd130ee3-807d-4f2f-c592-f856191f1dae.template

Description

This template deploys 3 node Vertica Eon mode database cluster with 1 Vertica Management Console with provisioning host --AWSMP::8e10071d-8a2f-4eb3-8bc8-d4d771a89011::dd130ee3-807d-4f2f-c592-f856191f1dae

Estimate cost

Link is not available

Details

Stack name:

DHS-TEST-01

Vertica Eon Mode Database Information

DatabaseName

verticadb

CommunalStorage

Dbadmin

dbadmin

DbadminPassword

.....

InstanceType

r4.xlarge

VolumeType

gp2

VolumeSize

50

IOPS

0

Vertica Management Console Information

McDbadmin

uidbadmin

McDbadminPassword

.....

MCInstanceType

c4.xlarge

AWSAuthenticate

IAM Role Instance Profile

Access Information

KeyName

sprogis

SSHLocation

0.0.0.0/0

Notification

EnableNotifications

No

Email

Other

SubnetAZ

us-east-1a

EULA

Yes

Options

Tags

No tags provided

Rollback Triggers

No monitoring time provided

No rollback triggers provided

Advanced

Notification

Termination Protection

Disabled

Timeout

none

Rollback on failure

Yes

Capabilities

i

The following resource(s) require capabilities: [AWS::IAM::Role]

This template contains Identity and Access Management (IAM) resources that might provide entities access to make changes to your AWS account. Check that you want to create each of these resources and that they have the minimum required permissions. [Learn more.](#)

☐ I acknowledge that AWS CloudFormation might create IAM resources.

Quick Create Stack

(Create stacks similar to this one, with most details auto-populated)

Cancel

Previous

Create

Feedback

English (US)

© 2008 - 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy

Terms of Use

• Review Selections

• Click “I acknowledge ...”

• Click “Create”

aws

Services

Resource Groups

dsprogis @ vertica

N. Virginia

Support

CloudFormation

Stacks

The redesigned AWS CloudFormation console is available now

We've completely redesigned the console to improve the overall look and feel. Try it out now and provide us feedback.

We have recently released a number of improvements in the new console, thanks to your feedback.

The changes include a new layout for faster access to information, resizable columns, and availability in 5 additional regions (AWS GovCloud (US-West and US-East), China (Beijing), China (Ningxia), EU (Stockholm)). Please tell us what you think!

Create Stack

Actions

Design template

Filter: Active

By Stack Name

Showing 14 stacks

	Stack Name	Created Time	Status	Drift Status	Description
<input checked="" type="checkbox"/>	DHS-TEST-01	2019-05-08 17:55:58 UTC-0400	CREATE_IN_PROGRE...	NOT_CHECKED	This templat
<input type="checkbox"/>	mlpotter-a-psa-0508	2019-05-08 15:40:04 UTC-0400	CREATE_COMPLETE	NOT_CHECKED	This templat
<input type="checkbox"/>	huam-ansible-test	2019-04-24 15:18:02 UTC-0400	CREATE_COMPLETE	NOT_CHECKED	This templat
<input type="checkbox"/>	NataliaStack2April19	2019-04-19 14:59:29 UTC-0400	CREATE_COMPLETE	NOT_CHECKED	This templat
<input type="checkbox"/>	Natalia-April18	2019-04-18 11:24:41 UTC-0400	CREATE_COMPLETE	NOT_CHECKED	This templat
<input type="checkbox"/>	JoeTest921-rc0	2019-04-16 07:54:35 UTC-0400	CREATE_COMPLETE	NOT_CHECKED	This templat
<input type="checkbox"/>	mlpotter4emichaud-a-psa-0409	2019-04-09 16:16:34 UTC-0400	CREATE_COMPLETE	NOT_CHECKED	This templat
<input type="checkbox"/>	NataliaApr8	2019-04-08 14:05:45 UTC-0400	CREATE_COMPLETE	NOT_CHECKED	This templat

Overview

Outputs

Resources

Events

Template

Parameters

Tags

Stack Policy

Change Sets

Rollback Triggers

Key	Value	Description	Export Name
No outputs found.			

Feedback

English (US)

© 2008 - 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy

Terms of Use

Select your Stack

Click Refresh Button

Select "Outputs" Tab

Observe connection options

aws

Services

Resource Groups

CloudFormation

Stacks

The redesigned AWS CloudFormation console is available now

We've completely redesigned the console to improve the overall look and feel. Try it out now and provide us feedback.

We have recently released a number of improvements in the new console, thanks to your feedback.

The changes include a new layout for faster access to information, resizable columns, and availability in 5 additional regions (AWS GovCloud (US-West and US-East), China (Beijing), China (Ningxia), EU (Stockholm)). Please tell us what you think!

Create Stack

Actions

Design template

Filter: Active

By Stack Name

Showing 14 stacks

	Stack Name	Created Time	Status	Drift Status	Description
<input checked="" type="checkbox"/>	DHS-TEST-01	2019-05-08 17:55:58 UTC-0400	CREATE_COMPLETE	NOT_CHECKED	This templat
<input type="checkbox"/>	mlpotter-a-psa-0508	2019-05-08 15:40:04 UTC-0400	CREATE_COMPLETE	NOT_CHECKED	This templat
<input type="checkbox"/>	huam-ansible-test	2019-04-24 15:18:02 UTC-0400	CREATE_COMPLETE	NOT_CHECKED	This templat
<input type="checkbox"/>	NataliaStack2April19	2019-04-19 14:59:29 UTC-0400	CREATE_COMPLETE	NOT_CHECKED	This templat
<input type="checkbox"/>	Natalia-April18	2019-04-18 11:24:41 UTC-0400	CREATE_COMPLETE	NOT_CHECKED	This templat
<input type="checkbox"/>	JoeTest921-rc0	2019-04-16 07:54:35 UTC-0400	CREATE_COMPLETE	NOT_CHECKED	This templat
<input type="checkbox"/>	mlpotter4emichaud-a-psa-0409	2019-04-09 16:16:34 UTC-0400	CREATE_COMPLETE	NOT_CHECKED	This templat
<input type="checkbox"/>	NataliaApr8	2019-04-08 14:05:45 UTC-0400	CREATE_COMPLETE	NOT_CHECKED	This templat

Overview

Outputs

Resources

Events

Template

Parameters

Tags

Stack Policy

Change Sets

Rollback Triggers

Key	Value	Description	Export Name
VSQl	( vsql -U dbadmin -h52.86.120.208 -w * ***)	Connect to database using vsql	
ClusterPrivateIPs	10.11.12.10,10.11.12.20,10.11.12.30	Internal IPs of the nodes in the cluster	
ManagementConsole	<a href="https://3.215.100.62:5450/webui">https://3.215.100.62:5450/webui</a> (login: uidbadmin, password: ****)	Access Management Console	
EULA	<a href="https://software.microfocus.com/en-us/about/end-user-agreement-terms">https://software.microfocus.com/en-us/about/end-user-agreement-terms</a>	End User License Agreement	
Documentation	<a href="https://my.vertica.com/docs/latest/HTML/index.htm">https://my.vertica.com/docs/latest/HTML/index.htm</a>	Vertica Documentation	

Feedback

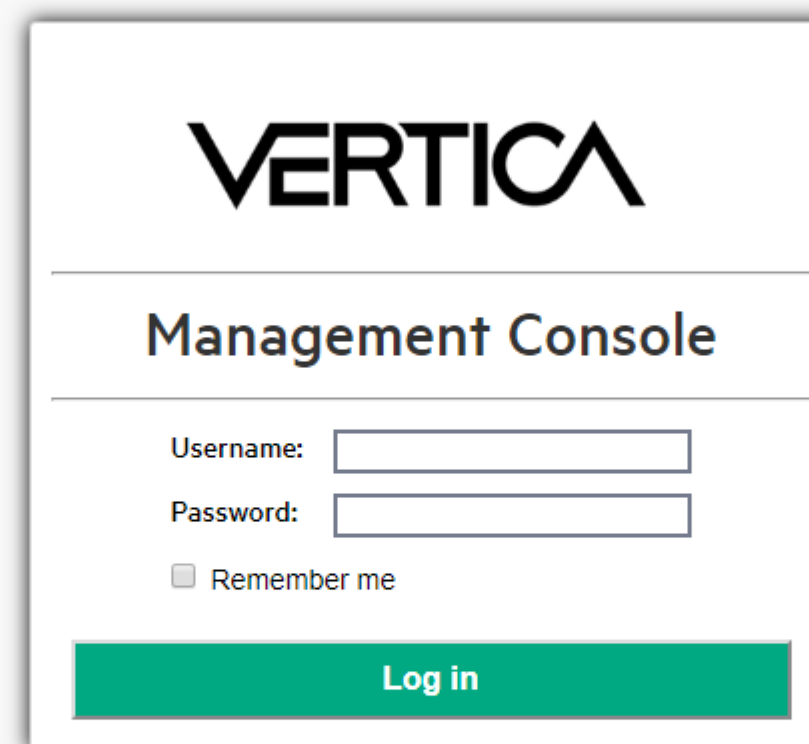
English (US)

© 2008 - 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy

Terms of Use

Click link to Management Console



The image shows a screenshot of the Vertica Management Console login interface. At the top is the 'VERTICA' logo. Below it is the title 'Management Console'. The login form includes a 'Username:' label with an input field, a 'Password:' label with an input field, and a 'Remember me' checkbox. A green 'Log in' button is positioned at the bottom of the form. Below the form, there is a small copyright notice: '© Copyright 2007 EntiT Software LLC, a Micro Focus company All Rights Reserved'.

Login using credentials specified in  
CloudFormation Stack creation



Vertica Management Console

uidbadminLog out?

Provision

Create or import a Vertica Database Cluster

Create a new Vertica Database Cluster

Provision & Revive an Eon Mode Database

Import a Vertica Database Cluster

Additional Import Options

Manage

See all monitored clusters and databases and their storage usage

View Your Infrastructure


MC Tools

MC Settings

Message Center

MC Diagnostics

Recent Databases

Database Name	Database Mode	IP	Node Count	Actions
 verticadb	Eon	10.11.12.10 ...	3	<a href="#">See Fast Tasks</a>   <a href="#">Go to database</a>

Documentation and Community

Access [online documentation](#)

Connect with the [Vertica Community](#)

Participate in our [blog](#)

Video Tutorials And Training

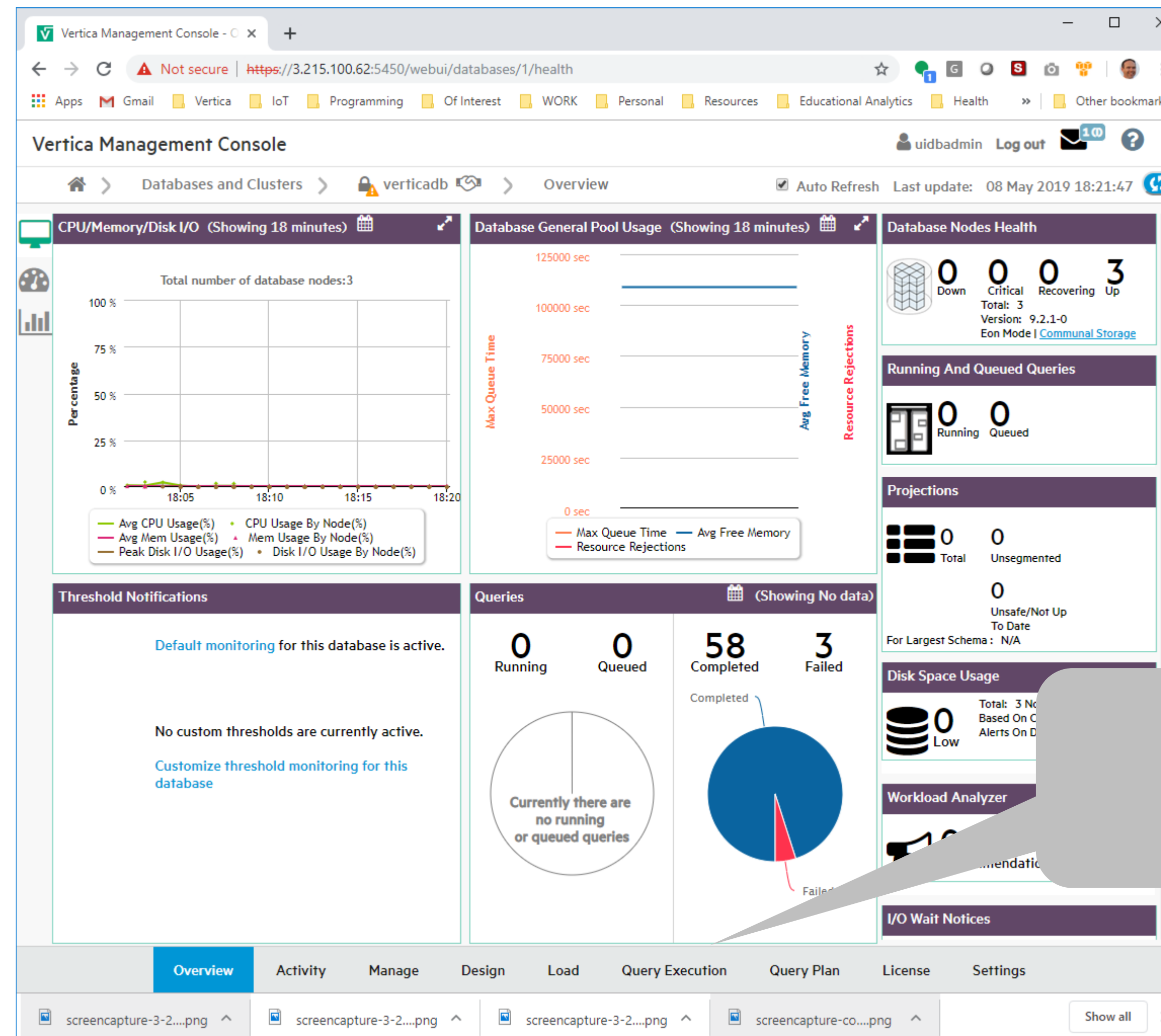
[Online training](#) is available to members of the Vertica community for free.

Visit the [Vertica Community channel](#) to view videos about using Vertica and the Management Console.

Vertica Management Console, v9.2.1 · Build: 9.2.1 - 0 · AMI: 9.2.1-0

© 2007 EntIT Software LLC, a Micro Focus company. All rights reserved.

Click “Go to database”



Click "Query Execution"

Vertica Management Console - C x

Not secure | https://3.215.100.62:5450/webui/admin/databases/1/sql

Apps Gmail Vertica IoT Programming Of Interest WORK Personal Resources Educational Analytics Health Other bookmarks

Vertica Management Console uidbadmin Log out

Databases and Clusters verticadb Query Execution

Query History

Clear all

Filter previous queries

Select \* from nodes;

1 Select \* from nodes;

2

Execute Query

Select \* from n

Query Results

Query Plan

Query Profile

Export Data

Auto-Resize all columns

Search query results

node_name	node_id	node_state	node_address	node_address_family	export_address	export_address_family
v_verticadb_node0001	45035996273704978	UP	10.11.12.10	ipv4	10.11.12.10	ipv4
v_verticadb_node0002	45035996273835368	UP	10.11.12.20	ipv4	10.11.12.20	ipv4
v_verticadb_node0003	45035996273835372	UP	10.11.12.30	ipv4	10.11.12.30	ipv4

3 rows | Execution time: 0.021s

Overview

Activity

Manage

Design

Load

Query Execution

Query Plan

License

Settings

screenshot-3-2....png

screenshot-3-2....png

screenshot-3-2....png

screenshot-co....png

Show all

Try entering SQL





# External Tables

# Create External Table of Parquet Data on S3

CloudFormation Manage x

Vertica Management Cor x

Not secure | https://34.231.219.222:5450/webui/admin/databases/1/sql

Apps Gmail Vertica Welcome to The MIT IoT Programming Of Interest WORK Personal Resources

Other bookmarks

Vertica Management Console uidbadmin Log out 6 0 ?

Databases and Clusters > VerticaDB > Query Execution

Query History Clear all

Filter previous queries

create EXTERNAL table ex\_table ( cal\_dt varchar NOT NULL, cmc\_chn\_str\_nbr int NOT NULL,

SELECT set\_config\_parameter('AWSRegion',... east-1'); SELECT set\_config\_parameter('AWSAuth','AKI

select node\_name from nodes;

select \* from nodes;

create EXTERNAL

Overview Activity Manage Design

```
1 create EXTERNAL table ex_table (
2   cal_dt varchar NOT NULL,
3   lane_nbr int NOT NULL,
4   pid_key int NOT NULL,
5   cmc_chn_nbr int NOT NULL,
6   ord_amt float NOT NULL,
7   ndc_ord_amt float NOT NULL,
8   tot_upc_qty int NOT NULL,
9   lane_mode_cd int NOT NULL,
10  loy_id_cnt int NOT NULL,
11  reject_nbr int NOT NULL,
12 )
13 AS COPY FROM 's3://aws-
```

```
1 create EXTERNAL table ex_table (
2   cal_dt varchar NOT NULL,
3   lane_nbr int NOT NULL,
4   pid_key int NOT NULL,
5   cmc_chn_nbr int NOT NULL,
6   ord_amt float NOT NULL,
7   ndc_ord_amt float NOT NULL,
8   tot_upc_qty int NOT NULL,
9   lane_mode_cd int NOT NULL,
10  loy_id_cnt int NOT NULL,
11  reject_nbr int NOT NULL,
12 )
13 AS COPY FROM 's3://aws-cluster-tests/parquet_data/hive013_uncompressed/orderupc/*' PARQUET;
```

Parquet File Format



# Create External Table of Parquet Data on S3

CloudFormation Manage x

Vertica Management Cor x

Not secure | https://34.231.219.222:5450/webui/admin/databases/1/sql

Apps Gmail Vertica Welcome to The MIT IoT Programming Of Interest WORK Personal Resources

Other bookmarks

Vertica Management Console uidbadmin Log out 6 0 ?

Databases and Clusters > VerticaDB > Query Execution

Query History Clear all

Filter previous queries

create EXTERNAL table ex\_table ( cal\_dt varchar NOT NULL, cmc\_chn\_str\_nbr int NOT NULL,

SELECT set\_config\_parameter('AWSRegion',... east-1'); SELECT set\_config\_parameter('AWSAuth','AKI

select node\_name from nodes;

select \* from nodes;

create EXTERNAL

Overview Activity Manage Design

```
1 create EXTERNAL table ex_table (
2   cal_dt varchar NOT NULL,
3   lane_nbr int NOT NULL,
4   pid_key int NOT NULL,
5   cmc_chn_nbr int NOT NULL,
6   ord_amt float NOT NULL,
7   ndc_ord_amt float NOT NULL,
8   tot_upc_qty int NOT NULL,
9   lane_mode_cd int NOT NULL,
10  loy_id_cnt int NOT NULL,
11  reject_nbr int NOT NULL,
12 )
13 AS COPY FROM 's3://aws-
```

```
1 create EXTERNAL table ex_table (
2   cal_dt varchar NOT NULL,
3   lane_nbr int NOT NULL,
4   pid_key int NOT NULL,
5   cmc_chn_nbr int NOT NULL,
6   ord_amt float NOT NULL,
7   ndc_ord_amt float NOT NULL,
8   tot_upc_qty int NOT NULL,
9   lane_mode_cd int NOT NULL,
10  loy_id_cnt int NOT NULL,
11  reject_nbr int NOT NULL,
12 )
13 AS COPY FROM 's3://aws-cluster-tests/parquet_data/hive013_uncompressed/orderupc/*' PARQUET ;
```

Note the use of File Globbing!

## 3.5 GBytes of Parquet Data

```
create EXTERNAL table ex_table (  
  cal_dt varchar NOT NULL,  
  lane_nbr int NOT NULL,  
  pid_key int NOT NULL,  
  cmc_chn_nbr int NOT NULL,  
  ord_amt float NOT NULL,  
  ndc_ord_amt float NOT NULL,  
  tot_upc_qty int NOT NULL,  
  lane_mode_cd int NOT NULL,  
  loy_id_cnt int NOT NULL,  
  reject_nbr int NOT NULL,  
  AS COPY FROM 's3://dsprogis/data/orderupc/*' PARQUET ;  
  cmc_chn_str_nbr int NOT NULL,  
  time_key int NOT NULL,  
  tran_nbr int NOT NULL,  
  str_tndr_typ_nbr int NOT NULL,  
  ss_ord_amt float NOT NULL,  
  unique_upc_qty int NOT NULL,  
  tot_ndc_qty int NOT NULL,  
  oper_id_nbr int NOT NULL,  
  non_loy_id_cnt int NOT NULL,  
  batch_id_nbr int NOT NULL )
```

## Fetch 10,000 rows

```
select * from ex_table;
```

How long did it take?

“ You mean I can query across more than one external file?! ”



“ You mean I can query across more than one external file?! ”



Actually, file globbing means more ...



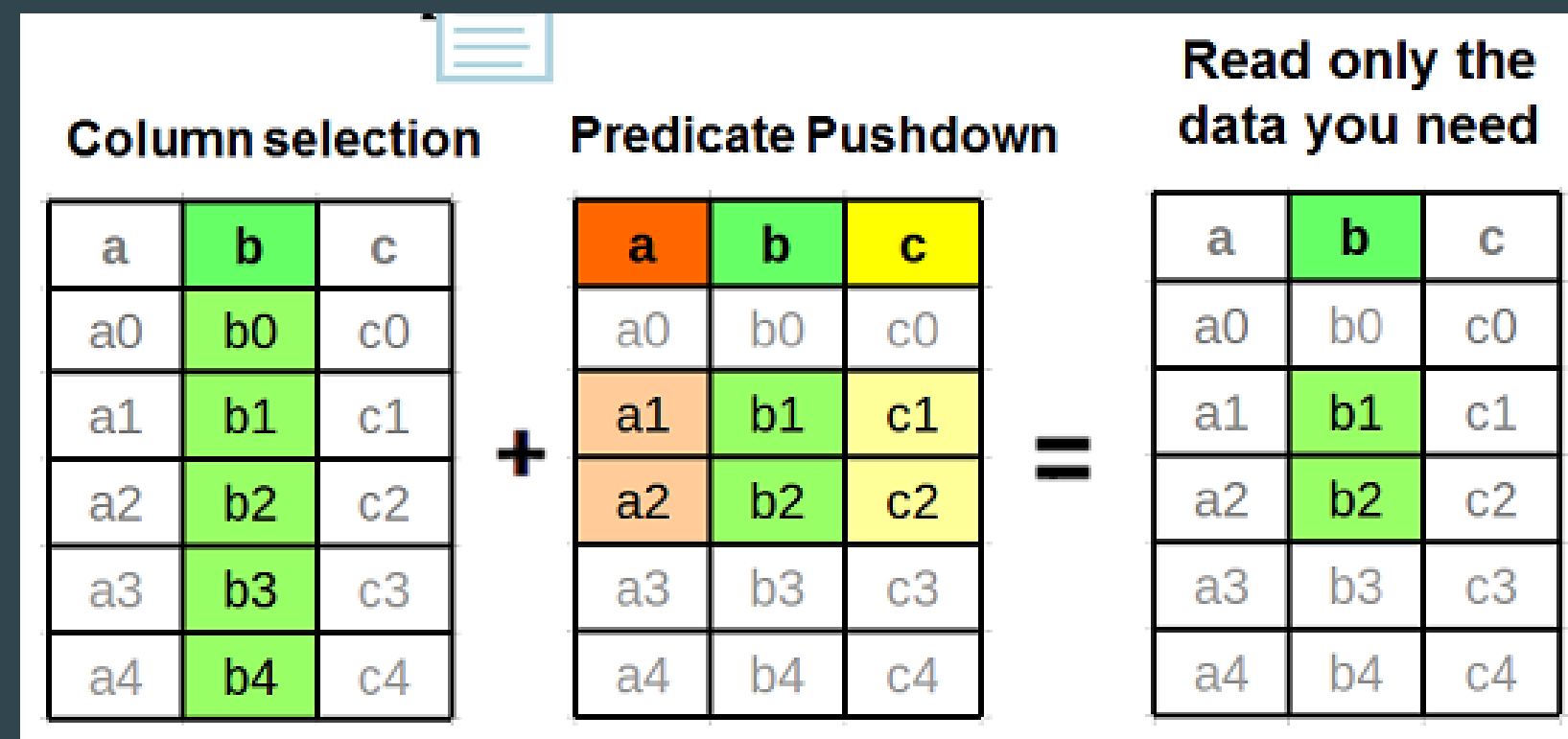
It means your external table is a “live table”:

```
12 /  
13 AS COPY FROM 's3://aws-cluster-tests/parquet_data/hive013_uncompressed/orderupc/*' PARQUET ;
```




# Vertica improves query performance of Parquet & Orc, through

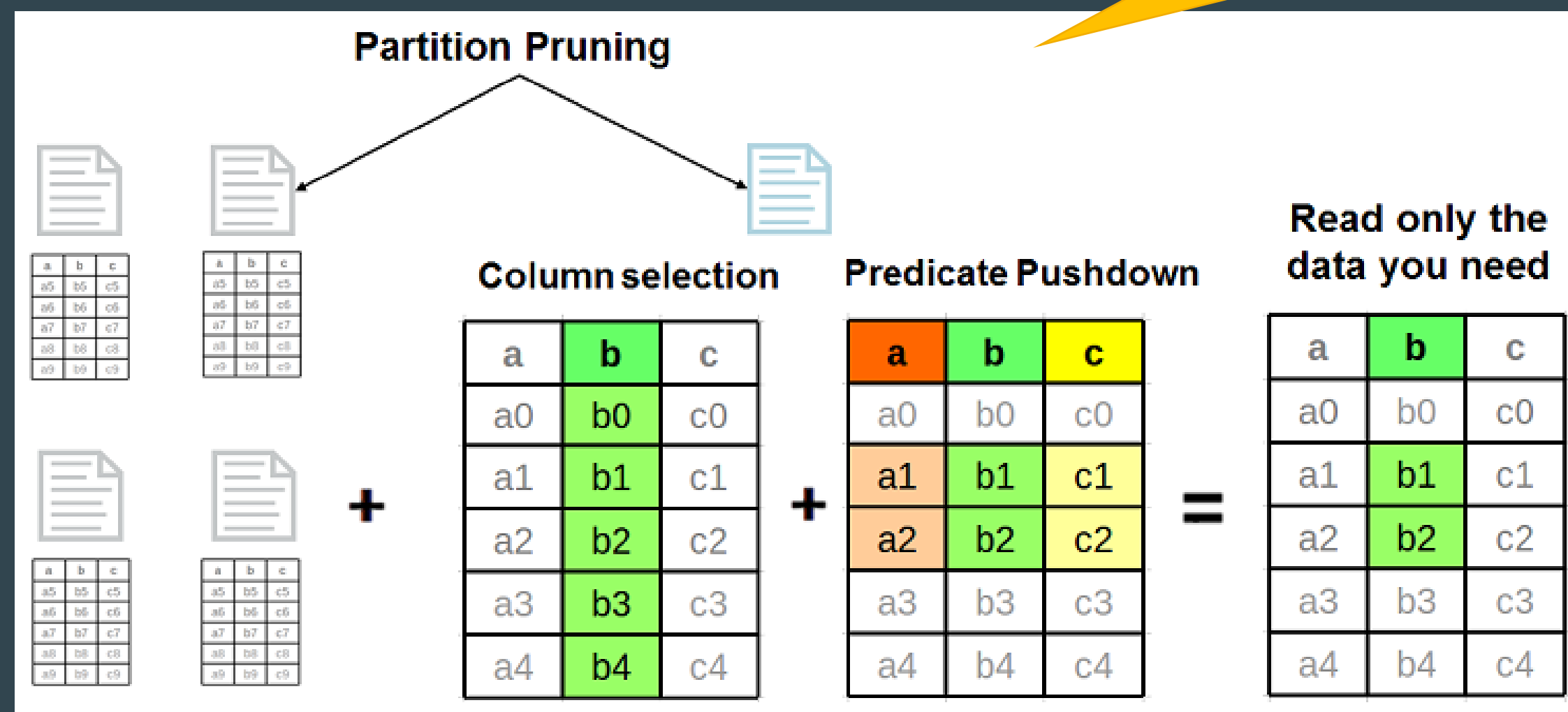
- column selection,
- predicate pushdown



# Parquet & Orc plus File Globbing

- column selection,
- predicate pushdown, and
- partition pruning.

File globbing adds the potential for Partition Pruning

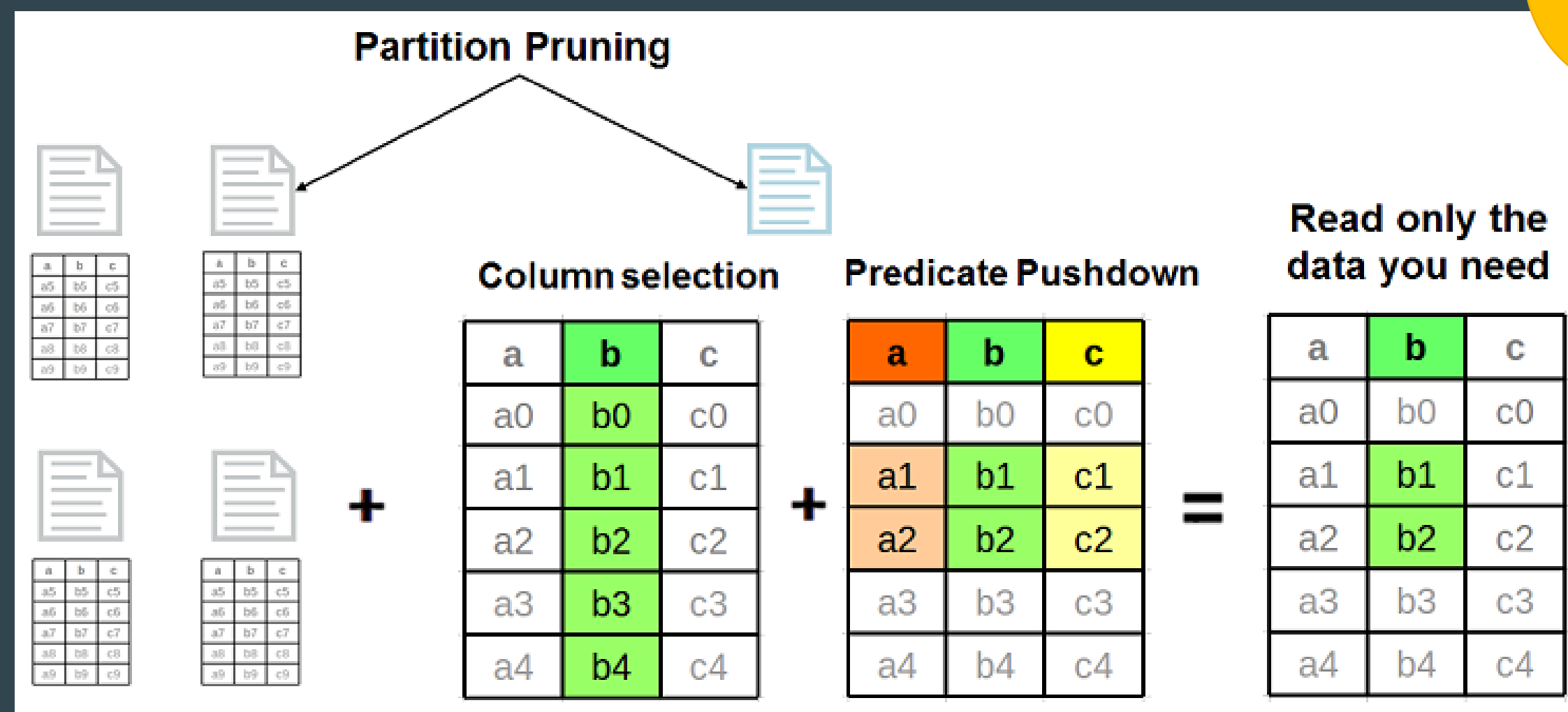


# Parquet & Orc plus File Globbing

- column selection,
- predicate pushdown, and
- partition pruning.

If your file naming convention includes:  
“.../part\_col=value/...”

Then Vertica will process  
“... where part\_col = value ...”  
by reading only the files that match.



Parquet & Orc on S3

File Globbing & Partition Pruning

Column Selection & Predicate Push-down

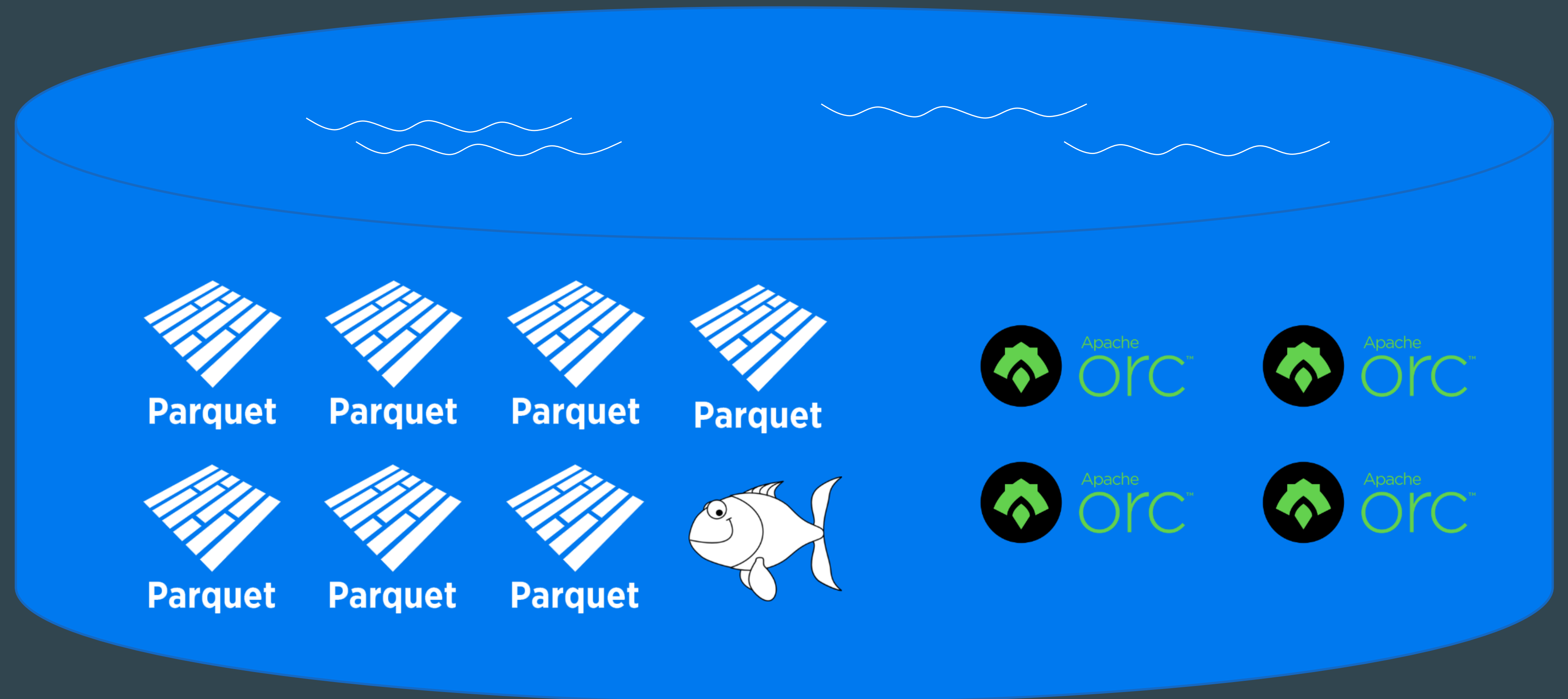
UDFS File Read Improvements

So what?



# S3 Data Lake Formats?

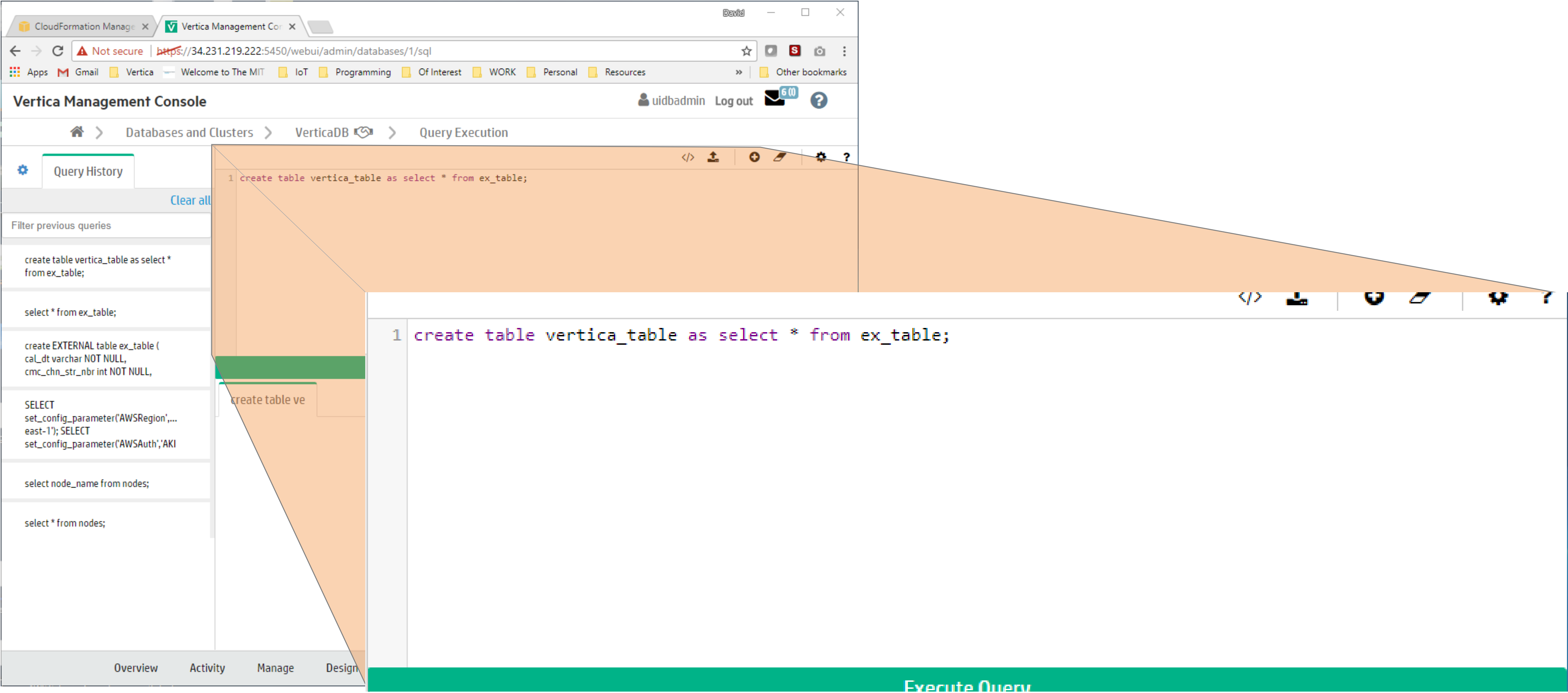
Performant because they are columnar.  
Efficient because they are compressed.



# Compare External to Internal Tables



# For Best Performance, Load the Data into Vertica.



## Copy external table to internal table

```
create table in_table as select * from ex_table;
```

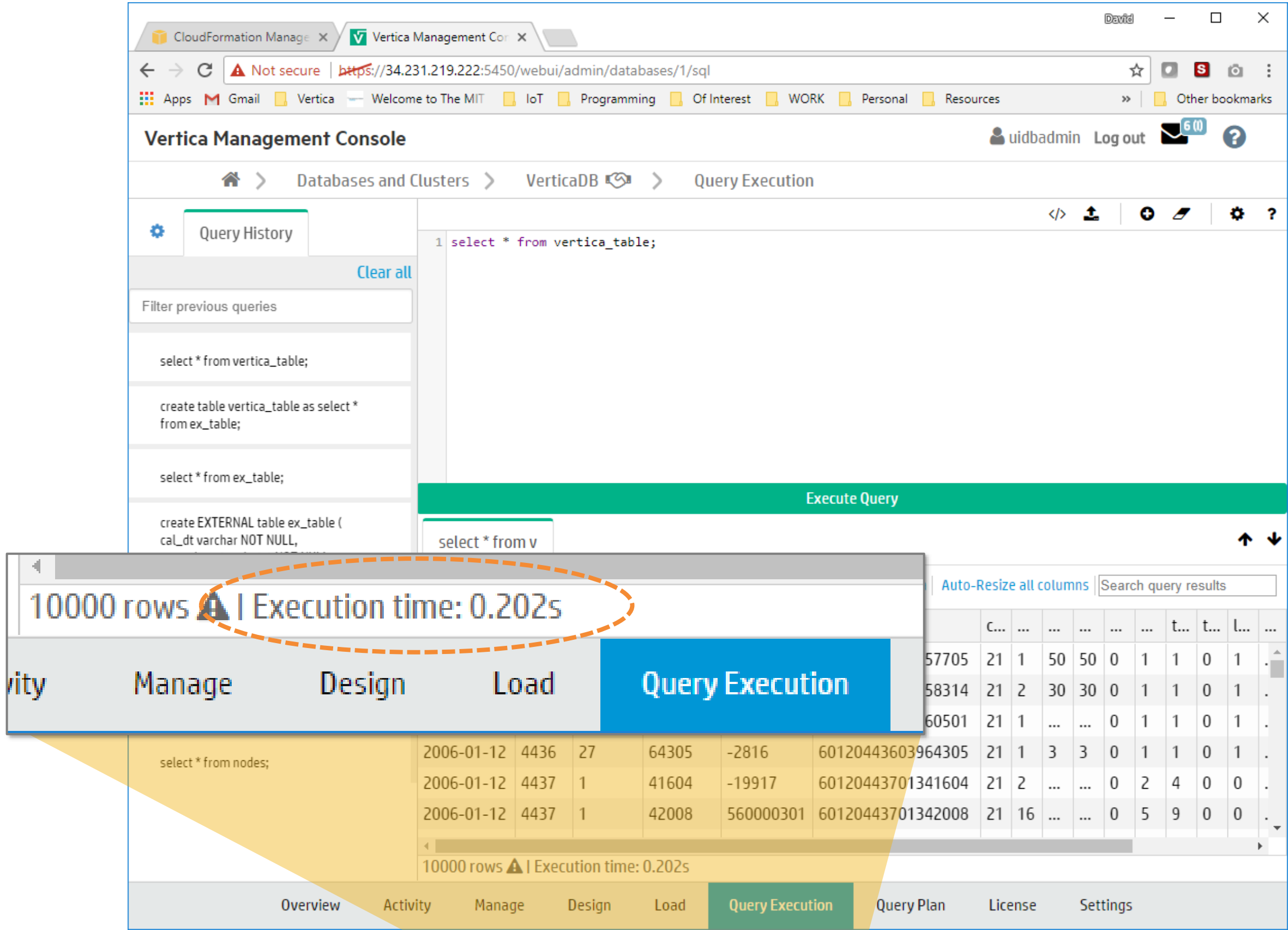
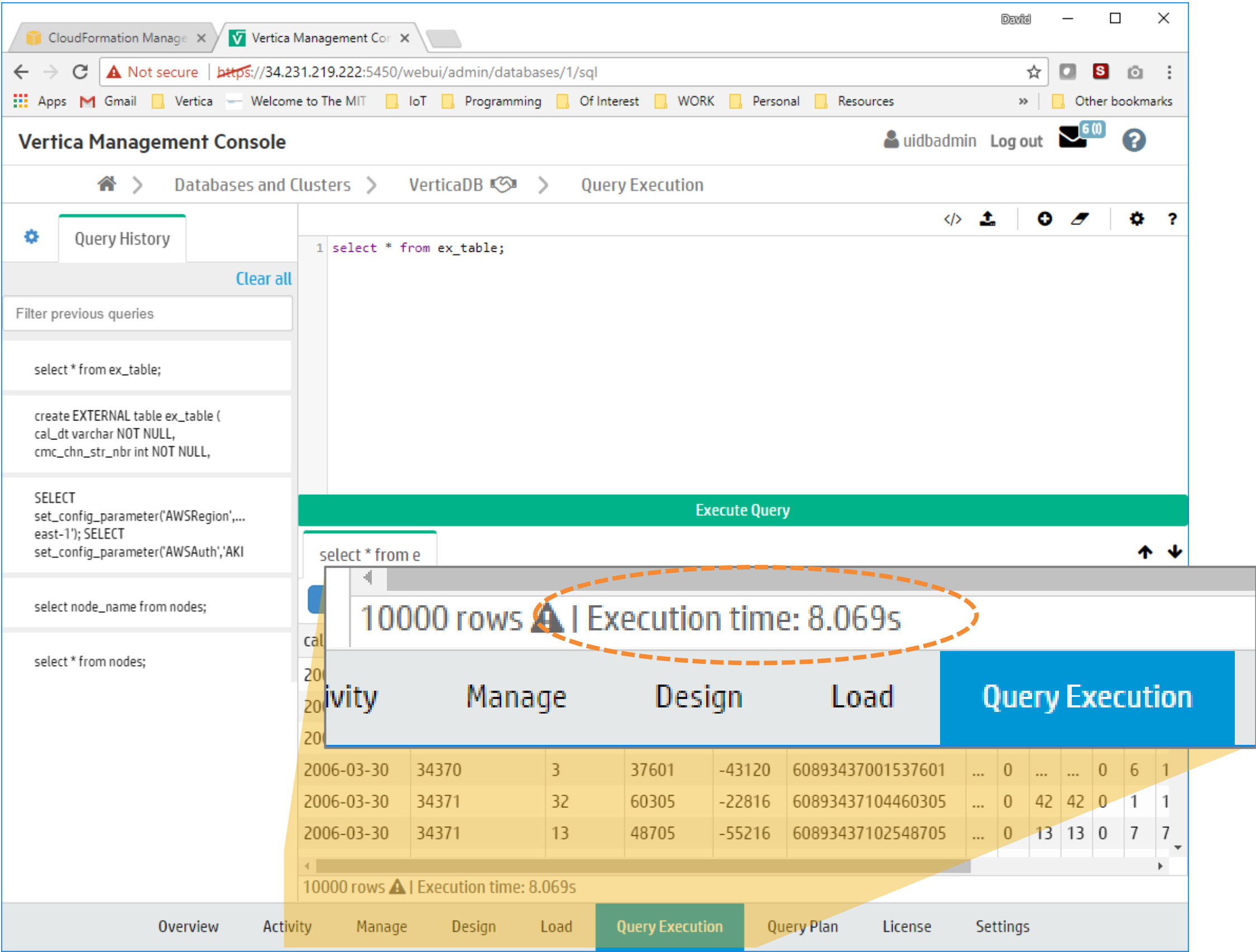


Fetch 10,000 rows

```
select * from in_table;
```

How long did it take?

# External Table Query of Parquet Data on S3 vs Internal



40X faster  
(Improvements will vary based on many factors.)





**More Learning**

## Other things to test

- Scaling Cluster (in Video, next slide)
- Hibernating (in Video, next slide)
- Creating Subcluster (example ahead)



# Workshop/Tutorial

<https://www.brighttalk.com/webcast/16439/307439>

The screenshot shows a web browser window displaying a BrightTalk webcast. The address bar shows the URL: [https://www.brighttalk.com/webcast/16439/307439?utm\\_campaign=webcasts-search-results-feed&utm\\_content=Eon&ut...](https://www.brighttalk.com/webcast/16439/307439?utm_campaign=webcasts-search-results-feed&utm_content=Eon&ut...). The BrightTalk logo is in the top left, with a search bar and navigation links for Summits, Communities, and My account. The main content area features a video player for the "Vertica Demo Series" titled "Get Under the Hood of Vertica: Eon Mode<sup>1</sup>, Separation of Compute and Storage". The video is presented by David Sprogis (Cloud Product Manager) and Tim Donar (Solution Engineer). The video player shows a progress bar at 00:06 of 58:33. Below the video are tabs for Attachments, Rate this, and Details. To the right of the video, there are buttons for "Share with your team" and "Watch later", followed by a "Recommended for you:" section. This section lists three recommended videos: "The Top 5 Security Lessons Learned from Transitioning to the Cloud" by Carla Arend, Arun Singh, and Nico Fischbach; "The future of smart connected (or IoT) devices, but what about security?" by Paul Lockley; and "Maxpert series: Schedule Optimization" by Daniel Bennett. At the bottom of the page, there are social media sharing icons for LinkedIn, Twitter, Facebook, and Google+, and a partial view of another video titled "GDPR: Balancing Customer Experience".

\* Slightly out-of-date instructions, contact David Sprogis for latest

## Example Subcluster

- Sample Subcluster Setup
- ALTER DATABASE VerticaDB DROP ALL FAULT GROUP;
- create fault group "mysubcluster01";
- ALTER FAULT GROUP "mysubcluster01" ADD NODE v\_verticadb\_node0001;
- ALTER FAULT GROUP "mysubcluster01" ADD NODE v\_verticadb\_node0003;
- ALTER FAULT GROUP "mysubcluster01" ADD NODE v\_verticadb\_node0005;
- create fault group "mysubcluster02";
- ALTER FAULT GROUP "mysubcluster02" ADD NODE v\_verticadb\_node0002;
- create fault group "mysubcluster03";
- ALTER FAULT GROUP "mysubcluster03" ADD NODE v\_verticadb\_node0004;
- 
- select parent\_name,member\_name from fault\_groups where parent\_name like 'subcluster%' order by parent\_name,member\_name;



Last Slide

Thank you