Best Practices For Running Databases on Kubernetes

March 15, 2024
Peter Zaitsev,
Founder, Percona
Big Picture
Proprietary Clouds bring Great Usability at Great Cost
Have We Been Here Before?
Open Source Catches Up Again

- Lock-in with Cloud Vendor
- Use Proprietary Solutions
- Highly Differentiated Cloud
- Freedom to Run Anywhere
- Use Open Source
- Cloud Is Commodity
- Customer
- Choice of Vendors
What is Cloud Computing?
An analogy: think of electricity services...

You simply plug into a vast electrical grid managed by experts to get a low cost, reliable power supply – available to you with much greater efficiency than you could generate on your own.

Power is a utility service - available to you on-demand and you pay only for what you use.
Open Source and Proprietary

Rise of Open Source

Kubernetes Operators are available for most popular Open Source Databases

Kubernetes is getting better for stateful applications

Kubernetes is universally available
Data on Kubernetes

https://dok.community/
Data on Kubernetes

DoK workload %s are already high, and expected to increase
Leaders are chomping at the DoK bit

% of data workloads on k8s

<table>
<thead>
<tr>
<th></th>
<th>Less than 10%</th>
<th>10%-24%</th>
<th>25%-49%</th>
<th>50%-74%</th>
<th>75%-100%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>11</td>
<td>37</td>
<td>43</td>
<td>8</td>
</tr>
</tbody>
</table>

Expected increase in data workloads on k8s

<table>
<thead>
<tr>
<th></th>
<th>Less than 10%</th>
<th>10%</th>
<th>20%</th>
<th>30%</th>
<th>40%</th>
<th>50% or more</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
<td>11</td>
<td>24</td>
<td>36</td>
<td>14</td>
<td>12</td>
</tr>
</tbody>
</table>

Satisfaction = Reality - Expectations
DoK is winning the expectations battle

<table>
<thead>
<tr>
<th>Workload</th>
<th>Very satisfied</th>
<th>Somewhat satisfied</th>
</tr>
</thead>
<tbody>
<tr>
<td>Databases</td>
<td>64</td>
<td>96</td>
</tr>
<tr>
<td>Analytics</td>
<td>61</td>
<td>95</td>
</tr>
<tr>
<td>Persistent storage</td>
<td>60</td>
<td>94</td>
</tr>
<tr>
<td>AI/ML</td>
<td>58</td>
<td>97</td>
</tr>
<tr>
<td>CI/CD</td>
<td>56</td>
<td>95</td>
</tr>
<tr>
<td>Streaming / messaging</td>
<td>55</td>
<td>96</td>
</tr>
</tbody>
</table>

© 2024 Percona

0. In general, how satisfied are you with using Kubernetes to run each of the following data workloads in your organization? Use a scale from 1 to 5 where 5 means “very satisfied” and 1 means “not at all..."
Quite a growth in Adoption!

KUBERNETES GROWTH AREAS

Focusing on non-application workloads, enterprises used an increasing variety of technologies. This reflects the need to enhance Kubernetes with better observability, security, and service-to-service communications. Other technologies enable specific use cases like CI/CD tools or databases.

Across all categories, open source projects rank among the most frequently used solutions.

Which data workloads on k8s

- Databases: 76
- Analytics: 67
- AI/ML: 50
- Persistent storage: 45
- Streaming / messaging: 39
- CI/CD: 31

76% Leaders
48% Leaders

Databases on Kubernetes
Power of Kubernetes Operators
Power of Kubernetes Operators

- **Day One Automation**
  - Installation and Initial Configuration

- **Day Two Automation**
  - Backups, Scaling, Self Healing, Upgrades
Percona Kubernetes Operators

Get look and feel for Basics?

Kubernetes with Operators is Easy and Powerful
https://per.co.na/PXCMinkube
Many Modern DBaaS Are Built on Kubernetes Operators
#1 Use Operators

For Production Deployment you need to ensure great “Day 2” Automation
Relying on Single Instance in Kubernetes environment is even more dangerous
#3 Keep Persistent Data Persistent

Persistent Volumes; Local Disk or Fast Remote Storage
Local NVMe storage is cheap!

**Local storage**
- i3.8xlarge
- 32 vCPU
- 244GB RAM
- 4x1900GB NVMe SSD included
- 1,650k Read + 720k Write IOPS

$1,822/month

**EBS storage**
- r5.8xlarge
- 32 vCPU
- 256GB RAM
- EBS 4x600GB
- 64k IOPS each

$15,915/month
50TB of data connected to single POD is not a good idea
Kubernetes or Not

Databases often need “Big Iron” more than Apps.
Or you may have non-uniform Performance and Severe Impact on other Pods
3 Node Cluster Running on Single Physical Node is not a Great High Availability Solution
#8 Tune your Database

Database Configuration, Indexes, Queries needs to be taken care of as usually
Query Analytics in Percona Monitoring and Management
Some Databases can be “Scaled Out” others only “Scaled-Up” and Scaled for Reads
Rescheduling Database Pod Can be Expensive, so better ensure it does not happen too often.
#11 Do not Expose your Database unless you have to

Unintended Publicly Accessible Data is leading cause of Security Leaks

© 2024 Percona
#12 Enable Encryption

Data at Rest and Data in Transit. Does not cause huge Overhead those days.
Great way to pass database access credentials to your application
Clustering Does not Eliminate need for backups. Do them. Good Operators make it Easy
Databases designed to be run on Cloud Native Infrastructure are Coming - Vitess, Neon, Yugabyte, TiDB
Arm can be increasingly performant and cost effective in the cloud
#17 Pick Right Kubernetes deployment mode

Managed and Self-Managed both have their merits
Spreading pods over more nodes than needed can be expensive.
#19 Observe…

`coroot:~#`

https://coroot.com
What are yours?
Beyond Kubernetes Operators?

- Day 1 and Day 2 Automation, Toil Reduction Similar to DBaaS
- UX is Different, Requires Kubernetes Expertise
Percona Everest – Open Source DBaaS bases on Kubernetes

https://docs.percona.com/everest/index.html
Future
Beyond Single Kubernetes Cluster

Open, Multi-Cloud, Multi-Cluster Kubernetes Orchestration
Automatic Volume Size Management

Can we automatically resize data volumes?
Auto Scaling
Auto Suspend
(For Development)

Suspend Idle Database Nodes, wake up when request comes in
Ease of Migration

Migrate databases to Kubernetes Easily
Thank you, Let’s Connect!

https://www.linkedin.com/in/peterzaitsev/
https://twitter.com/PeterZaitsev
http://www.peterzaitsev.com