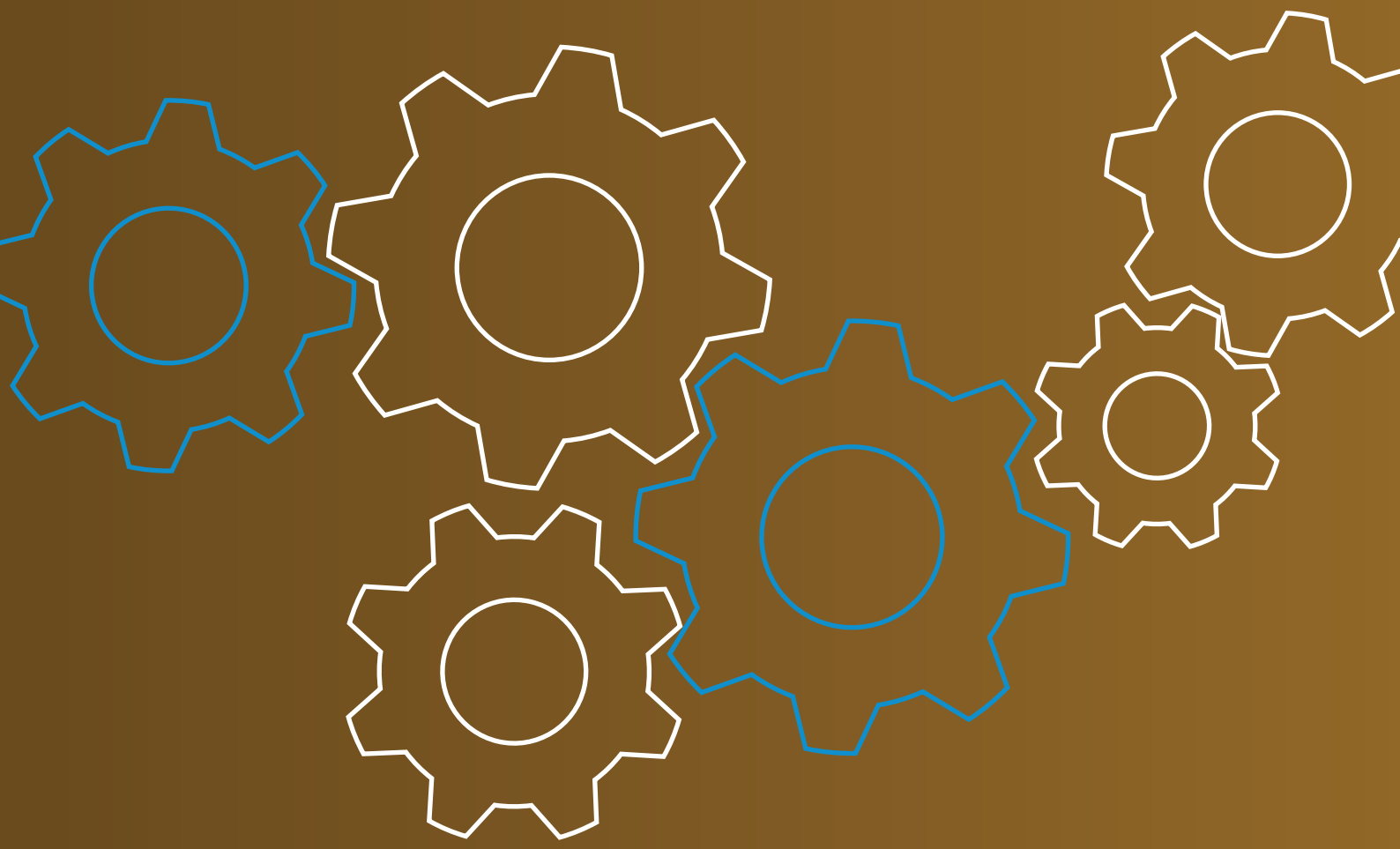
severalnines

Management and Automation of Open Source Databases



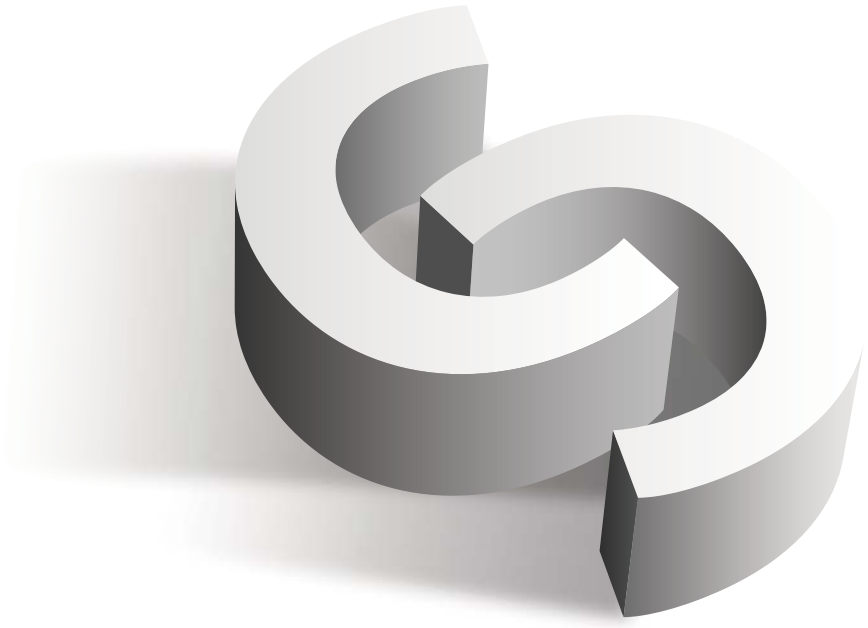


Table of Contents

Introduction

Databases are the repositories of a company's most critical information. Being stateful, they typically are the most complex part of the application stack. Databases are complex to install and manage, especially when clustering or high availability is involved. To operate a database usually requires specialised knowledge.

While large enterprises have made significant investments in database management tools, most of these have been targeted to be used with proprietary DBMSs like Oracle or SQL Server. However the database landscape in the modern data center is changing.

> By 2018, more than 70% of new in-house applications will be developed on an OSDBMS, and 50% of existing commercial RDBMS instances will have been converted or will be in process," predicts Gartner in a new research paper, State of Relational Open Source RDBMSs 2015.

Open Source DBs - the New Data Center Standard?

The use of open source databases had historically been associated with small, non mission-critical use cases. However, recent evidence suggests that enterprises are now turning to open source databases to reduce database costs and avoid supplier lock-in.

With organisations capturing and processing more data, and IT budgets staying flat at best, it is of no surprise that open source databases are proliferating in the data center.

According to DB-Engines, an online knowledge base that curates information about database management systems, three of the top five most popular databases today are open source.

This is driving large-scale adoption across all sizes of business. The question remains though - proprietary databases have been around for decades, and customers have access to a rich third-party ecosystem of management tools. How about open source databases? What type of tools are available to help customers operate these databases?

Rank			DBMS
Aug 2016	Jul 2016	Aug 2015	
1.	1.	1.	Oracle
2.	2.	2.	MySQL
3.	3.	3.	Microsoft SQL Server
4.	4.	4.	MongoDB
5.	5.	5.	PostgreSQL

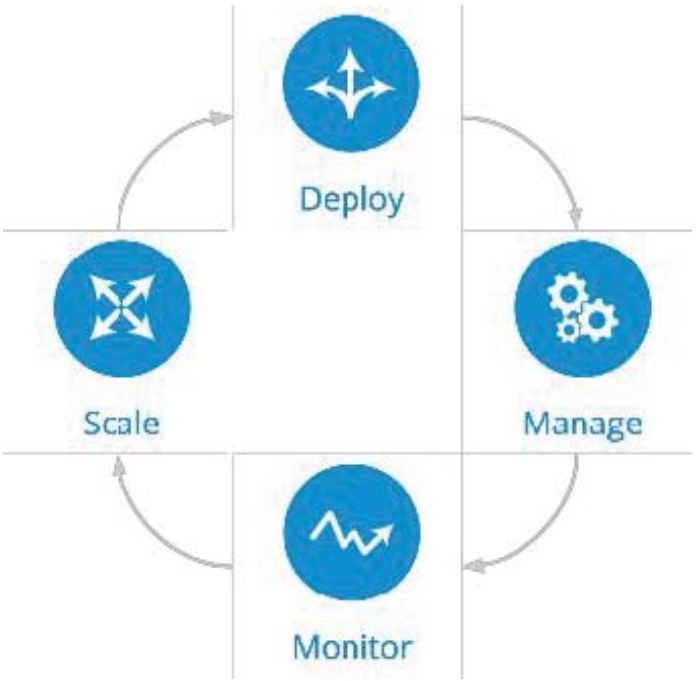
The Database Infrastructure Lifecycle

As the adoption of open source databases increases in the enterprise, especially for mission-critical applications, so does the need for robust and integrated tools. Operational staff need to able to manage everything from provisioning, capacity, performance and availability of the database environment. This is needed to minimize the risk for service outages or poor application performance.

Provisioning in larger enterprises typically happens across multiple environments, from development, testing/QA, staging to production and disaster recovery. Each system needs to be configured to suit its deployment environment and related applications.

Once in place, systems need to be monitored. Monitoring is a fundamental part of a mission-critical operation, it gives visibility into what is going on. This facilitates the user workflows that drive problem resolution and prevention, as well as performance optimization.

Next, operations would streamline tasks such as configuration, patching, upgrades, backup and restore, database cloning, security and compliance with company policies.



With infrastructure becoming increasingly software-defined, resources can be dynamically allocated to shifting application demands. Database capacity need to be adjusted accordingly, either by scaling out for more capacity or scaling back.

Tools - Build or Buy?

For each of the above stages, there are multiple tools available. Configuration management tools like Puppet, Chef, Ansible, Salt, etc. can automate deployments. Monitoring tools, including Nagios, Zabbix, Cacti and Munin, cover everything from simple healthchecks to realtime charting of metrics. Clustering software like Corosync and Pacemaker can help manage availability and failover of processes. There is an overabundance of tools that provide point solutions, which however leads to another set of problems. Data transfer between different tools, as well as different methodologies, processes, and user interfaces lead to new challenges in operating an overall database environment.

Some enterprises have done the heavy lifting, by integrating a mixture of publicly-available open source tools and in-house built components.

In reality, this often turns out to be a patchwork of loosely connected components that were introduced over time - unless the enterprise is ready to dedicate appropriate R&D resources.

A unified operations management platform that encompasses all the complexity of the deploy-monitor-manage-scale lifecycle would be the obvious solution to deliver simplicity to the enterprise.

Database Deployments - Beyond Chef and Puppet

It is often the case that a database must be deployed in multiple environments on its journey to production. Often, these processes are manual and error prone. For instance, a four-node MySQL Cluster needs at least 40 commands to get deployed, the commands being different depending upon the production environment. The shift towards service oriented architectures, virtualization and cloud environments is also contributing to a more diverse, and dynamically changing database landscape.

Unless there is a way to consistently replicate best practice configurations, one may end up with inconsistent deployments or faulty database configurations that lead to poor performance or availability.

Configuration management systems like Chef or Puppet are rapidly gaining popularity, as enterprises look at automating their infrastructure. These work relatively well with single-instance databases, the problem arises with multi-server setups that are replicated or clustered. The current tools have poor support for incremental and cross node coordination. Distributed databases can be complex, as there are different node types and roles involved. Deployment procedures usually call for orchestration across nodes, with tasks carried out in a specific order.

Monitoring - Why Not Just Use Nagios?

There are numerous enterprise monitoring solutions available to enterprises, as well as a myriad of open source solutions – each addressing specific pain points. For instance, in the open-source space: Graphite and Cacti provide trending, Nagios provides alerting and Statsd and Collectd gather raw metrics. Integrating these systems can be a daunting task.

Databases, though, need monitoring for specific concepts and components, and are not well served by the one-size-fits-all approach of generic monitoring tools. The complexity of the database requires far more sophisticated analysis of performance than is provided by the raw metrics of the database engine.

> According to a recent application performance management (APM) survey by Gleanster Research, 88 percent of respondents cited the database as the most common challenge or issue with application performance. Furthermore, 71 percent said their current APM tools only provide hints, but rarely identify the root cause of problems.

Dealing with performance problems is usually one of the biggest post-deployment nightmares faced by enterprises. For instance, DBAs spend an inordinate amount of time combing through cumbersome SQL logs to find problems.

A complete monitoring system would provide all the required pieces to create a holistic view of the database environment, from database to the underlying compute, network and storage resources. This includes real time data to know what is happening now, high resolution of data for better accuracy, pattern recognition and baseline alerting to report on emerging database problems as well deviations from normal behaviour.

Management - a Rising Tide of Continuous Scripting

Databases are being deployed everywhere, and they need to be managed. Management operations include rolling out configuration changes, patching, version upgrade or downgrade, backups, fixing broken nodes and maintaining availability, adding/removing nodes with subsequent reconfiguration of a running setup, cloning and so on.

However, these procedures tend to be very manual.

> According to a survey on database manageability by the Independent Oracle User Group in 2013, the majority (of DBAs) still overwhelmingly perform a range of tasks manually, from patching databases to performing upgrades. Depending on the task, between 38 per cent and 59 per cent of users were still carrying them out manually.

However, there are some tools available that automate individual aspects of database management.

Database Task	Available Tools
Configuration Management	Puppet, Chef, Ansible
Patching/Upgrades	BMC BladeLogic Database Automation
Availability/Failover	Corosync/Pacemaker, MHA
Adding/Removing nodes	Orchestration tools (Puppet, Chef) with custom scripts/playbooks/recipes
Database Cloning	Orchestration tools with custom scripts/recipes
Backup & Restore	MySQL Enteprise Monitor, Zmanda, Percona XtraBackup, storage-centric approaches (LVM/ EBS snapshops)

The problem with the above is that the enterprise ends up with an array of opportunistic tools that has been cobbled together.

Orchestration tools like Chef and Puppet could be applied to solve multiple problems, but these are generic tools. It would require extensive scripting in order to add the specific intelligence about the database processes.

If DBAs are still patching and provisioning manually, how are enterprises possibly going to handle the proliferation of open source databases in their datacenters?

Severalnines ClusterControl - a Systematic Approach to Operations

Enterprises are evolving their infrastructure to span across private datacenters and public clouds. As open source databases achieve broad adoption, these will become an important asset that need to be monitored around the clock - as any database downtime will have direct impact on service availability. IT Operations need to be able to evolve from manual practices to effective and automated management procedures.

DBAs have better things to do than piece together builds, follow checklists full of release commands, search the internet for sample configurations, copy files around on servers, and monitor running database processes. A management platform that systematically addresses all the different aspects of the database lifecycle will be more robust than patching together a number of point solutions.

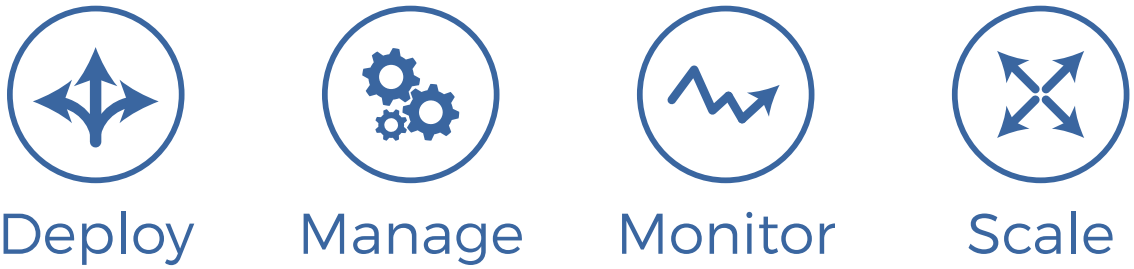
Severalnines provides a unified platform that automates the lifecycle of the database:

- Deployment of single servers or entire clusters on-premise or cloud environments
- Monitoring via a unified and comprehensive view of all database instances across all data centers
- Management via automated processes that are specifically written for the underlying databases
- Scaling out based on dynamic capacity needs

About Severalnines

Severalnines provides automation and management software for database clusters. We help companies deploy their databases in any environment, and manage all operational aspects to achieve high-scale availability.

Severalnines' products are used by developers and administrators of all skills levels to provide the full 'deploy, manage, monitor, scale' database cycle, thus freeing them from the complexity and learning curves that are typically associated with highly available database clusters. The company has enabled over 8,000 deployments to date via its popular ClusterControl solution. Currently counting BT, Orange, Cisco, CNRS, Technicolour, AVG, Ping Identity and Paytrail as customers. Severalnines is a private company headquartered in Stockholm, Sweden with offices in Singapore and Tokyo, Japan. To see who is using Severalnines today visit, http://severalnines.com/customers.



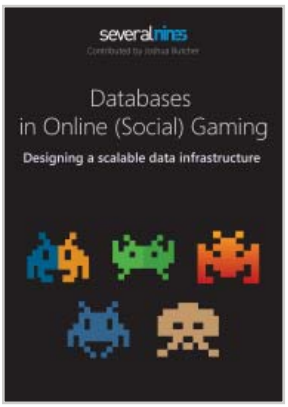
Related Resources from Severalnines

Whitepapers

MySQL Replication for High Availability

This tutorial covers information about MySQL Replication, with information about the latest features introduced in 5.6 and 5.7. There is also a more hands-on, practical section on how to quickly deploy and manage a replication setup using ClusterControl.

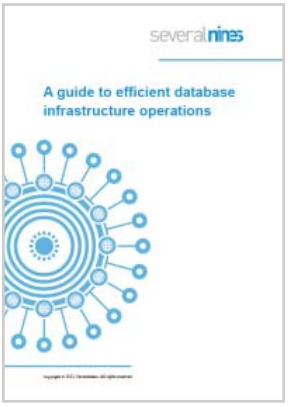
Download here



Databases in Online (Social) Gaming

This paper discusses the importance of databases for the gaming industry, what its requirements are in terms of database technology as well as a discussion on why MySQL is or should be the database of choice for anyone wanting to develop online social games that are reliable and stable in all their aspects.

Download here



A Guide to Efficient Database Infrastructure Operations

Taking control of their data is every company's number one job.

Database operations encompass a number of functions, including the initial deployment of a solution, configuration management, performance monitoring, SLA management, backups, patches, version upgrades and scaling.

Download here

severalnines

Deploy

Manage

Monitor

Scale