O'REILLY®

致敬 亚马逊云科技

云数据库简介

管理员指南

Wendy A. Neu、Vlad Vlasceanu、Andy Oram 和 Sam Alapati 著

报告

摆脱传统数据库的束缚

亚马逊云科技提供了最广泛的专用数据库选择，
您可以**节省成本**、**不断壮大**以及
实现**更快创新**

企业规模成本为商用数据库的 1/10

性能是常用替代方案的 3 至 5 倍

14 个以上的数据库引擎 - 比其他任何提供商都要多

亚马逊云科技 databases

了解更多信息：aws.amazon.com/databases

云数据库简介

管理员指南

Wendy A. Neu、Vlad Vlasceanu、
Andy Oram 和 Sam Alapati 著

北京 • 波士顿 • 法纳姆 • 塞巴斯托波 • 东京 O'REILLY®

云数据库简介

Wendy A. Neu、Vlad Vlasceanu、Andy Oram 和 Sam Alapati 著

版权所有 © 2019 O’Reilly Media Inc.保留所有权利。

在美国印刷。

出版社：O’ Reilly Media, Inc.

出版社地址：1005 Gravenstein Highway North, Sebastopol, CA 95472

可以购买 O’Reilly 图书供教育、商业或促销使用。大多数图书都有网络版(http://oreilly.com)。有关更多信息，请联系本公司/机构销售部门：800-998-9938 或 corporate@oreilly.com。

开发编辑：Jeff Bleiel
策划编辑：Jonathan Hassell
制作编辑：Katherine Tozer
文字编辑：Octal Publishing, LLC
内页设计：David Futato
封面设计：Karen Montgomery
插图：Rebecca Demarest

2019 年 9 月：　　第一版

第一版修订记录

2019 年 8 月 19 日：　　首次发行

O’Reilly 徽标是 O’Reilly Media, Inc. 的注册商标。《云数据库简介》、封面图片以及相关商业外观是 O’Reilly Media, Inc. 的商标。

本书中所表达的观点仅代表作者本人，并不代表出版商的观点。出版商及作者已尽最大努力确保本书中的信息和说明正确无误，但出版商及作者对错误或遗漏不承担任何责任，包括但不限于因使用或依赖本书造成的损失。使用本书中包含的信息和说明所产生的风险由您自行承担。若本书包含或描述的任何代码示例或其它技术，受开源许可或其它知识产权的约束，则您有责任确保您的使用符合此类许可和/或权利的要求。

本书是 O’Reilly 和亚马逊云科技合作的一部分。请参阅我们的编辑独立性声明。

978-1-492-04482-6

[LSI]

目录

1. 云中的数据库选项 1
迁移到云的总体效果 3
自行管理的数据库与托管数据库 5
原生云数据库 6
托管数据库的类型 7
托管数据库可能不太适用的情况 9
DBA 在托管数据库中的角色 9

2. DBA 在云中的角色不断变化 11
DBA 任务在云中发生了什么变化 13
云中数据和应用程序的安全性 19
基础设施即代码：充分利用云 23

3. 将数据库迁移到云 25
规划 26
数据移动 32
优化 35

4. 总结 39

第 1 章

云中的数据库选项

众所周知，云技术越来越受欢迎，而企业常常仅用纯商业指标，来衡量市场对云技术的追捧。譬如，近期的一份 Gartner 报告（需提交企业信息才能查看）发现，云中的数据库增长率为 68%，而各企业本地实体数据库方面，除了价格上涨以及所谓的“强制升级”外，各组织实体数据库数量几乎没有增长。

但是，计算和数据库访问的巨大转变，还体现在所提供的服务类型，以及计算专业工作的发展方面。当下，各种托管数据库已成为各个主要云供应商的产品的一部分。使用这些数据库，就能省去传统环境中由数据库管理员 (DBA) 执行的许多任务，在传统环境中，组织拥有自己的硬件，我们称之为本地环境。迁移到云会增加新的任务，改变一些现有任务，同时提供稍微不同的环境，以便理解许多任务。

云供应商紧跟数据库空间的变化，及时改变其数据库产品，以满足客户的各种需求。关系数据库（以事务和数据仓库的形式）与非关系数据库（例如键/值和文档存储）一起出现。

本报告有助于 DBA 及相关人员（例如数据科学家、数据架构师和应用程序开发人员）选择云产品。报告中概括性地介绍了 DBA 在云环境中应承担的职责。最后，本报告还提供了迁移指南。

本报告没有涵盖支持或反对迁移到云的观点，这是因为还有其它资源支持这一决策，而且该决策与数据库的特定特征，以及使用数据库的方式紧密相关。此外，尽管我们涉及的是当前占主导地位的云供应商的产品（Amazon Web Services (亚马逊云科技)、Microsoft Azure 和 Google Cloud Platform），但我们不评价，也不推荐特定的云产品。

本章包括以下内容：

- 简要回顾本地计算和云计算之间的差异，并着眼于与 DBA 相关的内容
- 托管数据库的定义
- 云提供商提供的托管数据库的类型
- DBA 在云中的角色概述

云产品通常对应于适用于传统本地部署的软件。这意味着您可以将数据，从您自己的设备迁移到云供应商提供的类似数据库。例如，您可以在云（而不是自己的数据中心）中运行 MySQL 或 Oracle，也可以使用原生云数据库，例如 Amazon Aurora。

您还应该能够涉足各个领域（本地和云），以适合您的组织的进度，将数据迁移到云。您还可以永久维护两个部署，一个是本地部署，另一个云部署。云通常是尝试全新事物的好地方，例如您从未创建过的大数据项目。

尽管本报告的大多数读者可能都熟悉云、云的优势以及使用或避免使用云的各种原因，但我们在下一节依然会总结构成本报告思想基础的云的特征。之后，我们将重点讨论数据库。

迁移到云的总体效果

将数据移到云时，您会遇到与本地部署不同的情形，这些情形会影响您的决策。以下是 DBA 感兴趣的主要更改：

- 采用即付即用的方式来运行各个服务器（在虚拟计算的术语中称为实例），费用按小时或按月递增。因此，如果您每晚都要运行商业智能 (BI) 应用程序，您可以在夜间启动一个数据库实例，持续运行 8 个小时，而对于一天中剩余的 16 个小时，您无需付费。借助某些服务（例如亚马逊云科技预留实例），您还可以在限定时间内以折扣价运行数据库。
- 硬件和相关基础设施由供应商负责处理。这会大大减轻系统和数据库管理员的工作负担。您可以选择数据库实例中需要的 CPU 功率、内存和存储容量，而不用配置自己的硬件。供应商通常会提供特殊的产品，这些产品可以通过与本地使用的策略相关的方式提高性能，例如，Microsoft Azure SQL 数据仓库提供固态硬盘 (SSD) 存储，Amazon DynamoDB 具有特殊的内存中缓存版本，等等。
- 主要的云供应商（亚马逊云科技、Azure 和 Google Cloud）不仅提供了虚拟机 (VM) 中的工作负载隔离，还提供了云中整个公司网络的隔离，这项服务称为虚拟私有云 (VPC)。就像虚拟机允许您的每台服务器高效且安全地使用计算资源一样，VPC 允许您在云中建立公司网络，从而节省资金，同时让云供应商管理大量的网络安全事务。数据库服务器可以参与这些 VPC。您还可以设置虚拟专用网络 (VPN)，以便在云的 VPC 与本地计算机之间进行通信。
- 云供应商提供了多个位置，这对于实现以下几个目的十分重要：提供地理位置上接近客户端用户群的服务器，在发生灾难的情况下提供冗余，允许负载均衡等等。出于隐私或其它法律原因，一些国家和地区（特别是在欧盟）会要求将某些类型的数据存储在其法律管辖范围内。区域（可能覆盖整个大陆的大部分）这些区域内可用区 (AZ) 的概念出现在了亚马逊云科技、Azure 和 Google Cloud 中。

因此，您可以跨区域内的多个可用区分配工作负载，复制数据和服务，并在冗余数据副本之间设置负载均衡，以提供弹性。一些云供应商为区域提供特殊的功能来满足法规需求，尤其是在安全性方面。所有主要的云提供商还提供内容交付网络 (CDN) 服务。

- 云供应商提供了先进的方式来扩展服务、提供容错能力以及执行负载均衡。如果您的数据库实例出现故障，且配置设置为高可用性，那么可能有新的实例可用。当旧实例出现故障时，客户端可能需要重新启动在该实例上运行的会话，但您的管理工作是配置容错机制，而不是在凌晨 2:00 起床，重新启动数据库。同样，如果数据库上的负载变高，新的实例会启动来共享负载，然后在不再需要时关闭。这称之为弹性扩展。
- 供应商通常提供命令行选项、图形界面和 API，每种选项都有计算机专业人员所熟知的优势和不足。API 允许自动化，因此可以减少您的工作量。
- 您原本需要担心的关于本地部署的某些安全性问题由供应商负责处理。显然，他们对谁能进出服务器室、谁能对服务器进行调试，拥有控制权。根据您使用其服务的方式，他们还可以保护您免受低级网络攻击，例如端口扫描、拒绝服务 (DoS) 攻击等。后面的章节列出了 DBA 仍需负责的安全性方面的事务。
- 云供应商内置了监控和性能工具，您可以轻松地将其连接到数据库中，而且这些工具总是会涌现出新的功能。第三方监控工具也可加以利用，也可以扩展平台的内置监控功能。

选择提供商的许多标准是通用的，而非特定于数据库：定价、围绕各种产品的第三方开发人员的生态系统、稳定性、对法律合规制度的支持、高端人工智能工具的附加值等等。

评估云产品时，请考虑企业的未来发展方向。例如，流式传输和实时数据处理在未来会占很大比重吗？ 几年后，哪些云产品会为您提供支持？

自行管理的数据库与托管数据库

现在，我们专门来探讨数据库。云中的选项分为两大类：

自行管理的数据库

在此云产品中，供应商仅提供硬件、用于运行您的虚拟机的管理程序、以及用于管理部署的 API 或其它工具。您需要自己创建一个虚拟机，来运行操作系统和您的应用程序。显然，您可以在虚拟机中运行任何您想运行的东西，包括您选择的数据库引擎。您需要自己执行大多数的管理任务，例如安装更新和配置所有联网选项。

托管数据库

在此云产品中，供应商不仅提供硬件，而且还提供服务器软件。大多数供应商同时提供传统数据库（例如 Oracle 和 MySQL）和供应商特定的原生云数据库。

自行管理的数据库更像是在本地运行您自己的数据库；托管数据库则会极大地改变您的工作方式。以下是两者的主要区别：

- 如果您使用的是托管数据库，则无需手动下载、安装、更新、配置或备份数据库；云提供商负责完成所有这些工作。您仍然可以修改参数，指示数据库如何运行（定序、游标、连接等）并控制决策，例如是否安装新版本以及保留备份多长时间等。
- 有时候，云供应商可以提供专有数据库的许可证，并将成本包含在云服务成本中。在其他情况下，您仍然需要负责获得许可证。如果您已经向专有数据库供应商购买了许可证，则有时候可以将其应用于托管数据库。
- 因为供应商负责安装并运行托管数据库，所以您只能采用供应商的选择。您无法使用供应商不支持的数据库引擎，甚至无法使用尚不支持的版本。但是，三大供应商（亚马逊云科技、Azure 和 Google）提供几个专有开源数据库以及他们自己的原生云数据库。

- 自行管理的数据库和托管数据库提供不同的配置参数。
- 托管数据库的大多数安全问题，均由供应商负责处理。但即便是托管数据库，重要的安全决策也还是由您来制定，例如向谁提供账户，以及在表或列上设置哪些限制等。监控和审计访问尝试通常也是您的工作。您还可以控制安全设置，例如哪些 IP 地址可以访问之类的。

因此，与本地部署和自行管理的数据库相比，托管数据库具有许多优势，但您也可以出于独特的原因运行自己的数据库。本报告的剩余部分将介绍托管数据库，因为这类数据库让 DBA 能够专注于对业务具有更长远影响的活动。

原生云数据库

托管数据库也可以分为两类：传统和原生云。云提供商经常将传统数据库（例如 Oracle、SQL Server、MySQL 和 PostgreSQL）作为托管数据库提供。如果您已经围绕某个本地数据库构建了组织，则可以迁移到云中相同的数据库，从而简化迁移。您基本不需要更改应用程序，而且可以使用熟悉的工具来管理数据库。您还可以混合使用不同供应商的产品，也可以维持本地版本的数据库。

不过，供应商已经投入了很大的精力来开发新产品，这些新产品有时候称为原生云数据库。供应商提供的证据表明，原生云数据库的性能更好、更容易扩展、而且从长远来看成本更低。因此，Autodesk 和 InfoScout 的评价表示，亚马逊云科技工程师已经使用他们自己的数据库 Amazon Aurora，解决了在云中管理关系数据库时遇到的许多扩展和效率问题。原生云数据库还采用了可以大规模扩展的设计，大规模扩展是关系数据库历来难以实现的任务。

托管数据库的类型

领域内大多数类型的数据库也以托管数据库的形式提供。此外，云供应商也开发了他们自己的原生云数据库，遵循行业共同趋势并提供性能优势。受支持的数据库的主要类型包括：

关系数据库

如前所述，一些云数据库是广泛使用的常见数据库的托管版本。例如，Azure 提供 Microsoft 的传统 SQL Server。Amazon 通过其 Amazon Relational Database Service (RDS) 支持 MariaDB、MySQL、Oracle、PostgreSQL 和 SQL Server。此类产品可以帮助您更轻松地从本地安装迁移数据库。

此外，供应商也创建了自己的数据库。例如，Azure 提供 Azure Cosmos DB，Google 提供 Cloud Spanner，而亚马逊云科技则提供 Amazon Aurora。

数据仓库

尽管这些通常都是关系数据库，但在内部和服务方面，与事务数据库有所不同。例如，事务数据库通常将一行的所有列存储在一起，以便您可以快速检索有关客户或产品的多个列。而云中的数据仓库往往是列式的，也就是说数据仓库按列（而不是按行）存储数据。这大大加快了常用仓库查询的速度，例如“告诉我居住在加利福尼亚州的所有客户的年龄”这样的常见查询。拥有数据仓库的云供应商提供的工具侧重于快速摄取和提取，从而方便在大数据应用程序中使用。这类产品包括 Amazon Redshift、Google BigQuery 和 Azure SQL 数据仓库。

非关系数据库

这个术语通常用来指与传统关系数据库不同、专为特殊

应用程序使用案例而构建的各种不同的数据存储。云供应商出于不同的目的提供以下各种数据库：

- 键/值数据库可以快速存储和检索值，而无需支持更复杂的操作。
- 文档数据库使用灵活的架构将数据存储为 JSON 文档，从而允许以应用程序中使用的格式存储和查询数据。
- 关系图数据库存储对象之间的关系，从而让您可以轻松地运行算法，例如查找连接最多的对象。
- 搜索数据库可以优化包含特定词语的文档的位置。
- 时间序列数据库使用时间戳记录事件，并针对与时间相关的任务（例如绘制随着时间推移发生的事件）进行了优化，可用于跟踪各种事件，例如设备读取或网络发帖等。
- 分类账数据库类似于增强的安全日志文件，可以记录金融交易等活动，对其进行签名，并使其不可变。

当然，可能还提供其它选项，包括内存中缓存数据库等。云供应商提供各种适用于结构化程度较低的存储（例如 BLOB 或文件）的服务：示例包括 Amazon Simple Storage Service (Amazon S3)、Google Cloud Storage 和 Microsoft Azure Blob Storage。这些服务提供了更实惠的存档存储方案，还提供了相关工具，使您轻松地将数据移入和移出存储。

本章不可避免地使用了普通措辞来描述数据库特性，其中包含许多像“一些”或“可能”这样模棱两可的词语。未明确描述的原因是，如果您仔细查看供应商的产品，则可

以发现其中存在细微的差异。有些产品比其它产品发展得更快。有些数据库比其它数据库更易于管理或迁移。在选择云产品时，您必须考虑所有这些因素。

托管数据库可能不太适用的情况

托管数据库服务不适用于两种情况。

数据库大小和 IOPS 超过了数据库限制

原生云数据库旨在进行扩展以满足极端需求。但是，对于以更传统的数据库引擎为基础的托管数据库，在支持的大小和每秒输入/输出操作 (IOPS) 方面通常会有一些上限。许多组织都可以适应这些限制。但是，如果您的数据库确实非常庞大，并且您的 IOPS 需求超出了提供商所能支持的范围，那么您可能会发现您所选择的托管数据库服务无法满足需求。您应该运行原生云数据库，或者通过租赁所需的虚拟机和存储，在云中运行自己的实例。

您需要数据库的完全控制权

如果对您来说最大限度地控制数据库很重要，那么您必须在云中管理自己的数据库。如果您需要云供应商不支持的特定数据库版本，或者需要使用提供商不支持的数据库功能或选项，那么您就需要自己管理了。

DBA 在托管数据库中的角色

正如本报告开头所提到的，迁移到云会改变您的工作方式，并且有时候会以微妙的方式改变。第 2 章将对此进行详细介绍，但我们必须要告诉您的是：虽然托管数据库会帮您完成硬件维护和系统配置等工作，您还有许多工作要做。以下是 DBA 通常在托管数据库上执行的一些任务：

- 确定各种要求，例如 CPU 功率和内存。
- 配置数据库运行，包括自动扩展、故障恢复和拍摄数据库快照。

- 选择托管数据库实例的区域和可用区。
- 使用供应商提供的命令行、控制台或 API 创建和处理数据库和表。
- 授予用户账户权限、设置组并控制对数据库以及数据库组件的访问。一些数据库通过 Identity and Access Management (IAM) 提供这种访问权限，而其它数据库则使用本地部署中提供的传统接口。
- 确定在需要将数据分发到多个系统时，如何将数据分段。
- 设置影响性能或弹性的数据库参数。
- 出于安全、性能和弹性目的，审计访问权限并监控数据库活动。
- 通过传统的提取、转换和加载 (ETL) 或最新的流数据工具，摄取和传输数据；数据传输可能涉及接受来自外部来源的数据、将数据提供给 Spark 或 Kafka 等大数据工具、将数据存储在云中或本地的数据仓库中，以及与其他工具建立所需的任何其它关系，以完成数据处理。
- 与数据架构师和应用程序开发人员一起确定架构或格式，并在关系或非关系数据存储中实施。

最令人兴奋的是，您将有更多时间查看数据库的业务环境，并思考如何提高您为组织管理的数据的价值。我们将在第 2 章中更详细地探讨这些课题。

第 2 章

DBA 在云中的角色不断变化

传统上，数据库管理员 (DBA) 受托负责与组织数据的存储和使用相关的几乎所有操作。DBA 的任务列表包括以下内容：

- 安装机架和布线
- 安装、修补和升级数据库软件
- 创建和配置数据库实例
- 管理用户、角色和权限
- 备份和恢复
- 保护数据安全，包括加密关键数据
- 迁移、摄取和导出数据
- 优化性能
- 监控数据库和排查问题
- 处理高可用性
- 协助数据库开发人员和分析人员完成与数据库相关的任务

从这个列表中可以看出，DBA 的工作通常围绕数据库操作、灾难管理、用户管理和性能优化进行。监控和问题排查大部分需要手动操作，而且通常是被动响应，而不是计划的工作。

这样一来，便没有多少时间可以用来执行价值更高的任务，例如设计数据库和优化依赖于数据库的应用程序。

自动化程度较低以及被动的问题处理方法，使得 DBA 难以妥善管理他们的各种数据库。他们大多数的日常工作都变成了对任务进行分类，并且始终需要处理最紧急的任务。

随着云计算从根本上改变了工作角色和组织经营的方式，DBA 原本传统的职责和任务也开始进行结构性重组。在云中，DBA 的许多传统任务已经完全消失或者大大减少。当然，云提供商会处理所有基础设施工作，例如安装和堆叠服务器、联网和存储。大部分备份和安全性工作也转为由云提供商负责。那么，DBA 在云中负责什么呢?

如果您要在云中运行自己的数据库，那么您仍将要执行许多传统的 DBA 任务，例如安装和修补软件、备份数据库，等等。本报告重点介绍托管数据库（如第 5 页“自行管理的数据库与托管数据库”中的定义），因为托管数据库比自行管理的数据库更具优势，而且 DBA 的角色也有相当大的变化。本章还将重点介绍关系数据库，这些数据库通常具有更多管理任务。

在云中使用托管数据库时，您可以将更多时间用在数据架构和数据库支持的应用程序上。现在，预测需求和改进数据库已经成为 DBA 的主要关注点。通过帮助团队更快地交付新功能，并主动优化应用程序性能，您所处的环境将大大改善，进而可以从组织的数据资产中获取更多价值。

其它任务还包括帮助建立批处理和实时数据管道，以摄取和转换批处理和流数据。构建、维护和优化高度分散的数据管道，是使用基于云的数据库的管理员的重要工作。您还可以投入更多时间来加强数据库安全，并确保遵守政府和行业机构提出的合规性要求。

在云中使用托管数据库服务时，传统 DBA 的角色会发生颠覆性改变。图 2-1 总结了 DBA 在本地和托管数据库服务上投入时间的区别。

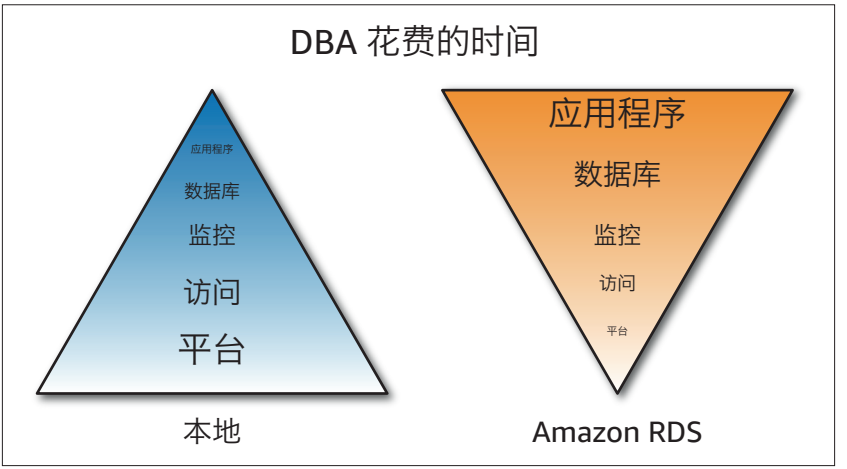

图 2-1. DBA 在云中的角色发生颠覆性改变

简而言之，将数据库迁移到云可以节省管理员和开发人员的时间，您可以将这些时间用于其他更重要的任务。由于云提供商会负责例行维护和备份等工作，因此您可以将更多精力放在业务目标上，例如简化关键业务流程。

在新的云环境中，熟悉所有相关的云提供商服务对于成功至关重要。其中包括各种存储选项、监控工具（通常不需要额外付费）、云安全功能以及其他内置云服务，您都可以利用这些服务，以经济高效、高性能且安全的方式，运行您的云数据库。

DBA 任务在云中发生了什么变化

将数据库迁移到云中后，DBA 的许多传统操作任务消失，有些任务仍然存在，但有所改变，还出现了一些新任务。不过，由于新工具和选项的出现，几乎所有本地任务都会有所改变。

由供应商接管的 DBA 任务

在云中使用托管数据库服务与使用本地数据库的一个主要区别是，DBA 无需实际访问服务器或直接控制安装便可执行管理任务。DBA 最初可能需要一些时间，才能适应这个新的现实。对于 DBA 来说，重要的是抓住这一机会扩展到本章中讨论的新任务。

仍然存在但有所改变的 DBA 任务

DBA 在云中仍然需要履行一些传统职责，但不论您使用托管数据库服务还是运行自己的数据库，这些职责都发生了细微的变化。

调配

托管数据库服务（例如 Amazon RDS）为您的数据库服务器提供了广泛的实例类型。您无需为数据中心订购具有特定内存和 CPU 要求的硬件系统，而是可以选择一种或多种实例类型，来提供满足需求的内存和 CPU 功率。将实例类型与工作负载进行匹配，这可以帮助降低为您不使用的资源付费的可能性。

扩展是另一项调配任务。您可以在虚拟机级别和数据库级别执行自动扩展，使系统可以在负载变高时添加实例，在负载减少时关闭实例，以免浪费过多容量。尽管自动扩展是一种非常有价值的自动化方式，但是您可能会遭受 DoS 攻击或遇到配置错误，这将导致启动大量实例并增加成本。

您还可以选择数据库用户附近的区域，以获得最佳响应时间。对于在云中进行配置而言，合理使用全球云提供商的产品很重要。

正如我们所讨论的，传统数据库，尤其是联机事务处理 (OLTP) 数据库，是出了名的难以扩展。随着用户数量、数据集大小和数据库操作复杂性的增加，查询响应时间也会

增加。业务增长可以体现在所有这些因素中，也可以体现在给数据库带来更多负担的应用程序的增长中。技能娴熟的 SQL 开发人员和 DBA 可以设置最佳数据库配置参数并优化 SQL 代码，在一定程度上改善这些问题。

但是，代码优化和最佳数据库设置的作用有限。因此，在许多情况下，扩展数据库仍然是难题，因为您无法轻松地动态更改服务器、RAM 或 CPU 的数量以满足不断变化的工作负载需求。所有这些资源都涉及资本预算支出，需要提前获得批准。在许多传统数据中心中，增置一台实体服务器，需要数周的时间。

ETL、数据摄取和数据导出

对于 DBA 来说，最重要的任务之一是以某种形式移动数据，通常是以某种方式转换数据以便提取重要字段，或者将数据转换成另一种格式。尽管 ETL 工具已经在这个领域独领风骚近 40 年，但是，Hadoop 和 Spark 等最近出现的流处理形式，带来了许多新的数据摄取工具。Apache Kafka 和 Amazon Kinesis 等消息代理商，在发布和订阅模型中，将数据流定向到多个使用者。大多数网站出于不同的目的使用不同代的工具，而您的职责是掌握每一种工具，并且使这些所有的工具互相兼容。

流数据给传统数据库带来了新的负担，因此迫使 DBA 以不同的方式调整参数。流式处理通常涉及更多进程，需要频繁执行很多小规模更新，而不是少量大规模更新。并发控制也很困难。

无论数据是进入本地数据库还是云存储，常规任务都一样。但是，云供应商提供他们自己的流数据版本，以及用于摄取和导出数据的工具。例如，如果您使用的是 Microsoft 基于云的数据仓库 SQL Data Warehouse，则您可以同时使用工具箱中的现有软件和 Microsoft Azure Data Factory (ADF) 中提供的工具。Amazon 提供了大量摄取工具，例如 Kinesis Data Firehose 和 Snowball。Google 摄取工具箱可以与许多其它来源的选项结合使用。

备份和快照

云数据库提供两种基本的数据库备份和恢复方法。系统通常默认开启自动备份，可用于时间点数据库恢复。

相反，快照备份需要您启动。在本地环境中，DBA 定期拍摄快照（通过在数据处于一致状态时转储数据库中的数据），以备份或复制数据存储。在云中，您仍然可以使用快照进行长期备份（例如磁带存储），复制到云中的其它区域，以及为了遵守相关规定而保留。

云中的托管数据库服务（例如 Amazon RDS）使拍摄快照变得很容易。您可以创建数据库实例（可以由多个数据库组成）的存储卷快照，并通过还原数据库快照创建数据库实例。

访问权限

在云中，您负责与数据和用户相关的一切工作：分发账户、创建组或角色以组合用户账户，以及指定对数据库各个组件的访问权限。我们将在第 19 页的“云中数据和应用程序的安全性”中详细介绍此课题。

开发和测试环境的维护

DBA 常见的一个任务是创建开发和测试环境。这通常需要 DBA 做许多工作。如果您定期拍摄数据库快照，则可以创建一个反映给定时间点数据库状态的开发或测试环境。Amazon、Google Cloud 和 Azure 均提供此类服务。

日志

DBA 需要数据库日志，来进行故障排除以及保证合规性和可审计性；这一点在云中仍然适用。图 2-2 显示了如何通过亚马逊云科技控制台轻松查看数据库实例的日志。当您有许多数据库需要监控和管理时，这个功能的作用就很明显了。

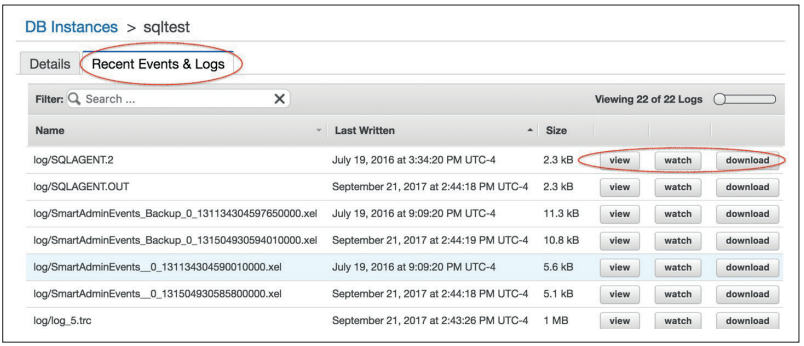

图 2-2. 通过 Amazon RDS 控制台查看数据库日志

监控和通知

云供应商提供大量指标信息，例如 CPU 利用率、查询延迟、查询吞吐量以及 I/O 和内存使用情况。他们还提供许多内置的监控功能，有些是免费的，有些是付费的。您还可以设置告警，以便在这些指标超出预测范围时迅速获悉。

许多云平台都会公开一些事件，例如故障转移、备份、数据库停止或重新配置。这些平台提供挂钩，使您可以在事件发生时设置通知或采取其它自动操作。

性能优化

优化数据库性能是 DBA 的传统关键任务之一；这项任务能将出色的 DBA 与一般管理员区分开。若要更好地优化性能，则需要深入了解存储、网络、内存和应用程序代码。将数据库置于云中并不能免除 DBA 在数据库方面的全部工作。无论是被动还是主动性能优化，都仍然是 DBA 的主要工作。不过，云供应商提供了一些简单的方法来获取信息，例如亚马逊云科技 RDS 性能详情。

在云中，您无需花太多时间处理与服务器和存储相关的典型性能问题，因为您可以根据工作负载轻松地纵向或横向扩展数据库。您可以将更多时间用在分析数据库等待上，以便改进数据库响应时间。

在传统的本地环境中，数据库自带一些性能工具，例如 EXPLAIN 命令和 MySQL 慢速查询日志。其它性能监控工具在数据库之外运行，您可以在云中设置一个系统来运行这些工具。但云供应商一直在努力开发云原生性能工具，通常不收取额外费用。您应该充分利用供应商提供的东西。

如果数据库在本地运行时性能不佳，那么不会因为迁移到云中便奇迹般变好！ 但在某些情况下，迁移可能会有所帮助，例如，当本地数据库环境无法支持应用程序对大量 IOPS 的需求时。

处理高可用性

如今，这项任务很关键，因为如果网站在两三秒钟内无法完成加载，使用者就会离开。在本地，这项任务需要提前购买足够的处理能力，还需要配置负载均衡器。在云中，弹性伸缩服务（autoscaling），并将数据库配置为在多个可用区中运行。如果一个可用区出现故障，则多可用区数据库可以自动执行故障转移至另一个可用区。同样，您可以在多个区域中提供数据，实现高可用性和本地存储。

DBA 在云中的新任务

在云环境中运行数据库时，您会遇到一些任务，在大多数情况下，数据中心没有这些任务。这些任务包括重新定位成本和在云中跟踪数据库许可证。

计划运营成本

云中的总体拥有成本 (TCO) 和投资回报率 (ROI) 基本计算发生了根本性变化。正如第 14 页的“调配”中所提到的，您通常只需为云中特定的实例和特定使用时间付费，而无需预先购买固定的基础设施。但是，您需要考虑许多与数据库没有直接关系的因素。例如，将数据库放在离用户较近的区域可以降低网络成本，甚至可能催生新业务。

供应商提供多种选择，可以满足您的使用量需求，并大大减少您的开支。例如，通过在亚马逊云科技中购买预留实例，您可以显著地降低成本。一般来说，通过认真计划，您可以预测自己在特定时间的需求，还可以避免本地数据中心超量调配。

对架构和应用程序设计的影响

基于云的数据库的架构与数据中心中传统本地数据库的架构有着本质的不同。在制定数据库部署计划时，您必须考虑多个地理区域和可用区、不同类型的网络设计以及其它云特有的特性。DBA 必须与应用程序开发人员保持密切联系，以便了解他们的需求，并以尽可能低的成本在云中提供最佳配合。

跟踪许可证合规性

专有数据库通过两种基本许可模式在云中提供：附带许可和自带许可 (BYOL)。如果您已经拥有许可证，并且系统允许使用 BYOL 模式，那么 BYOL 模式很有用。您可能需要咨询许可专家或供应商，以了解您的现有许可是否适用于云。

云中数据和应用程序的安全性

引入云计算后，管理员通常负责堆栈上的任务，例如，最底层的实体服务器安装和布线，以及顶层应用程序管理。如果使用托管数据库，则这些任务主要由云提供商负责，只有应用程序级别的任务由客户负责。如果使用自行管理的数据库，则大多数任务由客户负责，而供应商主要负责提供实体服务器和基本网络。

安全任务也符合这个堆栈概念，这催生了通常所说的安全责任共担模式。供应商在托管数据库中承担的安全任务比在自行管理的数据库中承担的多。简而言之，云提供商负责基础设施安全，而您仍负责云中数据和用户信息的安

全。因此，云供应商负责执行的任务包括以下各项：

- 实际保护数据中心和设备，包括监督员工
- 按照客户的指导原则，通过复制和备份确保冗余
- 修补和更新软件，包括数据库本身（如果使用的是托管数据库）
- 运行用于监控和审计访问权限的工具
- 提供网络和应用程序防火墙
- 防止 DoS 攻击
- 提供配置的身份管理服务

客户负责执行的安全任务包括：

- 通过供应商提供的 identity and access management (IAM) 系统设置用户和角色
- 为数据库中的用户分配访问权限，包括定义账户和角色，以及在用户不需要访问权限时删除该用户来随时更新账户
- 指定云供应商提供的防火墙规则、备份和其它参数
- 设置告警，并查看未经授权访问的日志
- 保护操作系统（如果您使用自己的虚拟机而不是托管数据库）
- 保护应用程序，例如防止 SQL 注入攻击
- 对静态和传输中的数据进行加密
- 对试图访问数据库或应用程序的用户进行适当的身份验证和授权
- 收集和检查日志
- 设置事件告警，并监控事件以检查是否存在异常

熟悉全部云安全功能，对于在云中运行数据库至关重要。在下面的小节中，我们将介绍 DBA 应该重点研究的关键云

安全功能：访问控制和 IAM、网络隔离和数据加密。

访问控制和 Identity and Access Management

云提供商通常擅长强大的访问控制机制。本报告中提到的供应商依赖集中式 IAM 来管理用户、安全凭证（密码、访问密钥和权限）以及控制用户可以访问哪些资源和服务的授权策略。管理员需要掌握 IAM，以便自己和用户访问云。此外，原生云数据库有时与通用云 IAM 工具集成。将数据库保护与云的 IAM 关联，是最简单、最安全的访问策略。

您将使用 IAM 定义用户账户，然后添加数据库特定的访问规则。使用 IAM，您可以授予不同的用户权限来执行不同的数据库操作。您还可以实施精细访问控制（大部分数据库都会提供），例如通过 IAM 对特定行或列的限制。

传统数据库在云中使用的工具，与您在本地数据库中使用的相同，例如 SQL 中的 GRANT 语句。但是，您可以将这些工具挂载到 IAM 中，以便在云和数据库中使用相同的用户账户，并通过这种集成获得额外的安全性和便利性。

网络隔离

保护网络中系统的一些常见云功能包括 VPC、防火墙和网络访问控制列表（ACL）。

正如我们在第 3 页“迁移到云的总体效果”中所讨论的，VPC 是云中的私有网络，用于服务器相互通信。在 VPC 中，您可以指定允许访问每个数据库的 IP 范围，来隔离数据库实例。创建 VPC 的组织可以完全控制其虚拟网络环境，并可以选择自己的 IP 地址范围、创建子网以及配置自己的路由表和网络网关。

此外，您还可以设置一个虚拟私有网关，将公司网络扩展

到您的 VPC 中，并允许通过您选择的 VPN 访问该 VPC 中的数据库实例。

大多数云提供商提供内置防火墙，帮助您控制网络对计算实例的访问。云提供商还可能提供私有或专用连接选项，将云消费者的办公室和本地部署环境，与云环境连接起来。您可以设置数据库安全组，来保护 VPC 内的数据库实例。安全组就是防火墙规则，作用在于控制对云数据库的网络访问。您还可以通过网络 ACL，允许或拒绝网络流量流入或流出子网。本地安全基础设施（例如防火墙）和入侵检测系统，能够检测通过 VPN 流入或流出 VPC 的任何网络流量。

备注

直接连接

您可以使用亚马逊云科技的 Direct Connect 或 Azure 的 ExpressRoute，将云外系统连接到云内系统，而不是使用 VPN 连接。他们利用电信运营商提供的私有网络链接，建立直接连接。仍然建议使用端到端加密，通常通过标准安全套接字层 (SSL)/传输层安全性 (TLS) 来完成。

数据加密

在云中，您可以轻松加密数据，为静态数据提供额外的保护。例如，如果您在 Amazon RDS 集群中启用加密功能，则数据库会以加密形式存储您创建的表中的所有数据。加密也适用于数据库备份。在云中，您也可以轻松加密传输中的数据。

对于必须满足诸如以下行业合规性要求的组织来说，加密是特别必要的：《美国健康保险流通与责任法案》(HIPAA) 医疗保健合规性要求、《萨班斯-奥克斯利法案》(SOX) 财务报表合规性要求，以及《支付卡行业数据安全标准(PCI DSS) 电子商务和零售业务标准。如果您保护加密密钥，则加密数据本身的丢失通常不会被视为可报告的安全事件。

云提供商通常提供托管服务来简化加密密钥的创建、控制和管理。因此，Amazon Key Management Service (KMS) 提供组织中所有密钥用户的集中视图。KMS 与亚马逊云科技 CloudTrail 集成来提供日志，日志显示整个组织的密钥使用情况信息，从而满足若干密钥法规和合规性要求。

基础设施即代码：充分利用云

当您将数据库迁移到云中时，总是希望尽量减少管理更改，因为您已经习惯了以某种方式做事。但是，这样做只是新瓶装旧酒，无法充分利用云提供的巨大好处。

与本地数据中心中的系统相比，基于云的系统有一个关键技术进步，那就是将基础设施视为代码的配置系统很容易获得。将基础设施视为代码使许多开发运营实践成为现实，这反过来又促进了开发人员和运营人员之间的密切协作，使他们能够大规模地自动交付应用程序。在本地，您当然可以安装自动化工具，例如配置管理系统（如 Puppet 和 Chef，这两者比较常用）和持续集成 (CI) 工具（如 Jenkins）。但此类工具中非常先进的那部分功能是内置于云中的。

例如，Amazon CloudFormation 随附的模板可帮助您通过建模将整个基础设施抽象为代码。您可以通过创建和配置资源（如数据库表和 (Amazon S3) 存储桶）来定义云基础设施，并将这些资源视为代码。您可以将Amazon CloudFormation 模板签入您的源代码控制系统，并采用与开发人员管理其代码文件相同的方式来管理这些模板。

CloudFormation 之类的工具提供了以下好处：

- 帮助为您的所有云资源建立单一事实来源。
- 可以与代码管理工具（如源代码控制系统）集成。
- 帮助自动执行开发和测试部署。
- 支持您的灾难恢复计划。

通过使用模板文件以编程方式创建资源，您可以在整个环境中实现可重复性和一致性。

第 3 章

将数据库迁移到云

与大多数组织变革一样，迁移到云并不是一蹴而就的。您应该选择一个项目在您选定的云提供商处进行尝试：一个提供了有用测试案例的现有项目，或者一个不受传统做法约束的新项目。

几乎所有迁移到云的组织都是先对非关键数据库进行概念验证型迁移。在实现了容易达成的目标后，或者成功迁移了一个从使用云获益的小型项目后，您可以将获得的经验教训推广到整个组织中的其它数据库。

本章将帮助您了解通常与成功的首次迁移相关的条件。许多内容都适用于关系数据库。尽管每个迁移都是独一无二的，但您在迁移时可能免不了要执行以下所有或大多数步骤：

规划

- 收集需求
- 确定满足需求的能力
- 评估要迁移的数据库以及可能需要对数据库或使用数据库的应用程序进行的更改
- 建立成功标准和回滚标准（故障保护）

数据移动

- 复制
- 合并自副本创建以来的更改
- 应用程序测试
- 转换
- 迁移后检查

优化

- 性能优化
- 设计高可用性
- 确定要记录和监控的事件
- 创建灾难恢复计划

我们将在下面的部分中介绍各个主要阶段。

规划

此阶段可帮助组织评估以下问题：

- 当前环境的相关元素：应用程序、数据库和关键工作负载
- 应用程序或工作负载是否会在云提供商的环境中正常运行
- 迁移将如何帮助实现业务目标
- 对系统最终状态的要求
- 在云中运行当前计算环境的成本，以及综合投资回报率，如第 18 页的“计划运营成本”中所述

迁移中的因素

当您想要将本地运行的数据库迁移到云中的托管服务时，您需要处理实体、软件和组织方面的问题。这需要注意以下几个级别的变化：

实体数据移动

您需要将数据本身放到云中。大型数据库的数据传输既费时又费钱，但有一些迁移服务可以提供帮助。

基础设施兼容性

针对您的传统本地系统检查云服务的各个方面。这篇亚马逊云科技博客中的一个案例研究提供了一个有趣的示例，，展示了测试迁移期间，由于意外的软件不兼容和轻微的错误，而可能出现的障碍。由于本地系统和云提供商在 SSL 实施方面的差异，这种特别的迁移被耽误了一段时间。

数据库兼容性

云数据库可能在某些影响迁移的功能上与本地数据库有所不同。您可能会在数据库字符集和存储过程权限等细微之处之处出错。同构迁移（例如，从一个 Oracle 数据库到另一个 Oracle 数据库）可能比异构迁移（例如，从 Oracle 数据库到 PostgreSQL 数据库或专有云供应商的数据库之一）更为顺利。第 31 页的“检查是否存在不兼容”说明了如何减少数据库之间的冲突。对于某些类型的迁移，云供应商可能会提供工具来减轻您的工作负担。

如果您当前正在本地运行旧版本的数据库，则可能是时候升级了。不要只选择同一关系数据库管理系统 (RDBMS) 的较新版本（此过程伴随有一定程度的痛苦），您还可以利用这次机会来，来尝试向展示出较好前景的新 RDBMS，进行异构迁移。

组织变化

有些团队中的人员（特别是 DBA 和应用程序开发人员）需要花时间来学习云工具和管理新实例，同时保持旧系统在本地运行，直到您准备好将所有内容迁移到云。您可能选择花费数月甚至数年的时间来完成整个迁移。您还可能将某些数据（或数据的副本）永久保留在本地。

主要迁移任务

我们建议执行以下步骤来实现成功迁移。

创建云迁移计划

云迁移计划应按您希望将所有数据库和应用程序迁移至云的顺序列出这些数据库和应用程序。最终评估计划应描述所有数据库的迁移计划。如果您在新的数据库环境中需要新的资源（软件、财务或人员方面），那么这也应包含在计划中。

确定执行迁移的人员

DBA 和开发人员应协作完成迁移工作，因为每个团队各有所长。在选择人员时，您还可以决定是使用云供应商提供的迁移服务，还是第三方公司提供的迁移服务。

教育任务

这包括一些宽泛的任务，例如了解云供应商工具的使用、目标数据库引擎的特定功能、数据库迁移的范围以及云数据库的架构。

创建云数据库架构

选择要使用的数据库的类型，并决定是使用托管数据库还是自行管理的数据库。做此选择时，请权衡所有相关因素，例如成本、性能、可靠性和可扩展性。

选择迁移过程

第 32 页的“迁移数据库”列出了在复制、备份/还原和专用迁移服务之间进行选择的标准。

对于自行管理的数据库，创建您的计算基础设施

如果您选择在云中管理自己的数据库，可通过供应商的计算服务（如 Amazon Elastic Compute Cloud (Amazon EC2)）创建虚拟实例并进行部署。

确定架构变更的机会

您可能会发现，在本地进行的选择不再适合云中的新环境。例如，您可能能够合并分区。

设置云数据库账户

在数据库中设置用户、角色和组之前，请通过 IAM 为云中的用户设置常规账户。

升级和测试应用程序

您可能需要重写代码才能使用新数据库。应用程序的某些迁移可能能够自动执行，就像数据库本身一样。

就绪情况评估

就绪情况评估可以帮助您估计成本、云数据库的架构、迁移计划以及云迁移对合规性法规的影响。云就绪情况评估的结果是一份详细的报告，说明了您的公司对于将数据库迁移到云的就绪情况。

以下是对云就绪情况评估中的关键步骤的简要说明。

利益相关者访谈

与应用程序开发人员、业务用户以及和组织团队中的数据使用有关的其他人员进行交流，这将帮助您确定与性能、高可用性以及基于云的数据库的功能有关的要求。

当前本地数据库的分析

重新研究当前的数据库，以确定数据的增长模式、备份和恢复策略，以及正在进行的数据导出和导入等等。了解当前的数据库使用模式，将帮助您决定在云中使用什么数据库，以及是否需要特定的数据库产品来获得所需的功能。

要迁移的数据库的优先级确定

选择要先迁移到云的数据库。一个重要的标准是需要进行的更改的程度，而这又需要与开发人员进行协调。

迁移成本分析

对迁移到云的成本与留在本地数据中心的成本进行彻底分析，以便完全了解云迁移的 TCO 和 ROI。

通过使用云数据库服务降低成本可能是您的关键目标之一。但是，一旦云服务选择不当，您就无法达到此目的。

安全性与合规性

有时，您需要特殊的区域或可用区来遵循服务组织控制 2 (SOC 2)、PCI DSS 或 US HIPAA 等标准。遗憾的是，云提供商通常会将特定的区域和可用区与这些法律要求进行匹配。此外，还有一些专用区域，例如亚马逊云科技 GovCloud（美国），这是一个受 FedRAMP 高基准和中等基准约束的隔离亚马逊云科技区域。最后，您需要检查组织的服务级别协议 (SLA) 是否能在所选的云设置中实现。这些 SLA 通常包括计划的维护、备份、恢复点目标 (RPO) 和恢复时间目标 (RTO) 的指标。

如果您已正确完成云评估，您还应该清楚地了解云迁移的范围。您的某些应用程序可能太旧了，需要完全重新设计才能迁移到云。如果您目前没有资源来重写应用程序，您可以将这些应用程序保留在本地数据库中，或者将整个服务器迁移到云中。

迁移步骤

您应该准确规划数据迁移的时间和方式，以及在之后的测试表明数据库无法正常工作时，回滚更改的准备工作。为每个步骤建立测试和成功标准。

应用程序更改

您需要与开发团队协作开展此工作。这包括将应用程序更改安排到开发过程和测试中。

自动执行迁移

此步骤很有价值，因为自动化使您可以根据需要多次

重复迁移，可以在遇到错误时修复错误，还可以更改参数。您还可以调整并重用自动化框架以进行进一步迁移。

检查是否存在不兼容

如果您对两个数据库使用相同的 RDBMS（例如 MySQL），请检查版本以确定您使用的是否是云版本不支持的功能。如果您要迁移到新数据库引擎，这些检查显然更有可能发现不兼容问题。

在执行向云数据库的异构迁移时，您会遇到向本地数据库执行此类迁移时遇到的相同转换问题；但是，您可以使用的解决方案更广泛。

数据库供应商本身通常会提供工具，来帮助将数据迁移到同类型的云数据库。例如，Oracle 提供 Data Pump、RMAN 和 SQL Developer 等数据迁移工具，来帮助您将 Oracle 数据库迁移到云中。此外，云提供商和第三方还会提供亚马逊云科技 Schema Conversion Tool 等服务来迁移架构。这些服务会向您说明是否可以转换架构，并在必要时提供有关解决方法的建议。

尽管如此，大多数异构数据库迁移仍然需要自动化工具和 DBA 的手动操作。如果您要迁移至亚马逊云科技数据库，可以从亚马逊云科技网站上获得具体的建议。

一些迁移工具会使用转换器将一个数据库的对象（例如 Oracle 存储过程）转换为非 Oracle 数据库（例如 PostgreSQL）。但是，由于每个数据库引擎都采用独特的编码实践，因此这些工具的工作可能不全面。

这就是为什么要在迁移到云之后对目标数据库执行严格的测试，以确保准确性和性能。特定 RDBMS 所采用的一些编

程技术在目标数据库中，可能没有完全对等的技术。迁移工具可能会在迁移过程中，将这些类型的转换问题标注出来。在许多情况下，这工具还可以向您显示为解决这些问题，并成功完成数据库迁移可能需要运行的代码。

数据移动

云提供商高度重视其潜在客户有关以最低成本和最少中断将本地数据库迁移到云的要求。每个供应商都提供加速迁移的工具。例如，亚马逊云科技提供大量论文来详细介绍如何从各种本地数据库迁移到亚马逊云科技，包括针对 Oracle、MySQL 和 PostgreSQL 的过程。

尽管如此，迁移到云还是可能涉及具有破坏性且成本高昂的停机。无论是由云供应商，还是由第三方提供商提供的迁移工具，都必须能够处理数据库各个方面的问题，例如架构、用户权限、触发器和存储过程。

我们建议您在早期迁移过程中将过程记录下来，因为您在这个过程中所学到的经验，将在您进行进一步迁移时，为您和您的同事提供指导。如果情况太过糟糕，以至于您决定使用其他供应商或者您根本不打算迁移，这些日志都将能提供重要的证据来支持您的决定。

迁移数据库

许多云提供商和第三方都提供非常值得考虑的迁移服务，例如 Microsoft Azure Database Migration Service 和Amazon Database Migration Service。这些服务的宣传都是快速、平稳和简单。但是，您可能更喜欢使用自己熟悉的现有流程，例如从备份还原到新的数据库环境和复制。

如果您使用数据库迁移服务，则可以在迁移到云的过程中确保源数据库完全正常运行，并最大程度地减少依赖数据库的应用程序的停机时间。这有助于实现许多迁移到云的组织的关键目标，那就是最大限度减少停机时间。图 3-1 说明了亚马逊云科技 Database Migration Service 如何创建表、加载数据并保持表与源数据库表同步。

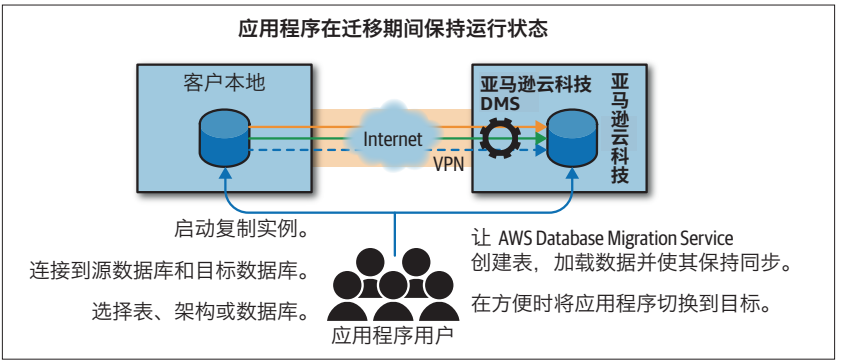

图 3-1. 在数据库迁移期间保持应用程序正常运行

迁移过程通常围绕典型的备份/还原过程进行。任务取决于您是否能够承受在迁移期间关闭数据库。如果您必须保持数据库正常运行，请按照以下基本顺序执行操作：

1. 在原始数据库运行时对其进行热备份。
2. 复制并在云中还原备份。
3. 启用复制功能，以将创建备份。

后对原始数据库所做的任何更改整合到云版本中。当然，如果您可以承受数据库停止运行，那么备份/还原会更简单一点。然后，您可以按照以下模式执行迁移：

1. 关闭原始数据库。
2. 执行备份。
3. 复制并在云中还原备份。
4. 在云中启动新版本。

但是，在准备将新的云版本投入生产之前，您可能需要花费时间来对其进行测试，以确保应用程序正常运行、安全性到位并且性能符合要求。因此，在准备信任新版本之前，您可能仍然需要重新启动原始版本，然后使用复制将新版本与原始版本同步。

迁移应用程序

迁移数据库和相关应用程序有两种基本策略：

直接迁移

将整个数据库按原样（包含其支持的所有应用程序）迁移至云中的新数据库。这是一种“换汤不换药”的应用程序迁移策略，因为您几乎没有使用原生云功能。

重新设计数据库

重新考虑您的本地数据库如何处理扩展和高可用性等管理任务，并调查云提供商在这些方面提供的产品。由于重新设计数据库的程序日益复杂，这个策略自然而然需要投入更多精力。您可能还需要对应用程序进行重新编码，以根据数据库策略的更改进行调整。

通常，客户一开始会执行“直接迁移”，然后再进行 重新设计和其他现代化改造。

迁移后检查

将数据完全复制到云数据库后，请验证目标数据库，以确保所有数据库对象都存在于云数据库中。几个测试可以确定迁移是否成功：

验证数据

此操作很简单，只需检查行数或运行校验和来确定是否存在丢失或损坏。更复杂的验证流程会考虑到异构迁移中固有的架构更改。遗憾的是，这种基本的验证常常会被忽略。

基本功能

执行端到端测试，以确保您成功执行了迁移并且系统按预期运行。测试数据库引擎和版本之间可能不同的功能的使用，例如触发器和存储过程。

性能测试

对新系统执行压力测试并对其性能执行基准测试，以了解如何缩短响应时间和增加吞吐量。

安全评估

通过执行漏洞评估和渗透测试，确保云系统的安全。

这些步骤也是向下一阶段“优化”的过渡。这个阶段紧跟在您将本地数据库和应用程序成功迁移到云之后。此阶段的关键任务是性能管理和成本优化。认真、持续地监控新系统至关重要，这样您就可以在遇到意外故障时，快速恢复到旧系统。

优化

现在，想必您的系统应该在平稳运行，并且可以庆祝团队出色完成了工作。本节总结了 DBA 和其他团队成员，在接下来需要完成的主要任务。

可用性

当然，这是您在任何情况下都必须满足的基本要求。幸运的是，迁移到云可以消除一些故障原因（前提是您正确配置故障服务重新启动功能）。此外，云供应商还提供许多工具来预测故障并向您发出故障告警。要利用这些工具，请执行本章中的规划（其中许多活动对于性能优化也很有用）：

灾难恢复 (DR) 计划

有了稳健的大型云供应商，如果当前可用区发生故障，您可以故障转移到新的可用区。但是，您需要制定计划和自动化流程，来进行故障转移。计划应该解决 RTO、RPO 和地理冗余问题。尽可能利用供应商提供的故障转移功能，而不是自己动手创建。

日志和系统监控

确定可以指示即将发生的问题和已经发生的问题的事件。您可以将这些事件整合到自动化跟踪中。跟踪应该传达足够的信息，让您了解故障的原因：例如，是源于用户操作（如重启服务），还是源于攻击或者环境中的其它变化。有些故障可被视为正常情况，并且可以通过您的自动化工具来解决；应该将这些记录下来，但不需要向管理员发出告警。

更改监控

管理员应该始终了解对数据库配置、实例大小或集群拓扑的哪些更改会影响可用性。现代开发环境使用稳健的流程，来进行更改跟踪和版本控制，这样每项更改都要经过审核过程，并且可以撤消。

系统测试

尝试确定系统中的弱点并预测故障。一些团队会执行“故障演练”，找出并消除潜在的故障源。大型站点可以承受故意关闭系统，从而观察恢复是否充分；这种测试称为混沌工程，通过 Netflix 的 Chaos Monkey 得到了普及。就像您定期执行还原以确保备份正常一样，您也应该测试恢复过程。

性能优化

性能可以从上述部分的实践中获益，尤其是监控。通过性能监控，您应该可以确定事件与数据库指标更改之间的关系，还应该能了解到预测性能趋势和实际性能趋势之间的差异。此外，还可以通过其它流程来进一步维护和改善性能。

工作负载测试

数据大小和复杂性的增加以及应用程序行为的变化都会影响性能。定期测试性能，以便在客户告诉您之前，就了解到性能下降情况；然后，您可以进行扩展

或其它更改来调整。在进行重大更改后，数据库缓存和表统计数据可能需要一段时间才能完成更新，这意味着查询计划和整体性能，可能需要一段时间才能恢复到预期水平。

分布式优化

在快速变化的应用程序环境中，开发人员比 DBA 更了解数据库查询。如果由他们负责查询测试和优化，他们可以比 DBA 更快地实现出色性能，从而让 DBA 可以负责长期的优化和规划。

了解了性能的重要性，云供应商已经插入跟踪信息并向用户公开了大量统计数据。例如，您可以使用亚马逊云科技 RDS 性能详情在控制面板上提取历史数据并查看各种图表，了解特定操作、用户或 SQL 语句对数据库施加的负载。您可以查阅控制面板，来回答一些简单的问题，例如“主机何时过载？”或者更复杂的问题，例如“什么原因使得这次查询的性能不太理想？”这样的问题。Azure 同时提供性能监控和自动优化。第三方供应商还提供工具来监控云中的性能。

适应云中的差异

最后，供应商自己将提供新的服务、新的虚拟机选项、新的硬件（例如 SSD）以及其它能为您带来巨大优势的更改。始终了解供应商可以为您做些什么，确保他们的流程已通过全面测试且标准化，这样您才能将工作委任给他们。与此同时，您也可以更新您的书面操作程序。最后，您可以向供应商的客户支持团队报告异常事件，以表明事件的紧急程度。

总结

迁移到云是一个长期过程。从小处着手，因为在这个过程中您会发现有很多东西需要学习。记日志并如实记录犯下

的所有错误和遇到的所有问题。如果旧系统中存在不佳实践或迁移期间发生重大错误，请不要感到尴尬，几乎每个组织都会遇到这些情况。记下这些问题是您可以为公司做的最好的事情。

希望您的一个或多个早期迁移顺利进行，并为向云的重大迁移做好准备。这样做可以在成本、灵活性和安全性方面获益。尤其重要的是，迁移到云将提供最新的计算环境，有助于吸引那些希望走在前沿的领先员工到您的公司。

第 4 章

总结

在 Amazon.com 推出了第一款主要的云产品后的很多年里，行业媒体一直在提系统管理员和 DBA 面临的一个问题：“使用云还是不使用云？” 不久，本地云和混合产品将纯粹的云解决方案，纳入考虑范围内，但是，选择总是更加复杂。正如本报告所介绍的，产品数量迅速成倍增加。DBA 评估数据库时必须同时考虑以下因素：

- 第三方供应商、本地或混合
- 关系型或众多非关系型之一
- 托管或自行管理
- 原生云（例如 Amazon Aurora）或跨平台（例如 MySQL）
- 是否利用性能增强功能，例如固态硬盘或缓存
- 云区域和可用区的实际位置
- 迁移便利性
- 您的员工需要和掌握的相关技能
- 供应商支持和声誉的其它方面

在查看所有选项之前，过早地限定在一个领域的选择上并非明智之举。通过接受一些额外的培训或大胆采用一种不熟悉的技术，您很可能会节省大量资金并改善客户体验。

除了列出选择数据库的基本标准，本报告还帮助您准备好应对您的职责和任务中可能会发生的变化，从而帮助您为迁移到云做好准备。通过迁移到云，您可以简化或消除某些职责和任务，但是您还需要学习新技术，并且需要开始以新方式思考目标，例如高可用性和优化。

在第一次迁移期间或通过在云中启动新项目，您会学到很多东西。每个项目都将向您阐明云数据库的格局，并为您的下一个项目提供思路。希望这份报告能为您在工作过程中，提供所需的注意事项。

关于作者

Wendy A. Neu 是亚马逊云科技专业服务首席顾问，致力于解决客户最棘手的问题，构建高质量、可扩展且架构合理的系统。她是亚马逊云科技数据库博客的定期撰稿人，并获得了亚马逊云科技、Oracle 和 Microsoft SQL Server 的认证。在加入 Amazon 之前，她曾在俄亥俄州辛辛那提市担任顾问，帮助客户将业务需求转化为可行的技术解决方案。

Vlad Vlasceanu 是亚马逊云科技首席数据库专家解决方案架构师，常驻加利福尼亚州圣莫尼卡市办事处。Vlad 帮助客户采用原生云数据库解决方案（例如，Amazon Aurora），并在亚马逊云科技上部署大规模、高性能的数据库架构。他致力于利用亚马逊云科技平台提供的最新最佳实践和功能，设计和实施可持续、经济高效且可扩展的数据库工作负载。在加入亚马逊云科技之前，Vlad 从事能源行业，负责设计和开发以消费者为中心且基于 Web 的应用程序以及数据驱动型应用程序，拥有超过 15 年的从业经历。Vlad 拥有贝勒大学的信息系统科学硕士学位。

Andy Oram 出版过许多著作，包括 O’Reilly 的 Linux 系列、具有开创性的《Peer-to-Peer》和畅销的《Beautiful Code》。Andy 还撰写了许多有关数据湖、Web 性能和开源软件等技术主题的报告。他的文章发表在《经济学家》、《美国计算机协会通讯》、《版权世界》、《信息技术与政治学杂志》、《Vanguardia Dossier》和《互联网法律与商业》上。他曾在 O’Reilly’s Open Source Convention、FISL（巴西）、FOSDEM、DebConf 和 LibrePlanet 等会议上发表过演讲。Andy 参加了美国计算机协会的政策组织 USTPC。他还为各种网站撰写关于医疗信息技术以及计算和政策问题方面的文章。

Sam R. Alapati 是位于德克萨斯州韦斯特莱克的 Solera Holdings 的数据管理员。他是大数据和 Hadoop 团队的成员。Sam 是由 Oracle Technology Network 授予认可的 Oracle ACE。他是《Modern Linux Administration》（O'Reilly，2018 年）的作者，并且还撰写了 20 多本有关数据库和系统管理的书籍。Sam 拥有与三大云提供商（亚马逊云科技、Microsoft Azure 和 Google Cloud Platform）合作的经验。