

WILEY



“十二五”
国家重点图书出版规划项目

国际先进工业技术译丛

RELIABILITY AND AVAILABILITY OF CLOUD COMPUTING

云计算实战 ——可靠性与可用性设计

【美】Eric Bauer Randee Adams 著
高巍 李洁 魏凯 郭亮 段世惠 唐浩 等 译



人民邮电出版社
POSTS & TELECOM PRESS

WILEY



国家重

国际先进工业技术译丛

RELIABILITY AND AVAILABILITY OF CLOUD COMPUTING

云计算实战 ——可靠性与可用性设计

【美】Eric Bauer Randee Adams 著
高巍 李洁 魏凯 郭亮 段世惠 唐浩 等 译

人民邮电出版社
北京

图书在版编目 (C I P) 数据

云计算实战：可靠性与可用性设计 / (美) 鲍尔
(Bauer, E.), (美) 亚当斯 (Adams, R.) 著 ; 高巍等译
-- 北京 : 人民邮电出版社, 2014. 1
(国际先进工业技术译丛)
ISBN 978-7-115-32932-5

I. ①云… II. ①鲍… ②亚… ③高… III. ①计算机
网络 IV. ①TP393

中国版本图书馆CIP数据核字 (2013) 第195891号

版 权 声 明

Eric Bauer, Randee Adams.

Reliability and Availability of Cloud Computing.

Copyright © 2012 by John Wiley & Sons Ltd.

All rights reserved. This translation published under license.

Authorized translation from the English language edition published by Wiley Publishing, Inc..

本书中文简体字版由 John Wiley & Sons Ltd 公司授权人民邮电出版社出版, 专有版权属于人民邮电出版社。

北京市版权局著作权合同登记号 图字: 01-2013-3654号



-
- ◆ 著 [美] Eric Bauer Randee Adams
 - 译 高巍 李洁 魏凯 郭亮 段世惠 唐浩 等
 - 责任编辑 刘洋
 - 责任印制 彭志环 焦志炜
 - ◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路11号
 - 邮编 100164 电子邮件 315@ptpress.com.cn
 - 网址 <http://www.ptpress.com.cn>
 - 三河市海波印务有限公司印刷
 - ◆ 开本: 787×1092 1/16
 - 印张: 16.5
 - 字数: 330千字 2014年1月第1版
 - 印张: 1-3 000册 2014年1月河北第1次印刷

著作权合同登记号 图字: 01-2013-3654号

定价: 59.00元

读者服务热线: (010)81055410 印装质量热线: (010)81055316
反盗版热线: (010)81055315

内容提要

本书在明确了云计算、虚拟化、服务可靠性以及服务可用性的关键术语和概念的基础上，通过对云计算服务可靠性及可用性的分析，结合云计算所涉及的虚拟化、分布式部署等技术特点，全面、深入、系统地探讨了云计算系统及服务可靠性的设计、计算、评估、优化方法。

本书具有较强专业性和技术性，同时还具有较强的可操作性，对云计算专业从业人员和科研工作者来说，是一本不可多得的学习与参考书籍。本书既适用于云计算系统的专业技术研发和工程实施人员，也适合于高等院校通信、电子、计算机等专业的师生，同时也可云服务用户提供有益的参考。

作者简介

Eric Bauer 是阿尔卡特朗讯的软件、解决方案和服务事业部的可靠性工程经理。他目前专门研究阿尔卡特朗讯的云计算相关产品、IP 多媒体子系统（IMS）以及其他解决方案的可靠性和可用性。在专注可靠性工程主题研究之前，**Bauer** 先生有 20 年的设计和开发嵌入式固件、网络操作系统、IP 电话互联网网络平台和光传输系统经验。他获得了超过 12 项美国专利，独立撰写了《可靠性设计：信息和基于计算机的系统》；与人合作撰写了《超越冗余：地理冗余如何改善计算机系统的服务可用性和可靠性》、《实用系统的可靠性》，并在贝尔实验室的技术期刊上发表了多篇论文。**Bauer** 先生先后获得了康奈尔大学的电气工程学士学位、普渡大学的电气工程硕士学位。他目前住在新泽西州 Freehold。

Randee Adams 是阿尔卡特朗讯的软件、解决方案和服务事业部的技术人员。她在 1979 年作为一名新数字 5ESS 交换机程序员加入贝尔实验室。**Adams** 女士曾在公司参与过许多项目（如软件开发、故障单管理、负载管理研究、软件交付、系统工程、软件体系结构、软件设计、工具开发等），涉及过许多领域（如数据库管理，最新变更/验证，公共信道信令，运营、监督、管理，可靠性和安全性）。目前，她正在专注阿尔卡特朗讯产品的可靠性。她在各种关于可靠性的内部论坛上做过演讲。**Adams** 女士与人合著了《超越冗余：地理冗余如何改善计算机系统的服务可用性和可靠性》。**Adams** 女士拥有亚利桑那大学的学士学位和伊利诺伊技术学院的计算机科学硕士学位。她住在伊利诺伊州 Naperville。

译者序

人类社会正在经历着信息爆炸的革命，互联网的出现更是加速了信息的产生和流动。近年来互联网流量一直在以每年 100%左右的速度增长，现在全世界一年所产生的信息量就超过了过去 5000 年产生的信息量总和。云计算正是在这样一个社会发展环境下应运而生的。

早在 20 世纪 60 年代，著名的计算机科学家 John McCarthy 就曾经预言：“未来计算机的发展，可能使计算能力变成像电话网一样构成一种公众服务设施……提供计算能力可能成为一个新兴的重要行业”。就在两三年之前，一些人还在抱着犹豫怀疑的态度看待云计算，认为云计算不过是 IT 领域又一次话题性炒作，会在不远的将来偃旗息鼓。但今天，云计算已经成为现实生活中的一部分，云计算服务已经形成了一个实实在在的产业。

目前，公共云服务已经成为互联网的重要基础性服务，并被认为是 IT 产业未来发展的重心。据统计，2013 年 5 月托管在亚马逊 AWS 云服务上的网站数量已经达到 1160 万个，占全球网站总数的 1.8%，北美大量的中小创新企业在亚马逊、Rackspace 等云服务企业的支持下，以低廉的成本构建了其 IT 及业务系统。同时，云计算服务已经成为全球 ICT 领域最具活力的增长点之一，并且正在逐步渗透到传统行业 IT 市场中，金融业、制造业、通信业、政府以及零售业已经成为国际上云服务市场中占比最大的 5 个领域。根据 Gartner 的预测，2013 年全球云计算服务市场规模达到 2460 亿美元，其中一般意义的 IaaS、PaaS、SaaS 服务市场将达到约 1290 亿美元，行业平均增长率高达 20%，远高于 ICT 产业 6%~7%的整体增长率水平。

我国的云计算市场在近几年也取得了长足的发展。据预测，2013 年我国云计算服务市场规模有望达到 63 亿元人民币，较 2012 年增长 80%以上，远高于同期国际增速水平。我国云服务的细分市场已形成了初步格局，互联网公司、电信运营商、IDC 服务商、IT 制造商、软件公司等都逐步加大在云计算方面的投入，一些公司已经形成了自己特色的云计算服务和产品。

但在云计算快速发展的同时，国内外服务商都在不同程度上发生过业务中断、数据泄露等安全事件，不断出现的云计算安全事件也在提醒着人们，要提高对云计算系统可靠性、安全性的关注。在这种情况下，为使国内相关人士全面了解国际上对云计算系统在可靠性和可用性设计方面的最新研究成果，我们选择了《云计算实战——可靠性与可用性设计》

一书介绍给国内的读者。本书为 Eric Bauer 先生与 Randee Adams 女士联合编著，他们均具有多年可靠性工程经验。

本书通过对云计算服务可靠性及可用性的分析，结合云计算所涉及的虚拟化、分布式部署等技术特点，全面、深入、系统地探讨了云计算系统及服务可靠性的设计、计算、评估、优化方法，既适用于云计算系统的专业技术研发和工程实施人员，也适合于高等院校通信、电子、计算机等专业的师生，同时也可作为云服务用户提供有益的参考，是一本不可多得的学习与参考书籍。

本书的翻译工作得到了工业和信息化部电信研究院通信标准研究所副所长石友康的指导。云计算发展与政策论坛秘书长、TGG（The Green Grid，绿色网格）中国委员会主席、工业和信息化部电信研究院互联网研究领域主席何宝宏和 DMTF（Distributed Management Task Force，分布式管理任务组）中国分会副理事长高巍通读了本书译稿的核心章节，并提出了非常中肯的意见。

本书第 1 章和第 14 章由栗蔚翻译，第 2 章和第 8 章由高巍翻译，第 3 章由曹峰翻译，第 4 章由郭亮翻译，第 5 章由魏凯翻译，第 6 章由田辉翻译，第 7 章由段世惠翻译，第 9 章由徐贵宝翻译，第 10 章由唐浩翻译，第 11 章由张恒升翻译，第 12 章由李洁翻译，第 13 章由陈凯翻译，其余部分由徐贵宝翻译。

另外，张德华、卢秋波、牛晓玲、陈默、刘媛对第 1~3 章进行了审校，曹蓟光、张杰、马军锋、赵越、叶永学对第 4~6 章进行了审校，赵锋、朱刚、董晓飞、葛裴、陈屹力对第 7~9 章进行了审校，刘述、朱晓云、宋菲、程楠、季磊对第 10~12 章进行了审校，马科、吴英桦、毕立波、张宇华、唐锡京对第 13~14 章进行了审校，孙明俊、陈凯、聂秀英、王子豪、刘佳良对其余部分进行了审校。全书由徐贵宝进行统稿并审校。

在翻译过程中，译者在追求技术术语准确性和理论正确性的同时，尽量做到通俗易懂。由于各章翻译人员不同，为保证用词习惯和语言风格等各方面的前后一致性，我们还专门由一人从头至尾进行通读修改。但由于时间仓促，加之译者水平有限，译稿还存在不尽如人意之处。敬请广大读者在阅读本书的过程中，提出宝贵意见和建议。本书责任编辑电子邮箱：liuyang@ptpress.com.cn。

在此，衷心感谢于涛、韩宝鱼、冉海芳、王敬敬等在本书翻译和出版过程中提供的大力协助和支持。还要衷心感谢人民邮电出版社高级策划编辑刘洋先生为本书出版所做的大量联系和沟通工作，没有他的努力，本译作也不可能顺利与读者见面。

本书翻译工作由工业和信息化部电信研究院承担的国家发展和改革委员会、财政部、工业和信息化部“云计算示范工程”项目课题“云计算公共技术服务平台”提供资金资助。

译 者

2013 年 9 月

前 言

云计算是一种为终端用户提供信息服务的新模式，这种模式相比传统的 IS/IT 部署模式有明显的优势，更为经济且能在更短时间内将服务推向市场。云计算由一组基本特征定义：按需自服务、宽带网络接入、资源池、快速弹性以及可度量服务。云服务商可以采用多种模式提供服务，包括基础设施即服务、平台即服务以及软件即服务；同时云部署选项包括私有云、社区云、公有云以及混合云。终端用户很自然地希望，通过云计算所提供的服务，至少具有同传统的服务实现方式一样的可靠性和可用性。基于云部署的应用，要想获得与传统部署相同的可靠性和可用性，需要面临一定的风险。本书在对此风险进行分析的同时，也分析了在云部署中存在的改进服务可靠性和可用性的机会。我们认为，服务可靠性和可用性所面临的风险来自于云计算的基本定义，也即其基本特征，而不是其他特定的虚拟化监管软件或云服务提供过程。因此，所有云服务提供过程和应用部署都应该基于这种更高层次分析的认知和建议。本书也在体系结构、测试以及工程实践方面提供了一些建议，以保障基于云所部署的应用能在服务的可靠性和可用性方面满足用户需求。

在传统的部署场景中，企业的应用都是直接安装于裸硬件之上的。利用虚拟化技术，企业能够将这些已有的应用迁移到具有硬件独立和服务器集中等更先进特征的场景中。采用虚拟化技术是云计算的一个共有特征，它使得云服务商可以更好地通过多个云用户来管理资源池的使用。本书也考虑了这种演化路径所伴随的可靠性和可用性风险，用于指导企业规划其应用向虚拟化演进路线，并在发布多个应用版本时实现全部云计算化。

适用读者

本书适用于信息技术系统和方案架构师、开发人员以及工程人员，也适用于技术销售、产品管理和质量管理专家。

本书结构

本书由 3 部分组成：第 1 部分“基本概念篇”，第 2 部分“分析篇”，第 3 部分“建议

篇”。第1部分“基本概念篇”中定义了云计算、虚拟化、服务可靠性以及服务可用性的关键术语和概念。第1部分包括3章。

第1章：云计算。本书采用由美国国家标准技术研究院所定义的云术语和分类。本章定义了云计算，并概述了云计算的本质和公有特征。同时也概述了云计算的标准服务、部署模式以及主要的云相关参与者的角色。对云计算的主要优势和风险进行了总结。

第2章：虚拟化。虚拟化是云计算的通用特征。本章概述了虚拟化技术，提供了用于虚拟化分析的架构模式，对虚拟化的应用与本地应用进行了分析和比较。

第3章：服务可靠性和服务可用性。本章定义了服务可靠性和可用性概念，概述了在传统部署中如何度量这些指标，以及如何将这些指标应用于基于虚拟化和云的部署环境中。由于电信行业有非常精确的量化标准指标来度量服务可靠性和可用性，本章介绍了一些电信行业的概念和术语，并将其应用于本书第2部分“分析篇”和第3部分“建议篇”之中。

第2部分“分析篇”在第1部分所介绍内容的基础上，系统地分析了云计算和虚拟化技术本质和公有特征及其导致的应用部署服务可靠性和可用性所面临的固有风险。

第4章：云计算的可靠性和可用性分析。本章分析了由云计算的本质和公有特征、业务模式以及云计算部署模式所导致的服务可靠性和可用性风险，会涉及服务迁移活动、弹性以及服务组合。在第2部分的后面几个章节中，将对这些已经明确出来的具体风险进行详细分析。

第5章：虚拟化的可靠性分析。本章采用标准的可靠性分析方法，分析了全虚拟化、OS虚拟化、半虚拟化、服务器虚拟化和共存。本章同时也对虚拟化和云计算的软件可靠性风险进行了分析。

第6章：硬件可靠性、虚拟化和服务可用性。本章考虑了当应用迁移至虚拟化和云化硬件平台时，硬件可靠性风险和责任是如何转移的，以及如何确定由硬件导致的服务中断时间。

第7章：容量和弹性。快速弹性作为云计算的本质特征，将把云用户从提前在几周或几个月前就要锁定资源的商业风险中解脱出来；然而快速弹性也引入了新的服务质量、可靠性和可用性风险，需要仔细应对。

第8章：服务编排分析。采用服务编排将使得IT服务管理的许多方面可以自动化，尤其是和系统容量管理相关的一些操作。本章回顾了云计算环境中基于策略的管理，并分析了随之而来的服务可靠性和可用性风险。

第9章：地理分布、地理冗余和灾难恢复。将应用实例在地理上分布式部署是云计算和灾难恢复最佳实践的共有特征。本章对在云部署应用中地理冗余的服务可用性进行了分析。

第3部分“建议篇”分析了如何利用技术，使部署在云上的应用服务可靠性和可用性

最大化；同时也分析了如何利用可靠性设计，确保虚拟化应用和基于云的解决方案能满足或超过传统部署方式的服务可靠性和可用性。

第 10 章：应用、解决方案和责任认定。本章研究了虚拟化应用如何适应业务解决方案，也解释了当应用迁移至云时应用服务中断的预算如何变化。本章也提出了 4 个可用于服务可用性的测量点，讨论了在各测量点进行测量将如何影响可计量性。

第 11 章：构建可靠系统的建议。本章论述了通过虚拟化和云部署来使服务可靠性和可用性最大化的体系架构与技术，通过一个简单案例研究说明了关键的体系结构点。

第 12 章：虚拟应用的可靠性设计。本章论述了如何进行虚拟化应用可靠性设计，以区别于传统应用。

第 13 章：云计算解决方案的可靠性设计。本章论述了如何进行云部署方案的可靠性设计，以区别于传统应用的解决方案。

第 14 章：总结。本章总结了在可靠性和可用性方面具有可执行性的分析、内幕和建议，以确保基于云的解决方案能够满足或超过传统部署的性能。

致谢

本书作者得到了阿尔卡特朗讯的工程师们的鼎力协助，特别是 Mark Clougherty、Herbert Ristock、Shawa Tam、Rich Sohn、Bernard Bretherton、John Haller、Dan Johnson、Srujal Shah、Alan McBride、Lyle Kipp 和 Ted East，他们具有广博的知识和深刻的洞察力。Joe Tieu、Bill Baker 和 Thomas Voith 仔细审阅了本书的早期书稿，并热诚地反馈了意见。Abhaya Asthana、Kasper Reinink、Roger Maitland 和 Mark Cameron 提供了很有价值的建议。本书源于 Gary McElvany 最初提出的有关体系结构的问题。如果没有 Tina Hinch、Werner Heissenhuber、Annie Lequesne、Vickie Owens-Rinn 和 Dor Skuler 强有力的管理，本书也不可能完成。

云计算是一项激动人心并且正在演进中的技术，还有许多方面需要去探索。如果读者对本书内容有评论或指正，或者对本书以后再版内容有什么建议，可以给作者发邮件，邮件地址是 Eric.Bauer@Alcatel-Lucent.com，Randee.Adams@Alcatel-Lucent.com 或者 pressbooks@ieee.org。

Eric Bauer
Randee Adams

目 录

第1部分 基本概念篇

第1章 云计算	3	2.3.2 部分虚拟化	17
1.1 云的基本特征	3	2.3.3 操作系统虚拟化	17
1.1.1 按需自服务	4	2.3.4 讨论	17
1.1.2 广泛的网络访问	4	2.4 虚拟机的生命周期	18
1.1.3 资源共享	4	2.4.1 虚拟机快照	20
1.1.4 快速弹性	5	2.4.2 克隆虚拟机	20
1.1.5 服务可度量	5	2.4.3 高可用性机制	21
1.2 共有的云特征	6	2.5 虚拟化的可靠性和可用性	
1.3 云计算到底是什么	6	风险	21
1.3.1 什么是数据中心	7	第3章 服务可靠性和	
1.3.2 云计算和传统数据中心		服务可用性	22
有何不同	8	3.1 错误和失效	22
1.4 服务模式	8	3.2 八成分框架	24
1.5 云部署模式	9	3.3 服务可用性	26
1.6 云计算中的角色	10	3.3.1 服务可用性指标	26
1.7 云计算的优势	12	3.3.2 MTBF 和 MTTR	27
1.8 云计算的风险	12	3.3.3 服务和网元影响中断	28
第2章 虚拟化	13	3.3.4 局部中断	29
2.1 背景	13	3.3.5 可用性等级	30
2.2 什么是虚拟化	14	3.3.6 中断归因	31
2.2.1 管理程序的类型	14	3.3.7 计划内宕机时间	32
2.2.2 虚拟和仿真	15	3.4 服务可靠性	33
2.3 服务器虚拟化	15	3.4.1 服务可靠性指标	33
2.3.1 完全虚拟化	16	3.4.2 缺陷事务	34

3.5 服务时延.....	35	3.8.1 控制和数据平面.....	45
3.6 冗余和高可用性	38	3.8.2 服务质量指标.....	45
3.6.1 冗余.....	39	3.8.3 等时性数据.....	46
3.6.2 高可用性	41	3.8.4 时延期望.....	47
3.7 高可用性和灾难恢复	43	3.8.5 流媒体质量损伤.....	47
3.8 流媒体服务.....	45	3.9 云计算的可靠性和可用性风险	48

第2部分 分析篇

第4章 云计算的可靠性和

可用性分析

4.1 服务可靠性和可用性的期望	51
4.2 云基本特征的风险	52
4.2.1 按需自助服务	52
4.2.2 宽带网络接入	52
4.2.3 资源池.....	52
4.2.4 快速弹性	53
4.2.5 测量服务	54
4.3 常见云特征的影响	55
4.3.1 虚拟化.....	55
4.3.2 地理分布	55
4.3.3 弹性计算	55
4.3.4 高级安全性	55
4.3.5 大规模.....	56
4.3.6 同质化.....	56
4.4 服务模式的风险	56
4.4.1 传统的责任归因	56
4.4.2 云应用的责任归因	57
4.5 IT 服务管理和可用性的风险.....	58
4.5.1 ITIL 概述	58
4.5.2 服务策略	59
4.5.3 服务设计	59
4.5.4 服务转移	60
4.5.5 服务运营	61
4.5.6 持续服务改进	61
4.5.7 IT 服务管理概述.....	62
4.5.8 服务编排的风险	62

4.5.9 IT 服务管理风险.....	63
----------------------	----

4.6 过程区域中断的风险.....	63
4.7 故障检测的注意事项.....	65
4.7.1 硬件故障.....	67
4.7.2 编程错误.....	67
4.7.3 数据的不一致性和错误.....	67
4.7.4 冗余错误.....	67
4.7.5 系统电源故障.....	67
4.7.6 网络错误.....	67
4.7.7 应用协议错误.....	68
4.8 部署模型的风险.....	68
4.9 IaaS 数据中心的期望值	68

第5章 虚拟化的可靠性分析

5.1 可靠性分析技术.....	70
5.1.1 可靠性框图.....	70
5.1.2 单点故障分析.....	71
5.1.3 故障模式影响分析 (FMEA)	71
5.2 虚拟化技术的可靠性分析	73
5.2.1 全虚拟化分析.....	73
5.2.2 虚拟操作系统分析	74
5.2.3 半虚拟化分析.....	74
5.2.4 VM 共存分析	75
5.2.5 讨论.....	77
5.3 软件故障率分析.....	77
5.3.1 虚拟化和软件故障率	77
5.3.2 虚拟机管理程序故障率.....	78

5.3.3 虚拟化和云的其他 软件风险	78	6.8.1 虚拟网络接口卡	97
5.4 恢复模型	78	6.8.2 虚拟局域网	98
5.4.1 传统的恢复选项	78	6.8.3 虚拟 IP 地址	99
5.4.2 虚拟化恢复选项	79	6.8.4 虚拟专用网络	99
5.4.3 讨论	82	6.9 虚拟硬件的 MTTR	99
5.5 应用架构策略	83	6.10 讨论	100
5.5.1 按需单用户模式	83	第 7 章 容量和弹性	101
5.5.2 单用户守护进程模式	84	7.1 系统负载基础知识	101
5.5.3 多用户服务器模式	84	7.1.1 特别事件的考虑	102
5.5.4 服务器整合模式	84	7.1.2 Slashdot 效应	103
5.6 虚拟化恢复方式的可用性建模	85	7.2 过载、服务的可靠性和 可用性	103
5.6.1 虚拟化单工架构的 可用性	85	7.3 传统的容量规划	104
5.6.2 虚拟化冗余架构的 可用性	86	7.4 云和容量	105
5.6.3 关键故障率	86	7.4.1 标称的云存储模型	106
5.6.4 故障覆盖率	87	7.4.2 弹性的期望	108
5.6.5 故障检测延迟	87	7.5 管理联机容量	110
5.6.6 切换延迟	87	7.6 容量相关的服务风险	112
5.6.7 切换成功率	87	7.6.1 弹性和弹性故障	112
5.6.8 建模与“快速故障”	88	7.6.2 部分容量故障	114
5.6.9 本地和虚拟化部署的 比较	88	7.6.3 服务延迟风险	115
第 6 章 硬件可靠性、虚拟化和 服务可用性	89	7.6.4 容量损伤和服务 可靠性	117
6.1 对硬件故障停机时间的期望	89	7.7 容量管理风险	117
6.2 硬件故障	89	7.7.1 脆弱的应用架构	118
6.3 硬件故障率	91	7.7.2 故障或监测数据不足	118
6.4 硬件故障检测	92	7.7.3 错误的容量决策	119
6.5 硬件故障遏制	93	7.7.4 不可靠的容量扩张	119
6.6 硬件故障的缓解	93	7.7.5 不可靠的容量缩减	119
6.7 通过虚拟化缓解硬件故障	95	7.7.6 转换速度不足	119
6.7.1 虚拟 CPU	95	7.7.7 缓慢的容量管理 决策	120
6.7.2 虚拟内存	96	7.7.8 资源库存耗竭	120
6.7.3 虚拟存储	96	7.7.9 云跳转失败	120
6.8 虚拟化网络	97	7.7.10 政策限制	120
		7.8 安全性和服务可用性	121

- 7.8.1 服务可用性的安全风险..... 121
- 7.8.2 拒绝服务攻击 122
- 7.8.3 DoS 攻击的防御 123
- 7.8.4 量化安全攻击对服务可用性的影响 123
- 7.8.5 建议..... 124
- 7.9 弹性扩张和收缩的架构 124
- 第 8 章 服务编排分析..... 126**
 - 8.1 服务编排定义..... 126
 - 8.2 基于策略的管理 127
 - 8.2.1 SLR 的作用 128
 - 8.2.2 服务可靠性和可用性的测量..... 128
 - 8.3 云管理..... 129
 - 8.3.1 云管理中快速弹性的作用..... 129
 - 8.3.2 云管理中云突发的作用..... 129
 - 8.4 服务编排在风险缓解中的作用 130

- 8.4.1 延迟..... 130
- 8.4.2 可靠性..... 130
- 8.4.3 监督管理..... 131
- 8.4.4 安全..... 131
- 8.5 小结..... 132

- 第 9 章 地理分布、地理冗余和灾难恢复 133**
 - 9.1 地理分布 VS 地理冗余 133
 - 9.2 传统的灾难恢复..... 134
 - 9.3 虚拟化和灾难恢复..... 135
 - 9.4 云计算和灾难恢复..... 136
 - 9.5 地理冗余恢复模型..... 137
 - 9.6 云和传统地理冗余的附加益处..... 138
 - 9.6.1 减少预期的宕机时间 ... 138
 - 9.6.2 缓解灾难性的网络元件故障..... 138
 - 9.6.3 减少未发现的和双工元件故障..... 139
 - 9.7 讨论..... 139

第 3 部分 建议篇

- 第 10 章 应用、解决方案和责任认定 143**
 - 10.1 应用配置场景 143
 - 10.2 应用的部署方案 144
 - 10.3 系统宕机时间预期 145
 - 10.3.1 传统的系统宕机时间预期..... 145
 - 10.3.2 虚拟化应用宕机时间预期..... 146
 - 10.3.3 IaaS 的硬件宕机时间预期..... 147
 - 10.3.4 云应用的宕机时间预算..... 149
 - 10.3.5 总结..... 151

- 10.4 最终的端到端解决方案的注意事项 151
 - 10.4.1 什么是端到端解决方案..... 152
 - 10.4.2 云消费者的特定架构... 153
 - 10.4.3 数据中心冗余..... 154
- 10.5 服务不佳的原因..... 155
- 10.6 解决方案服务的测量..... 157
- 10.7 管理可靠性和云计算服务 ... 160
- 第 11 章 构建可靠系统的建议... 161**
 - 11.1 虚拟化和云计算的架构..... 161
 - 11.1.1 软件映射到 VM..... 161
 - 11.1.2 服务负载分配..... 162

11.1.3 数据管理.....	162	12.2.3 多租户应用场景	191
11.1.4 软件冗余和高可用性 机制.....	163	12.2.4 虚拟化设备应用方案...	191
11.1.5 快速弹性.....	165	12.2.5 云部署应用场景	191
11.1.6 过载控制.....	165	12.3 可靠性要求.....	191
11.1.7 共享.....	165	12.3.1 通用可用性要求	192
11.1.8 多租户.....	166	12.3.2 服务的可靠性和延迟 要求.....	192
11.1.9 同步应用.....	166	12.3.3 过载要求.....	193
11.2 灾难恢复.....	167	12.3.4 在线容量增长和收缩...	194
11.3 IT 服务管理的注意事项.....	167	12.3.5 (虚拟化) 实时迁移 要求.....	195
11.3.1 软件升级和打补丁.....	167	12.3.6 系统转变活动的要求...	196
11.3.2 服务迁移活动的影响 分析.....	168	12.3.7 地理冗余和服务持续性 要求.....	196
11.3.3 通过 VM 迁移缓解服务 迁移活动的影响	168	12.4 可靠性定性分析.....	197
11.3.4 服务迁移活动的测试...	170	12.4.1 虚拟化应用的 SPOF 分析.....	197
11.3.5 流程错误最小化.....	170	12.4.2 虚拟化应用的故障模式 影响分析.....	198
11.3.6 服务编排的注意事项...	172	12.4.3 容量的增长和收缩 分析.....	198
11.4 许多分布式云 VS 少量 巨型云	172	12.5 可靠性定量预算与建模	199
11.5 硬件归因宕机时间最小化.....	173	12.5.1 可用性(宕机时间) 建模.....	199
11.6 架构优化.....	178	12.5.2 整体宕机时间预算及 目标.....	200
11.6.1 可靠性和可用性标准...	178	12.5.3 管理维护预算分配	200
11.6.2 可访问性优化.....	179	12.6 健壮性测试.....	200
11.6.3 高可用性、持续性、可靠 性和质量优化	181	12.6.1 基准健壮性测试	201
11.6.4 灾难恢复优化.....	181	12.6.2 高级主题: 虚拟化能更好 地测试健壮性.....	204
11.6.5 操作注意事项.....	181	12.7 稳定性测试.....	205
11.6.6 案例分析.....	182	12.8 实际性能分析.....	206
11.6.7 理论最优的应用架构...	186	12.9 可靠性路线图.....	207
第 12 章 虚拟化应用的可靠性 设计	188	12.10 硬件可靠性.....	208
12.1 可靠性设计.....	188	第 13 章 云计算解决方案的 可靠性设计	209
12.2 调整 DR 以适应虚拟化 应用	189	13.1 解决方案的可靠性设计	209
12.2.1 硬件独立性应用场景 ...	189		
12.2.2 服务器整合应用场景 ...	190		

- 13.2 解决方案范围和期望 210
- 13.3 可靠性需求 212
 - 13.3.1 解决方案的可用性需求 212
 - 13.3.2 解决方案的可靠性需求 212
 - 13.3.3 灾难恢复需求 213
 - 13.3.4 弹性需求 214
 - 13.3.5 明确的配置参数 214
- 13.4 解决方案建模与分析 215
 - 13.4.1 云数据中心部署的可靠性框图 215
 - 13.4.2 解决方案故障模式的影响分析 215
 - 13.4.3 解决方案服务转变活动的影响分析 216
 - 13.4.4 云数据中心的服务可用性（MP 2）分析 216
 - 13.4.5 聚合服务可用性（MP 3）建模 216
 - 13.4.6 恢复点目标分析 219
- 13.5 组件可靠性检查 219
- 13.6 解决方案的测试和验证 219
 - 13.6.1 健壮性测试 220
 - 13.6.2 服务的可靠性测试 220
 - 13.6.3 地理冗余测试 220
 - 13.6.4 弹性与调度测试 221
 - 13.6.5 稳定性测试 221
 - 13.6.6 在用服务测试 221
- 13.7 实际性能的跟踪和分析 222

- 13.7.1 云服务的测量 223
- 13.7.2 解决方案的可靠性路线图 224
- 13.8 解决方案可靠性的其他主题 225
 - 13.8.1 服务等级协议 225
 - 13.8.2 云服务提供商的选择 226
 - 13.8.3 书面的可靠性计划 226

第 14 章 总结 228

- 14.1 服务可靠性和服务可用性 229
- 14.2 故障问责和云计算 230
- 14.3 服务宕机时间因素 232
- 14.4 服务可用性测量点 234
- 14.5 云容量和弹性的考虑 235
- 14.6 最大化服务可用性 236
 - 14.6.1 降低产品归因的宕机时间 236
 - 14.6.2 降低数据中心归因的宕机时间 236
 - 14.6.3 降低 IT 服务管理的宕机时间 236
 - 14.6.4 降低灾难恢复的宕机时间 237
 - 14.6.5 优化云服务可用性 237
- 14.7 可靠性努力 238
- 14.8 结束语 239

缩略语 240

参考文献 244

第 1 部分

基本概念篇