



Jaumir Valença da Silveira Junior

Dealing with Device Data Overflow in the Cloud

Dissertação de Mestrado

Dissertation presented to the Programa de Pós-Graduação em Informática of the Departamento de Informática, PUC-Rio as partial fulfillment of the requirements for the degree of Mestre em Informática.

Advisor: Karin Koogan Breitman

Rio de Janeiro

March 2012



Jaumir Valença da Silveira Junior

Dealing with Device Data Overflow in the Cloud

Dissertation presented to the Postgraduate Program in Informatics of the Departamento de Informática, PUC-Rio as partial fulfillment of the requirements for the degree of Mestre em Informática. Approved by the following commission:

Prof. Karin Koogan Breitman

Advisor

Departamento de Informática - PUC-Rio

Prof. Jose Viterbo Filho

Universidade Federal Fluminense

Carolina Howard Felicissimo

Schlumberger

Prof. Simone Diniz Junqueira Barbosa

Departamento de Informática - PUC - Rio

Prof. José Eugenio Leal

Coordenador Setorial do Centro Técnico Científico - PUC-Rio

Rio de Janeiro, March 29th 2012

All rights reserved.

Jaumir Valença da Silveira Junior

BSc. in Informatics at Universidade Federal do Rio de Janeiro - UFRJ 1994.

Ficha Catalográfica

Da Silveira Junior, Jaumir Valença

Dealing with device data overflow in the cloud /
Jaumir Valença da Silveira Junior ; advisor: Karin Koogan
Breitman. – 2012.

64 f : il. (color.) ; 30 cm

Dissertação (mestrado)–Pontifícia Universidade
Católica do Rio de Janeiro, Departamento de Informática,
2012.

Inclui bibliografia

1. Informática – Teses. 2. Computação em
nuvem. 3. Banco de dados. 4. Engenharia de software. I.
Breitman, Karin Koogan. II. Pontifícia Universidade
Católica do Rio de Janeiro. Departamento de Informática.
III. Título.

CDD: 004

Acknowledgements

To my parents, for their everlasting unconditional support.

To my advisor, Prof. Karin Breitman, for the steady support and encouragement along these two years of work .

Abstract

Da Silveira Junior, Jaumir Valença; Breitman, Karin Koogan (advisor). **Dealing with Device Data Overflow in the Cloud**. Rio de Janeiro, 2012. 64p. MSc. Dissertation - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Cloud computing is rapidly becoming an important platform for research in Software Engineering. Despite the huge literature on commercial Cloud environments, there is, however, little research on how to capture, model, design and implement new software applications that can make intelligent use of the Cloud. In this paper we propose a new abstraction that explores a fundamental aspect of Cloud systems – data elasticity. The Container Database (CDB) abstraction provides a Cloud-based solution for scenarios where device local storage is not sufficient for manipulating data. To demonstrate the viability of the proposed approach we present an implementation of the CDB abstraction as an Object-Oriented API designed to work on Windows 7 and Windows Mobile Phone 7 Operation Systems.

Keywords

Cloud Computing; Databases; Software Engineering.

Resumo

Da Silveira Junior, Jaumir Valença; Breitman, Karin Koogan. **Utilizando a Nuvem para lidar com a sobrecarga de dados em dispositivos fixos e móveis.** Rio de Janeiro, 2012. 64p. Dissertação de Mestrado - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

A Computação em Nuvem torna-se a cada dia mais importante como plataforma para pesquisa na Engenharia de Software. Apesar da vasta literatura disponível para uso da Nuvem em ambientes comerciais, ainda há pouca pesquisa feita para que se modelem, desenhem e implementem novos aplicativos que façam uso inteligente da Nuvem. Nesta dissertação é proposta uma abstração que explora um aspecto fundamental dos sistemas em Nuvem - a elasticidade de dados. A abstração Container Database (CDB) provê uma solução baseada em Nuvem para a falta de espaço para armazenamento local de dados em dispositivos eletrônicos. Para demonstrar a viabilidade desta abordagem, é apresentada uma implementação da abstração CDB como uma API que funciona nos sistemas operacionais Windows 7 e Windows Mobile Phone 7.

Palavras-chave

Computação em Nuvem; Bancos de Dados; Engenharia de Software.

Index

INTRODUCTION	11
1.1. Context	11
1.2. Problem Scenarios	15
1.3. Work Contribution	16
2 . CURRENT TECHNOLOGIES	17
2.1. Cloud Services	17
2.2. The Windows NT File System and Memory-mapped files	19
2.3. System.Generics.Collections.Dictionary Class	20
2.4. LINQ to XML	21
3 .THE CONTAINER DATABASE ABSTRACTION	22
3.1. Container Database High-level Architecture	22
3.2. Basic Operations and Mechanisms	24
3.3. CDB Operations and the ACID Transaction Properties	28
4 . CONTAINER DATABASE OBJECT ORIENTED DESIGN	32
4.1. Container Database Key Abstractions	32
4.2. Object-Oriented Public Interface	39
4.3. Integration into Client Code	41
5 . API IMPLEMENTATION ON DESKTOP SYSTEMS	42
5.1. Types, persistence and access methods	42
5.2. Consistency and Rollback	45
5.3. Database limitations	46
6 IMPLEMENTATION FOR MOBILE SYSTEMS	48
6.1. Windows Phone Execution Model and Data Persistence	48
6.2. Windows Phone Thread Model – need for Asynchrony	49

6.3. Types, persistence and access methods	53
6.4. Database limitations	55
7 Conclusion	56
7.1. Lessons Learned	56
7.2. Limitations	57
7.3. Future Work	58
8 . Related Work	61
Bibliography	62

Figure Index

Figure 1: Storage growth expected to more than double a year	12
Figure 2: Data example showing MS Azure CloudBlob hierarchy.	18
Figure 3: Memory-mapped file and views	20
Figure 5: Dependency Graph for the CDB public interface	35
Figure 6: Not a good choice - metadata can occupy too much memory.	38
Figure 7: DataFile Tokens point to stored file metadata	39
Figure 8: CDB Public interface, with its classes and methods	40
Figure 9: Cloud-covered Computing abstractions	60

Table Index

Table 1: Main Index File Structure	43
Table 2: Operations Log	44
Table 3: Metadata file	45
Table 4: Rollback file	46
Table 5: Checking for successful operations in the Cloud	53
Table 6: Main Index File Structure	54
Table 7: Metadata file for the Windows Phone API	55
Table 8: Time spent to add+delete files in the local storage	57
Table 9: Time spent to add+delete files to the Cloud storage area	58