



**Marcello de Lima Azambuja**

**A Cloud Computing Architecture for Large  
Scale Video Data Processing**

**Dissertação de Mestrado**

Dissertation presented to the Postgraduate Program in Informatics of the Departamento de Informática do Centro Técnico da PUC-Rio, as partial fulfillment of the requirements for the degree of Mestre em Informática.

Advisor: Prof. Karin Breitman

Rio de Janeiro  
August 2011



**Marcello de Lima Azambuja**

**A Cloud Computing Architecture for Large  
Scale Video Data Processing**

Dissertation presented to the Postgraduate Program in Informatics of the Departamento de Informática do Centro Técnico da PUC-Rio, as partial fulfillment of the requirements for the degree of Mestre em Informática.

**Prof. Karin Breitman**

Advisor

Departamento de Informática – PUC-Rio

**Prof. José Viterbo Filho**

Universidade Federal Fluminense – UFF

**Carolina Howard Felicíssimo**

Pesquisadora da Schlumberger

**Prof. José Eugênio Leal**

Coordenador Setorial do Centro Técnico Científico  
PUC-Rio

Rio de Janeiro, August 31, 2011

All rights reserved

## Marcello de Lima Azambuja

The author is graduated in Electronics and Computer Engineering from Universidade Federal do Rio de Janeiro – UFRJ in 2005. Currently, Senior Product Manager at Globo.com.

### Bibliographic data

Azambuja, Marcello de Lima

A Cloud Computing Architecture For Large Scale Video Data Processing / Marcello de Lima Azambuja; advisor: Karin Breitman – 2011.

v., 62 f: il. ; 29,7 cm

1. Dissertação (Mestrado em Informática) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática, Rio de Janeiro, 2011.

Inclui bibliografia.

1. Informática – Teses. 2. Arquitetura de software. 3. Computação na nuvem. 4. Armazenamento. 5. Sistemas distribuídos. 6. Conteúdo gerado pelo usuário. I. Breitman, Karin. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título

CDD: 004

## Resumo

Azambuja, Marcello de Lima; Breitman, Karin. **Uma Arquitetura em Nuvem para Processamento de Vídeo em Larga Escala**. Rio de Janeiro, 2011. 62p. Dissertação de Mestrado - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

O advento da Internet coloca grandes desafios para o projeto de sistemas de submissão abertos, uma vez que elimina barreiras físicas e geográficas. A redução de custos, associada a passagem de mídias analógicas para digitais, dificultou a projeção de capacidades e recursos necessários para a construção de tais sistemas. Neste trabalho propomos uma arquitetura de software, baseada em computação na nuvem, que provê a escalabilidade necessária para a construção de sistemas de submissão abertos. Estes sistemas são caracterizados pela geração de grandes volumes de dados. Um caso real de uso é analisado utilizando o processamento de vídeos.

## Palavras-chave

Arquitetura de software; computação na nuvem; armazenamento; sistemas distribuídos; conteúdo gerado pelo usuário.

## Abstract

Azambuja, Marcello de Lima; Breitman, Karin (Advisor). **A Cloud Computing Architecture for Large Scale Video Data Processing**. Rio de Janeiro, 2011. 62p. MSc. Dissertation - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

The advent of the Internet poses great challenges to the design of public submission systems as it eliminates traditional barriers, such as geographical location and cost. With open global access, it is very hard to estimate storage space and processing power required by this class of applications. In this thesis we explore cloud computing technology as an alternative solution. The main contribution of this work is a general architecture in which to built open access, data intensive, public submission systems. A real world scenario is analyzed using this architecture for video processing.

## Keywords

Software architecture; cloud computing; storage; distributed systems; user generated content.

# Contents

<b>1</b>	<b>INTRODUCTION</b> .....	<b>8</b>
1.1	MOTIVATION .....	8
1.2	GOALS .....	9
1.3	THESIS ORGANIZATION .....	9
<b>2</b>	<b>BACKGROUND AND THE ONLINE VIDEO LANDSCAPE</b> .....	<b>11</b>
2.1	OVERVIEW .....	11
2.2	SEA CHANGE .....	11
2.3	RAMPING UP .....	12
2.4	SPLIT AND MERGE .....	13
2.5	DISTRIBUTION CHALLENGES .....	14
2.6	CONSUMER CHALLENGES .....	15
<b>3</b>	<b>AN ARCHITECTURE FOR PUBLIC AND OPEN SUBMISSION SYSTEMS IN THE CLOUD</b> .....	<b>17</b>
3.1	PROBLEM ADDRESSED .....	17
3.2	RESEARCH RELEVANCE .....	17
3.3	UNIQUENESS OF DESIGN AND IMPLEMENTATION .....	18
3.4	UNDERLYING IMPLEMENTATION TECHNIQUES AND USED TECHNOLOGIES .....	21
3.5	DESCRIPTION OF PRESENTATION .....	22
<b>4</b>	<b>CASE STUDY – THE BIG BROTHER REALITY SHOW</b> .....	<b>26</b>
4.1	INTRODUCTION .....	26
4.2	CLOUD COMPUTING .....	27
4.3	PROPOSED ARCHITECTURE .....	28
4.4	CASE STUDY .....	31
4.5	USER GENERATED CONTENT ARCHITECTURE .....	32
4.6	CONCLUSION .....	35
<b>5</b>	<b>AN ARCHITECTURE FOR DISTRIBUTED HIGH PERFORMANCE VIDEO PROCESSING IN THE CLOUD</b> .....	<b>37</b>
5.1	INTRODUCTION .....	37
5.2	BACKGROUND AND PROBLEM STATEMENT .....	38
5.3	DISTRIBUTED VIDEO .....	40
5.4	THE SPLIT STEP .....	42
5.5	THE PROCESS STEP .....	44
5.6	THE MERGE STEP .....	46
5.7	PERFORMANCE TESTS .....	47
5.8	DISCUSSION .....	50
5.9	FAULT TOLERANCE .....	51
5.10	PRIVATE CLUSTER .....	52
5.11	PUBLIC CLOUD DEPLOYMENT .....	54
5.12	CONCLUSION .....	55
<b>6</b>	<b>CONCLUSIONS</b> .....	<b>57</b>
6.1	CONTRIBUTION .....	57
6.2	FUTURE WORK .....	57
<b>7</b>	<b>REFERENCES</b> .....	<b>59</b>

## List of Figures

FIGURE 2.1. EVOLUTION OF MEDIA DISPLAY TYPES. ....	12
FIGURE 2.2. SPLIT AND MERGE ARCHITECTURE.....	13
FIGURE 3.1. ARCHITECTURE FOR THE PUBLIC SUBMISSION SYSTEM. ....	19
FIGURE 3.2. EXAMPLE OF INPUT MESSAGE USED BY THE APPLICATION. ....	20
FIGURE 3.3. EXAMPLE OF OUTPUT MESSAGE USED BY THE APPLICATION. ....	20
FIGURE 3.4. BIG BROTHER BRASIL'S OFFICIAL WEB PAGE. ....	22
FIGURE 3.5. SIGN-IN AND SIGN-UP WEB PAGE.....	23
FIGURE 3.6. CONTENT SUBMISSION FORM WEB PAGE. ....	23
FIGURE 3.7. AMAZON AWS MANAGEMENT CONSOLE SCREEN. ....	24
FIGURE 4.1. PROPOSED ARCHITECTURE FOR THE PUBLIC SUBMISSION SYSTEM. ....	29
FIGURE 4.2. EXAMPLE OF INPUT MESSAGE USED BY THE APPLICATION. ....	30
FIGURE 4.3. EXAMPLE OF OUTPUT MESSAGE USED BY THE APPLICATION. ....	30
FIGURE 4.4. CONFIGURATION FILE FOR EC2 INSTANCE.....	31
FIGURE 4.5. WEB SITE FOR THE BIG BROTHER REALITY SHOW.....	33
FIGURE 4.6. META-DATA INFORMATION FOR THE UGC.....	34
FIGURE 4.7. USER GENERATED CONTENT ARCHITECTURE.....	34
FIGURE 5.1. MAP REDUCE ARCHITECTURE.....	38
FIGURE 5.2. ENCODING SPEED FOR DIFFERENT H.264 IMPLEMENTATIONS.....	41
FIGURE 5.3. THE PROPOSED SPLIT&MERGE ARCHITECTURE.....	42
FIGURE 5.4. TOTAL ENCODING TIMES FOR DIFFERENT SEQUENCE DURATIONS (S). ...	48
FIGURE 5.5. PRIVATE CLUSTER SPLIT&MERGE ARCHITECTURE.....	53
FIGURE 5.6. PUBLIC CLOUD SPLIT&MERGE ARCHITECTURE .....	55