A Backup as a Service (BaaS) Software Solution

Heitor Medrado de Faria
heitor.faria@serpro.gov.br

Departamento de Ciência da Computação
Instituto de Ciências Exatas
Universidade de Brasília

11 de Setembro de 2017
Topics

1. Introduction
   - Objective

2. Literature Review
   - Cloud Computing
   - Disaster Recovery
   - Architecture
   - Cloud Backup Macro Requirements
   - The Overlying Backup Software

3. The Proposal
   - Scope
   - Objectives
   - Schedule
   - Bibliography
Topics

1. Introduction
   - Objective
2. Literature Review
3. The Proposal
Introduction

- Backup is the replica of any data that can be used to restore its original form [Guise]
- Digital data created worldwide more than doubles every two years - 44 zetabytes in 2020 [IDC]
- Backup is old and popular, but most backup systems are outdated [Russel et. al]
- Modern backup should provide integration to the growing Cloud [Russel]
Objective

The present study aims to propose and deploy a Backup as a Service software solution

- Cloud/backup parameters
- Cloud backup challenges
- Architectures
- BaaS system requirements
- Prototype construction and deploy
Topics

1. Introduction

2. Literature Review
   - Cloud Computing
   - Disaster Recovery
   - Architecture
   - Cloud Backup Macro Requirements
   - The Overlying Backup Software

3. The Proposal
Cloud Computing

- Shift data and computational services from individual devices to distributed architectures [Armbrust, Buyya]
- Global market revenues growing from U$ 67 billions by 2015 to U$ 241 billions by 2020 [Columbus, Ried et al.]
- 5 characteristics, 3 service categories, 4 deployment models [NIST]
Cloud Disaster Recovery Model

- DR is persistent problem for the CSP - must provide services to their customers even if a data center is down
- Data, application and system level [Khoshkholghi et al.]

Figure: Comparison of DR models [Raju et al.]
DR Parameters [Alhazmi and Malaiya]

\[ RPO \propto \frac{1}{FB} \]  

\[ RTO = \text{fraction of } RPO + \sum_{j_{\text{min}}}^{5} T_j \]  

- Minimum RPO and RTO.
- Minimal impact on the normal system operation.
- Should be geographically separated.
- Application shall be restored to a consistent state.
- DR solution must guarantee integrity, privacy and confidentiality.
Cloud DR Challenges

- Dependency.
- Cost (initializing, ongoing and potential disaster costs).
- Failure Detection.
- Security.
- Replication Latency.
- Limited Data Transmission Capabilities.
- Increasing Data Storage Demand.
- Lack of Redundancy.
Remote Backup to the Cloud [Camacho et al.]

Figure: A Cloud Environment for Backup and Data Storage Architecture
Local Backup from the Cloud

Figure: Local Backup Architecture [Javaraiah]
Geographical Redundancy and Backup (GRB)

Figure: GRB Architecture [Pokharel et al.]
Inter-Private Cloud Storage (IPCS)

Figure: The Inter-Private Cloud Storage architecture [Jian-hua and Nan]
Secure-Distributed Data Backup (SDDB)

Figure: Basic configuration of the SDDB Architecture [Ueno et al.]
**BaaS Provider Requirements** [Ismail et al., Rimal et al.]

- Autonomy.
- Cloud Scalability.
- Self-description.
- Fault-tolerance.
- Interoperability.
- Load Balancing.
- Multi-party.
- Multi-Tenancy.
- Optimal Provisioned Infrastructure Usage.
- Quality-of-Service.
- Standard Interface.
- Storage and Data Management.
- Workload Management.
BaaS User’s Requirements [Ismail et al., Rimal et al.]

- Adaptability and Learning.
- Automation.
- Backup Standardization.
- Data Consistency (Application specific, Crash and File-system).
- Data Integrity.
- Firewall Protected or NATed Clients Backup.
- Non-disruptive Backup.
- Service Level Agreements (SLAs).
- User-Centric Privacy.
- User Consumption-based Billing and Metering.
- User Experience (UX).
BaaS Candidate Solutions

Figure: Magic Quadrant for Data Center Backup and Recovery Software [Russel].
BaaS Candidate Solutions

Figure: Backup Software Solutions Google Queries for the Last 5 years [Google].
## BaaS Candidate Solutions Comparison

Table: BaaS focused Backup Software Comparison

<table>
<thead>
<tr>
<th>Solution</th>
<th>S3 Storage Support</th>
<th>Self-service portal?</th>
<th>Number of Plugins</th>
<th>$ U$, Monthly Price per Client</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arcserve</td>
<td>Yes</td>
<td>No</td>
<td>4</td>
<td>3.08-7.59</td>
</tr>
<tr>
<td>Bacula Enterprise</td>
<td>Yes</td>
<td>No</td>
<td>17</td>
<td>3.90-16.90</td>
</tr>
<tr>
<td>Bareos</td>
<td>Yes</td>
<td>No</td>
<td>5</td>
<td>5.46-52.45</td>
</tr>
<tr>
<td>Commvault</td>
<td>Yes</td>
<td>Yes</td>
<td>22</td>
<td>12.71-31.66</td>
</tr>
<tr>
<td>EMC</td>
<td>No</td>
<td>No</td>
<td>9</td>
<td>38.25 or more</td>
</tr>
<tr>
<td>Spectrum Protect</td>
<td>Yes</td>
<td>No</td>
<td>10</td>
<td>90.66-146.67</td>
</tr>
<tr>
<td>Veeam</td>
<td>No</td>
<td>Yes</td>
<td>7</td>
<td>12.25</td>
</tr>
<tr>
<td>Veritas Netbackup</td>
<td>No</td>
<td>No</td>
<td>14</td>
<td>10.58-332.91</td>
</tr>
</tbody>
</table>
The Overlying Backup Software

- *Bacula* is the 3rd most popular enterprise multi-platform server backup system, and open source [Google Trends]
- Full, differential, incremental, copy and migrate multiplexed jobs supported
- Compression, block deduplication, backups from FIFO, communication and data encryption [Sibbald]

Figure: Bacula traditional volume writing format.
Bacula Architecture

![Bacula Architecture Diagram](image)

Figure: Bacula Services Distribution and Network Ports [Preston].
Bacula Features

- General Features (organization, resource efficiency, architecture multi-thread etc.).
- Job Features (scheduler, job types, multiplexing etc.).
- Interface Features (CLI, GUI, REST, monitoring, multi-party etc.).
- Catalog Features (SQL, file browsing, pruning routines etc.).
- Storage Features (tape, Tape-Library, disks, NAS, NDMP, S3, object storage etc.).
- Data Reducing Features (compression, deduplication etc.).
- Security Features (backup server DR, backup files checksum, storage integrity verification etc.).
- Specific Applications Backup (virtual machines, databases, LDAP bases, system state etc.).
Topics

1. Introduction
2. Literature Review
3. The Proposal
   - Scope
   - Objectives
   - Schedule
   - Bibliography
Choosen Architecture
Remote Backup to the Cloud

Figure: A Cloud Environment for Backup and Data Storage Architecture.
**Choosen Requirements**

<table>
<thead>
<tr>
<th>Id</th>
<th>Requirement</th>
<th>Macro Requirement</th>
<th>Re-requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Shall contain interactive illustration or demonstration to instruct unaware users.</td>
<td>adaptability</td>
<td>and learning</td>
</tr>
<tr>
<td>9</td>
<td>Shall provide the usage of a Bacula single instance of the object code and database in order to support multiple customers.</td>
<td>multi-tenancy</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Shall be able to easily create new tenants</td>
<td></td>
<td>automation</td>
</tr>
<tr>
<td>16</td>
<td>Shall involve feedback of end-user on the interface design, and deploy</td>
<td></td>
<td>user experience</td>
</tr>
</tbody>
</table>
Objectives

This work aims to deploy a Backup as a Service solution under the Remote Backup to the Cloud architecture. In order to achieve what is aimed the following specific Objectives have been determined.

1. To enroll cloud backup challenges.
2. To find current proposed cloud backup architectures.
3. To determine what BaaS features must be added to traditional backup software.
4. To deploy, evaluate and validate the proposal prototype.
Activities

1. Architecture Design and Adjustment. Verify and document the technical specifications, boundaries and conditions of the experiment.
2. Determine performance metrics, system and workload parameters. Select evaluation techniques.
3. Prototype Development. Description and documentation of features deploy and inner workings.
4. Validation. Analyze and interpret data. Present the results.
5. Prototype Adjustments.
8. Scientific Papers Writing.
### Activities

Table: Research activities schedule.

<table>
<thead>
<tr>
<th>Act.</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

Note: The table indicates which months each activity is scheduled to take place. "X" marks the months the activity is planned for.
A Backup as a Service (BaaS) Software Solution

Heitor Medrado de Faria
heitor.faria@serpro.gov.br

Departamento de Ciência da Computação
Instituto de Ciências Exatas
Universidade de Brasília

11 de Setembro de 2017