



Ministério da Justiça



UnB



**Centro de Apoio ao
Desenvolvimento
Tecnológico**



Laboratório de tecnologias da tomada de decisão

Termo de Cooperação/Projeto:

**Acordo de Cooperação Técnica
FUB/CDT e MJ/SE
Registro de Identidade Civil –
Replanejamento e Novo Projeto Piloto**

Documento:

**RT Infraestrutura Tecnológica:
Planejamento da Prova de Conceito –
Banco de Dados**

Data de Emissão:

30/08/2015

Elaborado por:

**Universidade de Brasília – UnB
Centro de Apoio ao Desenvolvimento
Tecnológico – CDT
Laboratório de Tecnologias da
Tomada de Decisão – LATITUDE.UnB**

MINISTÉRIO DA JUSTIÇA

José Eduardo Cardozo
Ministro

Marivaldo de Castro Pereira
Secretário Executivo

Helvio Pereira Peixoto
Coordenador Suplente do Comitê Gestor do SINRIC

EQUIPE TÉCNICA

Ana Maria da Consolação Gomes Lindgren
Andréa Benoliel de Lima
Celso Pereira Salgado
Delluiz Simões de Brito
Elaine Fabiano Tocantins
Fernando Saliba Oliveira
Fernando Teodoro Filho
Guilherme Braz Carneiro
Joaquim de Oliveira Machado
José Alberto Sousa Torres
Marcelo Martins Villar
Raphael Fernandes de Magalhães Pimenta
Rodrigo Borges Nogueira
Rodrigo Gurgel Fernandes Távora
Sara Lais Rahal Lenharo

UNIVERSIDADE DE BRASÍLIA

Ivan Marques Toledo Camargo
Reitor

Paulo Anselmo Ziani Suarez
Diretor do Centro de Apoio ao
Desenvolvimento Tecnológico – CDT

Rafael Timóteo de Sousa Júnior
Coordenador do Laboratório de Tecnologias
da Tomada de Decisão – LATITUDE

EQUIPE TÉCNICA

Flávio Elias Gomes de Deus
(Pesquisador Sênior)
William Ferreira Giozza
(Pesquisador Sênior)
Ademir Agostinho de Rezende Lourenço
Adriana Nunes Pinheiro
Alysson Fernandes de Chantal
Amanda Almeida Paiva
Andréia Campos Santana
Antônio Claudio Pimenta Ribeiro
Carolinne Januária de Souza Martins
Daniela Carina Pena Pascual
Danielle Ramos da Silva
Diogenes Ferreira Reis Fustinoni
Fábio Lúcio Lopes Mendonça
Fábio Mesquita Buiati
Glaudson Menegazzo Verzeletti
Heverson Soares de Brito
Johnatan Santos de Oliveira
José Carneiro da Cunha Oliveira Neto
Kelly Santos de Oliveira Bezerra
Luciano Pereira dos Anjos
Luciene Pereira de Cerqueira Kaipper
Luiz Antônio de Souto Evaristo
Luiz Claudio Ferreira
Marco Schaffer
Marcos Vinicius Vieira da Silva
Pedro Augusto Oliveira de Paula
Roberto Mariano de Oliveira Soares
Sergio Luiz Teixeira Camargo
Soleni Guimarães Alves
Suzane Lais De Freitas
Valério Aymoré Martins
Vera Lopes de Assis
Wladimir Rodrigues da Fonseca

Projeto: MJ/SE-RIC	Emissão: 30/08/2015	Arquivo: 20150830 MJ RIC - RT Planejamento da Prova de Conceito - Banco de Dados	Pág.2/43
--------------------	---------------------	--	----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

HISTÓRICO DE REVISÕES

Data	Versão	Descrição
17/02/2015	0.1	Versão inicial.
30/02/2015	0.2	Versão com novos conteúdos sobre os servidores e banco de dados.
30/08/2015	0.3	Versão submetida à revisão final.



Universidade de Brasília – UnB
Campus Universitário Darcy Ribeiro - FT – ENE – Latitude
CEP 70.910-900 – Brasília-DF
Tel.: +55 61 3107-5598 – Fax: +55 61 3107-5590

Projeto: MJ/SE-RIC	Emissão: 30/08/2015	Arquivo: 20150830 MJ RIC - RT Planejamento da Prova de Conceito - Banco de Dados	Pág.3/43
--------------------	---------------------	--	-----------------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

SUMÁRIO

ABREVIACÕES	5
FIGURAS	6
TABELAS	7
1 INTRODUÇÃO	8
2 PROVA DE CONCEITO	10
3 SERVIÇOS	11
3.1 Serviços de Autenticação	11
3.2 Serviços de Cadastramento	12
3.3 Serviços de Informação.....	13
4 MODELO DE DADOS E VOLUMETRIA	15
4.1 Volumetria	15
4.2 Modelo de Dados / Mensageria	16
5 BANCO DE DADOS	24
5.1 Armazenadores de Dados.....	24
5.1.1 Armazenamento Chave-Valor (Key-Value Stores)	24
5.1.2 Armazenamento Orientado a Colunas (BigTable-style Databases)	25
5.1.3 Armazenamento Orientado a Documentos (Document Databases).....	26
5.1.4 Armazenamento Orientado a Grafos (Graph Databases)	28
5.2 Benchmarking YCSB (Yahoo! Cloud Serving Benchmark)	29
5.2.1 Carga A (Autenticação Biométrica)	33
5.2.2 Carga B (Autenticação OTP/Usuário e Senha).....	34
5.2.3 Carga C (Cadastramento das Impressões Digitais Brutas)	34
5.2.4 Carga D (Cadastramento da Imagem Bruta da Íris)	35
5.2.5 Carga E (Cadastramento da Imagem da Carteira de Identidade)	36
5.2.6 Carga F (Recuperação dos Dados Biográficos)	36
6 INFRAESTRUTURA	38
6.1 Infraestrutura Física	39
6.2 Arquitetura Lógica.....	40
7 CONCLUSÃO	41
REFERÊNCIAS.....	42

ABREVIações

ABIS	<i>Automated Biometric Identification System</i>
AFIS	<i>Automated Fingerprint Identification System</i>
CER	<i>Crossover Equal Rate</i>
DET	<i>Detection Error Trade-off</i>
EER	<i>Equal Error Rate</i>
FAR	<i>False Acceptance Rate</i>
FRR	<i>False Rejection Rate</i>
FTE	<i>Failure to Enroll</i>
GAR	<i>Genuine Acceptance Rate</i>
ROC	<i>Receiver Operating Characteristics</i>
SDK	<i>Software Development Kit</i>

FIGURAS

Figura 1 - Arquitetura Tecnológica da Prova de Conceito

38

Projeto: MJ/SE-RIC	Emissão: 30/08/2015	Arquivo: 20150830 MJ RIC - RT Planejamento da Prova de Conceito - Banco de Dados	Pág.6/43
--------------------	---------------------	--	-----------------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

TABELAS

Tabela 1 – Volumetria	16
Tabela 2 – Bancos de dados selecionados para a execução do benchmarking YCSB.....	29
Tabela 3 – Arquivo de criação das cargas de trabalho do YCSB.....	33
Tabela 4 – Carga A (Autenticação Biométrica).	34
Tabela 5 – Carga B (Autenticação OTP/Usuário e Senha).....	34
Tabela 6 – Carga C (Cadastramento das Impressões Digitais Brutas).....	35
Tabela 7 – Carga D (Cadastramento da Imagem Bruta da Íris).....	36
Tabela 8 – Carga E (Cadastramento da Imagem da Carteira de Identidade).....	36
Tabela 9 – Carga F (Recuperação dos Dados Biográficos).....	37
Tabela 10 – Infraestrutura Física.	39

1 INTRODUÇÃO

A Secretaria Executiva (SE/MJ), vinculada ao Ministério da Justiça (MJ), é responsável por viabilizar o desenvolvimento e a implantação do Registro de Identidade Civil, instituído pela Lei nº 9.454, de 7 de abril de 1997, regulamentado pelo Decreto nº 7.166, de 5 de maio de 2010.

Atualmente, a República Federativa do Brasil conta com sistema de identificação de seus cidadãos amparado pela Lei nº 7.116, de 29 de agosto de 1983. Essa lei assegura validade nacional às Carteiras de Identidade, ou Cédulas de Identidade; confere também autonomia gerencial às Unidades Federativas no que concerne à expedição e controle dos números de registros gerais emitidos para cada documento. Essa condição de autonomia, ao contrário do que pode parecer, fragiliza o sistema de identificação, já que dá condições ao cidadão de requerer legalmente até 27 (vinte e sete) cédulas de identidades diferentes. Com essa facilidade legal, inúmeras possibilidades fraudulentas se apresentam de maneira silenciosa, pois, na grande maioria dos casos, os Institutos de Identificação das Unidades Federativas não dispõem de protocolos e aparato tecnológico para identificar as duplicações de registro vindas de outros estados, ou até mesmo do seu próprio arquivo datiloscópico. Consoante aos fatos, os Institutos de Identificação não trabalham interativamente para que haja trocas de informações de dados e geração de conhecimento para manuseio inteligente e seguro para individualização do cidadão em prol da sociedade.

Com foco na busca de soluções para tais problemas, o Projeto RIC prevê a administração central dos dados biográficos e biométricos dos cidadãos no Cadastro Nacional de Registro de Identificação Civil (CANRIC) e ABIS (do inglês *Automated Biometric Identification System*), respectivamente. A previsão desse novo modelo sustenta a não duplicação de registros e a consequente identificação unívoca dos cidadãos brasileiros natos e naturalizados. O Projeto RIC, portanto, visa otimizar o sistema de identificação e individualização do cidadão brasileiro nato e naturalizado com vistas a um perfeito funcionamento da gestão de dados da sociedade, agregando valor à cidadania, à gestão administrativa, à simplificação do acesso aos serviços disponíveis ao cidadão e à segurança pública do país.

Nesse contexto, o termo de cooperação entre MJ/SE e FUB/CDT define um projeto que

Projeto: MJ/SE-RIC	Emissão: 30/08/2015	Arquivo: 20150830 MJ RIC - RT Planejamento da Prova de Conceito - Banco de Dados	Pág.8/43
--------------------	---------------------	--	----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.

É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

objetiva identificar, mapear e desenvolver parte dos processos e da infraestrutura tecnológica necessária para viabilizar a realização de uma prova de conceito do Registro de Identidade Civil – RIC no Brasil.

Resultante de um subconjunto das atividades previstas para inicialização da cooperação MJ/SE e FUB/CDT, o presente documento apresenta uma primeira abordagem da prova de conceito a ser realizada. São analisados diferentes aspectos que podem ser considerados na escolha da adoção de uma infraestrutura tecnológica, apresentando diretrizes para subsidiar essa escolha.

Projeto: MJ/SE-RIC	Emissão: 30/08/2015	Arquivo: 20150830 MJ RIC - RT Planejamento da Prova de Conceito - Banco de Dados	Pág.9/43
--------------------	---------------------	--	-----------------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

2 PROVA DE CONCEITO

Esta seção visa estabelecer um planejamento detalhado da prova de conceito do projeto tecnológico do Programa RIC. Primeiramente, é feita uma introdução sobre o que é uma prova de conceito, para que serve, quando fazer e por que é importante em um projeto de base tecnológica.

Feita essa introdução, dar-se-á ênfase aos processos e serviços que serão implementados no âmbito da prova de conceito. Por último, comentaremos sobre alguns *benchmarks* de processamento de grandes bases de dados.

PoC ou prova de conceito é um termo oriundo do inglês (*Proof of Concept*) e é utilizado para estruturar um modelo prático que consiga provar o que foi estabelecido na pesquisa teórica. Na área de Tecnologia da Informação, uma prova de conceito também se refere ao desenvolvimento de um protótipo para provar a viabilidade de uma arquitetura ou sistema a ser desenvolvido.

Na prática, uma prova de conceito permite demonstrar a metodologia, os conceitos, as arquiteturas lógicas e físicas e as tecnologias envolvidas na implementação do projeto. Normalmente, é uma atividade que deve ser trabalhada de forma colaborativa, envolvendo projetistas de arquitetura e infraestrutura, desenvolvedores de sistemas e, até mesmo, a participação dos utilizadores finais do sistema.

Entre os objetivos da prova de conceito, pode-se destacar:

- lista das tecnologias e conceitos envolvidos na prototipação;
- esboço do modelo conceitual e do fluxo de informação;
- lista de *softwares* e *hardwares* adequados ao projeto;
- plano de execução das atividades do protótipo;
- instrumentos de mitigação de erros técnicos.

A grande vantagem de se realizar uma prova de conceito é poder testar *hardware* e *software*, utilizando os resultados para a elaboração de uma proposta final de projeto, validando-a antes que seja executada na prática. Adicionalmente, é de suma importância que se realize um grande número de testes e combinações de configurações de arquitetura nessa etapa, uma vez que nas fases posteriores, o custo para a realização de mudanças pode ser inviável para a implementação final do projeto.

Projeto: MJ/SE-RIC	Emissão: 30/08/2015	Arquivo: 20150830 MJ RIC - RT Planejamento da Prova de Conceito - Banco de Dados	Pág.10/43
--------------------	---------------------	--	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.

É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

3 SERVIÇOS

A prova de conceito foi definida em torno de um conjunto restrito de serviços que poderiam ser demonstrados em um cenário real. Eles foram selecionados, visando a um entendimento geral do cadastramento de um novo cidadão na base de dados nacional.

Para a execução desta prova de conceito, foram selecionadas as capacidades dos três principais grupos de serviços, dentre um universo de doze listados em [1]¹.

3.1 Serviços de Autenticação

Referem-se, fundamentalmente, ao atendimento dos princípios de autenticação por meio de recursos de identificação, pesquisa e validação do acesso a partir de chaves do RIC.

Capacidade	Validar
Capacidade OASIS	<i>Verify</i>
Contexto Funcional	Identifica única e inequivocamente um registro no RIC. A mensageria de retorno informa apenas esta validade (S/N).
Mensagem de Entrada	<i>VerifyRequestMessage</i>
Mensagem de Saída	<i>VerifyResponseMessage</i>
Característica de Acesso	Alto volume de acesso de leitura.
Volumetria/SLA	100.000.000 requisições/dia.

Capacidade	Identificar
Capacidade OASIS	<i>Identify</i>
Contexto Funcional	Lista os prováveis elementos identificados em uma pesquisa nos registros do RIC (<i>Candidate</i>). A mensageria de retorno é a lista de dados básicos de pesquisa no RIC. Os dados básicos são limitados pela característica de acesso público.
Mensagem de Entrada	<i>IdentifyRequestMessage</i>
Mensagem de Saída	<i>IdentifyResponseMessage</i>

¹ RT Estudos e Análise, Modelagem e Projeto de Serviços e de Modelos Canônico de Dados

Característica de Acesso	Médio volume de acesso de leitura.
Volumetria/SLA	100.000 requisições/dia.

3.2 Serviços de Cadastramento

Referem-se, fundamentalmente, aos serviços que atendem a manutenção da base do RIC e a respectiva entrega de informações como serviço (dados como serviço – DaaS). Junto aos serviços de cadastramento e informação, encontra-se capacidade (ou serviço específico) para atender ao processo de agendamento de cadastramento que sugerimos ser suportada por base de dados distinta, mas integrada ao RIC.

Capacidade	Cadastrar
Capacidade OASIS	<i>Enroll</i>
Contexto Funcional	Permite o cadastramento (inclusão) na base de dados do RIC.
Mensagem de Entrada	<i>EnrollRequest</i>
Mensagem de Saída	<i>EnrollResponse</i>
Característica de Acesso	Alto volume de acesso de gravação.
Volumetria/SLA	200.000.000 requisições/3 anos.

Capacidade	Recuperar Número de RIC
Capacidade OASIS	<i>GetEnrollResults</i>
Contexto Funcional	Permite a recuperação do Número do RIC após o processo de deduplicação.
Mensagem de Entrada	<i>GetEnrollResultsRequest</i>
Mensagem de Saída	<i>GetEnrollResultsResponse</i>

Projeto: MJ/SE-RIC	Emissão: 30/08/2015	Arquivo: 20150830 MJ RIC - RT Planejamento da Prova de Conceito - Banco de Dados	Pág.12/43
--------------------	---------------------	--	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.

É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

Característica de Acesso	Alto volume de acesso de leitura.
Volumetria/SLA	200.000.000 requisições/3 anos.

Capacidade	Expirar RIC
Capacidade OASIS	<i>Expires</i>
Contexto Funcional	Permite mudar o <i>status</i> de um indivíduo cadastrado e deduplicado.
Mensagem de Entrada	<i>ExpiresRequest</i>
Mensagem de Saída	<i>ExpiresResponse</i>
Característica de Acesso	Baixíssimo volume de acesso de gravação.
Volumetria/SLA	Casos esporádicos.

3.3 Serviços de Informação

Referem-se, fundamentalmente, aos serviços que suportam o atendimento de terceiros no acesso a dados e situações no RIC. Ainda não é possível o desenvolvimento da total expectativa de capacidades bem como a apropriação dos aspectos tecnológicos necessários, porém este documento já aponta capacidades consideradas elementares para esse serviço.

Capacidade	Recuperar Dados Biográficos
Contexto Funcional	Efetua a recuperação de dados biográficos de um indivíduo, realizando, preliminarmente, um teste de validação desse indivíduo na base.
Mensagem de Entrada	<i>RetrieveBiographicInformationRequest</i>

Mensagem de Saída	<i>RetrieveBiographicInformation Response</i>
Característica de Acesso	Alto volume de acesso de leitura.
Volumetria/SLA	Casos esporádicos.

Capacidade	Atualizar Dados Biográficos
Contexto Funcional	Efetua a atualização de dados biográficos de um indivíduo, realizando, preliminarmente, um teste de validação desse indivíduo na base.
Mensagem de Entrada	<i>UpdateBiographicDataRequest</i>
Mensagem de Saída	<i>UpdateBiographicDataResponse</i>
Característica de Acesso	Baixo volume de acesso de leitura.
Volumetria/SLA	Casos esporádicos.

4 MODELO DE DADOS E VOLUMETRIA

Um ponto muito importante no planejamento do desenvolvimento de uma prova de conceito é a análise de volumetria. Trata-se de um estudo detalhado do volume de dados que será gerado no sistema. Uma boa análise de volumetria impacta nas decisões de arquiteturas físicas e lógicas, no armazenamento de dados, nos planos de *backup*, na comunicação na rede, no modelado dos dados, entre outros fatores. Portanto, nesse capítulo far-se-á uma breve análise do volume de dados utilizado na prova de conceito. Estimou-se um volume de dados para os seis primeiros anos de vida do Programa RIC.

4.1 Volumetria

Unidade de medida (<i>TeraBytes</i>)	Ano 1	Ano 2	Ano 3	Ano 4	Ano 5	Ano 6
Cadastramento (cumulativo)	35.000.000,00	70.000.000	105.000.000	140.000.000	170.000.000	200.000.000
Dados Biográficos estáticos	0,0065	0,0130	0,0196	0,0261	0,0317	0,0373
Dados biográficos não estáticos	0,0033	0,0065	0,0098	0,0130	0,0158	0,0186
Template da íris	0,04	0,08	0,12	0,16	0,19	0,23
Template da digital	1,96	3,91	5,87	7,82	9,50	11,18
Imagens da digital	48,89	97,79	146,68	195,58	237,49	279,40
Imagens da íris	16,30	32,60	48,89	65,19	79,16	93,13
Documentos de cadastramento	65,19	130,39	195,58	260,77	316,65	372,53
Dados de autenticação	0,0033	0,0065	0,0098	0,0130	0,0158	0,0186
Imagens da face	8,15	16,30	24,45	32,60	39,58	46,57
Pacote de cadastramento	162,98	325,96	488,94	651,93	791,62	931,32
Logs de auditoria	0,13	0,26	0,39	0,52	0,63	0,75
Logs de autenticação	4,69	26,29	84,49	187,75	375,51	938,77
	308,3487	633,5953	995,4541	1.402,3740	1.850,4042	2.673,9437

Tabela 1 – Volumetria

4.2 Modelo de Dados / Mensageria

Dada a especificação dos serviços a serem implementados e usados na PoC de Serviços, ainda se faz necessário dar a conhecer os padrões de mensagens XML, isto é, o modelo de mensageria canônica de cada serviço, o qual se dará a seguir

Observa-se que em algumas descrições do modelo de mensagens de entrada e saída dos serviços, detalhadas a seguir, encontramos a simbologia "..."; em seguida, geralmente, uma estrutura do tipo "<Image>". Isto indica que existe uma repetição de blocos de "<Image>" até o que é definido no modelo de execução do RIC. No máximo estão previstos 13 blocos, correspondendo a:

- 10 imagens de digitais (*ContentInfo=Finger, FingerPosition* de 1 a 10);
- 1 imagem com a foto (*ContentInfo=Face*);
- 1 imagem de documento de identificação (*ContentInfo=DocumentID*);
- 1 imagem de documento de residência (*ContentInfo=DocumentResidence*).

Os subtópicos a seguir exemplificam o *payload* XML de cada mensagem de requisição (<Processo>*Request*) e de resposta (<Processo>*Response*), definido nos quadros apresentados abaixo.

4.2.1. VerifyRequest

```
<soapenv:Envelope xmlns:soapenv="http://schemas.xmlsoap.org/soap/envelope/">
  <soapenv:Body>
    <Verify>
      <VerifyRequest>
        <GenericRequestParameters>
          <Application>RIC</Application>
          <ApplicationUser>1</ApplicationUser>
          <BIASOperationName>Verify</BIASOperationName>
        </GenericRequestParameters>
        <ProcessingOptions>
          <Option>DEDUPLICATION=DEFAULT</Option>
        </ProcessingOptions>
        <InputData>
          <NumeroRIC></NumeroRIC>
          <Images>
            <Image>
              <ContentInfo>Finger</ContentInfo>
              <ContentType>image/png;base64</ContentType>
            </Image>
          </Images>
        </InputData>
      </VerifyRequest>
    </Verify>
  </soapenv:Body>
</soapenv:Envelope>
```

Projeto: MJ/SE-RIC	Emissão: 30/08/2015	Arquivo: 20150830 MJ RIC - RT Planejamento da Prova de Conceito - Banco de Dados	Pág.16/43
--------------------	---------------------	--	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.


```

        <FingerPosition></FingerPosition>
        <ImageData></ImageData>
    </Image>
</Images>
</InputData>
</VerifyRequest>
</Verify>
</soapenv:Body>
</soapenv:Envelope>

```

4.2.2. VerifyResponse

```

<soapenv:Envelope xmlns:soapenv="http://schemas.xmlsoap.org/soap/envelope/">
  <soapenv:Body>
    <Verify>
      <VerifyResponse>
        <GenericRequestParameters>
          <Application>RIC</Application>
          <ApplicationUser>0</ApplicationUser>
          <BIASOperationName>Verify</BIASOperationName>
        </GenericRequestParameters>
        <ProcessingOptions>
          <Option>DEDUPLICATION=DEFAULT</Option>
        </ProcessingOptions>
        <ResultCode>OK</ResultCode>
      </VerifyResponse>
    </Verify>
  </soapenv:Body>
</soapenv:Envelope>

```

4.2.3. IdentifyRequest

```

<soapenv:Envelope xmlns:soapenv="http://schemas.xmlsoap.org/soap/envelope/">
  <soapenv:Body>
    <Identify>
      <IdentifyRequest>
        <GenericRequestParameters>
          <Application>RIC</Application>
          <ApplicationUser>1</ApplicationUser>
          <BIASOperationName>Identify</BIASOperationName>
        </GenericRequestParameters>
        <ProcessingOptions>
          <Option>DEDUPLICATION=DEFAULT</Option>
        </ProcessingOptions>
        <InputData>
          <NumeroRIC></NumeroRIC>
          <Images>
            <Image>
              <ContentInfo>Finger</ContentInfo>
              <ContentType>image/png;base64</ContentType>
              <FingerPosition></FingerPosition>
              <ImageData></ImageData>
            </Image>
            ...
          </Images>
        </InputData>
      </IdentifyRequest>
    </Identify>
  </soapenv:Body>
</soapenv:Envelope>

```

4.2.4. IdentifyResponse

```
<soapenv:Envelope xmlns:soapenv="http://schemas.xmlsoap.org/soap/envelope/">
  <soapenv:Body>
    <Identify>
      <IdentifyResponse>
        <GenericRequestParameters>
          <Application>RIC</Application>
          <ApplicationUser>0</ApplicationUser>
          <BIASOperationName>Identify</BIASOperationName>
        </GenericRequestParameters>
        <ProcessingOptions>
          <Option>DEDUPLICATION=DEFAULT</Option>
        </ProcessingOptions>
        <Candidates>
          <Candidate>0000-0000-0000</Candidate>
          ...
        </Candidates>
      </IdentifyResponse>
    </Identify>
  </soapenv:Body>
</soapenv:Envelope>
```

4.2.5. EnrollRequest

```
<soapenv:Envelope xmlns:soapenv="http://schemas.xmlsoap.org/soap/envelope/">
  <soapenv:Body>
    <Enroll>
      <EnrollRequest>
        <GenericRequestParameters>
          <Application>RIC</Application>
          <ApplicationUser>1</ApplicationUser>
          <BIASOperationName> Enroll </BIASOperationName>
        </GenericRequestParameters>
        <ProcessingOptions>
          <Option>DEDUPLICATION=DEFAULT</Option>
        </ProcessingOptions>
        <InputData>
          <ID>1</ID>
          <NomeUC></NomeUC>
          <Nome></Nome>
          <Sexo></Sexo>
          <DataNascimento></DataNascimento>
          <Nacionalidade></Nacionalidade>
          <Naturalidade></Naturalidade>
          <Pai></Pai>
          <Mae></Mae>
          <SircRGNumero></SircRGNumero>
          <SircRGEmisor></SircRGEmisor>
          <ContatoEndereco></ContatoEndereco>
          <ContatoCEP></ContatoCEP>
          <ContatoTelefone></ContatoTelefone>
          <ContatoEmail></ContatoEmail>
          <Images>
            <Image>
              <ContentInfo>Finger</ContentInfo>
              <ContentType>image/png;base64</ContentType>
              <FingerPosition></FingerPosition>
              <ImageData></ImageData>
            </Image>
            ...
          </Images>
        </InputData>
      </EnrollRequest>
    </Enroll>
  </soapenv:Body>
</soapenv:Envelope>
```

```
</InputData>
</EnrollRequest>
</Enroll>
</soapenv:Body>
</soapenv:Envelope>
```

4.2.6. EnrollResponse

```
<soapenv:Envelope xmlns:soapenv="http://schemas.xmlsoap.org/soap/envelope/">
  <soapenv:Body>
    <Enroll>
      <EnrollResponse>
        <GenericRequestParameters>
          <Application>RIC</Application>
          <ApplicationUser>0</ApplicationUser>
          <BIASOperationName> Enroll </BIASOperationName>
        </GenericRequestParameters>
        <ProcessingOptions>
          <Option>DEDUPLICATION=DEFAULT</Option>
        </ProcessingOptions>
        <Token>100000000000</Token>
      </EnrollResponse>
    </Enroll>
  </soapenv:Body>
</soapenv:Envelope>
```

4.2.7. GetEnrollResultsRequest

```
<soapenv:Envelope xmlns:soapenv="http://schemas.xmlsoap.org/soap/envelope/">
  <soapenv:Body>
    <GetEnrollResults>
      <GetEnrollResultsRequest>
        <GenericRequestParameters>
          <Application>RIC</Application>
          <ApplicationUser>1</ApplicationUser>
          <BIASOperationName>GetEnrollResults</BIASOperationName>
        </GenericRequestParameters>
        <ProcessingOptions>
          <Option>DEDUPLICATION=DEFAULT</Option>
        </ProcessingOptions>
        <InputData>
          <Token>100000000000</Token>
          <Images>
            <Image>
              <ContentInfo>Finger</ContentInfo>
              <ContentType>image/png;base64</ContentType>
              <FingerPosition></FingerPosition>
              <ImageData></ImageData>
            </Image>
          </Images>
        </InputData>
      </GetEnrollResultsRequest>
    </GetEnrollResults>
  </soapenv:Body>
</soapenv:Envelope>
```

4.2.8. GetEnrollResultsResponse

```
<soapenv:Envelope xmlns:soapenv="http://schemas.xmlsoap.org/soap/envelope/">
  <soapenv:Body>
    <GetEnrollResults>
      <GetEnrollResultsResponse>
```

Projeto: MJ/SE-RIC	Emissão: 30/08/2015	Arquivo: 20150830 MJ RIC - RT Planejamento da Prova de Conceito - Banco de Dados	Pág.19/43
--------------------	---------------------	--	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

```
<GenericRequestParameters>
  <Application>RIC</Application>
  <ApplicationUser>0</ApplicationUser>
  <BIASOperationName>GetEnrollResults</BIASOperationName>
</GenericRequestParameters>
<ProcessingOptions>
  <Option>DEDUPLICATION=DEFAULT</Option>
</ProcessingOptions>
<NumeroRIC>1000-0000-0001</NumeroRIC>
</GetEnrollResultsResponse>
</Verify>
</soapenv:Body>
</soapenv:Envelope>
```

4.2.9. ExpiresRequest

```
<soapenv:Envelope xmlns:soapenv="http://schemas.xmlsoap.org/soap/envelope/">
  <soapenv:Body>
    <Expires>
      <ExpiresRequest>
        <GenericRequestParameters>
          <Application>RIC</Application>
          <ApplicationUser>1</ApplicationUser>
          <BIASOperationName>Expires</BIASOperationName>
        </GenericRequestParameters>
        <ProcessingOptions>
          <Option>DEDUPLICATION=DEFAULT</Option>
        </ProcessingOptions>
        <InputData>
          <NumeroRIC></NumeroRIC>
          <Images>
            <Image>
              <ContentInfo>Finger</ContentInfo>
              <ContentType>image/png;base64</ContentType>
              <FingerPosition></FingerPosition>
              <ImageData></ImageData>
            </Image>
          </Images>
        </InputData>
      </ExpiresRequest>
    </Expires>
  </soapenv:Body>
</soapenv:Envelope>
```

4.2.10. ExpiresResponse

```
<soapenv:Envelope xmlns:soapenv="http://schemas.xmlsoap.org/soap/envelope/">
  <soapenv:Body>
    <Expires>
      <ExpiresResponse>
        <GenericRequestParameters>
          <Application>RIC</Application>
          <ApplicationUser>0</ApplicationUser>
          <BIASOperationName>Expires</BIASOperationName>
        </GenericRequestParameters>
        <ProcessingOptions>
          <Option>DEDUPLICATION=DEFAULT</Option>
        </ProcessingOptions>
        <ResultCode>OK</ResultCode>
      </ExpiresResponse>
    </Expires>
  </soapenv:Body>
```

Projeto: MJ/SE-RIC	Emissão: 30/08/2015	Arquivo: 20150830 MJ RIC - RT Planejamento da Prova de Conceito - Banco de Dados	Pág.20/43
--------------------	---------------------	--	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

</soapenv:Envelope>

4.2.11. RetrieveBiographicInformationRequest

```
<soapenv:Envelope xmlns:soapenv="http://schemas.xmlsoap.org/soap/envelope/">
  <soapenv:Body>
    <RetrieveBiographicInformation>
      <RetrieveBiographicInformationRequest>
        <GenericRequestParameters>
          <Application>RIC</Application>
          <ApplicationUser>1</ApplicationUser>
          <BIASOperationName>RetrieveBiographicInformation</BIASOperationName>
        </GenericRequestParameters>
        <ProcessingOptions>
          <Option>DEDUPLICATION=DEFAULT</Option>
        </ProcessingOptions>
        <InputData>
          <NumeroRIC></NumeroRIC>
          <Images>
            <Image>
              <ContentInfo>Finger</ContentInfo>
              <ContentType>image/png;base64</ContentType>
              <FingerPosition></FingerPosition>
              <ImageData></ImageData>
            </Image>
          </Images>
        </InputData>
      </RetrieveBiographicInformationRequest>
    </RetrieveBiographicInformation>
  </soapenv:Body>
</soapenv:Envelope>
```

4.2.12. RetrieveBiographicInformationResponse

```
<soapenv:Envelope xmlns:soapenv="http://schemas.xmlsoap.org/soap/envelope/">
  <soapenv:Body>
    <RetrieveBiographicInformation>
      <RetrieveBiographicInformationResponse>
        <GenericRequestParameters>
          <Application>RIC</Application>
          <ApplicationUser>1</ApplicationUser>
          <BIASOperationName>RetrieveBiographicInformation</BIASOperationName>
        </GenericRequestParameters>
        <ProcessingOptions>
          <Option>DEDUPLICATION=DEFAULT</Option>
        </ProcessingOptions>
        <InputData>
          <ID>1</ID>
          <NomeUC></NomeUC>
          <Nome></Nome>
          <Sexo></Sexo>
          <DataNascimento></DataNascimento>
          <Nacionalidade></Nacionalidade>
          <Naturalidade></Naturalidade>
          <Pai></Pai>
          <Mae></Mae>
          <SircRGNumero></SircRGNumero>
          <SircRGEmissor></SircRGEmissor>
          <ContatoEndereco></ContatoEndereco>
          <ContatoCEP></ContatoCEP>
          <ContatoTelefone></ContatoTelefone>
          <ContatoEmail></ContatoEmail>
        </InputData>
      </RetrieveBiographicInformationResponse>
    </RetrieveBiographicInformation>
  </soapenv:Body>
</soapenv:Envelope>
```

Projeto: MJ/SE-RIC	Emissão: 30/08/2015	Arquivo: 20150830 MJ RIC - RT Planejamento da Prova de Conceito - Banco de Dados	Pág.21/43
--------------------	---------------------	--	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

```
</InputData>
</RetrieveBiographicInformationResponse>
</RetrieveBiographicInformation>
</soapenv:Body>
</soapenv:Envelope>
```

4.2.13. UpdateBiographicDataRequest

```
<soapenv:Envelope xmlns:soapenv="http://schemas.xmlsoap.org/soap/envelope/">
  <soapenv:Body>
    <UpdateBiographicData>
      <UpdateBiographicDataRequest>
        <GenericRequestParameters>
          <Application>RIC</Application>
          <ApplicationUser>1</ApplicationUser>
          <BIASOperationName>UpdateBiographicData</BIASOperationName>
        </GenericRequestParameters>
        <ProcessingOptions>
          <Option>DEDUPLICATION=DEFAULT</Option>
        </ProcessingOptions>
        <InputData>
          <DataNascimento></DataNascimento>
          <Nacionalidade></Nacionalidade>
          <Naturalidade></Naturalidade>
          <Pai></Pai>
          <Mae></Mae>
          <SircRGNumero></SircRGNumero>
          <SircRGEmissor></SircRGEmissor>
          <ContatoEndereco></ContatoEndereco>
          <ContatoCEP></ContatoCEP>
          <ContatoTelefone></ContatoTelefone>
          <ContatoEmail></ContatoEmail>
          <Images>
            <Image>
              <ContentInfo>Finger</ContentInfo>
              <ContentType>image/png;base64</ContentType>
              <FingerPosition></FingerPosition>
              <ImageData></ImageData>
            </Image>
          </Images>
        </InputData>
      </UpdateBiographicDataRequest>
    </UpdateBiographicData>
  </soapenv:Body>
</soapenv:Envelope>
```

4.2.12. UpdateBiographicDataResponse

```
<soapenv:Envelope xmlns:soapenv="http://schemas.xmlsoap.org/soap/envelope/">
  <soapenv:Body>
    <UpdateBiographicData>
      <UpdateBiographicDataRequest>
        <GenericRequestParameters>
          <Application>RIC</Application>
          <ApplicationUser>1</ApplicationUser>
          <BIASOperationName>UpdateBiographicData</BIASOperationName>
        </GenericRequestParameters>
        <ProcessingOptions>
          <Option>DEDUPLICATION=DEFAULT</Option>
        </ProcessingOptions>
        <ResultCode>OK</ResultCode>
      </UpdateBiographicDataRequest>
    </UpdateBiographicData>
  </soapenv:Body>
</soapenv:Envelope>
```

Projeto: MJ/SE-RIC	Emissão: 30/08/2015	Arquivo: 20150830 MJ RIC - RT Planejamento da Prova de Conceito - Banco de Dados	Pág.22/43
--------------------	---------------------	--	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

```
</UpdateBiographicData>  
</soapenv:Body>  
</soapenv:Envelope>
```

5 BANCO DE DADOS

No que se refere aos serviços de armazenamento de dados, para o estabelecimento de uma infraestrutura com condições de suportar o esperado volume de acesso massivo, contínuo e distribuído baseado em serviços web - como idealizado para o RIC - é necessário que se discuta os problemas notórios envolvidos no armazenamento de dados, principalmente com o uso mais recente de técnicas que permitem o tratamento massivo de dados, técnicas essas que estão ora concentrados no estudo dos assuntos de BigData e de armazenadores NoSQL.

Portanto, nesse capítulo far-se-á uma explicação dos armazenadores de dados NoSQL e da ferramenta de benchmarking YCSB, bastante utilizada por administradores de bancos de dados na etapa de planejamento da estrutura física e lógica de armazenamento.

5.1 Armazenadores de Dados

A principal motivação para o desenvolvimento de armazenadores de dados, conhecidos como NoSQL, foi para dar suporte a implementação da Web 2.0, que aumentou o uso e a quantidade de dados armazenados em bases de dados.

Os armazenadores NoSQL não são projetados para abandonar SQL, a linguagem de consulta usada para recuperar informações de bancos de dados relacionais tradicionais, tais como Oracle e MySQL. Um nome melhor seria defini-los como soluções de "banco de dados não-relacionais" pois não usam o modelo de tabelas normalizadas de dados que sustentam as bases de dados relacionais.

Abaixo, detalhamos os quatro principais tipos de armazenadores NoSQL:

5.1.1 Armazenamento Chave-Valor (Key-Value Stores)

a) Conceituação

A ideia principal desse armazenador é usar uma tabela hash, na qual há uma chave única e um indicador de um dado ou de um item em particular. O modelo chave-valor é o mais simples e fácil de implementar, mas ele é ineficiente quando você somente está interessado em consultar ou em atualizar parte de um valor, entre outras desvantagens. É um modelo simples que permite a visualização do

Projeto: MJ/SE-RIC	Emissão: 30/08/2015	Arquivo: 20150830 MJ RIC - RT Planejamento da Prova de Conceito - Banco de Dados	Pág.24/43
--------------------	---------------------	--	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.

É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

banco de dados como uma grande tabela hash.

b) Exemplos:

- Dynamo DB (Amazon)
- Voldemort
- Oracle Berkeley DB
- Scalaris
- Redis
- Tokyo Cabinet/Tyrant.

c) Aplicações típicas:

Cache de conteúdo (foco em escalar para imensas quantidades de dados)

d) Modelo de dados:

Coleção de pares de chave-valor

e) Pontos fortes:

Pesquisas rápidas

f) Fraquezas:

Dados armazenados não têm schema.

5.1.2 Armazenamento Orientado a Colunas (BigTable-style Databases)

a) Conceituação

Foram criados para armazenar e para processar grandes quantidades de dados distribuídos em muitas máquinas. Ainda existem chaves, mas elas apontam para colunas múltiplas. As colunas são organizadas por família da coluna. É um estilo emergente de banco de dados que teve como referência o modelo BigTable do

Projeto: MJ/SE-RIC	Emissão: 30/08/2015	Arquivo: 20150830 MJ RIC - RT Planejamento da Prova de Conceito - Banco de Dados	Pág.25/43
--------------------	---------------------	--	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.

É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

Google, e que dentre suas características podemos destacar o particionamento vertical dos dados com forte consistência dos mesmos.

b) Exemplos:

- Apache HBase
- Cassandra
- Google BigTable
- Hyperbase
- Hypertable
- Apache Accumulo
- HPCC
- Splice Machine

c) Aplicações típicas:

Sistemas de arquivos distribuídos

d) Modelo de dados:

Colunas e famílias de colunas

e) Pontos fortes:

Pesquisas rápidas

Boa distribuição de armazenamento de dados

f) Fraquezas:

API de baixo nível.

5.1.3 Armazenamento Orientado a Documentos (Document Databases)

a) Conceituação

Projeto: MJ/SE-RIC	Emissão: 30/08/2015	Arquivo: 20150830 MJ RIC - RT Planejamento da Prova de Conceito - Banco de Dados	Pág.26/43
--------------------	---------------------	--	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.

É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

Foram inspirados por Lotus Notes e são similares ao armazenamento chave-valor. O modelo consiste basicamente de documentos versionados que são coleções de outras coleções de chave-valor. Os documentos semi-estruturados são armazenados em formato JSON. Bancos de dados de documentos são essencialmente o próximo nível do chave-valor, permitindo valores aninhados associados a cada chave. Bancos de dados de documentos suportam consultas mais eficientemente, pois armazenam coleções de documentos, ou seja, objetos com identificadores únicos e um conjunto de campos (strings, listas ou documentos aninhados), assemelhando ao modelo chave-valor, porém cada documento tem um conjunto de campos-chaves e valores destes campos.

b) Exemplos:

- CouchDB
- MongoDB.

c) Aplicações típicas:

Aplicações Web (Similar ao armazenamento chave-valor, mas o BD sabe qual é o valor)

d) Modelo de dados:

Coleções de coleções de chave-valor

e) Pontos fortes:

Tolerante a dados incompletos

f) Fraquezas:

Desempenho na pesquisa

Não apresenta uma sintaxe de pesquisa padrão

Projeto: MJ/SE-RIC	Emissão: 30/08/2015	Arquivo: 20150830 MJ RIC - RT Planejamento da Prova de Conceito - Banco de Dados	Pág.27/43
--------------------	---------------------	--	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.

É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

5.1.4 Armazenamento Orientado a Grafos (Graph Databases)

a) Conceituação

Neste modelo, em vez de tabelas com linhas e colunas e a rígida estrutura do SQL, um modelo gráfico flexível é usado e que pode, mais uma vez, ser escalado através de varias máquinas. Possui três componentes básicos (nós, relacionamentos, propriedades) que permitem que o SGBD possa ser visto como um multigrafo rotulado e direcionado, onde cada par de nós pode ser conectado por mais de uma aresta.

b) Exemplos:

- Neo4j
- InfoGrid
- Sones
- HyperGraphDB.

c) Aplicações típicas:

Redes Sociais, Recomendações (Foco em modelar a estrutura dos dados – interconectividade)

d) Modelo de dados:

Grafo de Propriedades – Nós

e) Pontos fortes:

Suporte direto a algoritmos gráficos tais como caminho mais curto, conectividade, relacionamentos em n graus, etc.

f) Fraquezas:

Tem que atravessar todo o gráfico para obter uma resposta definitiva.

Não facilita solução de problemas de agrupamento / agregação.

Projeto: MJ/SE-RIC	Emissão: 30/08/2015	Arquivo: 20150830 MJ RIC - RT Planejamento da Prova de Conceito - Banco de Dados	Pág.28/43
--------------------	---------------------	--	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.

É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

5.2 Benchmarking YCSB (Yahoo! Cloud Serving Benchmark)

Para a realização da prova de conceito dos bancos de dados, utilizou-se a ferramenta de benchmark YCSB (*Yahoo! Cloud Serving Benchmark*)² que oferece um conjunto pré-definido de cargas de trabalho (*workloads*) para obter métricas de rendimento em mais de dez bancos de dados diferentes [2].

YCSB está desenvolvido em Java e disponível na plataforma GitHub. A versão utilizada (0.1.4), dispõe de uma interface para uma série de bancos de dados, tais como: Cassandra, DynamoDB, GigaSpaces XAP, HBase, Infinispan, MongoDB, Oracle NoSQL Database, Redis e Voldemort³. Além disso, a ferramenta permite que o desenvolvedor programe uma interface para a inclusão de novos bancos de dados⁴.

O estudo realizado pela equipe técnica do Programa RIC sobre a Infraestrutura Tecnológica referente à banco de dados SQL, NewSQL e armazenadores NoSQL [3] balizou a decisão de utilizar bancos de dados para o planejamento da prova de conceito:

Armazenamento NoSQL	Bancos utilizados
Chave-valor	Dynamo DB, Voldemort
Orientado a coluna	Hbase, Cassandra
Orientado a documentos	CouchDB, MongoDB
Orientado a grafos	Neo4j, Infogrid
Armazenamento SQL	MySQL, PostgreSQL

Tabela 2 – Bancos de dados selecionados para a execução do benchmarking YCSB.

Além dos bancos de dados NoSQL, também incluímos no planejamento dessa prova

² <https://github.com/brianfrankcooper/YCSB>

³ <https://github.com/brianfrankcooper/YCSB/wiki>

⁴ <https://github.com/brianfrankcooper/YCSB/wiki/Adding-a-Database>

de conceito a utilização de dois tradicionais bancos de dados relacionais, sendo o MySQL e o PostgreSQL os selecionados.

O YCSB já vem pré-configurado com seis cargas de trabalho contemplando diferentes configurações de escrita e leitura de dados, podendo ser utilizado nos bancos de dados mencionados na Tabela 2. Além disso, qualquer pessoa pode modificar essas cargas de trabalho para configurações que mais se adaptem as necessidades de seus projetos.

A configuração de uma carga de trabalho é bastante simples e segue um formato pré-definido conforme ilustra a seguinte tabela.

```
# Yahoo! Cloud System Benchmark
# Workload Template: Default Values

# The name of the workload class to use
workload=com.yahoo.ycsb.workloads.CoreWorkload

# There is no default setting for recordcount but it is
# required to be set.
# The number of records in the table to be inserted in
# the load phase or the number of records already in the
# table before the run phase.
recordcount=1000000

# There is no default setting for operationcount but it is
# required to be set.
# The number of operations to use during the run phase.
operationcount=3000000

# The number of insertions to do, if different from recordcount.
# Used with insertstart to grow an existing table.
#insertcount=

# The offset of the first insertion
insertstart=0
```

The number of fields in a record

fieldcount=10

The size of each field (in bytes)

fieldlength=100

Should read all fields

readallfields=true

Should write all fields on update

writeallfields=false

The distribution used to choose the length of a field

fieldlengthdistribution=constant

#fieldlengthdistribution=uniform

#fieldlengthdistribution=zipfian

What proportion of operations are reads

readproportion=0.95

What proportion of operations are updates

updateproportion=0.05

What proportion of operations are inserts

insertproportion=0

What proportion of operations read then modify a record

readmodifywriteproportion=0

What proportion of operations are scans

scanproportion=0

On a single scan, the maximum number of records to access

maxscanlength=1000

The distribution used to choose the number of records to access on a scan

scanlengthdistribution=uniform

#scanlengthdistribution=zipfian

Should records be inserted in order or pseudo-randomly

insertorder=hashed

#insertorder=ordered

The distribution of requests across the keyspace

requestdistribution=zipfian

#requestdistribution=uniform

#requestdistribution=latest

Percentage of data items that constitute the hot set

hotspotdatafraction=0.2

Percentage of operations that access the hot set

hotspotopnfraction=0.8

Maximum execution time in seconds

#maxexecutiontime=

The name of the database table to run queries against

table=usertable

The column family of fields (required by some databases)

#columnfamily=

How the latency measurements are presented

measurementtype=histogram

#measurementtype=timeseries

The range of latencies to track in the histogram (milliseconds)

histogram.buckets=1000

Granularity for time series (in milliseconds)

timeseries.granularity=1000

Tabela 3 – Arquivo de criação das cargas de trabalho do YCSB.

O YCSB também permite que novas cargas de trabalho sejam criadas estendendo a classe `workload=com.yahoo.ycsb.workloads.CoreWorkload` de onde todos as cargas existentes são criadas. Portanto, para a prova de conceito do banco de dados do RIC, definimos seis cargas de trabalho personalizadas, detalhadas a seguir.

5.2.1 Carga A (Autenticação Biométrica)

Serviço de Referência	Autenticação Biométrica
Capacidade de Referência	Validar/ Verify
SLA esperado	7000/s
Proporção de Inserção (insert)	5%
Proporção de Atualização (update)	0%
Proporção de Exclusão (delete)	0%
Proporção de Consulta (Select)	95%
Quantidade de campos	1
Tamanho dos campos (em kb)	300
Quantidade de Registros no BD	100.000.000 de templates (escala de 5% de 2 bilhões)
Descrição	Operação de autenticação através da impressão digital. Os campos do banco de dados representam o template biométrico e o número de identificação do usuário. O tamanho estimado da base real é o número de brasileiros X 10 impressões digitais.
Comando	../bin/ycsb run jdbc -s -threads 30 -P ../workloads/workloada -P ../jdbc/mysql.properties > ArquivoSaida

Tabela 4 – Carga A (Autenticação Biométrica).

5.2.2 Carga B (Autenticação OTP/Usuário e Senha)

Serviço de Referência	Autenticação OTP/Usuário e Senha
Capacidade de Referência	Identificar / Identify
SLA esperado	7000/s
Proporção de Inserção (insert)	5%
Proporção de Atualização (update)	5%
Proporção de Exclusão (delete)	0%
Proporção de Consulta (Select)	90%
Quantidade de campos	2
Tamanho do campos (em kb)	1
Descrição	Autenticação simples através da recuperação da senha ou do endereço de e-mail/número do telefone celular para posterior envio do código OTP. Também testa a operação de validação do código OTP.
Comando	../bin/ycsb run jdbc -s -threads 16 -P ../workloads/POC-Teste02 -P ../jdbc/mysql.properties > WorkloadB

Tabela 5 – Carga B (Autenticação OTP/Usuário e Senha).

5.2.3 Carga C (Cadastramento das Impressões Digitais Brutas)

Serviço de Referência	Cadastramento das impressões digitais brutas
Capacidade de Referência	Cadastrar / Enroll
SLA esperado	14/s para cadastramento
Proporção de Inserção (insert)	50%
Proporção de Atualização (update)	0%
Proporção de Exclusão (delete)	0%
Proporção de Consulta (Select)	50%

Quantidade de campos	10
Tamanho do campos (em kb)	120
Quantidade de Registros no BD	200.000.000 (escala de 10% de 2 bilhões)
Descrição	O requisito aponta para o cadastramento de toda a população brasileira em dois anos. Levando em consideração os dias úteis, teria-se aproximadamente 400 mil operações de cadastramento por dia e aproximadamente 14 por segundo. “Todas” as impressões digitais armazenadas em seu formato bruto são transformadas em templates biométricos de forma assíncrona, por este motivo, a taxa de consultas é exatamente igual a de inserções.
Comando	../bin/ycsb run jdbc -s -threads 16 -P ../workloads/POC-Teste03 -P ../jdbc/mysql.properties > WorkloadC

Tabela 6 – Carga C (Cadastramento das Impressões Digitais Brutas).

5.2.4 Carga D (Cadastramento da Imagem Bruta da Íris)

Serviço de Referência	Cadastramento da Imagem bruta da íris
Capacidade de Referência	Cadastrar / Enroll
SLA esperado	14/s para cadastramento
Proporção de Inserção (insert)	50%
Proporção de Atualização (update)	0%
Proporção de Exclusão (delete)	0%
Proporção de Consulta (Select)	50%
Quantidade de campos	2
Tamanho do campos (em kb)	260
Quantidade de Registros no BD	40.000.000 (escala de 10% de 400 milhões)
Descrição	O requisito aponta para o cadastramento de toda a população brasileira em dois anos. Levando em consideração os dias úteis, teria-se aproximadamente 400 mil operações de cadastramento por dia e aproximadamente 14 por segundo. “Todas” as íris armazenadas em seu formato bruto são transformadas em templates biométricos de forma assíncrona, por este motivo, a taxa de consultas é exatamente igual a de inserções.

Comando	<pre>../bin/ycsb run jdbc -s -threads 16 -P ../workloads/POC-Teste04 -P ../jdbc/mysql.properties > WorkloadD</pre>
---------	---

Tabela 7 – Carga D (Cadastramento da Imagem Bruta da Íris).

5.2.5 Carga E (Cadastramento da Imagem da Carteira de Identidade)

Serviço de Referência	Cadastramento da Imagem da Carteira de Identidade
Capacidade de Referência	Cadastrar / Enroll
SLA esperado	14/s para cadastramento
Proporção de Inserção (insert)	99%
Proporção de Atualização (update)	0%
Proporção de Exclusão (delete)	0%
Proporção de Consulta (Select)	1%
Quantidade de campos	2
Tamanho do campos (em kb)	1000
Quantidade de Registros no BD	20.000.000 (escala de 10% de 200 milhões)
Descrição	O requisito aponta para o cadastramento de toda a população brasileira em dois anos. Levando em consideração os dias úteis, teria-se aproximadamente 400 mil operações de cadastramento por dia e aproximadamente 14 por segundo. “
Comando	<pre>../bin/ycsb run jdbc -s -threads 16 -P ../workloads/POC-Teste05 -P ../jdbc/mysql.properties > WorkloadE</pre>

Tabela 8 – Carga E (Cadastramento da Imagem da Carteira de Identidade).

5.2.6 Carga F (Recuperação dos Dados Biográficos)

Serviço de Referência	Recuperação dos Dados Biográficos
Capacidade de Referência	Cadastrar / Enroll
SLA esperado	7000/s

Proporção de Inserção (insert)	10%
Proporção de Atualização (update)	10%
Proporção de Exclusão (delete)	0%
Proporção de Consulta (Select)	80%
Quantidade de campos	15
Tamanho do campos (em kb)	10
Quantidade de Registros no BD	20.000.000 (escala de 10% de 200 milhões)
Descrição	A recuperação das informações biográficas tem comportamento complementar ao processo de autenticação. Entretanto, ao tempo que se faz a recuperação, novos cadastros continuam sendo inseridos e outros atualizados.
Comando	../bin/ycsb run jdbc -s -threads 16 -P ../workloads/POC-Teste06 -P ../jdbc/mysql.properties > WorkloadF

Tabela 9 – Carga F (Recuperação dos Dados Biográficos).

6 INFRAESTRUTURA

Nesse item, apresenta-se de forma detalhada a infraestrutura física e lógica sobre a qual será realizada a prova de conceito. A fim de suportar os serviços descritos no Capítulo 3, foi desenhada uma arquitetura de alto nível, composta por componentes que permitem simular de forma simplificada e controlada os sistemas envolvidos no registro, cadastramento e autenticação de um cidadão, em um eventual cenário real.

Assim, se procurou incorporar componentes que representassem, na prova de conceito, sistemas reais que participarão na futura implementação do Programa RIC, contemplando aqueles atualmente já existentes e os que virão a existir, dependendo da composição final da arquitetura dos requisitos finais e das soluções técnicas e organizativas que venham a ser adoptadas na implementação real do Programa.

A Figura 1 ilustra a arquitetura tecnológica que orientou a implementação da prova de conceito do projeto RIC.

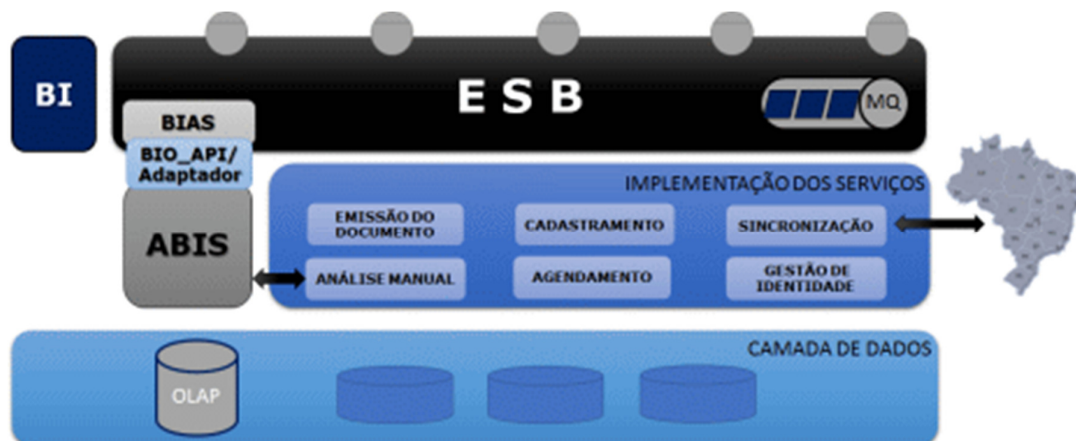


Figura 1 - Arquitetura Tecnológica da Prova de Conceito

Na arquitetura proposta, foram incluídas algumas simplificações, a fim de permitir a sua concretização no contexto da prova de conceito, por definição limitada no tempo e nos recursos disponíveis. Assim, foram tomadas simplificações, por exemplo, as seguintes:

- utilização de comunicações locais, dentro de uma mesma rede local;
- simplificação dos mecanismos de segurança;

Projeto: MJ/SE-RIC	Emissão: 30/08/2015	Arquivo: 20150830 MJ RIC - RT Planejamento da Prova de Conceito - Banco de Dados	Pág.38/43
--------------------	---------------------	--	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

- limitado número de registros.

Entretanto, alguns aspectos da arquitetura se aproximam bastante de uma situação realista, tais como:

- interação entre distintos serviços baseada em contratos bem definidos e protocolos baseados em padrões tecnológicos;
- implementação dos serviços em plataformas tecnológicas heterogêneas em diversos níveis (incluindo sistemas operacionais, bancos de dados, servidores, linguagens de programação e mecanismos de comunicação).

6.1 Infraestrutura Física

Nesse item, apresenta-se de forma detalhada a infraestrutura física sobre a qual será realizada a prova de conceito. A prova de conceito conta com um parque de *hardware* de 12 servidores, com as seguintes características e especificações técnicas:

<i>Hardware</i>			
Servidor	Qtde.	Marca e Modelo	Configuração
PowerEdge R730 Tipo I	2	Intel	Intel Xeon E52630V3, 2.4GHz, 20M Cache; Memoria: 128G; HD: 600G (8 discos), 10k.
PowerEdge R520 Tipo II	10	Intel	Intel Xeon E52420V2, 2,2GHz, 15M Cache; Memoria: 16 GB; SDRAM DDR3; 1333 MHz (PC3-10600); HD: 5 discos SAS de 2 TB e 7200 RPM.

Tabela 10 – Infraestrutura Física.

Dado o elevado número de funcionalidades presentes em um sistema de porte nacional, optou-se por utilizar, como suporte para a esta infraestrutura, a tecnologia de máquinas virtuais. Tal tecnologia permite que, em uma mesma máquina ou servidor, possam existir várias outras máquinas, denominadas virtuais, compartilhando os recursos físicos da máquina em questão. Na prova de conceito, foi utilizado o *software* VirtualBox da

fabricante Oracle. Dessa forma, os servidores estão dispostos nas seguintes configurações, a saber.

Máquinas Tipo I

	VM1	VM2	VM3	VM4	VM5	VM6	SVR1
Núcleos	3	3	3	6	6	6	12
Memória	12GB	36GB	36GB	12GB	36GB	36GB	48GB
Disco	SATA	SATA	RAID	SATA	SATA	RAID	RAID

Máquinas Tipo II

	VM7	VM8	VM9	VM10	SVR2
Núcleos	6	16	16	16	32
Memória	36GB	36GB	52GB	104GB	128GB
Disco	SAS	SAS	SAS	SAS	SAS

6.2 Arquitetura Lógica

Nesse item, apresentam-se os *softwares* utilizados em cada uma das etapas da prova de conceito.

Lógico	Características
Sistema Operacional	CentOS versão 6.6, kernel 2.6.32-504.16.2.el6.x86_64
Ferramenta de <i>Benchmarking</i>	Yahoo Cloud Service Benchmarking (YCSB), versão 0.1.4
Banco de Dados	MySQL Community Server 5.6.25, MemSQL
Virtualbox	Versão 5.0

7 CONCLUSÃO

O resultado de uma primeira avaliação deste documento permitiu o direcionamento inicial de uma Prova de Conceito (PoC) mínima, visando nortear a avaliação de quesitos específicos de infraestrutura, tais como: os produtos e ferramentas que suportam os modelos e arquétipos sugeridos, as questões de desempenho e curvas de ruptura na entrega de serviços, a identificação de requisitos mínimos de *hardware* de infraestrutura de armazenamento, entre outros.

É fator fundamental dessa Prova de Conceito, também, comprovar a eficácia e eficiência do modelo, dada a capacidade de resposta de um número de acesso previsto para o registro da totalidade de cidadãos.

Projeto: MJ/SE-RIC	Emissão: 30/08/2015	Arquivo: 20150830 MJ RIC - RT Planejamento da Prova de Conceito - Banco de Dados	Pág.41/43
--------------------	---------------------	--	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.

É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

REFERÊNCIAS

- [1] RT Estudos e Análise, Modelagem e Projeto de Serviços e de Modelos Canônico de Dados.
- [2] Cooper, Brian F., et al. "Benchmarking cloud serving systems with YCSB." Proceedings of the 1st ACM symposium on Cloud computing. ACM, 2010.

Projeto: MJ/SE-RIC	Emissão: 30/08/2015	Arquivo: 20150830 MJ RIC - RT Planejamento da Prova de Conceito - Banco de Dados	Pág.42/43
--------------------	---------------------	--	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a MJ/SE.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da MJ/SE.

Universidade de Brasília – UnB

Centro de Apoio ao Desenvolvimento Tecnológico – CDT

Laboratório de Tecnologias da Tomada de Decisão – LATITUDE

www.unb.br – www.cdt.unb.br – www.latitude.eng.br

