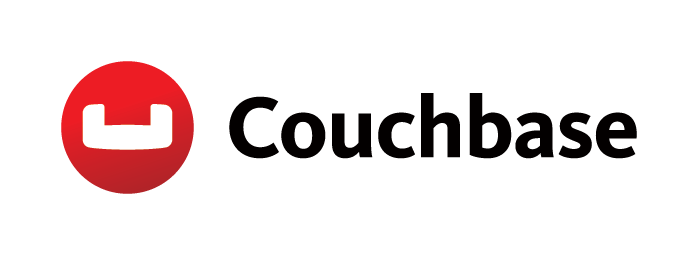


**MIT EM BIG DATA**

**BLOCO C – ARMAZENAMENTO HETEROGÊNEO DE DADOS**



**Anexos**

**Alunos: Milton Bahia**

**Roberto Rio**

**Vinicius Rosa**

**Sumário**

[ANEXO 1 - Teste de Stress 2](#_Toc451368969)

[ANEXO 2 -Integração com Ferramentas de Análises de Dados 5](#_Toc451368970)

[ANEXO 3 – Instalação e desinstalação do Couchbase Server 26](#_Toc451368975)

# ANEXO 1 - Teste de Stress

Executamos alguns testes de stress inserindo e lendo 1 milhão de registros no couchbase e para termos um parâmetro comparamos com um banco relacional e também testes comparativos entre standalone servers e servers em cluster. Para realização dos testes foi utilizado a ferramenta YCSB(Yahoo Clouding Serving Benchmark) para que pudéssemos ter um resultado menos parcial que utilizar utilitários de carga de cada ferramenta, correndo o risco de tornar os resultados tendenciosos. Abaixo seguem os resultados encontrados.

* Comparativos de troughput entre servidores standalone e cluster
* Comparativo de tempo de resposta entre servidores standalone e cluster
* Comparativos de troughput entre servidores Couchbase e MySQL
* Comparativo de tempo de resposta entre servidores Couchbase e MySQL

# ANEXO 2 -Integração com Ferramentas de Análises de Dados

O objetivo desta atividade foi integrar as ferramentas de análise de dados apresentadas no Bloco B com o CouchBase, abrangendo o ciclo completo de obtenção de dados, armazenamento, visualização e análise inicial dos dados.

3. **Obtenção dos dados**

Para esta atividade, utilizamos um conjunto de dados de Unidades Básicas de Saúde obtidos no Portal de Dados abertos do Governo Federal.

<http://dados.gov.br/dataset/unidades-basicas-de-saude-ubs>



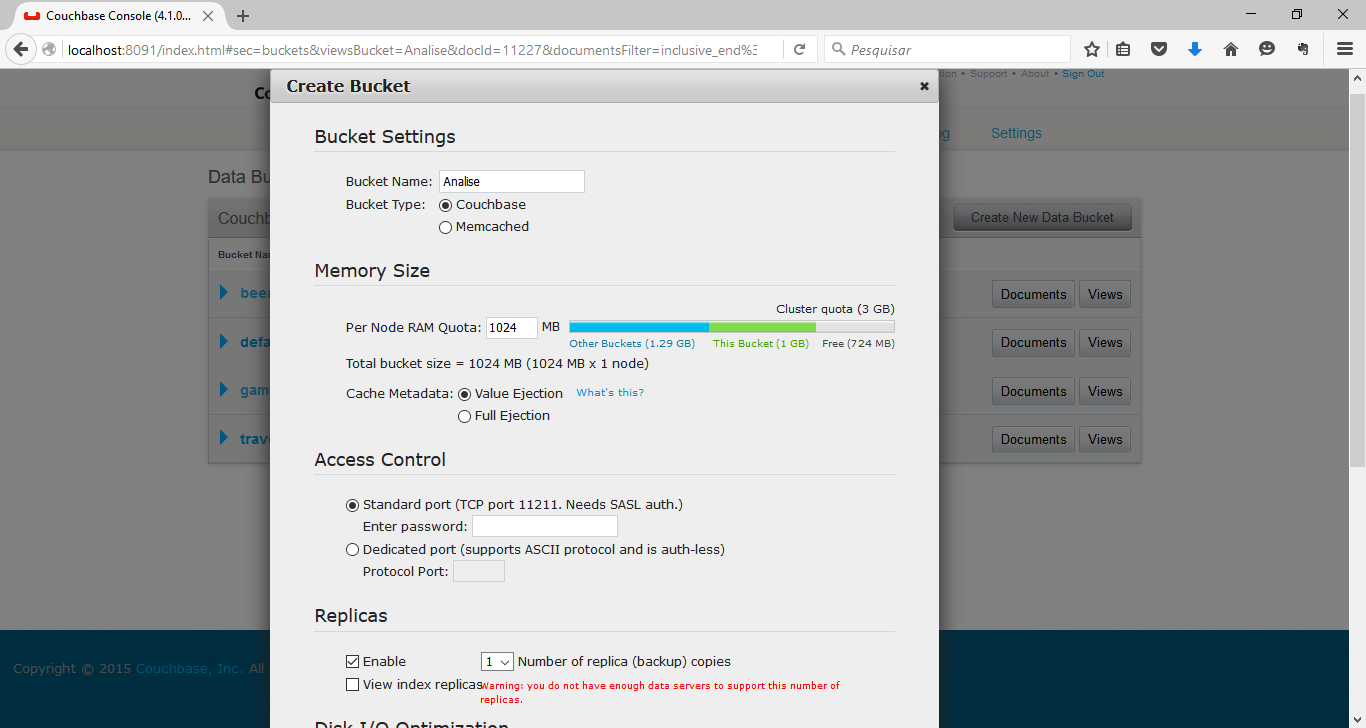
O conjunto de dados é disponibilizado em formato “csv”, juntamente com o dicionário de dados (vide tabela abaixo).

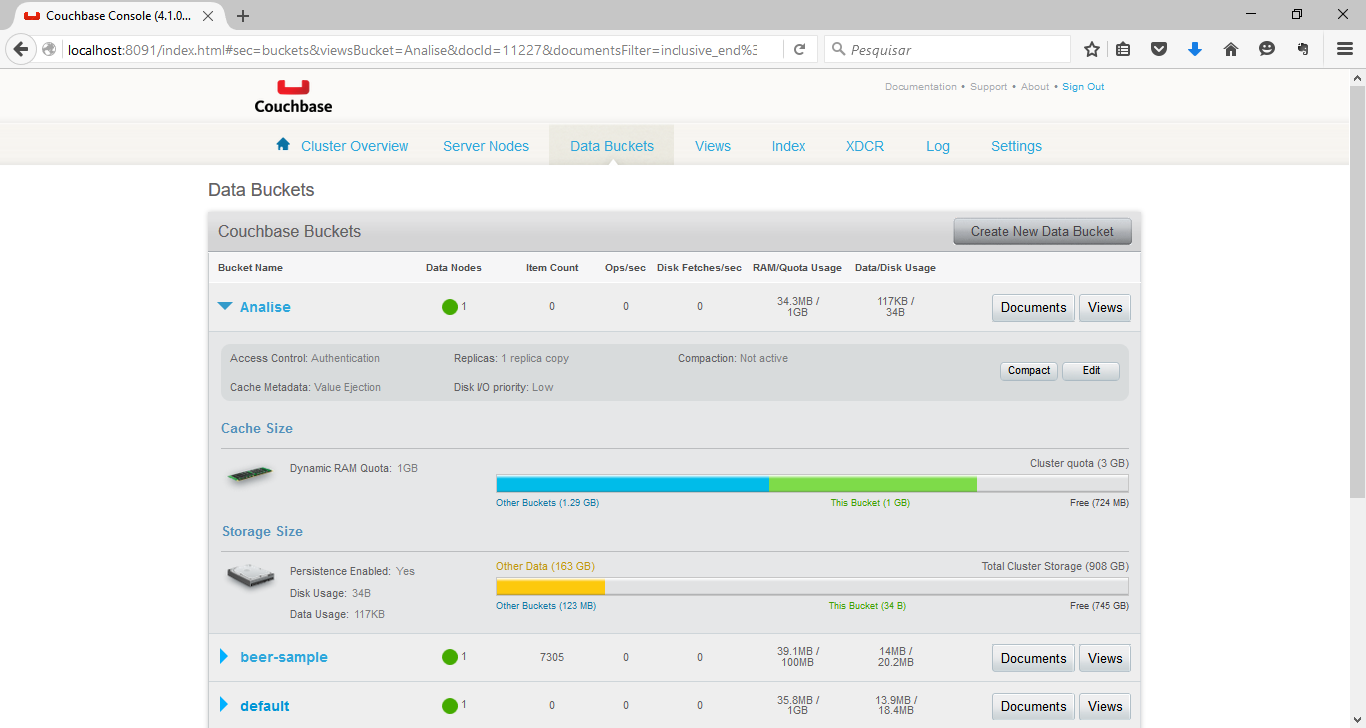
| **Nome do campo** | **Rótulo do campo CSV** | **Descrição** |
| --- | --- | --- |
| Latitude | vlr\_latitude | Representa a latitude em formato decimal. Considerar valores negativos.O sistema de referência espacial é o EPSG:4326. |
| Longitude | vlr\_longitude | Representa a longitude em formato decimal. Considerar valores negativos.O sistema de referência espacial é o EPSG:4326. |
| Código do Município | cod\_munic | Representa o código IBGE do município. |
| Código CNES | cod\_cnes | Representa o código CNES |
| Nome do Estabelecimento | nom\_estab | Representa o nome do estabelecimento. |
| Endereço | dsc\_endereco | Representa o endereço da unidade. |
| Bairro | dsc\_bairro | Representa o nome do Bairro. |
| Cidade | dsc\_cidade | Representa o nome da cidade |
| Telefone | dsc\_telefone | Representa o número do telephone |
| Situação em relação a estrutura física e ambiência | dsc\_estrut\_fisic\_ambiencia | Informa a situação em relação a estrutura física e ambiência |
| Situação em relação a adaptações para deficientes e idosos | dsc\_adap\_defic\_fisic\_idosos | Informa a situação em relação a adaptações para deficientes e idosos |
| Situação em relação aos equipamentos | dsc\_equipamentos | Informa a situação em relação aos equipamentos |
| Situação em relação aos Mediacamentos | dsc\_medicamentos | Informa a Situação em relação aos medicamentos |

1. **Criação de bucket para armazenamento dos dados**

Conforme arquitetura do CouchBase, cada conjunto de documentos pertencente a uma solução ou aplicação deve ser armazenado em um bucket.

Para esta atividade foi criado o bucket “Analise” através da interface de Administração Web.





1. **Importação de Dados**

Para importação de dados foi utilizada a ferramenta cbdocloader disponível no pacote do CouchBase. Esta ferramenta carrega um conjunto de documentos no formato JSON armazenados em uma pasta ou arquivo compactado (“.Zip”) para um bucket do Couchbase.

Referências:

* <http://developer.couchbase.com/documentation/server/4.1/cli/cbdocloader-tool.htm>
* <http://blog.couchbase.com/loading-json-data-couchbase>

Esta ferramenta é executada em linha de comando do sistema operacional.

**Conversão do conjunto de dado para o formato JSON**

Como o arquivo original era em formato CSV, foi necessário executar um procedimento de conversão para o formato JSON.

O cbdocloader não consegue importar um arquivo com múltiplos documentos JSON. Em função dessa característica, cada registro (linha) presente no arquivo original (ubs.csv) foi gravado em um arquivo distinto em formato JSON.

Segue o script em python utilizado para a preparação de dados para carga.

*Conteúdo do Arquivo converterjson.py*

import csv

import json

csvfile = open('C:\Users\Milton\Documents\MIT BigData\Bloco C\Couchbase\ubs.csv', 'r')

reader = csv.DictReader(csvfile)

i = 1

for row in reader:

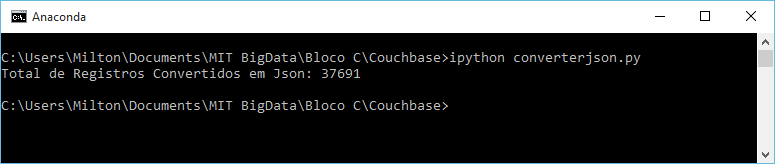
outfile = open('C:/Users/Milton/Documents/MIT BigData/Bloco C/Couchbase/saude/' + row["cod\_cnes"] + '.json', 'w')

json.dump(row, outfile)

outfile.close()

i += 1

print "Total de Registros Convertidos em Json: " + str(i)



**Importação dos Arquivos**

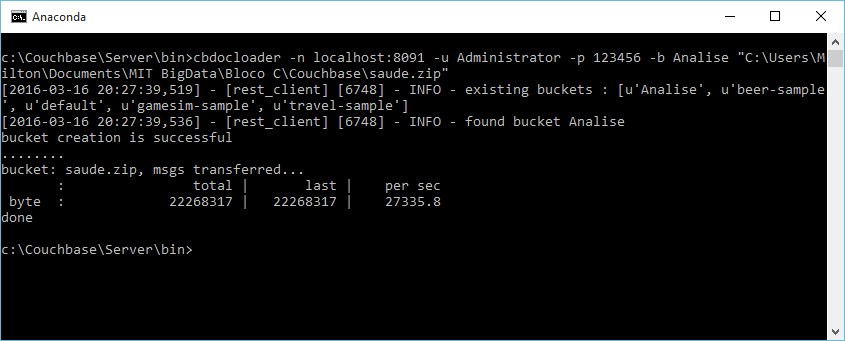
Após a conversão do arquivo original em formato CSV para múltiplos arquivos JSON, foi executado o procedimento de carga.

Segue sintaxe básica do cbdocloader:

cbdocloader -n [host]:8091 -u [Administrator] -p [password] -b [bucket-name] [source-file]

Segue comando utilizado para importação:

cbdocloader -n localhost:8091 -u Administrator -p 123456 -b Analise "C:\Users\Milton\Documents\MITBigData\Bloco C\Couchbase\saude.zip"



Observações:

* O arquivo saude.zip continha os arquivos convertidos em formato JSON.

Os dados foram carregados no bucket “Analise”

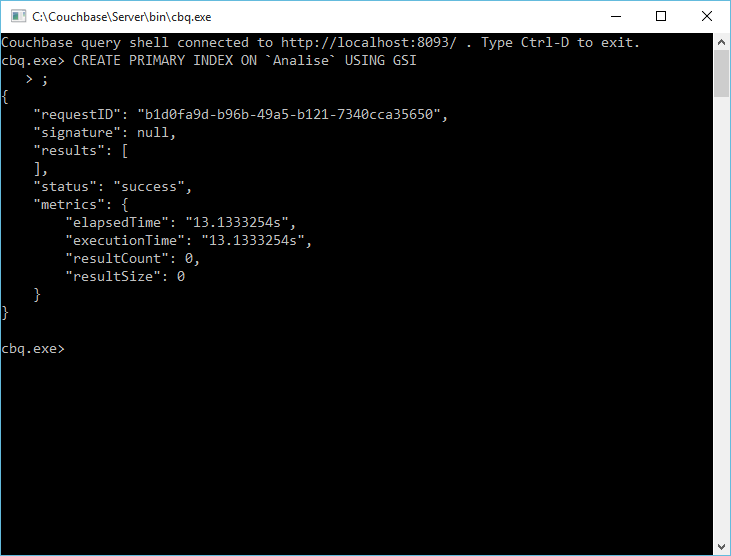
1. **Consultas N1QL na ferramenta CBQ**

Após o procedimento de importação foram realizadas algumas consultas diretamente na ferramenta de comando de linha CBQ. Esta ferramenta permite a execução de consultas N1QL.

1. Criação do Índice primário para o bucket Analise.

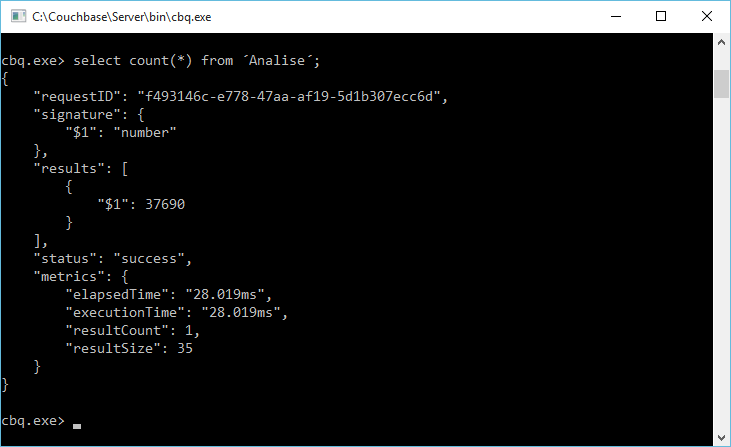
Para execução de consultas N1QL foi necessário criar o índice primário do bucket Analise.

Comando: CREATE PRIMARY INDEX ON `Analise` USING GSI



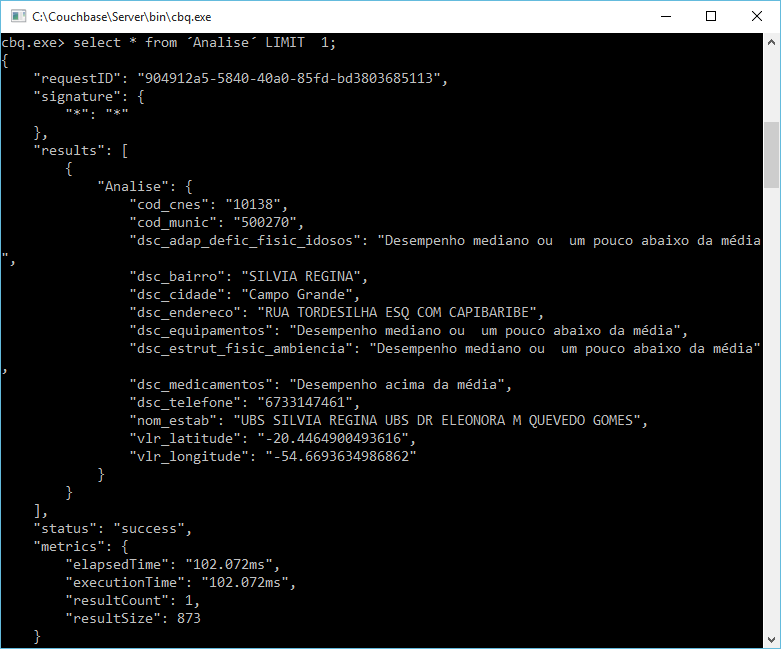
1. Verificação do total de documentos importados

Consulta: “Select count(\*) from ´Analise´;



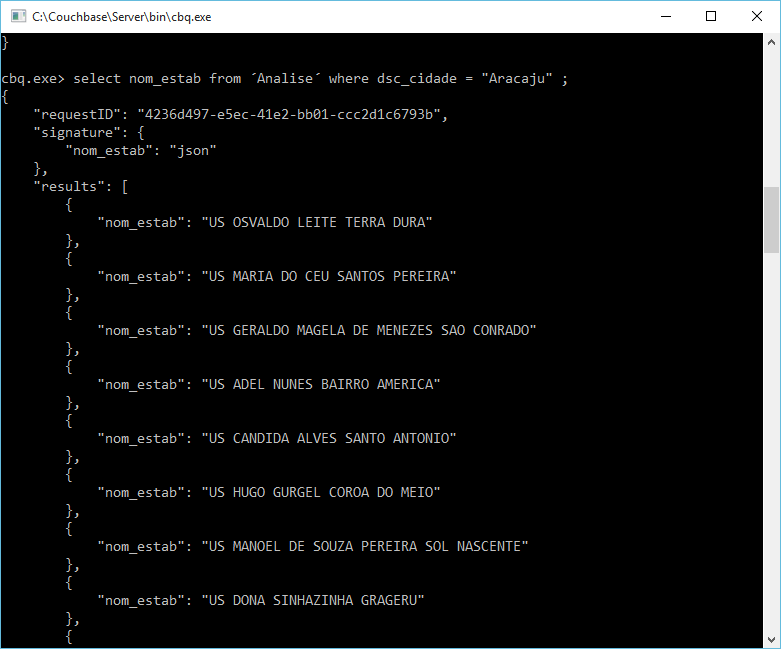
1. Consulta de um documento para verificação da sua estrutura.

Consulta: select \* from ´Analise´ LIMIT 1;



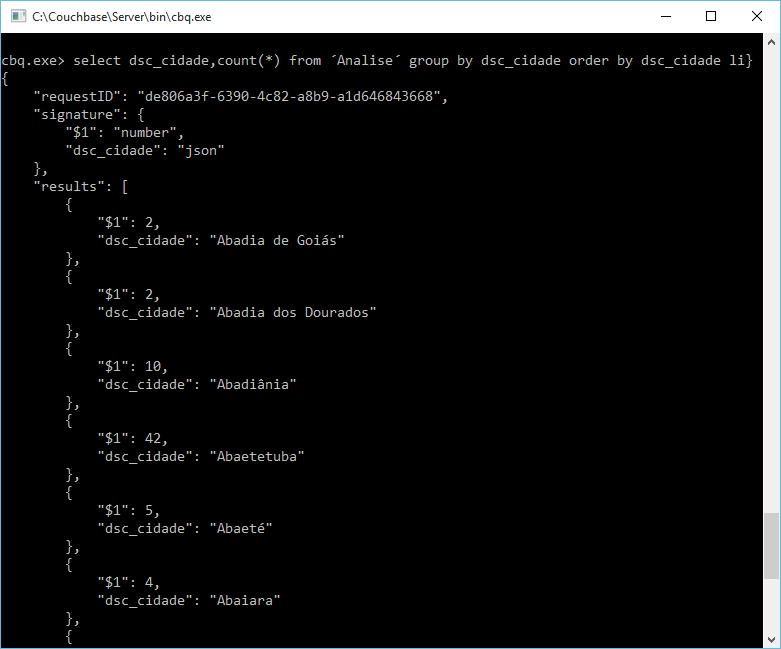
1. Consulta de todos os estabelecimentos na cidade de Aracaju

Consulta: select nom\_estab from ´Analise´ where dsc\_cidade = "Aracaju" ;



1. Total de postos por cidade

Consulta: select dsc\_cidade,count(\*) from ´Analise´ group by dsc\_cidade order by dsc\_cidade;



A clausula adicional “LIMIT 10” foi inserida para permitir a captura da tela.

1. **Criação de Views**

Foram criadas as seguintes views a partir dos dados importados no bucket Analise:

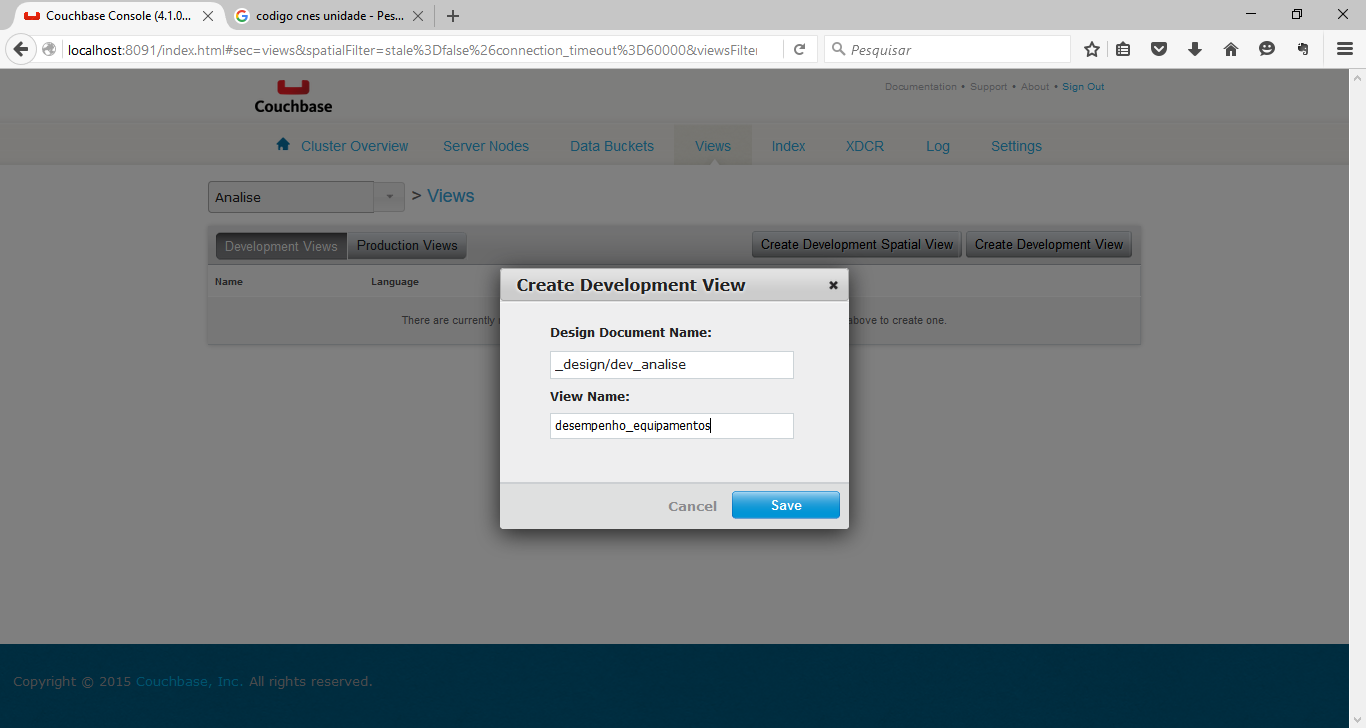
1. View desempenho\_equipamentos
2. View desempenho\_medicamento
3. Views postos

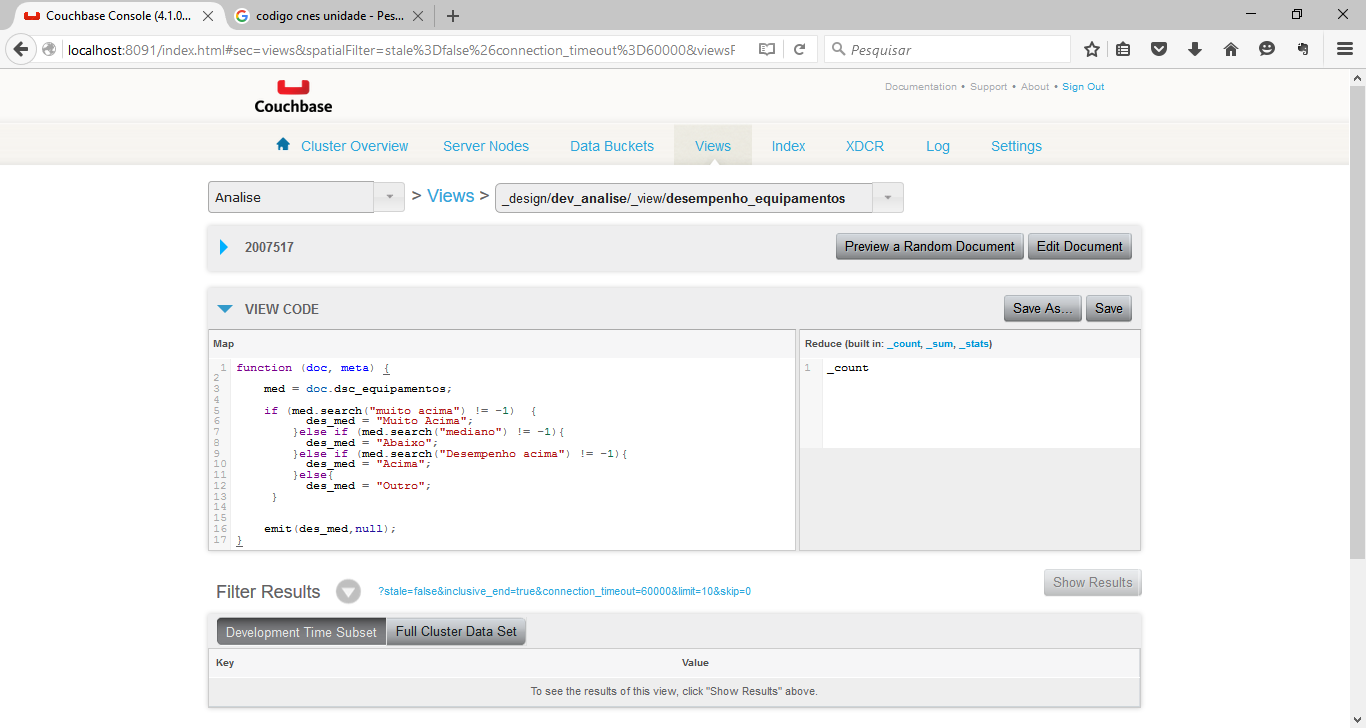
As views foram criadas segundo o procedimento indicado na documentação do Couchbase, primeiramente como views de desenvolvimento e posteriormente publicadas como views de produção.

**View desempenho\_equipamentos**

Apresenta a contagem de postos em cada status de situação em relação dos equipamentos: Muito Acima da Média, Acima da Média e Abaixo da Média.

1. Telas de Criação





1. Código da View

map

function (doc, meta) {

med = doc.dsc\_equipamentos;

if (med.search("muito acima") != -1) {

des\_med = "Muito Acima";

}else if (med.search("mediano") != -1){

des\_med = "Abaixo";

}else if (med.search("Desempenho acima") != -1){

des\_med = "Acima";

}else{

des\_med = "Outro";

}

emit(des\_med,null);

}

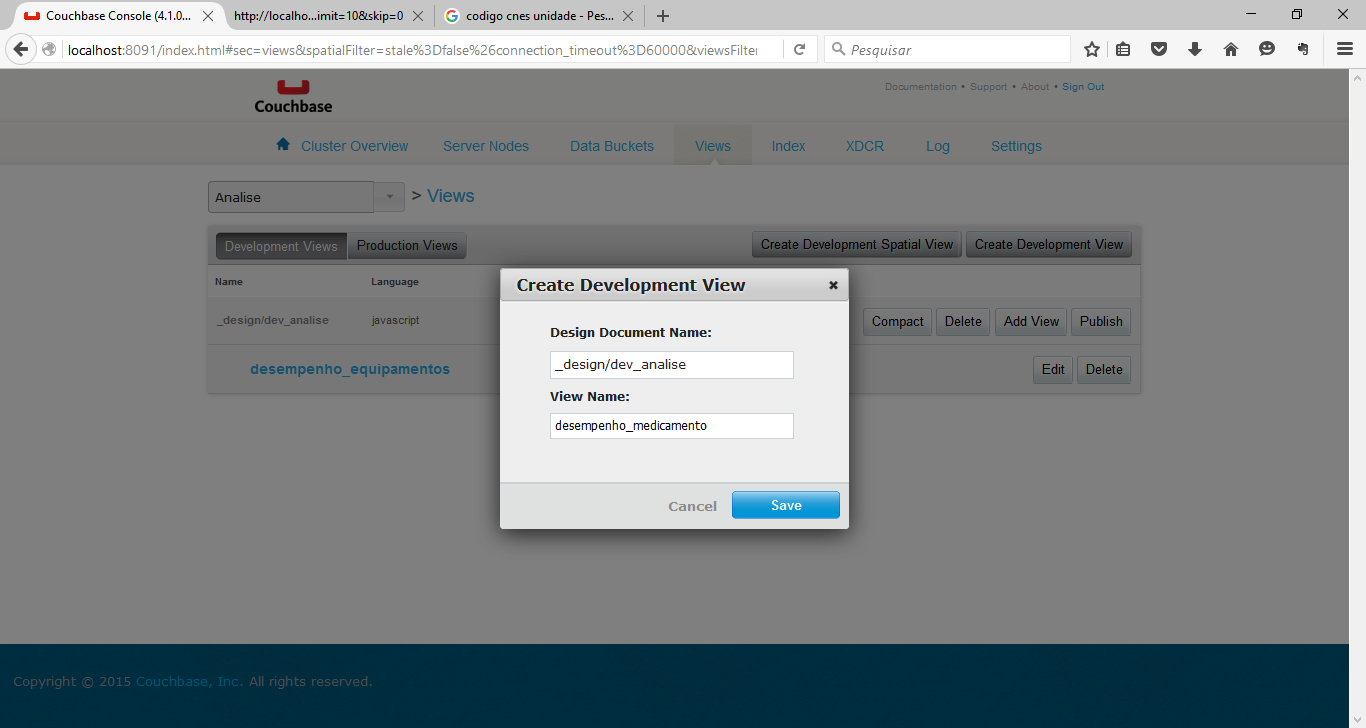
Reduce

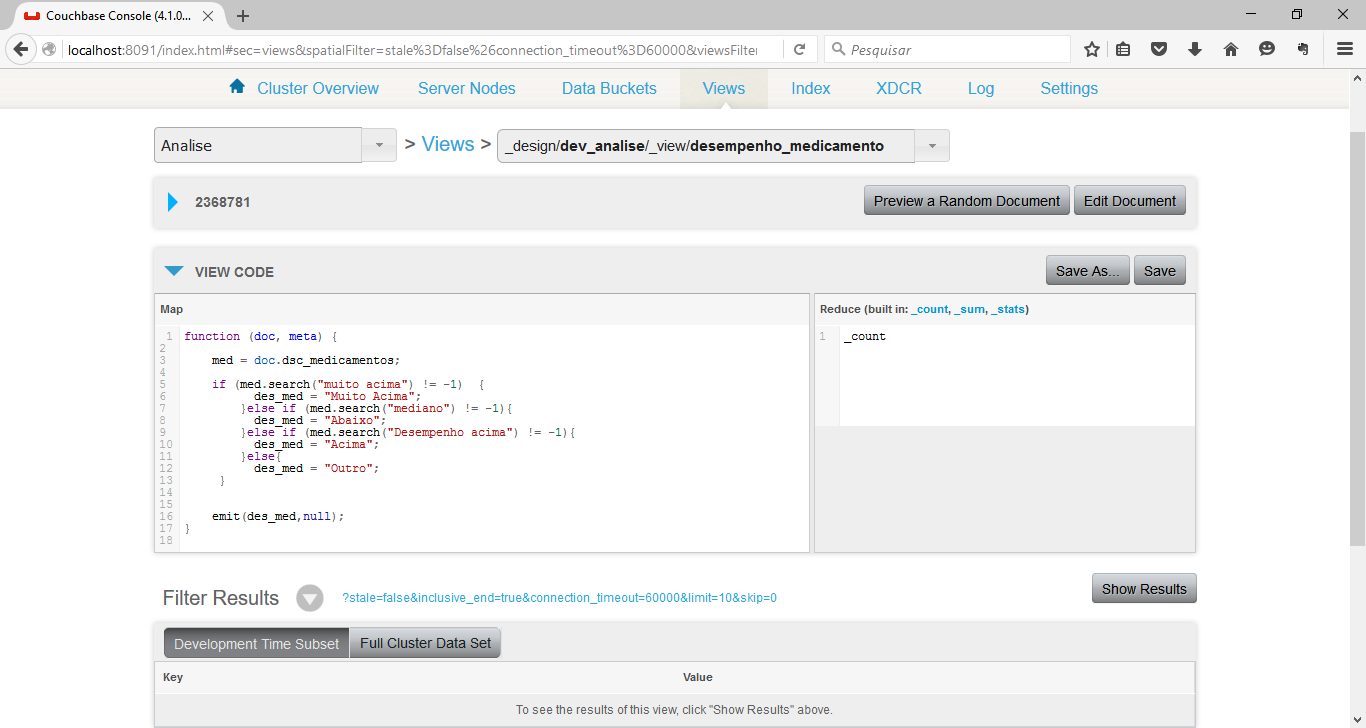
\_count

**View desempenho\_medicamento**

Apresenta a contagem de postos em cada status de situação em relação aos medicamentos: Muito Acima da Média, Acima da Média e Abaixo da Média

1. Telas de Criação





1. Código da View

Map

function (doc, meta) {

med = doc.dsc\_medicamentos;

if (med.search("muito acima") != -1) {

des\_med = "Muito Acima";

}else if (med.search("mediano") != -1){

des\_med = "Abaixo";

}else if (med.search("Desempenho acima") != -1){

des\_med = "Acima";

}else{

des\_med = "Outro";

}

emit(des\_med,null);

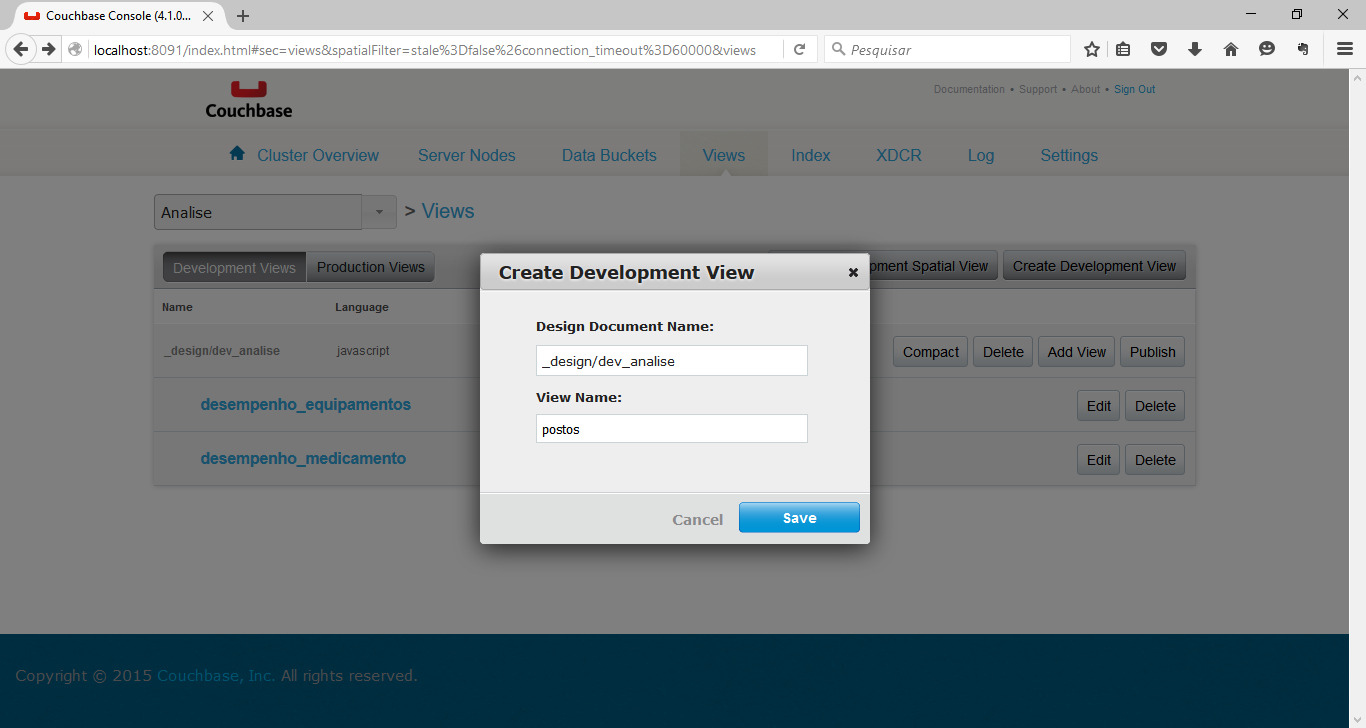
}

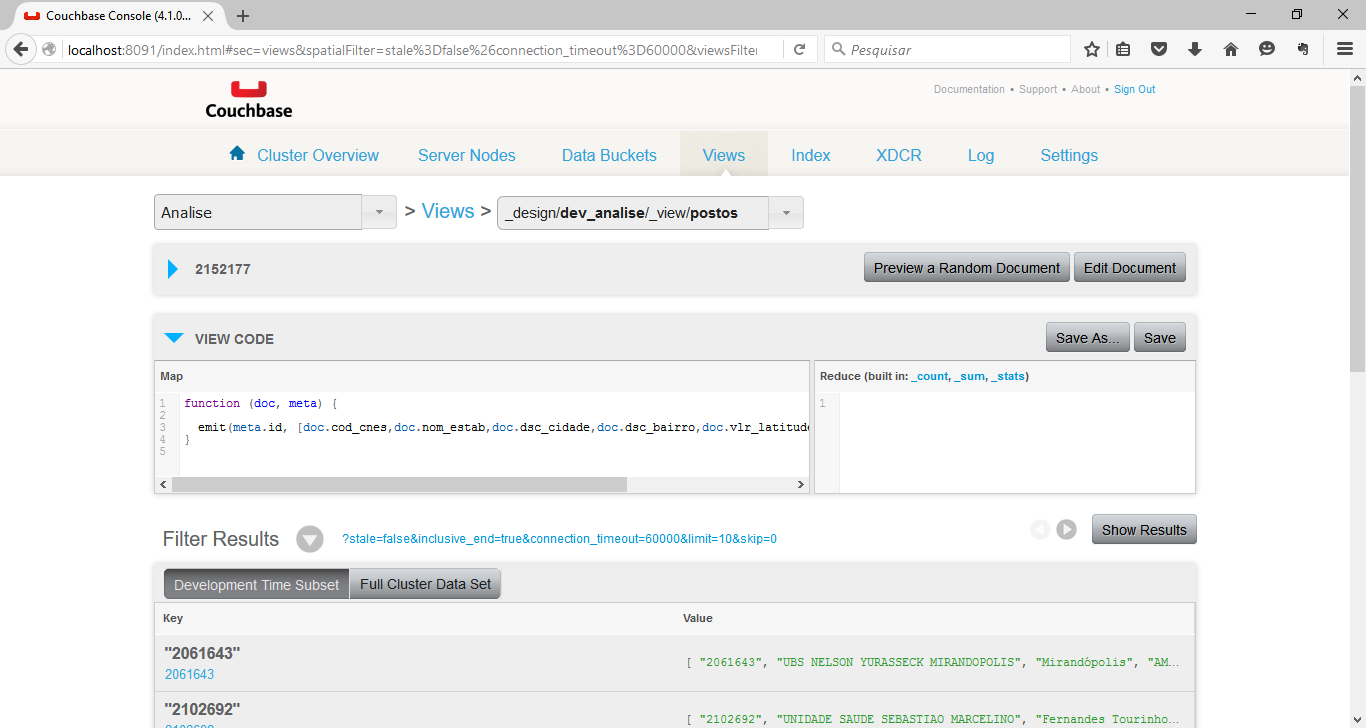
Reduce

\_count

**View postos**

Apresenta os dados das unidades básicas de saúde: cod\_cnes, nom\_estab, cidade, bairro, latitude e longitude.

1. Telas de criação



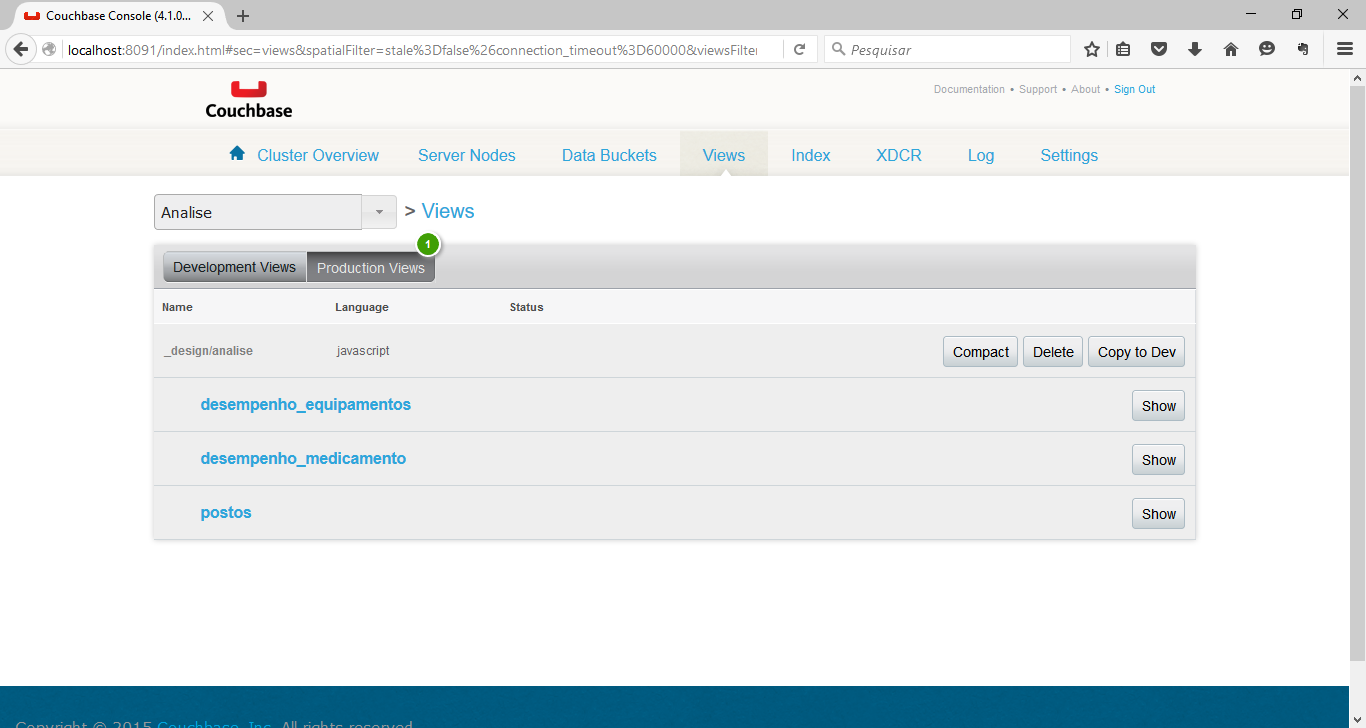
1. Código da View

Map

Fun­ction (doc, meta) {

emit(meta.id, [doc.cod\_cnes,doc.nom\_estab,doc.dsc\_cidade,doc.dsc\_bairro,doc.vlr\_latitude,doc.vlr\_longitude]); }

**Publicação das views em produção**



1. **Análise com Iphyton Notebook**

Através do IPhyton Notebook foram feitas algumas consultas explorando as diferentes formas de acesso disponíveis no Couchbase.

1. Consultas pela chave
2. Consultas N1QL
3. Consultas Map/Reduce Views

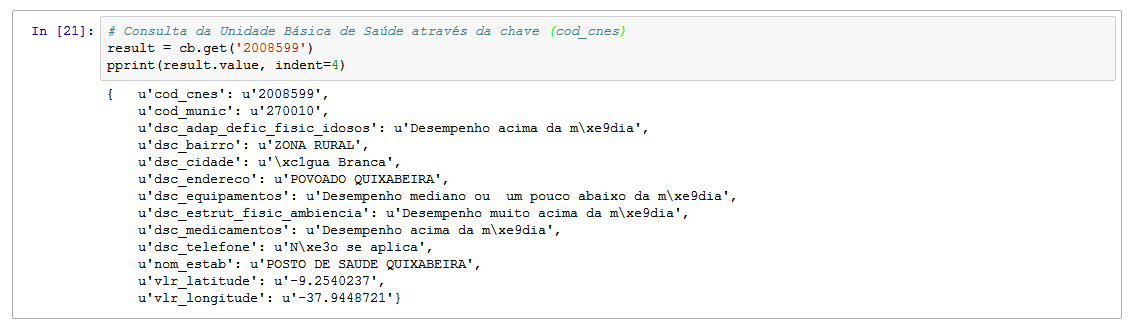
Neste documento apresentaremos apenas alguns trechos de código, entretanto o Notebook completo seguirá em anexo.

**Conexão com o Couchbase**

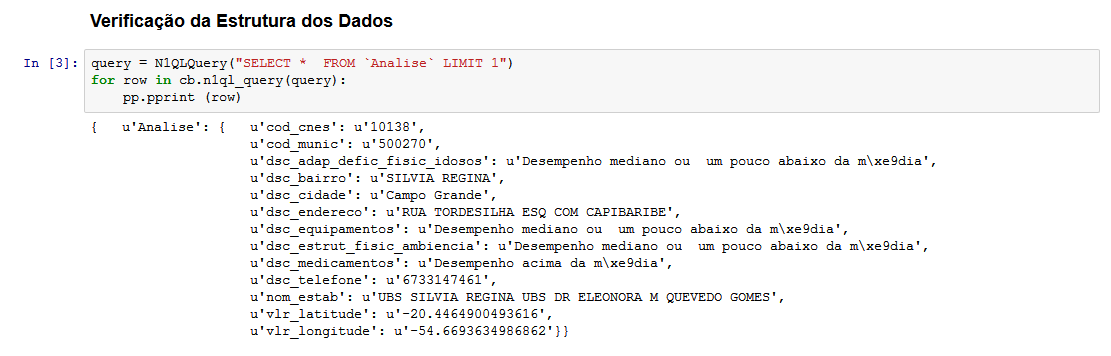


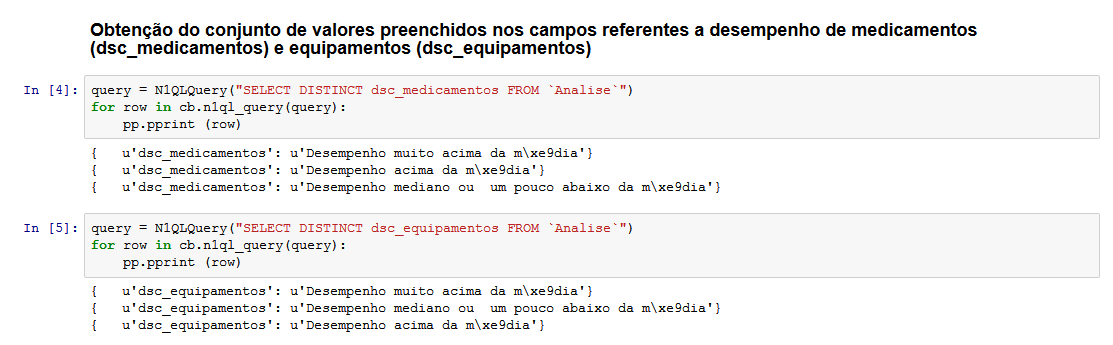
**Consulta através da chave**

Os dados das Unidades Básicas de Saúde foram importados como documentos no bucket Analise utilizando o Código CNES como Document\_ID.

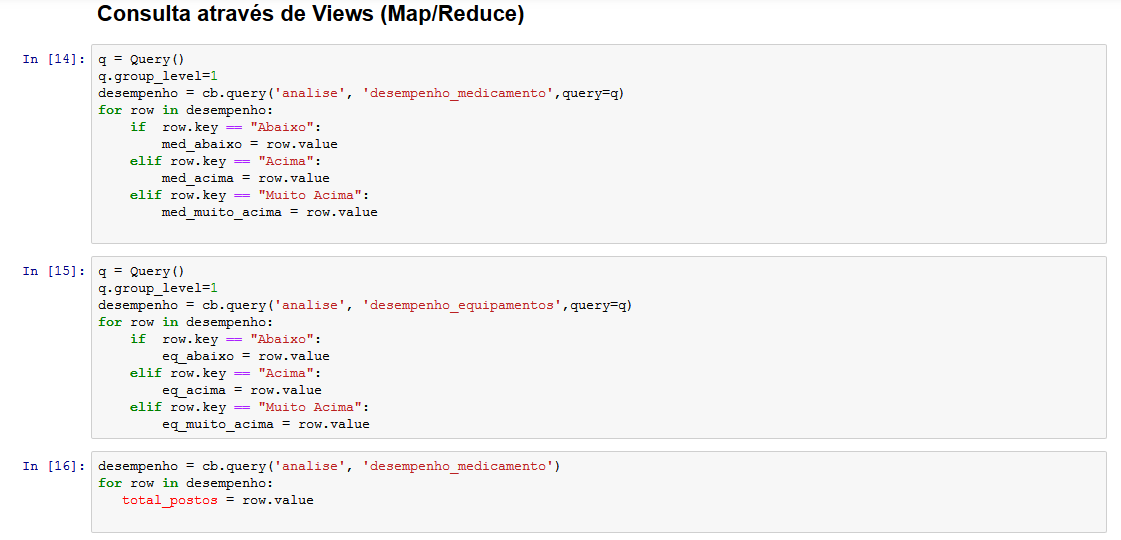


**Consultas N1QL**



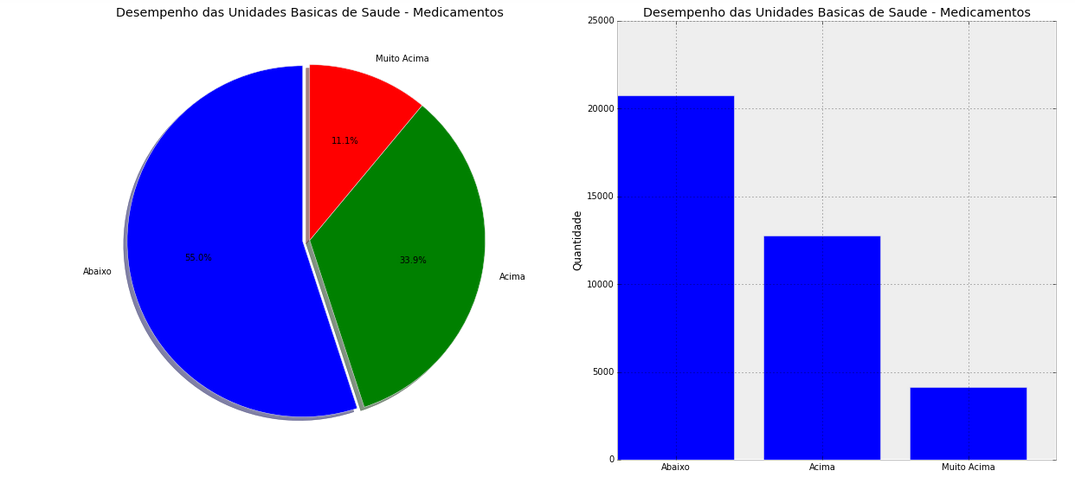


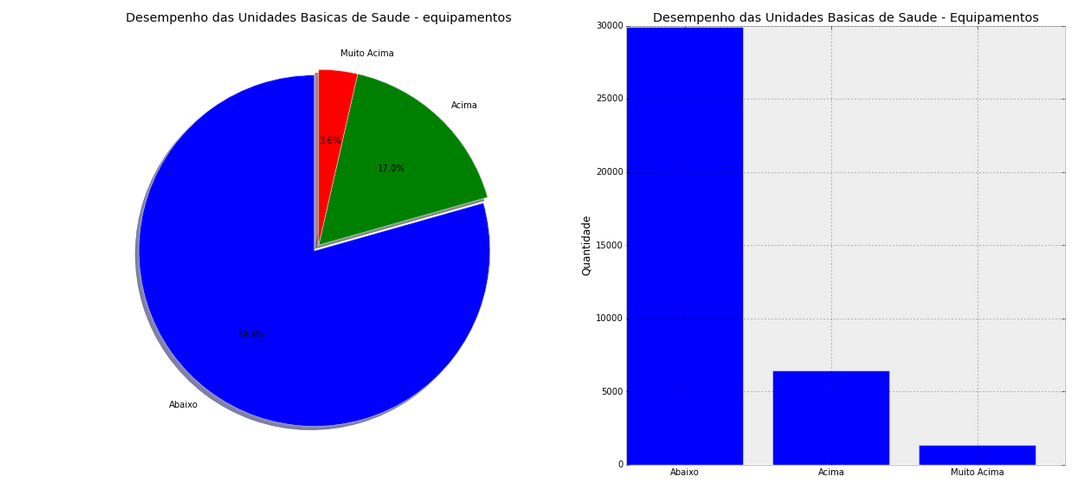
**Consulta através Map/Reduce Views**



**Geração de Gráficos de Desempenho de UBS**

Através dos resultados das consultas (N1QL / Views) foram gerados gráficos de desempenho das Unidades Básicas de Saúde.





**Visualização dos dados básicos das Unidades Básicas de Saúde em formato tabular**

Os dados das Unidades Básicas de Saúde foram consultados através da view postos. O resultado da view foi armazenado em um Dataframe (Pandas).



1. **Análise com o R**

Como ainda não existe um pacote R para acesso direto ao Couchbase, a consulta aos documentos armazenados pode ser realizada através da API REST. As views criadas no Couchbase são consultadas através dessa API. O formato da URL de consulta utiliza o seguinte padrão de formação:

http://[localhost]:8092/[bucket-name]/\_design/[ddoc-name]/\_view/[view-name]?[Parametros]

Como exemplo, segue URL de consulta da View de Desempenho de Equipamentos:

<http://localhost:8092/Analise/_design/analise/_view/desempenho_equipamentos?group_level=1>

A principal referência para realização desta atividade foi o artigo a seguir disponível no blog oficial do Couchbase:

* <http://blog.couchbase.com/incremental-mapreduce-analytics-r>

Através do RStudio foi feita uma reprodução parcial das consultas realizadas o IPython

* Consulta da View Desempenho de Equipamentos
* Geração de Gráfico de Barras do Desempenho de Equipamentos das Unidades Básicas de Saúde
* Consulta aos dados da Unidades Básicas de Saúde através da View Postos.

Foram utilizados os seguintes pacotes do R:

* Rjson
* Plyr
* Graphics

Segue Código do arquivo testecouchbase.R:

library("rjson", lib.loc="~/R/win-library/3.2")

# URL Base das Views do bucket Analise

urlBase <- 'http://localhost:8092/Analise'

# Função para obter os dados

getData <- function(subpath) {

fromJSON(file=paste(urlBase, subpath, sep=''))$rows

}

# Função que transforma os Dados (Json) em Dataframe

getFlatData <- function(sub, n=NULL) {

b <- plyr::ldply(getData(sub), unlist)

if (!is.null(n)) {

names(b) <- n

}

b

}

#Consulta a view de desempenho de equipamentos das Unidades Básicas de Saude

desempenho <- getFlatData('/\_design/analise/\_view/desempenho\_equipamentos?group\_level=1',c('desempenho','count'))

desempenho$desempenho <- factor(desempenho$desempenho)

desempenho$count <- as.numeric(desempenho$count)

# Plotar Grafico de Desempenho

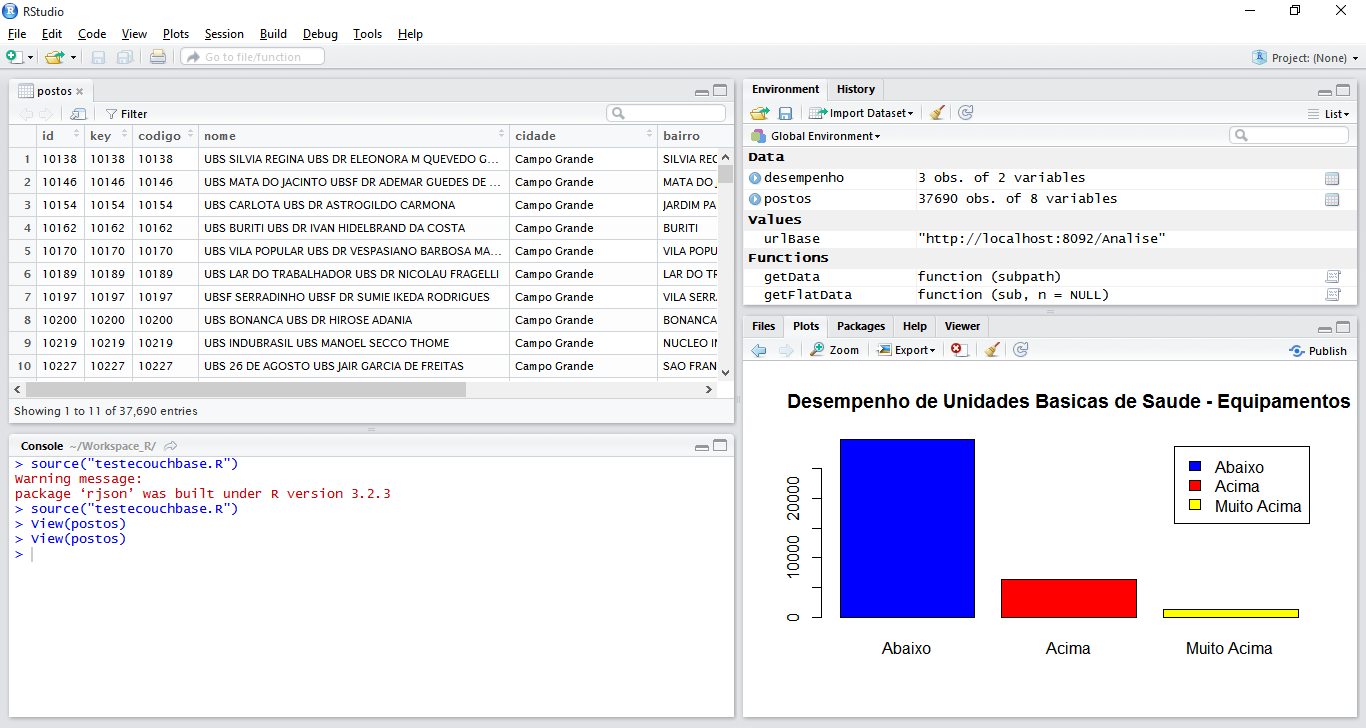
barplot(desempenho$count,col=c("blue","red","yellow"),legend=desempenho$desempenho,names.arg = desempenho$desempenho,border=TRUE)

title(main="Desempenho de Unidades Basicas de Saude - Equipamentos")

# Consulta a View com os dados das Unidades Basicas de Saude

postos <- getFlatData('/\_design/analise/\_view/postos',c("id","key","codigo","nome","cidade","bairro","latitude","longitude"))

Segue resultado obtido no RStudio:



# ANEXO 3 – Instalação e desinstalação do Couchbase Server

1. **Plataformas suportadas**

| **Plataforma** | **Versão** | **32 / 64 bit** | **Suportado para:** |
| --- | --- | --- | --- |
| Red Hat Enterprise Linux | 7 | 64 bit | Desenvolvimento, teste e produção |
| Red Hat Enterprise Linux | 6 | 64 bit | Desenvolvimento, teste e produção |
| SUSE Enterprise Linux\* | 11 | 64 bit | Desenvolvimento, teste e produção |
| Oracle Enterprise Linux | 6 | 64 bit | Desenvolvimento, teste e produção |
| Ubuntu Linux | 14.04 | 64 bit | Desenvolvimento, teste e produção |
| Ubuntu Linux | 12.04 | 64 bit | Desenvolvimento, teste e produção |
| CentOS | 7 | 64 bit | Desenvolvimento, teste e produção |
| CentOS | 6 | 64 bit | Desenvolvimento, teste e produção |
| Debian Linux | 7 | 64 bit | Desenvolvimento, teste e produção |
| Amazon Linux 2014 | 2013.03 | 64 bit | Desenvolvimento, teste e produção |
| Windows 2012 | R2 with SP1 | 64 bit | Desenvolvimento, teste e produção |
| Windows 2008 | R2 with SP1 | 64 bit | Desenvolvimento, teste e produção |
| Windows 10 |  | 32 bit and 64 bit | Desenvolvimento e teste |
| Windows 8 |  | 32 bit and 64 bit | Desenvolvimento e teste |
| Windows 7 |  | 32 bit and 64 bit | Desenvolvimento e teste |
| Mac OSX | 10.11 (El Capitan) | 64 bit | Desenvolvimento e teste |
| Mac OSX | 10.10 (Yosemite) | 64 bit | Desenvolvimento e teste |
| Mac OSX | 10.9 and 10.8 | 64 bit | Desenvolvimento e teste |

Tabela 3 - Plataformas Suportadas

1. **Exemplo de Instalação para Linux - Ubuntu**

O Couchbase Server é executado em vários sistemas Linux: Red Hat, CentOS, Ubuntu, Debian, SuSE Enterprise e Oracle Enterprise. Você pode baixar Couchbase Server diretamente a um sistema Linux usando o wget ou comando curl. Para exemplificar segue o exemplo de instalação Linux para Ubuntu:

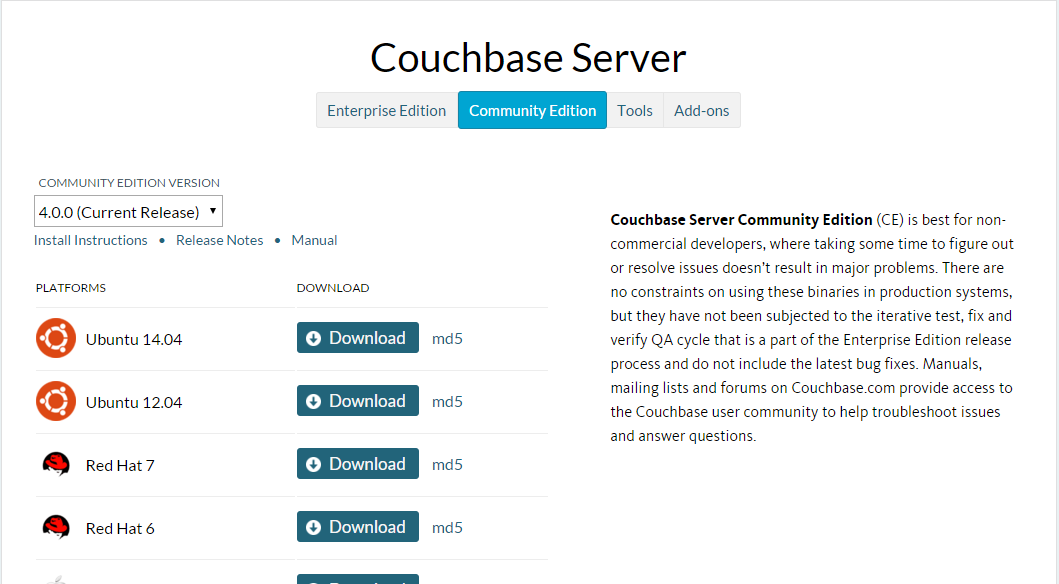


Figura 6 - Tela de Download - CouchBase para Ubuntu.



Figura 7 - Comando de Download - CouchBase para Ubuntu.



Figura 8- Comando de Instalação - CouchBase para Ubuntu

1. **Exemplo de Instalação para Windows**

Para instalar Couchbase Server no Windows, você precisa verificar os seguintes pré-requisitos:

* Certifique-se que você não tem nenhum software antinvírus em execução no computador antes de iniciar o processo de instalação.
* Verifique se você tem privilégios de administrador na máquina em que você está executando a instalação.
* Por padrão, no Windows TCP / IP tem número padrão de portas disponíveis para a comunicação do cliente.
* Você precisa seguir as informações para ajustar a configuração e aumentar o número de portas disponíveis.
* Por padrão, o sistema Windows não fornece um número adequado de portas efêmeras para clusters de Couchbase Server. Sem o número correto de portas efêmeras abertas, você pode experimentar erros durante o rebalanceamento, tempos de espera em clientes e falha nos backups. O instalador do Servidor Couchbase vai verificar automaticamente a sua configuração atual de portas e ajustá-lo, se necessário.

***Importante: O Couchbase Server usa pacote redistribuível Microsoft C ++, que é baixado automaticamente durante a instalação. No entanto, se outro aplicativo em sua máquina já está usando o pacote, o processo de instalação pode falhar. Para certificar-se de que o seu processo de instalação é concluído com êxito, encerre todas as outras aplicações em execução durante a instalação.***

Para instalar Couchbase Server no Windows usando assistente de instalação interativa siga os passos abaixo:

* No Windows Explorer, localize o arquivo Couchbase Server.exe baixado, que normalmente está localizado na pasta de downloads padrão, e clique duas vezes no arquivo executável.
* Você será solicitado com a tela de Instalação Localização. Você pode alterar o local onde o aplicativo servidor Couchbase está localizado, o que configura a localização do servidor e não o local onde os dados persistentes são armazenados.

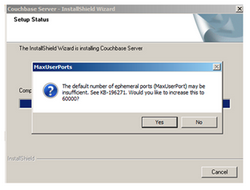


Figura 9 - Instalando parte 1 - CouchBase para Windows.

Clique em **Sim** para continuar. Sem um número suficiente de portas efemeras, um cluster Couchbase Server falhar durante reequilíbrio e backup. Se você já alterou essa configuração, você pode clicar em **Não**. O instalador exibe este painel para confirmar a atualização.

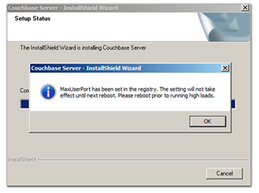


Figura 10 - Instalando parte 2 - CouchBase para Windows

* Após a instalação, siga as instruções de configuração do servidor.

1. **Exemplo de Desinstalação para Linux - Ubuntu**

Para desinstalar Couchbase Server no Ubuntu usando a ferramenta de linha de comando dpkg com o pacote DEB baixado, é necessário executar previamente os seguintes passos:

* Desligue Couchbase Server.
* Se a sua máquina é parte de um cluster ativo, reequilibrar o cluster para tirar o nó fora da configuração.
* Atualizar quaisquer clientes para apontar para um nó disponível no cluster Couchbase Server.

Feito o processo prévio inicia-se o processo de desinstalação do Couchbase Server, seguindo os passos:

* Login como root (super usuário) para concluir a instalação.
* Execute o seguinte comando: > sudo dpkg -r couchbase-server

***Nota: Pode ser necessário apagar os arquivos de dados associados à sua instalação. O local de instalação padrão é / opt. Se um local alternativo para os arquivos de dados foi especificado, cada diretório dados devem ser apagados individualmente a partir de seu sistema.***

1. **Exemplo de Desinstalação para Windows**

Para desinstalar Couchbase Server em um sistema Windows, você deve ter privilégios de administrador ou usuário avançado e é necessário executar previamente os seguintes passos:

* Desligue Couchbase Server.
* Se a sua máquina é parte de um cluster ativo, reequilibrar o cluster para tirar o nó fora da configuração.
* Atualizar quaisquer clientes para apontar para um nó disponível no cluster Couchbase Server.

Feito o processo prévio inicia-se o processo de desinstalação do Couchbase Server, seguindo os passos:

* Navegue para Iniciar> Configurações > Painel de controle
* Selecione Adicionar ou remover programas.
* Remover o software Couchbase Server.