BIG DATA

INTERNATIONAL DEBT STATISTICS – Statystyki dotyczące długu międzynarodowego

Adam Stańczyk, 202297

Kinga Sochacka, 20…

Prowadzący: dr inż Paweł Drzymała

**1. WSTĘP**

Strona z której wzięliśmy dane:

<https://datacatalog.worldbank.org/dataset/international-debt-statistics>

Opis kolumn które wybraliśmy do przetwarzania:

"Average grace period on new external debt commitments (years)" - Okres karencji to okres od daty podpisania pożyczki lub wydania instrumentu finansowego do pierwszego zleceniodawcy. Aby uzyskać średnią, okresy karencji dla wszystkich publicznych i publicznie gwarantowanych pożyczek zostały zważone kwotami pożyczek. Dług publiczny jest zewnętrznym obowiązkiem dłużnika publicznego, w tym rządu krajowego, podobszaru (lub agencji) oraz niezależnych organów publicznych. Dług publiczny gwarantowany to zewnętrzny obowiązek dłużnika prywatnego gwarantowany do spłaty przez podmiot publiczny.

"Commitments official creditors (COM current US$)" - Zobowiązania to kwota pożyczek długoterminowych, na które zostały podpisane umowy w danym roku. Dług z oficjalnych pożyczek od instytucji międzynarodowych (pożyczki wielostronne) i pożyczki od rządów (kredyty dwustronne). Pożyczki z organizacji obejmują pożyczki i kredyty z Banku Światowego, regionalnych banków rozwoju oraz innych wielostronnych i agencji. Wyłączone są pożyczki z funduszy zarządzanych przez organizację międzynarodową w imieniu pojedynczego darczyńcy, które są klasyfikowane jako pożyczki od rządów. Pożyczki rządowe obejmują pożyczki od rządów i ich agencji (w tym banków centralnych), pożyczki od niezależnych podmiotów oraz pożyczki bezpośrednie od oficjalnych agencji kredytów eksportowych. Długoterminowy dług zewnętrzny definiuje się jako dług, którego pierwotny lub przedłużony termin zapadalności przekracza jeden rok i który jest należny nierezydentom przez rezydentów danej gospodarki i podlega spłacie w walucie, towarach lub usługach. Dane znajdują się w dolarach amerykańskich.

"Concessional debt (% of total external debt)" - Koncesyjny dług do całkowitego zewnętrznego długu. Dług umowny definiowany jest jako pożyczki o pierwotnym dotacji w wysokości 25 procent lub więcej.

"Debt stock reduction (current US$)" - Redukcje zapasów zadłużenia pokazują kwotę, która została skompensowana ze stanu długu przy użyciu programów konwersji długu, takich jak wykupy i swapy lub zdyskontowana wartość długoterminowych obligacji, które zostały wyemitowane w zamian za zaległy dług. Obejmuje on efekt jakiejkolwiek operacji, która ograniczy zadłużenie inne niż restrukturyzacja zadłużenia, spłatę kapitału i umorzenie długów. w szczególności redukcja zadłużenia będzie obejmować wartość nominalną wykupionego długu, wartość nominalną zadłużenia zamienionego na kapitał własny (lub "" rozwój "), dowolną redukcję wartości nominalnej, która może wynikać z wymiany obligacji oraz dowolną wartość nominalną wynikającą z zamiany długu na obligacje dyskontowe. Dane znajdują się w bieżących dolarach amerykańskich.

"External debt stocks total (DOD current US$)" - Zadłużenie zagraniczne ogółem to dług wobec nierezydentów spłacanych w walucie, towarach lub usługach. Jest to suma publicznych, gwarantowanych i prywatnych nie gwarantowanych długów długoterminowych, krótkoterminowych długów i wykorzystania kredytu MFW. Dane znajdują się w bieżących dolarach amerykańskich.

"Interest payments on external debt total (INT current US$)" - Płatności odsetkowe to rzeczywiste kwoty odsetek wypłaconych przez pożyczkobiorcę w walucie, towarach lub usługach w danym roku. Obejmuje to odsetki zapłacone od zadłużenia długoterminowego, opłaty MFW i odsetki zapłacone od zadłużenia krótkoterminowego. Długoterminowy dług zewnętrzny to określony dług o pierwotnym lub przedłużonym terminie zapadalności powyżej jednego roku, który jest należny nierezydentom przez rezydentów gospodarki podlegającej spłacie w walucie, towarach lub usługach. Krótkoterminowe zadłużenie zagraniczne definiuje się jako dług o pierwotnym terminie zapadalności mniejszym o rok. Dostępne dane nie pozwalają na rozróżnienie między krótkoterminowymi długami publicznymi i prywatnymi. Dane znajdują się w bieżących dolarach amerykańskich.

"Short-term debt (% of total external debt)" - Zadłużenie krótkoterminowe obejmuje wszystkie zadłużenia o pierwotnym terminie zapadalności do jednego roku lub krótszym oraz odsetki za zwłokę z tytułu długu długoterminowego. zadłużenie zagraniczne to dług wobec nierezydentów podlegających spłacie w walucie, towarach lub usługach. Zadłużenie zagraniczne ogółem jest sumą publicznie gwarantowanego publicznego i prywatnego, nie wymagającego zabezpieczenia długu długoterminowego, wykorzystania kredytu MFW i zadłużenia krótkoterminowego.

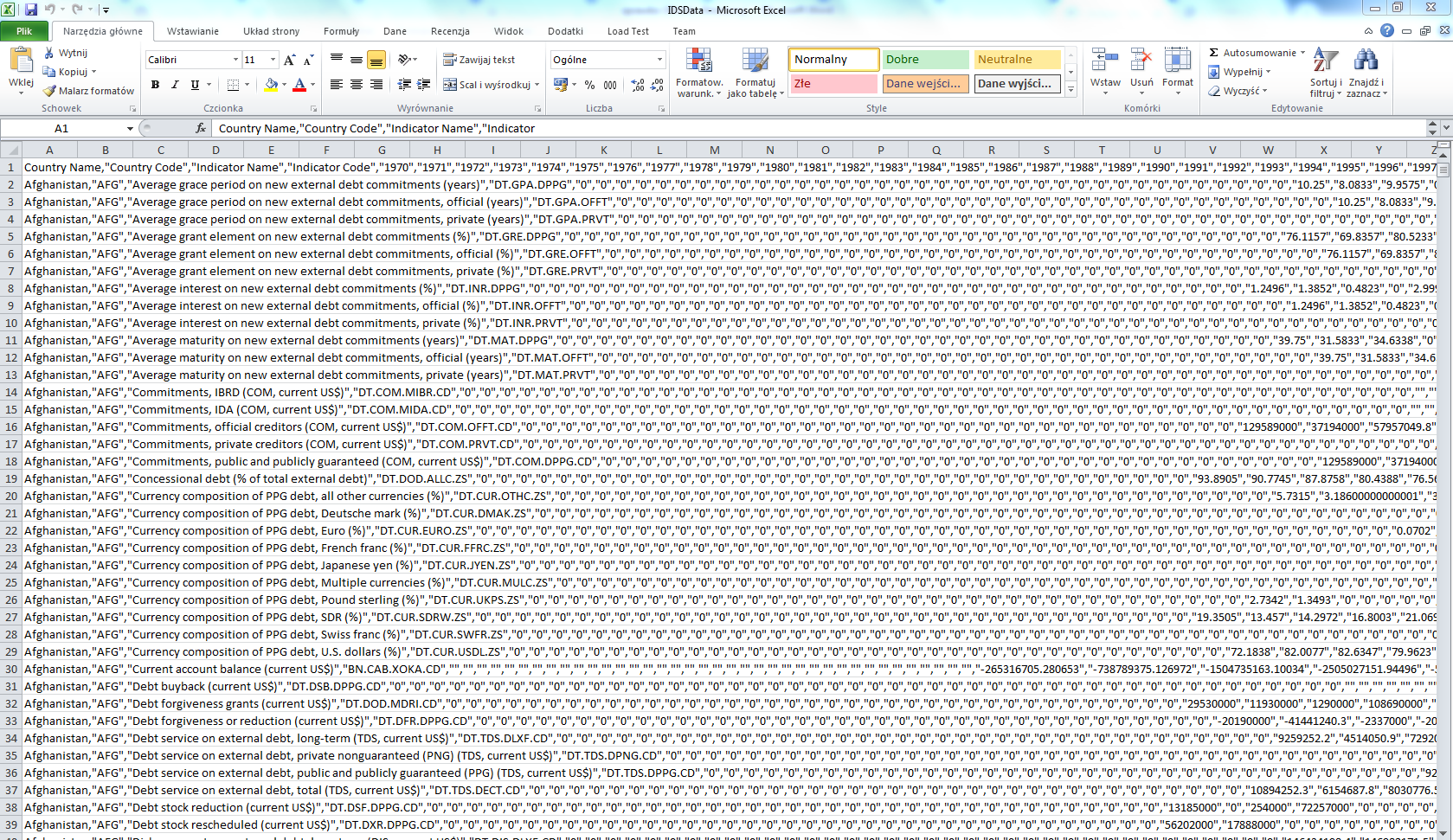
"Technical cooperation grants (current US$)" - Dotacje na współpracę techniczną obejmują niezależne dotacje na współpracę techniczną, które mają na celu sfinansowanie transferu technologii i zarządzania lub technologii w celu budowania ogólnych zdolności krajowych bez odwoływania się do konkretnych projektów inwestycyjnych; dotacje na współpracę techniczną w zakresie inwestycji, które są przyznawane w celu wzmocnienia zdolności do realizacji konkretnych projektów inwestycyjnych. Dane znajdują się w bieżących dolarach amerykańskich.

Statystyki które zamierzamy utworzyć:

Commitments official creditors (COM current US$) - > srednia z lat 2005 – 2015

Short-term debt (% of total external debt) \* External debt stocks total (DOD current US$) - > kwota długów krótkoterminowych, wybierzemy sobie jakiś rok

**2. PRZYGOTOWANIE DO PRZETWARZANIA W HIVE**



Rys. Dane przed przetworzeniem

Postanowiliśmy zredukować liczbę wierszy, wybraliśmy tylko potrzebne dane oraz kolumny lat, badamy dane od 2005 roku do 2015.

import csv

old\_file = 'IDSData.csv'

new\_file = 'IDS/data.csv'

i = 0

add\_to\_buff = [

"Average grace period on new external debt commitments (years)",

"Commitments official creditors (COM current US$)",

r"Concessional debt (% of total external debt)",

"Debt stock reduction (current US$)",

"External debt stocks total (DOD current US$)",

"Interest payments on external debt total (INT current US$)",

r"Short-term debt (% of total external debt)",

"Technical cooperation grants (current US$)",

]

content = []

# otwieramy plik z ktorego bierzemy dane

with open(old\_file, 'r', newline='') as csv\_file\_read:

reader = csv.reader(csv\_file\_read)

buff\_row = []

next(reader)

for row in reader:

buff\_row = []

for i in range(0, 2):

buff\_row.append(row[i])

string = row[2].replace(',',' ')

string = string.replace(" "," ")

if string in add\_to\_buff:

buff\_row.append(string)

else:

continue

buff\_row.append(row[3])

for i in range(40, 50):

if row[i]:

buff\_row.append(row[i])

else:

buff\_row.append(0)

content.append(buff\_row)

# dane zapisujemy do nowej csvki

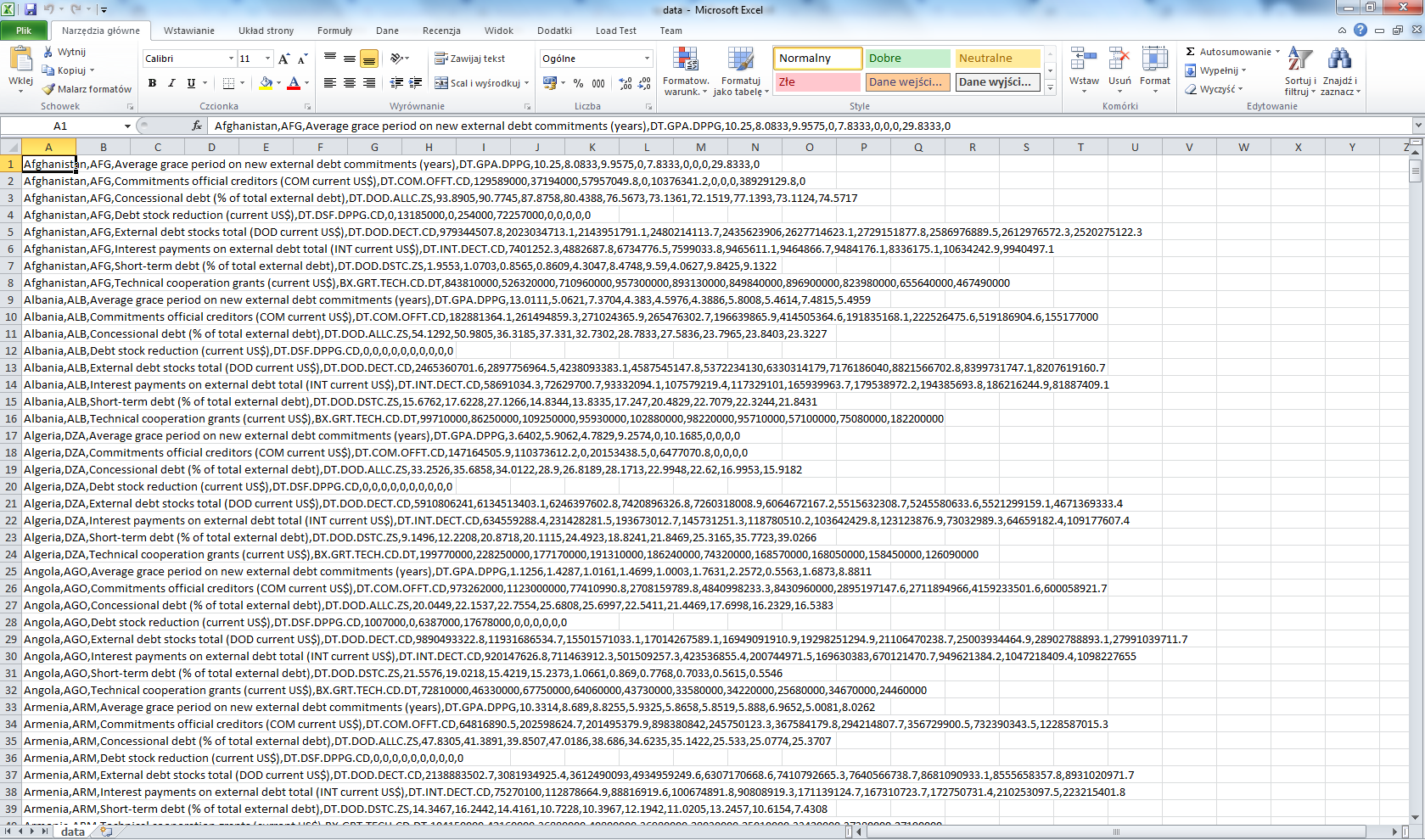
with open(new\_file, 'w', newline='') as csv\_file\_write:

writer = csv.writer(csv\_file\_write)

for row in content:

writer.writerow(row)

Listing Skryptu przetwarzającego dane



Rys. Po przetworzeniu

**3. Przetwarzanie w hivie**

Stworzyliśmy tabele w

CREATE TABLE ids\_data (CountryName string, CountryCode string, IndicatorName string, IndicatorCode string,

y2006 double, y2007 double, y2008 double, y2009 double, y2010 double, y2011 double,

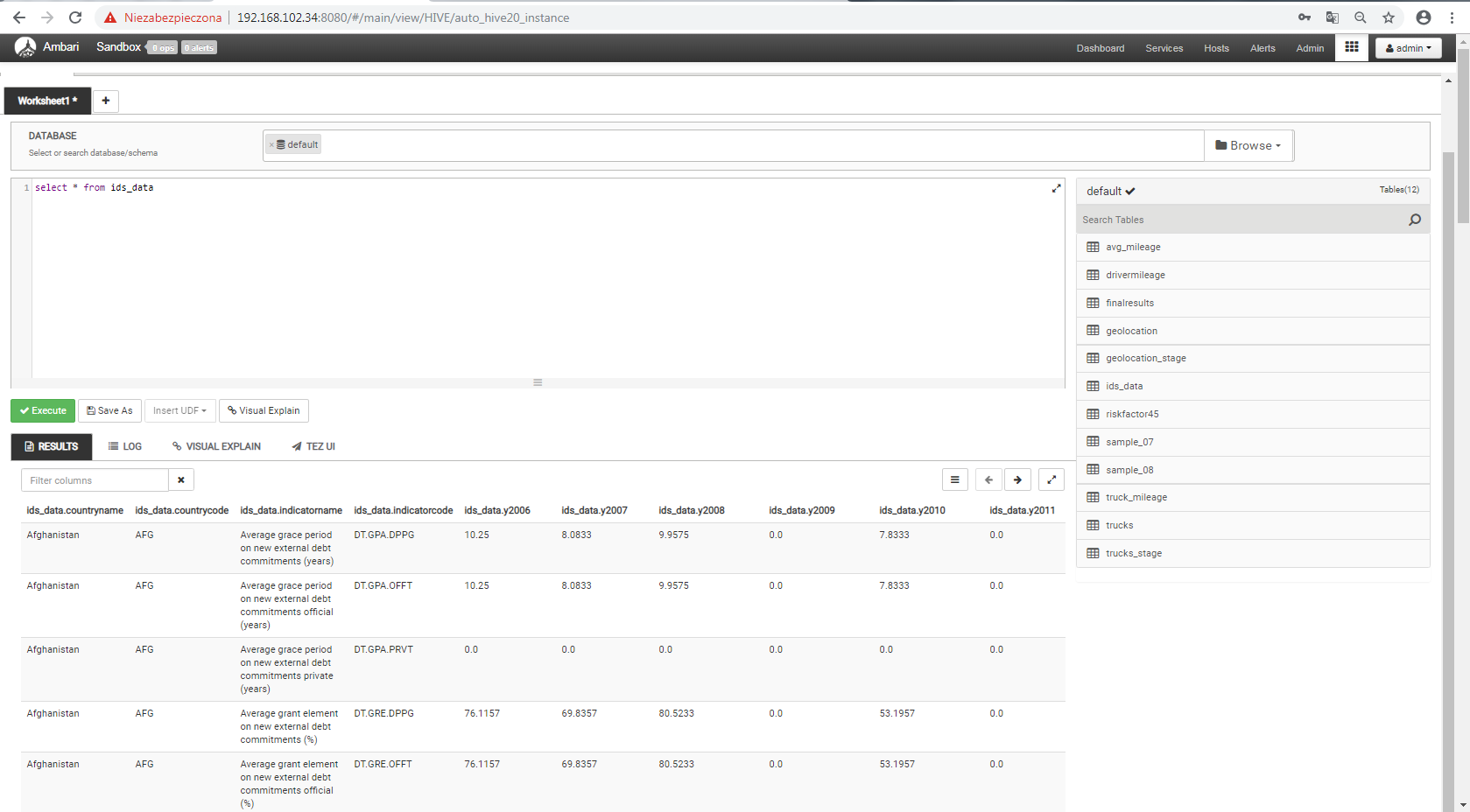
y2012 double, y2013 double, y2014 double, y2015 double)

row format delimited fields terminated by ',';

listing tworzenia tabeli

Uploadowalismy plik przez na stworzony na rysunku do katalogu/apps/hive/warehouse/ids\_data

Ids\_data.csv



Select wykonany po wstwieniu danych do tabeli ids\_data

**4. Przetwarzanie w PIGU**

Uruchamiamy piga komendą pig -useHCatalog

a = LOAD ‘ids\_data’ using org.apache.hive.hcatalog.pig.HCatLoader();

commitments = FILTER a BY indicatorname == ‘Commitments official creditors (COM current US$)’;

grace\_period = FILTER a BY indicatorname == ‘Average grace period on new external debt commitments (years)’;

concessional\_debt = FILTER a BY indicatorname == ‘Concessional debt (% of total external debt)’;

debt\_stock\_reduciton = FILTER a BY indicatorname == ‘Debt stock reduction (current US$)’;

external\_debt\_total = FILTER a BY indicatorname == ’External debt stocks total (DOD current US$)’;

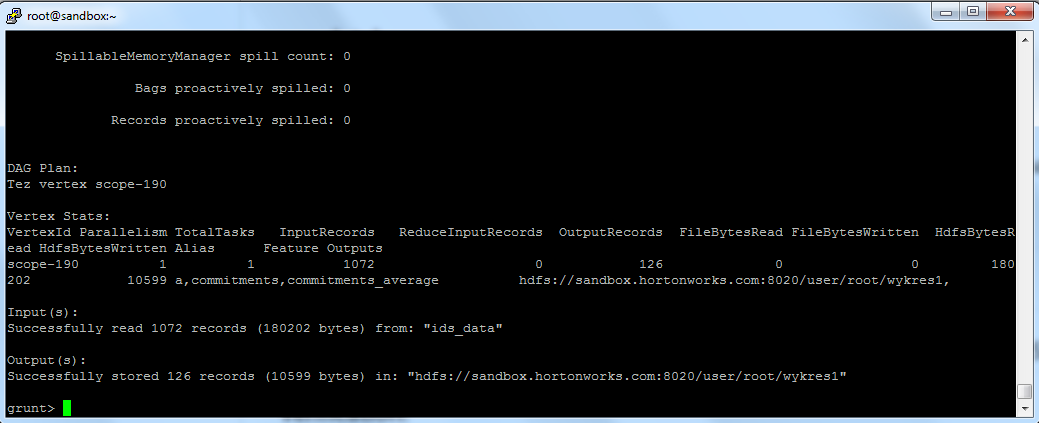
interest\_payments = FILTER a BY indicatorname == ‘Interest payments on external debt total (INT current US$)’;

short\_term\_debt = FILTER a BY indicatorname == ‘Short-term debt (% of total external debt)’;

commitments\_average = FOREACH commitments GENERATE (countryname, countrycode, indicatorname, ((y2006+y2007+y2008+y2009+y2010+y2011+y2012+y2013+y2014+y2015)/10) );

STORE commitments\_average INTO ‘wykres1’

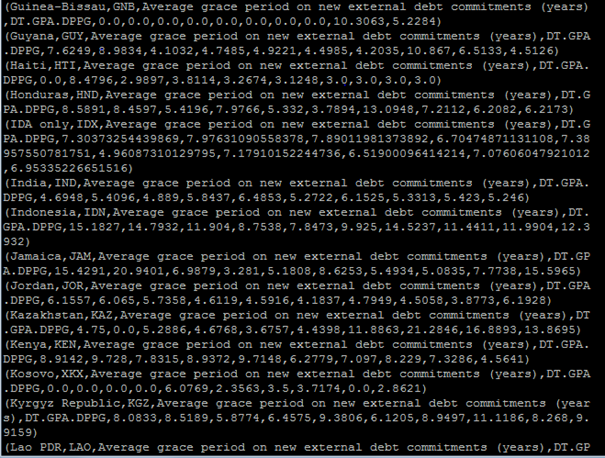
listing skryptu napisanego w pig latin



Rys. Komunikat po zapisaniu zmiennej pig

CREATE TABLE ids\_commitments(CountryName string, CountryCode string, IndicatorName string, srednia\_commitments double)

row format delimited fields terminated by ',');

Rys. Dump wszystkich danych

**5. PRZETWARZANIE W SPARKU**

Komendy:

%jdbc(hive) show tables

%spark2

val hiveContext = new org.apache.spark.sql.SparkSession.Builder().getOrCreate()

%spark2

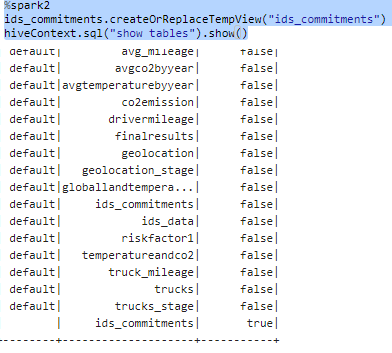
val ids\_commitments = hiveContext.sql("select \* from ids\_commitments")

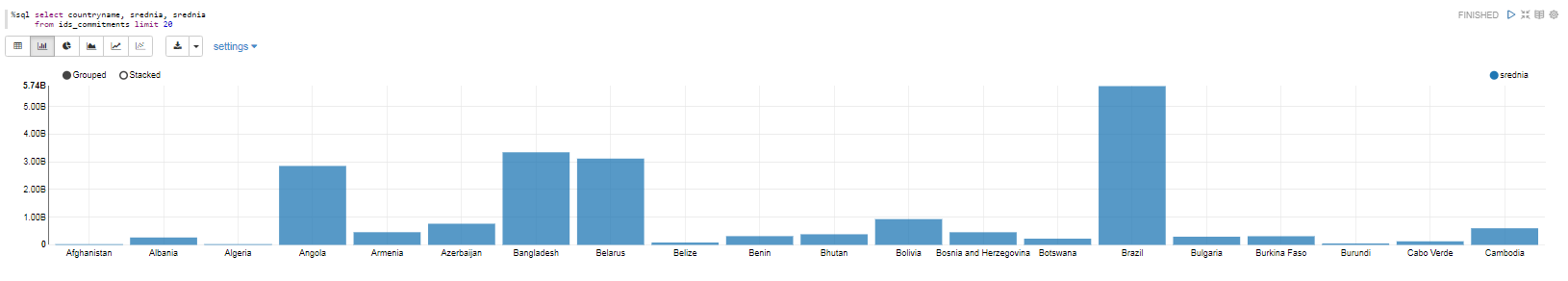
%spark2

ids\_commitments.createOrReplaceTempView("ids\_commitments")

hiveContext.sql("show tables").show()

Listing spark skrypty wszystkie





Jeszcze raz zrobimy tego skrina, można lepiej