

# Składowanie danych w systemach Big Data

## Testy poprawności działania rozwiązania

Kacper Skonieczka, 313505

Grzegorz Zakrzewski, 313555

10 stycznia 2024

Nr	Co jest testowane?	W jaki sposób?	Oczekiwany wynik	Potwierdzenie
1	Sprawdzenie procesu składowania plików .parquet z cenami energii na daną godzinę z ENTSOE Transparency Platform	Wykonanie odpowiednich flow w Apache NiFi i sprawdzenie zawartości na HDFS	Widoczne wśród plików takie, które zawierają datę i godzinę "2024-01-08 13" w nazwie	<pre>vagrant@node1:~\$ hdfs dfs -ls /user/project/germany_prices/   tail -n 5 SLF4J: Class path contains multiple SLF4J bindings. SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.7.6/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class] SLF4J: Found binding in [jar:file:/usr/local/apache-tez-0.9.1-bin/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class] SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation. SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory] drwxr-xr-x - root supergroup 0 2024-01-08 13:40 /user/project/germany_prices/prices_2024-01-08 13 drwxr-xr-x - root supergroup 0 2024-01-08 13:40 /user/project/germany_prices/prices_2024-01-08 13 drwxr-xr-x - root supergroup 0 2024-01-08 13:40 /user/project/germany_prices/prices_2024-01-08 13 drwxr-xr-x - root supergroup 0 2024-01-08 13:40 /user/project/germany_prices/prices_2024-01-08 13 drwxr-xr-x - root supergroup 0 2024-01-08 13:40 /user/project/germany_prices/prices_2024-01-08 13</pre>
2	Sprawdzenie procesu składowanie plików .parquet z ilością energii wiatrowej na daną godzinę z ENT-SOE Transparency Platform	Wykonanie odpowiednich flow w Apache NiFi i sprawdzenie zawartości w Apache HBase	Widoczne rekordy z <i>timestamp</i> z dnia 2024-01-08 sprzed godziny 14	<pre>hbase(main):016:0&gt; scan 'germany_prices', (LIMIT =&gt; 5, REVERSED =&gt; true) ROW 9 2024-01-07T23:00Z column=cf:end_date, timestamp=2024-01-08T13:46:04.295, value=2024-01-08T23:00Z 9 2024-01-07T23:00Z column=cf:position, timestamp=2024-01-08T13:46:04.295, value=9 9 2024-01-07T23:00Z column=cf:price_amount, timestamp=2024-01-08T13:46:04.295, value=121.11 9 2024-01-07T23:00Z column=cf:publication_date, timestamp=2024-01-08T13:46:04.295, value=2024-01-08 13 9 2024-01-07T23:00Z column=cf:start_date, timestamp=2024-01-08T13:46:04.295, value=2024-01-07T23:00Z 9 2024-01-06T23:00Z column=cf:end_date, timestamp=2024-01-08T13:46:05.330, value=2024-01-07T23:00Z 9 2024-01-06T23:00Z column=cf:position, timestamp=2024-01-08T13:46:05.330, value=9 9 2024-01-06T23:00Z column=cf:price_amount, timestamp=2024-01-08T13:46:05.330, value=81.79 9 2024-01-06T23:00Z column=cf:publication_date, timestamp=2024-01-08T13:46:05.330, value=2024-01-08 13 9 2024-01-06T23:00Z column=cf:start_date, timestamp=2024-01-08T13:46:05.330, value=2024-01-06T23:00Z</pre>
3	Sprawdzenie procesu zapisywania w Apache HBase najnowszych danych o cenach energii w Niemczech	Wykonanie odpowiednich flow w Apache NiFi i sprawdzenie zawartości na HDFS	Widoczne wśród plików takie, które zawierają datę i godzinę "2024-01-08 13" w nazwie	<pre>vagrant@node1:~\$ hdfs dfs -ls /user/project/germany_prices/   tail -n 5 SLF4J: Class path contains multiple SLF4J bindings. SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.7.6/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class] SLF4J: Found binding in [jar:file:/usr/local/apache-tez-0.9.1-bin/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class] SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation. SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory] drwxr-xr-x - root supergroup 0 2024-01-08 13:40 /user/project/germany_prices/prices_2024-01-08 13 drwxr-xr-x - root supergroup 0 2024-01-08 13:40 /user/project/germany_prices/prices_2024-01-08 13 drwxr-xr-x - root supergroup 0 2024-01-08 13:40 /user/project/germany_prices/prices_2024-01-08 13 drwxr-xr-x - root supergroup 0 2024-01-08 13:40 /user/project/germany_prices/prices_2024-01-08 13 drwxr-xr-x - root supergroup 0 2024-01-08 13:40 /user/project/germany_prices/prices_2024-01-08 13</pre>

4	Sprawdzenie procesu zapisywania w Apache HBase najnowszych danych o energii wiatrowej w Niemczech	Wykonanie odpowiednich flow w Apache NiFi i sprawdzenie zawartości w Apache HBase	Widoczne wśród danych, rekordy z timestamp z dnia 2024-01-08 sprzed godziny 14	<pre>hbase(main):001:0&gt; scan 'windOffShore', (LIMIT =&gt; 5, REVERSED =&gt; true) ROW                                COLUMN+CELL 9 2024-01-08T23:00Z                column=cf:end_date, timestamp=2024-01-08T13:55:59.596, value=2024-01-08T23:00Z 9 2024-01-08T23:00Z                column=cf:position, timestamp=2024-01-08T13:55:59.596, value=9 9 2024-01-08T23:00Z                column=cf:publication_date, timestamp=2024-01-08T13:55:59.596, value=2024-01-08 13 9 2024-01-08T23:00Z                column=cf:quantity, timestamp=2024-01-08T13:55:59.596, value=0 9 2024-01-08T23:00Z                column=cf:row_id, timestamp=2024-01-08T13:55:59.596, value=9 9 2024-01-08T23:00Z                column=cf:start_date, timestamp=2024-01-08T13:55:59.596, value=2024-01-07T23:00Z 9 2024-01-07T23:00Z                column=cf:end_date, timestamp=2024-01-08T13:56:07.068, value=2024-01-07T23:00Z 9 2024-01-07T23:00Z                column=cf:position, timestamp=2024-01-08T13:56:07.068, value=9 9 2024-01-07T23:00Z                column=cf:publication_date, timestamp=2024-01-08T13:56:07.068, value=2024-01-08 13 9 2024-01-07T23:00Z                column=cf:quantity, timestamp=2024-01-08T13:56:07.068, value=0 9 2024-01-07T23:00Z                column=cf:row_id, timestamp=2024-01-08T13:56:07.068, value=9 9 2024-01-07T23:00Z                column=cf:start_date, timestamp=2024-01-08T13:56:07.068, value=2024-01-06T23:00Z</pre>																																																																																																																																																			
5	Sprawdzenie dostępności z Apache Spark plików parquet z cenami energii na daną godzinę z ENTSOE Transparency Platform	Połączenie się z HDFS za pomocą odpowiedniej ścieżki i wyświetlenie części zawartości	Tabela jest zapełniona danymi o cenach energii	<pre>prices_da.show()</pre> <table><thead><tr><th></th><th>position</th><th>price_amount</th><th>start_date</th><th>row_id</th><th>end_date</th><th>publication_date</th></tr></thead><tbody><tr><td>1</td><td></td><td>106.91</td><td>2023-11-30T23:00Z</td><td>1_2023-11-30T23:00Z</td><td>2023-12-01T23:00Z</td><td>2024-01-08 22</td></tr><tr><td>3</td><td></td><td>99.1</td><td>2023-11-30T23:00Z</td><td>3_2023-11-30T23:00Z</td><td>2023-12-01T23:00Z</td><td>2024-01-08 22</td></tr><tr><td>5</td><td></td><td>96.07</td><td>2023-11-30T23:00Z</td><td>5_2023-11-30T23:00Z</td><td>2023-12-01T23:00Z</td><td>2024-01-08 22</td></tr><tr><td>19</td><td></td><td>199.16</td><td>2023-11-30T23:00Z</td><td>19_2023-11-30T23:00Z</td><td>2023-12-01T23:00Z</td><td>2024-01-08 22</td></tr><tr><td>20</td><td></td><td>168.56</td><td>2023-11-30T23:00Z</td><td>20_2023-11-30T23:00Z</td><td>2023-12-01T23:00Z</td><td>2024-01-08 22</td></tr><tr><td>7</td><td></td><td>120.03</td><td>2023-11-30T23:00Z</td><td>7_2023-11-30T23:00Z</td><td>2023-12-01T23:00Z</td><td>2024-01-08 22</td></tr><tr><td>10</td><td></td><td>222.63</td><td>2023-11-30T23:00Z</td><td>10_2023-11-30T23:00Z</td><td>2023-12-01T23:00Z</td><td>2024-01-08 22</td></tr><tr><td>13</td><td></td><td>197.09</td><td>2023-11-30T23:00Z</td><td>13_2023-11-30T23:00Z</td><td>2023-12-01T23:00Z</td><td>2024-01-08 22</td></tr><tr><td>16</td><td></td><td>184.4</td><td>2023-11-30T23:00Z</td><td>16_2023-11-30T23:00Z</td><td>2023-12-01T23:00Z</td><td>2024-01-08 22</td></tr><tr><td>21</td><td></td><td>142.37</td><td>2023-11-30T23:00Z</td><td>21_2023-11-30T23:00Z</td><td>2023-12-01T23:00Z</td><td>2024-01-08 22</td></tr><tr><td>22</td><td></td><td>125.13</td><td>2023-11-30T23:00Z</td><td>22_2023-11-30T23:00Z</td><td>2023-12-01T23:00Z</td><td>2024-01-08 22</td></tr><tr><td>23</td><td></td><td>118.4</td><td>2023-11-30T23:00Z</td><td>23_2023-11-30T23:00Z</td><td>2023-12-01T23:00Z</td><td>2024-01-08 22</td></tr><tr><td>24</td><td></td><td>108.13</td><td>2023-11-30T23:00Z</td><td>24_2023-11-30T23:00Z</td><td>2023-12-01T23:00Z</td><td>2024-01-08 22</td></tr><tr><td>7</td><td></td><td>98.93</td><td>2023-12-01T23:00Z</td><td>7_2023-12-01T23:00Z</td><td>2023-12-02T23:00Z</td><td>2024-01-08 22</td></tr><tr><td>14</td><td></td><td>114.58</td><td>2023-12-01T23:00Z</td><td>14_2023-12-01T23:00Z</td><td>2023-12-02T23:00Z</td><td>2024-01-08 22</td></tr><tr><td>18</td><td></td><td>144.04</td><td>2023-12-01T23:00Z</td><td>18_2023-12-01T23:00Z</td><td>2023-12-02T23:00Z</td><td>2024-01-08 22</td></tr><tr><td>22</td><td></td><td>114.83</td><td>2023-12-01T23:00Z</td><td>22_2023-12-01T23:00Z</td><td>2023-12-02T23:00Z</td><td>2024-01-08 22</td></tr><tr><td>24</td><td></td><td>106.01</td><td>2023-12-01T23:00Z</td><td>24_2023-12-01T23:00Z</td><td>2023-12-02T23:00Z</td><td>2024-01-08 22</td></tr><tr><td>3</td><td></td><td>99.88</td><td>2023-12-02T23:00Z</td><td>3_2023-12-02T23:00Z</td><td>2023-12-03T23:00Z</td><td>2024-01-08 22</td></tr><tr><td>17</td><td></td><td>122.2</td><td>2023-12-02T23:00Z</td><td>17_2023-12-02T23:00Z</td><td>2023-12-03T23:00Z</td><td>2024-01-08 22</td></tr></tbody></table> <p>only showing top 20 rows</p>		position	price_amount	start_date	row_id	end_date	publication_date	1		106.91	2023-11-30T23:00Z	1_2023-11-30T23:00Z	2023-12-01T23:00Z	2024-01-08 22	3		99.1	2023-11-30T23:00Z	3_2023-11-30T23:00Z	2023-12-01T23:00Z	2024-01-08 22	5		96.07	2023-11-30T23:00Z	5_2023-11-30T23:00Z	2023-12-01T23:00Z	2024-01-08 22	19		199.16	2023-11-30T23:00Z	19_2023-11-30T23:00Z	2023-12-01T23:00Z	2024-01-08 22	20		168.56	2023-11-30T23:00Z	20_2023-11-30T23:00Z	2023-12-01T23:00Z	2024-01-08 22	7		120.03	2023-11-30T23:00Z	7_2023-11-30T23:00Z	2023-12-01T23:00Z	2024-01-08 22	10		222.63	2023-11-30T23:00Z	10_2023-11-30T23:00Z	2023-12-01T23:00Z	2024-01-08 22	13		197.09	2023-11-30T23:00Z	13_2023-11-30T23:00Z	2023-12-01T23:00Z	2024-01-08 22	16		184.4	2023-11-30T23:00Z	16_2023-11-30T23:00Z	2023-12-01T23:00Z	2024-01-08 22	21		142.37	2023-11-30T23:00Z	21_2023-11-30T23:00Z	2023-12-01T23:00Z	2024-01-08 22	22		125.13	2023-11-30T23:00Z	22_2023-11-30T23:00Z	2023-12-01T23:00Z	2024-01-08 22	23		118.4	2023-11-30T23:00Z	23_2023-11-30T23:00Z	2023-12-01T23:00Z	2024-01-08 22	24		108.13	2023-11-30T23:00Z	24_2023-11-30T23:00Z	2023-12-01T23:00Z	2024-01-08 22	7		98.93	2023-12-01T23:00Z	7_2023-12-01T23:00Z	2023-12-02T23:00Z	2024-01-08 22	14		114.58	2023-12-01T23:00Z	14_2023-12-01T23:00Z	2023-12-02T23:00Z	2024-01-08 22	18		144.04	2023-12-01T23:00Z	18_2023-12-01T23:00Z	2023-12-02T23:00Z	2024-01-08 22	22		114.83	2023-12-01T23:00Z	22_2023-12-01T23:00Z	2023-12-02T23:00Z	2024-01-08 22	24		106.01	2023-12-01T23:00Z	24_2023-12-01T23:00Z	2023-12-02T23:00Z	2024-01-08 22	3		99.88	2023-12-02T23:00Z	3_2023-12-02T23:00Z	2023-12-03T23:00Z	2024-01-08 22	17		122.2	2023-12-02T23:00Z	17_2023-12-02T23:00Z	2023-12-03T23:00Z	2024-01-08 22
	position	price_amount	start_date	row_id	end_date	publication_date																																																																																																																																																	
1		106.91	2023-11-30T23:00Z	1_2023-11-30T23:00Z	2023-12-01T23:00Z	2024-01-08 22																																																																																																																																																	
3		99.1	2023-11-30T23:00Z	3_2023-11-30T23:00Z	2023-12-01T23:00Z	2024-01-08 22																																																																																																																																																	
5		96.07	2023-11-30T23:00Z	5_2023-11-30T23:00Z	2023-12-01T23:00Z	2024-01-08 22																																																																																																																																																	
19		199.16	2023-11-30T23:00Z	19_2023-11-30T23:00Z	2023-12-01T23:00Z	2024-01-08 22																																																																																																																																																	
20		168.56	2023-11-30T23:00Z	20_2023-11-30T23:00Z	2023-12-01T23:00Z	2024-01-08 22																																																																																																																																																	
7		120.03	2023-11-30T23:00Z	7_2023-11-30T23:00Z	2023-12-01T23:00Z	2024-01-08 22																																																																																																																																																	
10		222.63	2023-11-30T23:00Z	10_2023-11-30T23:00Z	2023-12-01T23:00Z	2024-01-08 22																																																																																																																																																	
13		197.09	2023-11-30T23:00Z	13_2023-11-30T23:00Z	2023-12-01T23:00Z	2024-01-08 22																																																																																																																																																	
16		184.4	2023-11-30T23:00Z	16_2023-11-30T23:00Z	2023-12-01T23:00Z	2024-01-08 22																																																																																																																																																	
21		142.37	2023-11-30T23:00Z	21_2023-11-30T23:00Z	2023-12-01T23:00Z	2024-01-08 22																																																																																																																																																	
22		125.13	2023-11-30T23:00Z	22_2023-11-30T23:00Z	2023-12-01T23:00Z	2024-01-08 22																																																																																																																																																	
23		118.4	2023-11-30T23:00Z	23_2023-11-30T23:00Z	2023-12-01T23:00Z	2024-01-08 22																																																																																																																																																	
24		108.13	2023-11-30T23:00Z	24_2023-11-30T23:00Z	2023-12-01T23:00Z	2024-01-08 22																																																																																																																																																	
7		98.93	2023-12-01T23:00Z	7_2023-12-01T23:00Z	2023-12-02T23:00Z	2024-01-08 22																																																																																																																																																	
14		114.58	2023-12-01T23:00Z	14_2023-12-01T23:00Z	2023-12-02T23:00Z	2024-01-08 22																																																																																																																																																	
18		144.04	2023-12-01T23:00Z	18_2023-12-01T23:00Z	2023-12-02T23:00Z	2024-01-08 22																																																																																																																																																	
22		114.83	2023-12-01T23:00Z	22_2023-12-01T23:00Z	2023-12-02T23:00Z	2024-01-08 22																																																																																																																																																	
24		106.01	2023-12-01T23:00Z	24_2023-12-01T23:00Z	2023-12-02T23:00Z	2024-01-08 22																																																																																																																																																	
3		99.88	2023-12-02T23:00Z	3_2023-12-02T23:00Z	2023-12-03T23:00Z	2024-01-08 22																																																																																																																																																	
17		122.2	2023-12-02T23:00Z	17_2023-12-02T23:00Z	2023-12-03T23:00Z	2024-01-08 22																																																																																																																																																	
6	Sprawdzenie dostępności z Apache Spark plików parquet z ilością energii wiatrowej na daną godzinę z ENTSOE Transparency Platform	Połączenie się z HDFS za pomocą odpowiedniej ścieżki i wyświetlenie części zawartości	Tabela jest zapełniona danymi o energii z wiatru	<table><thead><tr><th></th><th>position</th><th>quantity</th><th>start_date</th><th>row_id</th><th>end_date</th><th>publication_date</th></tr></thead><tbody><tr><td>42</td><td></td><td>4418.0</td><td>2023-12-11T23:00Z</td><td>42_2023-12-12T23:00Z</td><td>2023-12-12T23:00Z</td><td>2024-01-08 21</td></tr><tr><td>1</td><td></td><td>17153.0</td><td>2023-12-11T23:00Z</td><td>1_2023-12-12T23:00Z</td><td>2023-12-12T23:00Z</td><td>2024-01-08 21</td></tr><tr><td>43</td><td></td><td>4327.0</td><td>2023-12-11T23:00Z</td><td>43_2023-12-12T23:00Z</td><td>2023-12-12T23:00Z</td><td>2024-01-08 21</td></tr><tr><td>2</td><td></td><td>16647.0</td><td>2023-12-11T23:00Z</td><td>2_2023-12-12T23:00Z</td><td>2023-12-12T23:00Z</td><td>2024-01-08 21</td></tr><tr><td>44</td><td></td><td>4231.0</td><td>2023-12-11T23:00Z</td><td>44_2023-12-12T23:00Z</td><td>2023-12-12T23:00Z</td><td>2024-01-08 21</td></tr><tr><td>3</td><td></td><td>16141.0</td><td>2023-12-11T23:00Z</td><td>3_2023-12-12T23:00Z</td><td>2023-12-12T23:00Z</td><td>2024-01-08 21</td></tr><tr><td>4</td><td></td><td>15627.0</td><td>2023-12-11T23:00Z</td><td>4_2023-12-12T23:00Z</td><td>2023-12-12T23:00Z</td><td>2024-01-08 21</td></tr><tr><td>45</td><td></td><td>4145.0</td><td>2023-12-11T23:00Z</td><td>45_2023-12-12T23:00Z</td><td>2023-12-12T23:00Z</td><td>2024-01-08 21</td></tr><tr><td>5</td><td></td><td>15067.0</td><td>2023-12-11T23:00Z</td><td>5_2023-12-12T23:00Z</td><td>2023-12-12T23:00Z</td><td>2024-01-08 21</td></tr><tr><td>46</td><td></td><td>4107.0</td><td>2023-12-11T23:00Z</td><td>46_2023-12-12T23:00Z</td><td>2023-12-12T23:00Z</td><td>2024-01-08 21</td></tr><tr><td>6</td><td></td><td>14481.0</td><td>2023-12-11T23:00Z</td><td>6_2023-12-12T23:00Z</td><td>2023-12-12T23:00Z</td><td>2024-01-08 21</td></tr><tr><td>47</td><td></td><td>4068.0</td><td>2023-12-11T23:00Z</td><td>47_2023-12-12T23:00Z</td><td>2023-12-12T23:00Z</td><td>2024-01-08 21</td></tr><tr><td>8</td><td></td><td>13313.0</td><td>2023-12-11T23:00Z</td><td>8_2023-12-12T23:00Z</td><td>2023-12-12T23:00Z</td><td>2024-01-08 21</td></tr><tr><td>48</td><td></td><td>4026.0</td><td>2023-12-11T23:00Z</td><td>48_2023-12-12T23:00Z</td><td>2023-12-12T23:00Z</td><td>2024-01-08 21</td></tr><tr><td>51</td><td></td><td>4138.0</td><td>2023-12-11T23:00Z</td><td>51_2023-12-12T23:00Z</td><td>2023-12-12T23:00Z</td><td>2024-01-08 21</td></tr><tr><td>9</td><td></td><td>12735.0</td><td>2023-12-11T23:00Z</td><td>9_2023-12-12T23:00Z</td><td>2023-12-12T23:00Z</td><td>2024-01-08 21</td></tr><tr><td>11</td><td></td><td>11615.0</td><td>2023-12-11T23:00Z</td><td>11_2023-12-12T23:00Z</td><td>2023-12-12T23:00Z</td><td>2024-01-08 21</td></tr><tr><td>13</td><td></td><td>10513.0</td><td>2023-12-11T23:00Z</td><td>13_2023-12-12T23:00Z</td><td>2023-12-12T23:00Z</td><td>2024-01-08 21</td></tr><tr><td>19</td><td></td><td>7896.0</td><td>2023-12-11T23:00Z</td><td>19_2023-12-12T23:00Z</td><td>2023-12-12T23:00Z</td><td>2024-01-08 21</td></tr><tr><td>65</td><td></td><td>7645.0</td><td>2023-12-11T23:00Z</td><td>65_2023-12-12T23:00Z</td><td>2023-12-12T23:00Z</td><td>2024-01-08 21</td></tr></tbody></table> <p>only showing top 20 rows</p>		position	quantity	start_date	row_id	end_date	publication_date	42		4418.0	2023-12-11T23:00Z	42_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21	1		17153.0	2023-12-11T23:00Z	1_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21	43		4327.0	2023-12-11T23:00Z	43_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21	2		16647.0	2023-12-11T23:00Z	2_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21	44		4231.0	2023-12-11T23:00Z	44_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21	3		16141.0	2023-12-11T23:00Z	3_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21	4		15627.0	2023-12-11T23:00Z	4_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21	45		4145.0	2023-12-11T23:00Z	45_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21	5		15067.0	2023-12-11T23:00Z	5_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21	46		4107.0	2023-12-11T23:00Z	46_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21	6		14481.0	2023-12-11T23:00Z	6_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21	47		4068.0	2023-12-11T23:00Z	47_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21	8		13313.0	2023-12-11T23:00Z	8_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21	48		4026.0	2023-12-11T23:00Z	48_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21	51		4138.0	2023-12-11T23:00Z	51_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21	9		12735.0	2023-12-11T23:00Z	9_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21	11		11615.0	2023-12-11T23:00Z	11_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21	13		10513.0	2023-12-11T23:00Z	13_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21	19		7896.0	2023-12-11T23:00Z	19_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21	65		7645.0	2023-12-11T23:00Z	65_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21
	position	quantity	start_date	row_id	end_date	publication_date																																																																																																																																																	
42		4418.0	2023-12-11T23:00Z	42_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21																																																																																																																																																	
1		17153.0	2023-12-11T23:00Z	1_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21																																																																																																																																																	
43		4327.0	2023-12-11T23:00Z	43_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21																																																																																																																																																	
2		16647.0	2023-12-11T23:00Z	2_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21																																																																																																																																																	
44		4231.0	2023-12-11T23:00Z	44_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21																																																																																																																																																	
3		16141.0	2023-12-11T23:00Z	3_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21																																																																																																																																																	
4		15627.0	2023-12-11T23:00Z	4_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21																																																																																																																																																	
45		4145.0	2023-12-11T23:00Z	45_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21																																																																																																																																																	
5		15067.0	2023-12-11T23:00Z	5_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21																																																																																																																																																	
46		4107.0	2023-12-11T23:00Z	46_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21																																																																																																																																																	
6		14481.0	2023-12-11T23:00Z	6_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21																																																																																																																																																	
47		4068.0	2023-12-11T23:00Z	47_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21																																																																																																																																																	
8		13313.0	2023-12-11T23:00Z	8_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21																																																																																																																																																	
48		4026.0	2023-12-11T23:00Z	48_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21																																																																																																																																																	
51		4138.0	2023-12-11T23:00Z	51_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21																																																																																																																																																	
9		12735.0	2023-12-11T23:00Z	9_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21																																																																																																																																																	
11		11615.0	2023-12-11T23:00Z	11_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21																																																																																																																																																	
13		10513.0	2023-12-11T23:00Z	13_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21																																																																																																																																																	
19		7896.0	2023-12-11T23:00Z	19_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21																																																																																																																																																	
65		7645.0	2023-12-11T23:00Z	65_2023-12-12T23:00Z	2023-12-12T23:00Z	2024-01-08 21																																																																																																																																																	

7	Sprawdzenie, czy pliki GRIB2 z prognozami meteorologicznymi GEFS są prawidłowo pobierane do systemu plików HDFS	Uruchomienie procesorów Apache Nifi odpowiedzialnych za pobieranie pliku i sprawdzenie zawartości HDFS	W HDFS w katalogu <b>gefs/raw</b> znajduje się pobrany plik GRIB2	<pre>vagrant@node1:~\$ hadoop fs -ls /user/bigdataproject/gefs/raw SLF4J: Class path contains multiple SLF4J bindings. SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.7.6/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticMDCBinder.class] SLF4J: Found binding in [jar:file:/usr/local/apache-tez-0.9.1-bin/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticMDCBinder.class] SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation. SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory] Found 1 items -rw-r--r-- 1 root supergroup 12015805 2024-01-07 14:50 /user/bigdataproject/gefs/raw/gefs_raw_20240107.grib2</pre>
8	Sprawdzenie, czy surowe pliki GRIB2 są prawidłowo konwertowane do plików CSV	Uruchomienie procesora Apache Nifi odpowiedzialnego za konwersję plików GRIB2 do formatu CSV i sprawdzenie katalogu w HDFS	W HDFS w katalogu <b>gefs</b> znajduje się przekonwertowany plik CSV	<pre>vagrant@node1:~/bigdataproject\$ hadoop fs -ls /user/bigdataproject/gefs SLF4J: Class path contains multiple SLF4J bindings. SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.7.6/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticMDCBinder.class] SLF4J: Found binding in [jar:file:/usr/local/apache-tez-0.9.1-bin/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticMDCBinder.class] SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation. SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory] Found 2 items -rw-r--r-- 1 root supergroup 53614 2024-01-07 21:45 /user/bigdataproject/gefs/gefs_20240107.csv</pre>
9	Sprawdzenie, czy dane z plików CSV są prawidłowo wstawiane do tabeli Apache Hive	Uruchomienie procesorów Apache Nifi odpowiedzialnych za wstawienie zawartości pliku CSV do tabeli i sprawdzenie zawartości tej tabeli	W tabeli Apache Hive znajdują się dane z pliku CSV	<pre>hive&gt; select * from gefs where time = '2024-01-07' limit 5; OK 40.0 0.0 2024-01-07 00:00:00.0 2024-01-14 00:00:00.0 0.86 -0.69 40.0 0.5 2024-01-07 00:00:00.0 2024-01-14 00:00:00.0 1.15 0.12 40.0 1.0 2024-01-07 00:00:00.0 2024-01-14 00:00:00.0 2.1 1.4 40.0 1.5 2024-01-07 00:00:00.0 2024-01-14 00:00:00.0 3.08 2.17 40.0 2.0 2024-01-07 00:00:00.0 2024-01-14 00:00:00.0 3.88 2.57 5 rows selected (0.367 seconds) hive&gt;</pre>
10	Sprawdzenie, czy dane z tabeli Apache Hive są prawidłowo agregowane do tabeli Apache HBase	Uruchomienie procesorów Apache Nifi odpowiedzialnych za agregację najnowszych danych z tabeli Apache Hive i sprawdzenie zawartości docelowej tabeli Apache HBase	W tabeli Apache HBase znajdują się zagregowane dane	<pre>hbase(main):047:0&gt; scan 'gefs_agg' ROW 20240114      COLUMN=CELL column=agg:u10_avg_north, timestamp=2024-01-08T20:26:33.254, value=5.043117 20240114      column=agg:u10_avg_south, timestamp=2024-01-08T20:26:33.254, value=1.023 20240114      column=agg:u10_avg_total, timestamp=2024-01-08T20:26:33.254, value=3.1257756 20240114      column=agg:v10_avg_north, timestamp=2024-01-08T20:26:33.254, value=-1.0790476190476188 20240114      column=agg:v10_avg_south, timestamp=2024-01-08T20:26:33.254, value=-0.031095238 20240114      column=agg:v10_avg_total, timestamp=2024-01-08T20:26:33.254, value=-0.550408163265306 1 row(s) Took 0.0655 seconds hbase(main):048:0&gt;</pre>

11	Sprawdzenie, czy tabela Apache Hive z danymi GEFS przed agregacją jest dostępna z poziomu Apache Spark	Połączenie się z tabelą Apache Hive z użyciem odpowiedniego polecenia SQL w Apache Spark	Tabela jest dostępna i zawiera odpowiednie dane meteorologiczne	<pre>spark.sql("select * from gefs where latitude is not null;").show(n=10)</pre> <table><thead><tr><th> latitude </th><th>longitude </th><th>time </th><th>valid_time </th><th>u10 </th><th>v10 </th></tr></thead><tbody><tr><td>  45.0 </td><td>5.0 </td><td>2023-11-24 00:00:00 </td><td>2023-12-01 00:00:00 </td><td>0.25 </td><td>1.71 </td></tr><tr><td>  45.0 </td><td>5.5 </td><td>2023-11-24 00:00:00 </td><td>2023-12-01 00:00:00 </td><td>-1.06 </td><td>1.77 </td></tr><tr><td>  45.0 </td><td>6.0 </td><td>2023-11-24 00:00:00 </td><td>2023-12-01 00:00:00 </td><td>-1.0 </td><td>2.34 </td></tr><tr><td>  45.0 </td><td>6.5 </td><td>2023-11-24 00:00:00 </td><td>2023-12-01 00:00:00 </td><td>0.37 </td><td>2.29 </td></tr><tr><td>  45.0 </td><td>7.0 </td><td>2023-11-24 00:00:00 </td><td>2023-12-01 00:00:00 </td><td>0.53 </td><td>0.37 </td></tr><tr><td>  45.0 </td><td>7.5 </td><td>2023-11-24 00:00:00 </td><td>2023-12-01 00:00:00 </td><td>0.26 </td><td>-0.64 </td></tr><tr><td>  45.0 </td><td>8.0 </td><td>2023-11-24 00:00:00 </td><td>2023-12-01 00:00:00 </td><td>0.41 </td><td>-0.7 </td></tr><tr><td>  45.0 </td><td>8.5 </td><td>2023-11-24 00:00:00 </td><td>2023-12-01 00:00:00 </td><td>0.26 </td><td>-1.45 </td></tr><tr><td>  45.0 </td><td>9.0 </td><td>2023-11-24 00:00:00 </td><td>2023-12-01 00:00:00 </td><td>-0.61 </td><td>-1.2 </td></tr><tr><td>  45.0 </td><td>9.5 </td><td>2023-11-24 00:00:00 </td><td>2023-12-01 00:00:00 </td><td>-1.2 </td><td>-0.68 </td></tr></tbody></table> <p>only showing top 10 rows</p>	latitude	longitude	time	valid_time	u10	v10	45.0	5.0	2023-11-24 00:00:00	2023-12-01 00:00:00	0.25	1.71	45.0	5.5	2023-11-24 00:00:00	2023-12-01 00:00:00	-1.06	1.77	45.0	6.0	2023-11-24 00:00:00	2023-12-01 00:00:00	-1.0	2.34	45.0	6.5	2023-11-24 00:00:00	2023-12-01 00:00:00	0.37	2.29	45.0	7.0	2023-11-24 00:00:00	2023-12-01 00:00:00	0.53	0.37	45.0	7.5	2023-11-24 00:00:00	2023-12-01 00:00:00	0.26	-0.64	45.0	8.0	2023-11-24 00:00:00	2023-12-01 00:00:00	0.41	-0.7	45.0	8.5	2023-11-24 00:00:00	2023-12-01 00:00:00	0.26	-1.45	45.0	9.0	2023-11-24 00:00:00	2023-12-01 00:00:00	-0.61	-1.2	45.0	9.5	2023-11-24 00:00:00	2023-12-01 00:00:00	-1.2	-0.68
latitude	longitude	time	valid_time	u10	v10																																																																	
45.0	5.0	2023-11-24 00:00:00	2023-12-01 00:00:00	0.25	1.71																																																																	
45.0	5.5	2023-11-24 00:00:00	2023-12-01 00:00:00	-1.06	1.77																																																																	
45.0	6.0	2023-11-24 00:00:00	2023-12-01 00:00:00	-1.0	2.34																																																																	
45.0	6.5	2023-11-24 00:00:00	2023-12-01 00:00:00	0.37	2.29																																																																	
45.0	7.0	2023-11-24 00:00:00	2023-12-01 00:00:00	0.53	0.37																																																																	
45.0	7.5	2023-11-24 00:00:00	2023-12-01 00:00:00	0.26	-0.64																																																																	
45.0	8.0	2023-11-24 00:00:00	2023-12-01 00:00:00	0.41	-0.7																																																																	
45.0	8.5	2023-11-24 00:00:00	2023-12-01 00:00:00	0.26	-1.45																																																																	
45.0	9.0	2023-11-24 00:00:00	2023-12-01 00:00:00	-0.61	-1.2																																																																	
45.0	9.5	2023-11-24 00:00:00	2023-12-01 00:00:00	-1.2	-0.68																																																																	
12	Sprawdzenie, czy tabela Apache HBase z danymi GEFS po agregacji jest dostępna z poziomu warstwy analitycznej	Połączenie się z tabelą Apache HBase z użyciem dedykowanej biblioteki happybase i sprawdzenie zawartości rekordu dla odpowiedniego klucza	Są zwracane odpowiednie, zagregowane dane dla zadanego klucza	<pre>connection = happybase.Connection("localhost")  table = connection.table('gefs_agg') table.row(b"20231201")</pre> <pre>{b'agg:u10_avg_north': b'-1.4940259740259734',  b'agg:u10_avg_south': b'-0.22376190476190475',  b'agg:u10_avg_total': b'-0.8891383219954643',  b'agg:v10_avg_north': b'1.9530735930735934',  b'agg:v10_avg_south': b'1.195142857142857',  b'agg:v10_avg_total': b'1.5921541950113374'}</pre>																																																																		