

Apache Oozie



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



Celownik – najfiszce inwestycji
Projekt nr POKL.04.01.00-00-131/12 „Inżynieria wiedzy dla inteligentnego rozwoju”
współfinansowany przez Unię Europejską w ramach środków Europejskiego Funduszu Społecznego



UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



MATERIAŁY DYDAKTYCZNE I SZKOLENIOWE NA STUDIA PODYPLOMOWE ORAZ NA SZKOLENIA DYSTRYBUOWANE SĄ BEZPŁATNE

str. 1

Apache Oozie



- system harmonogramowania zadań w Hadoop,
- kolekcja akcji w formie grafu (DAG),
- “control dependency” (kolejna akcja nie może się zacząć, jeżeli wcześniejsza się nie skończyła),
- definicja zapisywana w pliku XML (za pomocą hPDL – Hadoop Process Definition Language),



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



Celownik – najfiszce inwestycji
Projekt nr POKL.04.01.00-00-131/12 „Inżynieria wiedzy dla inteligentnego rozwoju”
współfinansowany przez Unię Europejską w ramach środków Europejskiego Funduszu Społecznego



UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



MATERIAŁY DYDAKTYCZNE I SZKOLENIOWE NA STUDIA PODYPLOMOWE ORAZ NA SZKOLENIA DYSTRYBUOWANE SĄ BEZPŁATNE

Workflow w Oozie

- control flow nodes
 - definiują początek i koniec przepływu,
 - kontrolują ścieżkę wykonania,
- action nodes
 - inicjują wykonanie akcji



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



Celownik – najfiszca inwestycji
Projekt nr POKL.04.01.00-00-151/12 „Inżynieria wiedzy dla inteligentnego rozwoju”
wspofinansowany przez Unię Europejską w ramach środków Europejskiego Funduszu Społecznego

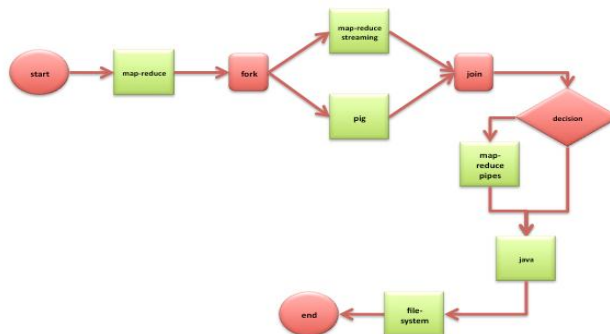


UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



MATERIAŁY DYDAKTYCZNE I SZKOLENIOWE NA STUDIA PODYPŁOMOWE ORAZ NA SZKOLENIA DYSTRYBUOWANE SĄ BEZPŁATNIE

Przykładowy workflow



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



Celownik – najfiszca inwestycji
Projekt nr POKL.04.01.00-00-151/12 „Inżynieria wiedzy dla inteligentnego rozwoju”
wspofinansowany przez Unię Europejską w ramach środków Europejskiego Funduszu Społecznego



UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



MATERIAŁY DYDAKTYCZNE I SZKOLENIOWE NA STUDIA PODYPŁOMOWE ORAZ NA SZKOLENIA DYSTRYBUOWANE SĄ BEZPŁATNIE

Action nodes

- uruchamiają zadania obliczeniowe i procesujące,
- zazwyczaj czekają, aż zadanie się skończy,
- dwa stany zakończenia: ok i err
- określają typ zadania do wykonania
 - MapReduce
 - skrypt
- dodatkowe elementy ułatwiające wykonanie
 - prepare element



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



Celownik – najfiszce inwestycji
Projekt nr POKL.04.01.00-00-151/12 „Inżynieria wiedzy dla inteligentnego rozwoju”
wspofinansowany przez Unię Europejską w ramach środków Europejskiego Funduszu Społecznego



UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



MATERIAŁY DYDAKTYCZNE I SZKOLENIOWE NA STUDIA PODYPŁOMOWE ORAZ NA SZKOLENIA DYSTRYBUOWANE SĄ BEZPŁATNE

Control flow nodes (1)

- start node
 - krok wejściowy dla przepływu,
 - wskazuje pierwszy node do wykonania,
- end node
 - jeżeli przepły dotrze do tego elementu oznacza poprawne zakończenie
- kill node
 - pozwala przepływowi na zniszczenie samego siebie,
 - podowuje zakończenie przepływu błędem



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



Celownik – najfiszce inwestycji
Projekt nr POKL.04.01.00-00-151/12 „Inżynieria wiedzy dla inteligentnego rozwoju”
wspofinansowany przez Unię Europejską w ramach środków Europejskiego Funduszu Społecznego



UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



MATERIAŁY DYDAKTYCZNE I SZKOLENIOWE NA STUDIA PODYPŁOMOWE ORAZ NA SZKOLENIA DYSTRYBUOWANE SĄ BEZPŁATNE

Control flow nodes (2)

- decision node
 - pozwala na wybranie gałęzi przepływu,
 - składa się z listy par <predykat, tranzycja>
- fork nodes, join nodes
 - fork rozdziela przepływ na kilka równoległych przepływów,
 - join czeka aż wszystkie gałęzie utworzone przez fork się zakończą,
 - fork i join muszą być używane razem



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



Celownik – najfiszce inwestycji
Projekt nr POKL.04.01.00-00-151/12 „Intynieria wiedzy dla inteligentnego rozwoju”
wspofinansowany przez Unię Europejską w ramach środków Europejskiego Funduszu Społecznego



UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



MATERIAŁY DYDAKTYCZNE I SZKOLENIOWE NA STUDIA PODYPŁOMOWE ORAZ NA SZKOLENIA DYSTRYBUOWANE SĄ BEZPŁATNE

Dostępne funkcje (1)

- podstawowe
 - timestamp()
 - concat()
- dające dostęp do atrybutów przepływu
 - wf:id() - workflowe ID
 - wf:name() - nazwa aplikacji
 - wf:user() - nazwa użytkownika, który uruchomił workflow



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



Celownik – najfiszce inwestycji
Projekt nr POKL.04.01.00-00-151/12 „Intynieria wiedzy dla inteligentnego rozwoju”
wspofinansowany przez Unię Europejską w ramach środków Europejskiego Funduszu Społecznego



UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



MATERIAŁY DYDAKTYCZNE I SZKOLENIOWE NA STUDIA PODYPŁOMOWE ORAZ NA SZKOLENIA DYSTRYBUOWANE SĄ BEZPŁATNE

Dostępne funkcje (2)

- funkcje HDFS
 - fs:exists() - czy dany URI istnieje
 - fs:isDir() - czy dany URI istnieje i jest katalogiem?
 - fs:blockSize() - rozmiar bloku dla danego pliku



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



Celownik – rozbija szary miazgę!

Projekt nr POKL.04.01.00-10-151/12 „Instynieria wiedzy dla inteligentnego rozwoju”

wspofinansowany przez Unię Europejską w ramach Europejskiego Funduszu Społecznego



UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



MATERIAŁY DYDAKTYCZNE I SZKOLENIOWE NA STUDIA PODYPŁOMOWE ORAZ NA SZKOLENIA DYSTRYBUOWANE SĄ BEZPŁATNE