

**Akademia Górniczo-Hutnicza
im. Stanisława Staszica w Krakowie**

Katedra Informatyki



PRACA MAGISTERSKA

PIOTR KOZŁOWSKI

**PRZENOŚNY KLASTER OBliczeniowy oparty na
urządzeniach SoC**

PROMOTOR:
dr hab. inż. Aleksander Byrski

Kraków 2015

OŚWIADCZENIE AUTORA PRACY

OŚWIADCZAM, ŚWIADOMY ODPOWIEDZIALNOŚCI KARNEJ ZA POŚWIADCZENIE NIEPRAWDY, ŻE NINIEJSZĄ PRACĘ DYPLOMOWĄ WYKONAŁEM OSOBIŚCIE I SAMODZIELNIE, I NIE KORZYSTAŁEM ZE ŹRÓDEŁ INNYCH NIŻ WYMIESZCZONE W PRACY.

.....
PODPIS

AGH
University of Science and Technology in Krakow

Department of Computer Science



MASTER OF SCIENCE THESIS

PIOTR KOZŁOWSKI

MOBILE COMPUTING CLUSTER BASED ON SoC DEVICES

SUPERVISOR:
Aleksander Byrski Ph.D

Krakow 2015

Podziękowania...

Spis treści

1. Wstęp.....	6
2. Obliczenia równoległe oraz rozproszone	8
2.1. Model architektur równoległych Flynnna.....	8
2.2. Modele programowania równoległego	10
2.3. Istniejące rozwiązania.....	13
2.4. Scala i Akka	15
3. Od SoC do IoT.....	20
3.1. SoC	20
3.2. IoT.....	23
4. Mobilny kластer obliczeniowy	25
4.1. Koncepcja	25
4.2. Platforma	25
4.2.1. Architektura	25
4.2.2. Wybrani aktorzy	27
4.2.3. Protokoły komunikacji.....	29
4.3. Inne wykorzystane technologie	31
5. Możliwości praktycznego zastosowania.....	33
5.1. Zastosowania	33
5.2. Równoległe algorytmy ewolucyjne	33
5.2.1. Algorytmy genetyczne	33
5.2.2. Model wyspowy	35
5.3. Problemy benchmarkowe.....	35
5.4. Ewolucyjna optymalizacja na platformie	35
5.5. Wyniki testów	40
6. Podsumowanie.....	43
6.1. Wnioski.....	43
6.2. Możliwości rozwoju	43

1. Wstęp

Motywacja

Obliczenia równoległe a także systemy rozproszone od lat są wykorzystywane przez różne ośrodki badawcze jak również przemysł do obliczeń wielkiej skali, a także wspomagania systemów wysokiej dostępności. Innymi obszarami zastosowań są różnego rodzaju symulacje, np. w biologii, chemii, astrofizyce lub medycynie a także przyśpieszenie krytycznych obliczeń w takich dziedzinach jak systemy obronne, przemysł komunikacyjny, meteorologia, astrodynamika, ponadto rozwiązywanie problemów optymalizacyjnych i przetwarzanie dużej ilości danych.

Przyczyniło się to do silnego rozwoju poruszanej tematyki, lecz niektóre koncepcje opracowane już w połowie XX wieku kiedy to trudno było przewidzieć skalę z jaką wzrasta ilość nowych urządzeń oraz danych do przetwarzania w dzisiejszych czasach napotykającą pewne bariery. Cały czas kreowane są nowe idee w myśl rozwoju współczesnych technologii co sprawia, że tematyka niniejszej pracy nie traci na aktualności. Takie koncepcje jak BigData lub IoT motywują do rozwoju prac związanych z wydajnym przetwarzaniem dużej ilości danych.

Potrzeby rynku oraz zwiększoną podaż spowodowały spadek cen urządzeń mobilnych, a także obniżenie stosunku ceny do mocy obliczeniowej. Urządzenia SoC zaczęto wykorzystywać na większą skalę w edukacji jak i domowych zastosowaniach. Pomimo, że proces miniaturyzacji oraz zwiększania wydajności procesorów osiąga bariery technologiczne poprzez rozwój systemów wieloprocesorowych udaje się nadal zwiększać efektywność obliczeń. Powstają aplikacje i technologie mające na celu połączenie zasobów w jedną całość oraz efektywne ich wykorzystanie. Zasoby te mogą być zgromadzone lokalnie lub rozproszone w różnych rejonach świata i połączone w sieci Internet. Natomiast dzięki elastycznym formom handlu jak na przykład sprzedaż zużytej mocy obliczeniowej oraz popularyzacji rozwiązań chmur obliczeniowych udaje się w coraz łatwiejszy sposób docierać do klienta indywidualnego oraz mniejszych firm. Możliwość użycia tego typu rozwiązań przez zwykłych użytkowników ma wpływ na tworzenie społeczności wspierających oraz testujących technologie o otwartym kodzie źródłowym.

Cel pracy

Celem ogólnym niniejszej pracy jest przygotowanie rozwiązań potrafiącego koordynować pracę klastra obliczeniowego złożonego z urządzeń SoC działających w sieci lokalnej.

Do celów szczegółowych należy rozpoznanie tematyki obliczeń równoległych w środowisku rozproszonym, zbadanie oraz wybór odpowiednich technologii dzięki którym możliwe będzie zaprojektowanie architektury a następnie zaimplementowanie platformy cechującej się mobilnością, skalowalnością oraz zapewniającej wsparcie w sytuacji awarii urządzeń pracujących w klastrze. Głównym zadaniem

powstałej platformy będzie zarządzanie obliczeniami prowadzonymi w klastrze, ich load-balancing a ponadto będzie ona również odpowiedzialna za komunikację urządzeń oraz monitoring stanu klastra czyli powinna być zdolna do informowania użytkownika jeżeli którykolwiek urządzenie opuści lub dołączy do klastra. Wynika z tego, że oprócz części platformy odpowiedzialnej za obliczenia i zarządzanie zadaniami oraz stanem klastra wymagany będzie interfejs użytkownika, dzięki któremu będzie on mógł zlecać zadania oraz mieć podgląd na stan działającego klastra. Dobór technologii oraz rozwiązań będzie brał pod uwagę ograniczone możliwości obliczeniowe typowych urządzeń SoC oraz mobilność powstałej platformy dlatego faworyzowane będą multiplatformowe technologie wysokiego poziomu, lecz na tyle wydajne aby możliwe było osiągnięcie zadowalającej efektywności obliczeń przy wszystkich ograniczeniach.

Ostatnim celem będzie przetestowanie opracowanego rozwiązania poprzez implementację jednego z modeli programowania równoległego wykorzystanego do prowadzenia wymagających obliczeń matematycznych. Zbadana zostanie skalowalność, wydajność oraz niezawodność. Następnie zostaną opracowane wyniki, a także zidentyfikowane możliwości zastosowań jak również perspektywy dalszego rozwoju powstałej platformy.

Opis rozdziałów

W pierwszym rozdziale została zawarta motywacja oraz cel pracy.

W kolejnym rozdziale zamieszczono wprowadzenie do tematyki obliczeń równoległych, wraz z podziałem architektur oraz opisem kilku popularnych modeli programowania jak również istniejących technologii.

W rozdziale trzecim opisano koncepcję IoT oraz kilka przykładowych urządzeń SoC.

Rozdział czwarty zawiera ideę przygotowanej platformy oraz jej szczegółowe implementacyjne.

W przedostatnim rozdziale przedstawiono możliwe zastosowania przygotowanego rozwiązania wraz z wynikami jego testów.

Ostatni rozdział, szósty, zawiera wnioski oraz naświetlono w nim możliwości dalszego rozwoju platformy stworzonej na potrzeby niniejszej pracy.

2. Obliczenia równoległe oraz rozproszone

Rozdział zawiera klasyfikację modelu architektur równoległych a także opis kilku wybranych technologii mających zastosowanie przy obliczeniach równoległych oraz rozproszonych.

2.1. Model architektur równoległych Flynn'a

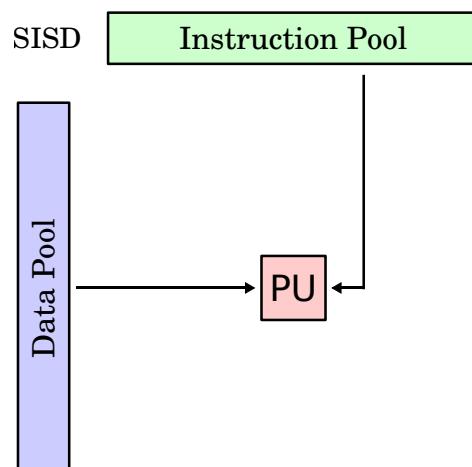
W latach 60-tych Michael Flynn sklasyfikował architektury komputerowe w czterech kategoriach biorąc pod uwagę ilość instrukcji oraz ilość strumieni danych przetwarzanych przez procesory. Poniżej opisano je w skrócie oraz pokazano schematy poglądowe.

SISD (ang. Single Instruction, Single Data)

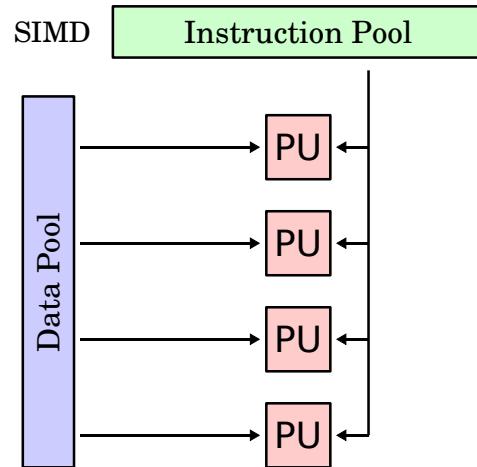
Jeden strumień danych jest przetwarzany sekwencyjnie przez jedną instrukcję w każdym kolejnym cyklu procesora (rysunek 2.1). Wyniki wykonania instrukcji są deterministyczne. Wykorzystany w komputerach skalarnych lub tzw. mainframe'ach. Przykład klasycznej architektury von Neumanna.

SIMD (ang. Single Instruction, Multiple Data)

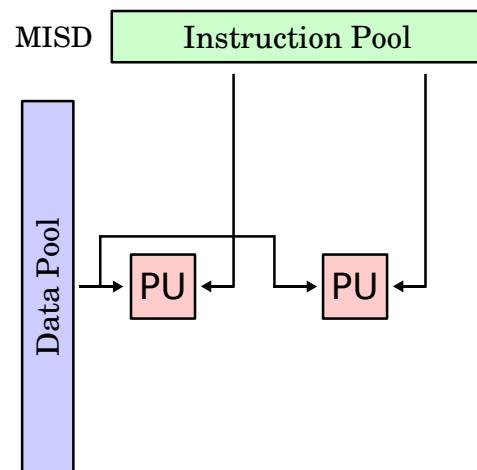
Ta sama instrukcja jest wykonywana równolegle na wielu procesorach używających oddzielnych strumieni danych (rysunek 2.2). Każdy procesor posiada swoją pamięć dla danych. Uzyskiwane wyniki są deterministyczne a instrukcje wykonywane są sekwencyjnie (lockstep). Model ten dobrze pasuje do danych, które cechują się pewną regularnością, co sprawdza się np. przy przetwarzaniu obrazów. Jako przykład wykorzystania tej architektury można tutaj podać komputery wektorowe lub współczesne procesory GPU wykorzystywane m.in. w kartach graficznych.



Rysunek 2.1: Schemat SISD [51].



Rysunek 2.2: Schemat SIMD [51].



Rysunek 2.3: Schemat MISD [51].

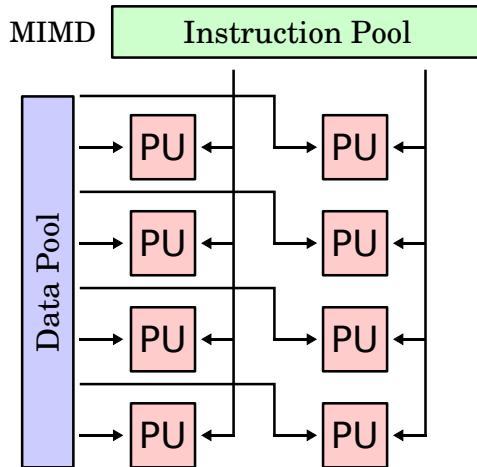
MISD (ang. Multiple Instruction, Single Data)

Wiele niezależnych instrukcji przetwarza równolegle ten sam strumień danych (rysunek 2.3). Model mający słabe zastosowanie komercyjne, mógłby zostać wykorzystany w przypadku potrzeby redundancji obliczeń.

MIMD (ang. Multiple Instruction, Multiple Data)

Wiele procesorów wykonuje równolegle instrukcje przetwarzające niezależne strumienie danych (rysunek 2.4). Wykonanie może być synchroniczne jak i niesynchroniczne jak również deterministyczne lub nie. Podejście bardziej elastyczne niż SIMD a co za tym idzie mające szersze zastosowanie. Jako przykład można tutaj podać systemy wieloprocesorowe, klastry lub gridy [69]. Większość współczesnych komputerów korzysta z tej architektury. Ta architektura może być stworzona z wielu komponentów SIMD.

Wyżej wymieniona klasyfikacja jest jedną z najstarszych oraz najbardziej ogólnych, będącą zazwyczaj punktem wyjścia do innych, bardziej szczegółowych klasyfikacji lub biorących pod uwagę inne aspekty zrównoleglania obliczeń. W klasycznej pozycji książkowej [69] rozważono ponadto



Rysunek 2.4: Schemat MIMD [51].

równoległość na różnych poziomach, takich jak instrukcje (Instruction-Level Parallelism), dane (Data-Level Parallelism) oraz wątki (Thread-Level Parallelism).

2.2. Modele programowania równoległego

Poniższy rozdział zawiera wstęp do kilku modeli programowania równoległego opartych na architekturze MIMD. Modele te istnieją jako abstrakcja do fizycznej architektury sprzętowej, co oznacza, że w niektórych przypadkach mogą być zaimplementowane na różnych architekturach sprzętowych.

SPMD (ang. Single Program, Multiple Data)

W tym modelu wiele procesorów wykonuje równolegle ten sam program przetwarzając różne strumienie danych. Jest to najbardziej powszechny model programowania równoległego, który może być wykonywany na większości współczesnych procesorach powszechnego użytku. W większości przypadków model ten nie wymaga synchronizacji tak jak w SIMD, gdyż procesory wykonują różne niezależne części programu i nie muszą się one wykonywać w tzw. lockstepie ([32]) tak jak w przypadku SIMD.

MPMD (ang. Multiple Program, Multiple Data)

Kolejny model programowania równoległego, który może mieć szersze zastosowanie aniżeli SPMD. Wiele niezależnych procesorów wykonuje równolegle conajmniej dwa niezależne programy.

Model SPMD implikuje kolejną klasyfikację ze względu na dostęp do pamięci biorąc pod uwagę pamięć współdzieloną oraz rozproszoną czyli Shared Memory oraz Distributed Memory.

Shared Memory (SM)

Procesy współdzielą jakiś obszar pamięci przechowujący dane, które odczytują lub zapisują asynchronicznie. Nie musi tutaj występować komunikacja pomiędzy procesami natomiast potrzebne są jakieś mechanizmy do kontroli dostępu do danych, takie jak synchronizacja, semafory, blokady (ang. lock) lub bariery pamięci zapewniające poprawność wykonywania operacji przy zachowaniu spójności danych. W związku z wymienionymi problemami programy korzystające z pamięci współdzielonej mogą być niedeterministyczne.

Distributed Memory (DM)

Każdy proces posiada swoją własną przestrzeń pamięci, co oznacza, że nie ma bezpośredniego dostępu do pamięci innych procesów. Co za tym idzie wymagana jest komunikacja pomiędzy procesami w celu wymiany informacji przez co mogą wystąpić jakieś opóźnienia w obliczeniach związane z narzutem komunikacji. Obszary pamięci nie muszą być podzielone sprzętowo, co oznacza, że ten model programowania może być zaimplementowany na sprzęcie posiadającym pamięć współdzieloną podzieloną logicznie pomiędzy procesy. Problemami które mogą wystąpić w opisywanym modelu jest load-balancing, podzielenie zadania na odpowiednie procesy jak również łączenie wyników zadań po zakończeniu obliczeń.

Poniżej wymieniono kilka innych modeli programowania równoległego z uwzględnieniem tego czy korzystają one z pamięci współdzielonej lub nie.

Data Parallel

Jest jedna zorganizowana globalna przestrzeń danych, np. w postaci tablicy na której operują procesy wykonujące się równolegle. Procesy wykonują operację na swojej wydzielonej części pamięci. Operacja może polegać na przykład na dodaniu jakieś liczby do kilku elementów tablicy. Jeżeli tablica posiada sto elementów a dziesięć różnych procesów wykonuje na niej jakieś operację, to każdy proces może mieć dostęp do dziesięciu elementów tablicy, co oznacza, że żaden proces nie będzie operował na tym samym elemencie tablicy. Wykorzystuje model SM.

Message Passing

W tym modelu procesy, zwane również taskami posiadają dane na których operują. Każdy task posiada unikalny identyfikator i komunikuje się z innymi taskami przesyłając jakieś informacje. Komunikacja musi być zazwyczaj skoordynowana, co oznacza, że po operacji wysłania, musi nastąpić odbiór wiadomości na innym tasku. Taski mogą wykonywać się równolegle, mogą być również tworzone dynamicznie. Każdy task może wykonywać się na oddzielnym procesorze, może być również umieszczony na innej fizycznej maszynie. Implementacją tego modelu jest standard MPI, który zostanie opisany później. Model ten wykorzystuje DM.

Threads

W tym modelu jeden „ciężki” proces może posiadać kilka wykonywanych wspólnie ścieżek, tzw. wątków. Dla przykładu założymy, że procesem jest program, który po starcie uruchamia kilka zadań wykonywanych równolegle przez różne wątki. Wykonywaniem równoległych wątków zajmuje się system operacyjny na którym ten program jest uruchomiony. Każdy wątek może posiadać swoje lokalne dane, które są prywatne i niewidoczne dla innych wątków ale może również współdzielić dane zaalokowane przez główny program z innymi wątkami, tzw. dane globalne. W przypadku gdy wątki potrzebują się ze sobą komunikować, mogą to zrobić właśnie przez dane globalne widziane z innych wątków. W takim przypadku jeśli chcemy uniknąć sytuacji, że dwa wątki próbują modyfikować ten sam obszar pamięci w tym samym czasie musi nastąpić synchronizacja dostępu do niego. Wątki mogą być tworzone i usuwane dynamicznie przez główny program i dzięki niemu mogą również współdzielić zasoby. Wykorzystuje model SM. Problemy które mogą wystąpić w przypadku tego modelu, to są blokady wątków (ang. deadlock), zagłodzenie wątków (ang. starvation) oraz tzw. race condition [74, 62, 69].

W modelu Shared Memory jeżeli procesy operują na wspólnych danych utrudnieniem jest kontrola ich dostępu oraz zapewnienie zadowalającej wydajności obliczeń. Różne technologie oraz modele programowania równoległego zapewniają rozwiązań, które czasem mogą okazać się niewystarczające lub źle wpływać na proces zrównoleglania. W przypadku gdy współzielony stan (pamięć) jest mutowalny, czyli może być modyfikowany przez wiele procesów rozwiązań powinna być synchronizacja dostępu. Jeżeli jednak każdy proces posiada odizolowany stan, który nie może być mutowalny, to synchronizacja jest niepotrzebna lecz tracimy również możliwość współdzielenia stanu. Jednym z rozwiązań tych problemów jest wykorzystanie podejścia z języków funkcyjnych takich jak Scala lub Erlang, czyli niemutowalnych obiektów. Modelem wykorzystującym niemutowalne obiekty do przesyłania komunikatów jest model aktorowy wykorzystany m.in. we wspomnianych wyżej językach funkcyjnych oraz w wielu innych do zrównoleglania obliczeń.

Model Aktorowy (ang. Actor Model)

W modelu aktorowym głównym bytem są aktorzy. Komunikują się oni ze sobą asynchronicznie poprzez wymianę wiadomości a więc są trochę podobni do tasków w modelu Message Passing. Po wysłaniu wiadomości nie muszą czekać na odpowiedź a kolejność dostarczania wiadomości jest nieistotna. Wiadomości są buforowane w skrzynkach odbiorczych. Każdy aktor przechowuje jakiś stan, który nie powinien być widoczny dla innych aktorów i może zostać zmieniony jedynie przez tego samego aktora. Model aktorowy wykorzystuje Shared Memory. Zaimplementowano go m.in. w Erlangu lub Scali. Zaletą tego modelu jest to, że nie trzeba się martwić o współdzielenie skomplikowanych stanów lecz z drugiej strony problem nie zawsze da się podzielić na mniejsze części i zaimplementować z wykorzystaniem aktorów co można traktować jako wadę tego podejścia [2].

STM (ang. Software Transactional Memory)

Wprowadza pojęcie atomicznych transakcji. Każdy dostęp do pamięci współzielonej objęty jest transakcją, której idea jest podobna do transakcji bazodanowych. Jest to alternatywa do blokowania dostępu do pamięci poprzez synchronizację. Transakcja obejmuje serię odczytów i zapisów do pamięci współzielonej a jej stan nie jest widoczny dla innych transakcji dopóki nie zostanie ona zakończona i za-komitowana. Wyróżnia się tutaj sekcje krytyczne zwane blokami atomicznymi. Jeżeli wykonanie takiego bloku się nie powiedzie zmiany są wycofywane, co wymusza przechowywanie starszej wersji danych. Transakcje mogą wykonywać się równolegle. Zaimplementowany w wielu językach programowania, m.in. w języku Scala [72, 9].

W literaturze można znaleźć jeszcze wiele innych modeli programowania równoległego wykorzystujących podobne koncepcje jak te opisane powyżej jak również wprowadzające zupełnie inne podejścia. Ostatnim modelem, który zostanie opisany jest Master/Slave ze względu na szersze wykorzystanie go w niniejszej pracy.

Master/Slave

Master/Slave może być implementacją modelu SPMD. W modelu Master/Slave jedno urządzenie lub proces zwane masterem kontroluje jedno lub więcej urządzeń lub procesów zwanych slave lub worker

[33]. Obliczenia wykonywane przez workery odbywają się równolegle, natomiast master odpowiada za load-balancing oraz łączenie wyników w jedną całość.

Może okazać się przydatny w następujących przypadkach:

- gdy chcemy mieć jeden proces, tzw. single-point of responsibility, który podejmuje decyzje lub koordynuje akcje w celu zachowania spójności w systemie,
- potrzebujemy zapewnić jeden punkt dostępu do zewnętrznego systemu lub jeden punkt wejścia do systemu z zewnątrz,
- istnieje potrzeba stworzenia scentralizowanego serwisu, np. zajmującego się routingiem,
- istnieje możliwość podzielenia zadania na kilka części, z których każda może być wykonywana równolegle.

Rozwiązanie to ma niestety również kilka wad, jak na przykład istnienie niebezpieczeństwa utworzenia tzw. wąskiego gardła co może spowodować problemy z wydajnością lub być tzw. single-point of failure. Jeśli urządzenie/proces posiadające rolę master ulega awarii, powinniśmy w jakiś sposób obsłużyć taką sytuację, np. poprzez ponowne wystartowanie procesu/urządzenia z rolą master aby zapewnić poprawne i niezawodne działanie komunikacji. Sytuacje takie mogą wprowadzić pewne opóźnienia w trakcie odzyskiwania sprawności systemu [22].

2.3. Istniejące rozwiązania

OpenMP (ang. Open Multi-Processing)

Standard używany na skalę przemysłową, korzystający z modelu Threads i pamięci współdzielonej. Dostępne są implementacje w językach C/C++ oraz Fortran. Pozwala na wielowątkowe obliczenia w których wiele wątków pracuje na tym samym zestawie danych. Pierwsza wersja standardu została opublikowana w roku 1997 dla języka Fortran [37].

MPI (ang. Message Passing Interface)

Popularny standard szeroko stosowany w obliczeniach równoległych jak również w przemyśle, powstał w roku 1994. Korzysta z modelu Message Passing i Distributed Memory. Najbardziej znane implementacje to OpenMPI lub MPICH dla języków C oraz Fortran. Zapewnia komunikację przez sieć dla danych przetwarzanych na różnych fizycznych maszynach [34].

CUDA (ang. Compute Unified Device Architecture)

Jest to technologia opracowana w 2007 roku przez firmę NVIDIA, producenta kart graficznych. Wykorzystuje układy GPU a co za tym idzie architekturę SIMD. Pozwala wykonywać obliczenia równoległe. Dostarcza SDK oraz API pozwalających na pisanie aplikacji z wykorzystaniem języka C/C++ lub Fortran. Ma zastosowanie w przetwarzaniu video, biologii obliczeniowej i chemii, astrofizyce oraz różnego rodzaju symulacjach lub analizach [35].

OpenCL (ang. Open Computing Language)

Technologia o otwartym kodzie źródłowym w przeciwieństwie do CUDA, tworzony przez grupę Khronos od 2009 roku. Wykorzystująca do obliczeń również procesory graficzne. Dodatkowo wspierająca obliczenia równoległe na zwykłych procesorach CPU. Podeczas gdy CUDA wykorzystuje karty graficzne firmy NVIDIA, OpenCL wspiera również platformy AMD, Intel, ARM oraz wiele innych. Technologia jest oparta o język C++ [36].

Java Concurrency Framework

Większość współczesnych obiektowych języków programowania wspiera pojęcie współbieżności poprzez wątki, które operują na pamięci współdzielonej. Nie inaczej jest z Java, w której od wersji 5 wprowadzono wiele usprawnień do wsparcia wielowątkowości m.in. relację happens-before, która wymusza, że jeden dostęp do pamięci musi nastąpić przed innymi. Poza wieloma mechanizmami wspomagającymi różne poziomy synchronizacji wprowadzono również wysokopoziomowe API, wiele kolekcji oraz typów danych z wbudowanymi mechanizmami wspierającymi współbieżność i ułatwiającymi programistom implementowanie obliczeń równoległych [31].

Poza frameworkami wymienionymi powyżej, których użycie oraz przygotowanie końcowego rozwiązania często wymaga dużo czasu, powstało również kilka gotowych systemów oferujących sprawdzone implementacje niektórych problemów i będące punktem wyjścia dla bardziej skomplikowanych rozwiązań.

Apache Hadoop

Jest to framework open-source napisany w Javie i wspierający obliczenia równoległe prowadzone na dużych ilościach danych w środowisku rozproszonym, takim jak klastry komputerowe. Pierwsza wersja została opublikowana w 2005 roku. Jego dwoma głównymi komponentami są HDFS (Hadoop Distributed File System) wykorzystujący technologię HBase i służący do przechowywania danych oraz MapReduce służący do przetwarzania danych. Kolejnym istotnym komponentem jest YARN pomagający zarządzać zasobami dostępnymi w klastrze oraz wykorzystywany do harmonogramowania zadań. Zamierzeniem twórców było stworzenie narzędzia dobrze skalowalnego, oraz zapewniającego wysoką dostępność poprzez wykrywanie oraz wsparcie urządzeń pracujących w klastrze [11].

HTCondor

Jest systemem służącym do zarządzania zasobami oraz kolejkowaniem zadań w architekturach rozproszonych. Jest to projekt open-source rozwijany przez Uniwersytet Wisconsin-Madison. Powstał w latach 80-tych. Może być uruchamiany na wielu platformach takich jak Windows czy Linux. Dostarcza mechanizmy zabezpieczeń takie jak autentykacja, autoryzacja czy szyfrowanie danych jak również mechanizmy do wspomagania administrowania systemem. Może być użyty do rozłożenia obciążenia na komputerach pracujących w klastrze jak również do wykorzystania wolnych zasobów podłączonych do niego komputerów zwykłych użytkowników w celu wykonywania obliczeń dużej skali i korzystając z okazji gdy komputery są nieobciążone. Wspiera równoległe wykonywanie zadań napisanych w stan-

dardzie MPI. Może również zarządzać zasobami znajdującymi się w chmurze obliczeniowej lub gridzie [27].

Wymienione wyżej narzędzia implementują różne modele programowania równoległego. Warto również wspomnieć o takich architekturach jak systemy gridowe lub klastry komputerowe, które są polem do testowania wspomnianych technologii oraz w dużym stopniu przyczyniły się do ich rozwoju.

Klastry

Klaster komputerowy jest zbiorem wielu komputerów połączonych zazwyczaj w sieci LAN mających na celu rozwiązywanie jakiegoś złożonego problemu obliczeniowego oraz współpracujących ze sobą. Zazwyczaj są używane do zwiększenia wydajności oraz dostępności jakichś usług. W porównaniu do gridów komputery pracujące w klastrze mają podobną architekturę oraz zazwyczaj zajmują się wykonywaniem podobnych, skoordynowanych zadań. Mogą posiadać dostęp poprzez sieć do magazynu danych, który można traktować jako pamięć współdzieloną. W typowym wykorzystaniu klastra istnieje dodatkowo maszyna klienta, która dystrybuje zadania do klastra i zbiera wyniki.

Gridy

Systemy które integrują heterogeniczne zasoby będące pod kontrolą różnych domen i połączone siecią komputerową WAN. Zapewniają narzędzia do zarządzania zasobami, uwierzytelniania oraz autoryzacji w celu rozwiązywania problemów dużej skali. Gridy są rozszerzeniem idei klastrów. Mogą zapewniać takie usługi jak globalny magazyn danych, aplikacje sieciowe, Software as a Service czyli pojęcia znane z chmur obliczeniowych [74, 65].

2.4. Scala i Akka

Technologiom Scala i Akka został poświęcony osobny rozdział, ze względu na wykorzystanie ich w części praktycznej niniejszej pracy. Scala posiada wsparcie dla dwóch modeli programowania równoległego takich jak STM oraz Actor Model, który został wykorzystany dzięki użyciu frameworka Akka. Z kolei rozszerzenia Akka, takie jak Akka Cluster wspierają użycie tego frameworka w klastrach komputerowych.

Scala

Scala jest językiem funkcyjnym rozwijanym od 2001 roku przez firmę Typesafe, której współtwórcą jest Martin Odersky na codzień pracujący na Uniwersytecie w Lozannie. Głównym aspektem przemawiającym na korzyść tego języka jest to, że do jego działania wystarczy maszyna wirtualna Javy. Jest to język obiektowy podobnie jak Java ale łączy również zalety języków funkcyjnych, które w ostatnim czasie stają się coraz bardziej popularne. Zamiarem twórców było stworzenie języka eleganckiego oraz zwięzłego syntaktycznie. Mimo, że Scala jest językiem dynamicznie typowanym zapewnia tzw. type safety [53]. Wprowadza również wiele innowacji oraz ciekawych rozwiązań do konstrukcji języka jak np. case classes, currying, zagnieżdżanie funkcji, DSL, tail recursion, słowo kluczowe lazy lub trait, konstrukcja podobna do interfejsu z języka Java ale mogącą posiadać częściową implementację. Scala

preferuje obiekty immutable. Obsługuje również funkcje wyższego rzędu oraz pozwala zwięzle definiować funkcje anonimowe, kładzie również duży nacisk na skalowalność. Dodatkową zaletą tego języka jest to, iż jest on w pełni kompatybilny z językiem Java, co oznacza, że możemy w nim używać bibliotek lub frameworków napisanych w Javie bez żadnych dodatkowych deklaracji. Ostatecznie program napisany w Scali jest komplikowany do kodu bajtowego Javy [55, 57, 20, 71]. Według autora niniejszej pracy do wad tego języka można zaliczyć m.in. to, że niektóre instrukcje da się wyrazić na wiele różnych sposobów, co utrudnia czytelność kodu oraz zwiększa trudność nauki tej technologii.

Akka

Jednym z kluczowych frameworków wykorzystanych w niniejszej pracy jest Akka. Akka jest projektem Open Source na licencji Apache 2 powstającym w roku 2009. Posiada wersję przeznaczoną dla języka Java jak również dla języka Scala. Stał się częścią implemenacji języka Scala od wersji 2.10. Projekt Akka mimo stosunkowo niedługiej obecności na rynku jest w pełni gotowy do zastosowań produkcyjnych. Cechuje się obecnością wielu interesujących z punktu widzenia niniejszej pracy rozszerzeń, które zostały opisane w jednym z następnych podrozdziałów, wyczerpującej dokumentacji oraz dużej i aktywnej społeczności rozwijającej ten produkt a także wsparciem komercyjnych firm. Akka używa modelu aktorowego aby zwiększyć poziom abstrakcji i oddzielić logikę biznesową od niskopoziomowego zarządzania wątkami oraz operacjami I/O [67, 19].

Podstawowym bytem w technologii Akka są aktorzy. Aktor zapewnia wysokopoziomową abstrakcję dla lepszej współprzeźności oraz zrównoleglenia operacji. Jest on lekkim, wydajnym oraz sterowanym zdarzeniami procesem, który komunikuje się z innymi aktorami za pomocą asynchronicznych i niemutowalnych wiadomości przechowywanych w skrzynkach odbiorczych, które posiada każdy aktor. Aktor enkapsuluje pewien stan i zachowania, które realizują określone zadania. Więcej informacji na temat koncepcji aktorów w projekcie Akka można znaleźć w pozycji [19].

Każdy aktor w systemie może być identyfikowany na kilka różnych sposobów, głównym z nich jest unikalny adres aktora dzięki któremu może zostać znaleziony przez innych aktorów np. w celu wysłania wiadomości. Aktorzy działają w obrębie systemu zwanego Actor System opisanego w pozycji [1]. System aktorowy można traktować jako pewną strukturę z zaalokowaną pulą wątków, stworzoną w ramach jednej logicznej aplikacji. Konfiguracją puli wątków zajmuje się Akka. Pewne ustawienia mogą być zmienione programatycznie lub w pliku *application.conf*, który jest głównym plikiem konfiguracyjnym systemu aktorowego. System aktorowy zarządza dostępymi zasobami i może mieć uruchomione miliony aktorów gdyż instancja każdego z aktorów zajmuje zaledwie około 300 bajtów pamięci.

Akka Framework dostarcza wszystkich zalet programowania reaktywnego [44] oraz zapewnia między innymi:

- współprzeźność, dzięki zaadaptowaniu modelu aktorowego, programista może zatem skupić się na logice biznesowej zamiast zajmować się problemami współprzeźności,
- skalowalność, asynchroniczna komunikacja pomiędzy aktorami dobrze skaluje się w systemach multiprocesorowych,

- odporność na błędy, framework Akka zapożyczył podejście z języka Erlang, co pozwoliło wykorzystać model *let it crash* [56] zwany też *fail fast* do zapewnienia niezawodnego działania systemu i skrócenia jego niedostępności,
- architekturę sterowaną zdarzeniami,
- transakcyjność,
- ujednolicony model programowania dla potrzeb wielowątkowości oraz obliczeń rozproszonych,
- Akka wspiera zarówno API języka Java jak i języka Scala [67].

Zastosowania projektu Akka są bardzo szerokie, można je odnaleźć chociażby w następujących dziedzinach:

- analiza danych,
- bankowość inwestycyjna,
- eCommerce,
- symulacje,
- media społecznościowe,
- batch processing,
- transaction processing (online gaming, social media, telecom, finanse, zakłady),
- aplikacje real-time,
- serwisy REST, SOAP, message hub.

Systemy w których potrzebujemy uzyskać wysoką przepustowość oraz małe opóźnienia są dobrym kandydatem do wykorzystania framework'a Akka [19].

Akka Cluster

Jest to rozszerzenie projektu Akka pozwalające zaimplementować komunikację pomiędzy urządzeniami działającymi w obrębie klastra. Zapewnia ono pewien poziom abstrakcji dla protokołu TCP/IP [8].

Dostarcza również odporny na awarie oraz zdecentralizowany serwis członkostwa (ang. membership service) oparty na protokole Gossip [40] oraz automatycznej detekcji niedziałających węzłów (ang. failure detector).

Pojęcia:

- węzeł (ang. node) - logiczny członek klastra, może istnieć wiele węzłów na jednej fizycznej maszynie, identyfikowany krotką: nazwa_hosta:port:uid,
- klaster (ang. cluster) - grupa węzłów zarejestrowana w serwisie członkostwa,

- lider (ang. leader) - węzeł odpowiedzialny za kluczowe akcje pozwalające zachować odpowiedni stan klastra w przypadku dołączania nowych lub awarii istniejących węzłów [19].

Lider jest tylko rolą jaką posiada dany węzeł, każdy węzeł może zostać liderem oraz każdy węzeł jest w stanie w sposób deterministyczny wyznaczyć lidera. Węzły mogą też posiadać inne role, które mogą się przydać np. w ograniczeniu zasięgu komunikacji.

Protokół Gossip wspomniany powyżej pozwala rozpropagować stan klastra do wszystkich jego węzłów, tak aby każdy węzeł miał takie same informację o pozostałych członkach klastra. Protokół ten pomaga uzyskać zbieżność stanu klastra we wszystkich jego węzłach w skończonym czasie.

Członkostwo w klastrze jest rozpoczynane komendą *join* wysyłaną do jednego z aktywnych członków klastra, gdy każdy węzeł klastra otrzyma informację o dołączającym węzle dzięki protokołowi Gossip, taki węzeł może zostać uznany za osiągalny. Może on jednak utracić ten stan gdy np. wystąpią jakieś problemy z siecią. Stan nieosiągalny może trwać jedynie określony czas po upływie którego jeśli węzeł nie powróci do pełnej sprawności traci on status członka klastra. Każdy węzeł poza nazwą hosta oraz portem jest identyfikowany dodatkowo unikalnym identyfikatorem, tzw. uid, co pozwala na uruchomienie kilku węzłów klastra na jednej fizycznej maszynie. Jeżeli członek klastra ulegnie awarii lub opuści klaszter na własne życzenie nie może on ponownie dołączyć do klastra dopóki system aktorowy uruchomiony na nim nie zostanie zrestartowany, co pozwoli na wygenerowanie nowego uid. Każdy węzeł może zmienić swój stan członkostwa lub może on zostać zmieniony automatycznie dzięki automatycznej detekcji niedziałających węzłów. Ustawienia detekcji niedziałających węzłów oraz protokołu Gossip są konfigurowalne, tzn. że można np. ustalić czas po którym węzeł zostanie uznany za nieosiągalny. Zdarzenia związane ze zmianą stanu węzłów klastra mogą być subskrybowane przez istniejących członków, co ułatwia implementację monitoringu stanu klastra. Więcej szczegółów na temat idei członkostwa w klastrze oraz dostępnych stanów węzłów można znaleźć w pozycjach [19, 70].

Kolejną ciekawą opcją w rozszerzeniu Akka Cluster jest wykorzystanie bibliotek Sigar [49] oraz integracja z nimi. Pozwalają one na dostęp do informacji systemowych takich jak zużycie CPU, wykorzystanie pamięci RAM, stan sieci. Mogą zostać wykorzystane do implementacji load balansingu lub monitorowania obciążenia urządzeń klastra [4, 3].

Cluster Singleton

Cluster Singleton jest rozszerzeniem projektu Akka zapewniającym jedną instancję aktora danego typu w obrębie klastra lub w obrębie węzłów z wybraną rolą. Może zostać wykorzystane do zaimplementowania modelu Master/Slave opisanego w jednym z poprzednich podrozdziałów. Rozszerzenie to dostarcza implementacji menedżera, który pozwala zarządzać instancjonowaniem aktora-singletona. Dostęp do działającej instancji jest możliwy z każdego urządzenia działającego w klastrze i odbywa się poprzez aktora pośrednika, tzw. proxy [22].

Distributed Publish Subscribe in Cluster

Rozszerzenie projektu Akka, które umożliwia komunikację między aktorami bez posiadania informacji na których konkretnie urządzeniach poszczególni aktorzy są uruchomieni, czyli lokacja aktora jest transparentna z punktu widzenia komunikacji. Dostarcza ono aktora-mediatora, który zarządza rejestracją innych aktorów do konkretnych kanałów komunikacji którymi są zainteresowani. Zachowanie to, można porównać z subskrypcją RSS [77]. Wiadomość opublikowana w danym kanale powinna zostać

dostarczona do wszystkich aktorów, którzy zostali w nim uprzednio zarejestrowani. Rozszerzenie to pozwala również wysłać wiadomość do jednego lub większej ilości aktorów pasujących do określonego wzorca [16].

3. Od SoC do IoT

Niniejszy rozdział zawiera charakterystykę urządzeń SoC oraz opis koncepcji IoT.

3.1. SoC

Urządzenia SoC (System-on-a-chip) są to układy, które oprócz głównego procesora opartego np. na architekturze ARM, zawierają cyfrowe i analogowe moduły składające się w jeden system elektroniczny. Poszczególne moduły zazwyczaj są tworzone przez różnych producentów od których firmy produkujące urządzenia SoC zamawiają sprzęt. W porównaniu do mikrokontrolerów charakteryzują się one posiadaniem większej ilości pamięci RAM oraz CPU o stosunkowo dużej mocy obliczeniowej, która jest wystarczająca do uruchomienia systemów operacyjnych podobnych do tych uruchamianych na komputerach PC. Charakteryzują się również posiadaniem interfejsów umożliwiających podpinanie urządzeń peryferyjnych tych samych co do zwykłych komputerów PC. Urządzenia SoC zyskały swą popularność wśród zwykłych użytkowników dopiero w ostatnich latach, głównie za sprawą taniego mini komputera jakim jest Raspberry Pi. Mają one zastosowanie m.in. jako systemy wbudowane, w dziedzinie Home Automation jak również wiele innych multimedialnych zastosowań ze względu na możliwości zbliżone do zwykłych PC oraz niewielkie rozmiary. Z założenia powinny mieścić się na jednej płytce drukowej, być energoszczędne oraz tanie w produkcji seryjnej.

Typowe urządzenie SoC powinno zawierać:

- procesor CPU,
- pamięć RAM, ROM, EEPROM lub FLASH,
- układy czasowo-licznikowe,
- kontrolery transmisji szeregowej lub równoległej,
- przetworniki analogowo-cyfrowe lub cyfrowo-analogowe,
- obwody zarządzania zasilaniem.

Spotykane są również urządzenia zawierające procesory GPU oprócz jednostki CPU [50, 64].

Poniżej wymieniono kilka popularnych urządzeń SoC.

ParallelA

Urządzenie posiada dwurdzeniowy procesor ARM o taktowaniu 800MHz, wielordzeniowy procesor Epiphany o architekturze RISC oraz 1GB pamięci RAM. Dodatkowo posiada złącze kart MicroSD,

USB 2.0, port Ethernet, HDMI oraz możliwość uruchomienia systemów operacyjnych z rodziny Linux. Urządzenie jest wielkości karty kredytowej a jego cena oscyluje w granicach 100\$. Istnieje kilka różnych specyfikacji tego urządzenia różniących się ilością rdzeni procesora Epiphany się oraz ceną. Dzięki obecności procesora Epiphany urządzenie dobrze sprawdza się w obliczeniach równoległych, które mogą zostać zaimplementowane przy wykorzystaniu takich technologii jak OpenMP, MPI lub OpenCL. Projekt został zapoczątkowany na platformie Kickstarter w 2012 roku [38].

Intel Galileo

Urządzenie stworzone przez firmę Intel bazujące na procesorze tej samej firmy klasy x86 Pentium o taktowaniu 400MHz (32 bity) wraz z 256MB pamięci RAM. Jest kompatybilne z systemami operacyjnymi takimi jak Windows, Linux lub Mac OS a także z wieloma rozszerzeniami czy bibliotekami platformy Arduino. Cena oraz rozmiary urządzenia są podobne jak w przypadku urządzeń Parallel. Urządzenie to posiada również złącze Ethernet, MicroSD oraz mini PCI Express. Nie posiada natomiast karty dźwiękowej ani procesora graficznego [28].

Sharks Cove

Urządzenie promowane przez firmę Microsoft i kompatybilne z Windowsem 10 oraz systemem Android. Pracuje pod kontrolą procesora Intel Atom o taktowaniu 1.33GHz, wyposażone jest również w zintegrowany procesor graficzny Intel HD Graphics, 1GB pamięci RAM, kartę dźwiękową oraz m.in. w złącze MicroSD, HDMI, USB 2.0 oraz Ethernet. Posiada trochę większe rozmiary od innych opisywanych tutaj urządzeń. Cena urządzenia również jest wyższa, bo wynosi około 300\$ [48].

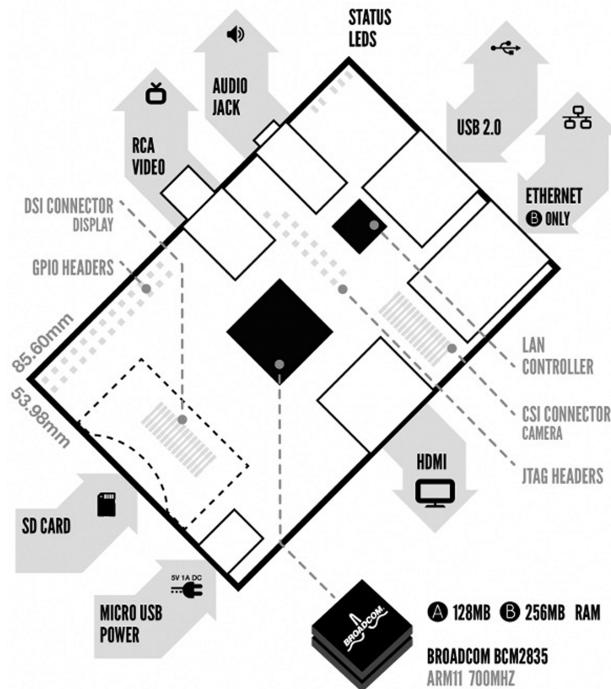
Raspberry Pi

Raspberry Pi jest jednym z urządzeń SoC szerzej wykorzystywanych w niniejszej pracy. Jest to urządzenie stworzone przez fundację non-profit oparte na architekturze ARM. Pierwsza wersja została wydana w roku 2012.

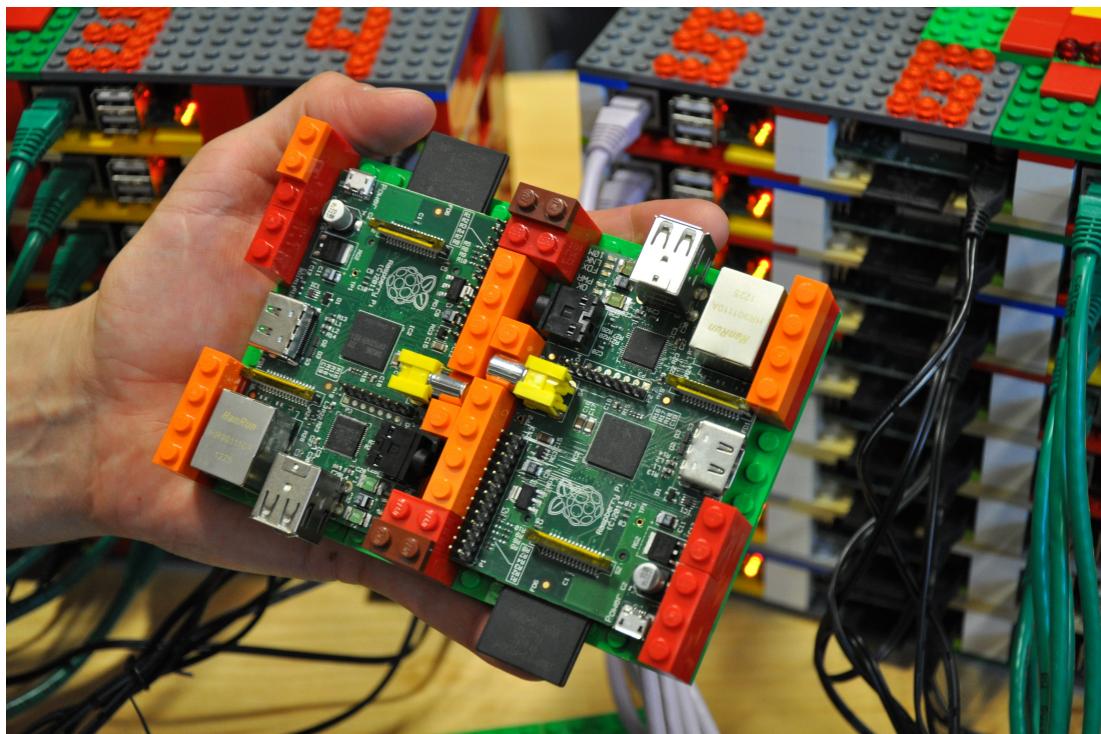
Na rysunku 3.1 przedstawiono schemat poglądowy urządzenia.

Urządzenie Raspberry Pi w wersji B posiada jednostkę CPU o taktowaniu 700MHz oraz 512MB pamięci RAM. Poza procesorem CPU jest również wyposażone m.in. w jednostkę GPU, dwa porty USB, złącze kart SD, złącze Ethernet 100Mb/s oraz złącze MicroUSB wykorzystane do zasilania. W trakcie obciążenia pobiera około 4W prądu. Na tym urządzeniu może być zainstalowana m.in. jedna z wielu dystrybucji Linuxa. Urządzenie to może służyć zarówno jako platforma deweloperska jak i centrum domowej rozrywki. Znalazło również zastosowanie w budowie klastrów obliczeniowych. Przykładem może być praca opisana w pozycji [59] oraz widoczna na rysunku 3.2. Pełną specyfikację urządzenia można znaleźć w pozycji [23].

W roku 2015 został udostępniony do dystrybucji nowy model urządzenia oznaczony wersją 2, który różni się od poprzedniej wersji m.in. tym, że posiada 1GB pamięci RAM oraz czterordzeniowy procesor ARM o taktowaniu 900MHz a także możliwość uruchomienia poza dystrybucjami Linuxa również systemu Windows w wersji 10. Cechuje się też mniejszym rozmiarem - wielkość karty kredytowej. Cena tej wersji urządzenia wynosi 35\$ [42].



Rysunek 3.1: Schemat Raspberry Pi model B [[23]].



Rysunek 3.2: Klaster Raspberry Pi [43].

Arduino

Jest to rodzina mikrokontrolerów bazująca na technologii Atmel AVR i bardzo popularna wśród użytkowników. Nie są to co prawda urządzenia SoC, ponieważ posiadają znacznie mniej pamięci, słabsze procesory oraz mniejszą ilość wbudowanych modułów ale są oparte na otwartej platformie, która jest łatwo programowalna, posiada dużą społeczność zainteresowanych osób przyczyniających się do jej rozwoju oraz wiele darmowych rozszerzeń a także bibliotek ułatwiających ich użycie oraz poszerzających możliwości tych urządzeń. Dzięki temu są one często wykorzystywane w połączeniu z urządzeniami SoC lub do sterowania innymi urządzeniami np. w Home Automation [12].

Dzięki popularyzacji, miniaturyzacji oraz obniżeniu cen urządzeń SoC oraz mikrokontrolerów rozwinięła się idea Internet Of Things. Firmy które wypuszczają nowe urządzenia na rynek starają się aktywizować społeczności wokół tych urządzeń, np. poprzez udostępnianie darmowych urządzeń we wczesnej fazie promocji zainteresowanym użytkownikom lub uczelniom. Dla przykładu można podać akcję Intel w przypadku urządzenia Galileo, gdzie aż 50 tys. egzemplarzy trafiło do tysiąca uczelni na całym świecie a wśród nich również do 18 polskich placówek [29].

3.2. IoT

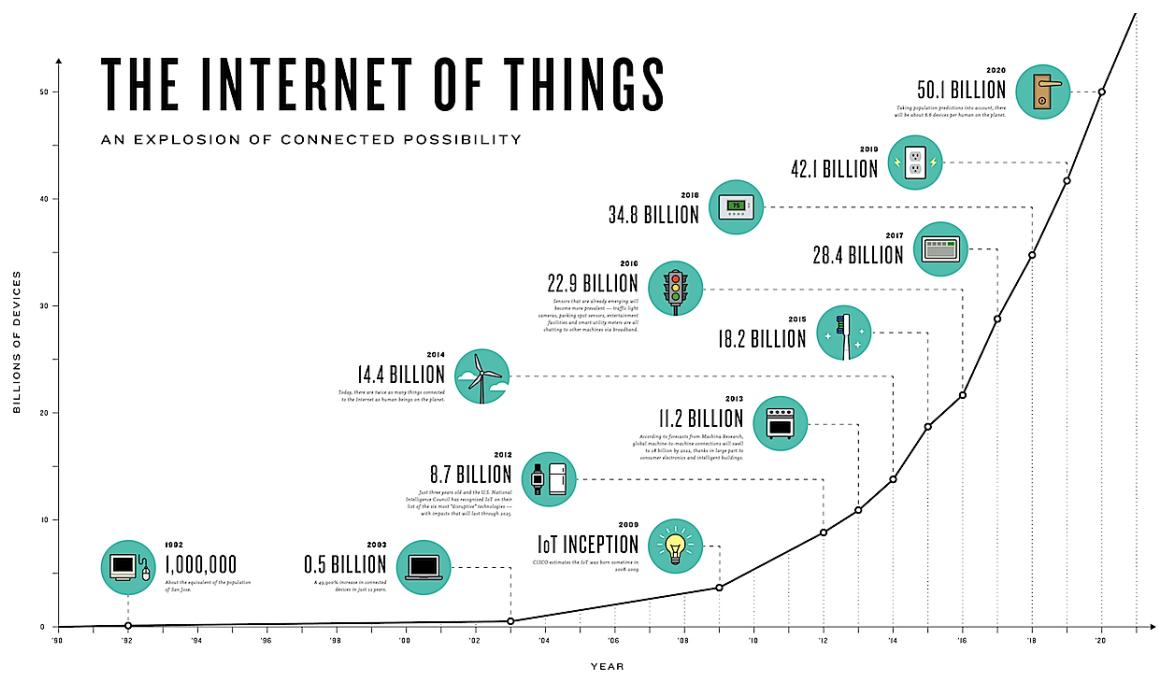
Współczesne systemy komputerowe stają się coraz mniejsze oraz tańsze, co skutkuje obecnością komputerów w prawie każdej dziedzinie życia. Dzięki rozwojowi urządzeń mobilnych, bezprzewodowej transmisji, technologiom działającym w chmurze zwiększającym zasięg Internetu coraz więcej urządzeń ma do niego dostęp. Skutkuje to tym, że ilość danych do przetwarzania ciągle wzrasta jak również wzrasta skala komunikacji tych urządzeń. Do tego typu urządzeń zaliczają się już nie tylko komputery osobiste i urządzenia mobilne takie jak smartfony, tablety czy inteligentne zegarki ale również urządzenia gospodarstwa domowego jak telewizory, lodówki, artykuły oświetleniowe i grzewcze oraz wiele innych urządzeń posiadających jakieś sensory i zbierających dane również na skalę przemysłową.

Koncepcja IoT (ang. Internet Of Things) rozwinięła się wokół wszystkich urządzeń podłączonych do Internetu. Zauważono tutaj wiele nowych możliwości jak również tendencję do wzrostu liczby urządzeń połączonych z Internetem. Firma Cisco szacuje, że do roku 2020 liczba takich urządzeń może przekroczyć 50 bilionów [14]. Na rysunku 3.3 przedstawiono poglądowy wykres tego zjawiska.

Istnieje również inne pojęcie zwane IoE (ang. Internet of Everything) zapoczątkowane przez firmę Cisco, które czasem jest używane zamiennie do IoT. Jest obecna strona internetowa zawierająca licznik pokazujący przybliżoną liczbę urządzeń podłączonych do Internetu, który aktualnie wskazuje ponad 16 bilionów i cały czas rośnie [14].

Wiele firm komercyjnych zauważa w tych prognozach sposób na zysk dlatego tanie, miniaturowe platformy komputerowe takie jak urządzenia SoC są bardzo mocno promowane.

Internet oraz wiele współczesnych technologii zostało wymyślonych o wiele wcześniej aniżeli koncepcja IoT więc nie wszystkie są gotowe na taką ilość urządzeń oraz danych. Pozostawia to miejsce do badań nad nowymi rozwiązaniami oraz istniejącymi problemami, których ilość będzie rosła wraz ze wzrostem skali zjawiska IoT [30, 60].



Rysunek 3.3: Prognozy ilości urządzeń podłączonych do Internetu [13].

4. Mobilny klaster obliczeniowy

Rozdział ten przedstawia opis przygotowanego rozwiązania.

4.1. Koncepcja

Na rysunku 4.1 zaprezentowano początkową koncepcję przedstawiającą sieć lokalną i działający w niej klaster obliczeniowy w skład którego wchodzą urządzenia Raspberry Pi, gdzie jedno z nich pełni rolę Master natomiast pozostałe pełnią rolę Slave. W klastrze znajduje się również uruchomiony serwer WWW służący do komunikacji z użytkownikiem za pomocą przeglądarki internetowej.

Na urządzeniach uruchomione są aplikacje oparte o technologie JVM, które komunikują się ze sobą za pomocą protokołu TCP/IP.

Opisac na podstawie notatek/pytan/todosow z jakimi problemami sie zmagano.

load-balancing, wybór nowego mastera, wysyłanie progresu, persystencja, zapewnienie jednej instancji mastera, wykrywanie awarii, reakcje w przypadku awarii, jak konfigurować zadania, Zapisywać stan aktora na każdym node lub do zdalnej bazki czy przesyłać dane do jednego node'a i tam zapisywać, np. we frontend, Gdzie persystować dane, co potrzebujemy zapisywać, cała populacja co kilka ewaluacji (chyba nie po każdej?) czy może tylko najlepszego osobnika? - po jakiej ustalonej liczbie ewolucji można zrobić snapshot, czyli zapisać populację do bazki

czemu odpowiada aktor

4.2. Platforma

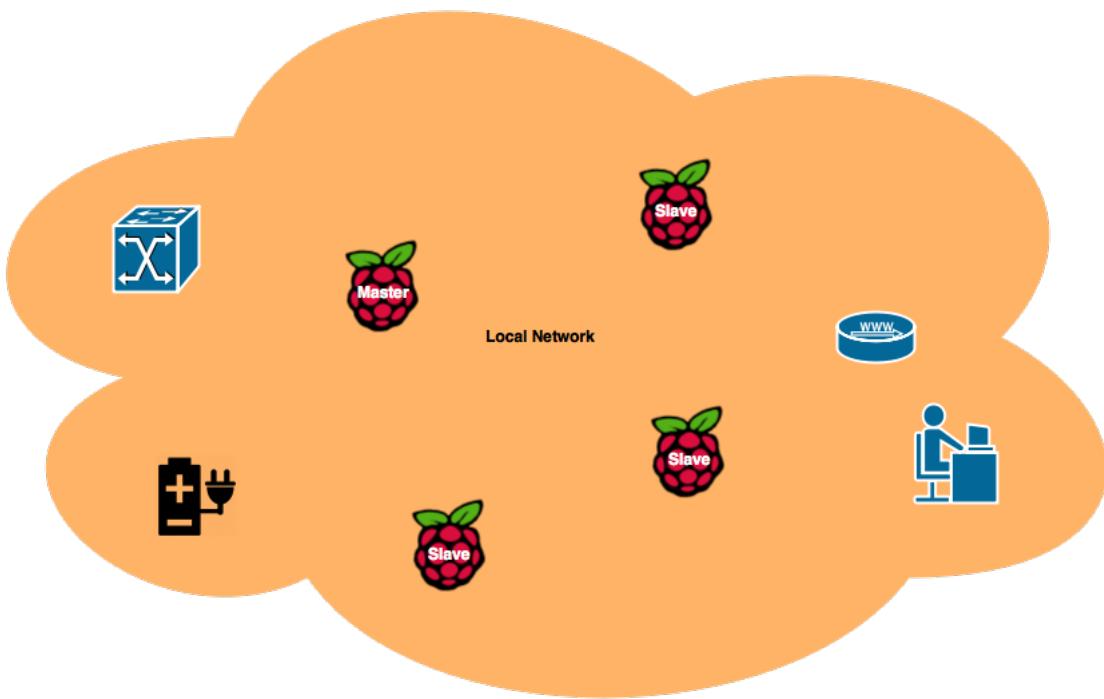
Niniejszy podrozdział omawia wybrane aspekty przygotowanej platformy, takie jak architektura, protokoły komunikacji oraz sposób przydzielania zadań.

4.2.1. Architektura

Na początek warto wyjaśnić kilka pojęć.

Węzeł klastra jest to instancja aplikacji niezwiązana z fizycznym urządzeniem. Co oznacza, że na jednym urządzeniu może zostać uruchomionych kilka węzłów, czyli kilka aplikacji. W skład platformy wchodzą właściwie dwie aplikacje nazwane *frontend* oraz *backend*. Posiadają one różne odpowiedzialności jak i różne rodzaje uruchomionych na nich aktorów, których dokładniejszy opis znajduje się w kolejnym podrozdziale.

Aplikacja *frontend* odpowiada za interakcję z użytkownikiem dzięki takim technologiom jak Play Framework, AngularJS, WebSocket oraz protokołowi HTTP. Komunikuje się ona z aplikacją *backend*

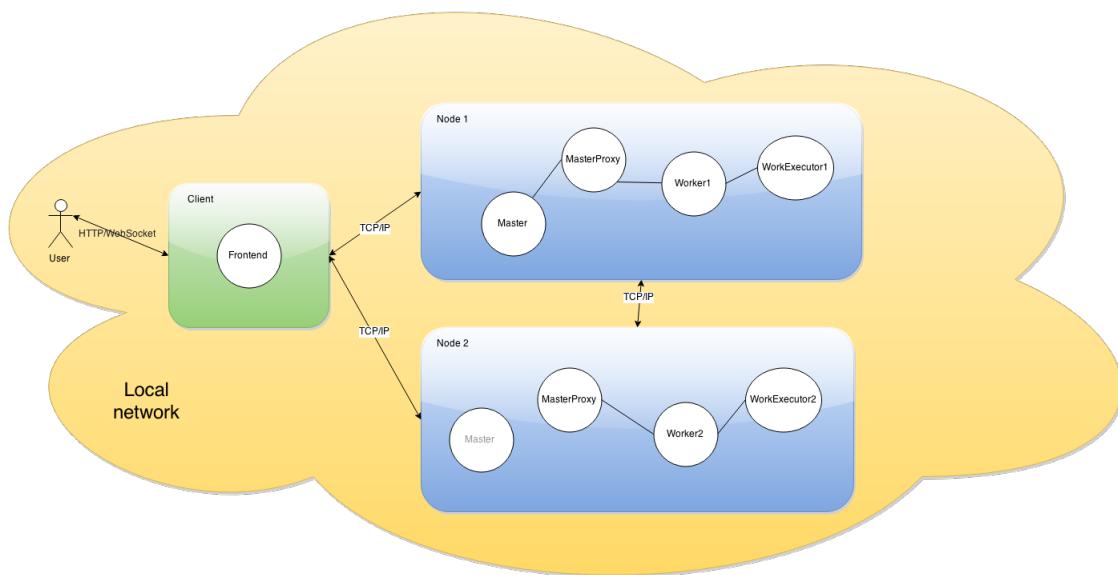


Rysunek 4.1: Wygląd klastra

w celu zlecenia wykonania zadania oraz odebrania wyników. Posiada ona interfejs webowy. Instancja aplikacji jest traktowana jako osobny węzeł klastra.

Aplikacja *backend* nie posiada interfejsu użytkownika. Zajmuje się realizacją zleconych zadań. Instancja tej aplikacji również odpowiada jednemu węzlowi klastra. W klastrze może być uruchomionych wiele takich instancji, po jednej na każdym urządzeniu lub jeśli urządzenie posiada wystarczająco dużo mocy obliczeniowej może posiadać ich kilka. Każda uruchomiona aplikacja posiada uruchomionego workera oraz tylko jedna z nich posiada również uruchomionego mastera, gdyż w klastrze powinna być tylko jedna aktywna instancja mastera.

Na rysunku 4.2 przedstawiono podgląd architektury, gdzie zielonym kolorem zaznaczono aplikację *frontend* natomiast niebieskim aplikację *backend*. Założono, że każda aplikacja jest uruchomiona na osobnym urządzeniu pracującym w klastrze, są dwa urządzenia wykonujące obliczenia oraz jedno urządzenie klienckie. Na wspomnianym rysunku pokazano także kilka aktorów uruchomionych na wybranych urządzeniach. Założono również, że urządzenia pracują w sieci lokalnej z wykorzystaniem technologii Ethernet. Użytkownik komunikuje się za pomocą protokołu HTTP z urządzeniem klienckim, które z kolei przesyła odpowiednie żądania do urządzenia posiadającego aktywnego mastera. Urządzenie klienckie nie musi posiadać informacji o tym gdzie dokładnie jest aktywny master, ponieważ dostęp do mastera odbywa się poprzez proxy. Natomiast aplikacja kliencka powinna mieć podany adres IP przynajmniej jednego urządzenia pracującego w klastrze. Może mieć informację o kilku adresach IP urządzeń pracujących w klastrze, co pozwala uniknąć problemów gdy jedno z urządzeń nie odpowiada. Najłatwiejszym sposobem podania adresów urządzeń pracujących w klastrze jest użycie dodatkowych argumentów podczas uruchamiania aplikacji. Szczegóły tych ustnień oraz sposobu uruchamiania ap-



Rysunek 4.2: Podgląd architektury.

likacji wchodzących w skład przygotowanej platformy zostały opisane w dodatku A załączonym do niniejszej pracy.

Urządzenia wykonujące obliczenia w klastrze komunikują się ze sobą za pomocą protokołu TCP/IP, którym zarządza Akka. Odkąd aktywny master posiada listę wszystkich zarejestrowanych workerów potrafi on w łatwy sposób nawiązywać z nimi kontakt oraz informować ich o nadchodzących zadaniach od klienta. Z kolei komunikacja w drugą stronę, np. gdy worker skończył wykonywać zadanie i chce wysłać zgromadzone wyniki odbywa się poprzez proxy, ponieważ worker nie wie na którym urządzeniu znajduje się aktywny master. Gdy master otrzyma wyniki zadania od realizującego je workera odsyła je do klienta za pomocą rozszerzenia *Distributed Publish Subscribe in Cluster* opisanego w jednym z następnych podrozdziałów. Co oznacza, że nie musi posiadać informacji o adresie IP klienta, ponieważ opublikowane wyniki trafią tylko do zainteresowanych odbiorców.

Na rysunku 4.2 pokazano jedynie kilka najważniejszych aktorów działających na urządzeniach będących częścią klastra. Zostali oni opisani razem z pozostałymi, którzy również są aktywni podczas działania platformy w następnym podrozdziale wraz z podziałem na aplikacje w których są wykorzystywani.

4.2.2. Wybrani aktorzy

Poniżej wymieniono kilka aktorów działających w klastrze z uwzględnieniem tego w jakiej aplikacji wchodzącej w skład przygotowanej platformy są oni uruchomieni.

Aplikacja *backend*:

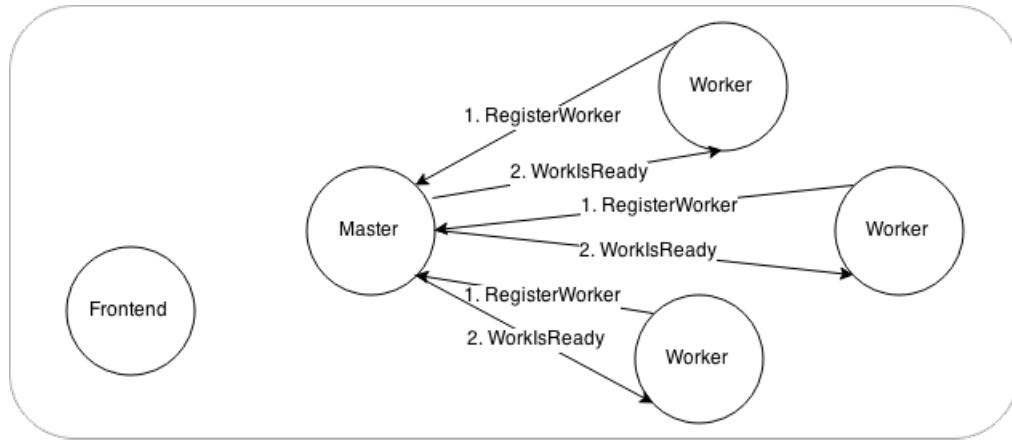
- *ClusterSingletonManager*, aktor dostarczony wraz z rozszerzeniem *Akka Cluster Singleton*, który odpowiada za to aby w klastrze była aktywna tylko jedna instancja mastera. Zostaje on uruchomiony na wszystkich węzłach lub tylko na węzłach z wybraną rolą. Instancjonuje aktora singletona na najstarszym węźle, czyli takim, który dołączył do klastra najwcześniej. Monitoruje on

również stan dostępności węzła na którym uruchomiony jest aktor singleton i w przypadku awarii startuje nową instancję mastera na innym węźle klastra.

- ClusterSingletonProxy, aktor współpracujący z poprzednim opisanym aktorem i będący pośrednikiem (ang. proxy) zapewniającym dostęp do aktora singletona czyli w tym przypadku mastera. Jest on uruchomiony na każdym węźle, który potrzebuje komunikować się z masterem i działa jak router przekierowujący wszystkie przychodzące wiadomości do aktora singletona. Jeśli aktor ten jest niedostępny np. ze względu na awarię lub jest w trakcie odzyskiwania sprawności, zadaniem aktora ClusterSingletonProxy jest przechowywanie wszystkich odebranych wiadomości aż do momentu ponownego nawiązania komunikacji z masterem [22].
- DistributedPubSubMediator, aktor ten zostaje uruchomiony na wszystkich węzłach klastra lub tylko na węzłach z wybraną rolą. Zarządza rejestracją innych aktorów do wybranych kanałów komunikacji oraz replikuje tą wiedzę pośród inne instancje pracujące na pozostałych urządzeniach klastra, tak aby zachować spójność tych danych w obrębie klastra. Zajmuje się również wysyłaniem wiadomości z każdego urządzenia klastra do zarejestrowanych aktorów pracujących na jakimkolwiek urządzeniu w obrębie klastra [16].
- Master, jest to tzw. aktor singleton. Oznacza to, że w obrębie klastra powinna być aktywna tylko jedna instancja tego aktora. Jest to zapewnione poprzez aktora ClusterSingletonManager opisanego powyżej. Dostęp do mastera uzyskuje się poprzez proxy, czyli aktora ClusterSingletonProxy z każdego urządzenia pracującego w klastrze. Głównymi zadaniami tego aktora są: przydzielanie zadań workerom, zarządzanie ich rejestracją, odbieranie wyników i przesyłanie ich do klienta, reagowanie w sytuacji awarii jednego z workerów oraz zapisywanie wyników do bazy.
- Worker, jest to aktor realizujący zadania otrzymane od mastera po uprzedniej rejestracji w serwisie mastera. Każdy worker uruchamia jedną instancję aktora wykonującego właściwe obliczenia, tzw. WorkExecutor, po to aby zachować odpowiedzialność podczas komunikacji z masterem. Zajmuje się on również procesem migracji, gdzie znowu jest wykorzystywany mechanizm publikowania w wybranym kanale dostarczony wraz z rozszerzeniem *Distributed Publish Subscribe*.
- WorkExecutor, aktor wykonujący właściwe obliczenia czyli realizujący zadanie zlecone przez klienta. Jest on instancjonowany przez workera i komunikuje się tylko z nim.

Aplikacja *frontend*:

- Frontend, aktor z którym komunikuje się użytkownik. Odbiera on wyniki zadań od mastera oraz przesyła mu zadania do wykonania. Do komunikacji z masterem używa proxy, natomiast wyniki odbiera dzięki rozszerzeniu *Distributed Publish Subscribe* oraz byciu zarejestrowanym w wybranym kanale komunikacji.
- Metrics, aktor subskrybujący zdarzenia związane ze zmianą różnych metryk klastra, np. obciążenia CPU lub pamięci RAM. Dzięki połączeniu WebSocket przesyła te informacje na widok.
- Monitor, aktor obserwujący zdarzenia związane z członkostwem w klastrze. Jeżeli jakiś węzeł dołącza do klastra lub go opuszcza taka informacja jest aktualizowana na widoku w aplikacji webowej.



Rysunek 4.3: Rejestracja Workerów.

Tablica 4.1: Opis komunikatów - rejestracja workerów.

Nazwa komunikatu	Opis
RegisterWorker	Komunikat rejestrujący workera o danym ID w serwisie mastera.
WorkIsReady	Informacja o tym, że master posiada zakolejkowane zadanie po które dany worker może się zgłosić.

Poza wyżej wymienionymi aktorami działającymi w aplikacji *frontend*, posiada ona również kilka podobnych aktorów co aplikacja *backend*, m.in. ClusterSingletonProxy oraz DistributedPubSubMediator, którzy służą do komunikacji z masterem oraz odbieraniem wyników realizowanych zadań.

4.2.3. Protokoły komunikacji

Na rysunkach 4.3 oraz 4.4 przedstawiono dwa najważniejsze protokoły definiujące kolejność oraz rodzaje przesyłanych wiadomości podczas komunikacji na linii Frontend - Master - Worker - WorkExecutor. Widoczni są tam również aktorzy biorący lub nie udział we wspomnianej komunikacji.

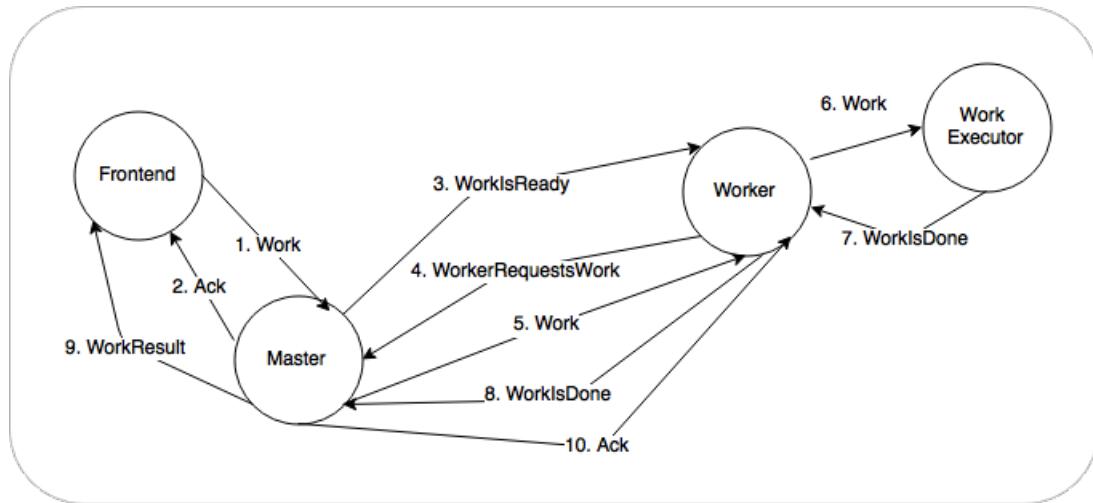
Na rysunku 4.3 przedstawiono proces rejestracji workerów w serwisie mastera, po to aby master mógł ich poinformować o tym, że posiada jakieś zadanie, które trzeba zrealizować.

W tabeli 4.1 przedstawiono opis komunikatów zawartych na rysunku 4.3.

Na rysunku 4.4 pokazano przepływ informacji podczas zlecania oraz realizacji zadania. Dla uproszczenia przedstawiono przypadek zadania zrealizowanego bez przesyłania częściowych wyników. W przypadku zadań trwających długo lub gdzie czas trwania nie jest z góry określony tzw. *long-running jobs* wprowadzono również komunikat *WorkInProgress*, który jest przesyłany cyklicznie w trakcie wykonywania zadania i zawiera jego częściowe wyniki, można go umiejscowić pomiędzy komunikatami *Work* a *WorkIsDone*.

W tabeli 4.2 przedstawiono opis komunikatów przesyłanych w trakcie procesu realizacji zadań.

Do odpowiedniego rozłożenia obciążenia pomiędzy pracującymi workerów początkowo rozważano podejście gdzie to master decyduje, któremu workerowi przydzielić zadanie bazując między innymi na metrykach takich jak obciążenie CPU lub zajętość pamięci RAM. Jednak okazało się ono nieefektywne w implementacji testowanych algorytmów genetycznych, ponieważ te metryki zmieniały się zazwyczaj z opóźnieniem pozostawiając workera w stanie bezczynności przez nieokreślony czas. Zdecydowano się



Rysunek 4.4: Protokół Master/Worker.

Tablica 4.2: Opis komunikatów przesyłanych w trakcie realizacji zadania.

Nazwa komunikatu	Opis
Work	Komunikat przechowujący parametry konfiguracyjne zadania jak również jego ID.
Ack	Potwierdzenie odebrania poprzedniego komunikatu.
WorkIsReady	Informacja o tym, że master posiada zakolejkowane zadanie po które dany worker może się zgłosić.
WorkerRequestsWork	Komunikat przechowujący ID workera i informujący mastera, że dany worker jest wolny i może zająć się kolejnym zadaniem.
WorkIsProgress	Komunikat przechowujący częściowe wyniki zadania jak również jego ID oraz ID workera.
WorkIsDone	Komunikat przechowujący WorkResult oraz ID workera.
WorkResult	Komunikat przechowujący końcowe wyniki zadania jak i jego ID.

zatem zaimplementować podejście zwane *Work Pulling Pattern* [54] w którym to wolny worker zgłasza się po zadanie a nie oczekuje na akcję mastera.

4.3. Inne wykorzystane technologie

Akka Persistence

Rozszerzenie Akka Persistence pozwala na zapisywanie stanu aktorów w bazie danych. Pozwala to na odtworzenie stanu aplikacji w przypadku awarii i zapewnia w pewnym stopniu niezawodność działania systemu [6]. Dostępnych jest wiele różnych dodatków dedykowanych dla różnych technologii bazodanowych. W niniejszej pracy wykorzystano dodatek przeznaczony do pracy z bazą MongoDB [7].

MongoDB

MongoDB jest bazą NoSQL opartą na dokumentach typu JSON, co pozwala na elastyczność oraz przejrzystość w przechowywaniu danych. Jest to technologia Open Source. Jako, że jest to baza NoSQL posiada ona elastyczny schemat danych. Wspiera replikację oraz sharding a także map reduce. Technologia ta została wykorzystana ze względu na jej łatwe użycie, dobrą skalowalność oraz wydajność przy wykorzystaniu stosunkowo niewielkich zasobów obliczeniowych [18].

Kolejnymi argumentami przemawiającymi na korzyść tej technologii są łatwa dostępność dokumentacji oraz integracja z innymi technologiami wykorzystanymi w niniejszej pracy.

Play Framework i AngularJS

Technologie Play Framework oraz AngularJS zostały wykorzystane do przygotowania aplikacji webowej będącej interfejsem klienta dzięki któremu może sterować on działaniem platformy, monitorować ją oraz oglądać wyniki obliczeń.

Play Framework jest to technologia server-side, działająca zarówno z językiem Java jak i Scala. Integruje ona komponenty oraz API potrzebne do stworzenia aplikacji webowej opartej o wzorzec MVC. Bazuje on na lekkiej, bezstanowej architekturze, niskim zużyciu zasobów, wysokiej skalowalności oraz programowaniu reaktywnym. Do komunikacji wykorzystuje metody protokołu HTTP oraz pozwala na implementację serwisów w technologii REST. Play zapewnia również wsparcie dla technologii WebSocket wykorzystanej w niniejszej pracy oraz łatwą integrację z frameworkiem Akka [21, 75].

Z kolei AngularJS jest to technologia stworzona z użyciem języka JavaScript, ze wsparciem komercyjnych firm takich jak Google. Rozszerza możliwości języka HTML oraz CSS a ponadto pozwala tworzyć tzw. SPA (ang. Single Page Applications). W niniejszej pracy wykorzystano ją do prezentowania informacji odebranych z aplikacji serwerowej oraz interakcji z użytkownikiem za pomocą przeglądarki internetowej [17, 63].

Wykorzystane szablony

Platforma Typesafe w której skład wchodzą technologie wykorzystane w niniejszej pracy, takie jak Play Framework, Akka oraz Scala dostarcza bardzo wyczerpującą dokumentację wraz z wieloma przykładami oraz szablonami na licencji Public Domain.

Rozwiązanie stworzone na potrzeby niniejszej pracy bazowało początkowo na dwóch szablonach, pierwszy zawierał przykład monitoringu urządzeń podłączonych do klastra natomiast drugi dystrybucję zadań pomiędzy urządzenia [39, 52].

5. Możliwości praktycznego zastosowania

W niniejszym rozdziale opisano wykorzystanie przygotowanej platformy w praktyce, sposób jej użycia oraz zasady działania a także wyniki zaimplementowanych rozwiązań.

5.1. Zastosowania

Poniżej wymieniono kilka możliwych zastosowań przygotowanej platformy:

- Symulacje.
- Obliczenia równolegle.
- Problemy optymalizacyjne.

5.2. Równoległe algorytmy ewolucyjne

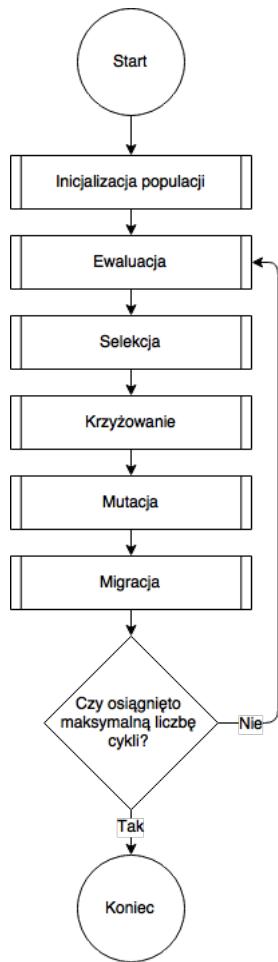
W rozdziale opisano kilka algorytmów użytych podczas implementacji rozwiązania stworzonego na potrzeby niniejszej pracy.

5.2.1. Algorytmy genetyczne

Algorytmy genetyczne zaliczają się do grupy algorytmów ewolucyjnych. Zostały opracowane przez Johna Hollanda w latach siedemdziesiątych i są inspirowane teorią ewolucji Darwina tudiżew ewolucją biologiczną [61]. Istnieje wiele różnych prac traktujących o wykorzystaniu algorytmów genetycznych, różnych ich odmianach czy modyfikacjach oraz badaniu ich efektywności. Przykładem może być praca magisterska zawarta w pozycji [61] lub trochę starsza pozycja książkowa [66]. Opis implementacji algorytmów genetycznych można znaleźć w pozycji [76].

Pojęcia

- Gen - cecha mająca wpływ na jakość rozwiązania.
- Chromosom - indywidual, osobnik prezentujący jedno z rozwiązań problemu, składa się z genów.
- Populacja - zbiór chromosomów, prezentuje zbiór rozwiązań danego problemu.
- Migracja - wymiana osobników pomiędzy populacjami.
- Selekcja - algorytm wyboru osobników pozostałych w danej populacji w następnym cyklu.



Rysunek 5.1: Schemat blokowy ewolucji.

- Krzyżowanie - tworzenie osobników potomnych w trakcie trwania ewolucji na podstawie innych, wybranych osobników - rodziców.
- Mutacja - modyfikacja osobnika zawierająca element losowości, mająca na celu próbę ekploracji nowych obszarów rozwiązań.
- Funkcja przystosowania (ang. fitness) - zwana też funkcją celu, określa jakość danego rozwiązania.
- Ewaluacja - wyliczenie wartości funkcji fitness dla każdego osobnika populacji czyli inaczej ocena jakości znalezionych rozwiązań.
- Ewolucja - powtarzający się cykl w trakcie którego następuje rozwój populacji oraz w skład którego wchodzi m.in. ewaluacja, selekcja, krzyżowanie, mutacja oraz migracja.
- Generacja - jeden cykl algorytmu.

Na rysunku 5.1 przedstawiono schemat blokowy ewolucji.

Zastosowanie

Algorytmy ewolucyjne sprawdzają się tam gdzie nie jest dobrze znana przestrzeń rozwiązań ale został określony sposób oceny jakości rozwiązania. Znajdują one zastosowanie przy rozwiązywaniu

problemów NP, np. problemu komiwojażera w którym trzeba znaleźć najkrótszą drogę łączącą wszystkie miasta tak, aby odwiedzić każde miasto conajmniej raz. Innymi zastosowaniami mogą być choćby poszukiwania ekstremów funkcji, których nie da się obliczyć analitycznie lub przeszukiwanie dużych przestrzeni rozwiązań jak np. w przypadku problemu grupowania. Algorytmy genetyczne są mniej zależne od wstępnej inicjalizacji zadania oraz mniej skonne do znajdywania rozwiązań lokalnych zamiast optymalnych [10, 61].

5.2.2. Model wyspowy

Jest to odmiana algorytmów ewolucyjnych przystosowana do obliczeń równoległych oraz rozproszonych. Zamiast jednej dużej populacji rozważamy tutaj kilka podpopulacji zwanych wyspami. Każda wyspa może być traktowana jako osobna populacja, ponieważ ewolucja na niej zachodzi w izolacji od pozostałych wysp. Co pewien okres populacje znajdujące się na wyspach wymieniają się między sobą pewną ilością osobników w procesie migracji. Pozwala to przyśpieszyć znajdywanie rozwiązania. Istotna jest tutaj topologia połączeń pomiędzy wyspami pozwalająca na wymianę osobników, metoda selekcji migrujących osobników, ich wielkość oraz odstęp pomiędzy kolejnymi migracjami. Badania różnych parametrów tych zjawisk można znaleźć m.in. w pracy [61].

5.3. Problemy benchmarkowe

Napisac o trudnych problemach wielomodalnych i napisac ze rastrigina wykorzystalem jako swój benchmark.

Algorytmy genetyczne opisane w poprzednim podrozdziale zostały wykorzystane do znalezienia minimum funkcji Rastrigina.

Funkcja Rastrigina jest funkcją niewypukłą. Ze względu na to, że posiada wiele minimów lokalnych oraz jedno globalne dla $x = 0$, $f(x) = 0$ bywa trudna w optymalizacji, ponieważ algorytmy optymalizacyjne mogą utknąć w którymś z lokalnych minimów tracąc szansę na znalezienie minimum globalnego. Funkcja ta często znajduje zatem zastosowanie jako funkcja testująca algorytmy optymalizacyjne. Została ona również użyta jako problem testowy w niniejszej implementacji.

Wzór funkcji Rastrigina przedstawiono poniżej

$$f(x) = An + \sum_{i=1}^n [x_i^2 - Acos(2\pi x_i)], \quad (5.1)$$

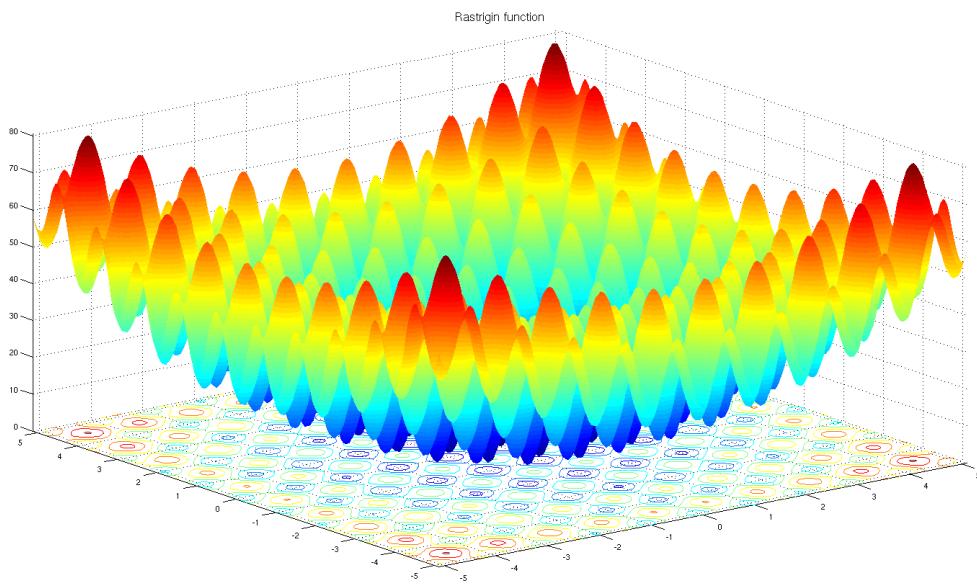
gdzie $A = 10$, $x_i \in [-5.12, 5.12]$ a n oznacza wymiar funkcji [68].

Wykres funkcji Rastrigina przedstawiono na rysunkach 5.2 oraz 5.3.

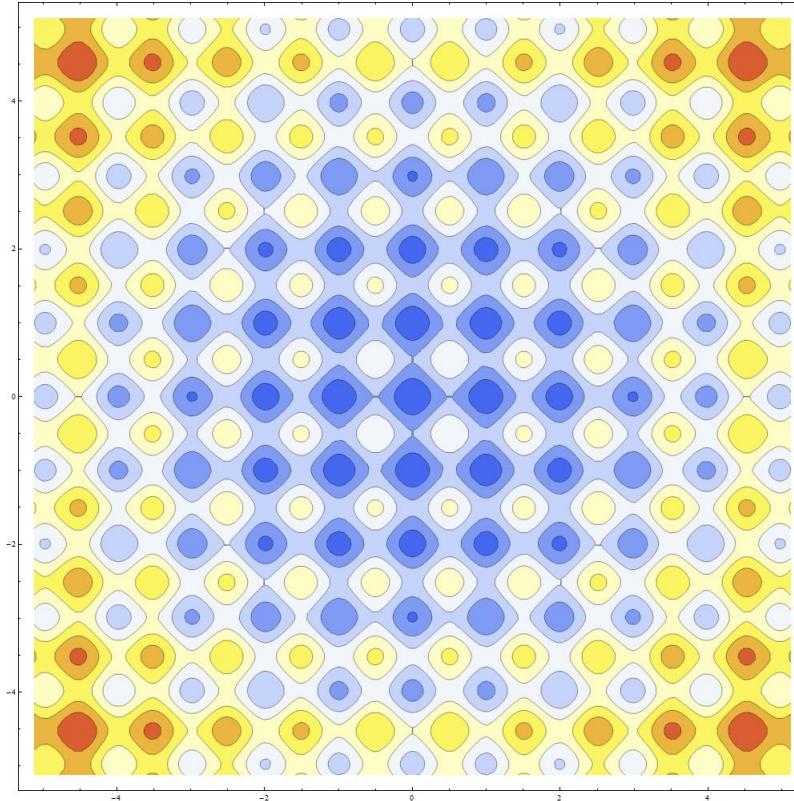
5.4. Ewolucyjna optymalizacja na platformie

....

Rozwiązanie stworzone na potrzeby niniejszej pracy przetestowano dzięki zaimplementowaniu algorytmów genetycznych w odmianie wyspowej opisanych w poprzednim rozdziale.



Rysunek 5.2: Wykres funkcji Rastrigina 3D [26].



Rysunek 5.3: Wykres funkcji Rastrigina 2D [26].

Tablica 5.1: Parametry zadania.

Nazwa parametru	Opis
Dimension	Wymiar optymalizowanej funkcji ciągłej.
Cycles	Maksymalna ilość cykli ewolucji.
Initial population size	Początkowy rozmiar populacji (podczas krzyżowania i mutacji liczba osobników populacji zostaje wzrasta).
Maximum population size	Maksymalny rozmiar populacji.
Snapshot frequency	Częstotliwość persystowania wyników obliczeń (w tym przypadku osobników populacji). *
Migration frequency	Częstotliwość występowania migracji. *
Mutation parameter	Parametr mutacji, określa wpływ mutacji na geny.
Migration factor	Procent migrującej populacji.

* częstotliwość jest określana przez ilość cykli

Implementacja ta, została oparta na implementacji algorytmów genetycznych przedstawionej w książce Scala for Machine Learning [73] oraz dostosowana do optymalizacji funkcji ciągłej.

W przypadku problemu Rastrigina gen jest liczbą typu Double, chromosomy zawierają listę genów czyli obrazują punkty w przestrzeni rozwiązań. W przypadku dwuwymiarowej funkcji Rastrigina jest to lista dwuelementowa. Natomiast populacja zawiera listę chromosomów.

Zadanie jest konfigurowane parametrami przedstawionymi w tabeli 5.1.

Migracja odbywa się poprzez wybranie pewnej ilości najlepszych osobników populacji oraz wysłanie ich do innego węzła klastra. Wykonuje się ona co określoną liczbę cykli konfigurowalną w parametrach zadania. Do wyboru są dwa rodzaje selekcji węzła do którego wysyłani są migrujący osobnicy. Pierwszy to jest wybór losowy a drugi metodą round-robin, czyli wysyłanie migrującej części populacji na zmianę do każdego kolejnego węzła [45]. Migracja odbywa się bez użycia mastera, do jej działania wykorzystano rozszerzenie framework'a Akka o nazwie *Distributed Publish Subscribe in Cluster* opisane w poprzednim rozdziale.

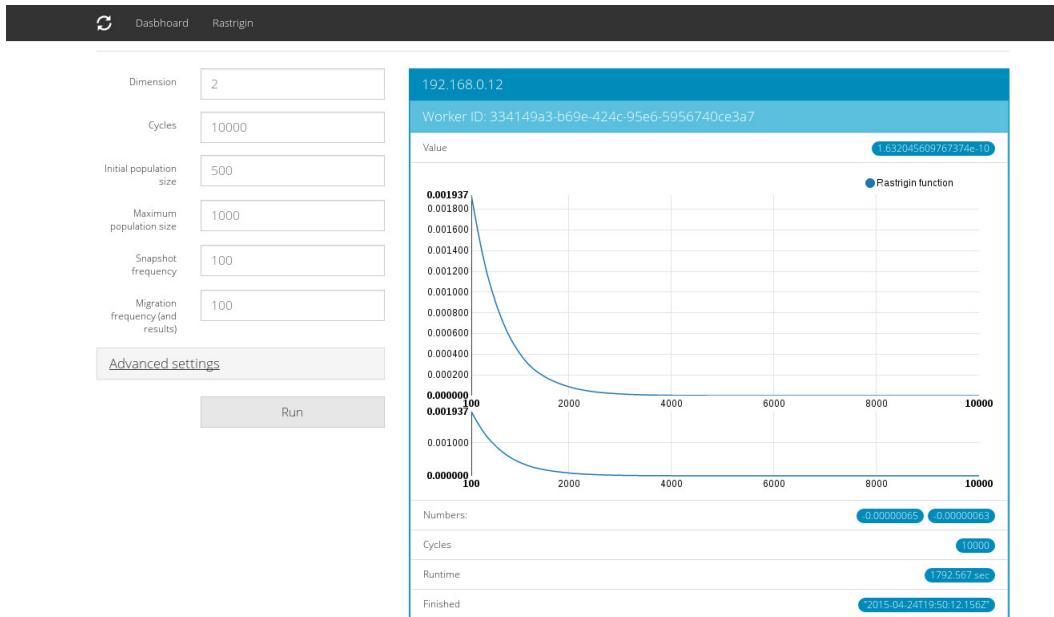
Inicjalizacja populacji odbywa się w każdym węźle klastra osobno. W przypadku problemu Rastrigina opisanego w kolejnym podrozdziale, początkowa populacja składa się z osobników zainicjalizowanych losowymi liczbami z przedziału $[-5.12, 5.12]$.

Na rysunku 5.4 zaprezentowano wygląd interfejsu użytkownika pozwalającego na podgląd wyników rozpoczętych zadań oraz rozpoczęcie nowych. Widoczne jest tutaj każde z urządzeń wykonujących obliczenia w klastrze i są one identyfikowane po adresie IP oraz ID workera.

Wyniki prezentowane są w czasie rzeczywistym przy użyciu technologii WebSocket [41]. Dla każdego zadania rysowany jest wykres przedstawiający wartość funkcji Rastrigina zmieniającą się w trakcie trwania ewolucji, czyli w kolejnych cyklach algorytmu. Wykres narysowano za pomocą biblioteki D3.js [15].

Monitoring

Na rysunku 5.5 przedstawiono interfejs klienta prezentujący obecny stan klastra. Wylistowano na nim również listę urządzeń pracujących w klastrze, z uwzględnieniem aktywnego mastera, adresu IP wraz z portami, dostępność urządzenia oraz kilka metryk, m.in. zużycie CPU oraz pamięci RAM.



Rysunek 5.4: Rozpoczęcie zadania.



Rysunek 5.5: Podgląd stanu klastra.

Tablica 5.2: Lista zdarzeń sterujących stanem zadania.

Nazwa stanu	Opis
WorkAccepted	Stan zadania po przyjęciu go przez mastera oraz wysłaniu potwierdzenia do klienta.
WorkStarted	Stan zadania przekazanego do realizacji do workera.
WorkInProgress	Stan zadania w trakcie trwania obliczeń, przechowujący częściowe wyniki zadania.
WorkCompleted	Stan zadania otrzymany po zakończeniu obliczeń, przechowujący końcowe wyniki.
WorkerFailed	Stan zadania zakońzonego niepowodzeniem.
WorkerTimedOut	Stan zadania ustawiany jeśli worker pracujący nad danym zadaniem nie odeśle wyników w określonym czasie.

Stan klastra propagowany jest za pomocą protokołu Gossip opisanego w poprzednim rozdziale. Jeżeli jakieś urządzenie nie odpowiada przez określony czas w wyniku np. problemów z siecią, jest ono uznawane jako niedostępne, czyli nie może już przyjmować żadnych zadań ale może jeszcze powrócić do pracy w klastrze jeżeli odzyska sprawność w ciągu najbliższych kilku sekund. Jeżeli nie, w takim przypadku jest ono usuwane z listy urządzeń pracujących w klastrze. Po usunięciu urządzenia z klastra nie może ono dołączyć ponownie do klastra dopóki aplikacja uruchomiona na tym urządzeniu nie zostanie zrestartowana. Pozwala to zapobiec sytuacji, gdy w trakcie problemów z siecią, dane urządzenie odłączy się na chwilę od klastra aktywując własnego mastera a następnie po powrocie do klastra będą aktywne dwa mastery. Dzieje się tak dlatego, że w przypadku gdy urządzenie utraci kontakt z pozostałymi urządzeniami, nie może ono z pewnością stwierdzić co było przyczyną awarii i zakłada, że jest jedynym urządzeniem w klastrze.

W trakcie awarii urządzenia podczas wykonywania obliczeń dzięki występowaniu migracji opisanej w jednym z kolejnych podrozdziałów nie tracimy wszystkich wyników pracy a jedynie te uzyskane od poprzedniej migracji. Stan mastera jest zapisywany w bazie danych, co pozwala na jego odtworzenie w przypadku awarii urządzenia z aktywnym masterem. Do bazy danych persystowane są zserializowane zdarzenia zgodnie ze wzorcem Event Sourcing opisany w pozycjach [24, 25]. Zdarzenia te odzwierciedlają stan przyjętego zadania. Listę zdarzeń oraz ich opis przedstawiono w tabeli 5.2.

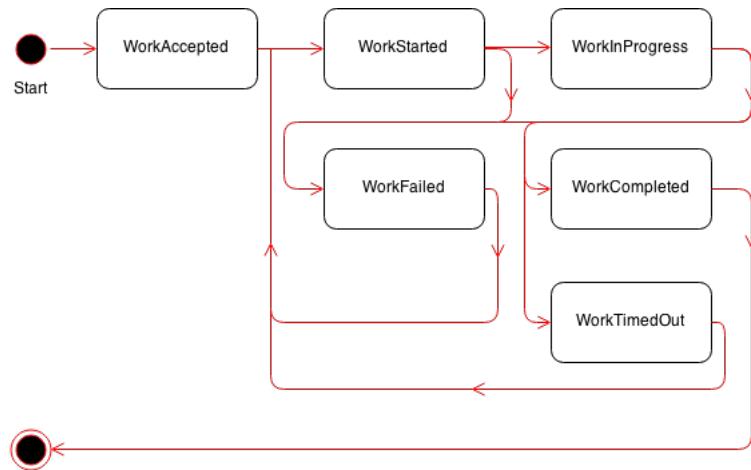
Dla przykładu, jeżeli klient zadał zadanie i zmienił ono stan na WorkAccepted (czyli zostało zaakceptowane ale nie przesiane jeszcze do żadnego workera) a w tym samym czasie nastąpiła jakaś awaria przerwująca pracę mastera, kolejna instancja mastera zostanie aktywowana na innym urządzeniu i odtworzy stan mastera, który uległ awarii, co oznacza, że klient nie będzie musiał ponownie rozpoczynać zadania, lecz jego poprzednio rozpoczęte zadanie zostanie przekazane do realizacji do wolnego workera.

Diagram prezentujący dopuszczalne zmiany stanów zadania przedstawiono na rysunku 5.6.

Dystrybucja

Do łatwej dystrybucji platformy między różnymi urządzeniami wykorzystano dodatek projektu Akka o nazwie Microkernel oraz rozszerzenie Sbt o nazwie sbt-native-packager.

Akka Microkernel oraz sbt-native-packager dostarczają mechanizmu archiwizowania dzięki któremu można udostępniać aplikację jako pojedyńczy plik bez potrzeby uruchamiania lub instalowania dodatkowych aplikacji, dostarczania zależności lub ręcznego tworzenia skryptów startowych. Pozwalają



Rysunek 5.6: Przejścia pomiędzy stanami zadania.

również na wystartowanie systemu aktorowego używając klasy z metodą statyczną main [5, 46].

Umożliwia to łatwą instalację oraz uruchomienie platformy na różnych urządzeniach klastra. Do uruchomienia platformy na danym urządzeniu wystarczy aby posiadało ono zainstalowany system wraz z wirtualną maszyną Javy oraz archiwum z platformą, które po rozpakowaniu pozwala uruchomić aplikację za pomocą jednego skryptu wykonywalnego. Żadne dodatkowe oprogramowanie nie jest wymagane.

5.5. Wyniki testów

Do przeprowadzenia testów zostało wykorzystanych kilka urządzeń Raspberry Pi. Jako problem testowy wykorzystano optymalizację wielowymiarowej funkcji Rastrigina opisanej w poprzednim podrozdziale.

Ze względu na ograniczoną precyzję obliczeń nie udało się uzyskać wyniku równego poszukiwanemu minimum, czyli $f(x) = 0$.

Najlepsze uzyskane wyniki oscylowały w okolicach $2e - 30$.

Do obliczeń wykorzystano typ Double, który w języku Scala cechuje się podwójną precyzją, zapisaną na 64 bitach [71].

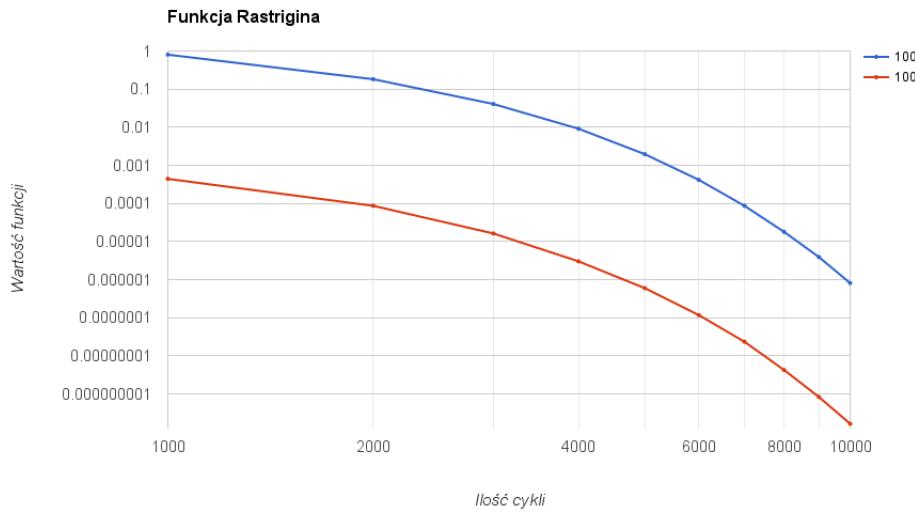
Na poniższych wykresach przedstawiono wyniki oraz czasy ich uzyskania w zależności od różnych ustawień początkowych zadania.

Rozważono tutaj problem dla jednego urządzenia w celu porównania go z bardziej istotnym przypadkiem dla celów niniejszej pracy a mianowicie wielu urządzeń pracujących w klastrze, który rozważono w następnym podrozdziale poświeconeemu skalowalności rozwiązania.

Na rysunku 5.7 przedstawiono wykres wartości funkcji Rastrigina 2D gdzie współczynnik mutacji wynosi 0.2 a całkowita ilość cykli 10000. Aby zwiększyć czytelność wykresu użyto skali logarytmicznej.

Na wykresie uwzględniono również różne maksymalne wielkości populacji, co zaznaczono innym kolorem.

Wykres na rysunku 5.7 pokazuje, że w początkowych cyklach ewolucji populacji gdy osobnicy nie są jeszcze dobrze rozwinięci kolejne rozwiązania znajdywane są wolniej. Szybkość znajdywania nowych rozwiązań przyśpiesza wraz ze wzrostem ilości cykli oraz wzrostem wielkości populacji.



Rysunek 5.7: Całkowity czas wykonania ... Znaleziony punkt ..., wartość funkcji ...

Rysunek 5.8: Całkowity czas wykonania ... Znaleziony punkt ..., wartość funkcji ...

Na rysunku ... przedstawiono wykres wartości funkcji dla większego współczynnika mutacji wynoszącego 0.4.

TODO sprobować zmieścić rozne współczynniki mutacji na jednym wykresie, z roznymi odcieniami danego koloru a drugi wykres zrobic dla 8 wymiarow.

Powtorzyc dla 0.4.

Powtorzyc dla 0.8.

Podsumowując, zauważono, że większy współczynnik mutacji przyśpiesza znajdywanie nowych rozwiązań.

Warto dodać, że podczas przeprowadzania testów w klastrze pracowały dwa typy urządzeń Raspberry Pi, model B oraz B plus, który jest nowszy od modelu B oraz posiada większą ilość zaistalowanej pamięci RAM [23].

Dostępna w urządzeniach pamięć RAM ma dość duży wpływ na czas obliczeń. Czas uzyskiwany przez nowszy model urządzenia, który posiada więcej pamięci był około 30% (**sprawdzić ile dokładnie**) lepszy.

Skalowalność

Kolejnymi dwoma parametrami konfiguруjącymi zadanie są współczynnik migracji oraz częstotliwość jej występowania.

Mają one bezpośredni wpływ na skalowalność platformy. Współczynnik migracji podawany jest w zakresie 0..100 i oznacza procent migrującej populacji. Natomiast parametr określający częstotliwość występowania migracji wyznacza co ile cykli następuje migracja. Jego wartość może wynosić zero, wtedy migracja nie występuje lub jakiś ułamek maksymalnej liczby cykli.

Bez występowania migracji osobników pomiędzy wyspami (węzłami klastra) mogłyby zaistnieć sytuacja, gdy pomimo pracy kilku urządzeń w klastrze każde z nich uzyskałoby podobne wyniki w zbliżonym czasie, co oznaczałoby, że równie dobrze obliczenia mogłyby być przeprowadzane tylko na jednym urządzeniu, bo wzrost liczby urządzeń nie wpłynął znaczco na szybkość znajdywanych wyników.

Dzięki zastosowaniu migracji, gdy do klastra zostaje włączone nowe urządzenie i zaczyna wykonywać obliczenia, to po odebraniu osobników migrujących z innej wyspy, którzy z dużym prawdopodobieństwem będą zawierali lepszej jakości rozwiązania aniżeli nowo zainicjalizowani osobnicy na aktualnej wyspie ewolucja takiej wyspy zostaje przyśpieszona. Przy wysokiej częstotliwości migracji oraz dużym procencie migrującej populacji istnieje ryzyko, że wyspy zostaną zdominowane przez osobników o podobnych cechach, co może prowadzić do odnalezienia jedynie lokalnego minimum zamiast minimum globalnego.

Tutaj chciałbym pokazać jaki jest wpływ migracji oraz jej parametrów na wyniki obliczeń porównując wyniki z poprzedniego podrozdziału uzyskane dla jednego urządzenia..

Wykres ... Na wykresie przedstawiono innym kolorem różną częstotliwość migracji dla 10% migrującej populacji.

4 urządzenia, 2 wymiary.

4 urządzenia 8 wymiarów.

Powtarzyc dla 20% migrującej populacji.

40%.

Wydajność

Średnie zużycie pamięci RAM oraz obciążenie procesora w trakcie wykonywania obliczeń (w zależności od różnych ustawień początkowych) oraz w stanie bezczynności.

Narzut komunikacji.

Częstość wykonywania zapisów do bazy, przesyłania wyników częściowych do klienta oraz wpływ wielkości populacji.

Pozostałym parametrem konfigurującym zadanie jest częstość przesyłania wyników częściowych do klienta co ma wpływ również na częstość wykonywania zapisów do bazy.

6. Podsumowanie

W poniższym rozdziale zostało zawarte podsumowanie niniejszej pracy. Opisano również wnioski oraz nakreślono możliwości dalszego rozwoju.

6.1. Wnioski

W ramach niniejszej pracy udało się spełnić początkowe założenia i przygotować platformę oferującą możliwość uruchomienia obliczeń równoległych na wielu urządzeniach pracujących w klastrze, dostarczającą narzędzi do monitoringu, rozwiązującą problem load-balancingu zadań, skalowalności a także zapewniającą wsparcie w sytuacji awarii tychże urządzeń poprzez reorganizację oraz odzyskiwanie sprawności czyli powrót do stanu klastra sprzed wystąpienia awarii. Platforma pozwala również na prowadzenie efektywnych obliczeń mimo ograniczonej mocy obliczeniowej urządzeń SoC. Wykorzystanie tych urządzeń oraz technologii działających na maszynie wirtualnej Javy zapewnia mobilność tego rozwiązania.

Technologia Akka okazała się kluczowa z punktu widzenia niniejszej pracy idealnie wkomponowując się w wybraną tematykę. Dzięki rozwiązaniu wysokiego poziomu ułatwiały one tworzenie nowych komponentów oraz dostarczyły pewnych strategii oraz wzorców zwiększając tym samym przejrzystość jak i możliwości dalszego rozwoju przygotowanego rozwiązania.

Niniejsza praca pozwoliła również autorowi na zapoznanie się z tematyką problemów występujących w klastrach oraz rozwiązaniami stosowanymi w celu zrównoleglenia lub rozproszenia obliczeń. Kolejnym ciekawym aspektem niniejszej pracy, który pozwolił na eksplorację nowych obszarów wiedzy z punktu widzenia autora jest względnie nowa dziedzina w programowaniu a mianowicie programowanie reaktywne i technologie z platformy Typesafe w połączeniu z urządzeniami SoC oraz Cluster Computing.

6.2. Możliwości rozwoju

Rozwiązanie stworzone na potrzeby niniejszej pracy posiada kilka ciekawych aspektów nad którymi badania mogą być kontynuowane. Kilka z nich zostało opisanych poniżej.

Wiele aplikacji klienckich

W testowanym rozwiążaniu rozważono tylko jedną aplikację kliencką, będącą jedynym źródłem zadań w klastrze. W kolejnych etapach prac nad stworzoną platformą warto byłoby się pokusić o zaimplementowanie przypadku gdzie klientów jest więcej i w różnym stopniu obciążają oni urządzenia

pracujące w klastrze a także nad sposobem dysponowania wolnymi węzłami tak, aby jeden klient nie zajął wszystkich dostępnych węzłów podczas gdy reszta nie byłaby w żaden sposób obsługiwana.

Sieć rozległa

Na potrzeby niniejszej pracy rozważono działanie urządzeń klastra w sieci lokalnej lecz przygotowane rozwiążanie mogłoby się również sprawdzić w sieci globalnej Internet, gdzie węzły klastra mogłyby być uruchomione na różnego typu urządzeniach znaczaco oddalonych od siebie i mających do siebie dostęp za pomocą protokołu TCP/IP. Ciekawym aspektem tutaj byłaby zróżnicowana przepustowość połączeń, ponieważ w takim przypadku trzeba byłoby dobrać odpowiednie ustawienia detekcji awarii i arbitralne wybranie dozwolonego czasu opóźnień mogłoby być niełatwym zadaniem.

Wykorzystanie GPU

Do obliczeń wykonywanych w obrębie platformy wykorzystano jedynie procesory CPU lecz większość urządzeń SoC posiada również jednostki GPU, które są w tym czasie niewykorzystywane. Warto byłoby się zastanowić nad potencjalnym wykorzystaniem tych jednostek np. przy użyciu takich bibliotek jak ScalaCL (opartej o OpenCL) [47], które pozwalają na wykonywanie kodu napisanego w Scala na procesorach graficznych i równoległy prowadzeniu obliczeń na różnych architekturach. W roku 2011 Martin Odersky współtwórca języka Scala oraz jego zespół pracujący na uniwersytecie w Lozannie uzyskali wielomilionowy grant od Europejskiej Rady ds. Badań Naukowych (ERBN) dzięki któremu mogli kontynuować pracę nad problemami zrównoleglania obliczeń w różnych modelach programowania jak np. OpenMP, MPI, CUDA, OpenCL [58] co potwierdza, że prace nad podobnymi rozwiązaniami są prowadzone już od wielu lat.

Inne algorytmy

Kolejną kwestią którą można rozważyć w planach dalszego rozwoju jest zaimplementowanie innych algorytmów, które będą w stanie równie dobrze wykorzystać potencjał stworzonej platformy. Warto również poeksperymentować z różnymi modyfikacjami algorytmów genetycznych, takimi jak różne metody selekcji lub krzyżowania, aby uzyskać jeszcze większą efektywność obliczeń.

Optymalizacja

Urządzenia SoC na których testowano przygotowaną platformę posiadają ograniczone możliwości obliczeniowe, co oznacza, że uruchamiane na nich aplikacje powinny być zoptymalizowane pod tym kątem. Wykorzystanie pamięci RAM oraz procesora jest tutaj bardzo istotne i może w znaczny sposób wpływać na szybkość obliczeń. Nie można zatem pozwolić sobie na zbędne operacje i nadmierną zajętość pamięci, co może wydłużyć wykonywanie obliczeń. Warto byłoby użyć jakiegoś narzędzia do profilowania JVM, aby sprawdzić czy nie ma żadnych wycieków pamięci, a także przenalizować zrzut pamięci (tzw. heap dump) pod kątem instancjonowanych obiektów oraz zajętości pamięci. Można byłoby również wykonać tuning GC (ang. Garbage Collector) oraz spróbować dobrać jego optymalne parametry.

IoT

W ostatnich latach coraz bardziej wzrasta znaczenie urządzeń podłączonych do Internetu. Coraz więcej sprzętu codziennego użytku posiada zainstalowane układy procesorowe o wystarczającej mocy obliczeniowej, aby móc obsłużyć systemy mobilne takie jak Android. Stwarza to zatem pole do popisu oraz nowe pomysły wykorzystania stworzonego rozwiązania w wersji bardziej rozproszonej.

Bibliografia

- [1] Actor systems.
- [2] Akka actor model.
- [3] Akka cluster specification.
- [4] Akka cluster usage.
- [5] Akka microkernel.
- [6] Akka persistence.
- [7] Akka persistence mongo.
- [8] Akka remoting.
- [9] Akka stm.
- [10] Algorytm genetyczny.
- [11] Apache hadoop.
- [12] Arduino.
- [13] Broadband by the numbers.
- [14] Cisco - ioe.
- [15] D3.js.
- [16] Distributed publish subscribe in cluster.
- [17] Dokumentacja angularjs.
- [18] Dokumentacja bazy mongodb.
- [19] Dokumentacja frameworka akka dla języka scala.
- [20] Dokumentacja języka scala.
- [21] Dokumentacja play framework dla języka scala.
- [22] Dokumentacja projektu akka cluster singleton.

- [23] Dokumentacja raspberry pi.
- [24] Event sourcing.
- [25] Event sourcing pattern.
- [26] Funckja rastrigina.
- [27] Htcondor.
- [28] Intel galileo.
- [29] Intel galileo idzie na studia.
- [30] Iot.
- [31] Java concurrency framework.
- [32] Lockstep.
- [33] Master/slave.
- [34] Mpi.
- [35] Nvidia cuda.
- [36] Opencl.
- [37] Openmp.
- [38] Parallela board.
- [39] Play akka cluster sample.
- [40] Protokół gossip.
- [41] Protokół websocket - rfc6455.
- [42] Raspberry pi 2.
- [43] Raspberry pi at southampton.
- [44] Reactive manifesto.
- [45] Round-robin.
- [46] Sbt native packager plugin.
- [47] Scalacl.
- [48] Sharks cove.
- [49] Sigar.
- [50] Soc.

- [51] Taksonomia flynna.
- [52] Template akka distributed workers.
- [53] Type safety.
- [54] Work pulling pattern.
- [55] T. Alexandre. *Scala for Java Developers*. Packt Publishing, 2014.
- [56] J. Armstrong. *Programming Erlang: Software for a Concurrent World (Pragmatic Programmers)*. The Pragmatic Bookshelf, 2013.
- [57] D. M. Bruce Eckel. *Atomic Scala*. Mindview LLC, Crested Butte, CO, 2013.
- [58] Z. M. A. R. T. S. A. H. P. O. M. O. K. Chafi, Hassan; DeVito. Language virtualization for heterogeneous parallel computing, 2010.
- [59] S. J. Cox. Iridis-pi: a low-cost, compact demonstration cluster. *Cluster Computing*, Czerwiec 2014.
- [60] F. daCosta. *Rethinking the Internet of Things: A Scalable Approach to Connecting Everything*. Apress Open, 2014.
- [61] M. Dudek. Badanie efektywności wielopopulacyjnego algorytmu ewolucyjnego. Akademia Górnictwo-Hutnicza w Krakowie, 2009.
- [62] I. Foster. *Designing and Building Parallel Programs: Concepts and Tools for Parallel Software Engineering*. Addison-Wesley, 1995.
- [63] A. Freeman. *Pro AngularJS*. Apress, 2014.
- [64] S. Furber. *ARM System-on-Chip Architecture*. Pearson Education, 2000.
- [65] F. Gebali. *Algorithms and Parallel Computing*. A John Wiley & Sons, Inc., Publication, 2011.
- [66] D. E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Longman Publishing Co., Styczeń 1989.
- [67] M. K. Gupta. *Akka Essentials*. Packt Publishing, 2012.
- [68] D. S. H. MÄžhlenbein and J. Born. The parallel genetic algorithm as function optimizer. *Parallel Computing*, 17:619–632, 1991.
- [69] D. A. P. John L. Hennessy. *Computer Architecture A Quantitative Approach*. Morgan Kaufmann, 2012.
- [70] C. B. M. Z. Marc Shapiro, Nuno PreguiÃ, ca. A comprehensive study of convergent and commutative replicated data types. *INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE*, 2011.
- [71] L. S. Martin Odersky and B. Venners. *Programming in Scala, Second Edition*. Artima, Grudzień 2010.

- [72] N. S. Maurice Herlihy. *The Art of Multiprocessor Programming*. Morgan Kaufmann, 2008.
- [73] P. R. Nicolas. *Scala for Machine Learning*. Packt Publishing, Grudzień 2014.
- [74] P. S. Pacheco. *An Introduction to Parallel Programming*. Morgan Kaufmann, 2011.
- [75] J. Richard-Foy. *Play Framework Essentials*. Packt Publishing, 2014.
- [76] D. WHITLEY. A genetic algorithm tutorial. *Statistics and Computing*, 4:65–85, 1994.
- [77] Wikipedia. Rss.