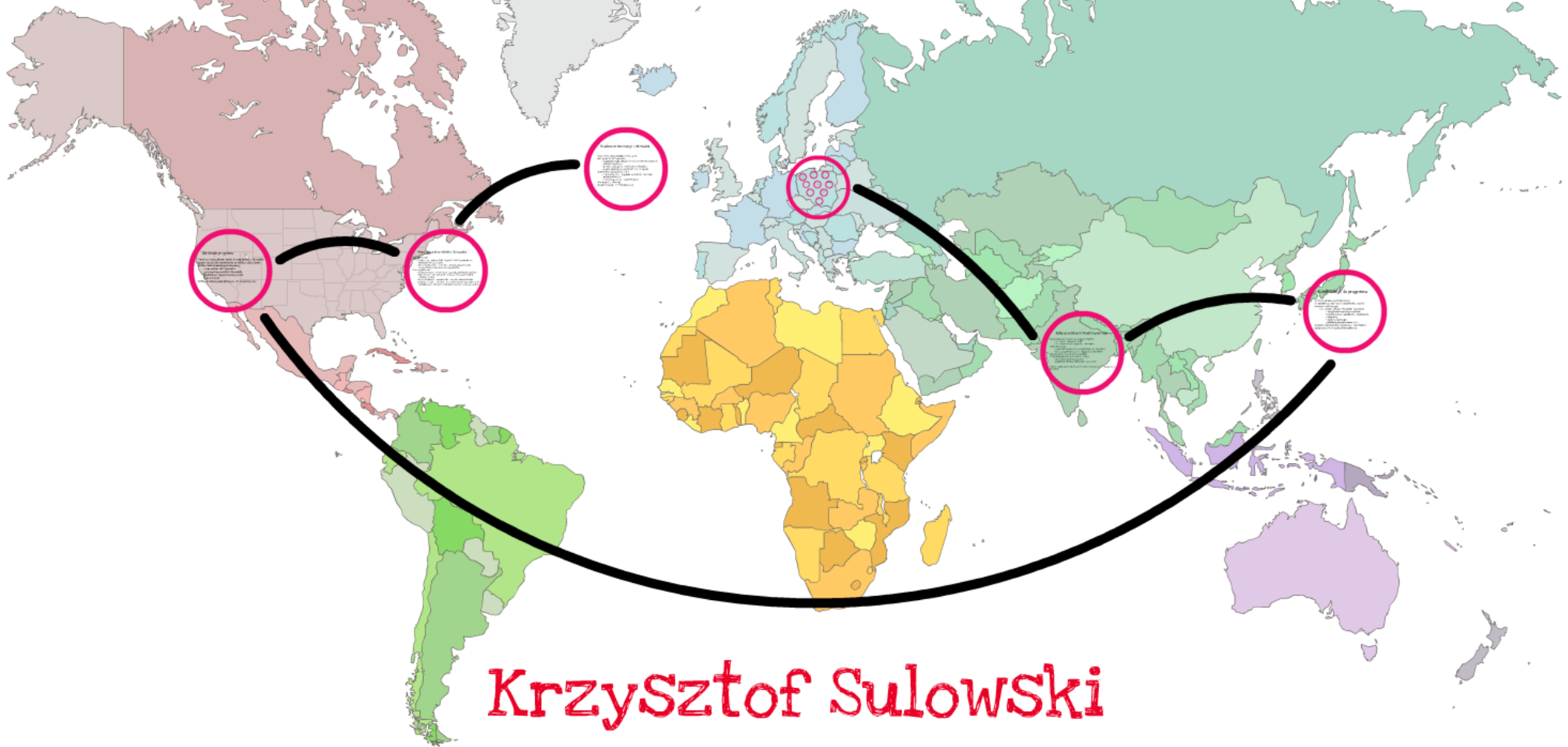


Program tworzenia prac magisterskich przy współpracy z firmą 9livesdata



Krzysztof Sulowski
9livesdata

Program tworzenia prac magisterskich przy współpracy z firmą 9livesdata



Krzysztof Sulowski
9livesdata

Podstawowe informacje o 9livesdata

Polska firma R&D z siedzibą na Mokotowie

R&D produktu NEC Hydrastor

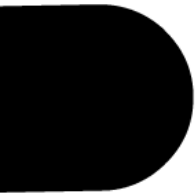
- najszybszy/najbardziej skalowalny na świecie system do deduplikacji danych
- powstał wg projektu konsultantów 9livesdata
- System sprzedawany przez NEC w USA i Japonii

Liczne publikacje i patenty w USA

- konferencja FAST - najlepsza na świecie w zakresie pamięci masowych
- konferencja SYSTOR - R&D IBM, Izrael

Nowe projekty - startupy

Węcej informacji - www.9livesdata.com



Prace studentów MIMUW w 9liveSdata

Prace doktorskie:

- Leszek Gryz - listopad 2012 - Impact of Data Organization on Distributed Storage System
- Michał Kaczmarczyk - w trakcie - Reducing impact of data fragmentation caused by in-line deduplication

Prace magisterskie:

- Maciej Kowalczyk - marzec 2009 - Upgrading distributed systems
- Piotr Skowron - czerwiec 2009 - Resource Management Unit in HYDRAStor System
- Kornel Jakubczyk - wrzesień 2010 - Marauder disks detection
- Dominik Borowiec - w trakcie - Correlation-based compression of the statistics generated by the distributed, secondary storage system

Założenia programu

Praca mgr na podstawie badań we współpracy z 9livesdata

Wynagrodzenie dla magistranta za każdą godzinę badań

Możliwe zastosowania wyników pracy

- w produkcie NEC Hydrastor
- w nowych produktach 9livesdata
- publikacje z wyników pracy w USA
- patent w USA

Możliwość dalszej współpracy po obronie pracy mgr

Kwalifikacja do programu

Od IV roku studiów informatycznych

We współpracy z opiekunem magistranta z uczelni

Rozmowa kwalifikacyjna

- ok. 4 godzin w biurze 9livesdata, wymagania:
 - umiejętność logicznego myślenia
 - programowanie współbieżne i rozproszone
 - algorytmy
 - systemy operacyjne
 - podstawy programowania w C++

Wymagane zaangażowanie na poziomie \geq 24h/tydzień

Zgłoszenia z CV: recruiting@9livesdata.com

Katalog możliwych tematów prac mgr

10 potencjalnych tematów na kolejnych slajdach

- 7 w ramach współpracy z NEC
- 3 w ramach nowych projektów - startupów

Poufność tematów

- rozpowszechnianie tylko do potencjalnych kandydatów
- do wykorzystania tylko we współpracy z 9livesdata

Zaprezentowane tematy to tylko podzbiór

Prezentacja dodatkowych tematów wymaga

- zakwalifikowania do programu
- podpisania NDA (non-disclosure agreement)

Możliwe rozszerzenie niektórych tematów do pracy dr we współpracy z 9livesdata

Temat 1: Systemy pilnowania do profesjonalizacji i transkrypcyjności

Wspieranie systemów pilnowania jakości w celu profesjonalizacji i transkrypcyjności

- systemy do monitorowania jakości
- systemy do monitorowania jakości
- systemy do monitorowania jakości

Temat 2: Kwalifikacje i kompetencje

Definicja i znaczenie kwalifikacji i kompetencji

- kwalifikacje i kompetencje
- kwalifikacje i kompetencje
- kwalifikacje i kompetencje

Temat 3: Ścieżki kariery i rozwoju

Definicja i znaczenie ścieżki kariery i rozwoju

- ścieżki kariery i rozwoju
- ścieżki kariery i rozwoju
- ścieżki kariery i rozwoju

Temat 4: Wzrost i rozwój człowieka

Definicja i znaczenie wzrostu i rozwoju człowieka

- wzrost i rozwój człowieka
- wzrost i rozwój człowieka
- wzrost i rozwój człowieka

Informacje z CV

rekrutacja@prezi.pl
www.prezi.pl

Temat 5: Kompetencje i umiejętności

Definicja i znaczenie kompetencji i umiejętności

- kompetencje i umiejętności
- kompetencje i umiejętności
- kompetencje i umiejętności

Temat 6: Wykorzystanie narzędzi i technologii

Definicja i znaczenie wykorzystania narzędzi i technologii

- wykorzystanie narzędzi i technologii
- wykorzystanie narzędzi i technologii
- wykorzystanie narzędzi i technologii

Temat 7: Kompetencje i umiejętności

Definicja i znaczenie kompetencji i umiejętności

- kompetencje i umiejętności
- kompetencje i umiejętności
- kompetencje i umiejętności

Temat 8: Wykorzystanie narzędzi i technologii

Definicja i znaczenie wykorzystania narzędzi i technologii

- wykorzystanie narzędzi i technologii
- wykorzystanie narzędzi i technologii
- wykorzystanie narzędzi i technologii

Temat 1: System plików do przetwarzania strumieniowego

Istniejące systemy plików mają istotne wady dla przetwarzania strumieniowego

- zbyt duża fragmentacja
- zbyt mała przepustowość
- kosztowne operacje synchroniczne

Zadania tematu

- stworzenie nowego systemu plików lub warstwy przechowywania danych na Linuksie
- ewaluacja tego systemu w systemie komercyjnym

Temat 2: koordynacja zarządzania zasobami rozproszonymi

Obecne zarządzanie zasobami ma pewne ograniczenia

- brak koordynacji między węzłami
- nie są brane pod uwagę aplikacje zewnętrzne
- mała elastyczność przy zmianie konfiguracji

Zadania tematu

- ulepszenia zarządzania zasobami
- ewaluacja ulepszonego zarządzania w systemie komercyjnym

Temat 3: skalowalna odbudowa danych po awarii

Obecny sposób odbudowy danych ma wąskie gardło

- po Szybkiej wymianie węzła, cała odbudowa przebiega na tym węźle

Zadania tematu

- usunięcie wąskiego gardła przy odbudowie danych
- ewaluacja nowego sposobu odbudowy w systemie komercyjnym

Temat 4: deduplikacja i kompresja metadanych

Systemy do deduplikacji kopii zapasowych efektywnie deduplikują dane użytkownika, ale nie metadane
Po pewnym czasie, większość nowych zapisów to metadane

- deduplikacja i kompresja metadanych pozwoli na dalszą poprawę oszczędności przestrzeni dyskowej

Zadania tematu

- opracowanie metod deduplikacji i kompresji metadanych
- ewaluacja tych metod na podstawie zebranych śladów zapisów danych oraz w systemie komercyjnym

Temat 5: kompresja statystyk na podstawie ich korelacji

Statystyki systemu rozproszonego są podstawą do analizy jego zachowania i obecnego stanu

Niestety, statystyki zajmują bardzo dużo miejsca
wiele statystyk jest silnie ze sobą skorelowanych

Zadania tematu

- opracowanie metod wykrywania korelacji statystyk a następnie ich kompresji na tej podstawie
- ewaluacja tych metod na podstawie zebranych śladów zapisów danych oraz w systemie komercyjnym

W TRAKCIE REALIZACJI

Temat 6: usuwanie przyczyny błędów przed ich wystąpieniem

Po wykryciu błędu w czasie testów lub u jednego klienta należałoby uaktualnić kod u pozostałych klientów. Uaktualnienie kodu może mieć miejsce tylko okresowo (np. raz na 6 miesięcy), chyba, że prawdopodobieństwo wystąpienia błędu u danego klienta jest wysokie.

Zadania tematu

- opracowanie automatycznych metod oceny prawdopodobieństwa wystąpienia danego błędu u klienta na podstawie analizy statystyk i raportów systemu
- implementacja detektorów sytuacji prowadzących do konkretnych błędów
- ewaluacja w systemie komercyjnym

Temat 7: automatyczna optymalizacja dyrektyw `#include`

Czas budowy dużego programu mocno zależy od struktury dyrektyw `#include`

Zadania tematu

- opracowanie automatycznych metod upraszczania struktur dyrektyw `#include` w celu zmniejszenia czasu budowy programu
- ewaluacja tych metod przy budowie programu komercyjnego w rozproszonym systemie budowania
- możliwość rozszerzenia istniejących narzędzi np. `cppclean` Googla

Tematy w ramach startupów

Modyfikacja systemu plików na Linuksie

Narzędzia do poprawy jakości programowanie współbieżnego

Video peer-to-peer

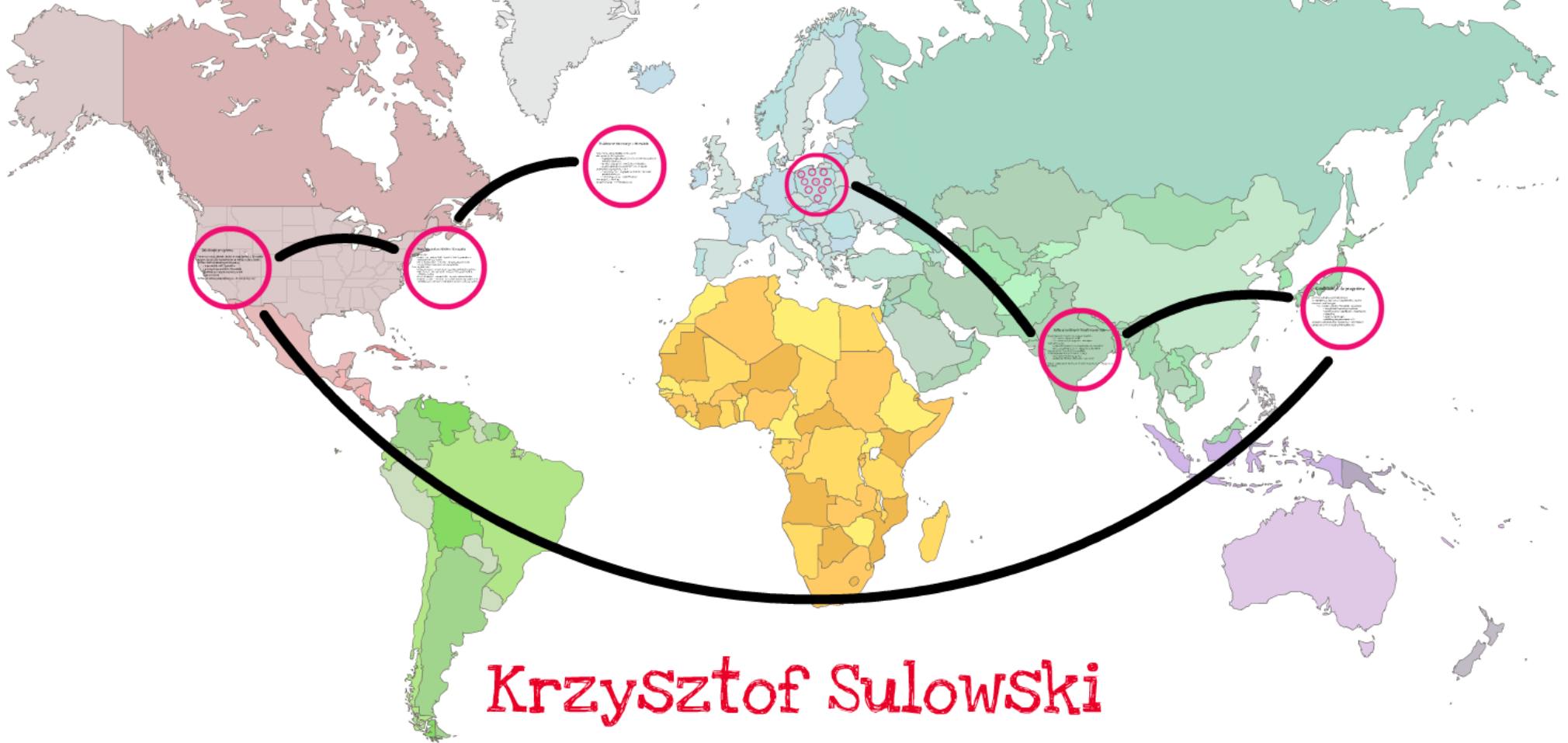
Zgłoszenia z CV:

recruiting@9livesdata.com

lub

ksulowski@9livesdata.com

Program tworzenia prac magisterskich przy współpracy z firmą 9livesdata



Krzysztof Sulowski
9livesdata