

Algorithm Configuration in the Cloud: A Feasibility Study

Daniel Geschwender Frank Hutter **Lars Kotthoff**
Yuri Malitsky Holger H. Hoos Kevin Leyton-Brown

18/02/2014

Supported by an Amazon Web Services research grant, European Union FP7 grant 284715 (ICON), Compute Canada and an Emmy Noether Grant by the DFG.

Motivation

- ▷ running algorithms on large problems requires substantial resources
- ▷ algorithm configuration often requires substantial resources
- ▷ public clouds available at relatively small cost
- ▷ large number of different machine configurations in clouds
- ▷ many cloud providers offer middleware to facilitate setup for common tasks

Motivation

“The use of cloud computing is growing, and by 2016 this growth will increase to become the bulk of new IT spend, according to Gartner, Inc. 2016 will be a defining year for cloud as private cloud begins to give way to hybrid cloud, and nearly half of large enterprises will have hybrid cloud deployments by the end of 2017.”

<http://www.gartner.com/newsroom/id/2613015>

Questions

- ▷ virtualised hardware – runtime measurements reliable?
[Kotthoff 2013], [Lampe et al. 2013], [Schad, Dittrich, Quiané-Ruiz 2010]
- ▷ different hardware – configurations on one infrastructure applicable on others?

Infrastructure setup

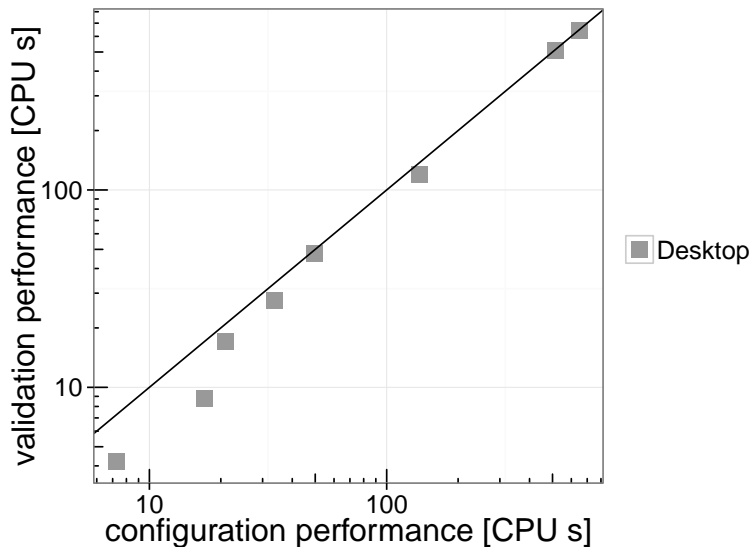
- ▷ Desktop machine (4 cores, 6 GB RAM)
- ▷ UBC cluster (8 cores, 16 GB RAM)
- ▷ UCC cluster (8 cores, 12 GB RAM)
- ▷ Amazon EC2 cloud with instance types
 - ▷ c1.xlarge (8 cores, 7 GB RAM)
 - ▷ m1.medium (1 core, 3.5 GB RAM)
 - ▷ m3.2xlarge (8 cores, 30 GB RAM)
- ▷ Microsoft Azure cloud with medium instance type (2 cores, 3.5 GB RAM)

Configuration setup

- ▷ configurators:
 - ▷ ParamILS [Hutter et al. 2009]
 - ▷ SMAC [Hutter, Hoos, Leyton-Brown 2011]
- ▷ scenarios:
 - ▷ AutoWeka-* [Thornton et al. 2013]
 - ▷ SPEAR-SWV [Hutter et al. 2007]
- ▷ setup:
 - ▷ 8 independent runs with different seeds
 - ▷ 1 CPU day to find best configuration
 - ▷ validate each found configuration on all infrastructures

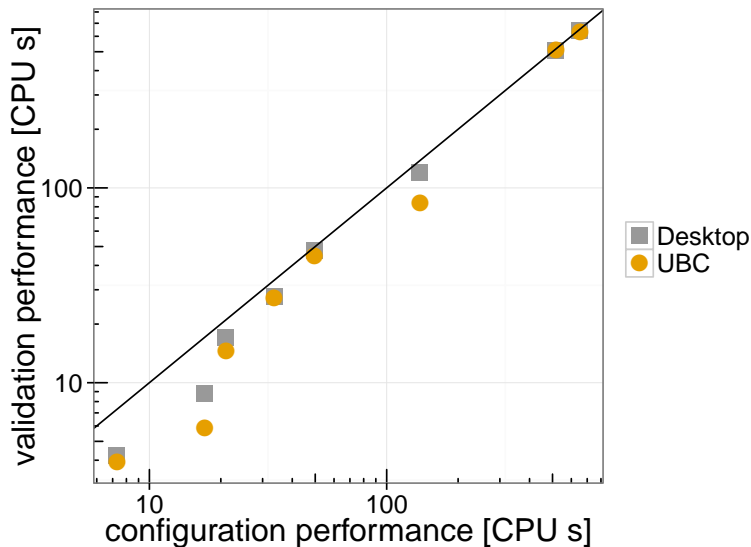
Results

ParamILS for Spear-SWV configured on UCC



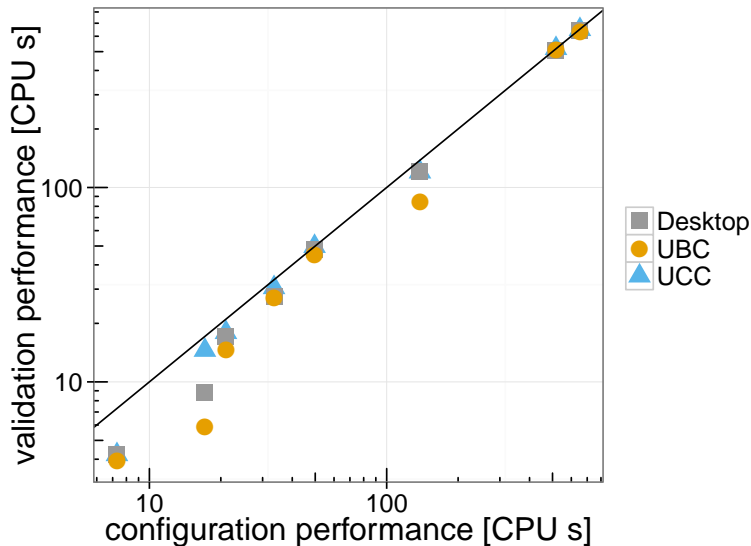
Results

ParamILS for Spear-SWV configured on UCC



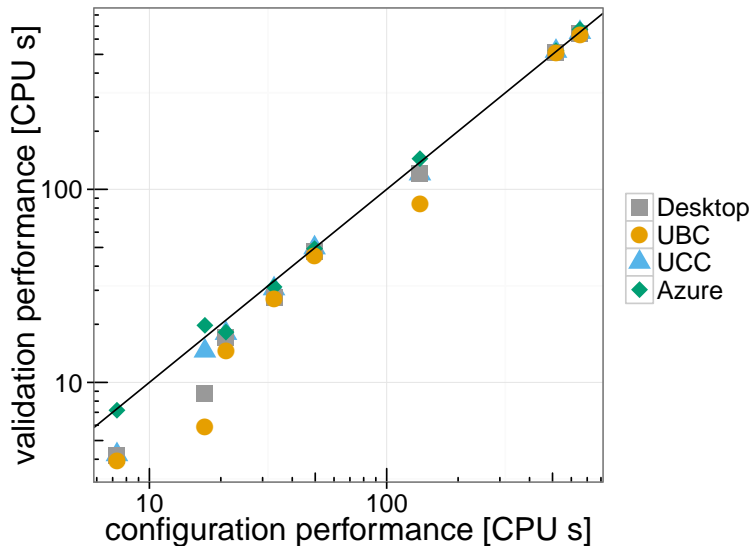
Results

ParamILS for Spear-SWV configured on UCC



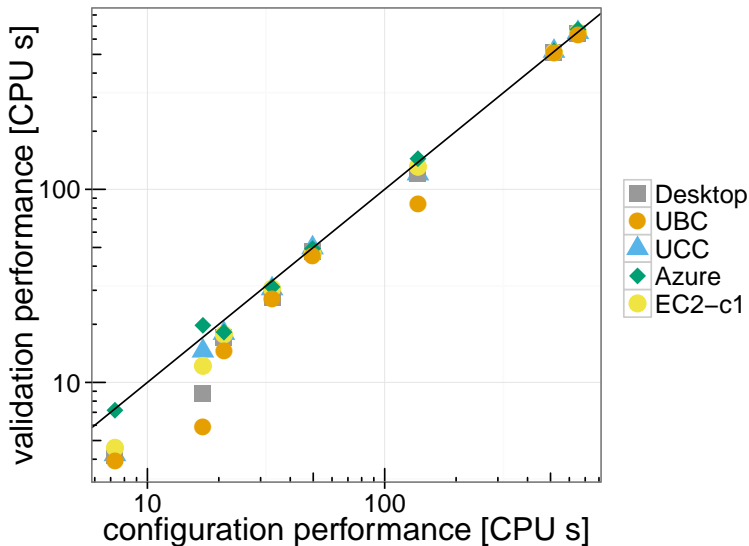
Results

ParamILS for Spear-SWV configured on UCC



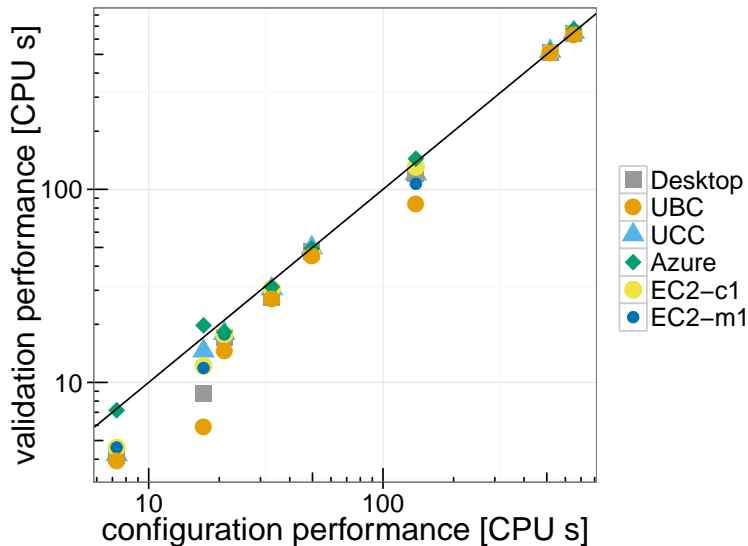
Results

ParamILS for Spear-SWV configured on UCC



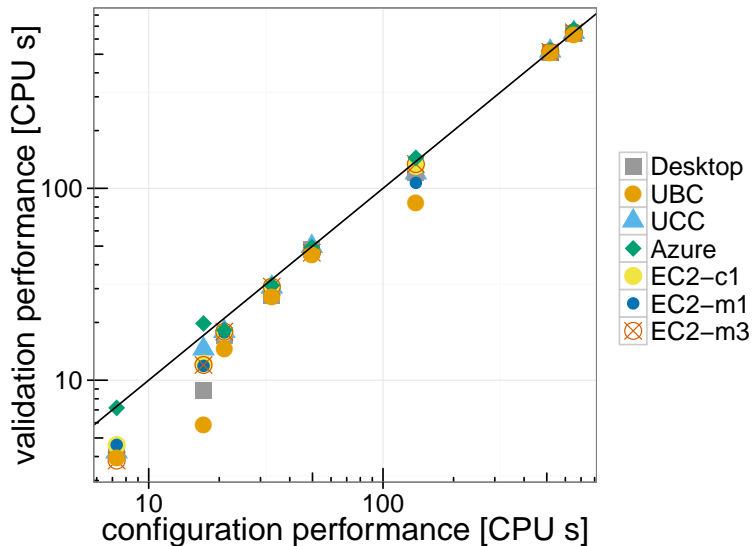
Results

ParamILS for Spear-SWV configured on UCC



Results

ParamILS for Spear-SWV configured on UCC



Results for other configurators/scenarios

- ▷ other results qualitatively similar
- ▷ SMAC
 - ▷ found better & more robust configurations
 - ▷ similar validation noise for found configurations
 - relative to variation across runs, validation noise is larger
- ▷ AutoWeka-*: less variation across compute infrastructures

Cost

single configuration scenario

≈\$15 for 8 different configurations and validation

Amazon cloud

≈\$100 to configure and ≈\$100 to validate per instance type

Microsoft cloud

\$138 to configure and \$94 to validate

Conclusions

- ▷ algorithm configuration on virtualised hardware (cloud) is possible
- ▷ configurations found on one infrastructure transfer to another
- ▷ larger virtual machine instances are more reliable