Extreme Scale Cloud Computing

Justin Y. Shi | shi@temple.edu

December 11, 2014

NSFCloud Workshop on Experimental Support for Cloud Computing

21st Century Software Challenge

Extreme scale mission critical **software** engineering principles

Distributed Computing Fallacies At Work

(Peter Deutsch'94, James Gosling'97, Blog by Arnon Rotem-Gal-Oz'2012)

1. **Network is reliable**: the **Virtual Circuit** concept.
2. **Latency is zero**: Amdahl's and Gustafson's **Laws**.
3. **Bandwidth is infinite**: Amdahl's and Gustafson's **Laws**.
4. **Network is secure**: Well-know ports for **web services**.
5. **Topology does not change**: **Explicit-parallel** programming.
6. **There is one admin**: **Security** at perimeters only.
7. **Transport cost is zero**: Amdahl's and Gustafson's **Laws**.
8. **Network is homogeneous**: **Explicit-parallel** programming.

New Fallacies

1. **Cloud applications are more reliable**: They are not. They are **more likely to crash** due to resource sharing but they are easier to restart.
2. **Single Truth Model is sufficient** for identity resolution: It is not. “Data cleansing” before analytics **erases** digital data evidences.

Broad practices bred great software engineering **conjectures**:

1. **CAP**: Consistency, Availability and Partition tolerance probably can only be satisfied partially.
2. **Single point failures** probably can never be all eliminated
3. **True scalability** is probably impossible (gaining performance and reliability at the same time as we upscale processing infrastructure).
4. An **architecture inflection point** was claimed (CCC Community White Paper: 21st Century Architecture)

“Half a truth is often a great lie” – Ben Franklin

- It appears that software industry tends to make overly optimistic assumptions
- Academia follows
- The 8 fallacies have been documented for more than a decade. The software industry as it stands is not sustainable.
- More worrisome are the newly surfaced ones. This “communal debt” has generated more debts.

Service Software Engineering Due Diligence

- **Direct program-program communication** is **NOT** reliable.
- **Well-known ports** should **NOT** be used for secured web services.
- **Performance-only** or reliability-only technologies are **INSUFFICIENT**.
- **Scalability** **needs** a new definition.
- **Approximation match** **needs** a definition.
- **Single truth model** is insufficient for identity resolution in multiple social network datasets

21st Century Service Software Discipline

- Statistic multiplexing is the only proven technique for solving the impossibility communication problem theoretically.
- Statistic multiplexed computing is probably the only sustainable software engineering principle for extreme scale mission critical service computing.
- Statistic multiplexed computing probably also holds the key for solving the long standing Internet service security problems.
- HPC Clouds are the ultimate testbeds for sustainable HPC programming.

Preliminary Results: SC14 Research Exhibit (Booth#3643)