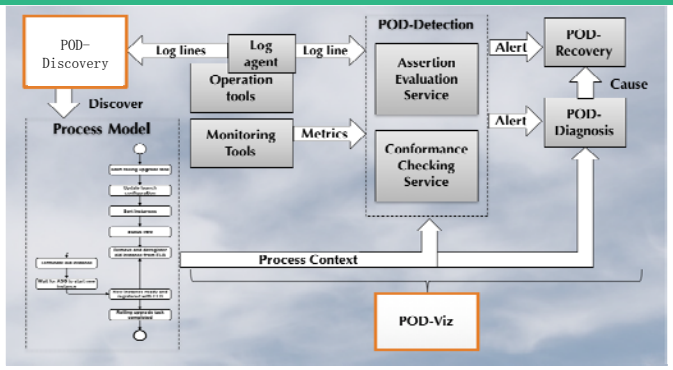
POD-Discovery & POD-Viz

Detecting and diagnosing errors during Cloud operations

www.data61.csiro.au

Detecting and diagnosing errors during Cloud operations based on *process knowledge*:

- Discovering the process from log events
- Visualizing progress and errors at runtime
- See tool demonstrations



Problem

- System logs are voluminous, distributed, low-level, noisy and diverse
- How to discover and track processes from these logs?
- How to visualize progress and errors to operators?

Solution

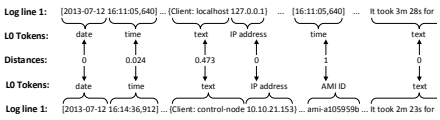
- Apply **process mining** techniques from business process management (BPM)
  - Exploit operation log characteristics, e.g., extensive amount of logging
  - Collect and correlate multiple logs
  - Discover process model from logs
- Use process knowledge to detect deviations and anomalies
- Trigger automatic diagnosis and even recovery
- Visualize relevant information

Next Steps & Impact

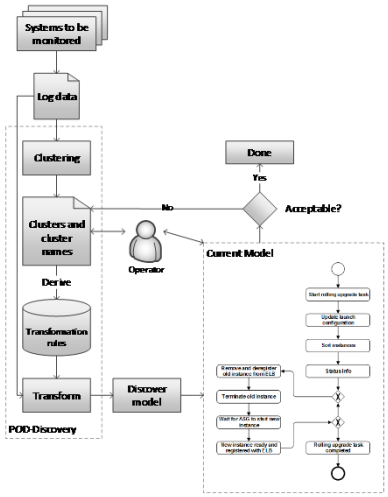
- Next: unprecedented fine-grained log analysis by considering timing profiles and numerical invariants
  - Tackled multi-instantiation, to deal with 100s of machines in parallel
- Application to diverse data sets
- Actively looking for trial projects
- POD-Discovery & POD-Viz hardened and (soon to be) released
- Research publications:
  - ACM SAC 2015 paper
  - DSN 2015 demo paper in review

POD-Discovery

- Pre-process logs so BPM tools for discovery can handle the data
  - Tokenize log lines
  - Calculate distance per log line pair with token-type specific functions

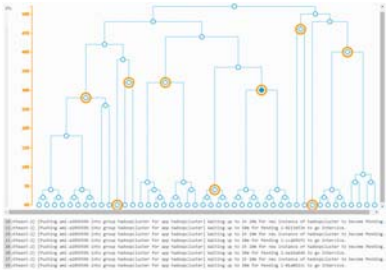


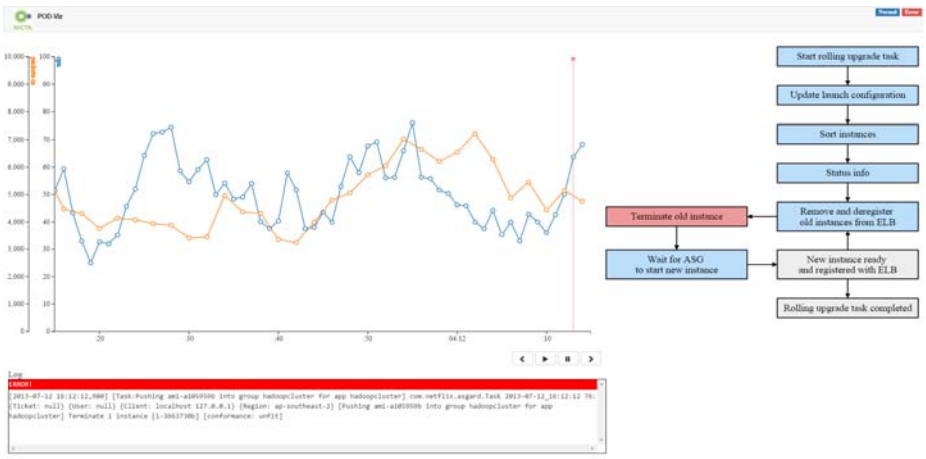
- Hierarchical agglomerative clustering (HAC)
- Interactive dendrogram allows control over clustering



POD-Viz

- Goal: Support operators in identifying and diagnosing errors quickly
- Visualize:
  - Process model & instance progress
  - Timeline of events
  - Errors and log lines
- Allow pause, skipping back / forward and replay





FOR FURTHER INFORMATION
Ingo Weber
e Ingo.Weber@data61.csiro.au