



Performance Analysis with PCP (Performance Co-Pilot) and R

Paul Evans
October 2015

Outline

- PCP Overview
 - Introduction to Performance Co-Pilot
 - PCP Basics
- R Overview
 - What is R?
 - Interfacing PCP and R
- Analysis of a demo workload run on GFS2
- Future work
- Questions?

Performance Co-Pilot

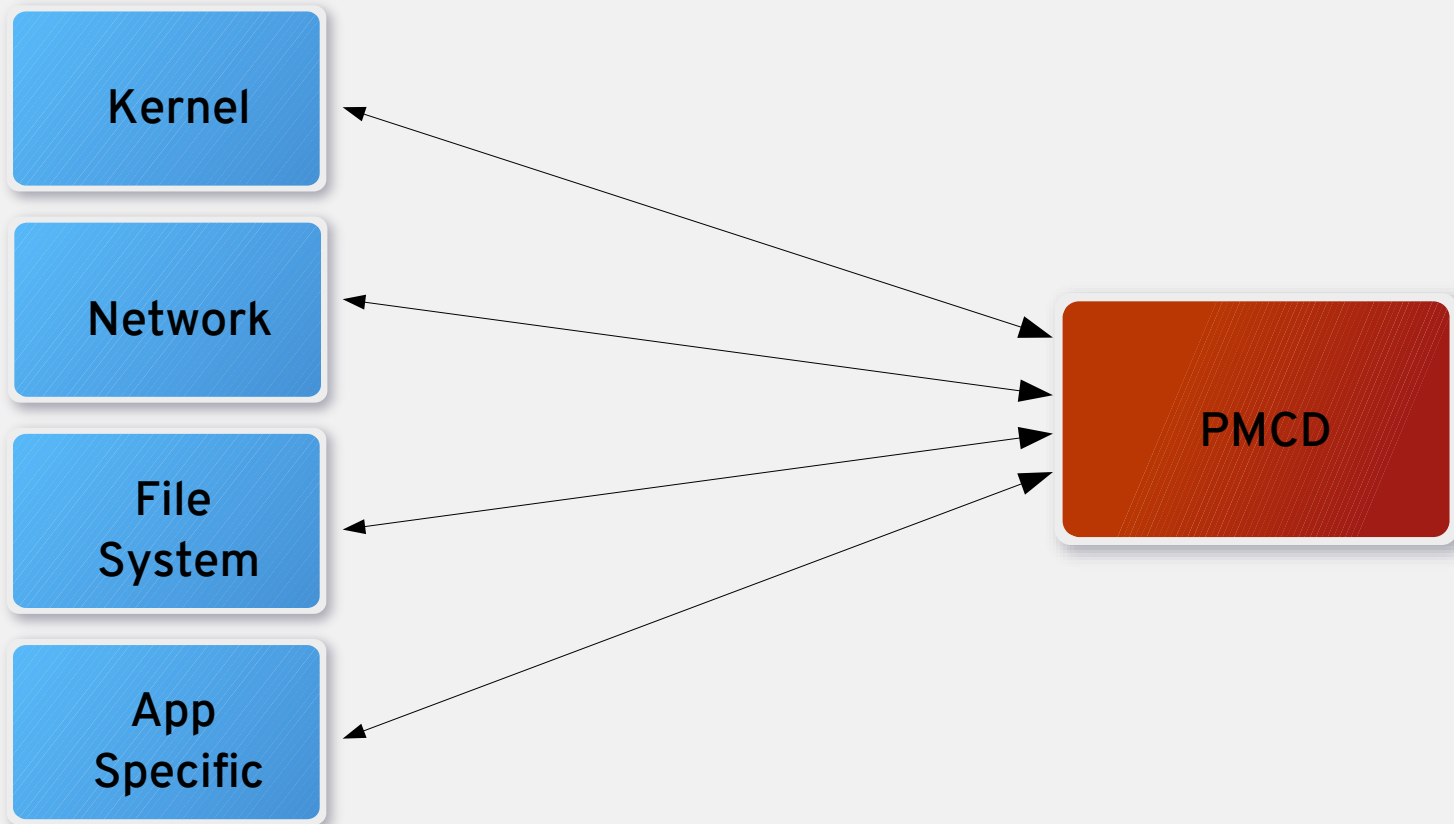
- What is PCP?
 - Open source toolkit
 - System-level analysis
 - Live and historical
 - Extensible (monitors, collectors)
 - Distributed
 - Cross platform
 - Consistent unit measurement

Performance Co-Pilot

- Two main underlying components
 - Performance Metric Domain Agents (Agents)
 - Know how to gather information
 - E.g. Kernel, Mail, DB, File Systems, App
 - Performance Metric Collection Daemon (PMCD)
 - Organizes, collects, manages metric information

Performance Co-Pilot

Agents

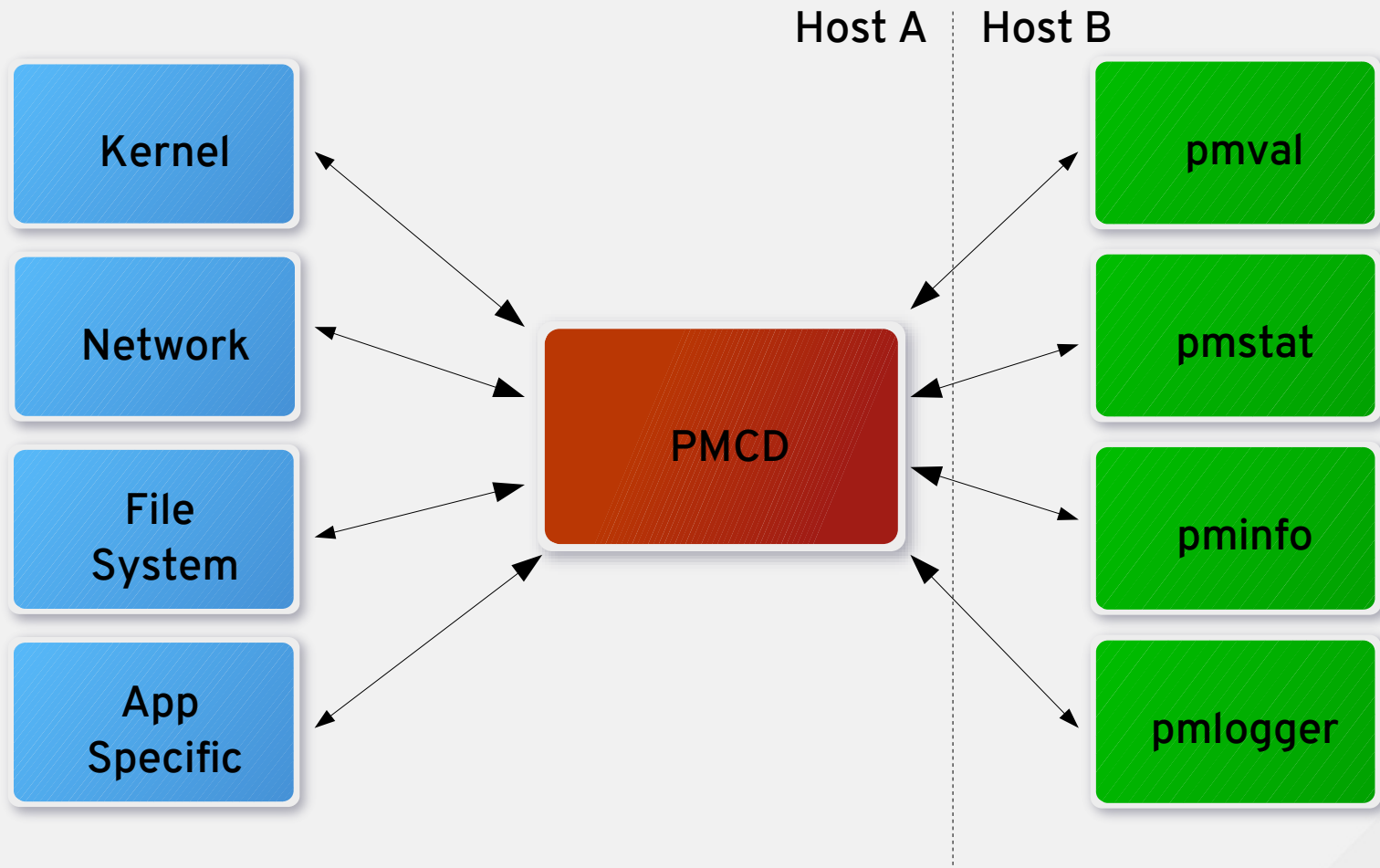


Performance Co-Pilot

Number of metrics available in PCP?

- A large amount (~1500 by default in Fedora)
- Variation between what they are measuring
 - Kernel
 - Network
 - Database
 - Webserver
 - File Systems
- How are they named?
 - Performance Metrics Name Space (PMNS)
 - `disk.all.read_bytes`

Performance Co-Pilot



Performance Co-Pilot

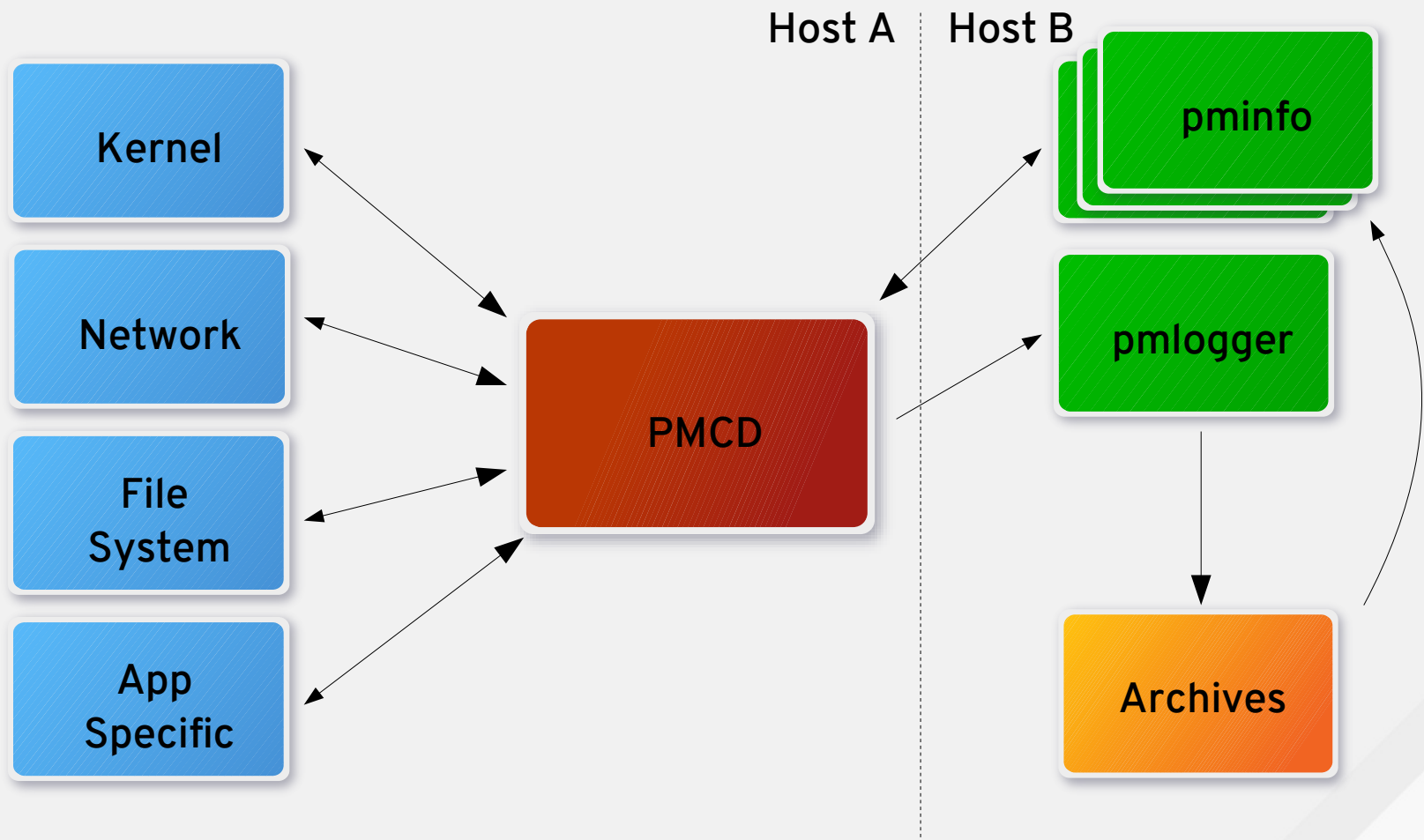
Where to start?

Pminfo – display information about metrics

```
$ pminfo -t disk.all
```

disk.all.read	[total read operations, summed for all disks]
disk.all.write	[total write operations, summed for all disks]
disk.all.total	[total (read + write) operations, summed for all disks]
disk.all.read_bytes	[count of bytes read for all disk devices]
disk.all.write_bytes	[count of bytes written for all disk devices]
disk.all.total_bytes	[count of bytes read and written for all disk devices]

Performance Co-Pilot



Performance Co-Pilot

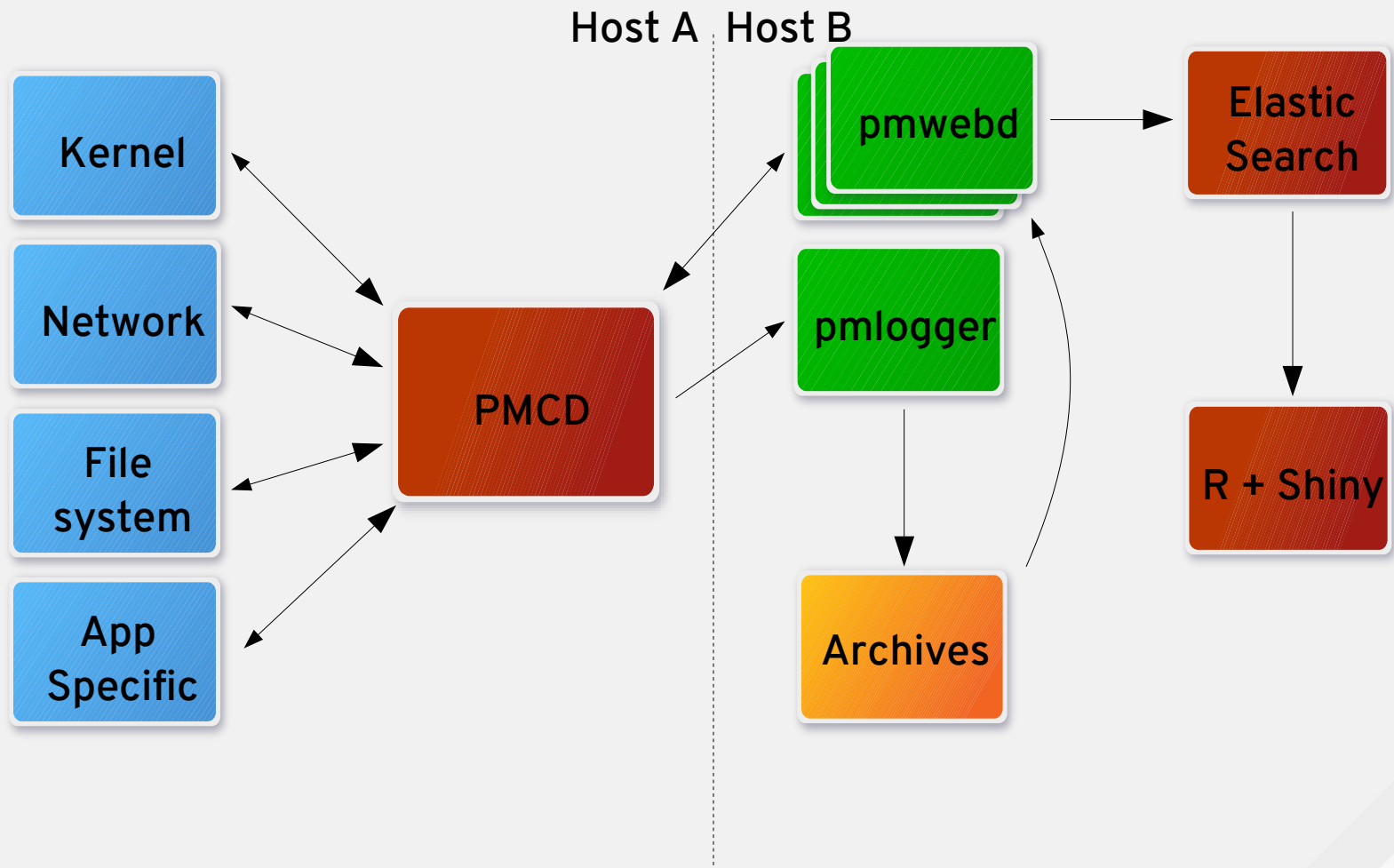
pmlogger creates logs for future analysis

- Allows us to use tools on older data, retrospectively
- Default logging profile about 5mb a day, rotates and compresses
- Organized

R Project for Statistical Computing

- What is R?
 - Environment for statistical computing and graphics
 - Programming language designed expressly for data analysis
 - Extendable through scripts
 - Open source with a large community
 - Cross platform

Interfacing PCP and R



Interfacing PCP and R

- pmwebd
 - Network daemon that bridges PCP client API to web applications through a JSON API
 - Can be used to grab both live events (PMCD) and from PCP archives
- Elastic Search
 - Open source analytic engine
 - Used to store our intermediate PCP data
 - Pre-existing data import scripts for R
- R Server + Shiny
 - Data Analysis and modeling of our metrics
 - Shiny to output our charts as a simple web app

Demo workload with pmchart

R Diagrams

Future Work

- Evaluate further workloads suitable for analysis
- Streamlining the capture process
 - R can read import JSON data
 - Create R library to read data directly from PCP
 - Both live and archived data

Questions?



THANK YOU