

Visualizing Map-Reduce

who is Paul Baclace?

- * worked on Nutch at archive.org with Doug Cutting prior to splitting Nutch into Hadoop and Nutch crawler in Jan. 2006

- * submitted code patches and bugs (HADOOP-9)

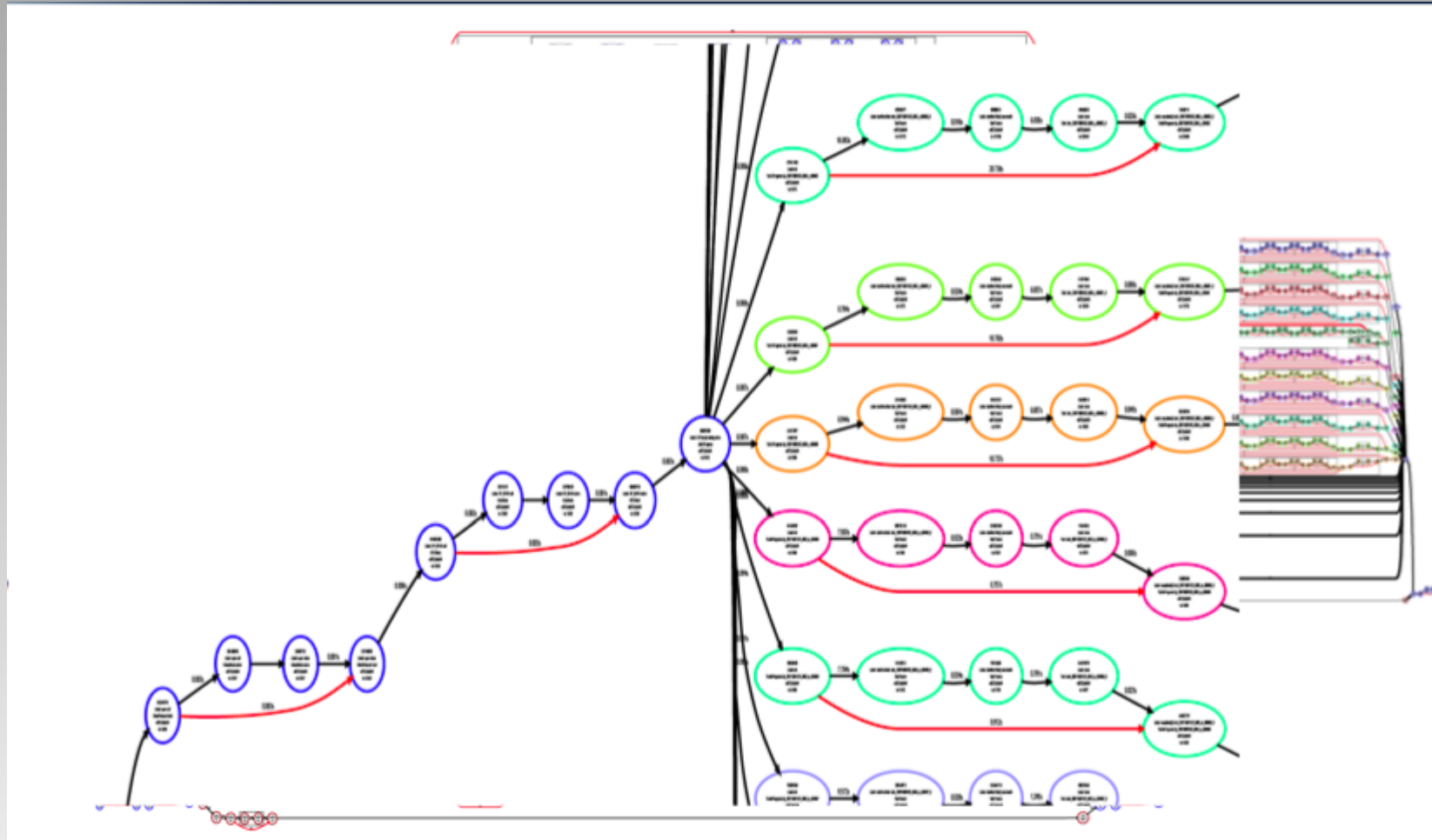
- * currently consulting on Web datamining using Hadoop, Heritrix, and Cascading.

Visualizing Map-Reduce

Previous Visualization Efforts

- * Ganglia: machine level oriented
- * xtrace: too detailed
- * SALSA / Mochi: nice swim-lanes, not interactive

X-Trace example



Mochi example

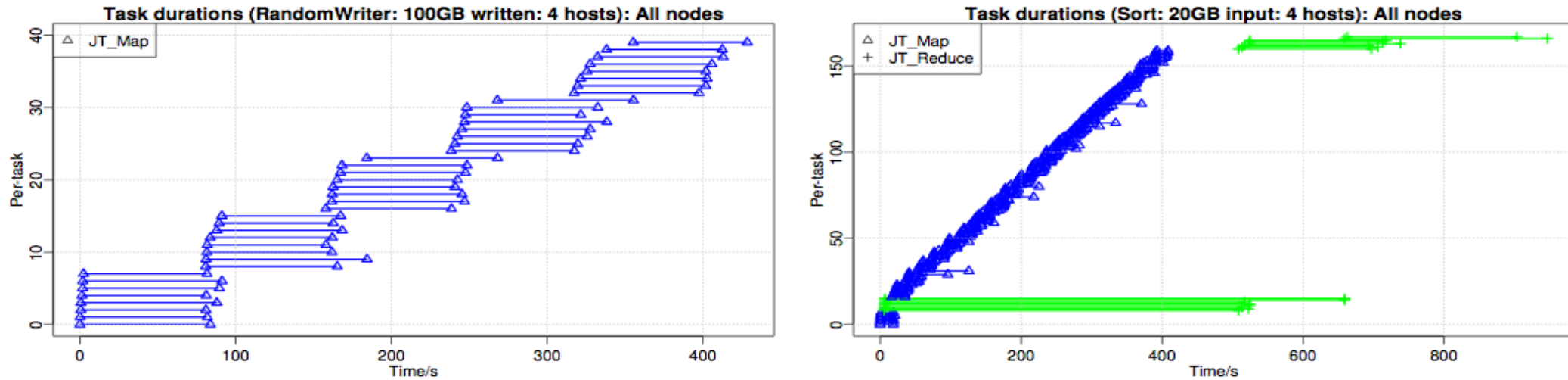
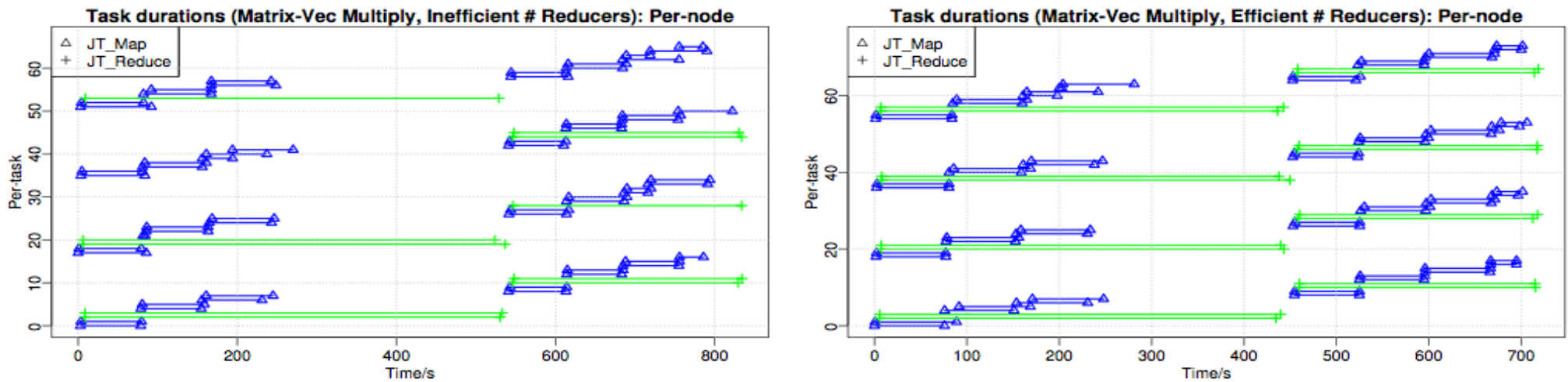


Figure 5: Summarized *Swimlanes* plot for RandomWriter (top) and Sort (bottom)

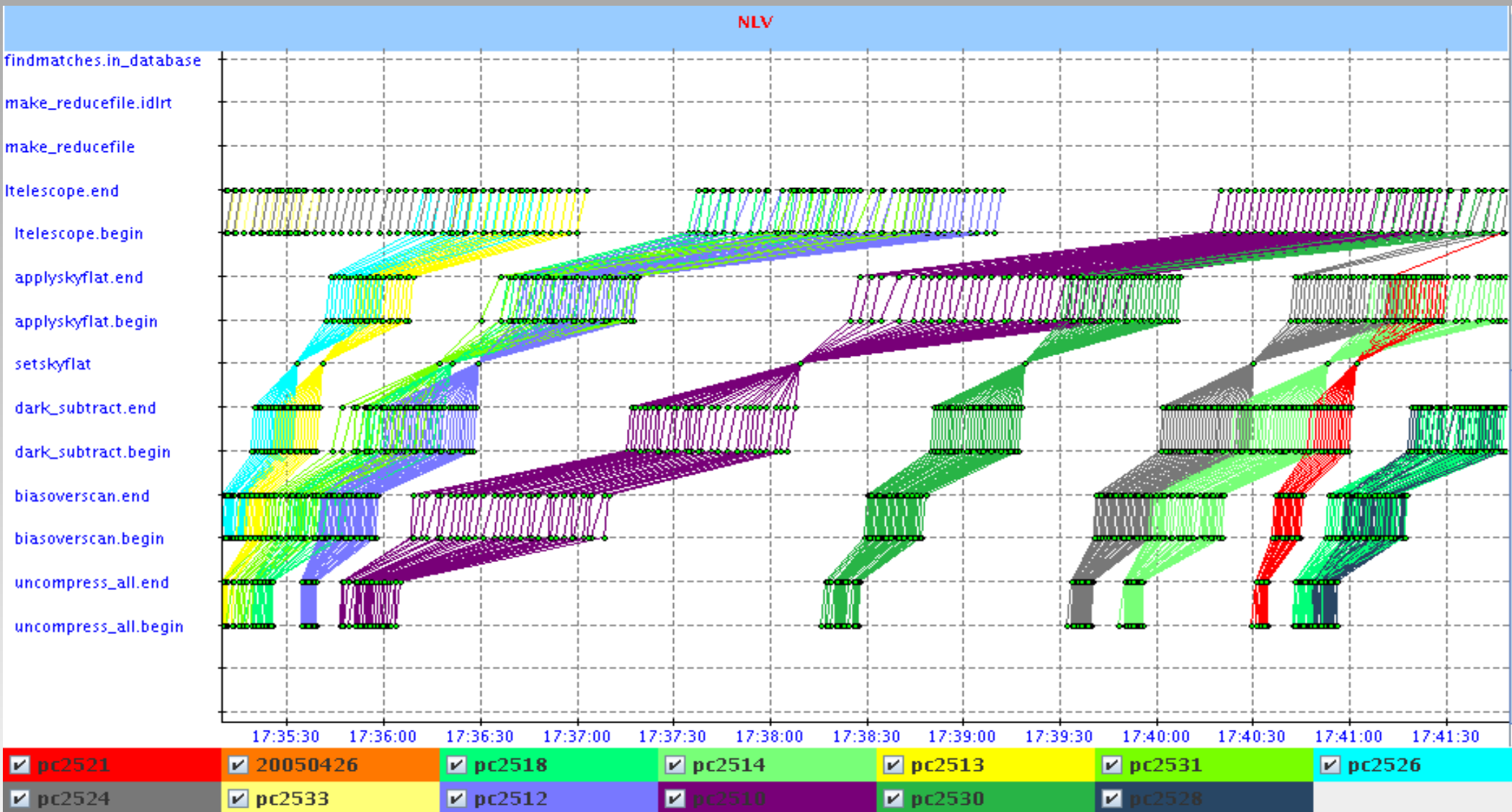


Supercomputer Visualization

Why not NetLogger?

- * Used at Lawrence-Berkeley Labs
- * Since 1999
- * Had Interactive “nlv” collecting dust
- * Now only uses R to make static charts
- * Never applied to Map-Reduce

NetLogger example



Java NetLogger Visualizer

Process

1. Each slave post-processes logs
2. Logs are sorted, merged, filtered
3. Visualizer reads a “universal log format” file with:
 attrName=attrValue ...

JNLV

Status

- * Java nlv never previously released as open source.
- * Contains BSD-style and LGPL code
- * Log event collection is via a bash script
- * bugs fixed and arbitrary annotations added
- * Still compatible with NetLogger

URLs

New Name: Phatvis (PHaseTimeVIS)

Project Page:

<http://assembla.com/spaces/phatvis>

NetLogger Background:

<http://dsd.lbl.gov/NetLoggerWiki/>