Three things you really should know about DAGMan Allstars

Alain Roy
OSG Software Coordinator
Condor Team Member

Dagman Allstars really rocks!

- Funk band, formed in 1998, released one album
- Sadly, separated in 2001
- Founder, Dan Monceaux, now does animation

2011-03-dagman.ppt

Home | Themes | Tables | Charts | SmartArt | Transitions | Animations | Slide Show | Review

Slides | Font | Paragraph | Insert | Format | Slide Show

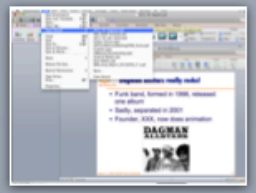
New Slide | Layout | Section | Text | Picture | Shape | Media | Arrange | Quick Styles | Play

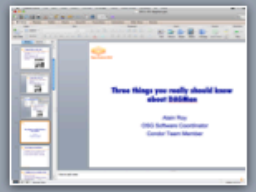
Slides | Outline

1

2

3

4


5


6

Open Science Grid

Dagman Allstars really rocks!

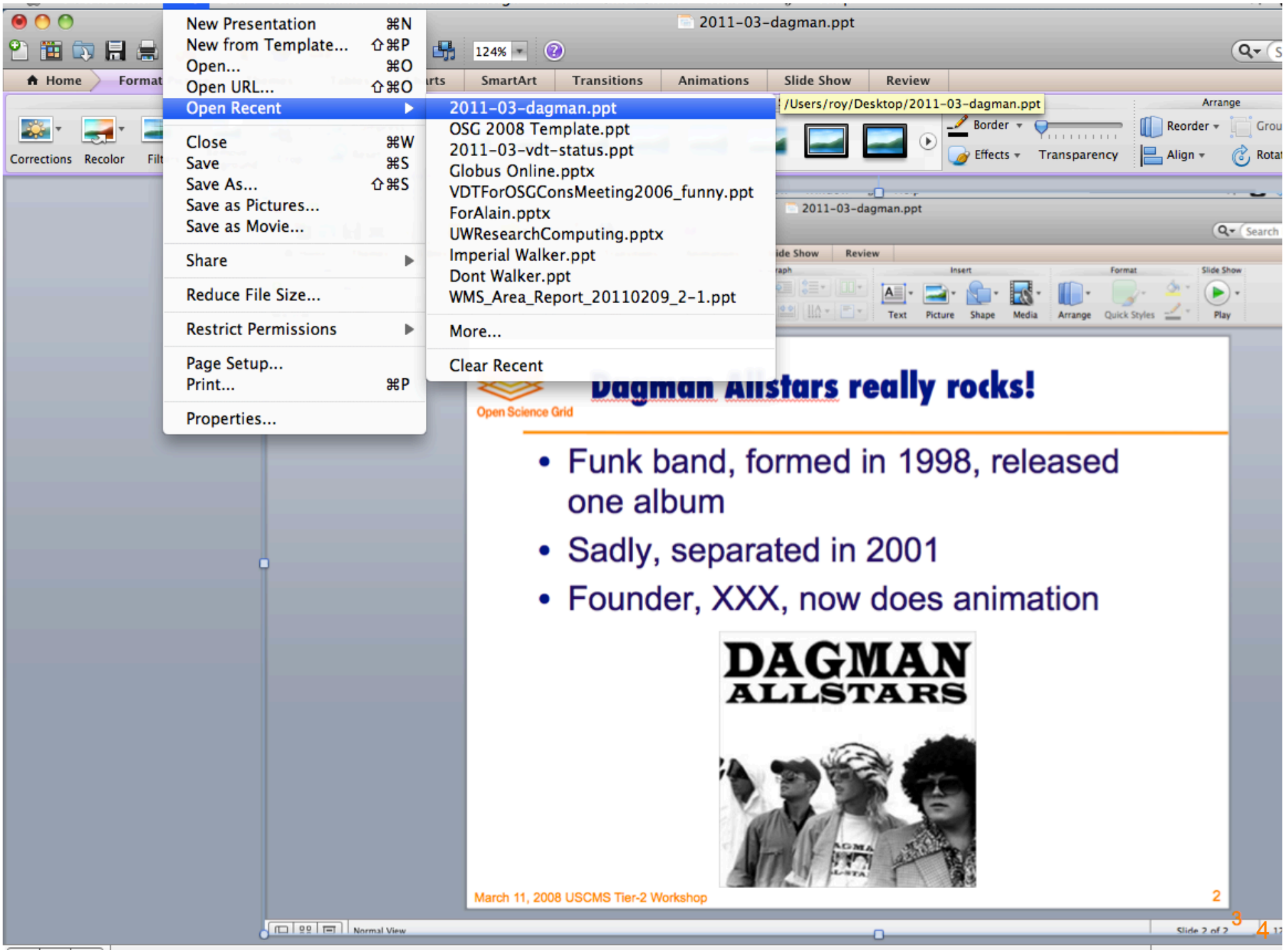
- Funk band, formed in 1998, released one album
- Sadly, separated in 2001
- Founder, Dan Monceaux, now does animation

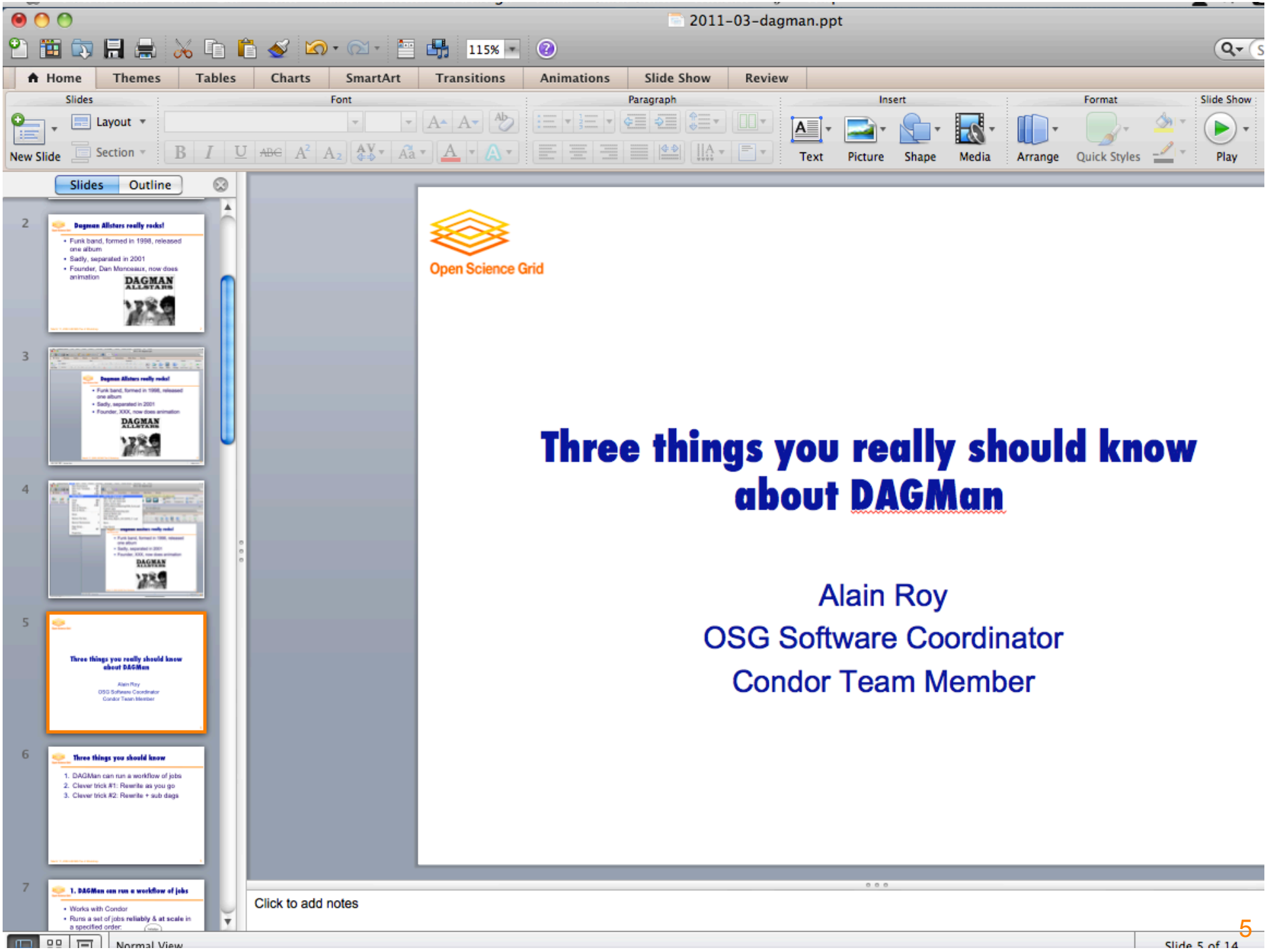
March 11, 2008 USCMS Tier-2 Workshop

Click to add notes

Presenter View

Slide 2 of 15


New Presentation ⌘N
New from Template... ⇧⌘P
Open... ⌘O
Open URL... ⇧⌘O
Open Recent
Close ⌘W
Save ⌘S
Save As... ⇧⌘S
Save as Pictures...
Save as Movie...
Share
Reduce File Size...
Restrict Permissions
Page Setup...
Print... ⌘P
Properties...
2011-03-dagman.ppt
OSG 2008 Template.ppt
2011-03-vdt-status.ppt
Globus Online.pptx
VDTForOSGConsMeeting2006_funny.ppt
ForAlain.pptx
UWResearchComputing.pptx
Imperial Walker.ppt
Dont Walker.ppt
WMS_Area_Report_20110209_2-1.ppt
More...
Clear Recent
2011-03-dagman.ppt
Home
Format
SmartArt
Transitions
Animations
Slide Show
Review
/Users/roy/Desktop/2011-03-dagman.ppt
Corrections
Recolor
Border
Effects
Transparency
Reorder
Align
Text
Picture
Shape
Media
Arrange
Quick Styles
Play
Open Science Grid
Dagman Allstars really rocks!
• Funk band, formed in 1998, released one album
• Sadly, separated in 2001
• Founder, XXX, now does animation
DAGMAN ALLSTARS
March 11, 2008 USCMS Tier-2 Workshop
2
Normal View
Slide 2 of 2


2011-03-dagman.ppt
115%
Home
Themes
Tables
Charts
SmartArt
Transitions
Animations
Slide Show
Review
Slides
Font
Paragraph
Insert
Format
Slide Show
New Slide
Layout
Section
Text
Picture
Shape
Media
Arrange
Quick Styles
Play
Slides
Outline
Open Science Grid
Three things you really should know about DAGMan
Alain Roy
OSG Software Coordinator
Condor Team Member
Click to add notes
Normal View
Slide 5 of 14

Three things you really should know about DAGMan

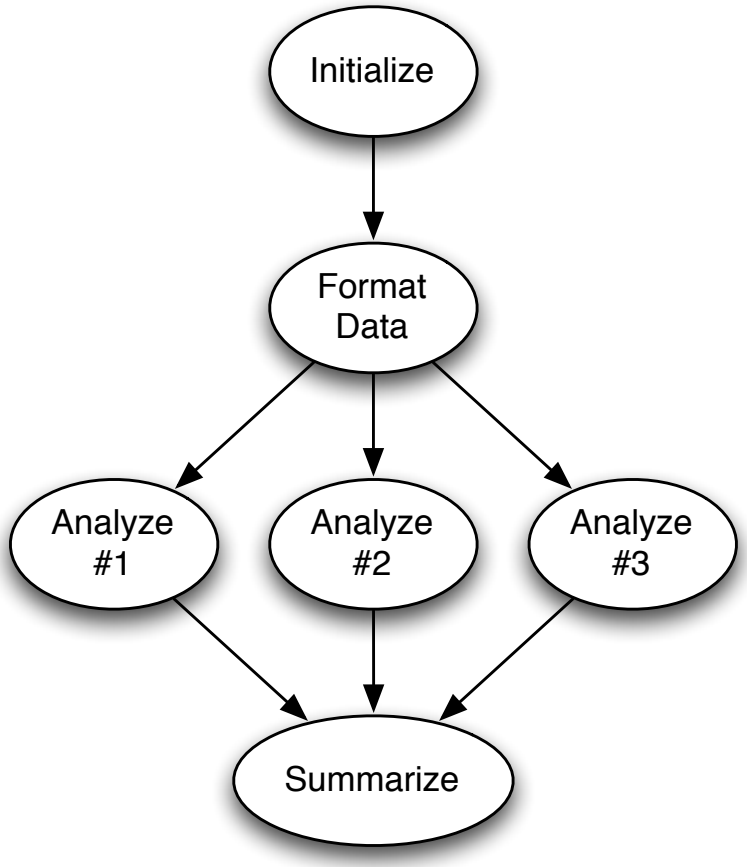
Alain Roy
OSG Software Coordinator
Condor Team Member

Three things you should know

1. DAGMan can run a workflow of jobs
2. Clever trick #1: Rewrite as you go
3. Clever trick #2: Rewrite + sub dags

1. DAGMan can run a workflow of jobs

- Works with Condor
- Runs a set of jobs **reliably** & **at scale** in a specified order:



Open Science Grid

1a. Works with Condor

- Runs entirely within Condor:
  - DAGMan itself is a reliable Condor job
  - Each job can run in/on
    - Local submission computer
    - Local Condor pool
    - Condor-G to Globus, CREAM, etc…
    - GlideinWMS pool
- DAGs are relatively simple
  - No loops
  - No conditionals

1b. Runs reliably

- If DAGMan itself fails, Condor restarts it
- If DAG is interrupted, DAGMan resumes based on saved state (logs)
  - Rescue DAG

Open Science Grid

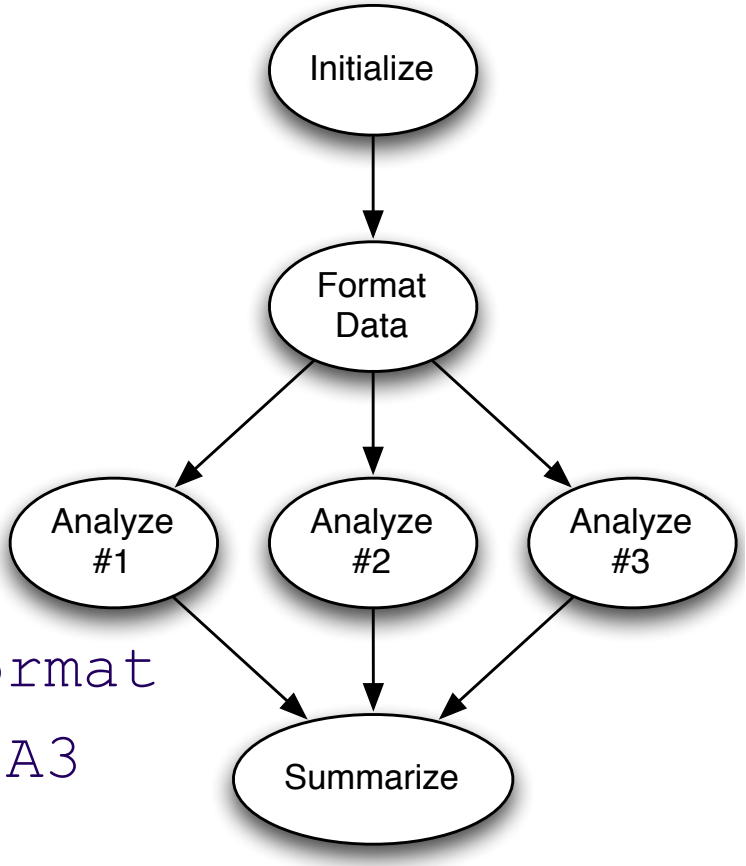
1c. Runs at scale

- Examples of scale:
  - We've run DAGs with 1,000,000 nodes
  - LIGO has run real workflows with 500,000+ nodes
  - My colleague helps local scientists run DAGS of 1,000 to 5,000 nodes every day
- Scaling depends on the details
  - Can be finely tuned to throttle various aspects of workflow

Easy to specify

```
JOB Initialize init.sub
JOB Format     format.sub
JOB A1         a1.sub
JOB A2         a2.sub
JOB A3         a3.sub
JOB Summarize  s.sub

PARENT Initialize CHILD Format
PARENT Format CHILD A1 A2 A3
PARENT A1 CHILD Summarize
PARENT A2 CHILD Summarize
PARENT A3 CHILD Summarize
```

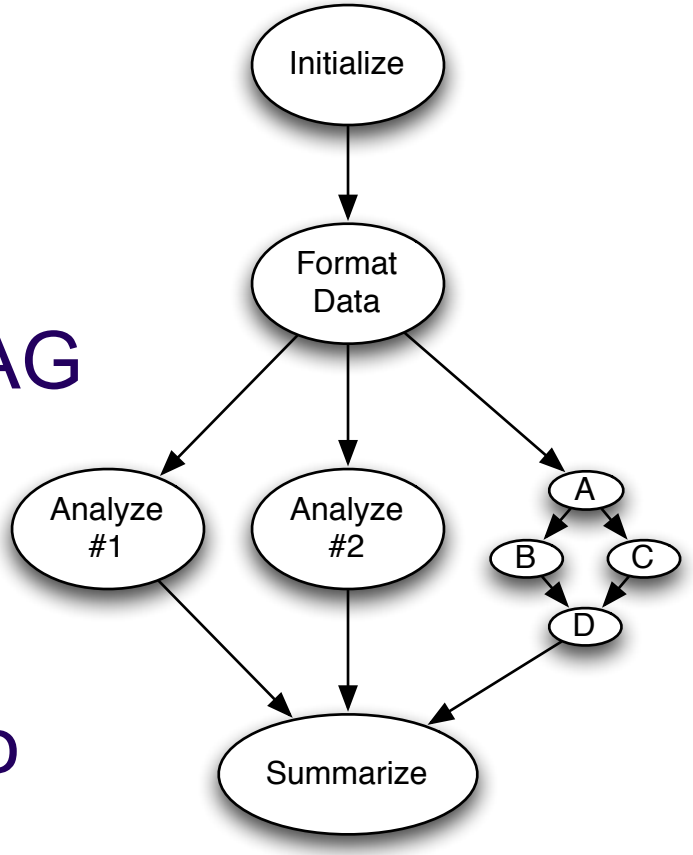


2. Clever trick #1: Rewrite as you go

- Each node in the workflow can have:
  - Pre-script: Runs just before node
  - Post-script: Runs just after node
- The pre-script can edit the node itself to change its behavior
  - Change parameters of jobs based on previous results, etc…

3. Clever trick #2: Rewrite + sub dags

- A single node in the workflow can be an entire DAG (sub-DAG).
  - Separate specification
  - Separate DAGMan process
  - But it acts like a single node
- But you can rewrite that DAG before running it, so it's the right size, shape, etc!
- Colleague uses this to dynamically adjust DAGs to meet needs, as they run.



Open Science Grid

Conclusion

- Start with a relatively simple construct
  - Workflows
  - No loops
  - No conditional
  - Reliable and easily scaled
- Add two features
  - Ability to run pre-script
  - Ability to run DAG as a node
- End up with very flexible workflow system

Open Science Grid

Questions?

- I could have said a lot more about DAGMan
  - Variable substitution…
  - Exactly how to write a DAG…
  - Sub-DAGs vs. splices…
- But hopefully this was simple and inspirational
- Ask me questions now, or until Thursday @ noon.