Cytel

Shaping the Future of
Drug Development

Benefits of Grid Computing in Clinical Industry

PhUSE SDE 2017 - Mumbai

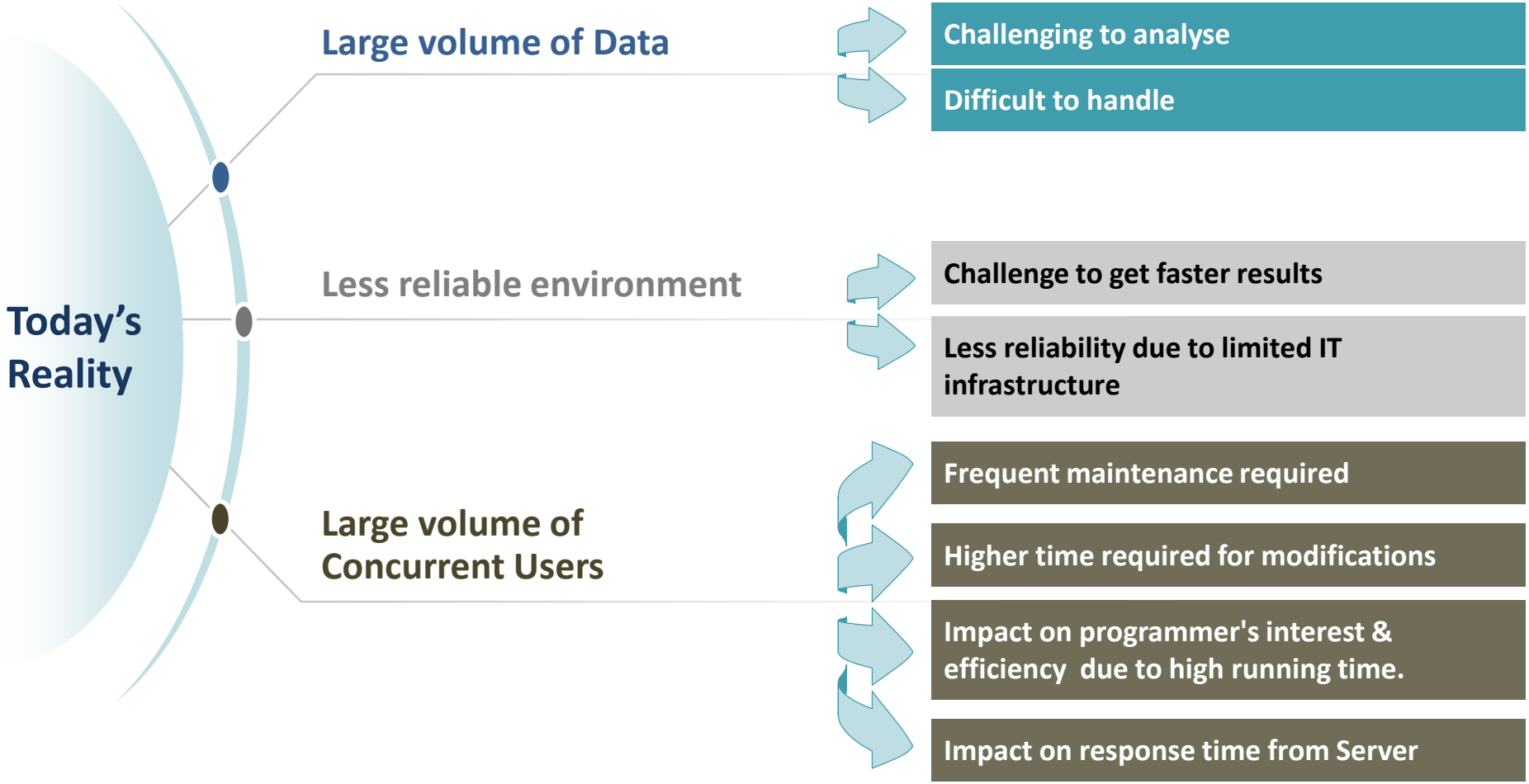
Disclaimer

The opinions expressed in this presentation and on the following slides are solely those of the presenter and not necessarily those of Cytel. Cytel does not guarantee the accuracy or reliability of the information provided herein.

Presentation Objective

- Overview of Grid Computing
- Concept of Parallel Processing
- Case Study
- Parallel Processing by SAS program
- Merits & Demerits
- Q&A

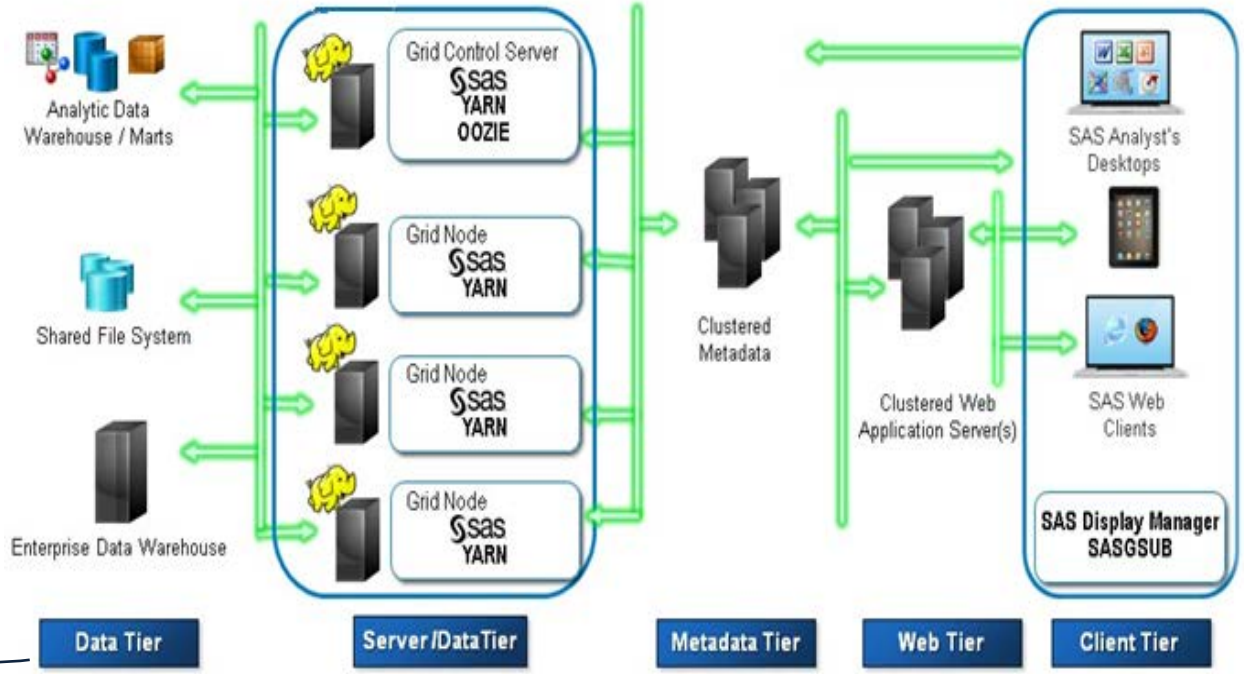
Grid Computing
Overview



Grid Computing
Overview

- Necessity of reduced available processing time vs higher volume of data, one must have to find ways to increase efficiency while keeping cost down.

- Grid computing can reduce overall elapsed time while leveraging existing hardware and software.



✓ **Ability to reduce overall processing time**

✓ **Ability to leverage existing hardware**

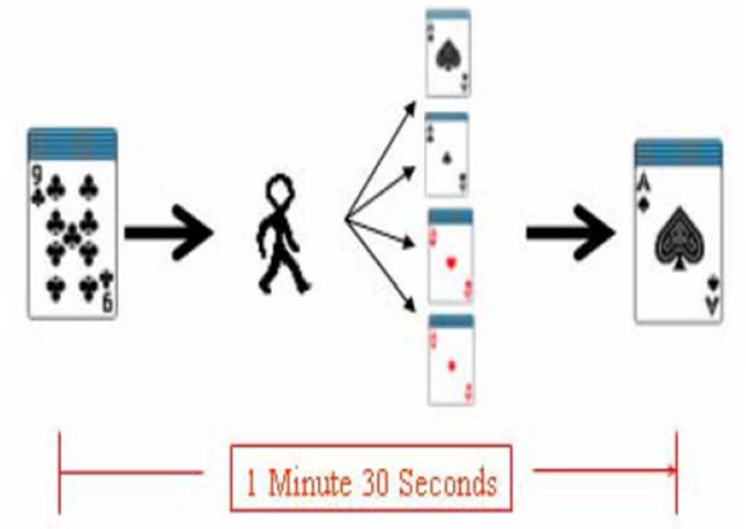
Concept of Parallel Processing

Overview

Parallel Processing

Grid Computing enables the ability to divide a task (SAS Program) into subtasks and these subtasks can be executed in a parallel fashion.

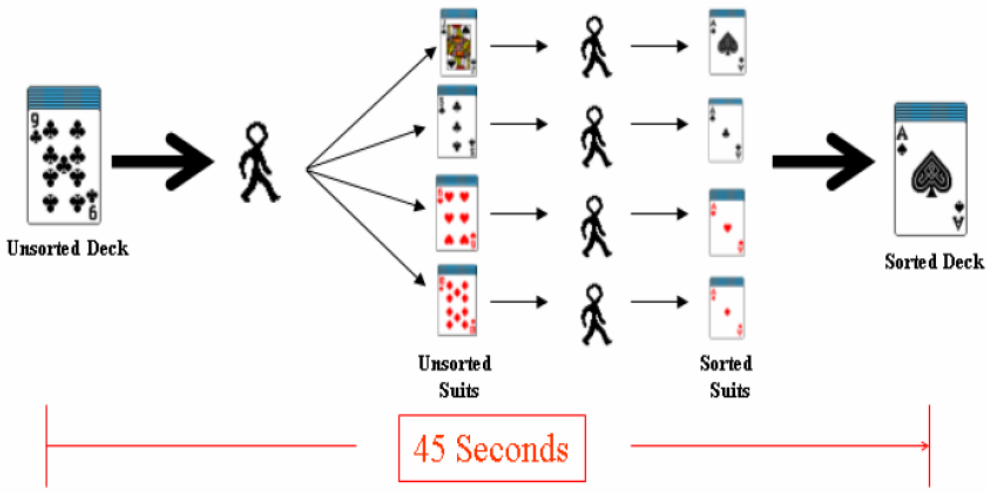
A great way to illustrate the concept of Grid Computing is by using a common deck of playing cards.



- Cards need to be sorted from Ace to King within each suit.
- Task is assigned to one person.
- Time taken → 1 min. 30 sec.

Concept of Parallel Processing

Overview



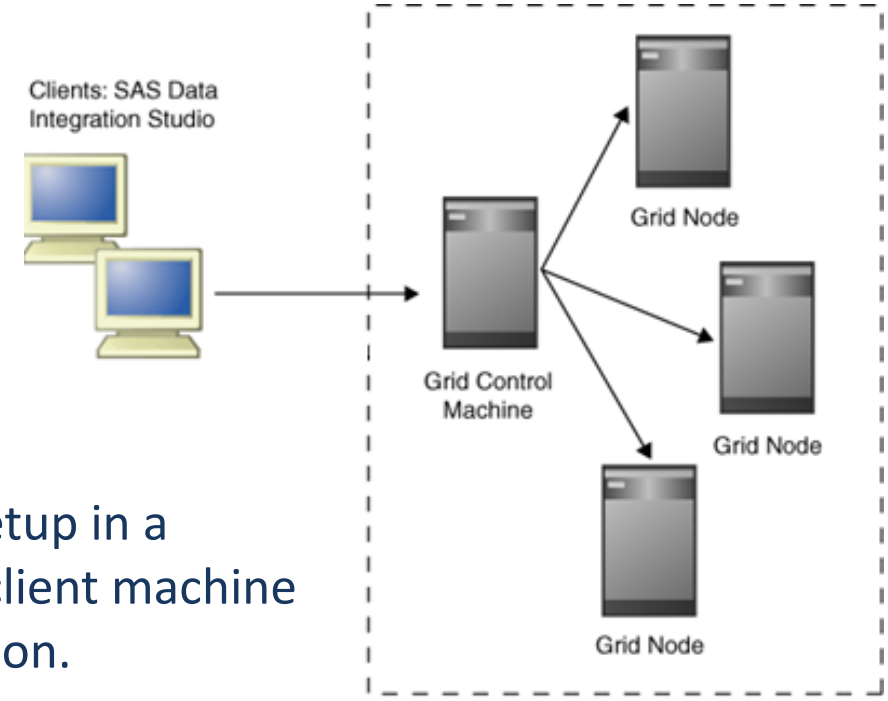
- Task is assigned to more than one person.
- Time taken → 45 sec.

- The benefit is gained based on the fact that some of the sub-tasks are performed at the same time or in parallel.
- The only difference in the Information Technology World is, we use computers instead of people in a Grid Computing environment.

Concept of Parallel Processing
Infrastructure

Infrastructure

- Client
- Node
- Grid Controller Server



Prerequisites

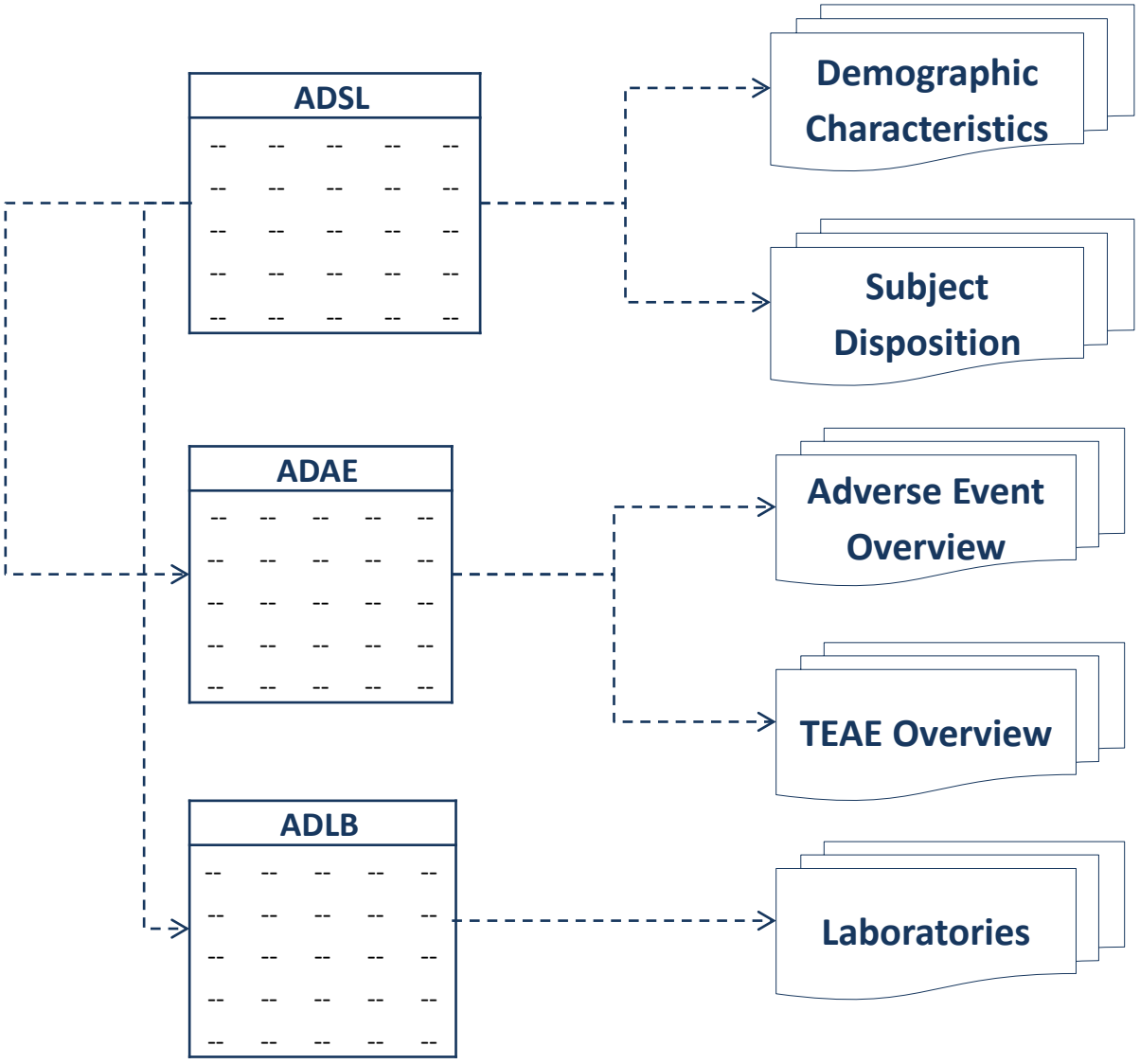
- Each of these Nodes must be setup in a particular fashion to allow the client machine to make the necessary connection.
- SAS must be installed on all machines to be used on the “Grid” (including the client machine).
- SAS “**Spawner**” service will need to be started and configured on a node in case of PC environment.

Concept of Parallel Processing

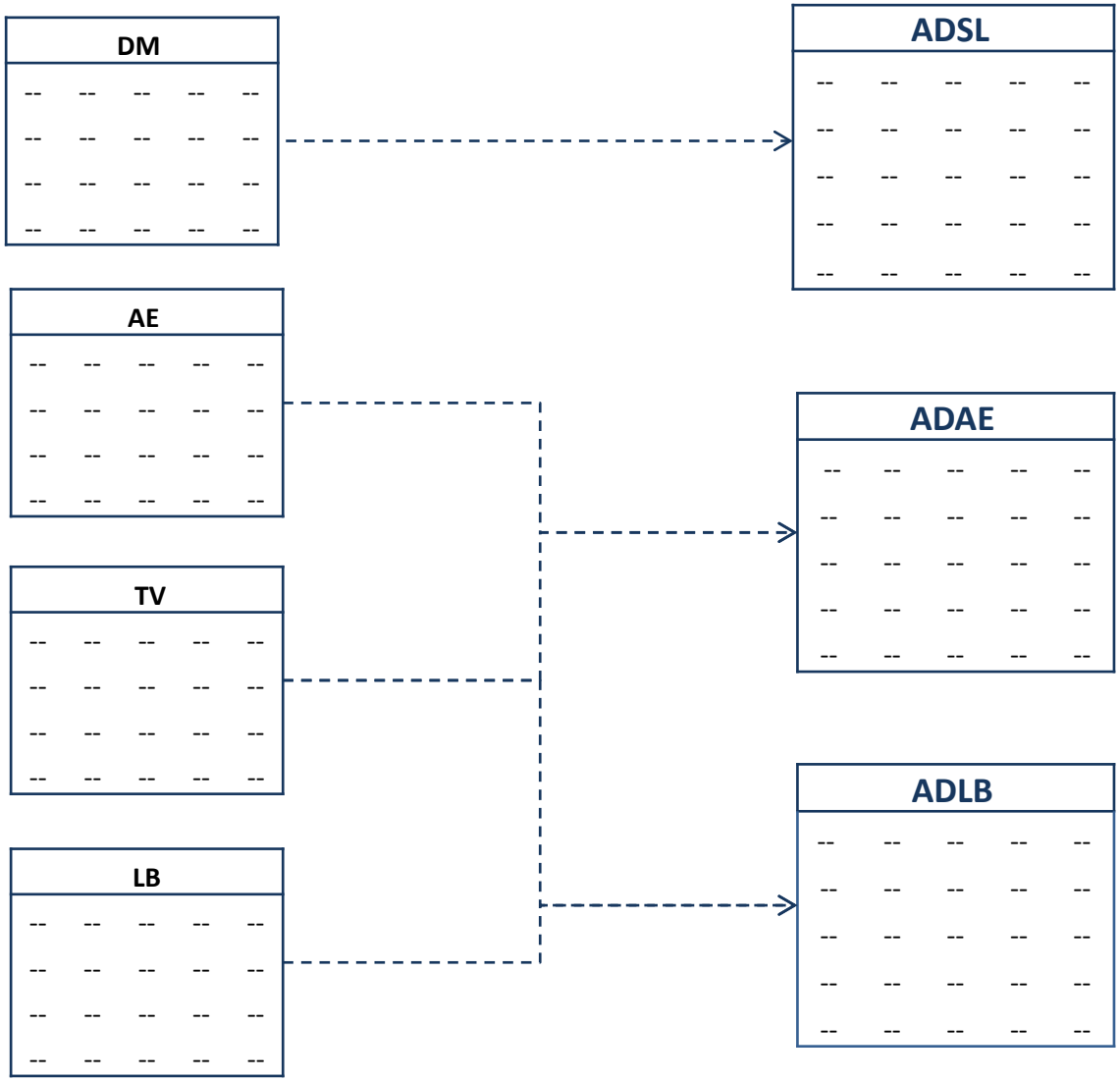
Infrastructure

- The Grid Controller Software is installed on the Grid Controller Server which helps to manage the Grid.
- The Grid Controller Software must manage three different aspects of the Grid:
  - ✓ **Nodes**
  - ✓ **Process Flow (Sub-tasks)**
  - ✓ **Job submission**
- In this presentation we will use a SAS\AF prototype to illustrate the various requirements of the Grid controller.
- Grid Infrastructure also has SAS Grid Environment Manager which helps to monitor all tasks submitted to Grid.

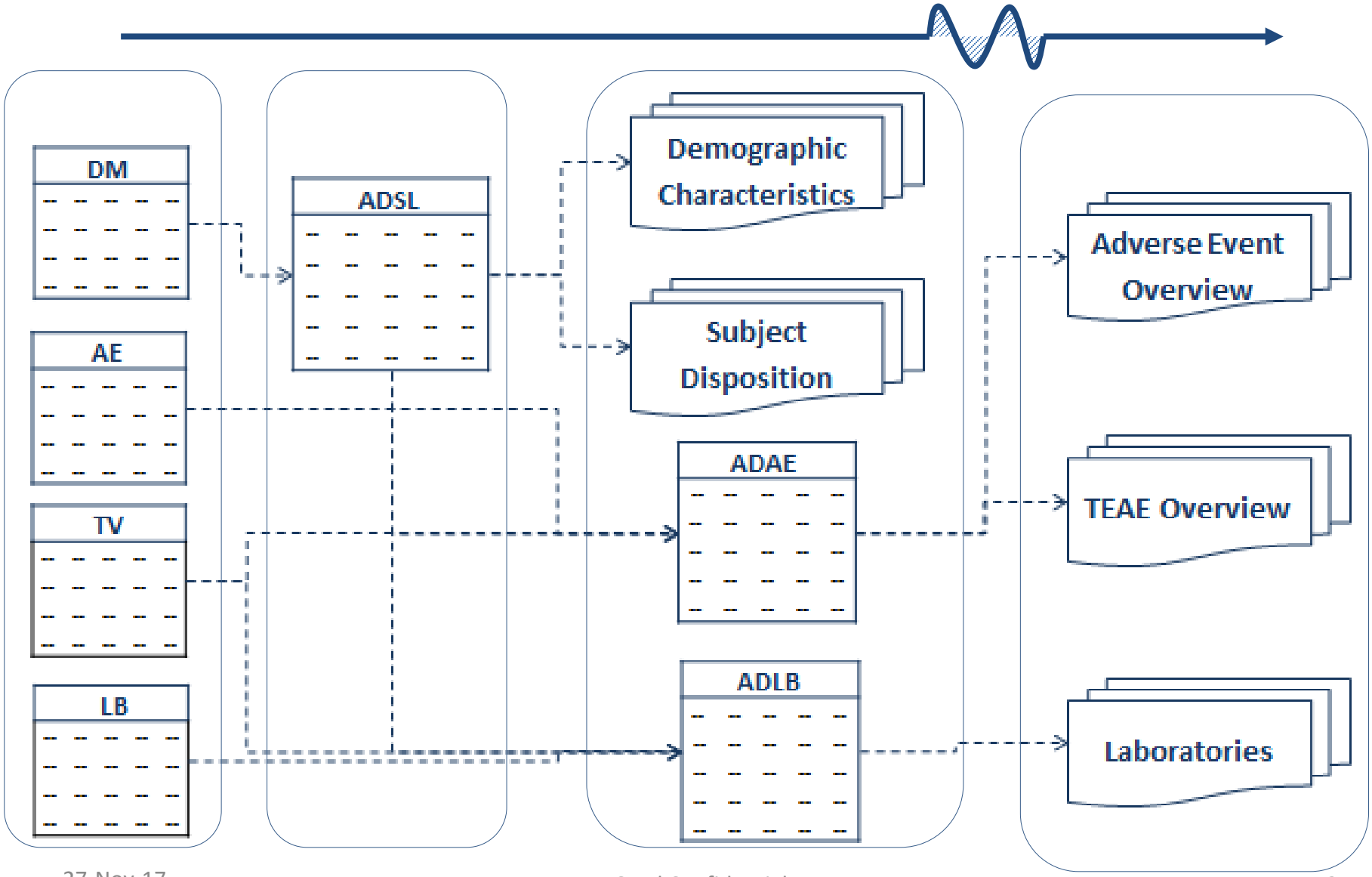
Case Study
Overview



Case Study
Overview



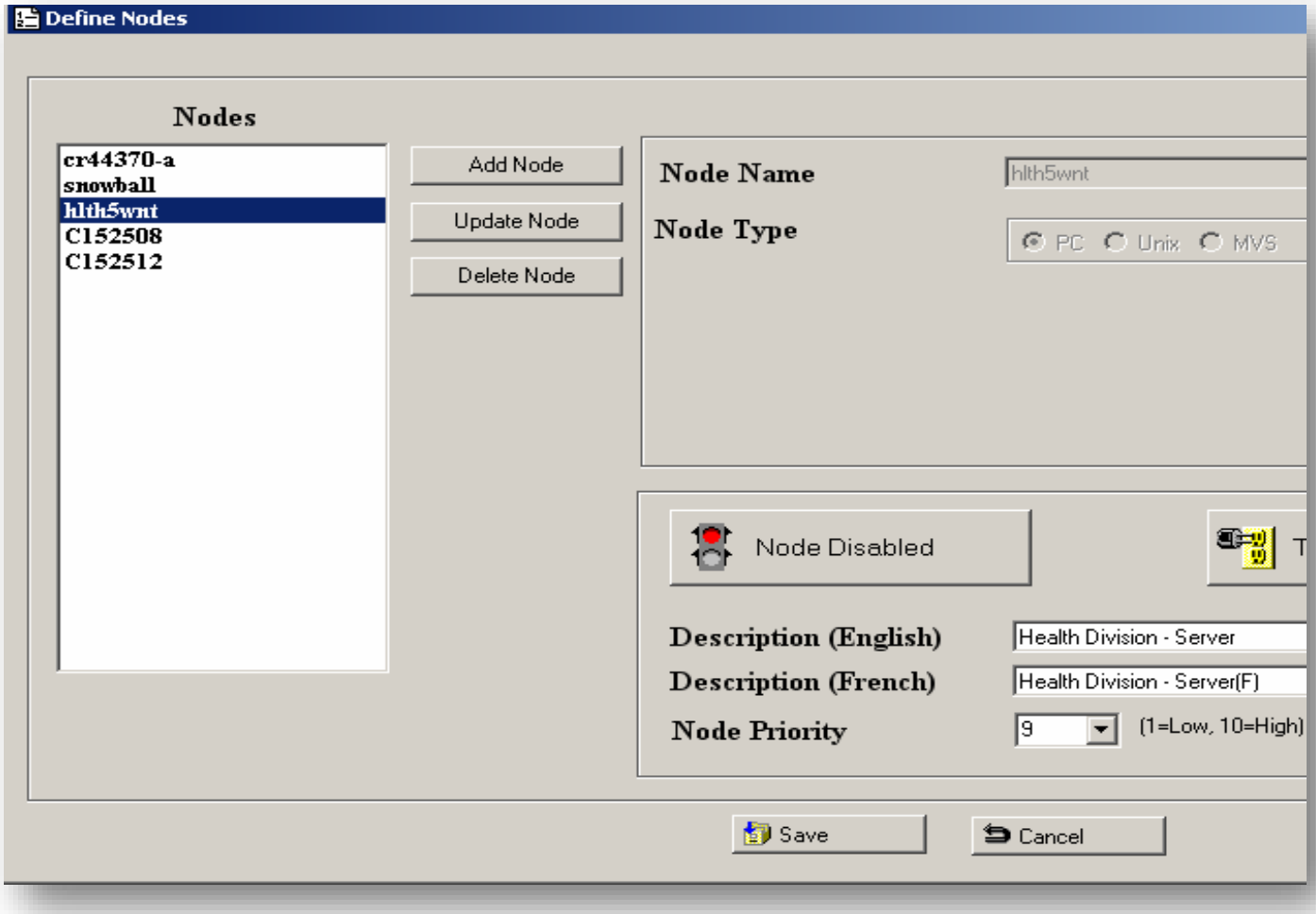
Case Study
Overview



Case Study
Implementation

Nodes

The Grid controller software must know about all nodes that are considered connected to the grid.



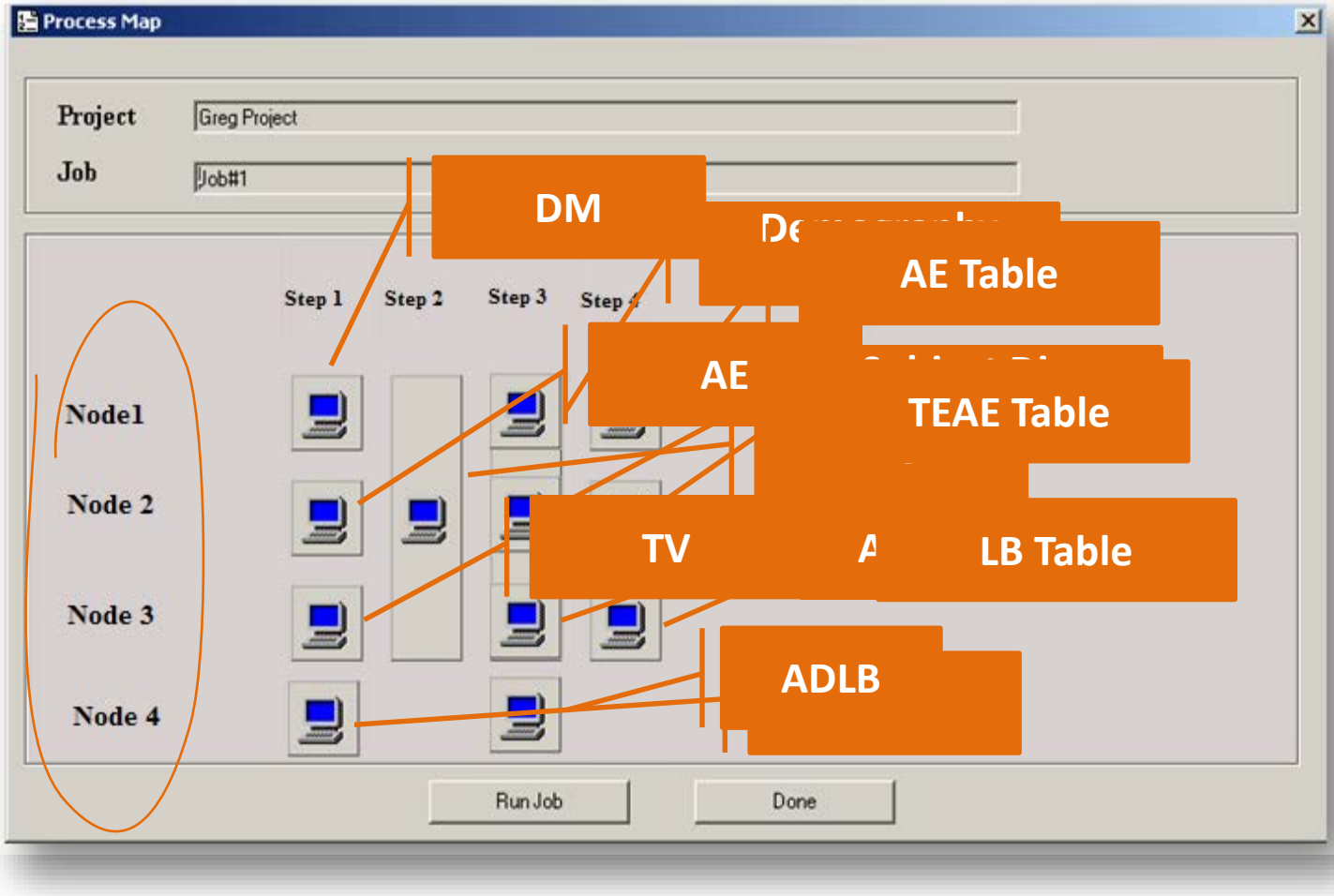
Therefore at run-time, the Grid controller software can match more complex sub-tasks to the more complex nodes (optimization issue).

Process Flow

- Grid computing is to break down tasks into sub-tasks.
- Subtasks can be processed in parallel fashion, thus reducing the overall elapsed processing time.
- In the SAS/AF prototype, the user has the ability to create a visual representation of the sub-tasks, including the process flow (icons).
- Once the visual representation is complete, the user then has the ability to define a separate SAS program for each visual object.
- It is then necessary for the user to identify ways to split the main SAS program.

Case Study
Implementation

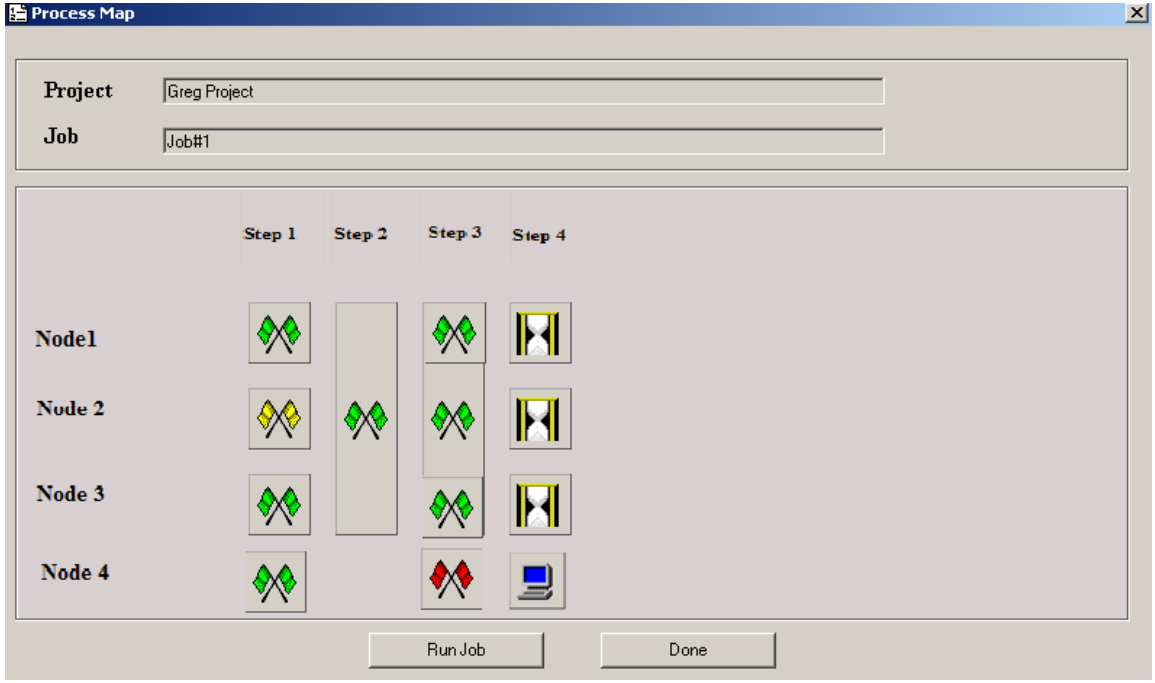
- Sub-tasks in any given column can be run at the same time or in parallel.
- Sub-tasks in a given column cannot run until their predecessor sub-tasks (the sub-task(s) in the column to the left) have completed.



Case Study
Implementation

Job Submission

The users can see the status of any given sub-task in real-time.



Completed successfully

Completed with warnings

Not completed due to errors

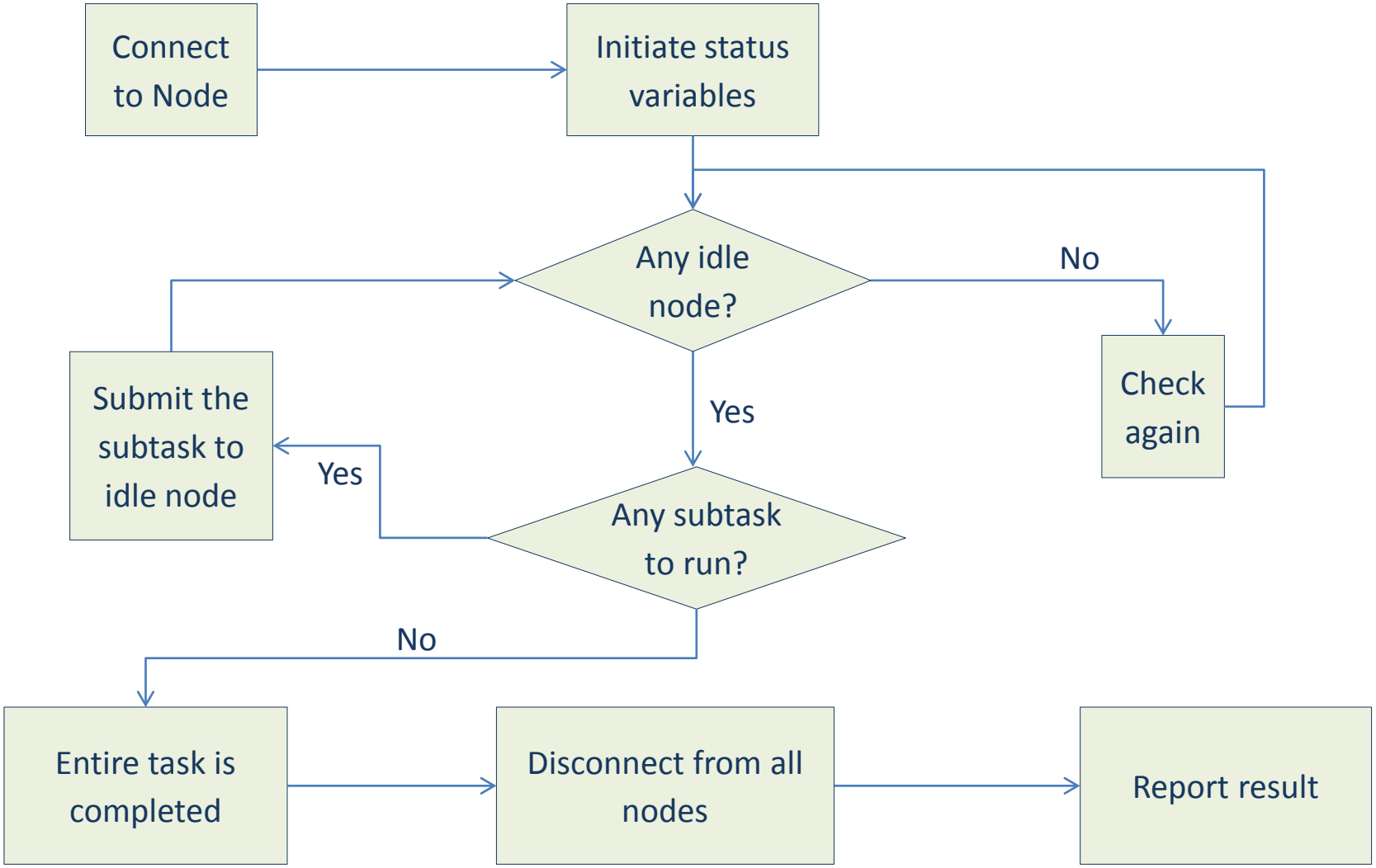
Currently running on node

Waiting to be run

Job Summary

Project	Job	Submission Number	Status	Elapsed Time	Start Time	End Time	Number of Nodes	Number of Tasks	Submitted by
1	1	7	E	0:00:02	2:29:51 PM	2:29:54 PM	3	12	mclegry

Case Study
Backend Algorithm



Parallel Processing By SAS Program

Normal SAS Program

```
/*mmmmmmmmmmmm
     SDTM
mmmmmmmmmmmmmm*/

%Include "\\Server1\Study_Name\Pgm_SDTM\DM";

%Include "\\Server1\Study_Name\Pgm_SDTM\AE";

%Include "\\Server1\Study_Name\Pgm_SDTM\LB";

%Include "\\Server1\Study_Name\Pgm_SDTM\TV";


/*mmmmmmmmmmmm
     ADAM
mmmmmmmmmmmmmm*/

%Include "\\Server1\Study_Name\Pgm_ADAM\ADSL";

%Include "\\Server1\Study_Name\Pgm_ADAM\ADAE";

%Include "\\Server1\Study_Name\Pgm_ADAM\ADLB";


/*mmmmmmmmmmmmmm
     Tables
mmmmmmmmmmmmmmmm*/

%Include "\\Server1\Study_Name\Pgm_TLF\Tables\T_DM";

%Include "\\Server1\Study_Name\Pgm_TLF\Tables\T_DS";

%Include "\\Server1\Study_Name\Pgm_TLF\Tables\T_AE";

%Include "\\Server1\Study_Name\Pgm_TLF\Tables\T_TEAE";

%Include "\\Server1\Study_Name\Pgm_TLF\Tables\T_LB";
```

Parallel Processing By SAS Program

Modified SAS Program

```
/*mmmmmmmmmmmmmmm
     Step 1
mmmmmmmmmmmmmmmmm*/

Option Autosignon SASCMD = "!SASCMD";
Rsubmit Step_1 Wait = No;

%Include "\\Server1\Study_Name\Pgm_SDTM\DM";

Endrsubmit;


Option Autosignon SASCMD = "!SASCMD";
Rsubmit Step_2 Wait = No;

%Include "\\Server1\Study_Name\Pgm_SDTM\AE";

Endrsubmit;

Option Autosignon SASCMD = "!SASCMD";
Rsubmit Step_3 Wait = No;

%Include "\\Server1\Study_Name\Pgm_SDTM\LB";

Endrsubmit;

Option Autosignon SASCMD = "!SASCMD";
Rsubmit Step_4 Wait = No;

%Include "\\Server1\Study_Name\Pgm_SDTM\TV";

Endrsubmit;
```

Parallel Processing By SAS Program

```
/*mmmmmmmmmmmmmmm
     Step 1
mmmmmmmmmmmmmmmmm*/

Option Autosignon SASCMD = "!SASCMD";
Rsubmit Step_1 Wait = No;

%Include "\\Server1\Study_Name\Pgm_SDTM\DM";

Endrsubmit;


Option Autosignon SASCMD = "!SASCMD";
Rsubmit Step_2 Wait = No;

%Include "\\Server1\Study_Name\Pgm_SDTM\AE";

Endrsubmit;

Option Autosignon SASCMD = "!SASCMD";
Rsubmit Step_3 Wait = No;

%Include "\\Server1\Study_Name\Pgm_SDTM\LB";

Endrsubmit;

Option Autosignon SASCMD = "!SASCMD";
Rsubmit Step_4 Wait = No;

%Include "\\Server1\Study_Name\Pgm_SDTM\TV";

Endrsubmit;
```

```
/*mmmmmmmmmmmmmmm
    Step 2
mmmmmmmmmmmmmmmmm*/

Option Autosignon SASCMD = "!SASCMD";
Rsubmit Step_5 Waitfor Step_1;

%Include "\\Server1\Study_Name\Pgm_ADAM\ADSL";

Endrsubmit;
```

Parallel Processing By SAS Program

```
/*mmmmmmmmmmmmmm
    Step 2
mmmmmmmmmmmmmmmm*/

Option Autosignon SASCMD = "!SASCMD";
Rsubmit Step_5 Waitfor Step_1;

%Include "\\Server1\Study_Name\Pgm_ADAM\ADSL";

Endrsubmit;
```

```
/*mmmmmmmmmmmmmm
    Step 3
mmmmmmmmmmmmmmmm*/

Option Autosignon SASCMD = "!SASCMD";
Rsubmit Step_6 Waitfor Step_2 Step_4 Step_5;

%Include "\\Server1\Study_Name\Pgm_ADAM\ADAE";

Endrsubmit;

Option Autosignon SASCMD = "!SASCMD";
Rsubmit Step_7 Waitfor Step_3 Step_4 Step_5;

%Include "\\Server1\Study_Name\Pgm_ADAM\ADLB";

Endrsubmit;

Option Autosignon SASCMD = "!SASCMD";
Rsubmit Step_8 Waitfor Step_5;

%Include "\\Server1\Study_Name\Pgm_TLF\T_DM";

Endrsubmit;

Option Autosignon SASCMD = "!SASCMD";
Rsubmit Step_9 Waitfor Step_5;

%Include "\\Server1\Study_Name\Pgm_TLF\T_DS";

Endrsubmit;
```

Parallel Processing By SAS Program

```
/*mmmmmmmmmmmmmmm
    Step 3
mmmmmmmmmmmmmmmmm*/

Option Autosignon SASCMD = "!SASCMD";
Rsubmit Step_6 Waitfor Step_2 Step_4 Step_5;

%Include "\\Server1\Study_Name\Pgm_ADAM\ADAE";

Endrsubmit;

Option Autosignon SASCMD = "!SASCMD";
Rsubmit Step_7 Waitfor Step_3 Step_4 Step_5;

%Include "\\Server1\Study_Name\Pgm_ADAM\ADLB";

Endrsubmit;

Option Autosignon SASCMD = "!SASCMD";
Rsubmit Step_8 Waitfor Step_5;

%Include "\\Server1\Study_Name\Pgm_TLF\T_DM";

Endrsubmit;

Option Autosignon SASCMD = "!SASCMD";
Rsubmit Step_9 Waitfor Step_5;

%Include "\\Server1\Study_Name\Pgm_TLF\T_DS";

Endrsubmit;
```

```
/*mmmmmmmmmmmmmmm
    Step 4
mmmmmmmmmmmmmmmmm*/

Option Autosignon SASCMD = "!SASCMD";
Rsubmit Step_10 Waitfor Step_6;

%Include "\\Server1\Study_Name\Pgm_TLF\T_AE";

Endrsubmit;

Option Autosignon SASCMD = "!SASCMD";
Rsubmit Step_11 Waitfor Step_6;

%Include "\\Server1\Study_Name\Pgm_TLF\T_TEAE";

Endrsubmit;

Option Autosignon SASCMD = "!SASCMD";
Rsubmit Step_12 Waitfor Step_7;

%Include "\\Server1\Study_Name\Pgm_TLF\T_LB";

Endrsubmit;
```

Merits & Demerits

Merits	Demerits
Avoid purchasing of large six figure SMP servers.	Memory hungry applications that can't take advantage of MPI you may be forced to run on a large SMP.
Efficient usage of idle resources.	Fast interconnect between compute resources (gigabit Ethernet at a minimum).
No failure in job submission.	Some applications may need to be tweaked to take full advantage of the new model.
Satisfy all the policies.	Licensing across many servers may make it prohibitive for some apps.
Upgradation without scheduled downtime.	Good tools for managing change and keeping configurations in sync with each other can be challenging
Parallel Processing can be easier.	Political challenges associated with sharing resources (especially across different admin domains). Many groups are reluctant with sharing resources even if it benefits everyone involved.

References

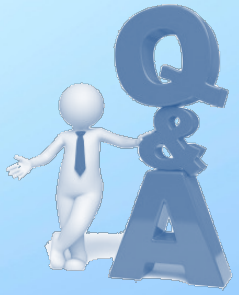
http://www2.sas.com/proceedings/sugi31/212-31.pdf

http://support.sas.com/documentation/cdl/en/gridref/67371/PDF/default/gridref.pdf

Cytel

Shaping the Future of
Drug Development

Thank You !!



Contact

Milan Adesara
Senior Statistical Programmer
Milan.adesara@cytel.com
+91 990.497.8778