

# Integrated Genomics Applications Pipeline (iGAP) on the GRID - Software Review

Atif Shahab

Bioinformatics Institute

Singapore

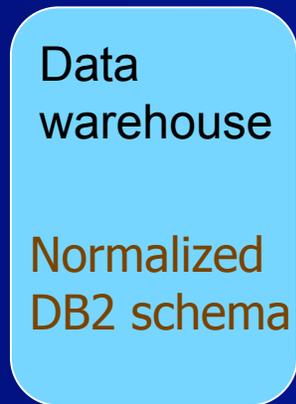
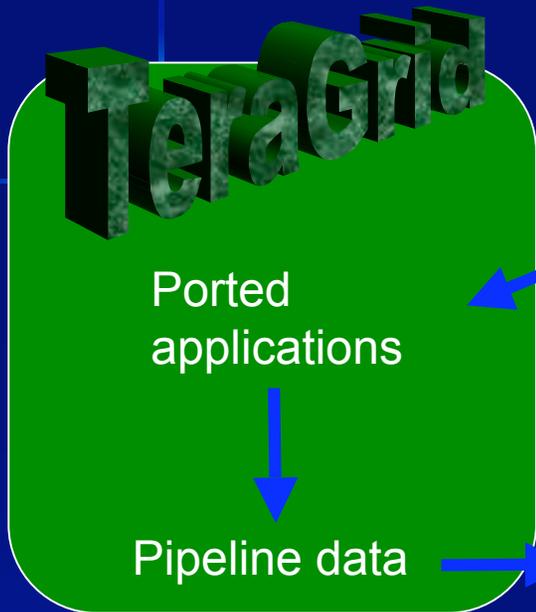


# Agenda

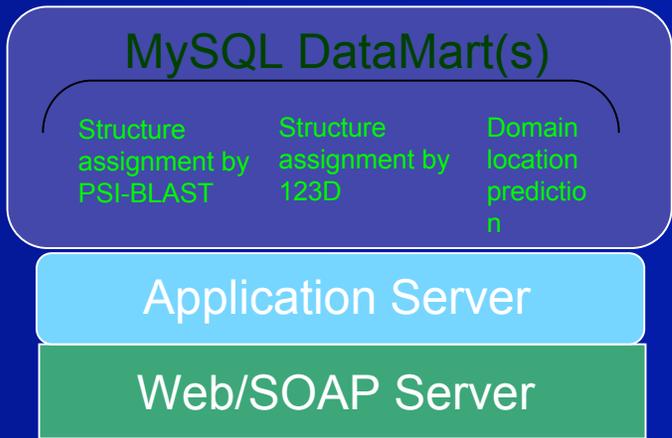
- Objective
- Overview
- Current Status
- Process Flow
- Meta Scheduling
- Future Development



Sequence data from genomic sequencing projects



Load/update scripts

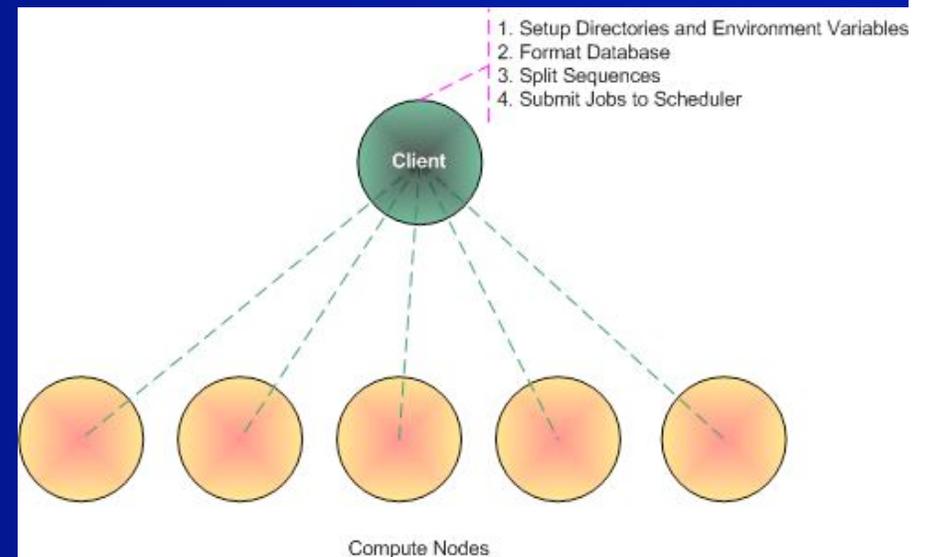


Retrieve Web pages & Invoke SOAP methods

Some Technical Detail Mapped to the Topology

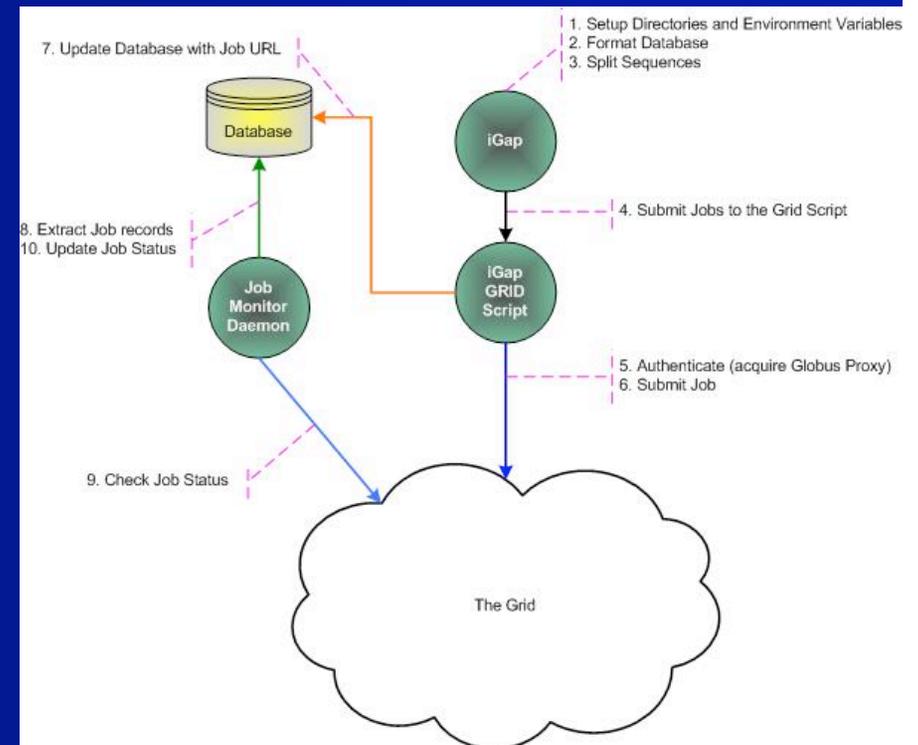
# Overview

- Cluster based
- Processes tightly coupled with each other
- Lack of job cum workflow management



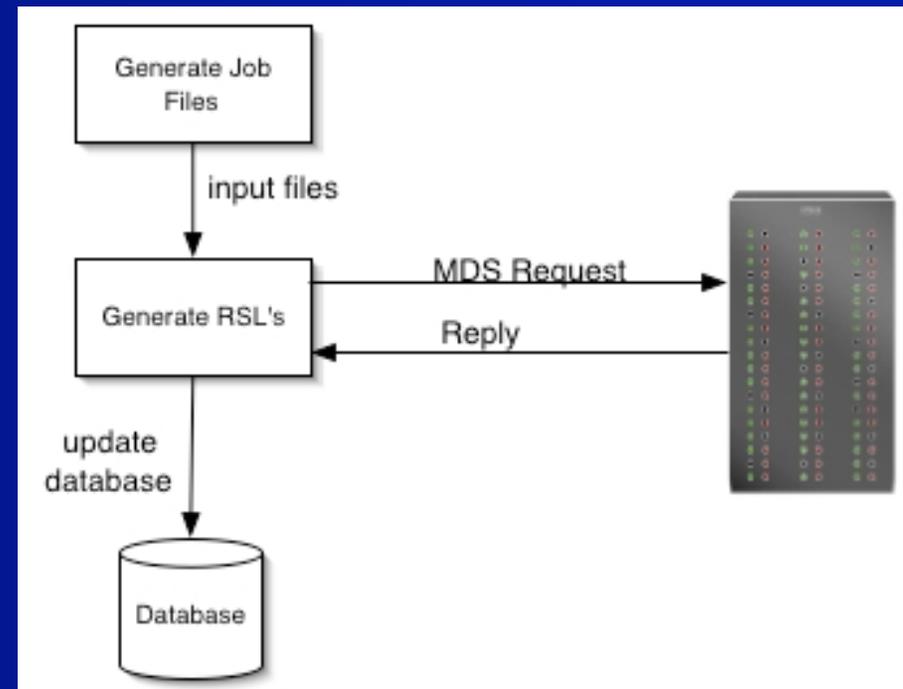
# Current Status

- Extended beyond cluster to incorporate the grid.
- Decoupled various application processes.
- Job monitoring (grid)
- Automatic certificate renewal
- File management



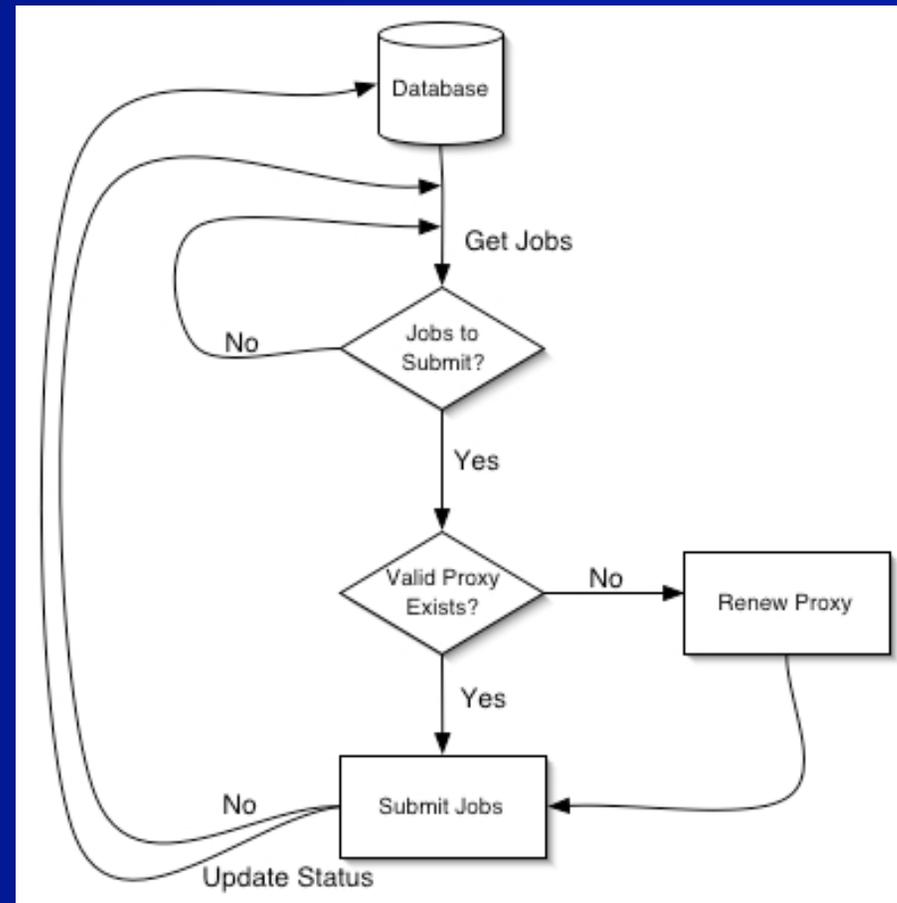
# Process Flow - Generation

- Generates normal job files.
- Queries MDS servers for scratch space path (GLUE schema)
- Generates appropriate RSL
- Updates the database.



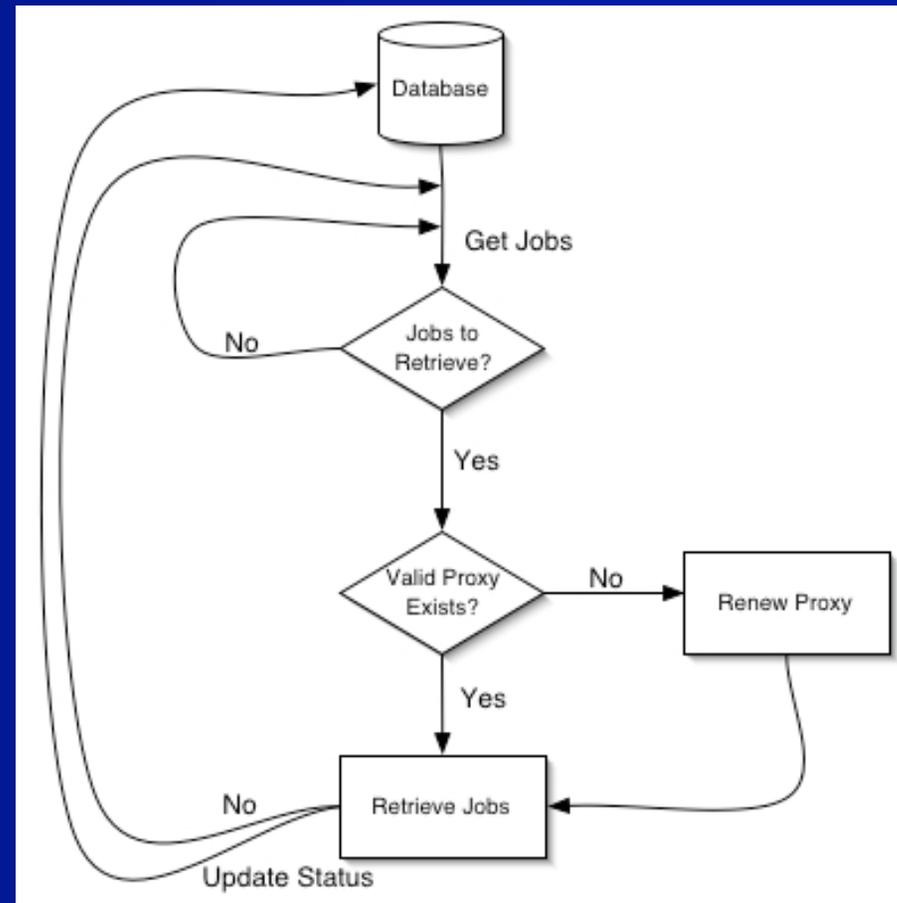
# Process Flow - Submission

- Obtain lists of jobs ready for submission
- Revalidate the proxy if expired or expiring
- Create necessary scratch space (/scratch/BII/pdb/atif/20030215/)
- Ftp necessary files and submit the RSL/job

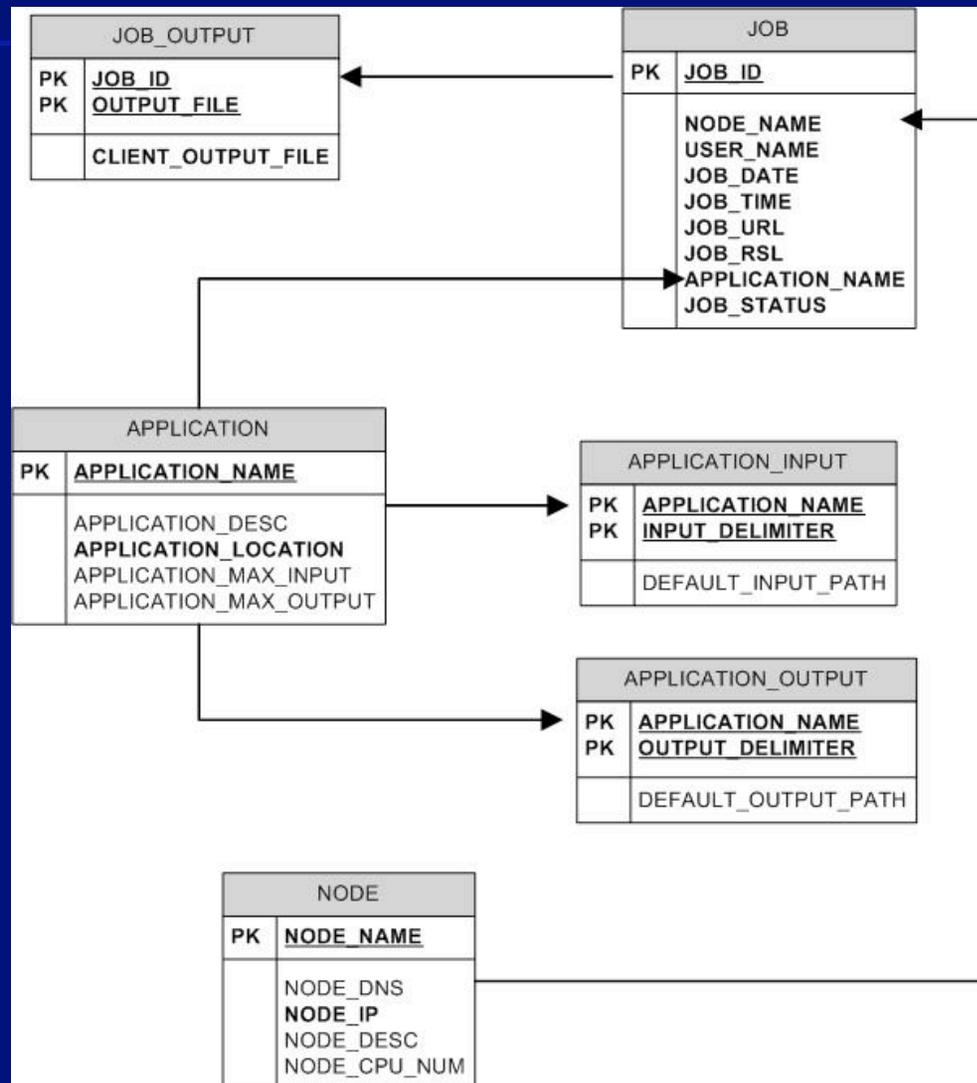


# Process Flow - Retrieval

- Obtain list of submitted jobs
- Revalidate the proxy if expired or expiring
- Retrieve required files and store locally
- Update database(!)

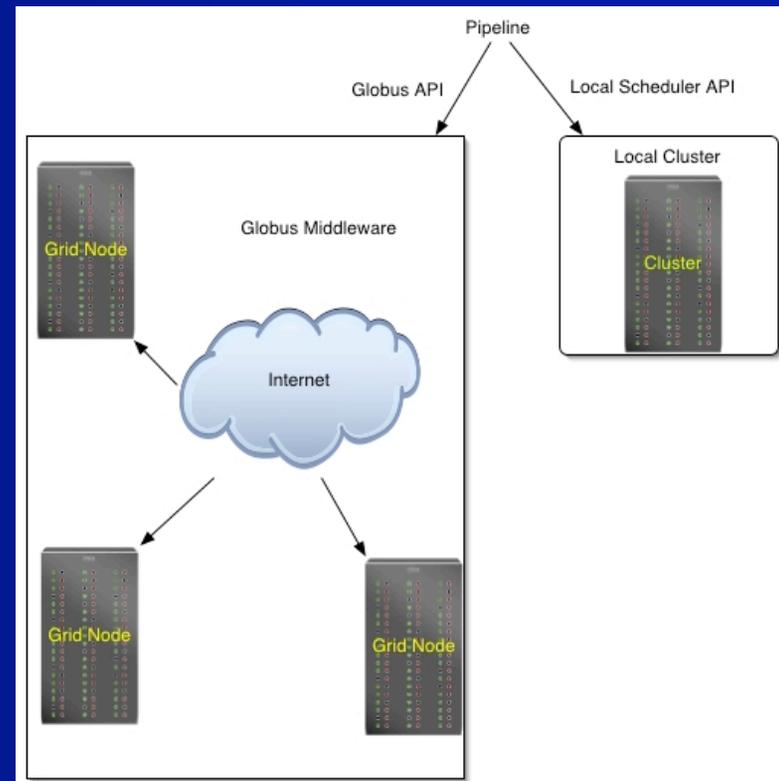


# Process Flow - Database



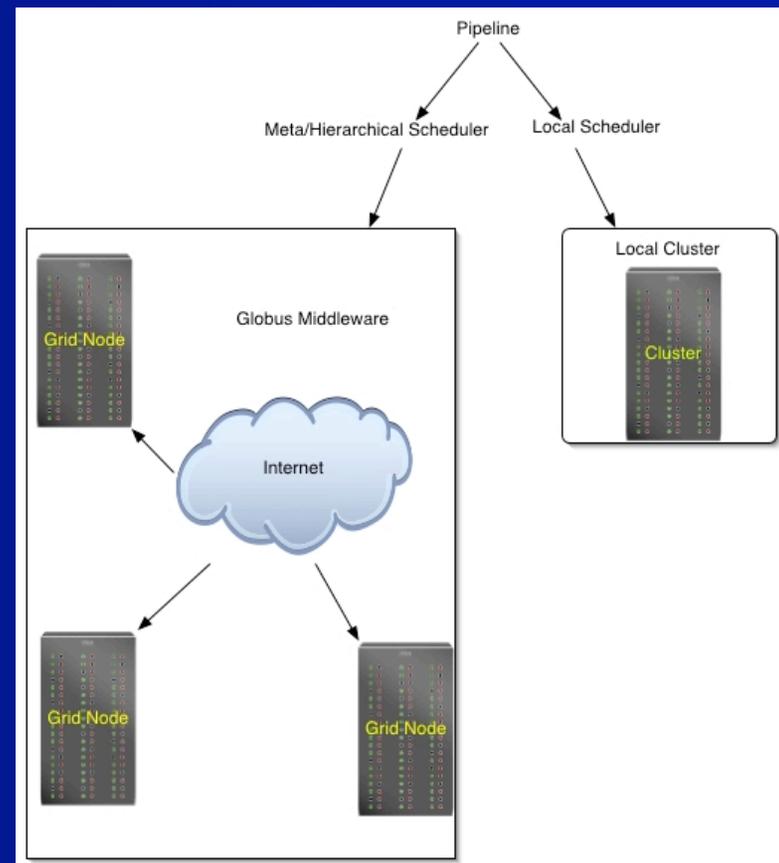
# Meta Schedulers - 1

- Assumes absence of meta schedulers.
- Implements rudimentary scheduler within the pipeline.
- Manages jobs and associated file transfers.
- Uses MDS for resource discovery and RSL creation.



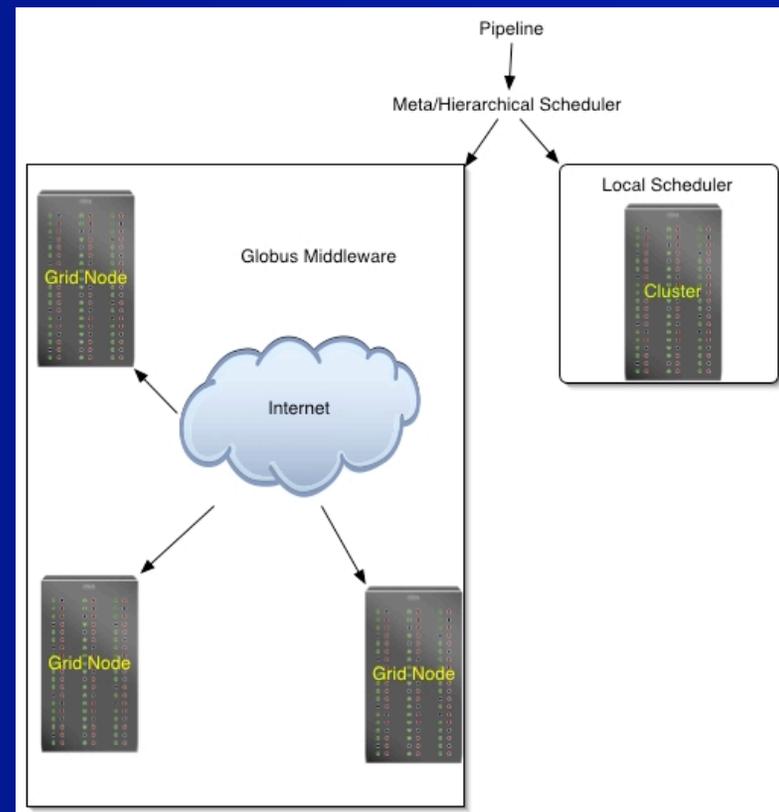
# Meta Schedulers - 2

- Modularise scheduling
- Interface with meta-schedulers
- Still fair bit of job management done by the pipeline
- APST, GRIDX, SILVER



# Meta Scheduler - 3

- Total abstraction of scheduling from the pipeline
- Pipeline's main focus becomes work flow management
- APST, GRIDX, SILVER, SGE



# Future Developments

- Further decoupling of the stages
- Integration with meta schedulers
- Explore web services
- Focus on Workflow
- Web Management Tools
- Unit Tests

# Acknowledgements

- BII

- Larry Ang
- Anand Sairam
- Yong How Choong

- SDSC

- Wilfred Li
- Mason J. Katz