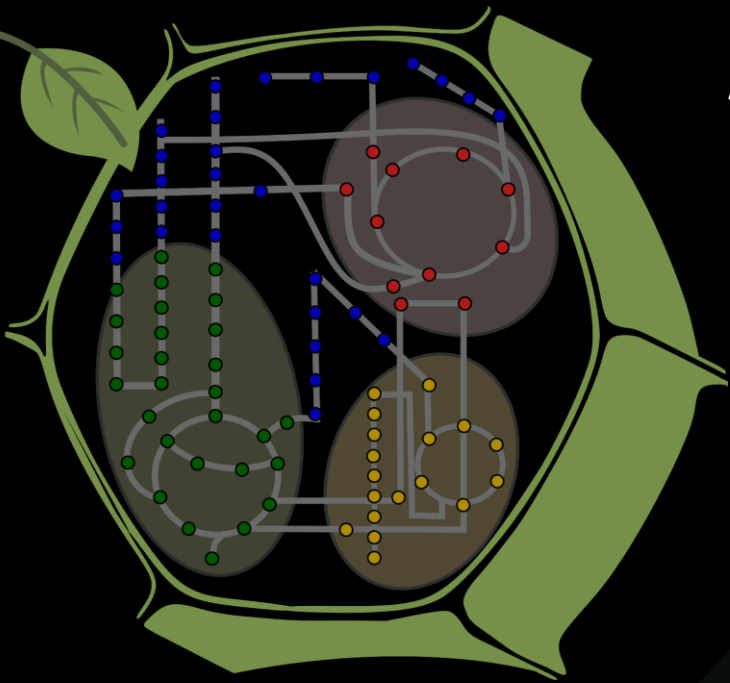
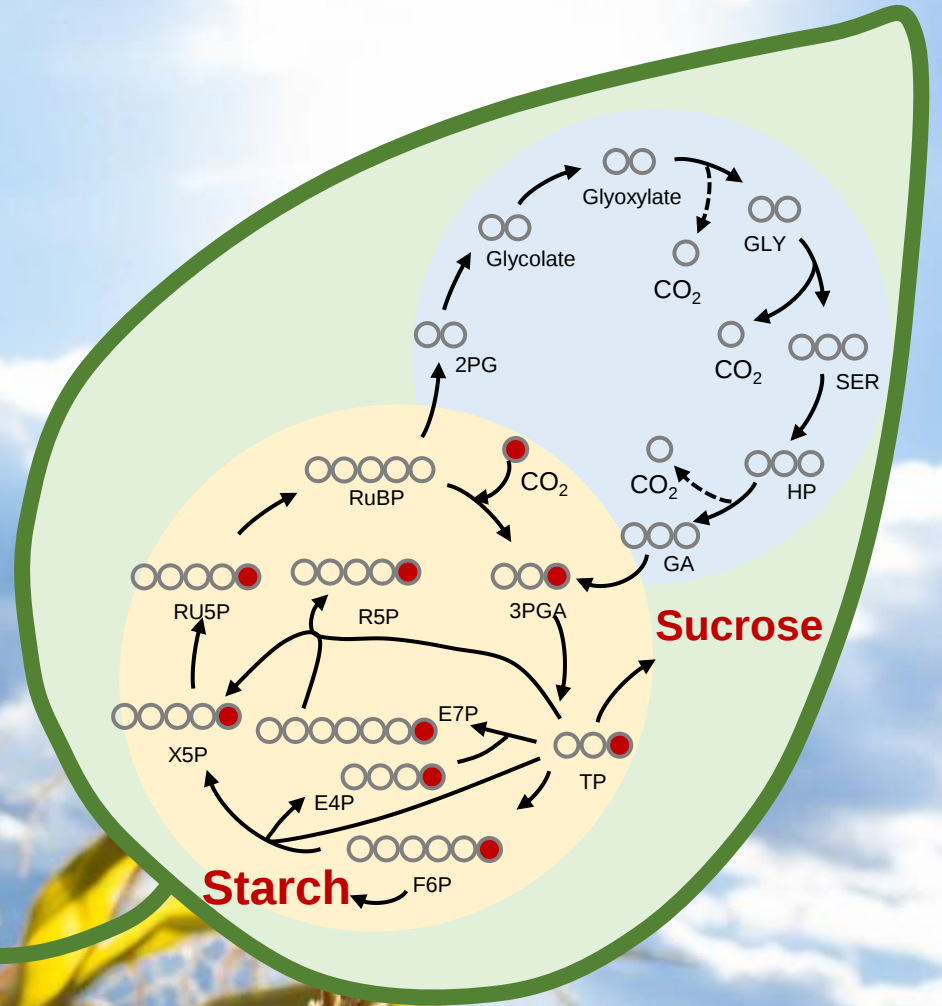




A Cloud-based Workflow to Run Parallel Simulations for Metabolic Flux Analysis

Presenter: Xinyu Fu

**Berkley Walker's Lab
MSU-DOE Plant Research Laboratory**



Metabolic flux analysis (MFA) is a powerful approach for quantifying intracellular metabolic reaction rates in plant metabolic networks and fundamental to metabolic engineering and improving crop fitness.

Project Aim

Metabolic Model

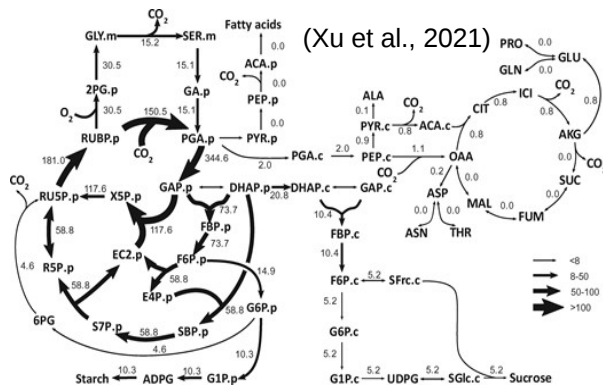


Experimental Data

Parameter
Fitting

MATLAB
INCA

Flux Map



(Xu et al., 2021)

Metabolic flux analysis (MFA) is compute-intensive
non-linear regression with Monte Carlo Bootstrap

HPC solution

1000 Simulations/ Model
1-5 cpu hours / Simulation
Thousands of jobs on HPC



- Require maintenance
- Variable queue wait times
- CPU Usage Limits
- Not easy to share data outside

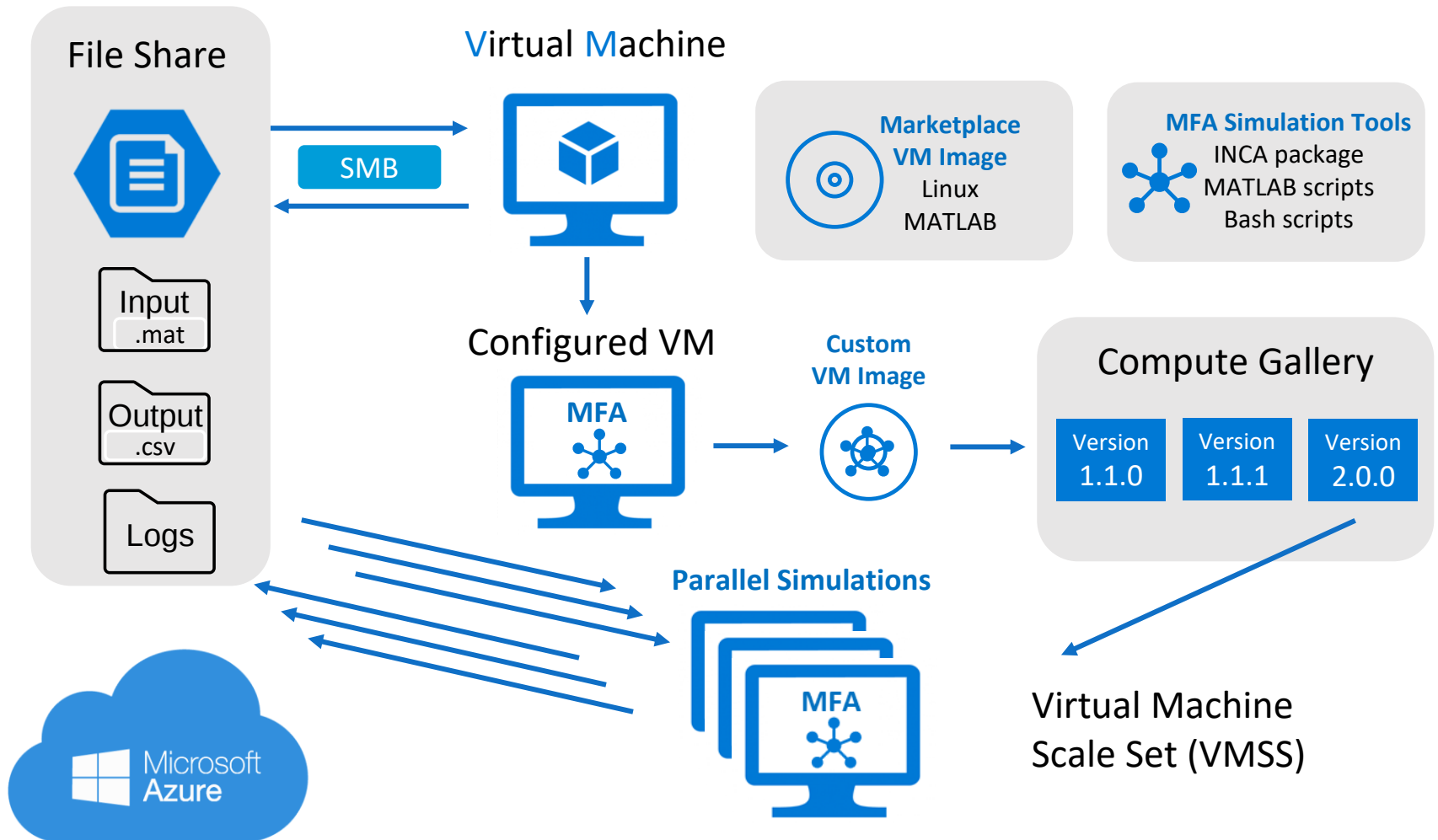


Cloud solution?

Explore the scalable
computing capacity for
parallel simulations



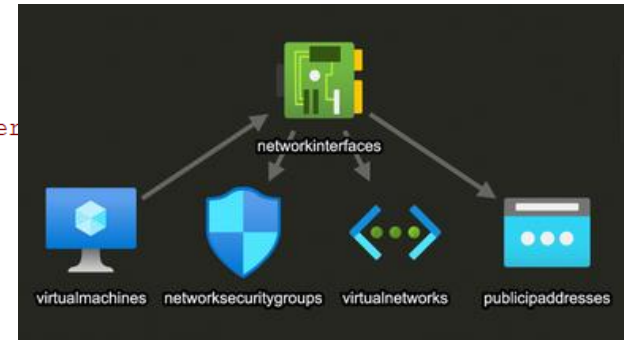
Cloud-base Workflow



Successes and Challenges of the Project

```
133 {
134   "type": "Microsoft.Compute/virtualMachines",
135   "apiVersion": "2021-07-01",
136   "name": "[parameters('virtualMachineName')]",
137   "location": "[parameters('location')]",
138   "plan": {
139     "name": "[variables('sku')]",
140     "publisher": "[variables('publisher')]",
141     "product": "[variables('offer')]"
142   },
143   "dependsOn": [
144     "[concat('Microsoft.Network/networkInterfaces/', parameters('networkInterfaceName'))]"
145   ],
146   "properties": {
147     "hardwareProfile": {
148       "vmSize": "[parameters('virtualMachineSize')]"
149     },
150     "storageProfile": {
151       "osDisk": {
152         "createOption": "fromImage",
153         "managedDisk": {
154           "storageAccountType": "[parameters('osDiskType')]"
155         }
156       },
157       "imageReference": {
158         "id": "/subscriptions/f58c4ece-3049-4806-a791-2cab500bd937/resourceGroups/rg-matlab/"
159       }
160     },
161     "networkProfile": {
162       "networkInterfaces": [
163         {
164           "id": "[resourceId('Microsoft.Network/networkInterfaces', parameters('networkInterfaceName'))]"
165         }
166       ]
167     },
168     "osProfile": {
```

JSON-based Azure Resource Manager (ARM) template for automatic deployment of cloud resources



Successes and Challenges of the Project

Azure Command-Line Interface (CLI)

```
az vm create \
  --resource-group $resourceGroup \
  --name matlabVM \
  --size Standard_F4s_v2 \
  --storage-sku Standard_LRS \
  --admin-username matlabuser \
  --generate-ssh-keys \
  --image mathworks-inc:matlab-ref-arch-18a-v1-linux-disk:matlab-ref-arch-20a-vm-linux-sku:latest \
  --plan-name matlab-ref-arch-20a-vm-linux-sku \
  --plan-product matlab-ref-arch-18a-v1-linux-disk \
  --plan-publisher mathworks-inc
```

Must supply purchase plan information for Marketplace-based images via ARM or CLI

MATLAB licensing credential expires after 2 days

```
az vmss create \
  --resource-group $resourceGroup \
  --name $vmssName \
  --image $image \
  --instance-count $instanceCount \
  --vm-sku $vmSize \
  --nsg $nsgName \
  --subnet $subnetName \
  --vnet-name $vnetName \
  --load-balancer $lbName \
  --upgrade-policy-mode $upgradePolicyMode \
  --admin-username $adminName \
  --generate-ssh-keys \
  --specialized \
  --plan-name $planName \
  --plan-product $planProduct \
  --plan-publisher $planPublisher
```

CLI documentation is not detailed enough

Comparison of VM sizes

Running a single simulation using different VM sizes

VM Size	vCPUs	RAM	Max IOPS	Time	Monthly Cost
B1ms	1	2	640	1.1	\$15
B2s	2	4	1280	1.0	\$30
B4ms	4	16	2880	1.1	\$121
A1_v2	1	2	2x500	2.9	\$31
DS1_v2	1	3.5	3200	1.2	\$47

- B1ms is the most cost-effective VM size for MFA simulations
- A typical plant model may require 2000 CPU hours and cost ~ \$40

Takeaways

- **On-demand scalability**

VM scale sets let you create and manage thousands of VMs based on custom image, great for parallel simulation tasks

- **Easy and manageable data sharing**

Direct mount of File Shares to authorized computers via SMB/NFS

- **Reproducibility**

Infrastructure for MFA can be templated (ARM) and scripted (CLI)

- **Cost**

Testing different options to optimize cost and monitoring bills often

- **Future use cases of the cloud-based MFA workflow**

- Allows researchers with limited/no HPC access to do and learn MFA
- Share intermediate data with the collaborators for them to decide whether to terminate the current analysis and improve the model.

Acknowledgements

- MSU's Institute for Cyber-Enabled Research (ICER) and the ITS Analytics and Data Solutions (ADS) group and organizers (Brian O'Shea, Danielle Barnes, Chelsea Beck, Starr Portice) for the fellowship
- Pat Bills and Mahmoud Parvizi for guidance and support
- My fellows for inspirations

