Microsoft

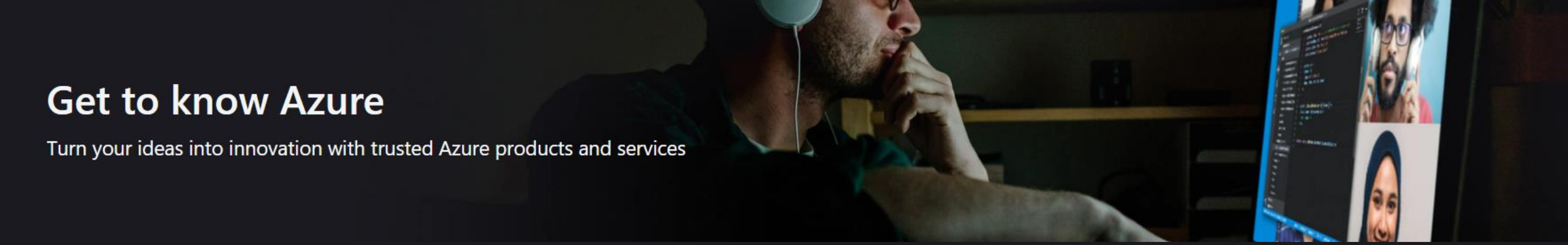
Capacity Planning for Microsoft Azure Datacenters Using R & RStudio Connect

Paul Chang, Sarah George, Ivy Huang, Yayun Jin, Riti Kumari, Sajay Suresh, Serge Sverdlov
Amit Gandhi (Airbnb), Jingyi Li, Yantao Wang

Agenda

- Capacity planning process for Microsoft Azure Datacenters
- How we use RStudio Connect and Microsoft Azure
- Datacenter Capacity Dashboard Demo

Capacity planning process for Microsoft Azure Datacenters



Get to know Azure

Turn your ideas into innovation with trusted Azure products and services

Azure is Microsoft's Cloud Platform

200
Products and Cloud Services

1+ billion
Customers Served

20 million
Companies on Azure

60+
Regions Served Worldwide

90+
Regulatory Compliance Offerings

95%
Fortune 500 Companies On Azure

$1 billion
Investment in cybersecurity per year

Azure global infrastructure

60+ regions, more than any other cloud provider

Explore the globe

Global Infrastructure Components

165,000 miles
Fiber-optic Network

200+
Physical Datacenters

A lot of
Land

Azure Datacenter Key Features

High Availability

Low Latency

Scalability

Latest Cloud Technologies

Your Data is Safe and Secure!

Data stays entirely in Microsoft Network

IP traffic never enters public internet

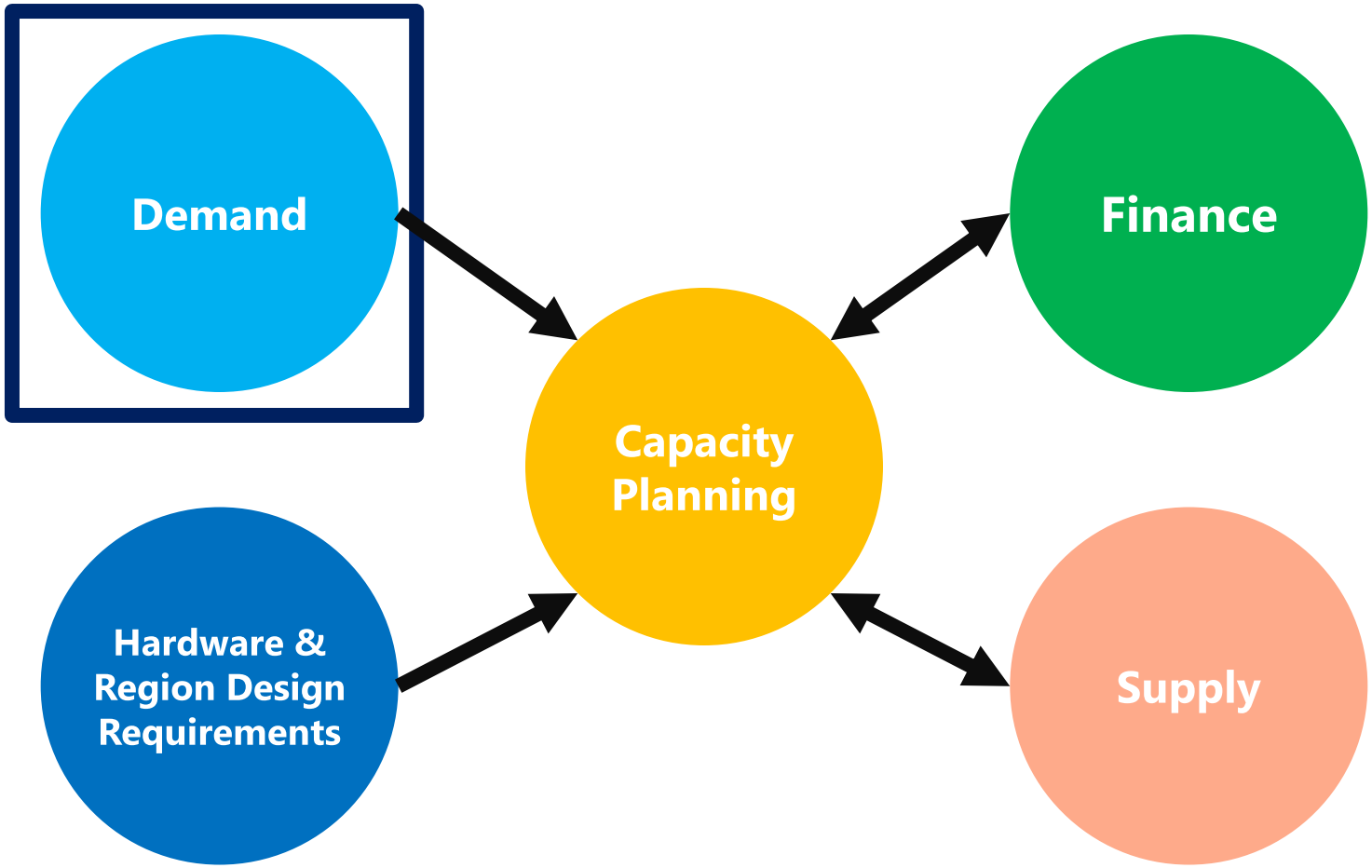
Capacity Planning: Overview

- Group of Economists and Data Scientists

- Group of Program Managers

- Long range capacity plan for Microsoft Azure Datacenters
  - 1 to 7 year range
  - Long lead times
  - $1B+ investments
  - Worldwide

Workflow Summary



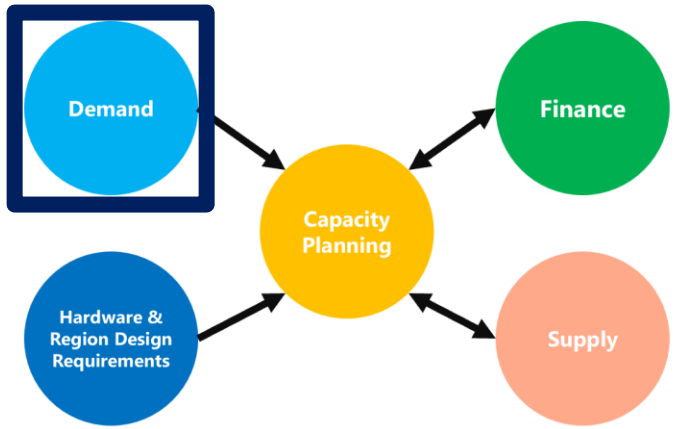
Demand Inputs

MS Engineering
Bing, Office 365, XBox

Organic Data
Economic Data
Azure Revenue

Large Customers
Governments
Large Enterprises

Sales Targets



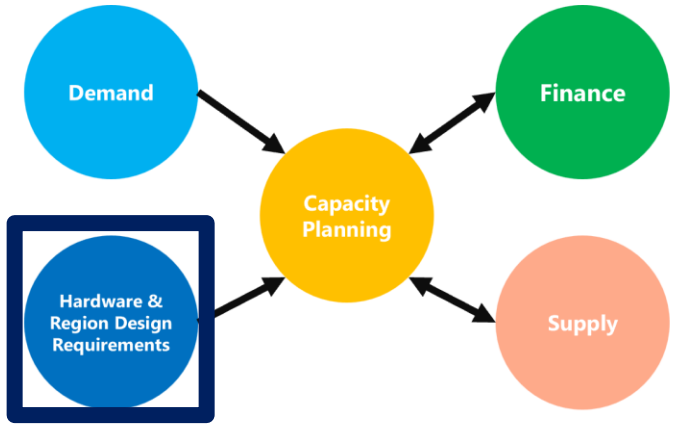
Hardware and Region Design Requirements

Region Design
Availability Zones
Disaster Recovery

Hardware Roadmap
CPU/GPU Generations

Networking

Azure Services
Active Directory
DNS, Firewall, etc.



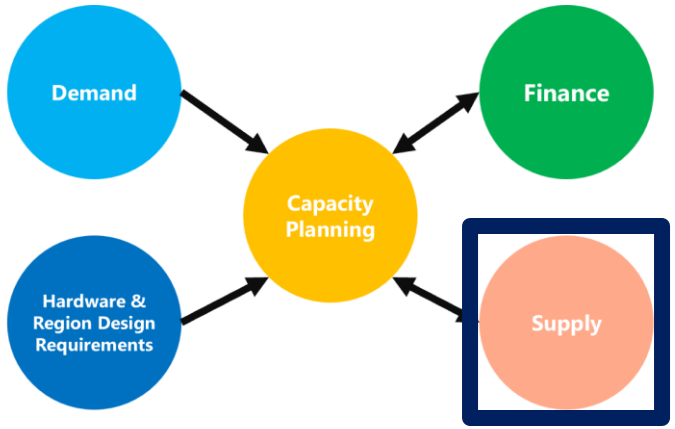
Supply Inputs and Outputs

Build Datacenters
Land, Energy

Lease Datacenters

Supply Mitigation

Supply Availability
Costs



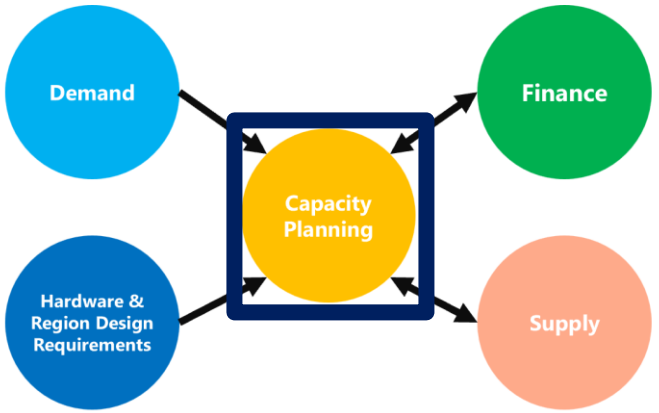
Capacity Planning (Our Team)

Revenue Forecasts
Tech Diffusion (Comin)
Hierarchical Time Series
Gravity Model of Trade

Hardware Forecasts
Server Power Efficiency
Server Cores per CPU

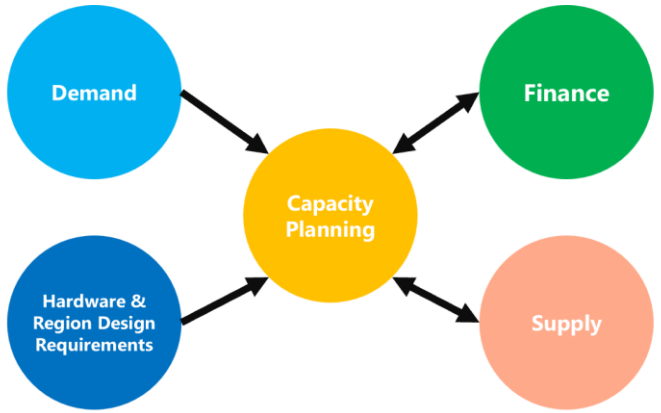
Capacity Principles
Capacity Tranche
Capacity Buffer
(Lead Time, Volatility)

Demand Shaping
Between Zones in Region
Between Regions in Geo
Costs/Supply Availability



Capacity Planning: Process Summary

- Demand
  - Organic, Large Customers, Microsoft Engineering Groups
- Hardware and Region Design Requirements
  - Server and networking requirements for Azure Services and Microsoft Engineering Groups
  - Location and Specifications for Datacenters
- Supply
  - Availability of supply
  - Lease vs Build
- Finance
  - Be cost effective



Capacity Planning: Top Considerations

Accuracy

- Model accuracy
- Empower humans to easily verify data

Capacity Planning: Top Considerations

Accuracy

- Model accuracy
- Empower humans to easily verify data

Timeliness

- Produce plans on a monthly basis
- Robust and agile infrastructure, data pipelines, and development cycle

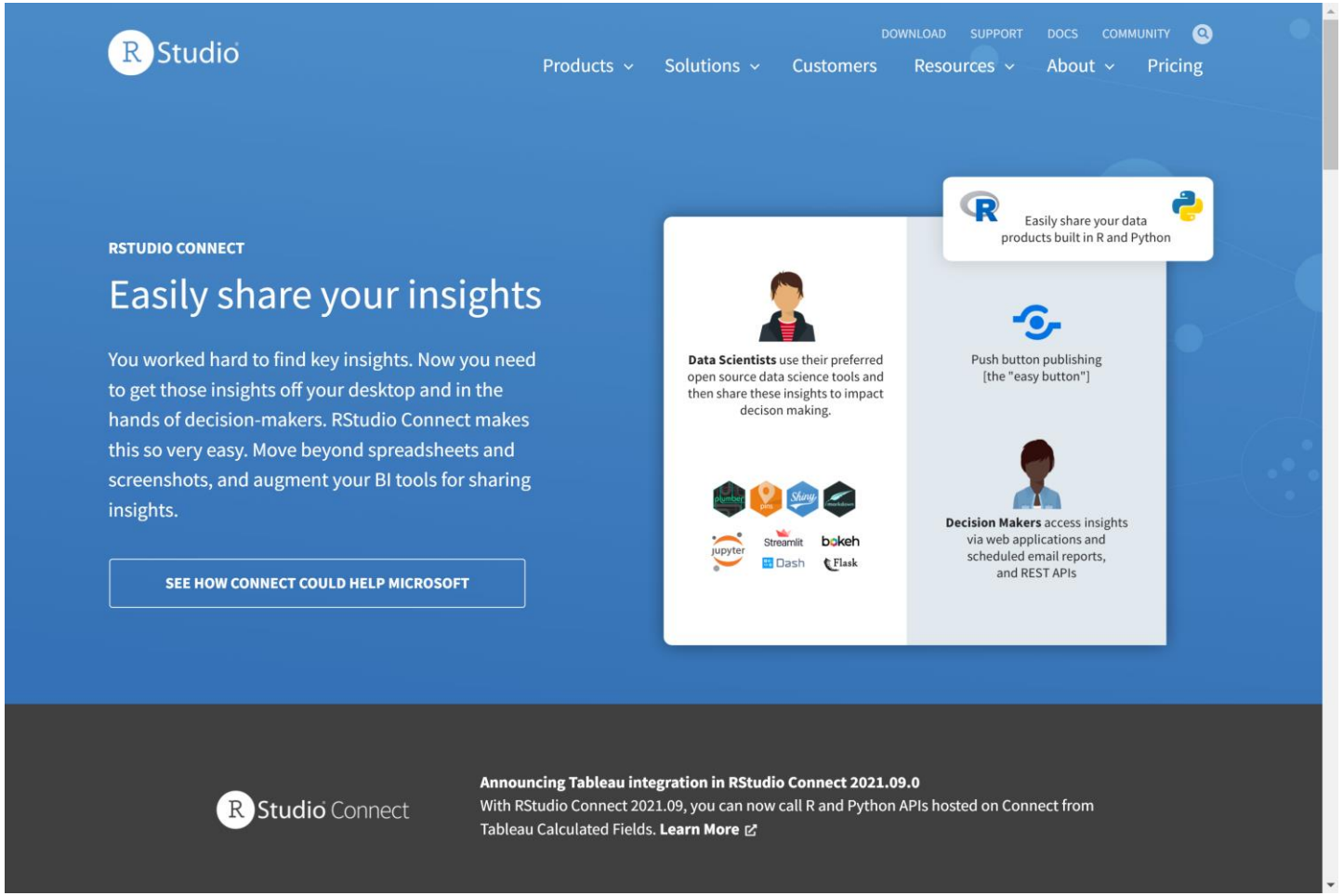
Human Intelligence and Business Strategy

- Add new inputs or override existing inputs
- Quickly model and incorporate human intelligence and business strategy

Explainability

- Clear and concise explanations to our partners and stakeholders
- Build trust

How we use RStudio Connect and Microsoft Azure



Before using RStudio Connect - Overview

- Models developed using R and C# by data scientists
- Models deployed in data pipelines by engineers using Databricks
  - Important to note most of our data is mid-sized (GBs, not TBs or PBs)
- Insights primarily shared via PowerPoint presentations
  - Some Power BI
  - Data shared with other teams via a database table
- Feedback from stakeholders
  - Incorporate feedback by modifying models

Before using RStudio Connect – Pain Points

- Slow End-to-End development time
  - ~3 days from Model Development to Stakeholder Feedback
  - Long delay to getting feedback on latest model changes
- Relatively slow computation time
  - Databricks is great for big data but relatively slow for mid-sized data
- Hard to do what-if scenarios
- Hard to create visualizations of results
  - Visualization is needed for both stakeholders and ourselves
  - Visualization is needed to help debug model code

RStudio Connect and Microsoft Azure – Overview

- Data scientists own entire model development lifecycle
  - Data Ingestion jobs
  - Model Development
  - Model Deployment
  - Visualization of Insights
- Engineers
  - Infrastructure (Data Storage, RStudio Connect, Active Directory)
  - Architecture
  - Data Compliance
  - Zero Trust Security

RStudio Connect and Microsoft Azure – Advantages

- Empower data scientists to directly communicate actionable insights to stakeholders
  - E2E Model Development Lifecycle
  - Visualization via Shiny apps
- Quick End-to-End development time
  - As little as 1 hour from Model Development to Stakeholder Feedback
- Easy collaboration between Data Scientists
  - Sharing datasets via pins
  - Sharing code via R packages
- Easy to do what-if scenarios
  - Data shared with stakeholders via Plumber APIs
  - Plan of Record still stored in database table

RStudio Connect: Return on Investment

On-Time Delivery of Monthly Capacity Plan

- Consistent on-time publication of capacity plan
- Prevent delays in downstream stages of datacenter development

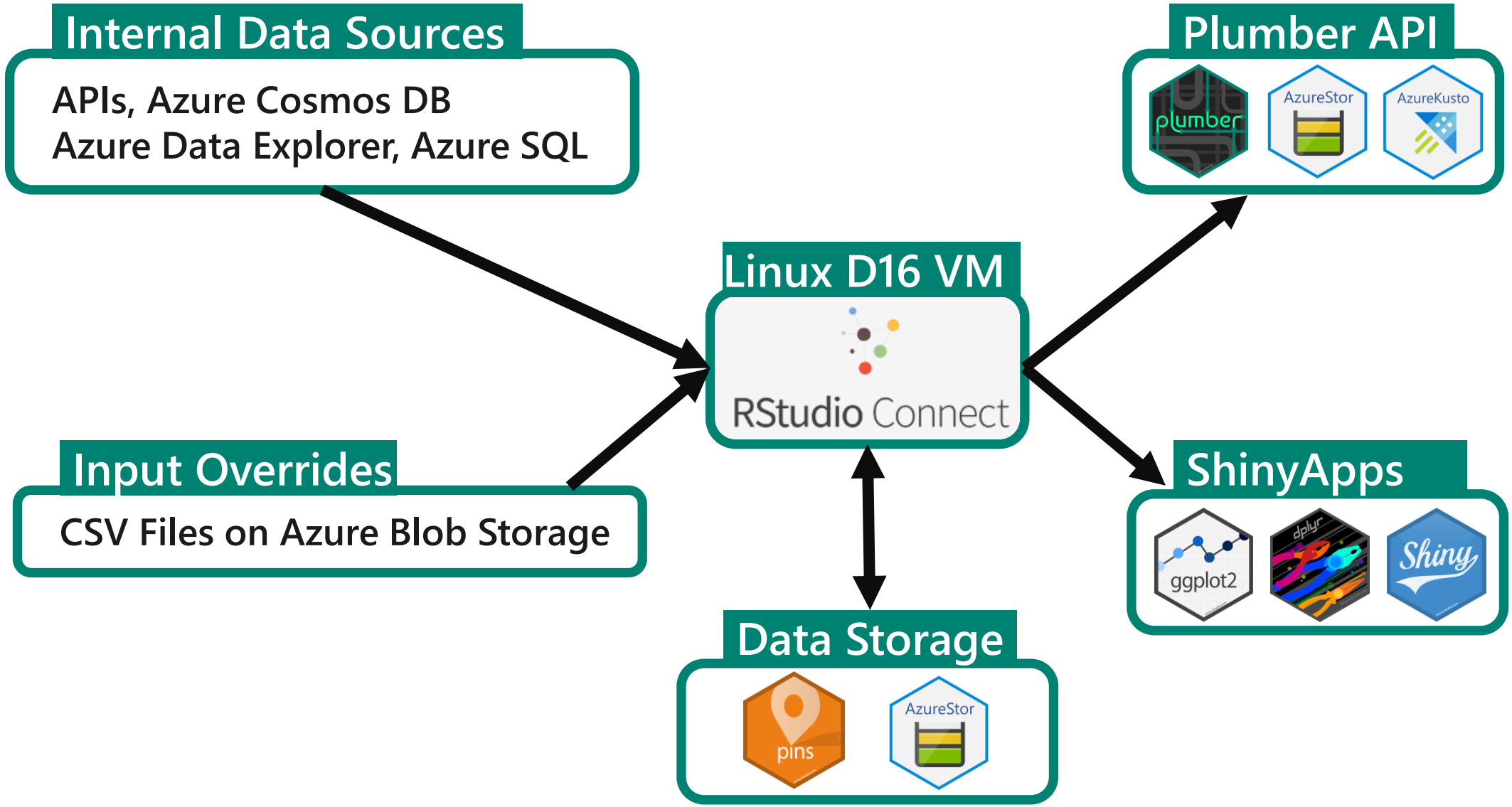
Transparent and Explainable

- Shiny Apps help explain the different factors driving the capacity plan
- Helped foster much greater transparency and collaborations among multiple teams

More Productive Data Scientists

- End-to-end ownership of data science process
- Quickly model and incorporate human intelligence and business strategy

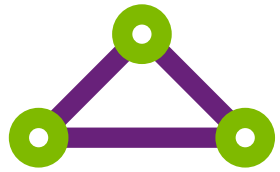
Capacity Planning: Dataflow



Capacity Planning: Microsoft Services Most Used

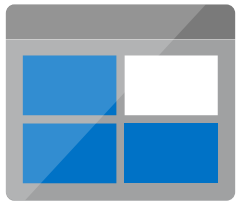
Azure AD

Azure Data Explorer

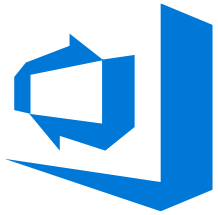


Azure ExpressRoute

Azure SQL Database



Azure Blob Storage



Azure DevOps

Azure Key Vault

GIT

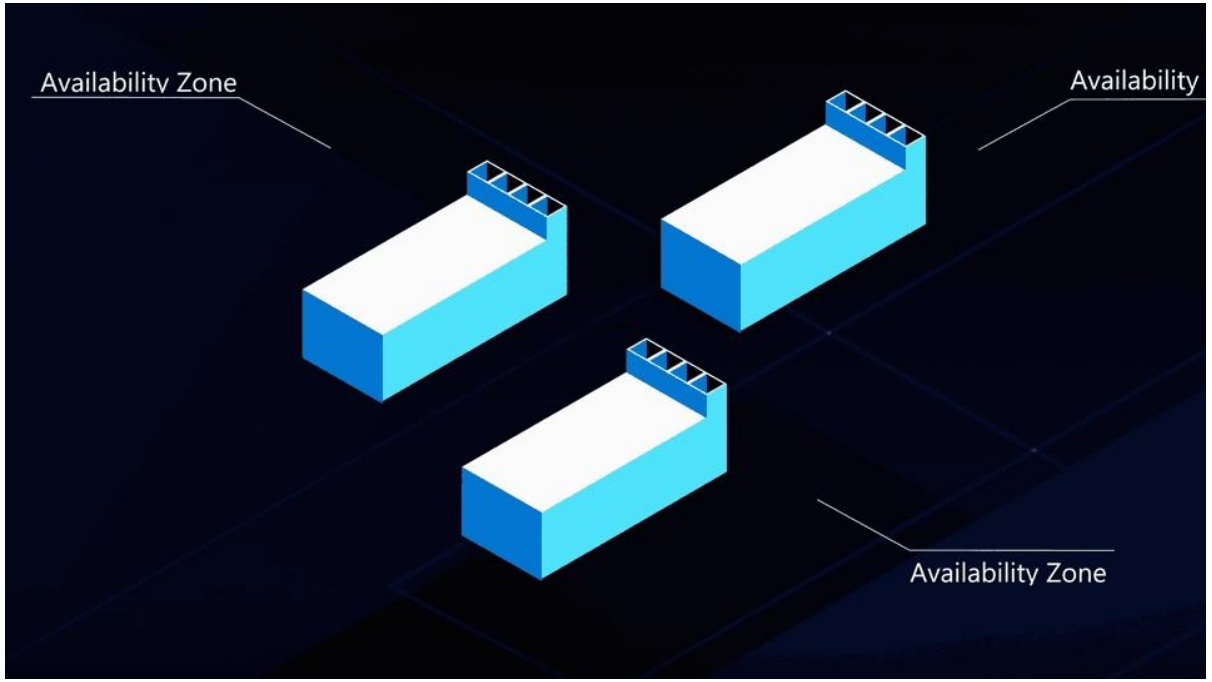
Azure VMs

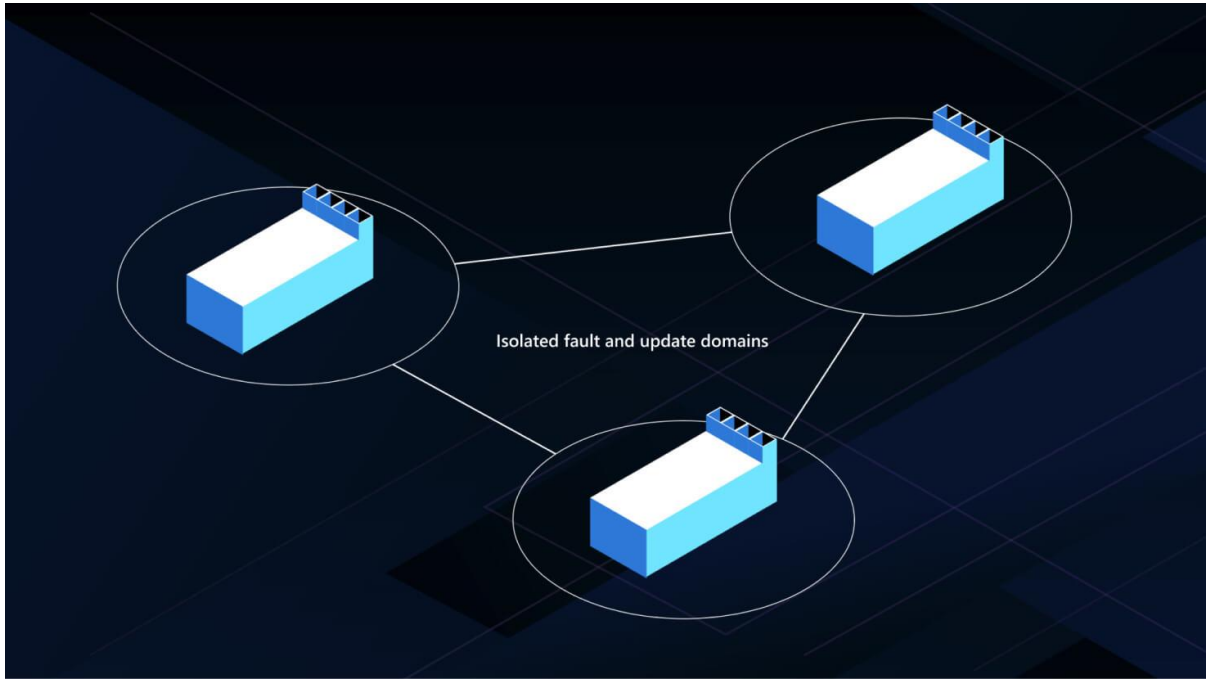
Azure Functions

Excel

Azure Synapse

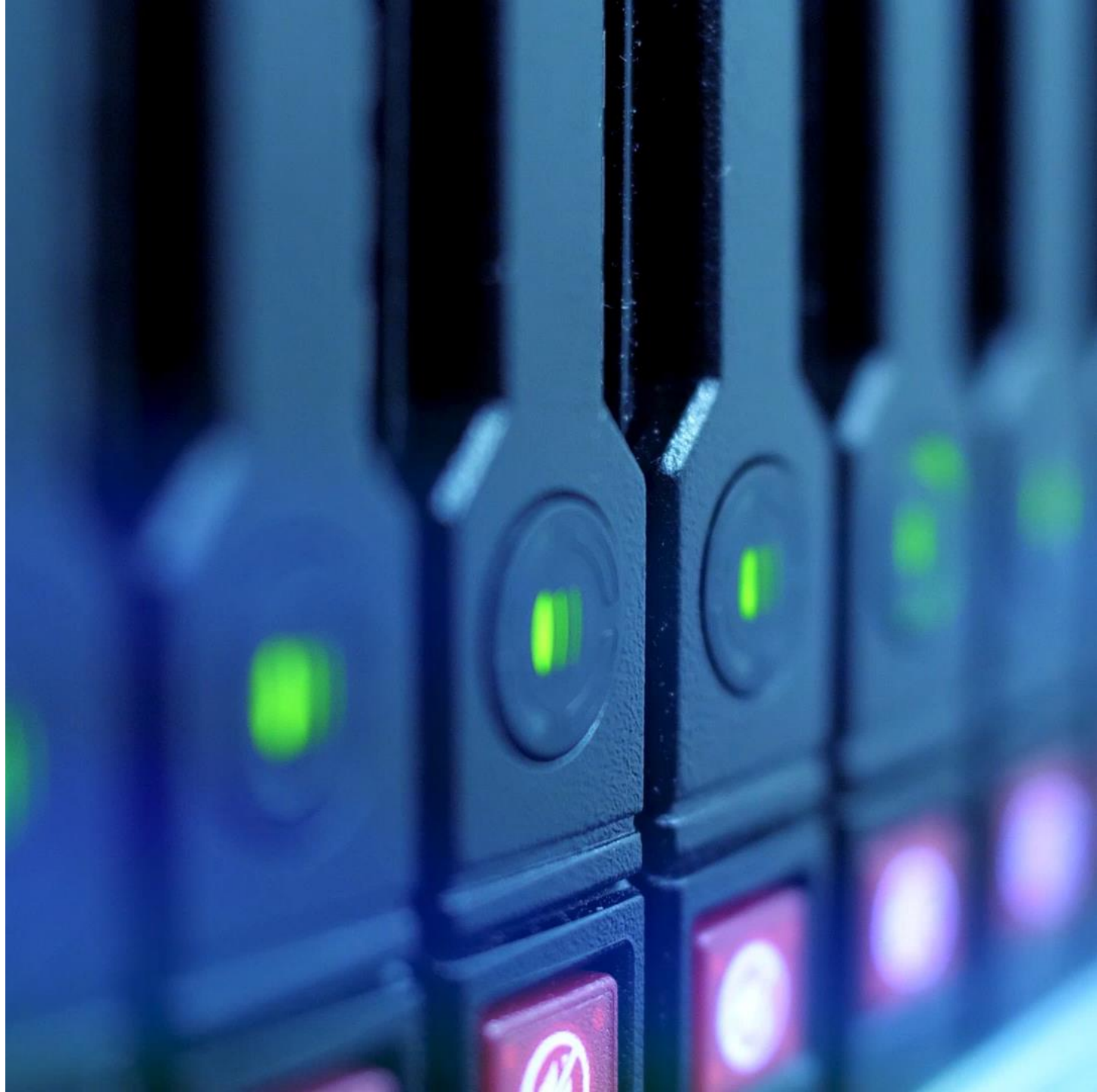
Intro to Azure Availability Zones





- Availability zones are physically and logically separated datacenters with their own independent power source, network and cooling.
- We split demand into 3 buckets:
  - Pinned to particular zone
  - Replicated across all 3 zones
  - Discretionary; can be shaped to zones with more excess supply, maybe lower costs, etc.

Datacenter Capacity Dashboard Demo



DataCenter Capacity

- LRP Comparer
- Pin Comparer
- Zonal Simulator

New Board

demo-20220101

Old Board

demo-20211201

Comparer

Execute

Results

Download

View

Table

Regional

Summarize

Metro

Filters

Metro

Geo

GeoArea

RegionType

Metrics

All

Years

2022 2023 2024 2025

Delta

0 10

0 1 2 3 4 5 6 7 8 9 10

Comparison

Grouped		June 2022		June 2023		June 2024		June 2025	
Metro	Metric	Value	Delta	Value	Delta	Value	Delta	Value	Delta
▼ Hala (3)	Regional	30.5	-0.2	40.6	-0.0	51.2	0.8	59.9	0.6
	- Organic	12.8		16.7	0.7	21.2	1.6	25.4	1.6
	- Networking	5.0		5.2	0.0	5.5		5.7	
	- Engineering	12.6	-0.2	18.7	-0.7	24.5	-0.9	28.7	-1.0
▼ Minas Tirith (3)	Regional	91.7	2.3	123.2	1.7	141.8	3.3	152.0	3.3
	- Organic	43.6	0.1	53.0	0.5	61.1	0.5	71.0	0.5
	- Networking	7.8		16.2	0.0	16.7	0.0	17.0	0.0
	- Engineering	40.3	2.3	54.0	1.3	64.0	2.8	64.0	2.8

DataCenter Capacity

- LRP Comparer
- Pin Comparer
- Zonal Simulator

Pin: revenue_forecast

New Board: demo-20220101

Old Board: demo-20211201

Comparer: Execute

Results: Download

Comparison

Metro	Metric	ForecastDate	demo-20220101	demo-20211201	DiffValue	Changed
Hala	Revenue_Forecast	2021-12-01	25675337	23775283	1900054	True
Hala	Revenue_Forecast	2022-01-01	26393247	24497676	1895571	True
Hala	Revenue_Forecast	2022-02-01	27121376	25230762	1890614	True
Hala	Revenue_Forecast	2022-03-01	27879628	25987331	1892297	True
Hala	Revenue_Forecast	2022-04-01	28670585	26772574	1898011	True
Hala	Revenue_Forecast	2022-05-01	29479506	27584081	1895425	True
Hala	Revenue_Forecast	2022-06-01	30298572	28418730	1879842	True
Hala	Revenue_Forecast	2022-07-01	31126886	29272677	1854209	True
Hala	Revenue_Forecast	2022-08-01	31970473	30141365	1829108	True
Hala	Revenue_Forecast	2022-09-01	32832640	31022881	1809759	True
Hala	Revenue_Forecast	2022-10-01	33713980	31917955	1796025	True
Hala	Revenue_Forecast	2022-11-01	34612372	32829965	1782407	True
Hala	Revenue_Forecast	2022-12-01	35526789	33762079	1764710	True
Hala	Revenue_Forecast	2023-01-01	36457298	34717257	1740041	True
Hala	Revenue_Forecast	2023-02-01	37405066	35698251	1706815	True
Hala	Revenue_Forecast	2023-03-01	38370736	36703722	1667014	True
Hala	Revenue_Forecast	2023-04-01	39354437	37728237	1626200	True
Hala	Revenue_Forecast	2023-05-01	40355775	38762275	1593500	True
Hala	Revenue_Forecast	2023-06-01	41374798	39799488	1575310	True
Hala	Revenue_Forecast	2023-07-01	42411994	0	42411994	True

1–20 of 62 rows

Previous 1 2 3 4 Next

DataCenter Capacity

- LRP Comparer
- Pin Comparer
- Zonal Simulator

Data

Board: demo-20220101

Metro/Region: Hala/Kree Empire

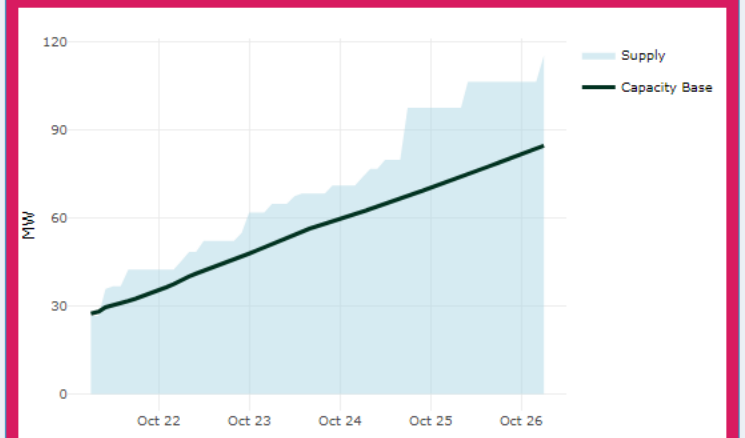
AZ1 Buffer %: 0 — 100 — 200

AZ2 Buffer %: 0 — 100 — 200

AZ3 Buffer %: 0 — 100 — 200

Download

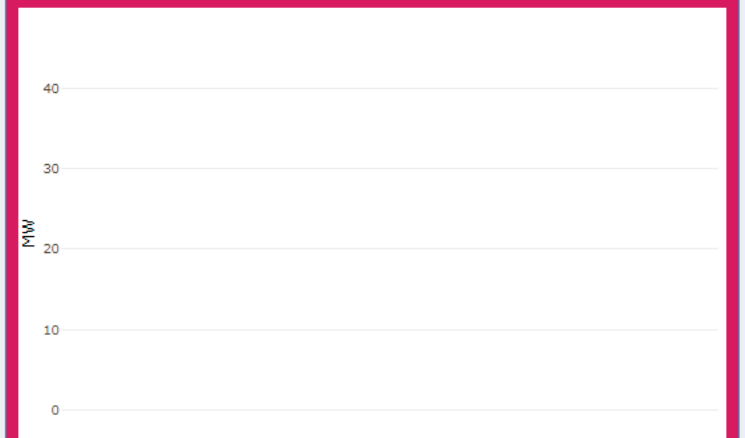
Regional



MAX ZONAL SUPPLY RISK INDEX
3

TOTAL FY24 P99 MW
56

Non-AZ



NON-AZ FY24 P99 MW
0

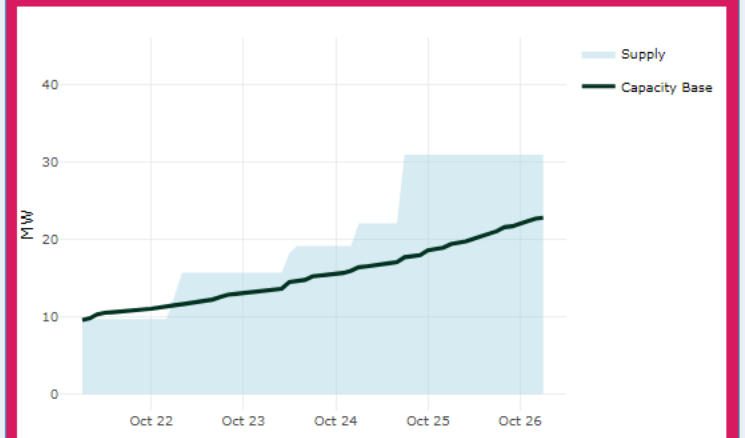
Model Parameters

AZ1 Max Allocation %: 0 — 100

AZ2 Max Allocation %: 0 — 100

AZ3 Max Allocation %: 0 — 100

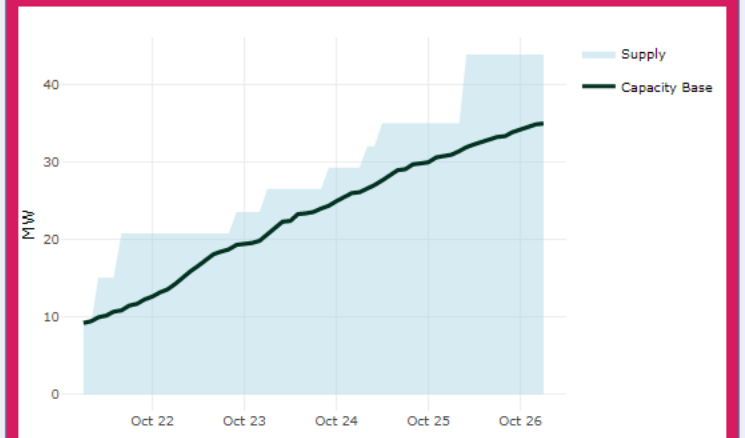
AZ01



AZ01 SUPPLY RISK INDEX
3

AZ01 FY24 P99 MW
15

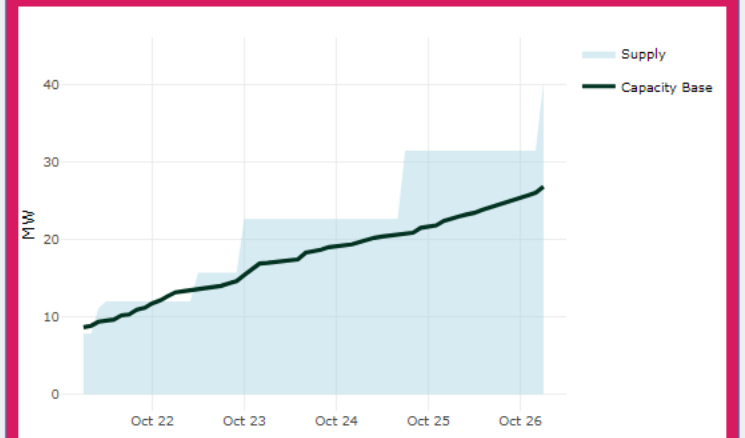
AZ02



AZ02 SUPPLY RISK INDEX
3

AZ02 FY24 P99 MW
23

AZ03



AZ03 SUPPLY RISK INDEX
3

AZ03 FY24 P99 MW
18

How to get started?

- Open an Azure Account here:
  - https://azure.microsoft.com/en-us/
- Create an Azure Virtual Network
- Create a Linux VM in that virtual network
  - Linux (ubuntu 18.04)
  - Standard D16s v3 (16 vcpus, 64 GiB memory)
- Install RStudio Connect on Linux VM:
  - https://www.rstudio.com/products/connect/
- (Optional?) Provision ExpressRoute to connect your virtual network to your on-prem network
- (Optional?) Connect to Azure Active Directory via SAML
  - Getting Started with SAML in RStudio Connect – RStudio Support

Thanks to Our Extended Team and Partners!

Microsoft

Akash Barnwal, Murthy Bhavaraju, Austin Dowling, Gitanjali Dusija, Mike Eck, Pat Filoteo, Betsy Han, Luciano Hunziker, Shantanu Jha,

Sid Rao + **all our internal partners & stakeholders**

RStudio

Rachael Dempsey, Tom Mock, Mitch Tanney

Microsoft