

云端科研计算实战分享

张洋，AWS 解决方案架构师

徐晓，AWS 教育行业总监

议程

- 为什么要用云计算
- 认识计算环境
- 构建计算环境
- AWS科研支持项目

从1到万
Unlimited infrastructure

Why Cloud for HPC?

Concurrent Clusters on-demand
灵活细化定义任务

包年+SPOT实例
Low cost with flexible pricing

有效配置
Efficient clusters

Faster time to results
业务驱动结果

Increased collaboration
随处访问

Defining HPC – example use cases

Clustered (Tightly coupled)



- Fluid dynamics
- Weather forecasting (天气预测)
- Materials simulations
- Crash simulations

- Seismic processing
- Metagenomics
- Astrophysics
- Deep learning (深度学习)

Data Light

Minimal requirements for high performance storage

Data Heavy

Benefits from access to high performance storage

- Risk simulations
- Molecular modeling
- Contextual search
- Logistics simulations

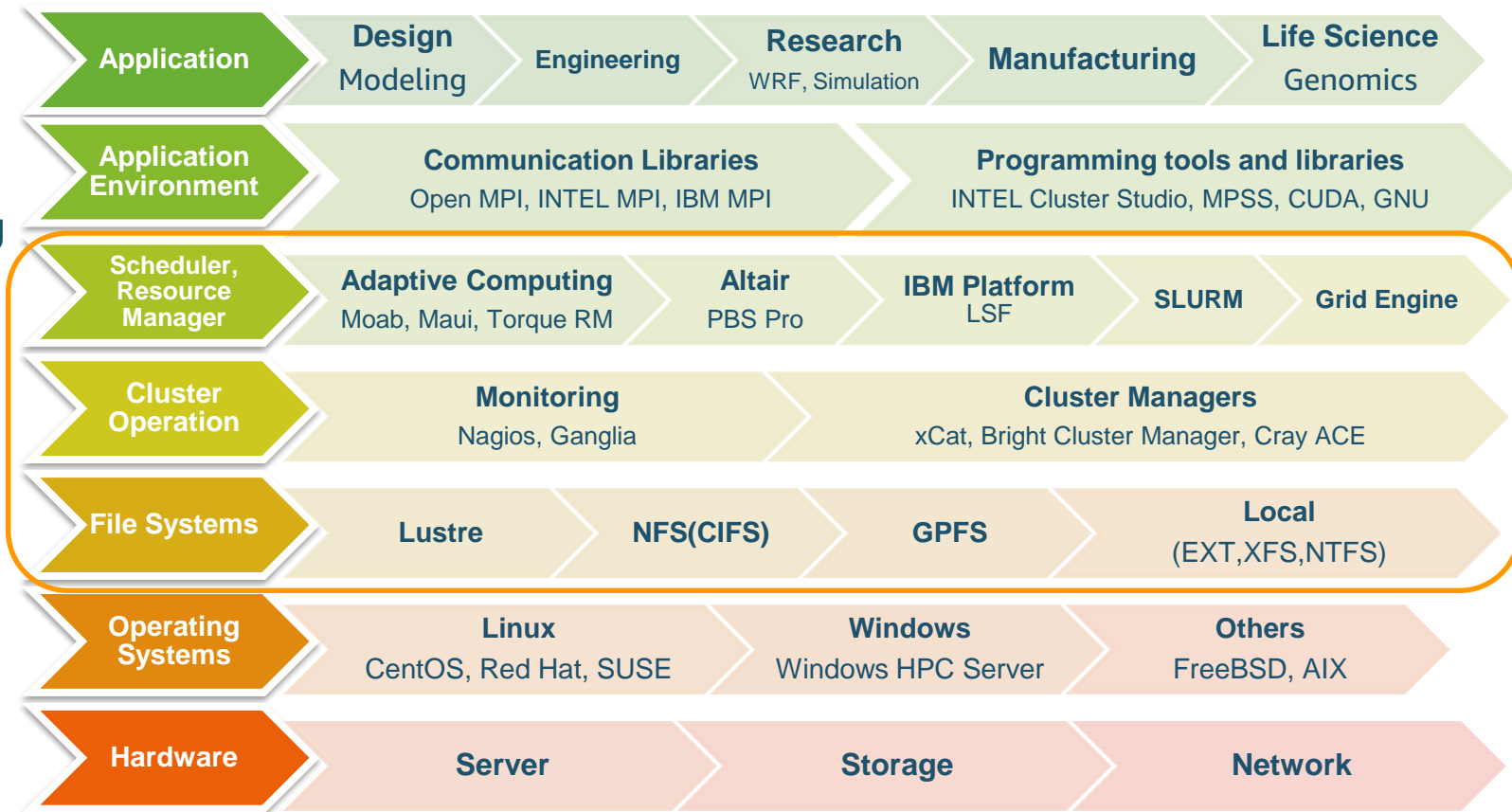
- Animation and VFX (渲染)
- Genomics (基因)
- Semiconductor verification
- Image processing/GIS

Distributed / Grid (Loosely coupled)



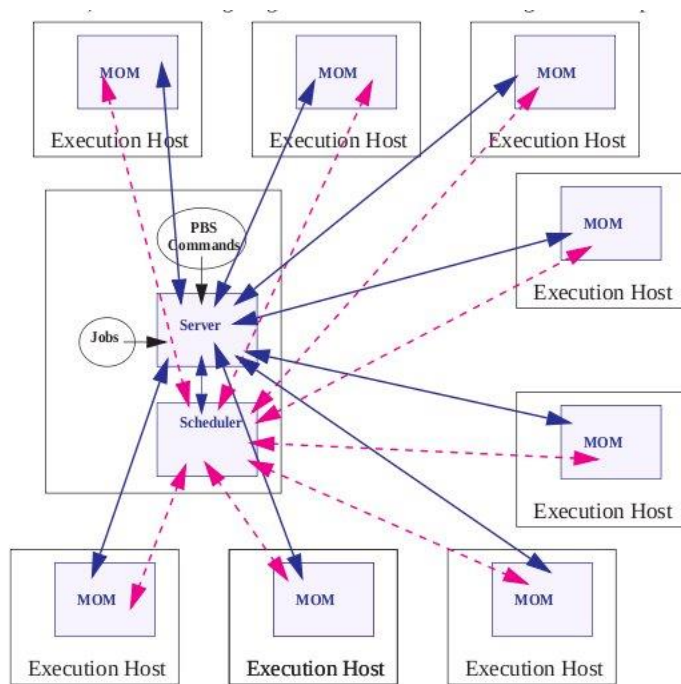
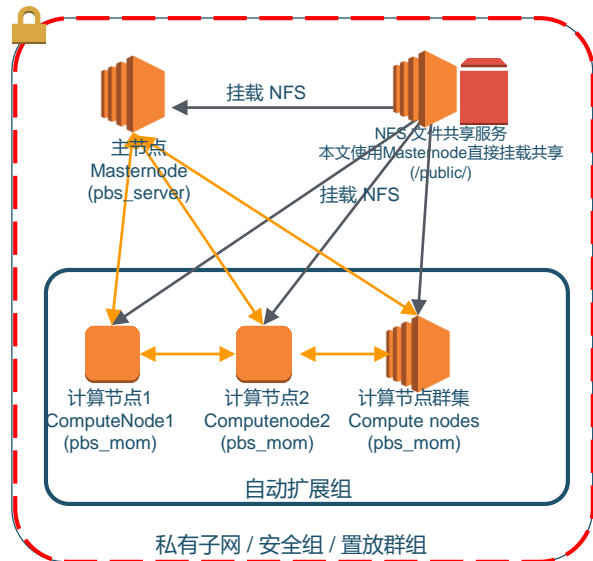
HPC Stack

- HPC
- Clustering
- Grid
- Distributing

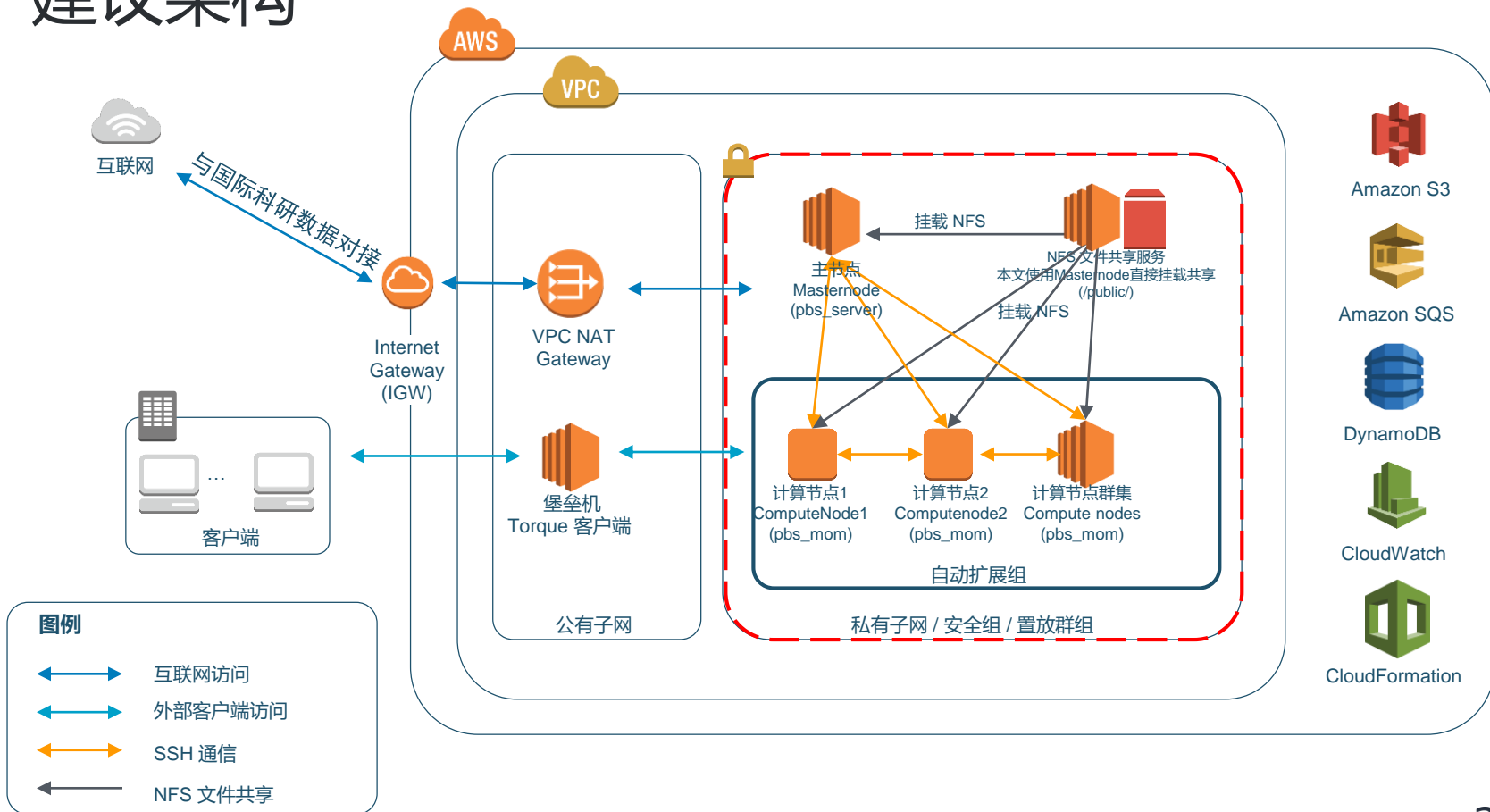


Torque Resource Manager (PBS) 架构

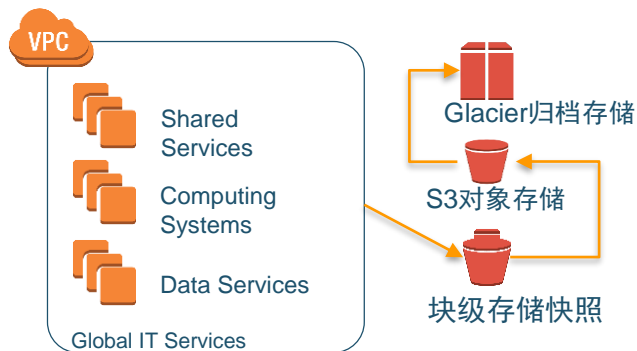
Master-Worker模型



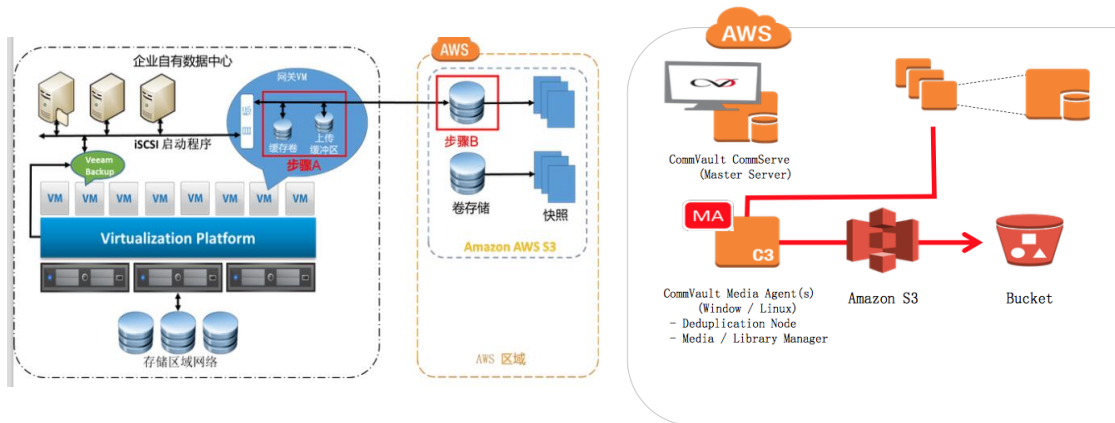
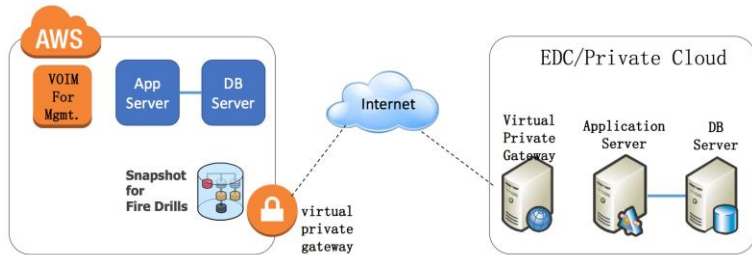
建议架构



数据生命周期



	Technology	Usage	Target Scenario
Disaster Recovery	InfoScale	Failover b/w AWS and on premise or b/w AZs for most complex LOB	Traditional EDC customers who lack second EDC or wants to remove second EDC
	Resilience Platform	Lightweight failover solution b/w VMWare and AWS	VMWare-centric customers (can run side-by-side with InfoScale)
Application High Availability	Storage Foundation	Optimized Volume management for Oracle RAC, SAP and Informatica	Running mission critical applications on AWS
Data Protection	NetBackup	Backup EDC data to AWS and restore back to EDC	EDC customers who expect minimize DP cost or requires remote DP capability
	Information Map (to be landed in China)	Monitor and tuning DP solution by manage backup metadata	Enterprise customers who have multiple EDC and cloud data across region protected by NBU



海量数据如何批量上传、下载？



E-ink shipping
label



80 TB
10G network



防雨防尘

电磁、温度防护

端到端加密

AWS中国（宁夏）区域由西云数据运营
AWS中国（北京）区域由光环新网运营



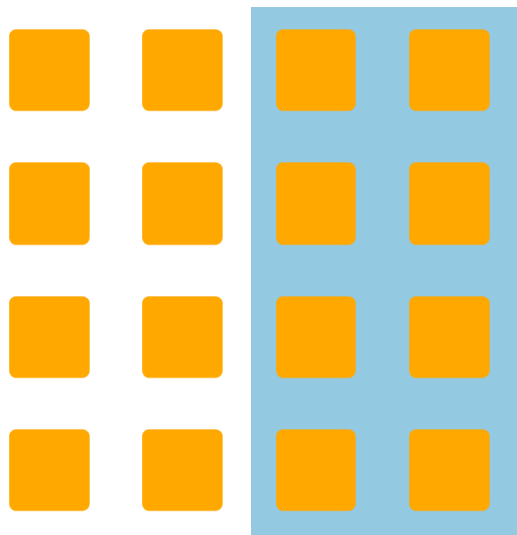
AWS技术助力科研

- 存储配置
 - 常用数据EBS(块级数据)
 - 非常用数据S3(文件, 归档数据)
- 安全配置
 - 所有用户数据系统不共享
 - 物理数据多重备份
- 网络配置：
 - 内部环境25Gb/s高速带宽
 - 实例不限制带宽
 - 国际线路优化
- 每年环境更新: CPU, GPU, FPGA等

SPOT型实例应用案例 – 弹性群集扩展

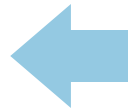
用预期的花费满足基本
运算需求

核心节点使用按需
实例，标准的EC2
计价方式



用更低的价钱超越基本
运算需求

任务节点使用SPOT实
例，相比标准计价方式
最高节省大于80%



实例类型	数量	出价价格
m3.xlarge	1	
c3.4xlarge	2 调整大小	0.30
c3.4xlarge	2 调整大小	

SPOT实例细节

近3折的价格!*



购买选项

- 单次请求竞价实例
- Spot Fleet 竞价型队列，协助启动不同大小的竞价型实例满足运算要求
- Spot Blocks 带有指定持续时间的竞价型实例(1-6 小时)

承诺租期

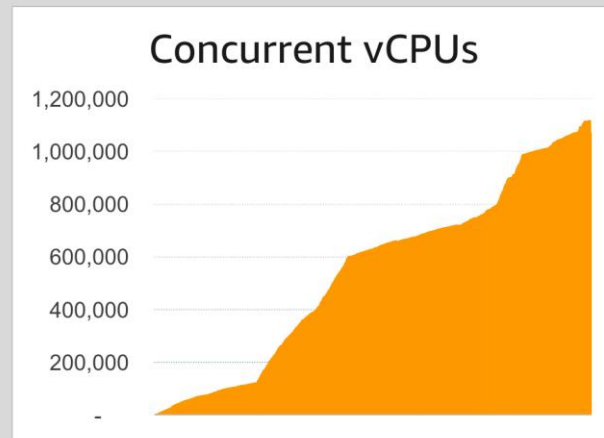
- 无须

SPOT实例: 物尽其用



ON AVERAGE, **EVERY WEEK**, AWS CUSTOMERS ARE USING MORE COMPUTE CAPACITY ON AMAZON EC2 SPOT INSTANCES THAN CUSTOMERS IN 2012 WERE RUNNING ACROSS **ALL OF AMAZON EC2**.

CLEMSON UNIVERSITY RAN THE LARGEST HIGH PERFORMANCE CLUSTER ON SPOT WITH **1,100,000 CONCURRENT vCPUs**



AWS支持高校教育创新项目计划

AWS教育项目

- 在云领域有前沿经验的教授开放的课程内容
- 云相关课题和AWS服务、实验和培训
- 免费AWS资源
- 可分享最佳实践个性化虚拟化社区
- 职业发展路径和工作机会



<https://www.awseducate.com>

AWS云计算科研项目

- 发起和支持云驱动技术的开发
- 一次性的提供免费的AWS 资源用于科研，例如，EC2 (计算) or S3 and EBS (存储) 及其他，例如Amazon Kinesis
- 所支持的项目通常会生成可重复利用的成果，例如软件镜像或开放数据集。
- 减少尝试云的风险



https://aws.amazon.com/cn/events/educate_cloud/research-credits/

创业支持

- AWS服务补贴（3千~10万人民币）
 - AWS Support（商业计划1月~1年）
 - 业务与技术在线培训
 - 免费云技术上手实验培训
 - 系统架构师一对一指导
- **具体内容因组织而异**

aws  云创计划

<https://www.amazonaws.cn/activate/>

Thank you!