On Cloud Nine:
How to be happy migrating your in-memory computing platform to the cloud

Stephen Darlington

3 June 2019

Hello

- Define terms
- Deployment options
- Best Practices

Photo by Jon Tyson on Unsplash

GridGain Company Confidential

What is GridGain?

Combines the performance and scale of in-memory computing together with the disk durability and strong consistency in one system

Co-located Computations

Brings the computations to the servers where the data actually resides, eliminating need to move data over the network
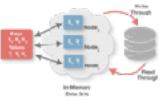
Distributed Key-Value

Read, write and transact with fast key-value APIs

Distributed SQL

Horizontally, fault-tolerant distributed SQL database that treats memory and disk as active storage tiers

ACID Transactions

Supports distributed ACID transactions for key-value as well as SQL operations
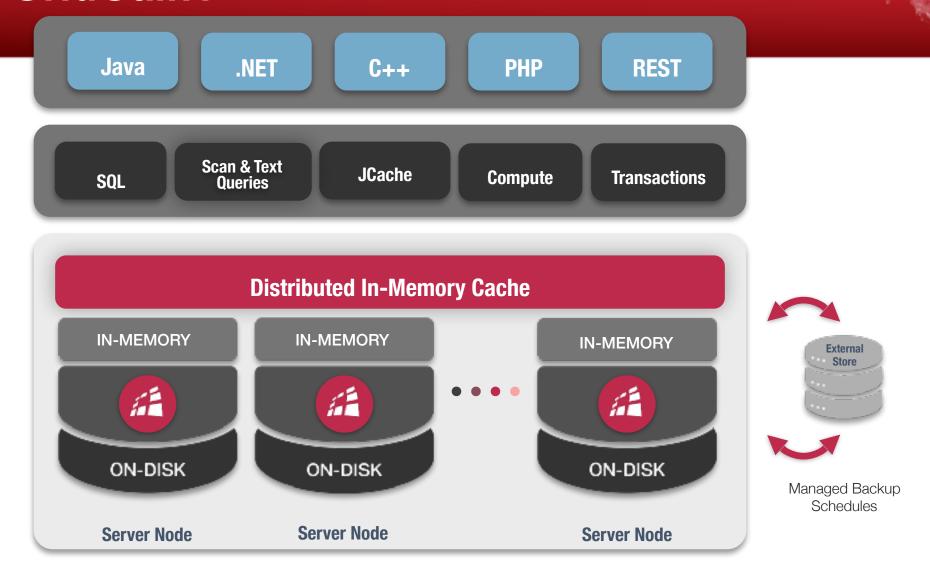
Machine and Deep Learning

Set of simple, scalable and efficient tools that allow building predictive machine learning models without costly data transfers (ETL)

GridGain

What is GridGain?

What is "The Cloud"?

There is no cloud
it's just someone else's computer

GridGain

According to SalesForce

1. Flexibility
2. Disaster Recovery
3. Automated software updates
4. Capital-expenditure free
5. ~~Increase collaboration~~
6. Work from anywhere
7. ~~Document control~~
8. Security
9. Competitiveness
10. Environmentally friendly

https://www.salesforce.com/uk/blog/2015/11/why-move-to-the-cloud-10-benefits-of-cloud-computing.html

According to LifeWire

- Cost efficiency
- Almost unlimited storage
- Backup and recovery
- Automatic software integration
- East access to information
- Quick Deployment

- Technical Issues

- Security
- Prone to attack

https://www.lifewire.com/cloud-computing-explained-2373125

According to IBM

- Scalability
- Storage options
- Control choices
- Tool selection
- Security features
- Accessibility
- Speed to market
- Data security
- Savings on equipment
- Pay structure

- Streamlined work
- Regular updates
- Collaboration
- Competitive edge

https://www.ibm.com/cloud/learn/benefits-of-cloud-computing

Scalability

GridGain

Flexibility

Photo by Yannes Kiefer on Unsplash

It's not in your data centre

Photo by Taylor Vick on Unsplash

Best Practice 1: Tooling

Photo by Lachlan Donald on Unsplash

GridGain Company Confidential

Kubernetes is…

"…an open-source system for automating deployment, scaling, and management of containerized applications."

kubernetes

GridGain

"A container is a standard unit of software that packages up code and all its dependencies so the application runs quickly and reliably from one computing environment to another"
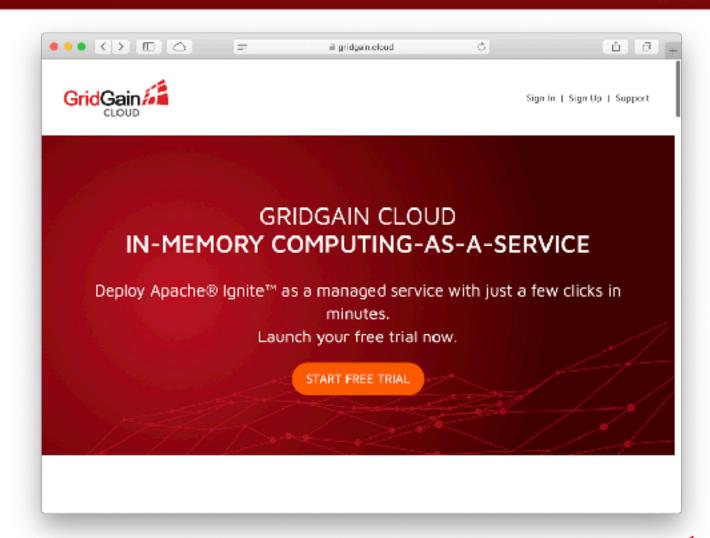

By dotCloud, Inc. - File:Docker (container engine) logo.png,
Apache License 2.0, https://commons.wikimedia.org/w/index.php?curid=52332268

GridGain Company Confidential

GridGain

Don't deploy: GridGain Cloud

- Fully Managed In-Memory Computing Built on Apache Ignite
- Web Console
- Data access
    - REST
    - JDBC/ODBC
    - Ignite thin-clients
- "Up and running in minutes"

AWS

- EC2
- ECS
- EKS
- On-demand
- Spot instances
- Reserved instances
- Dedicated hosts
- Fixed performance
- Burstable
- Cluster Networking

- Intel
- ARM
- General purpose (7 options)
- Compute (3 options)
- Memory optimized (7 options)
- Accelerated (4 options)
- Storage optimized (4 options)

Azure

- Virtual Machines
- Virtual Machine Scale Sets
- Azure Kubernetes Service
- Container Instances

GridGain

Other options

2019 © GridGain Systems   *GridGain Company Confidential*

Best Practice 2: Memory Sizing

- For data-grid use cases, aim for the "big memory" options
- Consider resilience
  - Better to have four 256Gb nodes than one 1Tb node
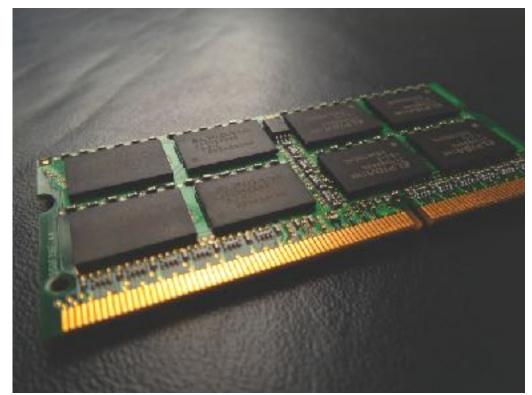- Use the Sizing Calculator (https://apacheignite.readme.io/docs/capacity-planning)

Photo by Franck V. on Unsplash

GridGain Company Confidential

GridGain

Best Practice 3: Disk Storage

- Trade offs
  - SSDs in EC2 instances are fast but…
  - EBS is more flexible
  - For Azure, use SSD
- Use StatefulSets in Kubernetes

Photo by Vincent Botta on Unsplash

GridGain Company Confidential

GridGain

Best Practice 4: Images

- Use custom images or containerisation!
- Make it easy to create / kill new GridGain nodes
- Automate as much as possible

This Photo by Unknown Author is licensed under CC BY-SA

Best Practice 5: Performance

- Predictable versus maximum
  - Reserved
  - Spot
- Scale automatically
  - "Horizonal Pod Autoscaler" with Kubernetes
  - EC2 Auto Scaling in AWS
  - Scale Sets in Azure
  - But remember licencing

Photo by Alessio Lin on Unsplash

Best Practice 6: Security

- TLS/SSL between nodes
- Disk encryption
- Firewall ports
- Use GridGain security options for authentication, authorisation and auditing

Photo by Victor Garcia on Unsplash

GridGain Company Confidential

GridGain

Best Practice 7: Data location

- Where is your data?
- How do you get it to "the cloud"?
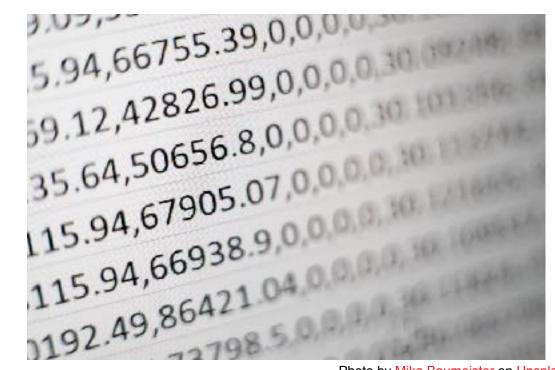 - Deltas versus full extracts
  - Migrate everything?
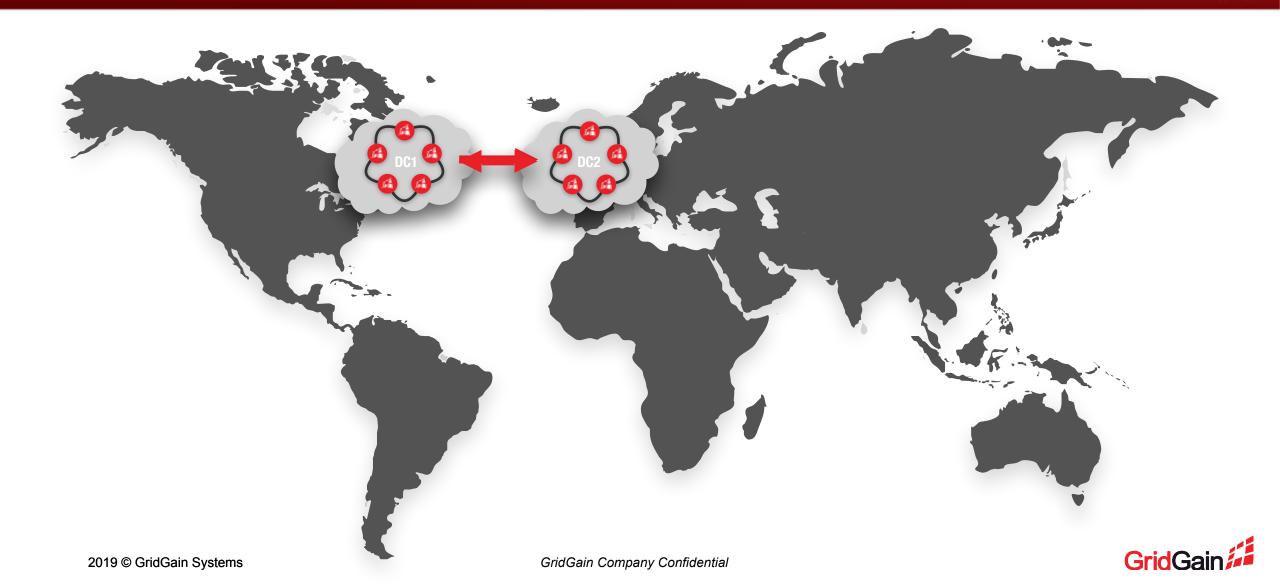

Photo by Mika Baumeister on Unsplash

GridGain Company Confidential

Best Practice 8: Migration

- Use Data Centre replication

Photo by Jason Hafso on Unsplash

GridGain Company Confidential

Best Practice 8: Migration

What have we learned?

- The cloud *is* different from on-prem
- Best practices
  - Tooling
  - Memory sizing
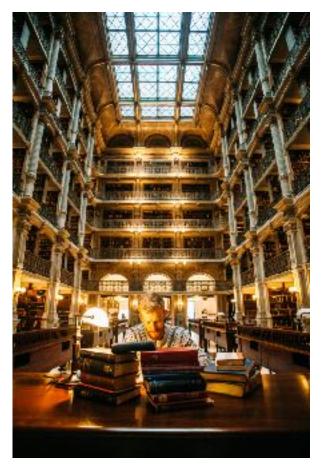 - Disk storage
  - Use custom images
  - Performance
  - Security
  - Data location
  - Migration

Photo by Elijah Hail on Unsplash

Thank you
Stephen Darlington, Senior Consultant, GridGain Systems
stephen.darlington@gridgain.com
@sdarlington

GridGain