

Porting Research Pipelines into Clouds

Architectural considerations (1/3)

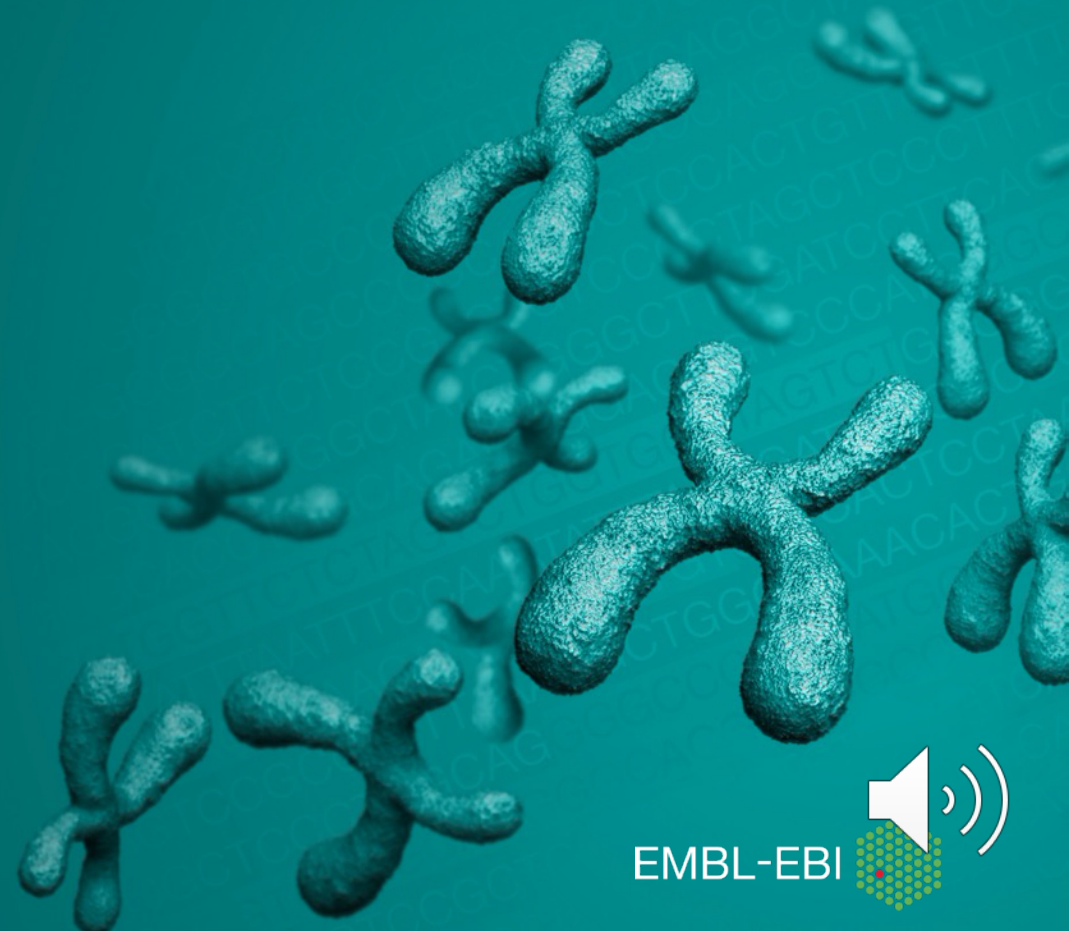


David Yuan, Ph.D.

Cloud Bioinformatics Application Architect

Technology and Science Integration

European Bioinformatics Institute, EMBL



Porting into clouds

Cloud overview

Why clouds

What the *-aaS

Which clouds

Container & orchestration

Important considerations

Portability

Scalability

High availability

Disaster Recovery

Maintainability

Research pipelines

Cost, budget & funding

Data-driven architecture

Lift-n-shift vs. cloud-native

Monitoring



Cloud overview – why?

Research pipelines

- Archive of sequence data, images, publications or ontology information
- Pipelines to analyse data
- Services to aggregate other research tools or databases

Good candidates for the cloud!

- You know your pinch-points.
- Cloud is mature and fast-evolving.
- Lift-n-shift is possible.
- Being cloud-native provides benefit way over cost.

Pros

- Stable infrastructure
- Global collaboration by default
- Flexible resource management
- Potential cost reduction
- Latest and greatest technology stack

Cons

- Accounting model is very different.
- The whole field is still growing.
- Beginners often face steep learning curves.



Cloud overview – why?

Research pipelines

- Archive of sequence data, images, publications or ontology information
- Pipelines to analyse data
- Services to aggregate other research tools or databases

Good candidates for the cloud!

- You know your pinch-points.
- Cloud is mature and fast-evolving.
- Lift-n-shift is possible.
- Being cloud-native provides benefit way over cost.

Pros

- Stable infrastructure
- Global collaboration by default
- Flexible resource management
- Potential cost reduction
- Latest and greatest technology stack

Cons

- Accounting model is very different.
- The whole field is still growing.
- Beginners often face steep learning curves.



Cloud overview – why?

Research pipelines

- Archive of sequence data, images, publications or ontology information
- Pipelines to analyse data
- Services to aggregate other research tools or databases

Good candidates for the cloud!

- You know your pinch-points.
- Cloud is mature and fast-evolving.
- Lift-n-shift is possible.
- Being cloud-native provides benefit way over cost.

Pros

- Stable infrastructure
- Global collaboration by default
- Flexible resource management
- Potential cost reduction
- Latest and greatest technology stack

Cons

- Accounting model is very different.
- The whole field is still growing.
- Beginners often face steep learning curves.

Cloud overview – why?

Research pipelines

- Archive of sequence data, images, publications or ontology information
- Pipelines to analyse data
- Services to aggregate other research tools or databases

Good candidates for the cloud!

- You know your pinch-points.
- Cloud is mature and fast-evolving.
- Lift-n-shift is possible.
- Being cloud-native provides benefit way over cost.

Pros

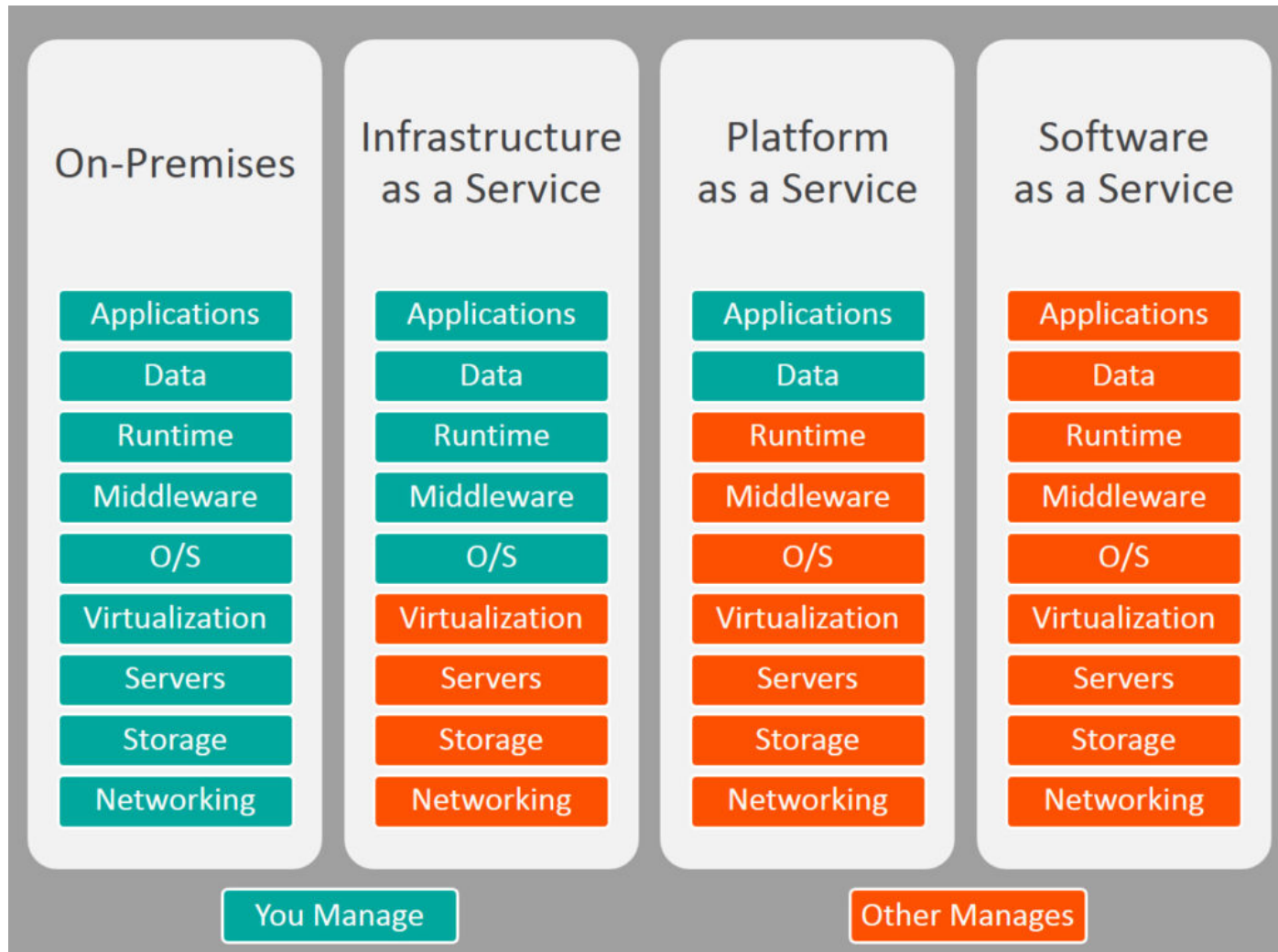
- Stable infrastructure
- Global collaboration by default
- Flexible resource management
- Potential cost reduction
- Latest and greatest technology stack

Cons

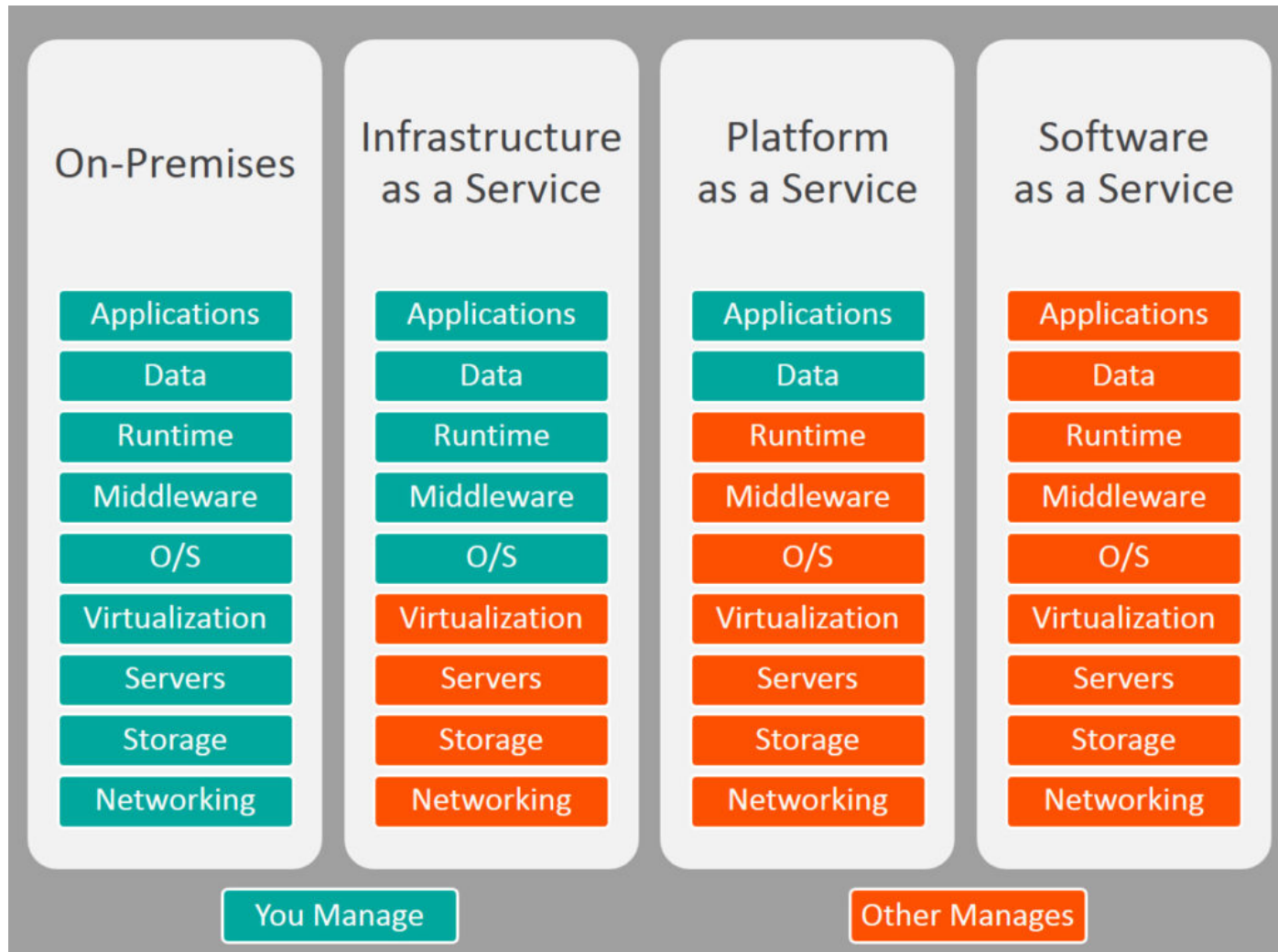
- Accounting model is very different.
- The whole field is still growing.
- Beginners often face steep learning curves.



Cloud overview – what?



Cloud overview – what?

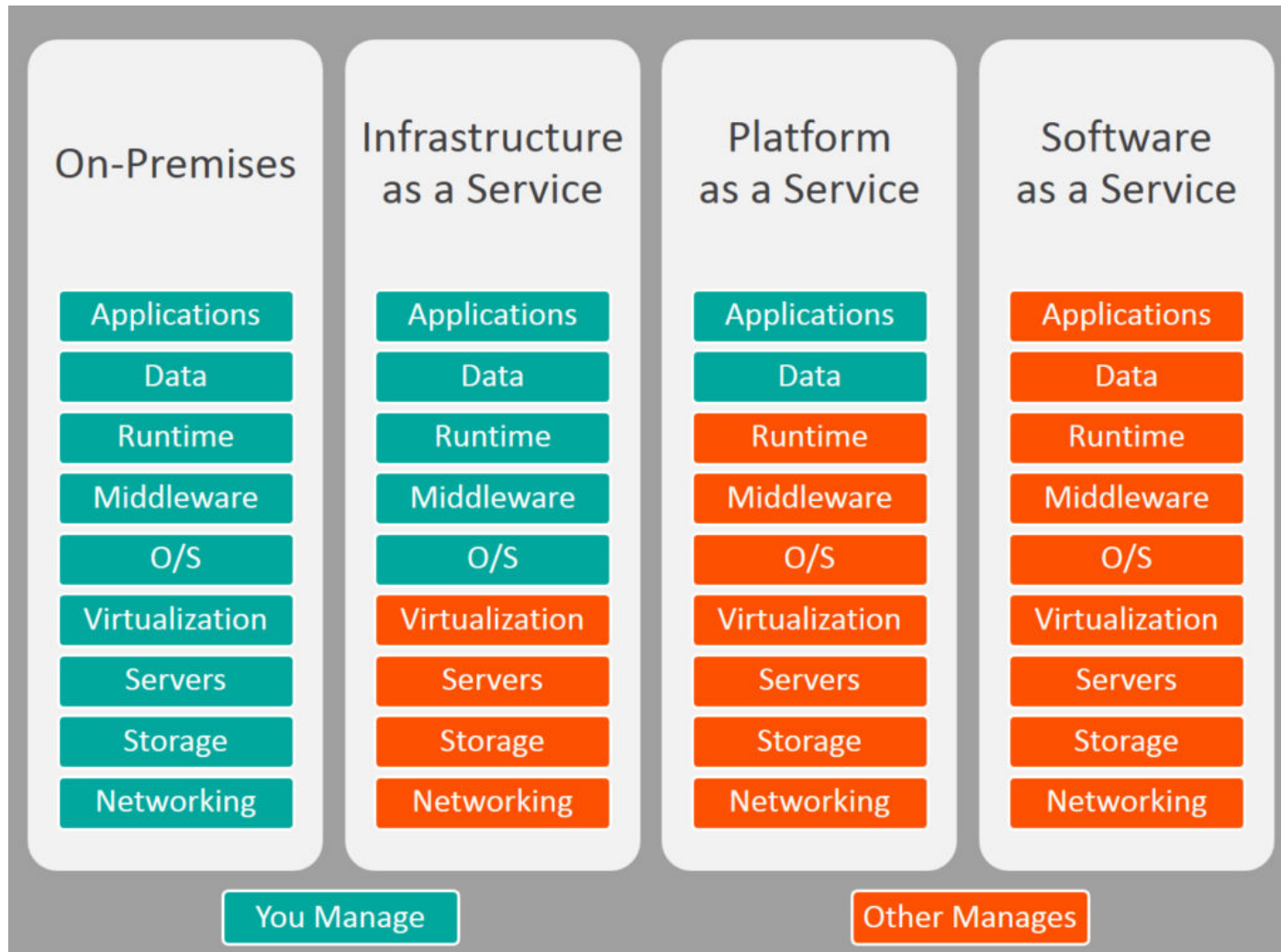


* - as a Service

- Infrastructure as a Service (IaaS)
 - OpenStack
 - GCP, AWS, MSA
 - RKE on OpenStack
 - GKE, EKS, AKS
- Platform as a Service (PaaS)
 - AWS Lambda
 - Azure App Service
- Software as a Service (SaaS)
 - AWS Route53
 - Oracle Autonomous Data Warehouse Cloud



Cloud overview – what?

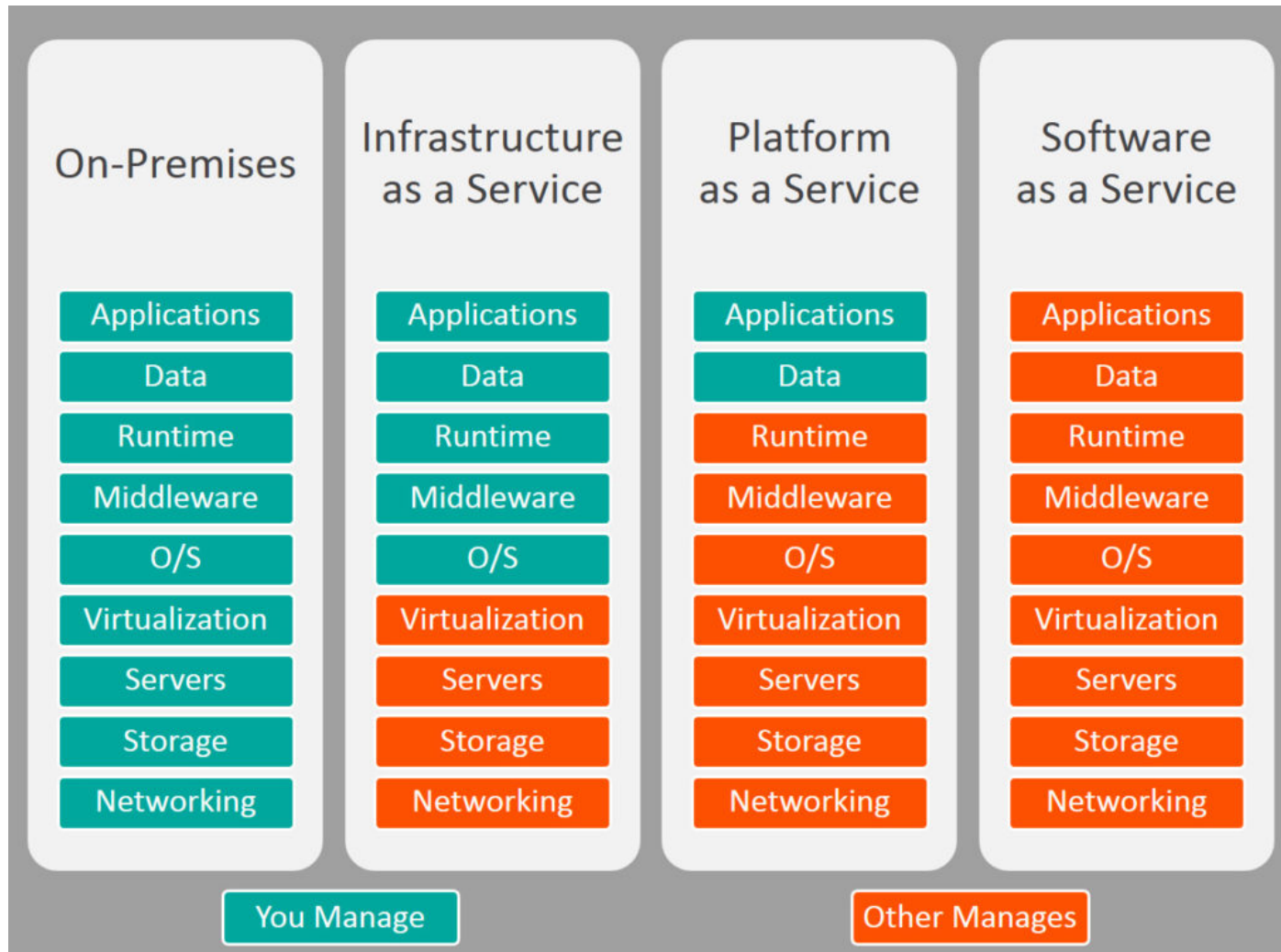


* - as a Service

- Infrastructure as a Service (IaaS)
 - OpenStack
 - GCP, AWS, MSA
 - RKE on OpenStack
 - GKE, EKS, AKS
- Platform as a Service (PaaS)
 - AWS Lambda
 - Azure App Service
- Software as a Service (SaaS)
 - AWS Route53
 - Oracle Autonomous Data Warehouse Cloud



Cloud overview – what?

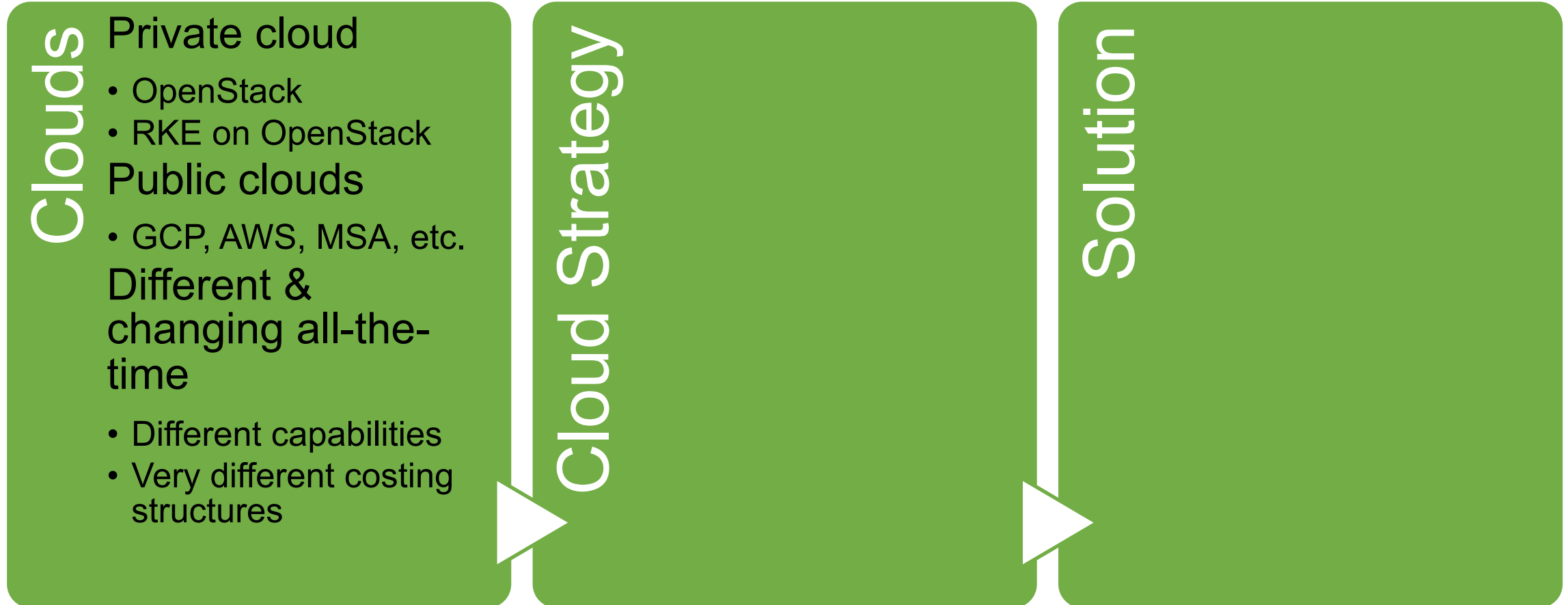


* - as a Service

- Infrastructure as a Service (IaaS)
 - OpenStack
 - GCP, AWS, MSA
 - RKE on OpenStack
 - GKE, EKS, AKS
- Platform as a Service (PaaS)
 - AWS Lambda
 - Azure App Service
- Software as a Service (SaaS)
 - AWS Route53
 - Oracle Autonomous Data Warehouse Cloud



Cloud overview – which?



Cloud overview – which?

Clouds

Private cloud

- OpenStack
- RKE on OpenStack

Public clouds

- GCP, AWS, MSA, etc.

Different & changing all-the-time

- Different capabilities
- Very different costing structures

Cloud Strategy

Cloud APIs

- Vendor API – latest & greatest but vendor lock-in
- Open source API – less up-to-date but more cloud-agnostic
- Both – cloud centric

Pipelines in clouds

- Cloud-agnostic
- Cloud-native
- Easy-to-use for programmers and researchers

Solution



Cloud overview – which?

Clouds

Private cloud

- OpenStack
- RKE on OpenStack

Public clouds

- GCP, AWS, MSA, etc.

Different & changing all-the-time

- Different capabilities
- Very different costing structures

Cloud Strategy

Cloud APIs

- Vendor API – latest & greatest but vendor lock-in
- Open source API – less up-to-date but more cloud-agnostic
- Both – cloud centric

Pipelines in clouds

- Cloud-agnostic
- Cloud-native
- Easy-to-use for programmers and researchers

Solution

Docker & K8S

- Well-known & proven
- Advantages over VMs
- New paradigm

CI/CD toolchains

- For best practice

Cloud consultancy

- Architectural advises
- Cloud-native designs



Cloud overview – container & orchestration

Docker & Kubernetes

- De-facto standards of runtime & orchestration
- Docker
 - Runtime architecture
 - Packaging tool
- Kubernetes
 - Orchestration engine

Benefit over VMs

- Light-weight
- Very high portability
- Seamless integration with CI/CD
- Across hardware boundaries
- Portability, scalability, high availability, disaster recovery & maintainability

Growing pains

- More difficult to use
- Dependent on VMs in some clouds
- Tricky integration with POSIX filesystems

Best practices

- KISS principle
- Security
 - Official Docker images
 - Non-root ID
- Compute, data & configuration
 - Stateless container
 - Data on storage
 - StatefulSet for configuration



Cloud overview – container & orchestration

Docker & Kubernetes

- De-facto standards of runtime & orchestration
- Docker
 - Runtime architecture
 - Packaging tool
- Kubernetes
 - Orchestration engine

Benefit over VMs

- Light-weight
- Very high portability
- Seamless integration with CI/CD
- Across hardware boundaries
- Portability, scalability, high availability, disaster recovery & maintainability

Growing pains

- More difficult to use
- Dependent on VMs in some clouds
- Tricky integration with POSIX filesystems

Best practices

- KISS principle
- Security
 - Official Docker images
 - Non-root ID
- Compute, data & configuration
 - Stateless container
 - Data on storage
 - StatefulSet for configuration



Cloud overview – container & orchestration

Docker & Kubernetes

- De-facto standards of runtime & orchestration
- Docker
 - Runtime architecture
 - Packaging tool
- Kubernetes
 - Orchestration engine

Benefit over VMs

- Light-weight
- Very high portability
- Seamless integration with CI/CD
- Across hardware boundaries
- Portability, scalability, high availability, disaster recovery & maintainability

Growing pains

- More difficult to use
- Dependent on VMs in some clouds
- Tricky integration with POSIX filesystems

Best practices

- KISS principle
- Security
 - Official Docker images
 - Non-root ID
- Compute, data & configuration
 - Stateless container
 - Data on storage
 - StatefulSet for configuration



Cloud overview – container & orchestration

Docker & Kubernetes

- De-facto standards of runtime & orchestration
- Docker
 - Runtime architecture
 - Packaging tool
- Kubernetes
 - Orchestration engine

Benefit over VMs

- Light-weight
- Very high portability
- Seamless integration with CI/CD
- Across hardware boundaries
- Portability, scalability, high availability, disaster recovery & maintainability

Growing pains

- More difficult to use
- Dependent on VMs in some clouds
- Tricky integration with POSIX filesystems

Best practices

- KISS principle
- Security
 - Official Docker images
 - Non-root ID
- Compute, data & configuration
 - Stateless container
 - Data on storage
 - StatefulSet for configuration

