# 

# Hadoop Architecture Metamorphosis

Alex Campos, Digital Data Strategist ac@cloudera.com | @campossalex



### Let's think about analytics 15 years ago...

#### Needs







#### **Tech**

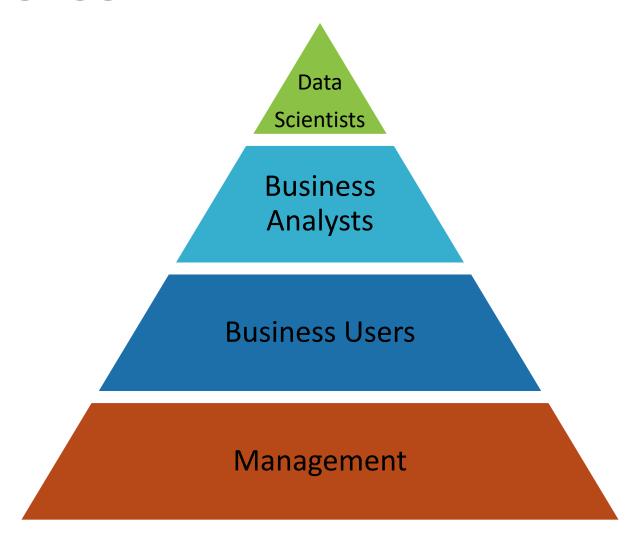




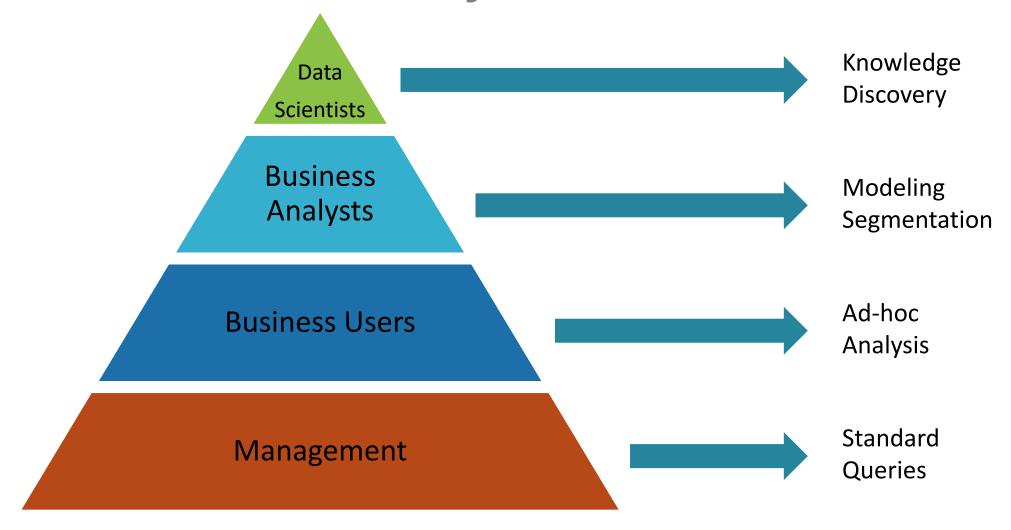


# How analytic looks like today?

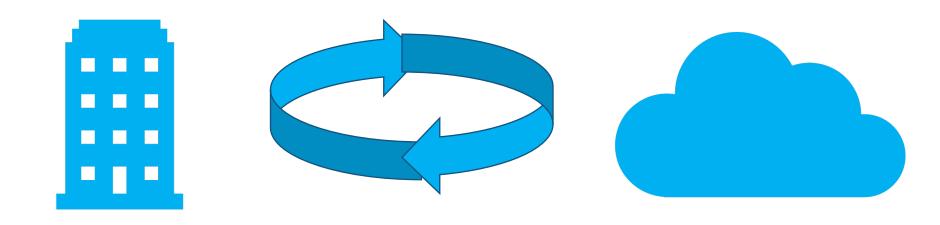
#### **Data Audience**



#### **Multi-functional Analytics**



#### We live in a hybrid world



#### **Technology Evolution**

#### CLOUD EXPERIENCE



Easy to use, self-service, on-demand, elastic, consumption

#### COMPUTE & STORAGE



Separation in public and private clouds for increased performance

#### **KUBERNETES**& CONTAINERS



Adoption as standard operating environment for flexibility and agility

### Decoupling the coupled

#### **Monolithic Environments**

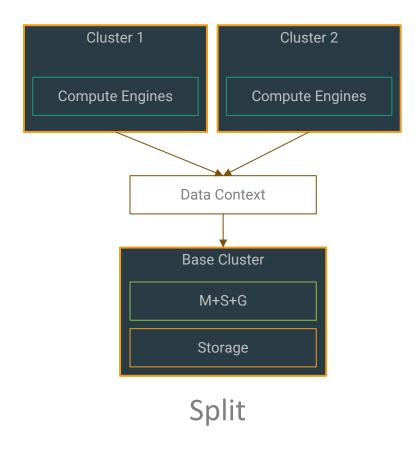


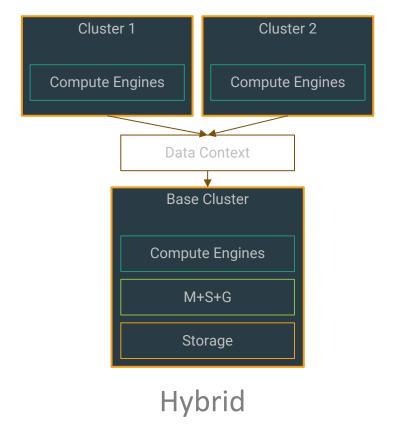
- Scaling out (horizontal)
- Data locality
- Storage and compute together

#### **Architecture Shift**

Self-Service **Data Science Business Operational Data** Intelligence **Applications** Data Analytics **Artificial Intelligence** and Reporting Dashboard, Reporting **Operational Data Data Science** Discovery, Analytics and Traditional BI **Advanced Analytics** Services **Data Exploration** Data Unification (Data Catalog, Governance, Security, Metadata) Data Analytic **Data Science Engines** Operational Warehouse **Model Pipeline Data Engines** Sandbox Data Repository (Data Layers) **Data Engineering Pipelines Streaming and Batch Data Ingestions** 

#### **Multi-cluster**





\* M+S+G = Metadata + Security + Governance

#### Choose the Right Infra

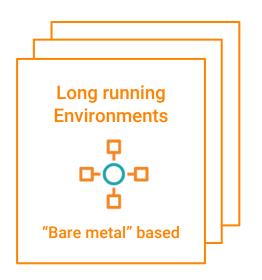
## Long running Environments

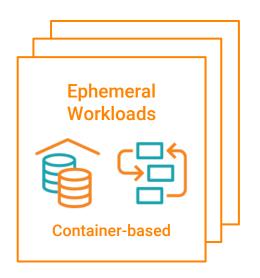
- Static resource allocation
- "Bare metal" based
- Traditional workloads
- More options for tuning

## Ephemeral Workloads

- Dynamic resource allocation
- Container-based
- Auto scaling
- Specific workloads

#### Right Infra





**Data Context** 

**Data Storage Layer** 

#### **Multi-tenancy Strategy**

#### Organizational

- Map human beings (user, team, department)
- Manage internal chargeback (billing)
- Isolation based on user or group experience

#### **Analytics Nature**

- Type of workload (adhoc, ML, query)
- Optimize resource/infrastructur e efficiency
- Isolation based on noisy neighbor

#### **SLA-oriented**

- Isolation based on workload priorities
- Trades strong SLA versus low-priority workloads

### Cloudera Data Platform

#### Cloudera Data Platform



From the Edge to Al

#### CDP, the first Enterprise Data Cloud



Hybrid, Multi-Cloud



Multi-Function & Open

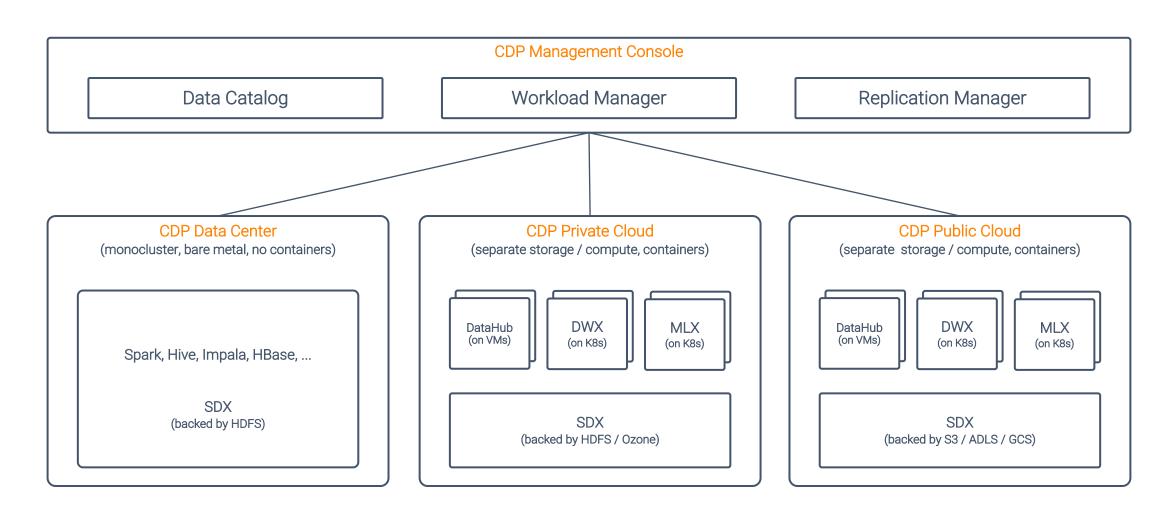


Secure & Governed



Cloud Experience

#### **Hybrid Multi-Cloud Architecture**



#### Thanks!!!

Alex Campos | ac@cloudera.com

https://www.linkedin.com/in/campossalex/ @campossalex