Zeta Architecture

Jim Scott – Director, Enterprise Strategy & Architecture

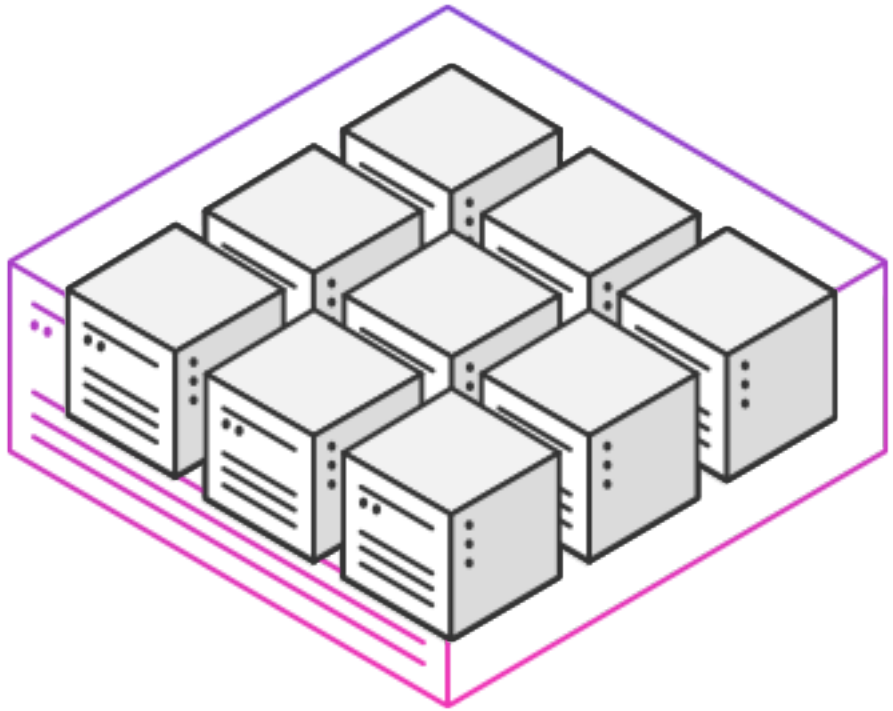
Agenda

- Current State
  - History
  - Moving Forward
- The Next Enterprise Architecture
- Business Implications
- Concrete Implementations

Current State

Study History to Prepare for the Future

- A data center was built
- The servers were statically partitioned
- If we want to break the cycle we have to break the partitions and become dynamic



Understanding the Why's

- Isolation of resources
  - Assists in troubleshooting
  - Prevents the analytics team from impacting production
- Maximum throughput of an application
  - Guaranteed volume (maximum): compute, memory and storage
- Business Continuity
  - We know exactly what is backed up, when, and where
  - Difficult to perfect and to test

Issues with Isolated Workloads

- Segregated servers lead to under utilized hardware
  - Wasted capacity and energy

- Complicated processes to move data to the required processing servers
  - Operational impact, including extra monitoring
  - Time delays moving data (not real-time)
  - Troubleshooting time when there are issues

- Difficult to thoroughly test DEV vs. QA vs. Production
  - Environments have different shapes and sizes
  - They will not have identical configurations

Goals Moving Forward

- Leverage all existing hardware
- Create isolation in a different way
- Improve production operational processes
- Fix process of moving from DEV to QA to Production
- Support **real-time** business continuity

The Next “Last” Enterprise Architecture

The Next Generation Enterprise Architecture

- Dynamic compute resources
- Common storage platform
- Real-time application support
- Flexible programming models
- Deployment management
- Solution based approach
- Applications to operate a business

** This is a pluggable architecture*

Technologies That Work

We Will Call This Architecture…

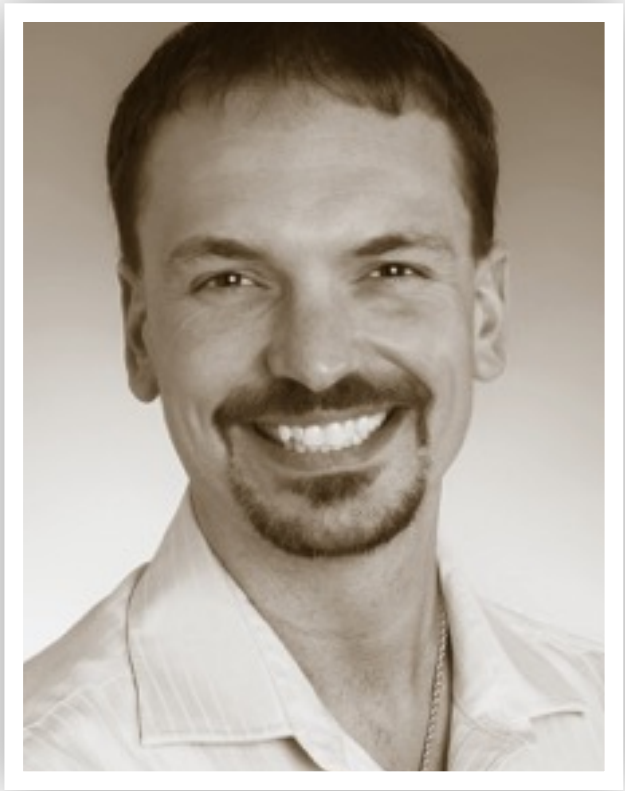
What's in a Name

- The letter Z is the last letter in the English alphabet, but Zeta is not the last letter of the Greek alphabet
  - But this is the last generalized architecture you will need.
- Sixth letter of the Greek alphabet
  - Hexagon represents the 6 surrounding pieces
- Zeta represents the number 7
  - 7 total components in this architecture
  - Components work with a global resource manager

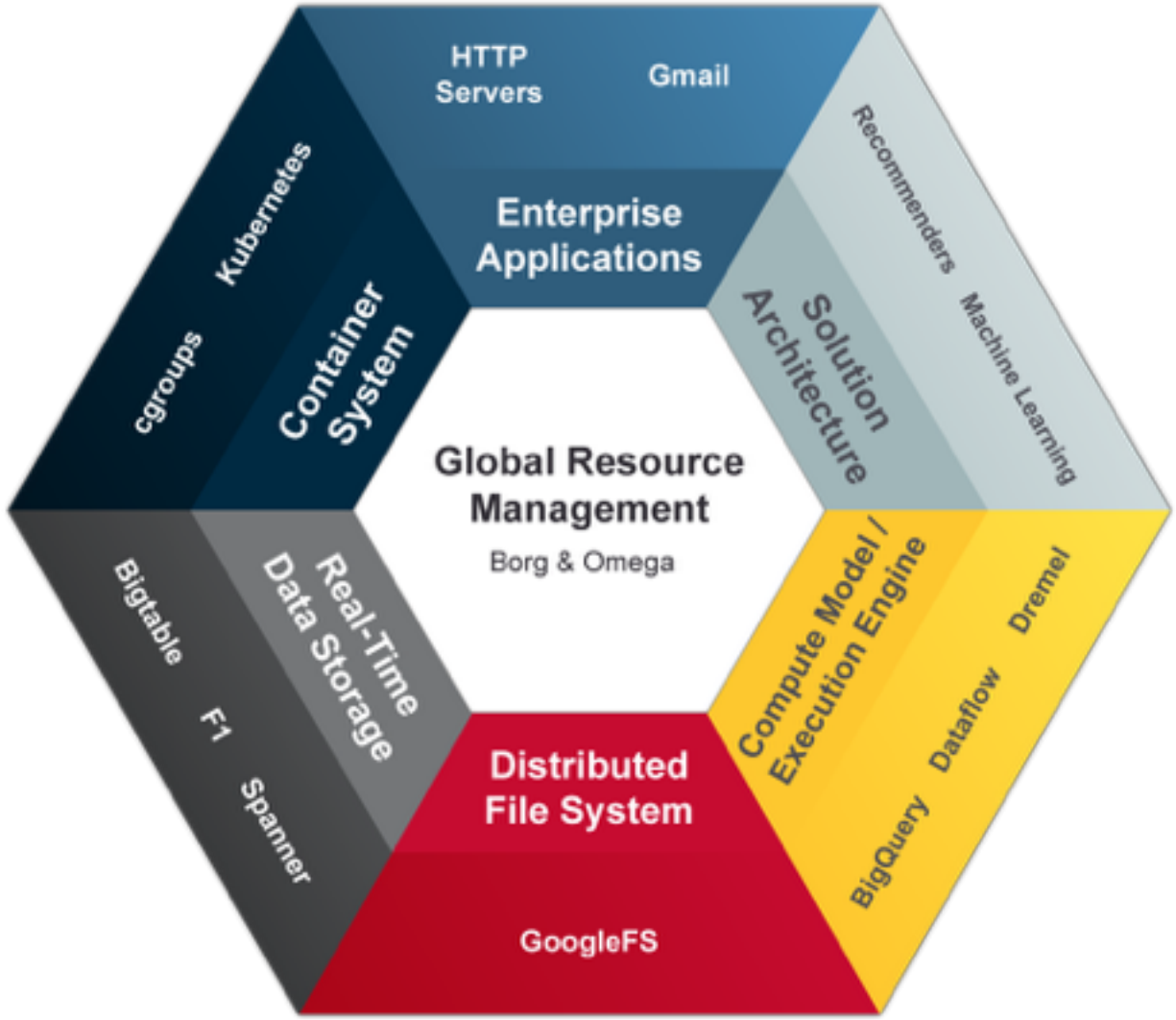
Origin Story of the Zeta Architecture

- Cultivated by Jim Scott
  - Created the pretty diagrams
  - Put a nice name on it
  - Documented the concepts

- Not really a new concept
  - Google pretty much pioneered these technology concepts
  - They have never really discussed it cohesively in this way



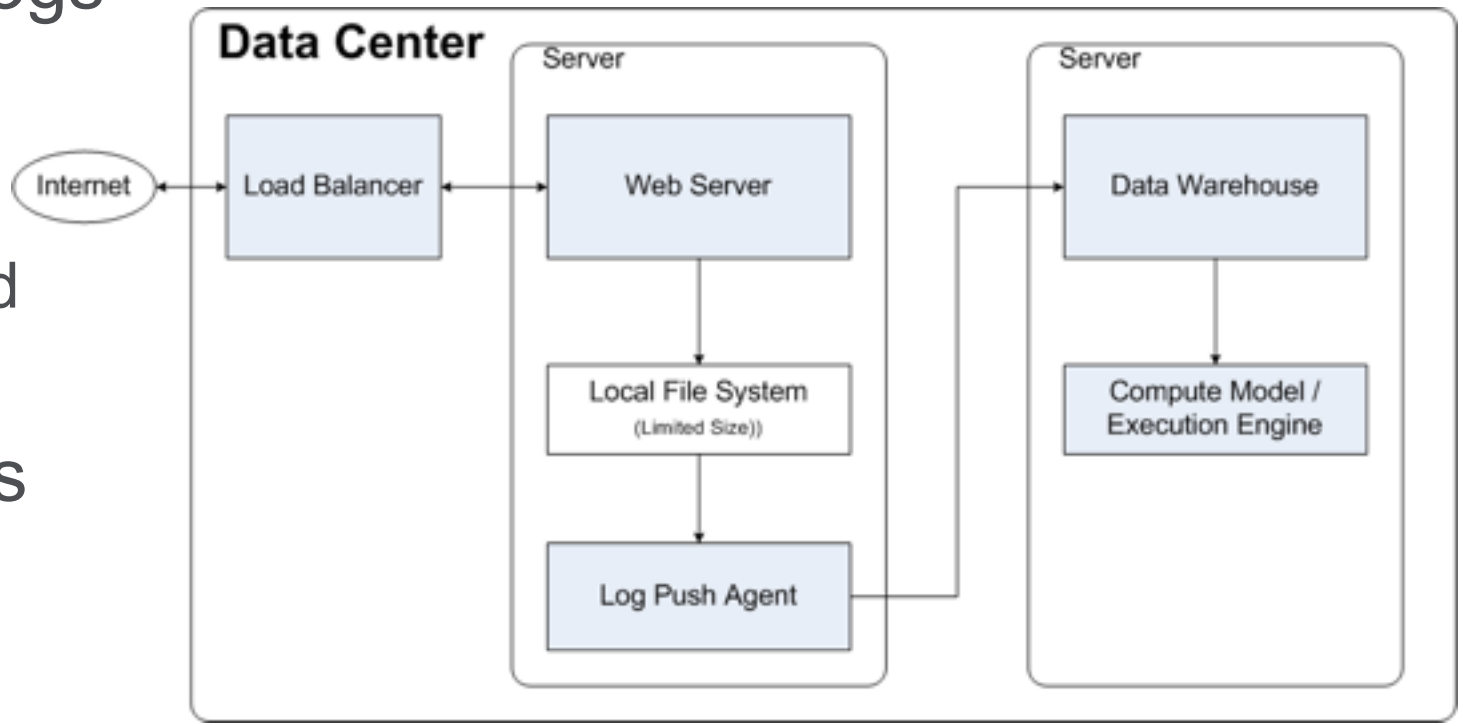
Zeta Architecture at Google



Concrete Implementations

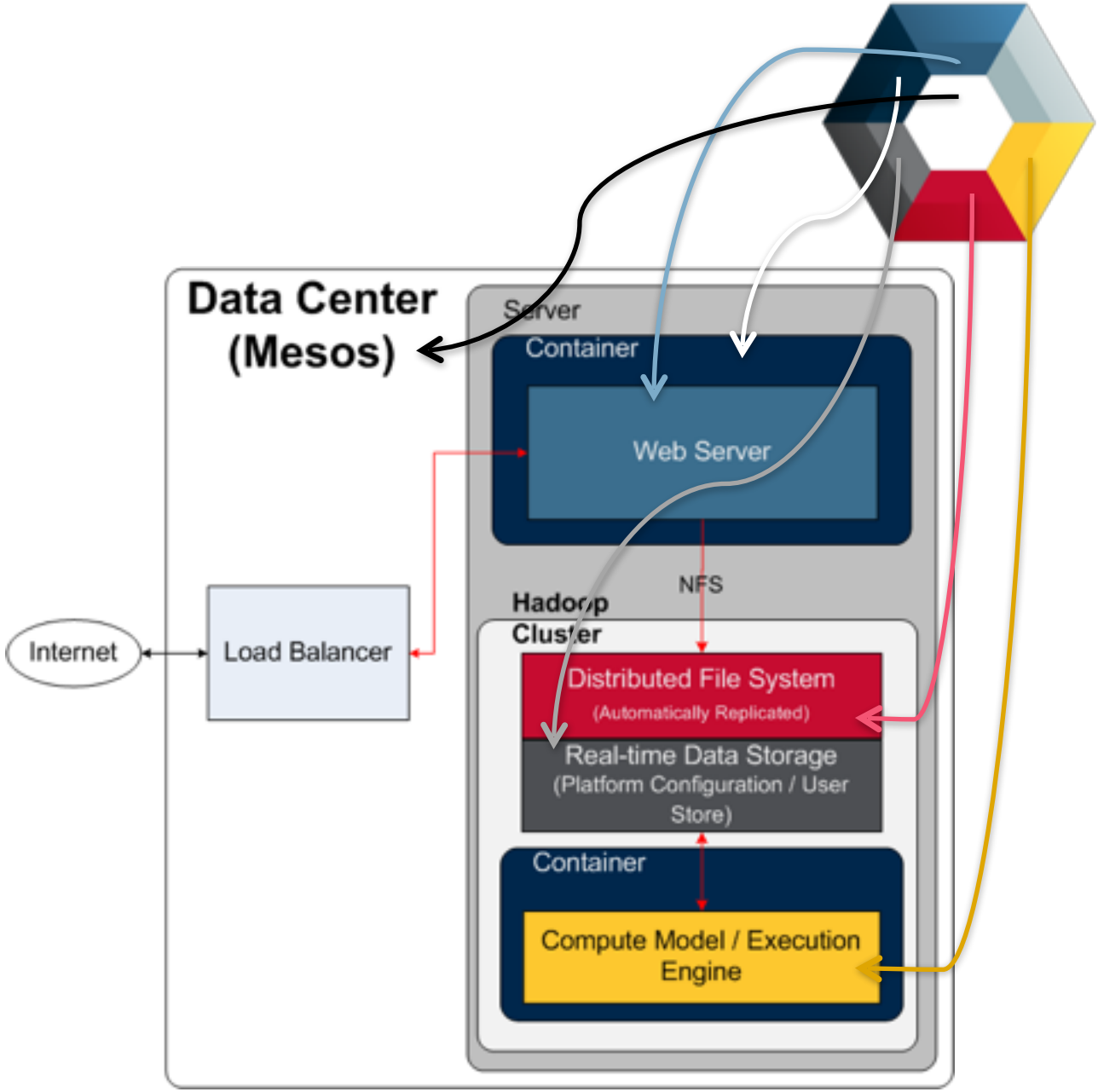
Web Server Logs

- Web server generates logs
- Land on local disk
  - Logs periodically rotated
- Shipped to other servers
- Run jobs on logs

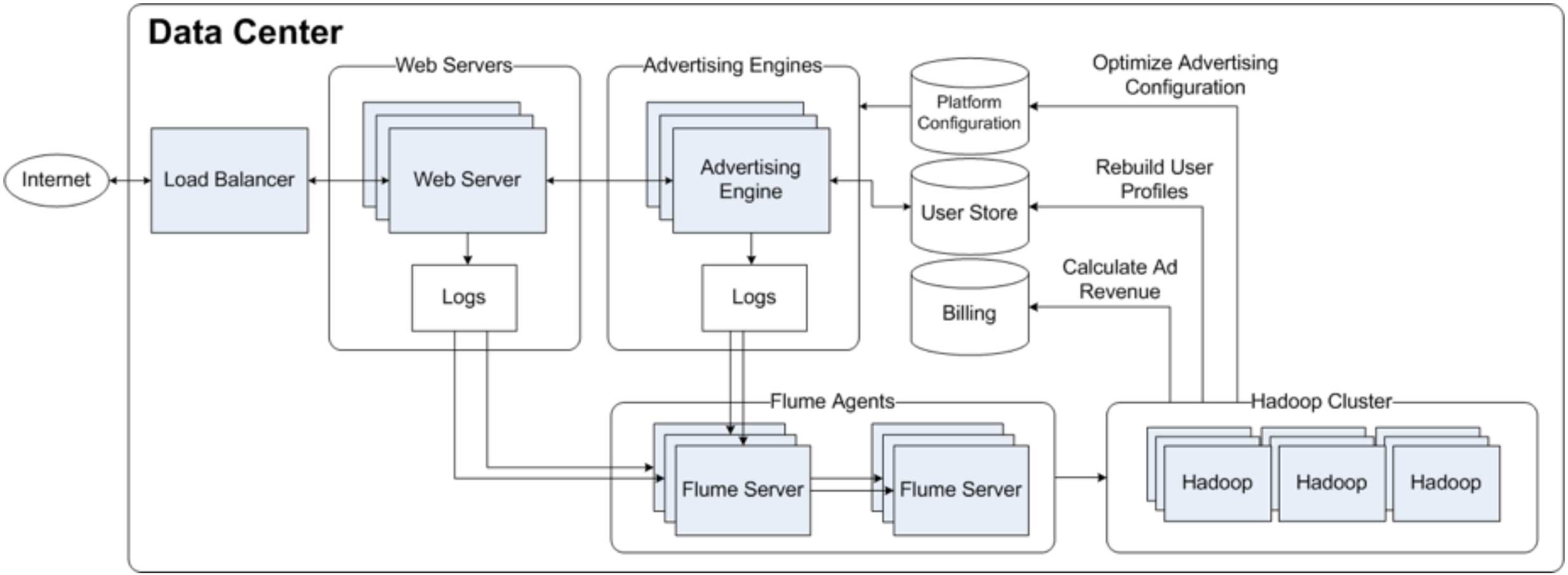


Web Server Logs

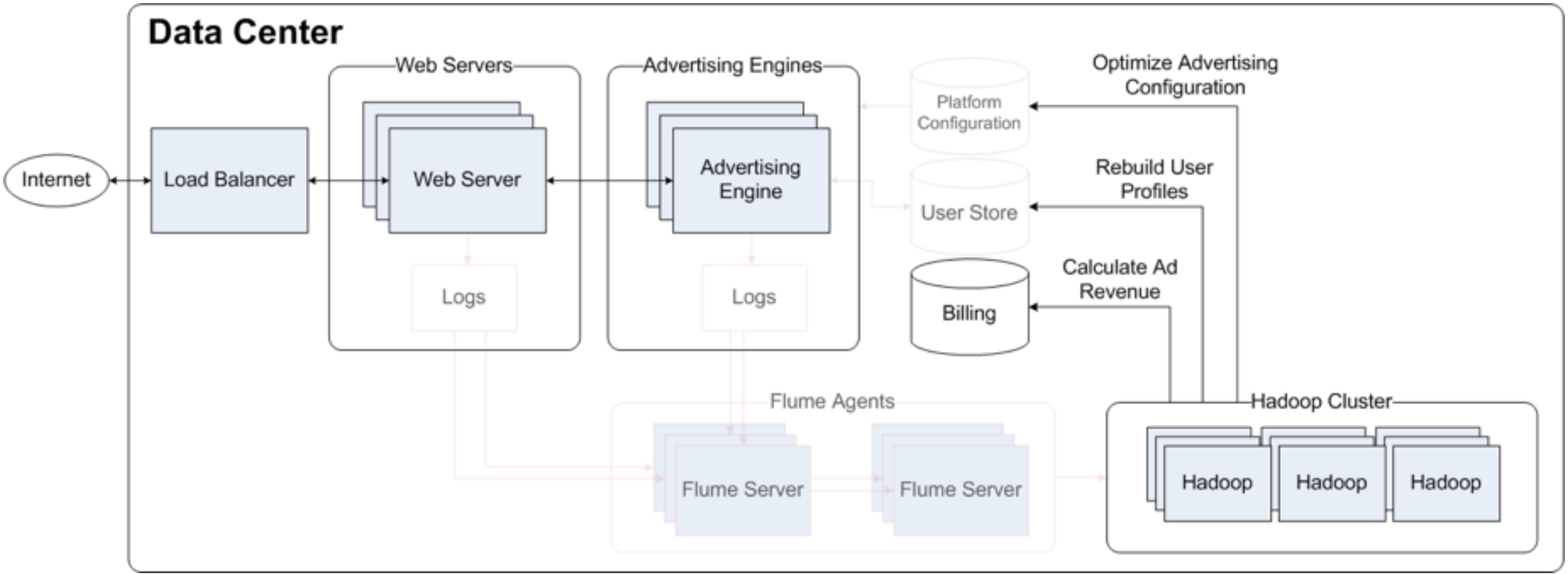
- Web server generates logs
- Land on DFS
  - Logs still rotate
  - Logs now tolerant of a server failure prior to rotation
  - Logs are instantaneously available for computation
- Run jobs on logs
  - Data locality



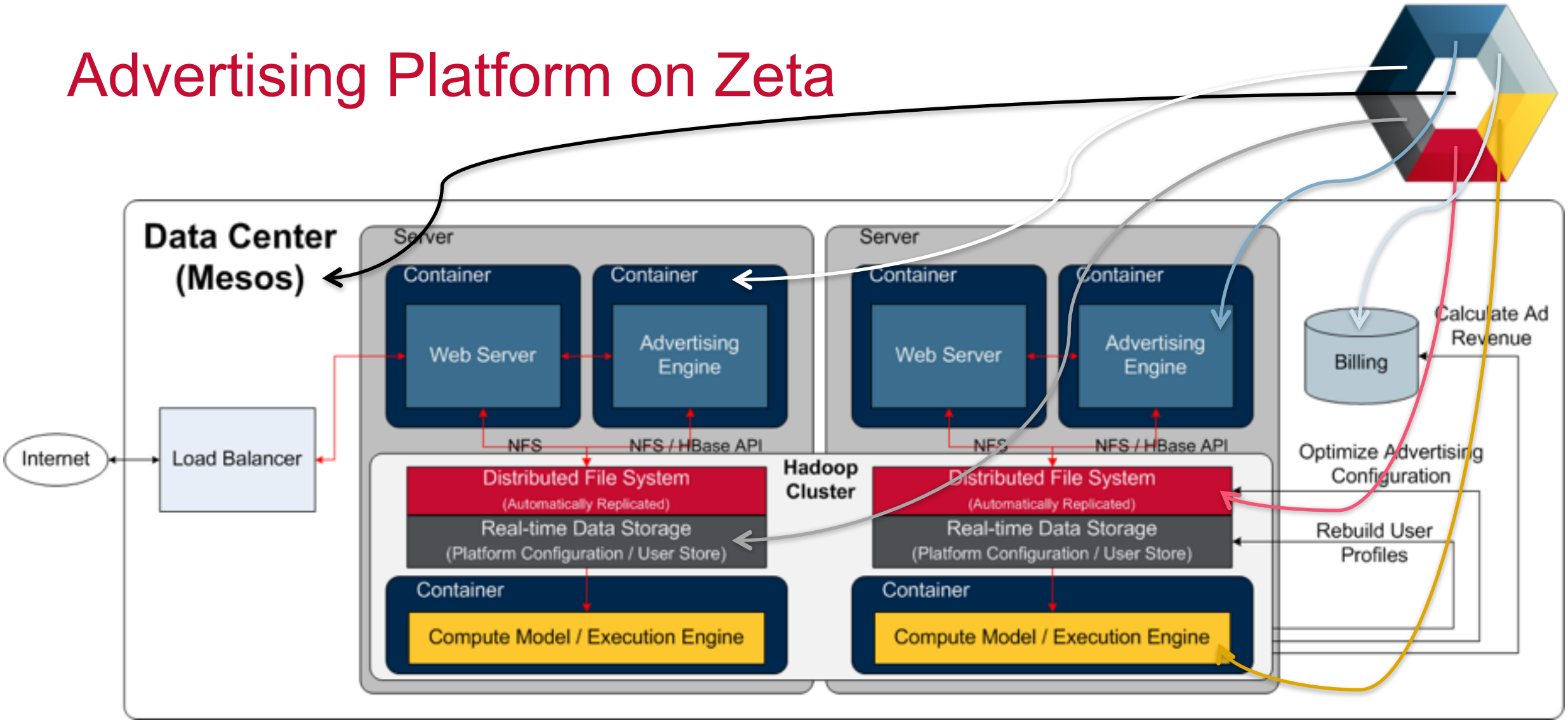
Advertising Platform



Advertising Platform - Simplified



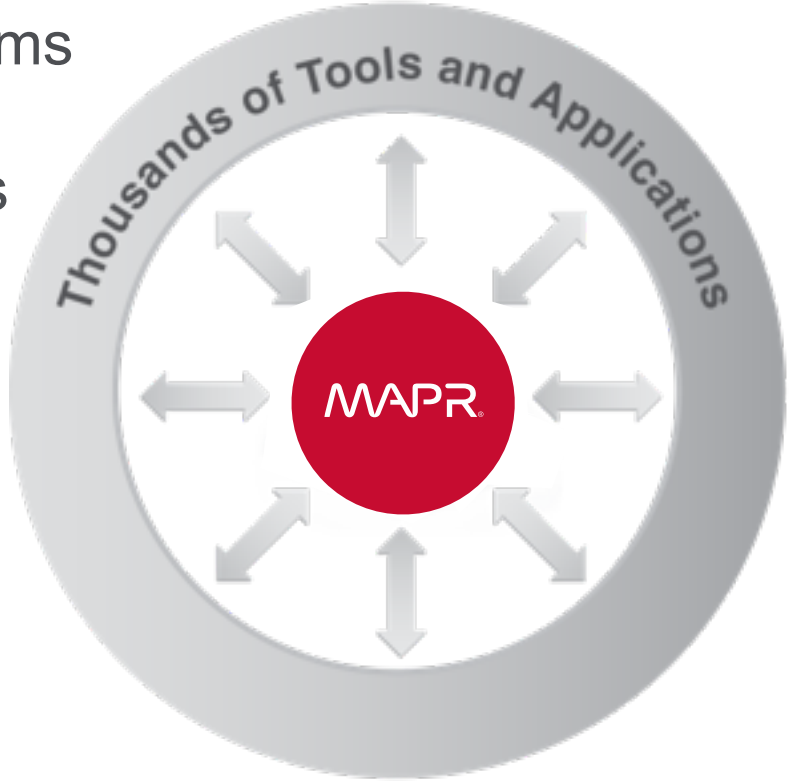
Advertising Platform on Zeta



Business Implications

Integration of Existing Systems

- Use standards like NFS to connect existing systems
- Pluggable security models fit into your companies current standards
- Database that have been tested to work
  - MySQL, Vertica, Sybase IQ
  - ElasticSearch, SOLR
- May work, but not tested
  - Oracle, DB2, SQL Server, PostgreSQL
    - Applications in this architecture can still use them
    - If they start supporting these technologies things change



Rethink the Data Center

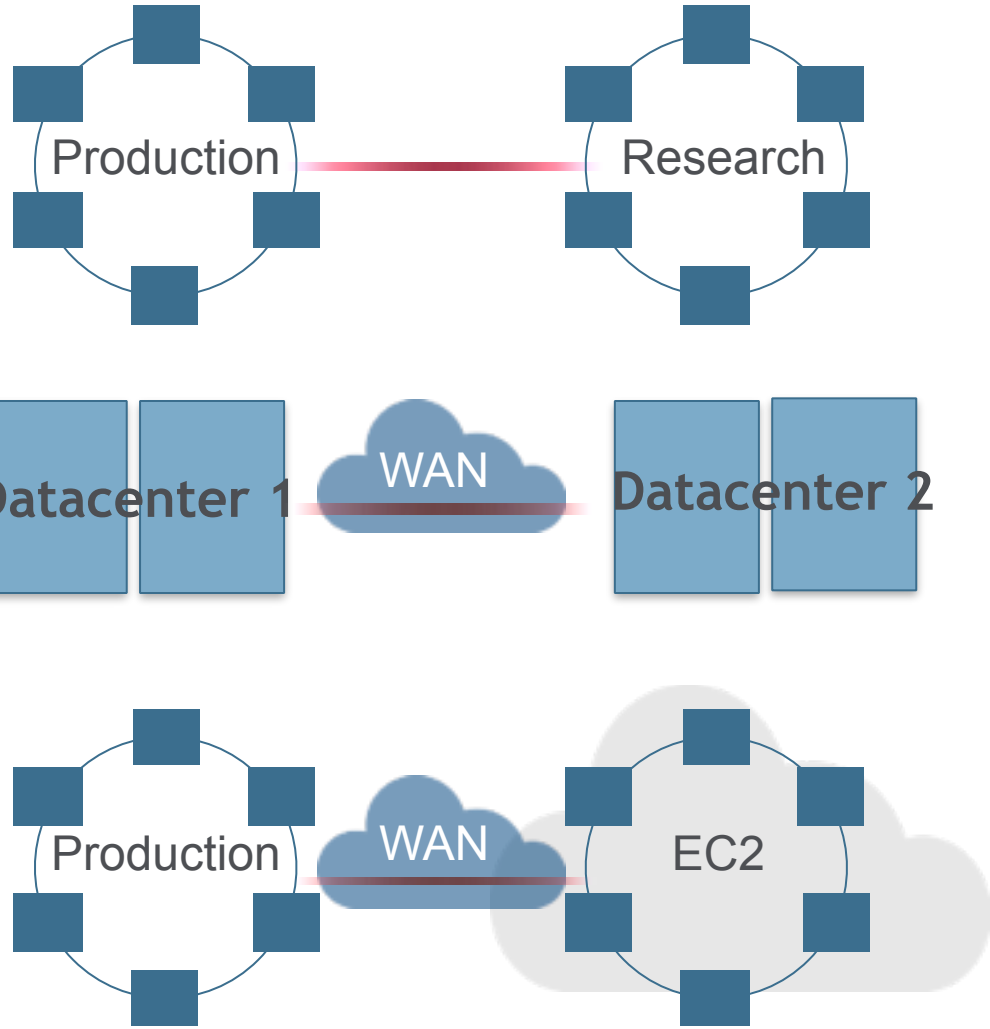
- All Servers
  - Run Mesos
  - Participate in the Distributed File System
- Dynamic Allocation of Resources
  - Spin up more web servers
  - Custom Business Applications
  - Big Data Analytics
- Data Locality
  - No more shipping data
  - Store and process the data where it was created

Simplified Architecture

- Less moving parts
  - Less things to go wrong
- Better resource utilization
  - Scale any application up or down on demand
- Common deployment model (new isolation model)
  - Repeatability between environments (dev, qa, production)
- Shared file system
  - Get at the data anywhere in the cluster
  - Simplifies business continuity

Business Continuity

- Resilience
  - Redundancy
  - High Availability
  - Spare Capacity
- Recovery
  - Snapshots
  - Disaster Recovery
- Contingency
  - Protect against the unforeseen
  - Multisite Capability



Platform-wide Security and Compliance

- Authentication, Authorization, Auditing
  - Users and jobs
  - All tiers
- Data protection
  - Wire-level encryption between servers
  - Masking
- Regulatory Compliance
  - Automatic expiration of “old” data
  - Data locality supported by distributed file system

Net Benefit

- Reduced operating expenses (OPEX)
  - Better utilization of available capacity and data center space
- Reduced capital expenses (CAPEX)
  - Less total hardware needed
- Improves time to market
  - Streamlined deployments
  - Environments become consistent and predictable
- Delivers a competitive advantage
  - Via platform scaling
  - Performance improvements

Recap

- Saves valuable time and money
- Enables stronger business continuity capabilities
- Google has been doing this for years
  - Real-time is the crux of everything Google does
- Time for the rest of us to operate at Google scale
  - The technologies are there and they play together nicely
  - Process changes must occur internally to achieve this architecture
- This approach will become the “traditional” way of thinking
  - Don’t get beat to it by your competitors

Go Forth and Implement the Zeta Architecture

Q&A

Engage with us!

@kingmesal

maprtech

mapr-technologies

MapR

jscott@mapr.com

maprtech