In-Memory Computing SUMMIT NORTH AMERICA 2018

Huawei's story of leveraging GridGain as a distributed caching service on its public cloud environment

Paul Chen

Chief Architect,

Cloud Services Research and Development,

Huawei Technologies Canada Lab

HUAWEI

Agenda

- Huawei Public Cloud Overview
- DCS Caching Architecture & Usage Patterns
- Caching Engines & Use Cases
- Public Cloud Caching Performance/Latency Summary
- Current Challenges
- Hybrid & Private Cloud Use Cases and Challenges
- Things to Explore

Huawei Public Cloud Overview

In-Memory Computing SUMMIT NORTH AMERICA 2018

Huawei Public Cloud Overview

Smart City | Digital manufacturing | Internet finance | e-Government

SAP Infosys sopra esri ptc esri get it right

60+ Solutions

SAP On Cloud | HPC | Web& Mobile | Cloud Office | IoT | Cloud Communication | … | Dedicated IT hosting | Cloud DR | Migration | FCS

14 categories 100+services

Database	Application	EI	Enterprise application	Cloud communication	DevCloud	Video	IoT
Computing	Storage	Network	Security	Management and deployment	Development and testing		

Chip | Server | Storage | Network | Software (openstack)

Co-operation

860+ solution partners for business innovation, and **2900+ service partners** for E2E services including consultancy, deployment and O&M

Solutions

Generic-specific solution, to adapt to industry business and optimize services

Services

High-performance ECSs and BMSs guarantee cloudification of critical businesses.

Heterogeneous computing capacity supports artificial intelligent applications.

Enterprise-class storage, DB, and data analysis services deeply dig into values of data.

Security: Anti-DDoS, WAF, and DBSS guarantee business security.

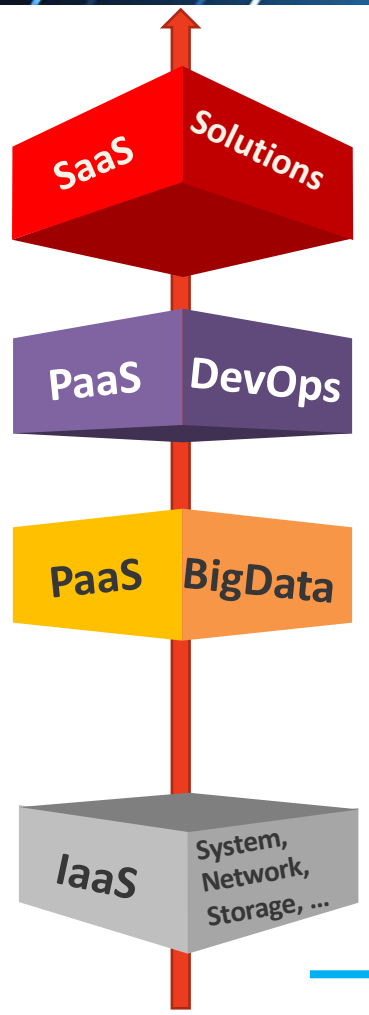
Software & Hardware

Atlas heterogeneous hardware, HPC, AI, and latest GPU and FPGA improve the computing capability.

Customized CPU, NVMe SSD card, smart NIC, RDMA, InfiniBand network and security chipset

In-Memory Computing SUMMIT NORTH AMERICA 2018

Huawei Cloud Services



Enterprise Apps: Workspace

Enterprise Cloud Comm.: IM, Meeting, CloudIPCC, VoiceCall, MSGSMS

IoT: IoTPlatform

Management & Deployment: CRS, IAM, ICMS, LTS, CTS

App Builder: WBS, LSB, MBS, SBS, DMB, RBS, WES, RES, AMS

Application: SMN, ServiceStageFunctionSt...

Dev Cloud: Collaboration, CloudIDE, CloudRelease, CloudBuild, CodeCheck, CloudPipeline, MobileTest, ProjectMan, CodeHub

Data Analysis: MRS, DPS, DIS, MLS, DWS, CDM, UQuery, OCR Forms

Database: DCS, RD...

Security: Anti-DDoS, AAD, WebScan, KMS, WAF, ARS, SSA, HIDS, SIS, SAS, HVD, SCS, HWAF, WTP, DBSS

Computing: ECS, AS, IMS, BMS, DeC, DeH, CCE

Storage: EVS, OBS, VBS, DES, SFS, DESS, CDN, CSBS

Network: VPC, ELB, DNS, DirectConnect, VPN

PaaS
DCS
Distributed Caching Services

7 categories 18 services — Sept. 2015

10 categories 45 services — Dec. 2016

13 categories 85+ services — Sept. 2017

14 categories 100+ services — Jan. 2018

In-Memory Computing SUMMIT NORTH AMERICA 2018

Architecture & Usage Patterns

In-Memory Computing SUMMIT NORTH AMERICA 2018

DCS Architecture

App developers

Manage my caching instance

DMZ

Caching Service Dashboard

PRV

Caching Service Broker (manager)

Resource Scheduling & Deployment

Caching service providers

Provision service instances

Apps

uses

Resources are isolated per tenant

Tenant Resources

VMs

Shared Resources

Bare-metal (x86/ARM)

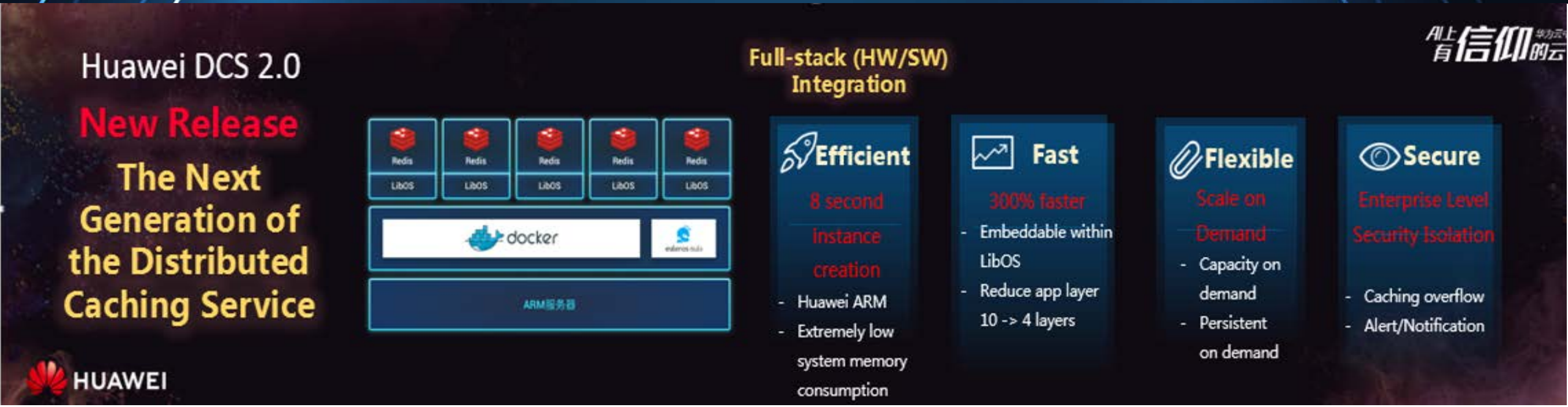
Caching Engines

- GridGain
- Redis

Horizontal Scale On-demand

App users

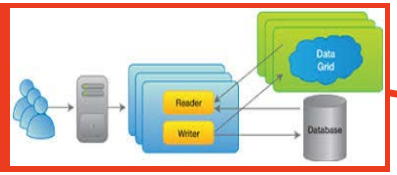
DCS 2.0 Released



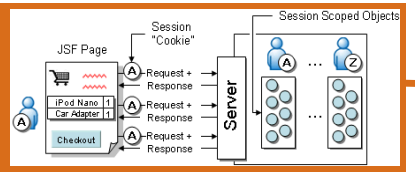
- Faster, more flexible and more secure
  - 8 seconds to create a caching instance
  - Caching operations 300% faster (leveraging seamless HW/SW/OS integration)
  - Scale on demand (add new caching capacity dynamically)
  - Strong Security: strong multi-tenant isolation; SLA warranty via caching overflow, cache persistency and alert/notification

Caching Usage Patterns

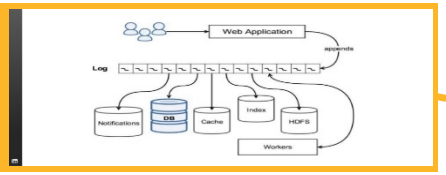
01 Side Cache



02 HTTP Session Replication



03 Change Data Capturing



04 Write-through/Write-behind/Map-reduced

05 SQL-like Query

```
lower(string) LIKE 'o%';
string LIKE '_n_';
string LIKE '_n%';
string LIKE '%';
string LIKE '%\%';
string LIKE '%\_%';
```

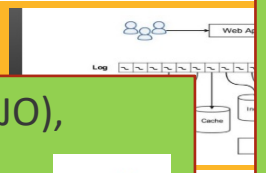
DCS Caching Usage: Side Cache

01 Side Cache

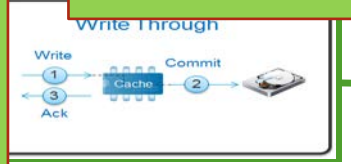
- e-commerce & Websites
- Public services
- Social media (e.g. feeds)
- Network games
- Search engines

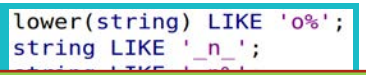
02 HTTP Session Replication

03 Change Data

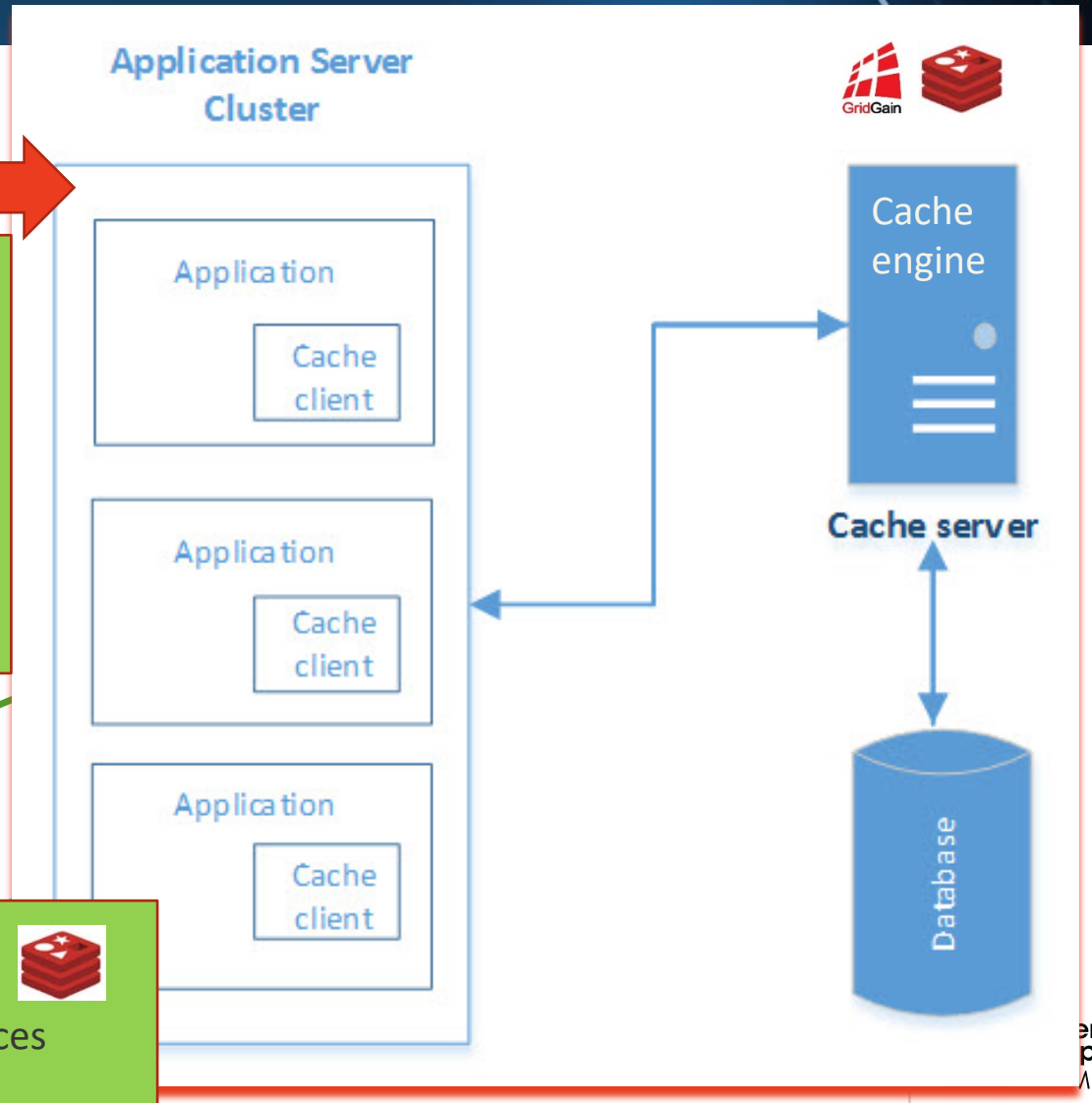


- Objects/Classes (e.g. POJO),
- SQL like queries
- Transaction controls
- Locking strategy control
- Multi-language supports
- Customization & Serializations



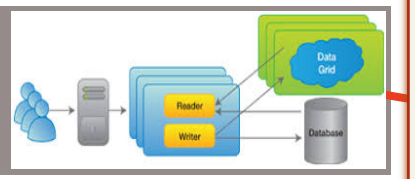


- Simple Key/Value
- Cloud native client
- Redis/Memcached Interfaces
- Redis objects (MSET e.g.)

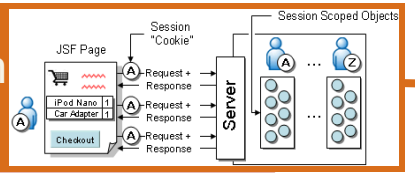


DCS Caching Usage: HTTP Session

01 Side Cache

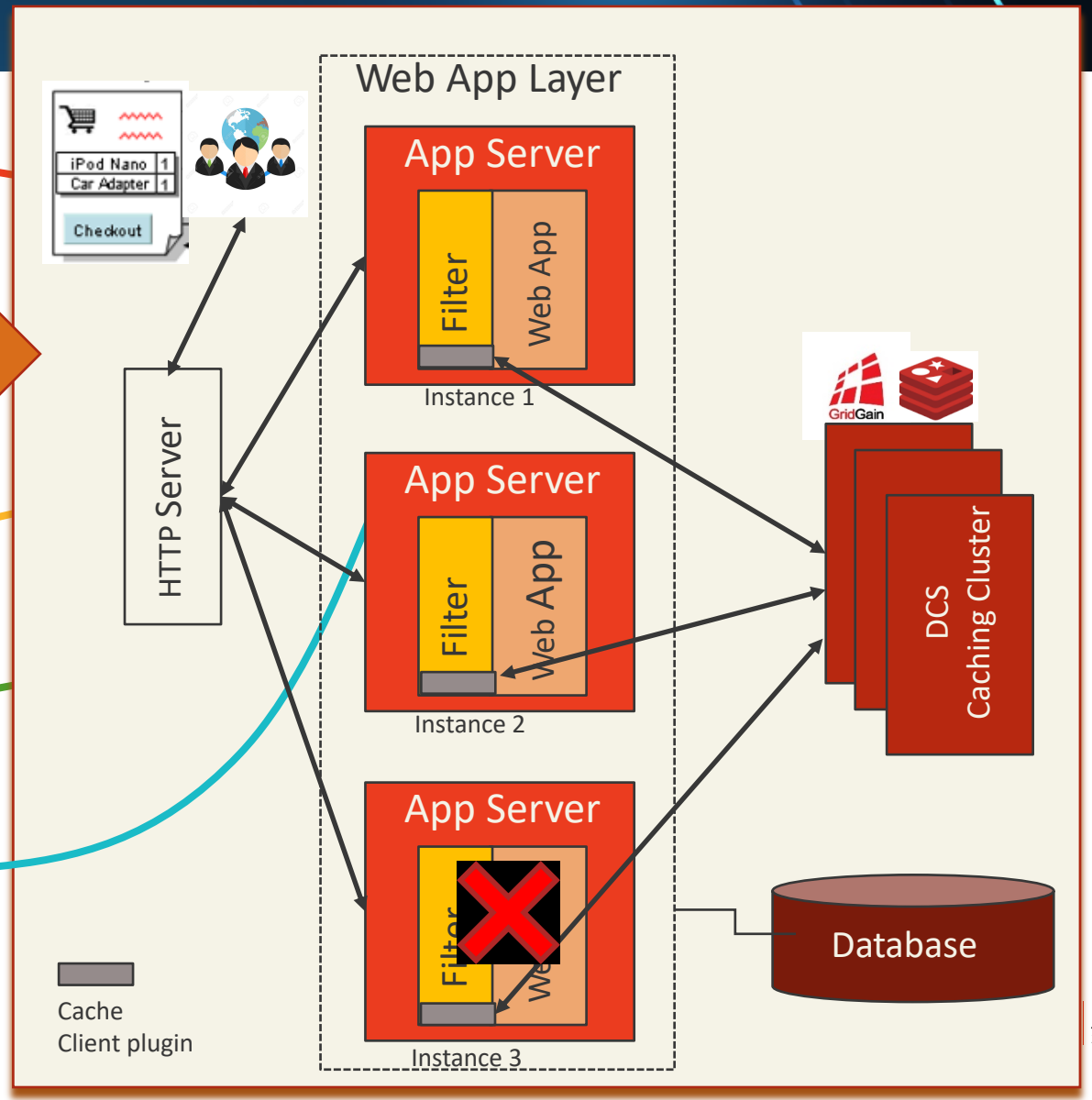


02 HTTP Session Replication



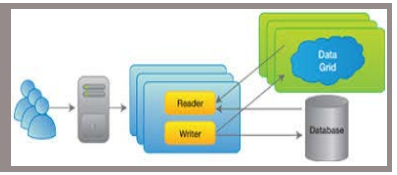
- Session persistent
  "HTTP Session objects cached on DCS"
  - User/login profile and session, user objects, session data (shopping cart, store catalogues, browsing histories ..)

- Session failover
  "App instances were down or restarted"
  - Survived from instance restart
  - Fast warm-up time

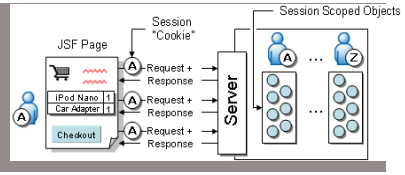


DCS Caching Usage: Data Grid

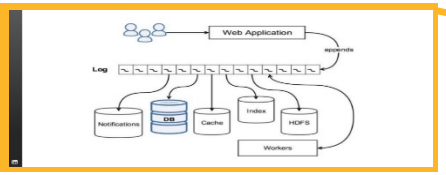
01 Side Cache



02 HTTP Session Replication



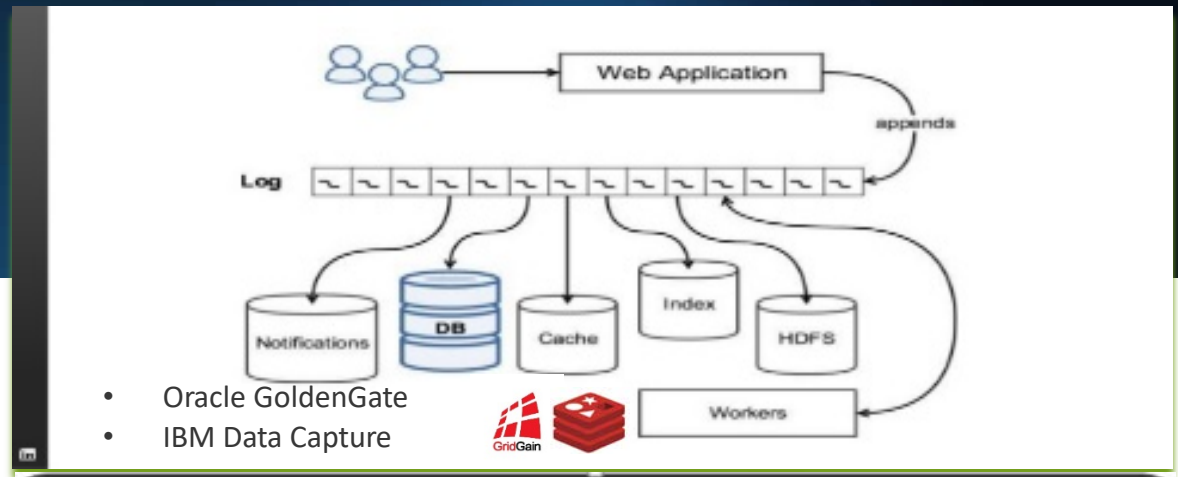
03 Change Data Capturing



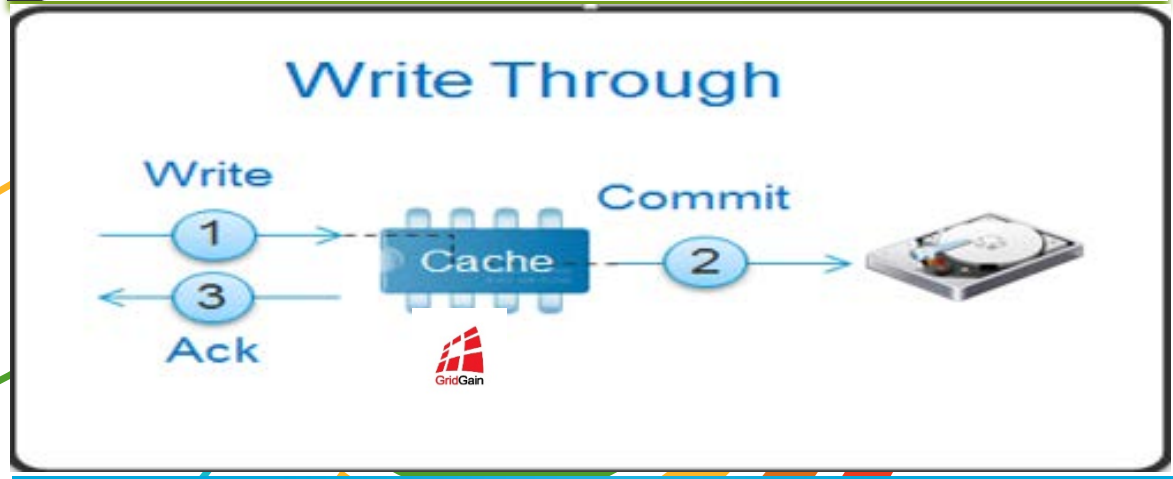
04 Write-through/Write-behind/Map-reduced

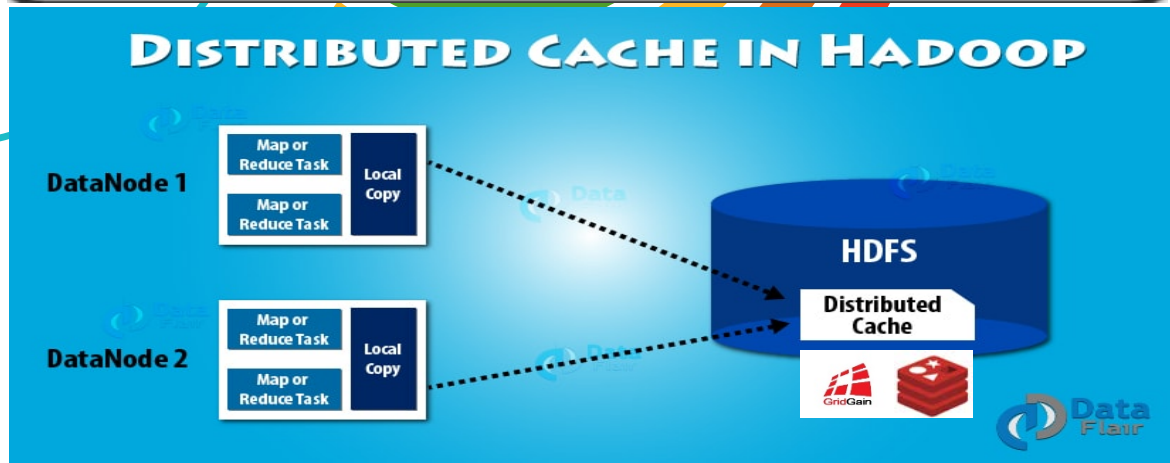
05 SQL-like Query

```
lower(string) LIKE 'o%';
string LIKE '_n_';
string LIKE '_n%';
string LIKE '%';
string LIKE '%\%';
string LIKE '%\_%';
```



- Oracle GoldenGate
- IBM Data Capture



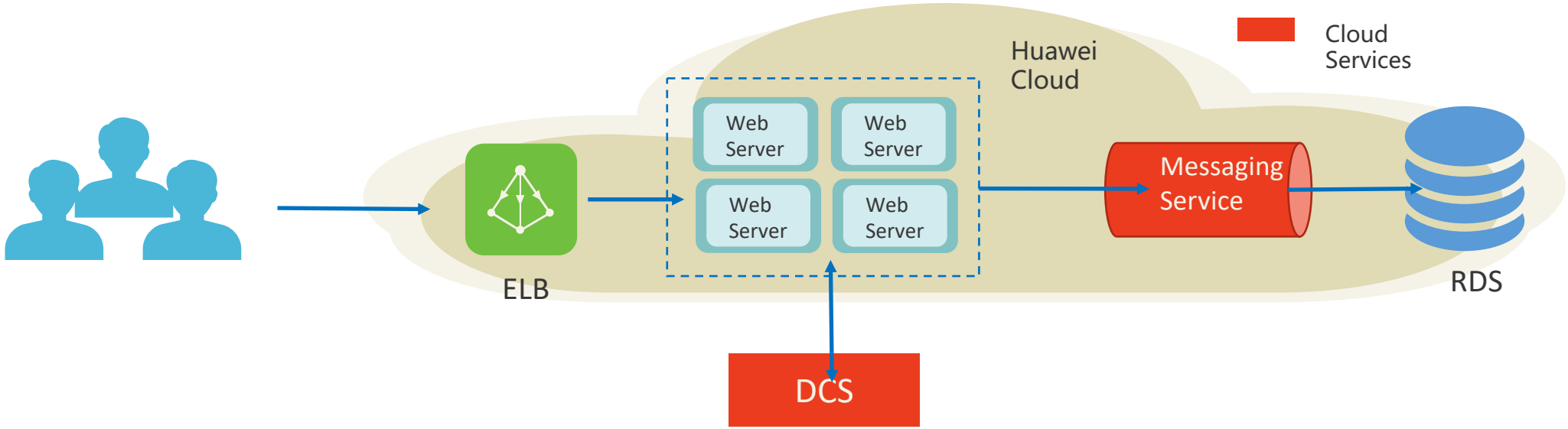


Engines & Use Cases

In-Memory Computing SUMMIT NORTH AMERICA 2018

DCS Use Case 1

- A public service agency (App was deployed on Huawei public cloud)
- > 50 ,000 concurrency => Database becomes a bottleneck
- Impact significantly on business during the request peak due to DB latency

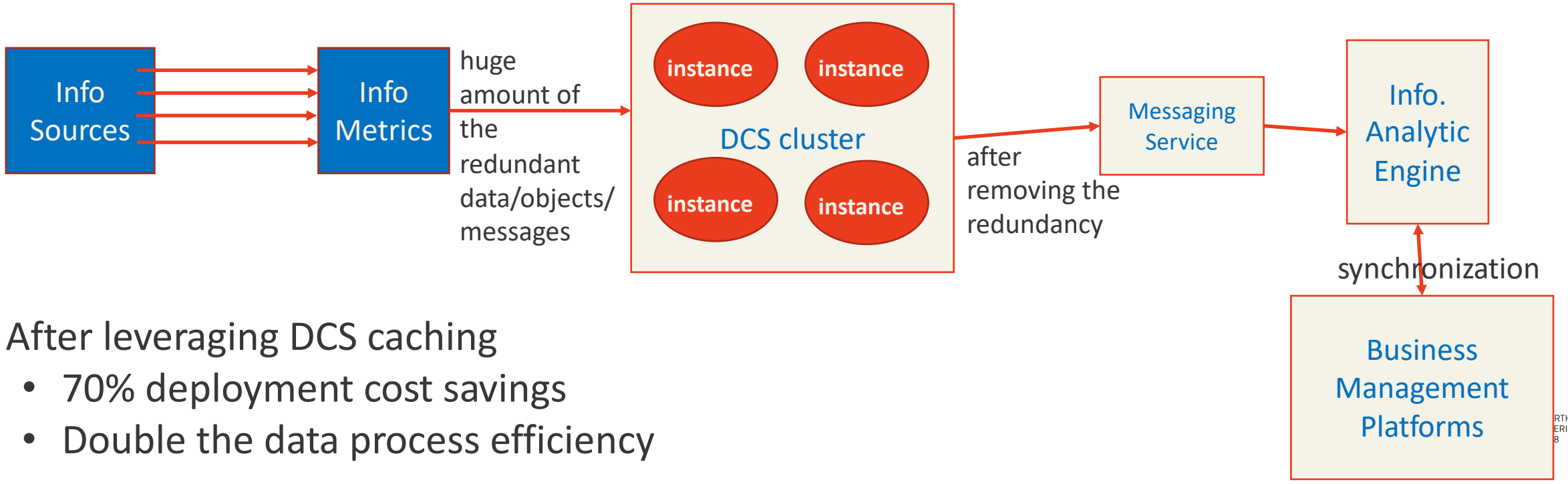


❑ After leveraging DCS caching

- Performance and concurrency improved 10 times

DCS Use Case 2

- A search engine provider (in Asian pacific)
- Huge amount of business data to collect and analyze (e.g. news, social media, blogs, chat groups, online forum…) – increase exponentially
- Large amount of collected data were redundant – significantly increase the process, modeling and analysis time – became “low performance” and “inefficient”



❑ After leveraging DCS caching

- 70% deployment cost savings
- Double the data process efficiency

DCS GridGain & Redis Engine Performance/Latency

- Clustered nodes
- 1 full async replica
- 9 million requests
- 1 K per object or value
- > 200 connections

Note: the following result is for reference purpose only – not for comparison)

- The different test tools used (Yardstick vs. memtier)
- The different cached objects measured (Java objects vs.. MSET)
- The different heap requirements (Java vs.. n+on-Java)

GridGain Engine (Enterprise v8.4.1)

Nodes	Replica	Threads	heap	Requests	CPU Usage %		*MEM Usage*		Network Mbps	Latency msec	Performance (Average per node)
					Driver	Server	Driver	Server			
Use Case 1：2 clients 1 replica，increases # of nodes and # of client connections											
9	1	360	8G	9000,000	498	252	1.56G	5.72G	60	2.01	95417

Redis Engine (v4.0.11)

Nodes	Replica	Threads	heap	Requests	Network Mbps	Latency msec	Performance (Average per node)
Use Case 1：2 clients 1 replica，increases # of nodes and # of client connections							
8	1	320	64G	1000,000		1.5	91795

Challenges

Challenges

- IMDG ecosystem buildup on public cloud
- Enterprise cloud transformation (private -> hybrid, private -> public cloud)
- Migration across the different cloud providers
- Smart cache (more reliable, predictable, intelligent, interoperable)
  e.g. user doesn't care what caching engines are used, but elastically picked by the intelligent behind the scene based on use cases
  (Redis engine ←→ GridGain engine)
- *Hardware optimization (FPGA, AEP …, Cache offload)*

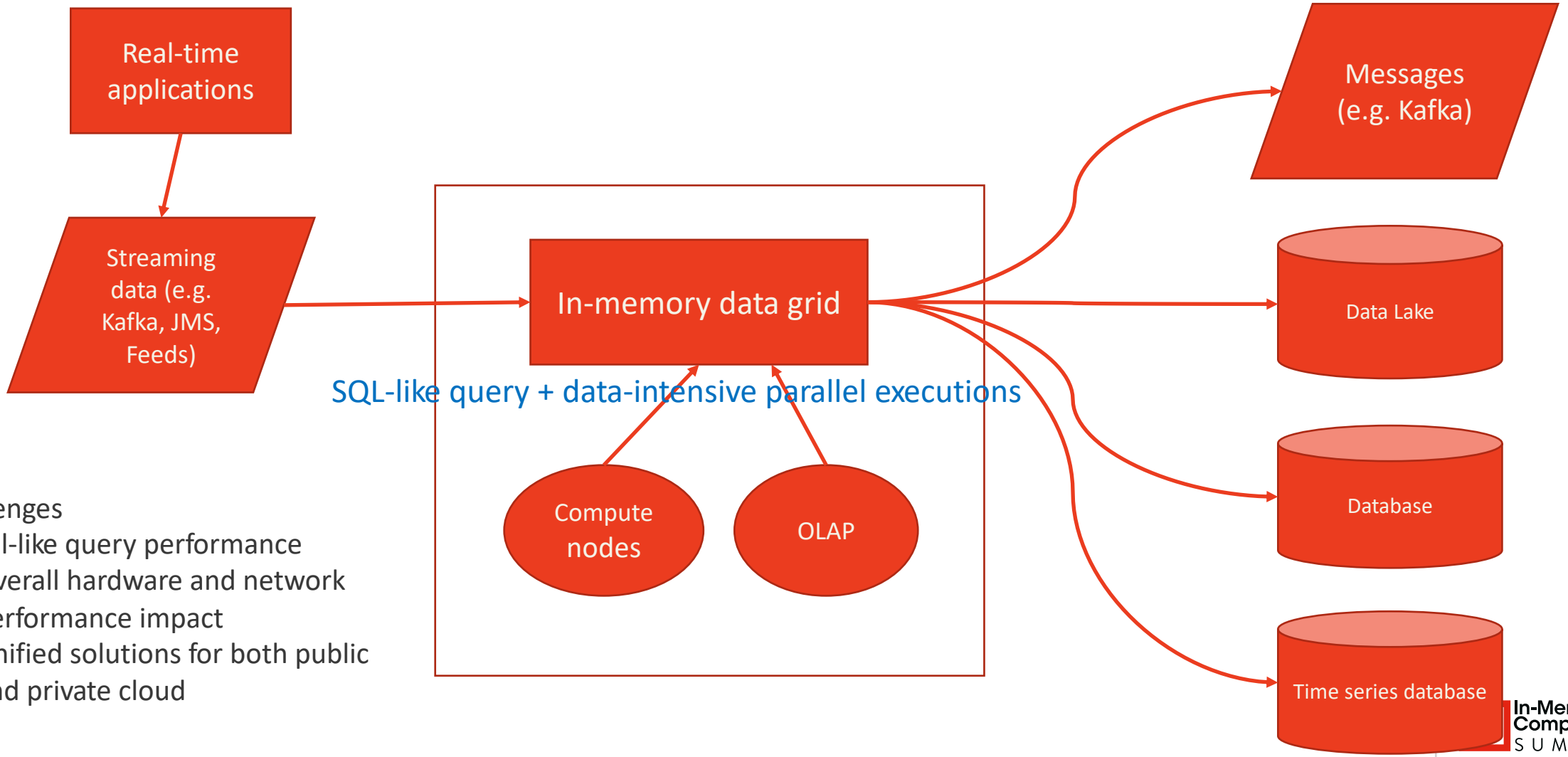
Things to Explore

In-Memory Computing SUMMIT NORTH AMERICA 2018

Things to Explore

- Write-through/Write-behind
- Data change capturing
- Smart cache (OLAP,,. Caching streaming data and real-time data analytics)
- Migrate caching services seamlessly from one cloud provider to another
- AEP (non-volatile memory (NVM) technology)

Private/Hybrid Cloud Use Cases & Challenges



Challenges

- sql-like query performance
- Overall hardware and network performance impact
- Unified solutions for both public and private cloud

In-Memory Computing SUMMIT NORTH AMERICA 2018

Thank you!

HUAWEI