

IT R&D Global Leader

Advanced Multitenant Analytics Framework Supporting Big Data Sharing and Utilization

2016 . 12 . 01.

Heesun Won

Contents



1

Motivation

2

Project Introduction

3

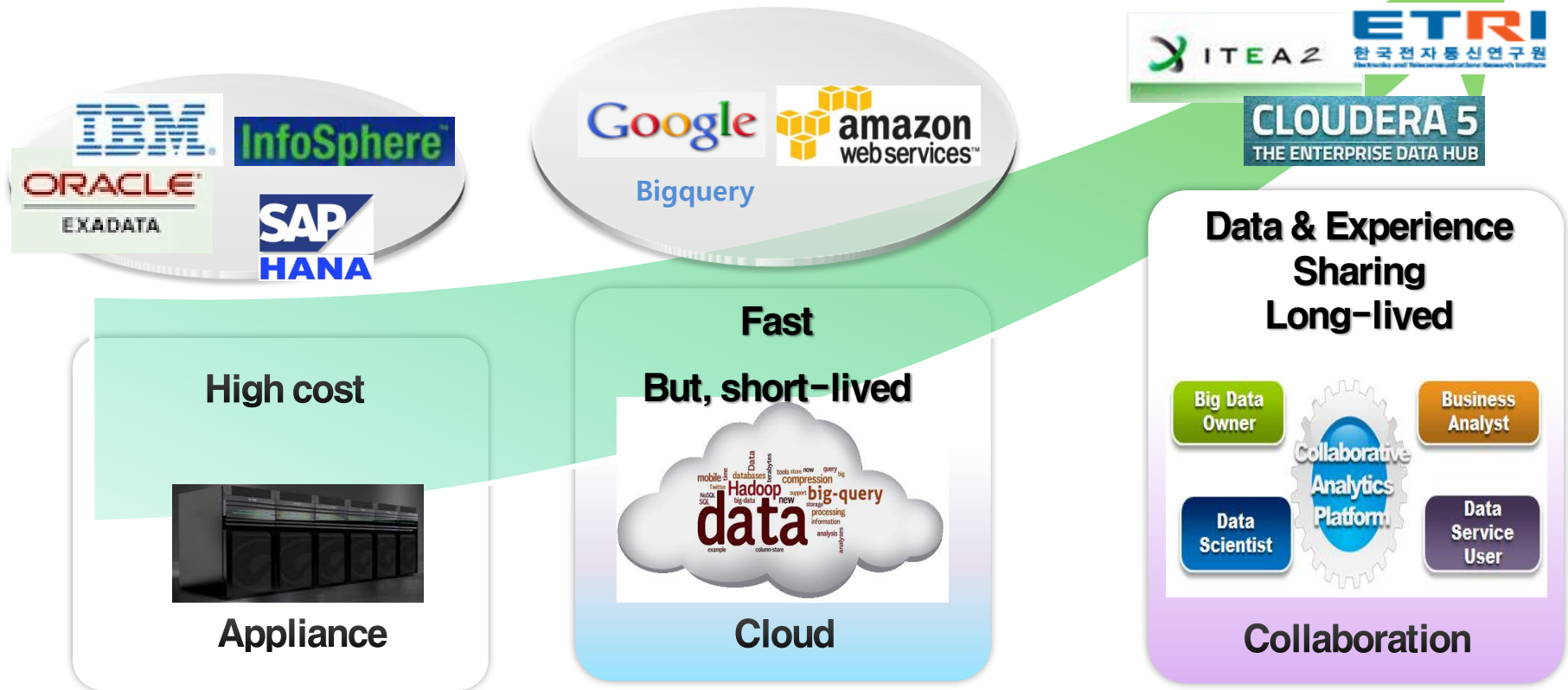
Technology Details

4

Demonstration

Motivation

- Barriers for using big data
 - ▶ Big data cluster is expensive to construct and difficult to operate efficiently.
 - ▶ Data is insufficient and not shared.
- The solution is big data as a service at a low cost via cloud



Needs

- Big data service via cloud environment at a low cost
- Secure and shared access to data, services, ...
- Easy to use tools and data portal



[Big data service concept]

Project Objectives



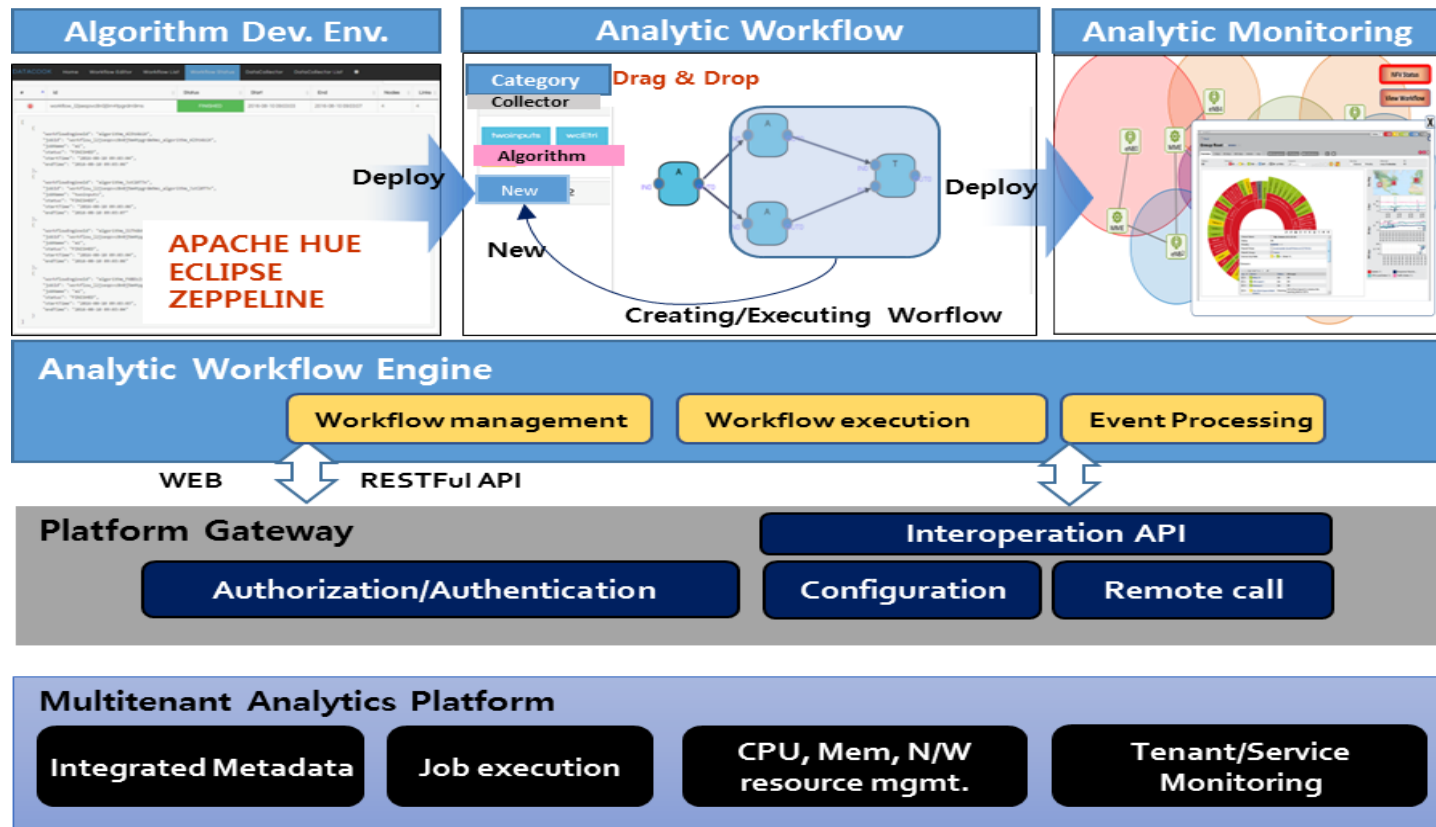
Development of Multitenant Analytics Framework for Big Data Service Supporting Data Storing, Managing and Analyzing on Cloud Environment

- ✓ A generic PaaS architecture
- ✓ Access point to data community
- ✓ Lowering investment entry barrier to big data processing
- ✓ Secure and privacy ensured access to data
- ✓ Open catalog of data services and analytics software

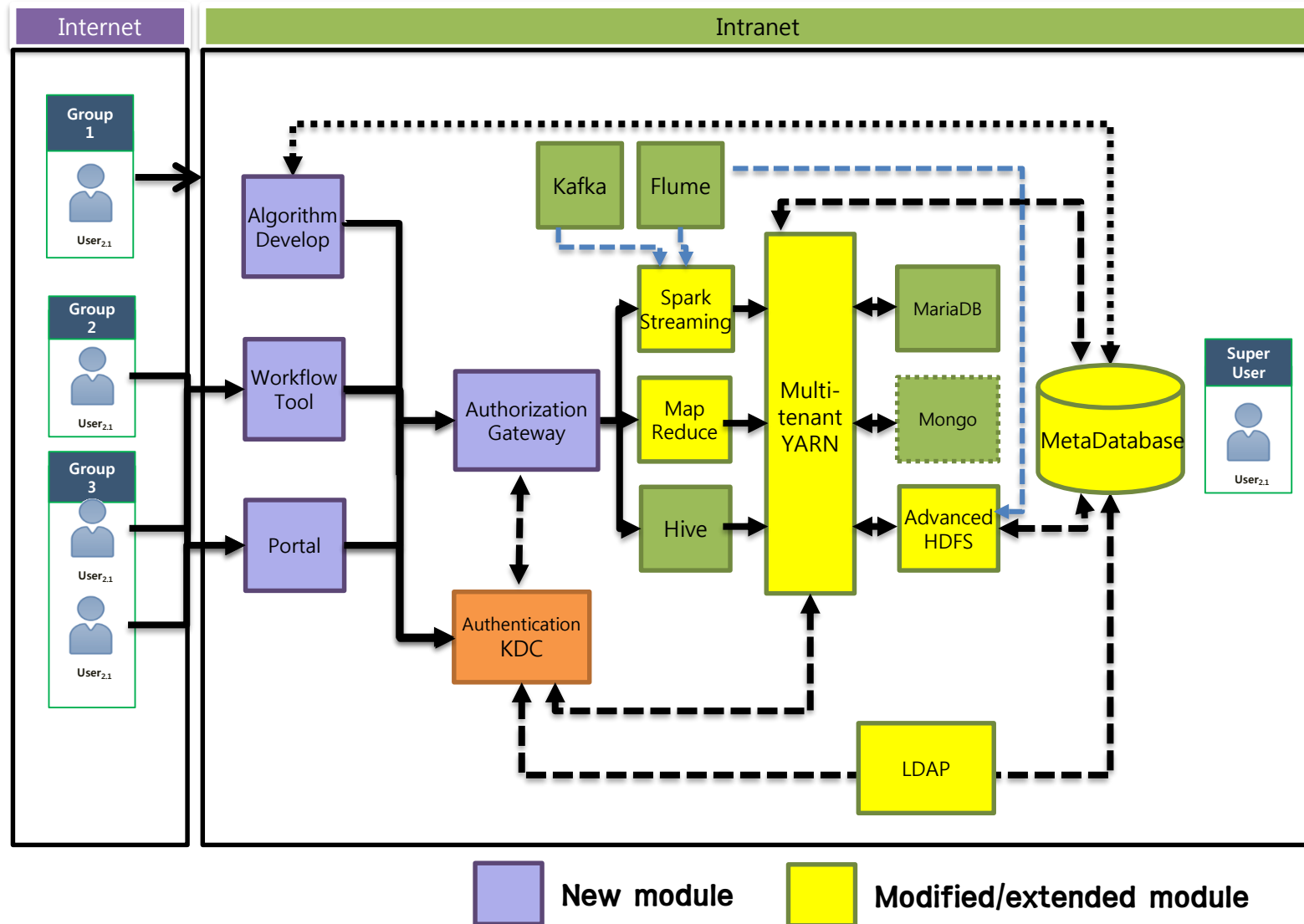
‘A Data scientist won’t have to worry about infrastructure.’

Overall System Architecture

- Collaboration Workspace for general users, data scientists and IT experts
- Gateway for platform interoperation with secure access control
- Multitenant Analytics Platform with enhanced extensibility, utilization & availability

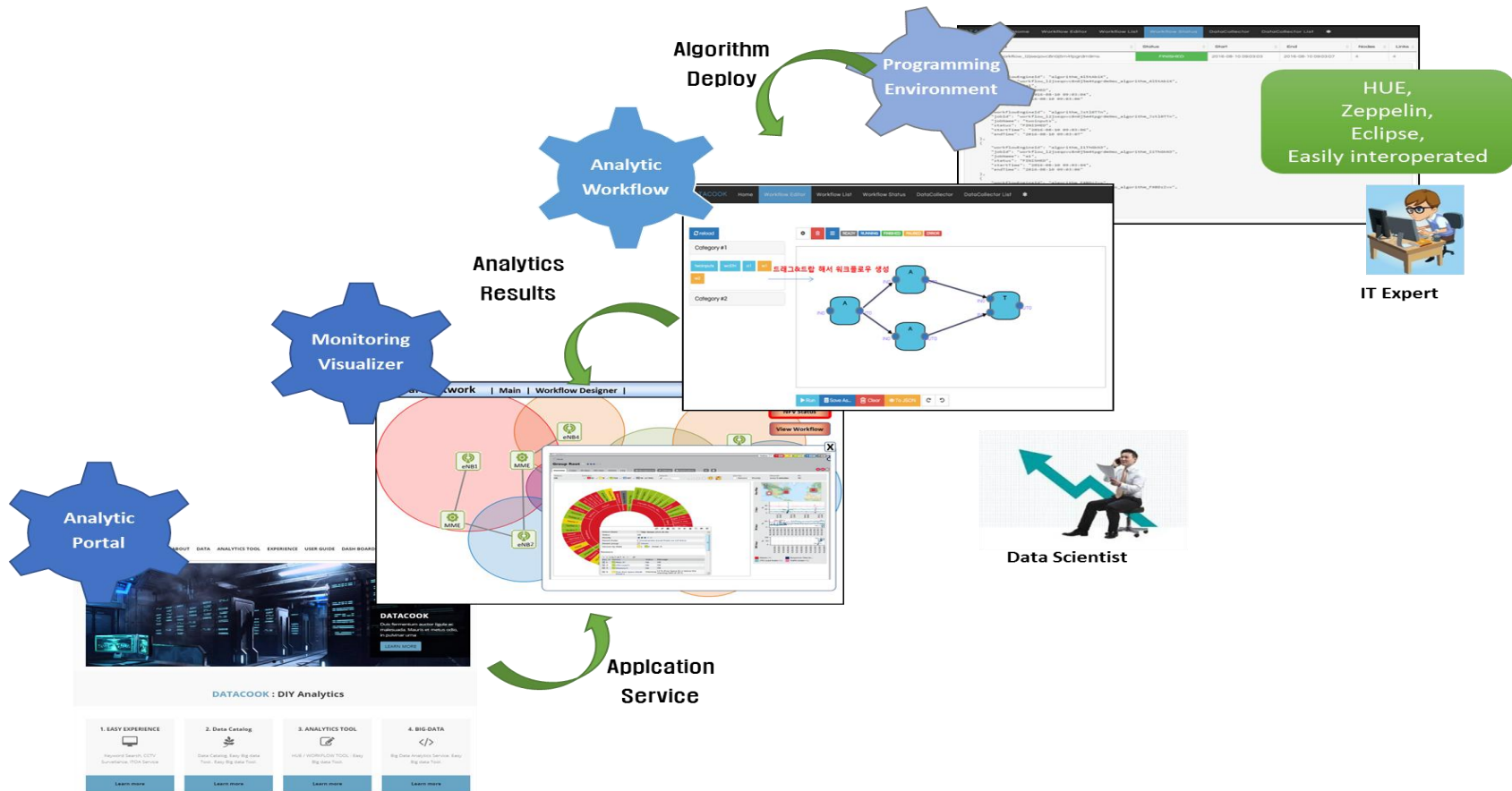


Dynamic System View



Collaboration Workspace

- Big Data portal for data uploading, sharing, viewing,...
- Workflow analytic tool supporting data analysis to implement their ideas
- Program (low-level algorithm) development environment

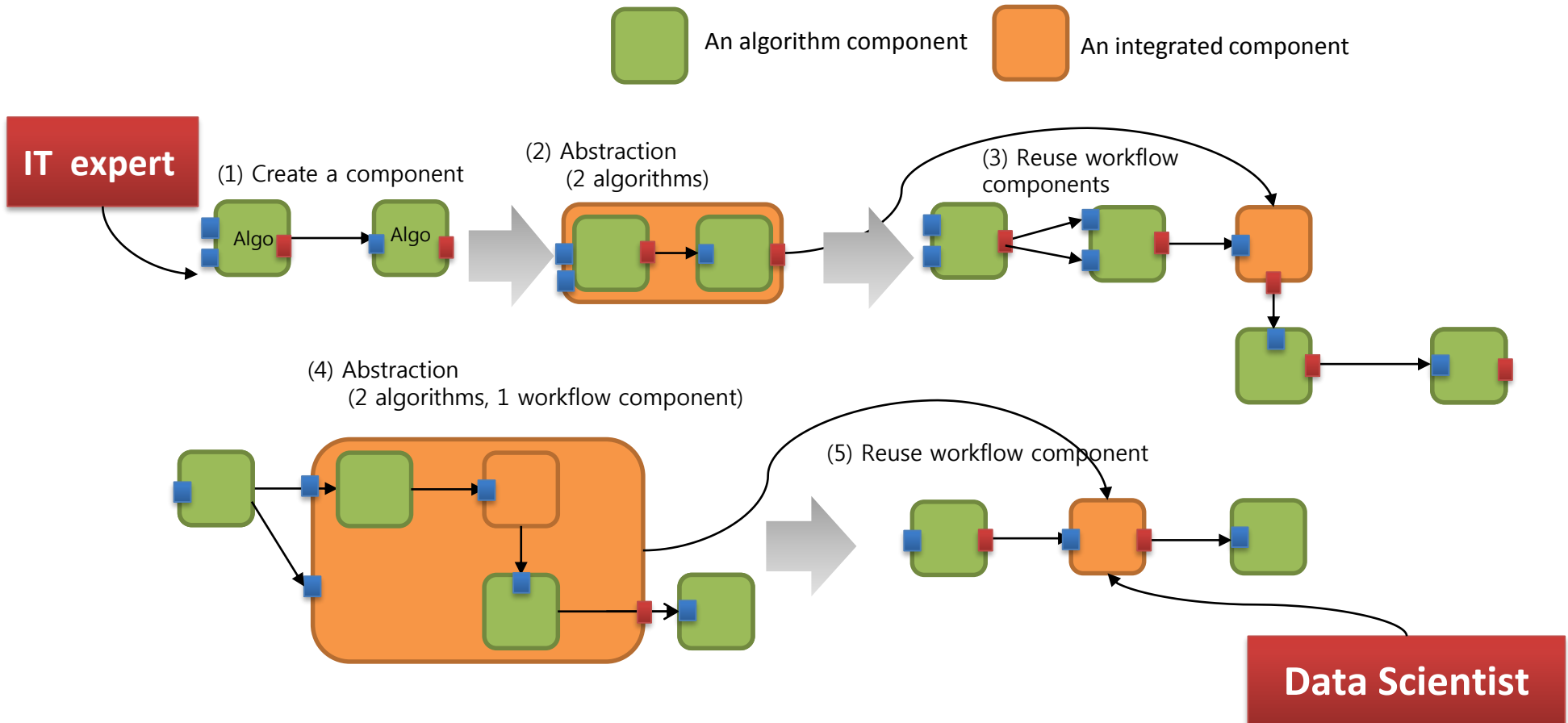


Integratable Analytic Workflow



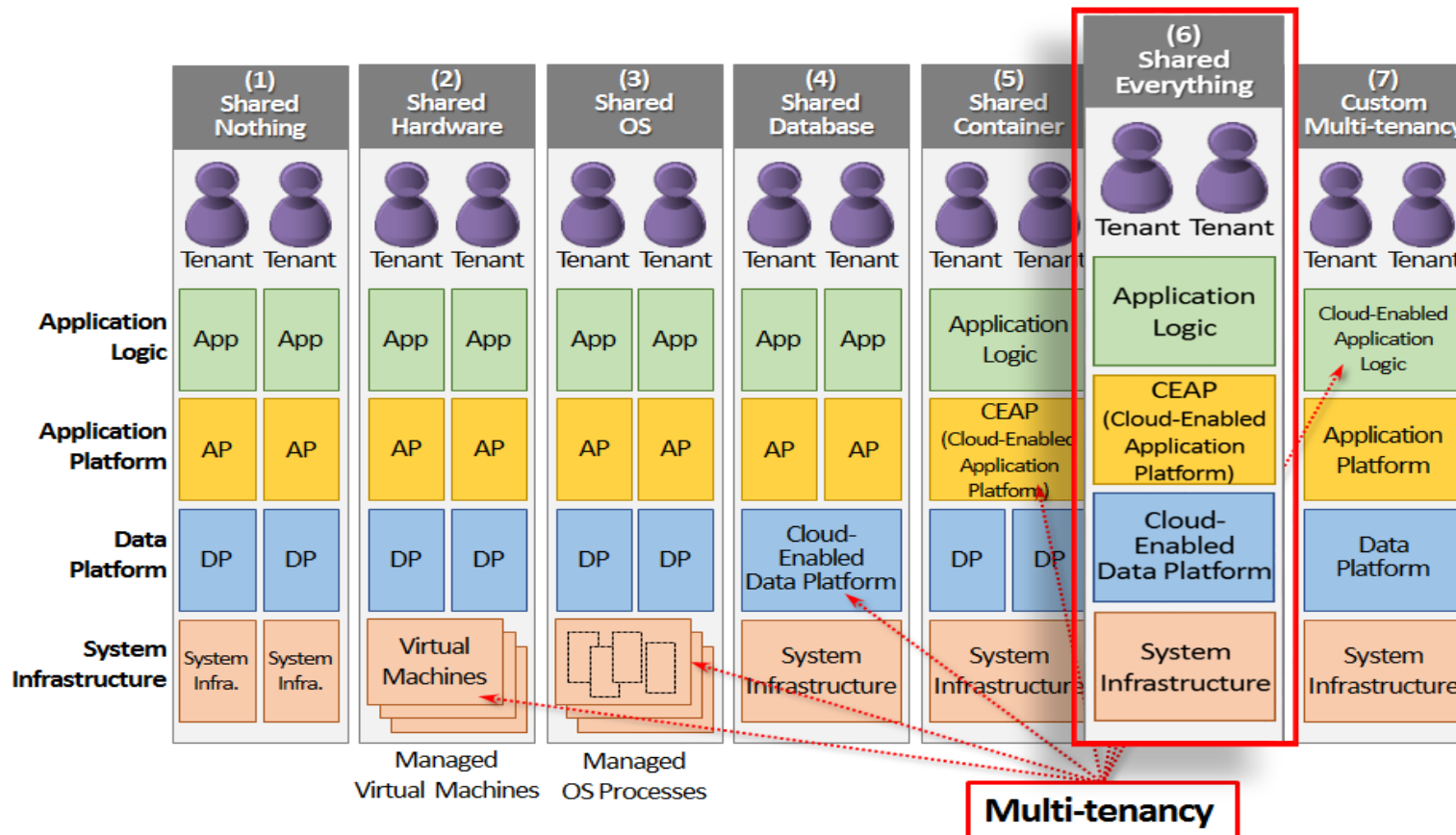
Workflow components can be

- ▶ An algorithm, a set of algorithms or a set of components



Multitenancy Concept for Cloud Computing

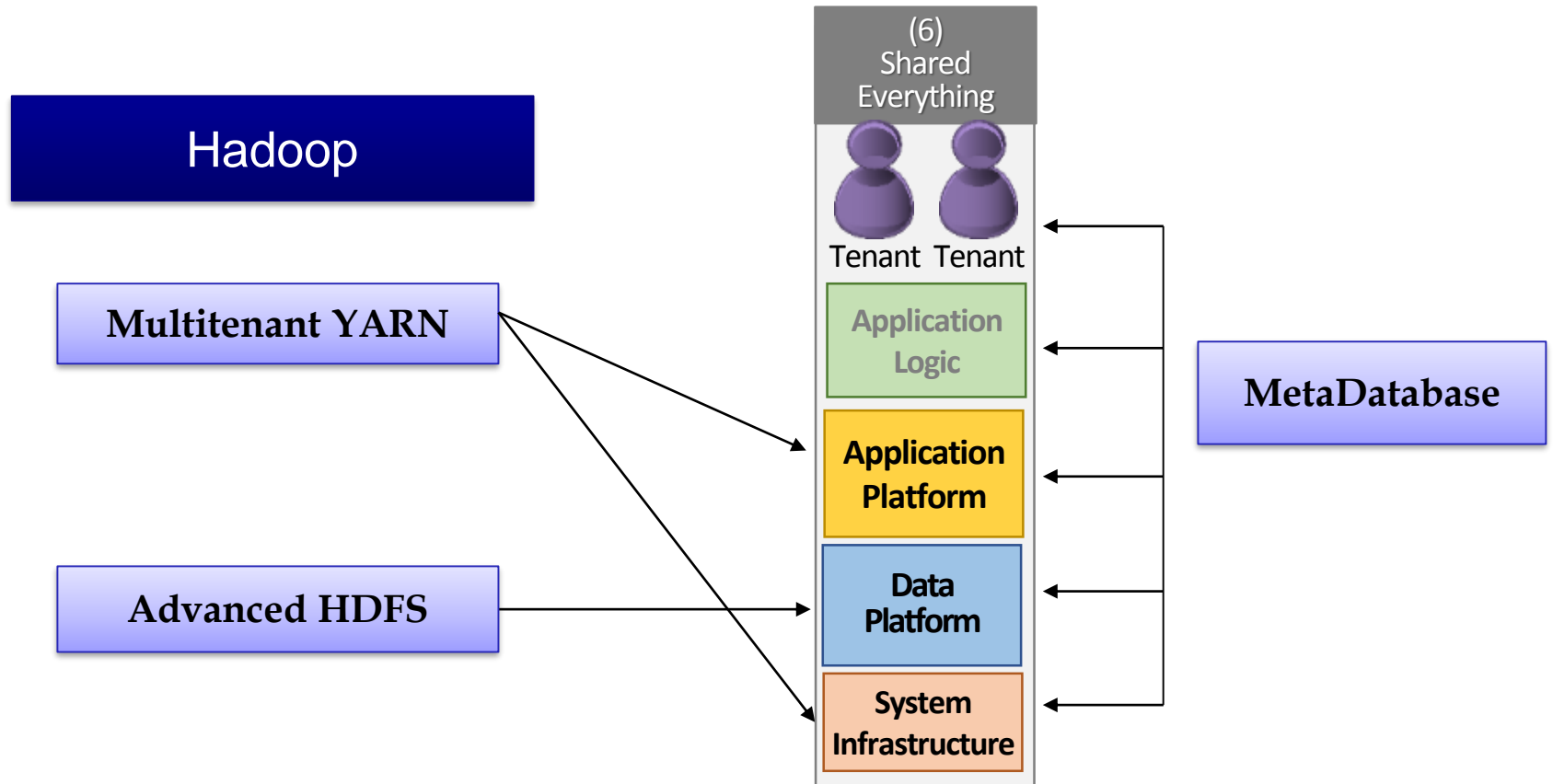
- Architectural concept for resource sharing in a single system across tenants
- Key factors: resource management, access control, audit
- Highly efficient and cost-effective system



[Gartner, 2012]

Project Goal

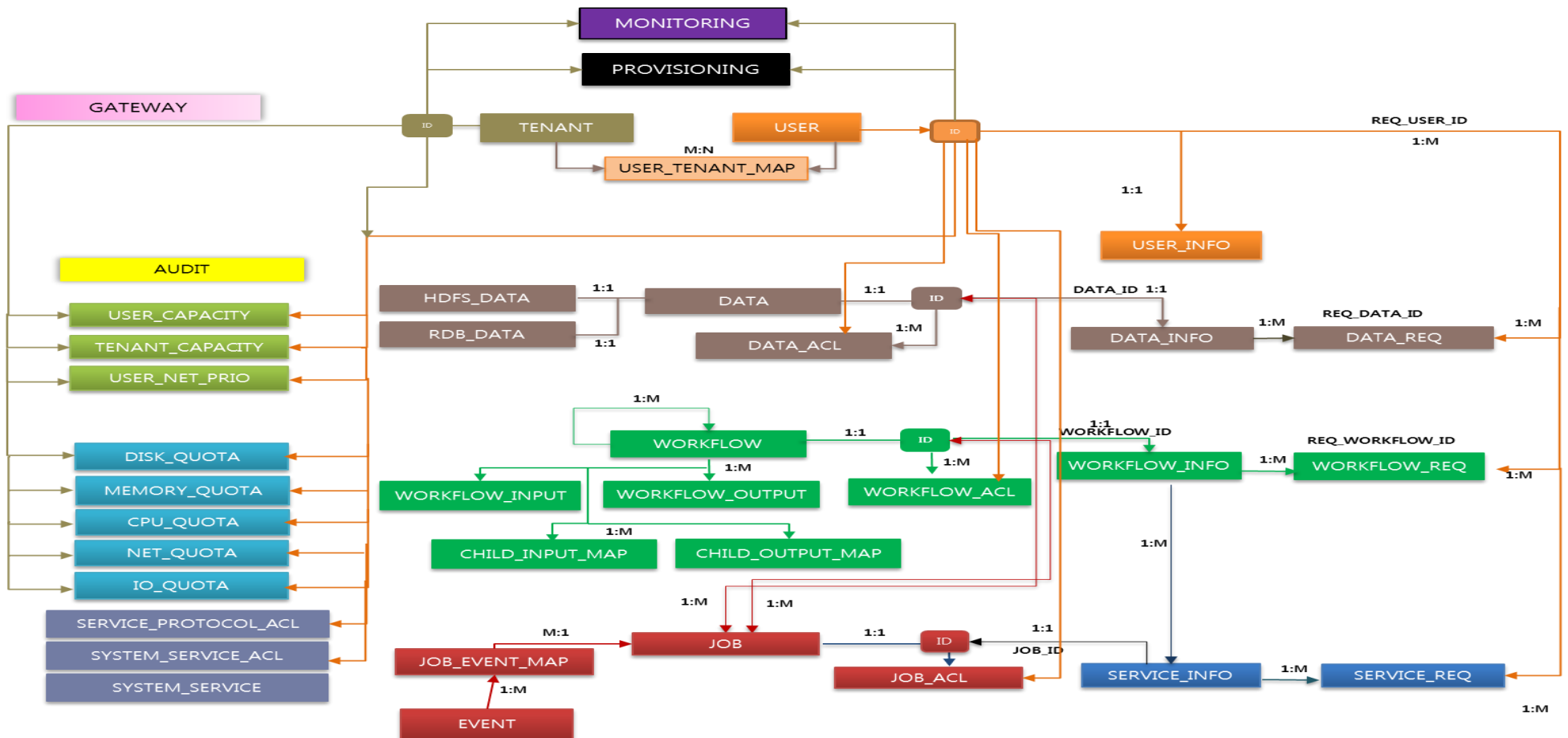
- Quota-based resource allocation per tenant, user and service
- Resource scheduling for high utilization
- Integrated Metadata for efficient system management, operation and maintenance
- Access control based on authorization/authentication



Integrated MetaDatabase

● Integrated metadata management using RDBMS

- ▶ Metadata for authorization/authentication, access control, resource management.



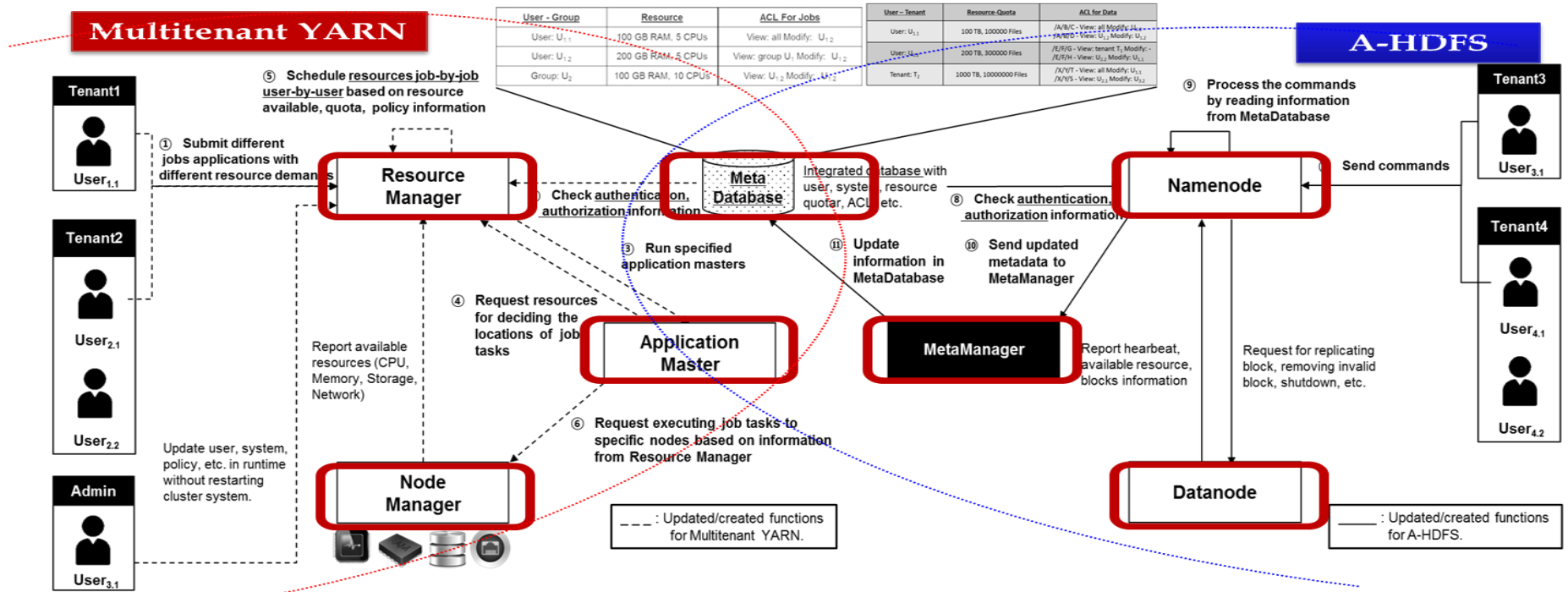
Multitenant Analytics Platform

- **Advanced multitenancy**

- ▶ Fine-grained resource allocation and scheduling optimization
- ▶ Secure computing environment

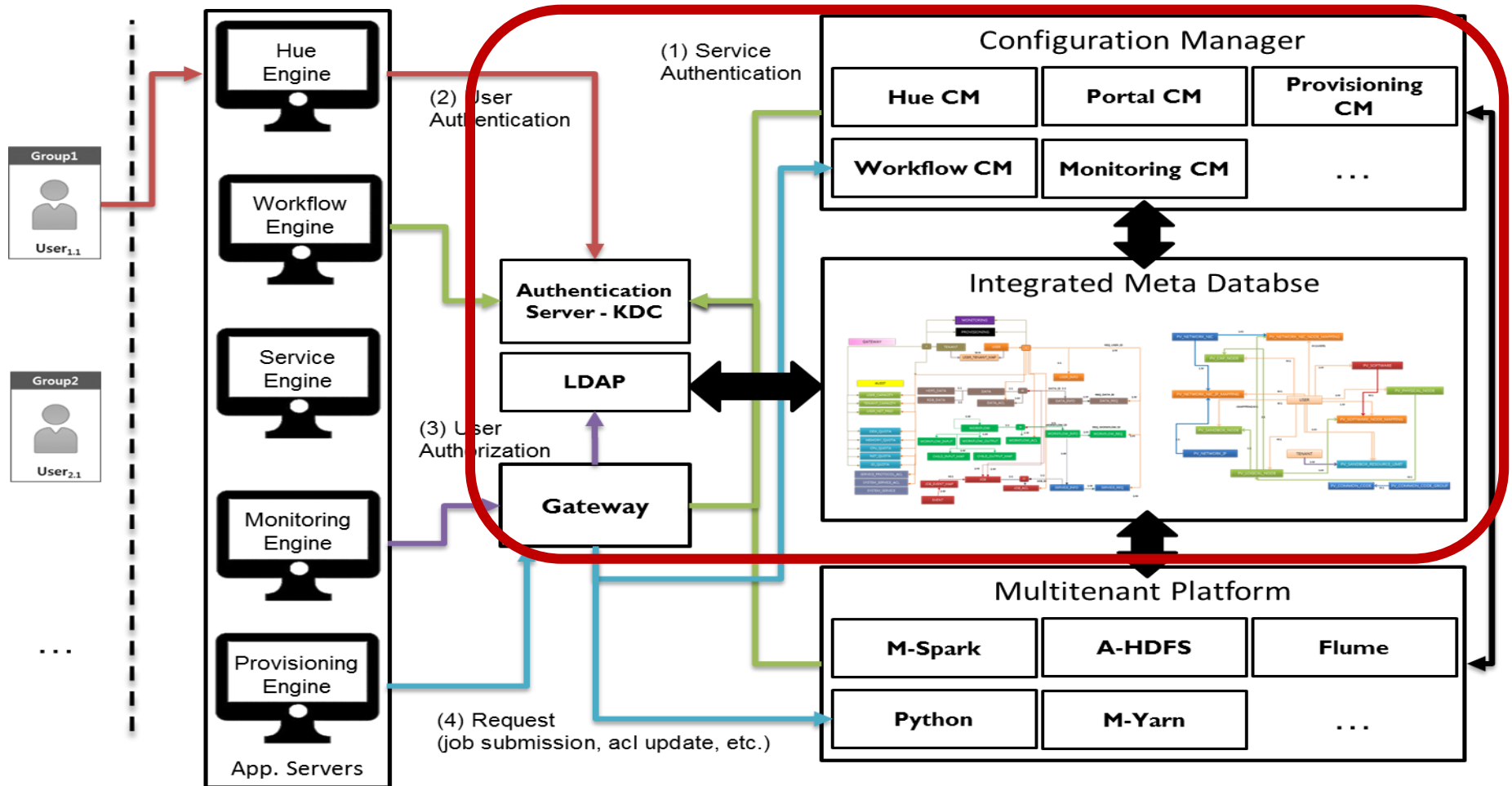
- **Improved scalability, robustness, availability**

- ▶ Moving metadata in memory and scattered files to DBMS
- ▶ Efficient management based on Metadatabase



Platform Gateway

- Authorization/Authentication based on MetaDatabase
- Restful APIs for Interoperation between platform and client services
- Configuration for integration of platform and client services



Technical Presentation

● Strata+Hadoop World, Singapore, 2015 & 2016

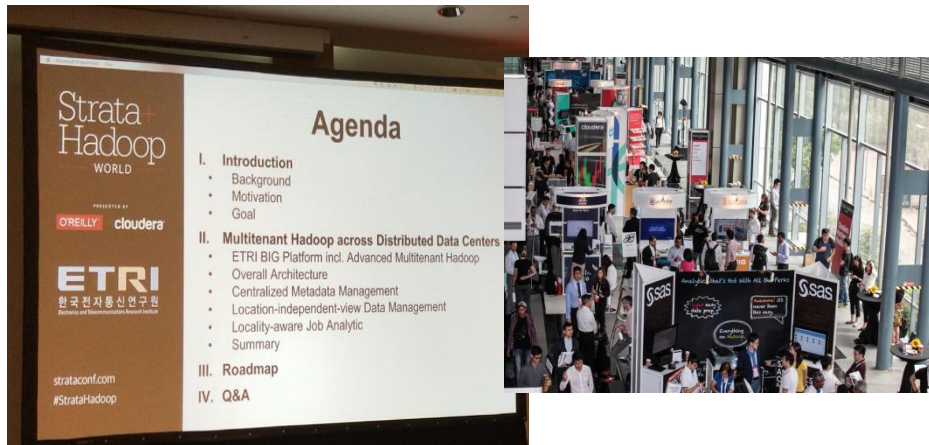
▶ Hadoop Platform Session Presentation

“Multitenant Hadoop across Geographically Distributed Data Centers”

“Unified Metadata Management for Scalability, Integrity and Reliability across Geographically Distributed Data Centers”



Strata+Hadoop World 2015 (12/'15, Singapore)




Strata+Hadoop World 2016 (12/'16, Singapore)




● EUREKA Co-summit 2015



● Korea EUREKA Day 2015



KOREA
EUREKA DAY
2015



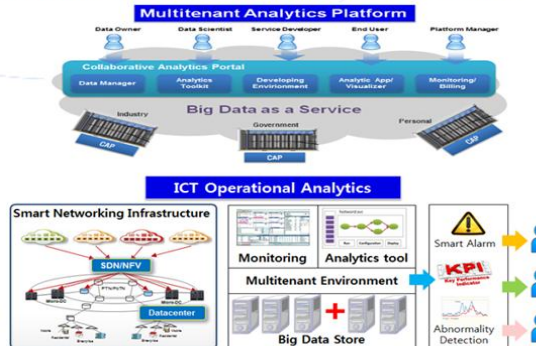
ETRI
Electronics and Telecommunications
Research Institute

Multitenant Analytics Platform for ICT Operational Analytics

- **Technology Overview**

A General Cloud Analytics Platform with Web Collaboration Framework for Data Ecosystem Which Provides Secure and Fair Access Control to Data and Computing Resources in Cloud.

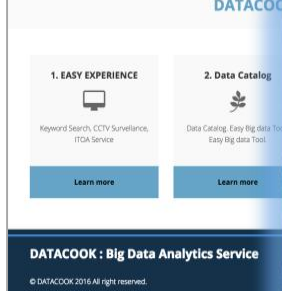
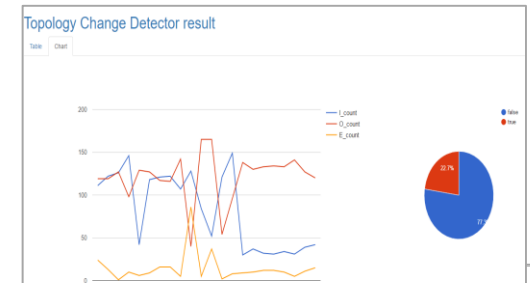
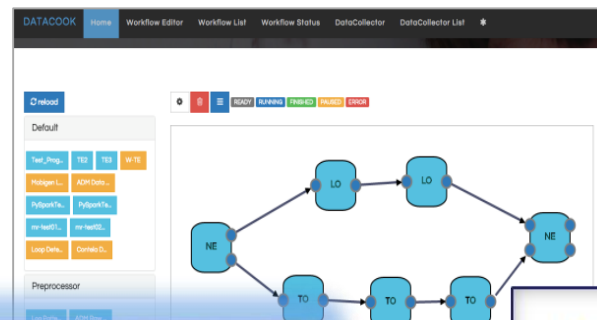
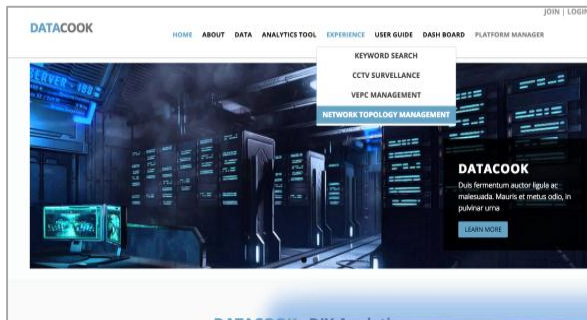
Operational Analytics Services to Make A Virtual Network and Cloud Infrastructure Intelligent.
- **Core Technologies**
 - Extension of Hadoop Ecosystem for Security and Resource Control based on Integrated Metadata for **Multitenancy**.
 - Prototype of Abnormal Pattern Detection With Collecting Cluster Logs and Primary Design of Operational Analytics Tool.
 - Web Collaboration Framework for Data Ecosystem (e.g. data owner, data scientist, service developer, platform manager, etc.)
- **Application Area and Advantages**
 - Data Marketplace Solution to Provide Secure and Fair Data Sharing
 - Cloud/NFV/SDN Management & Operation System with Big Data



Demonstration

● Big data portal, Analytic workflow, Services

- ▶ Intelligent CCTV monitoring
- ▶ Network topology monitoring



<Big Data>

▪ City-wide Surveillance Center

Workflow

▪ Prioritize CCTV for monitoring

▪ More accurate and fast recommendation



Thank you!!

Contact: hswon@etri.re.kr