Current Status of Open Source Big Data Business in Japan

2017/11/16
Yukio Yoshida
Research & Development Department
Hitachi Solutions, Ltd
0. Introduction
0-1. Self Introduction

【Career】
✓ At First, engaged in software development of financial terminal
✓ From 2000, responsible for the business development of the Linux / OSS.

【Current business】
✓ Support for business development using OSS
  • Excavation and evaluation validation of new technology / OSS
  • Start-up support of business solutions
  • Focus Area:
    Cloud Infrastructure (OpenStack, Spark),
    Big Data Infrastructure (Hadoop, NoSQL),
    Enterprise (PostgreSQL, OpenCOBOL)

【Activities outside of the company】
✓ Japan OSS Promotion Forum Vice Chairman
✓ Open License Laboratory Director
✓ OSS consortium Vice Chairman
etc
## 0–2. Corporate Data

<table>
<thead>
<tr>
<th>Corporate Name</th>
<th>Hitachi Solutions, Ltd.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head Office</td>
<td>4-12-7 Higashishinagawa, Shinagawa-ku, Tokyo 140-0002 Japan</td>
</tr>
<tr>
<td>Telephone</td>
<td>03-5780-2111</td>
</tr>
<tr>
<td>Representative</td>
<td>Setsuo Shibahara, President and Chief Executive Officer</td>
</tr>
<tr>
<td>Founded</td>
<td>September 21, 1970</td>
</tr>
<tr>
<td>Capital stock</td>
<td>20 billion Yen</td>
</tr>
<tr>
<td>Number of employees</td>
<td>11,572 (Consolidated)</td>
</tr>
</tbody>
</table>
Pathways to new frontiers; with Unique Strengths

Hitachi Solutions will develop new solutions and business models by working together with our customers; through businesses that deliver social innovation and that are based on dependable technologies. We will leverage all our strengths, from global ideas that transcend national and cultural boundaries, and our ability to execute solutions together with the overall strength of the Hitachi Group – opening a door to a new era.
The Hitachi Solutions Group will realize the development of customer businesses and social innovation through collaboration between our global bases while leveraging our extensive experience, technologies and ability to provide solutions.
0–5. Global Network

- **Hitachi Solutions Europe Ltd.**
  - Employees: 220
  - London

- **Zhejiang Hitachi Solutions Software Services Co., Ltd.**
  - Employees: 160
  - Beijing, Shanghai, Jiaxing, Guangzhou

- **Hitachi Solutions India Pvt. Ltd.**
  - Employees: 650
  - Chennai

- **Hitachi Solutions Asia Pacific Pte Ltd.**
  - Employees: 130
  - Hitachi Solutions (Thailand), Ltd.
  - Hitachi Solutions Philippines Corporation

- **Hitachi Solutions America, Ltd.**
  - Employees: 370
  - San Francisco, Toronto

- **Hitachi Solutions Canada, Ltd.**
  - Employees: 150
  - Manila, Bangkok

(As of March 31, 2017)
1. Market Trend
Big data market trend in Japan

Big data technology / service market in Japan

- **2016**
  - $2.3B+
  - 8.0%

- **2021**
  - $3.4B+
  - 8.4%
Most of the Big Data Platforms are composed of 3 functions:
“Data Collection”, “Data Store” and “Data Analysis”
2. Approach to OSS business
Our OSS Solutions

Provide one stop service from planning phase to operation phase

- **PLAN**
  - Consultation
  - Validation and evaluation

- **DESIGN**
  - Server Design / Integration / Migration service

- **DEVELOPMENT**

- **INTEGRATION**
  - Red Hat Solution
  - Big Data Utilization support solutions (Hadoop, Talend)
  - SDN network operation efficiency solution
  - Open Source System Monitoring Solution

- **OPERATION**
  - Support service
  - OSS management solution (Operation agency service, Maintenance service)

- OSS management solution (Protex Analysis and verification service)
Big Data Utilization support solutions

① Consulting Service
- PoC · Support feasibility verification
- Support architectural design
- Provide general technical consultation at the time of introduction

② System Integration
- Construction of Bigdata platform
- Development of Bigdata related applications
- Provide optimal architecture design and operation design support

③ Support
- Total support the product handled by the Big Data Utilization Platform
  Cloudera Enterprise, Talend Enterprise, Asakusa Framework, Privacy Data Anonymizing Platform
3. Use Case
Use case No.1 (Securities company)

Used customers: Securities company planning and marketing department
Used customers: Increase analysis area, increase added value of products and services · Expand market share
Utilization scene: Non-structured data such as access log and SNS in addition to ERP and CRM data

Benefits

- Optimize IT Cost ⇒ Reduce maintenance costs as a hybrid configuration with DWH and realize future IT cost optimization by flexibly scaling out according to the amount of data
- Expansion of analysis use ⇒ Realize effective marketing and product development by utilizing non-structural data such as various logs and customer's opinions
Customer issues solved by Hadoop (why Hadoop ?!)

Customers of the issues and circumstances of the past

- Data utilization that finds new values that customers demand
- Most of the company's data (80%) is nonstructural data
- Today's DWH reaches its limit
- Location of dormancy data
- Corporate data is divided / noncommon

Resolution by Hadoop

- Functions suitable for search type analysis are standard equipment
- Petabyte class mass data and structured / nonstructured data can also be stored and used
- It can be used as a complementary system of DWH, solving performance issues and cost issues
- Start by saving data easily, then flexibly deal with high-speed / advanced utilization (processing · BI connection etc.)
- Accumulate and consolidate any data in the enterprise and can process to analyzable data at high speed
Use case No.2 (Finance industry)

- Used customers: financial institution (bank)
- Customer business: Financial business innovation research through data utilization
- Utilization scene: Cleansing and structuring of massive unstructured logs on Internet banking sites? Patterning of production lines and operation time for each user, data separation and staying time data

**Benefits**

- Analyze and grasp the status and reasons of actual operation for each user of Internet banking ⇒ Improve the rate of closure by appropriate marketing
- Mere log information is visualized as a user’s business line or pattern, ⇒ Secondary analysis by data scientist, new points of improvement rate increase points and target persons are derived
Use case No.3 (Electric Power Company)

Used customer: Electric power company
Customer Business: Providing the optimum price menu by fee calculation and simulation
Utilization scenes: Data storage infrastructure corresponding to increase in meter reading data accompanying smart meterization? Electric charge calculation simulation for electric power liberalization (using mass meter data)

- Scale up to meter data that increases to more than 100 TB in 3 years
- Acceleration of simulation calculation by distributed processing
- Improve Hadoop application development with Asakusa F / W

Benefits

- Data storage that can be scaled out
  ⇒ Optimized for the amount of data to be accumulated
- Acceleration of simulation calculation by distributed processing
  ⇒ Efficiency improvement of work such as examination of fee structure by shortening calculation processing
- Introduction of high efficiency batch processing development framework (Asakusa FW) ⇒ Optimization of development cost

Performance comparison
(400 M records: 37 GB)

<table>
<thead>
<tr>
<th></th>
<th>Unit</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>RDB</td>
<td>1</td>
<td>123 Hrs</td>
</tr>
<tr>
<td>Hadoop, (Asakusa FW)</td>
<td>5</td>
<td>21 Hrs</td>
</tr>
</tbody>
</table>

Development efficiency
About 50%
Use case No.4 (Electoric Power Company)

Used customers: Event venue operator
Customer business: visualization and prediction of congestion situation in quasi real time at event venue
Utilization scene: Real-time location information collection, processing, accumulation
Information transmitted tens of thousands of messages per minute is subjected to micro batch processing by Spark job

**Benefits**

- Real-time grasp of user location in the venue ⇒ By visualizing and presenting the results of aggregation / analysis, users can consider effective booth tour
- Periodic collection of user location information ⇒ Can be used for examining the following venues during the event venue management
Use case No.5 (Distribution industry)

**Used customers:** Distribution industry  
**Customer Business:** Manufacture and sale of soft drinks  
**Utilization scene:** A new analysis application is built on Hadoop in order to improve the accuracy of vending machine sales prediction. Using Spark greatly shortens processing time for complex data shaping / correction and prediction logic.

**Benefits**  
- Increase replenishment accuracy ⇒ Reduce the rate of missing items for each item and make it possible to replenish the appropriate quantity  
- Optimize visit routes / times ⇒ Reduce wasteful visits, optimize visiting routes, number of visits, and improve worker efficiency

**Hadoop (Cloud)**

- Accelerate complicated correction processing, shaping processing, and prediction logic with Spark! (Performance requirement within 1 hour)  
- Hive also runs in-memory!
Thank you
감사합니다.
谢谢