Big Data in China Railway Application and Development

Prof. Ping LI
China Academy of Railway Sciences

11/30/2016, St Petersburg
Content

1. Introduction on China Railway
2. Key Requirements on Big Data
3. Architecture of Big Data Platform of CR
4. Typical Application of Big Data in CR
1 Introduction

China national policy on big data

- Action plan for promoting the development of big data, 2015, The State Council
- The 13th Five-year Plan for national economic and social development, 2016-2020
- Major projects to promote the development of big data, 2016, National Development and Reform Commission (NDRC)
1 Introduction
2 Key Requirements on Big Data
2 Key Requirements on Big Data

① Passenger transportation marketing
- CRM
- Product Design
- Precise marketing
- Service optimization
- Route planning

② Freight transportation marketing
- CRM
- Transport volume forecast
- Logistics optimization
- Multi-modal transportation

③ Mobile and fixed infrastructure
- Fault detection
- Fault relevance analysis
- Fault forecast
- Maintenance decision making

④ Security
- Analysis of causes
- Risk and related analysis
- Early warning of Security risk

⑤ Enterprise Operation
- Economic circumstance analysis
- Economic benefit analysis
- Productivity analysis
- ROI analysis
3 The Reference Architecture of Big Data Platform

<table>
<thead>
<tr>
<th>Principles</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Big Data Requirement</strong></td>
<td>➢ Massive data loading, processing and analysis</td>
</tr>
<tr>
<td></td>
<td>➢ Variety data loading, processing and analysis</td>
</tr>
<tr>
<td></td>
<td>➢ Key data loading, processing and analysis</td>
</tr>
<tr>
<td><strong>Application Requirement</strong></td>
<td>➢ Scalability</td>
</tr>
<tr>
<td></td>
<td>➢ Availability</td>
</tr>
<tr>
<td></td>
<td>➢ Security</td>
</tr>
<tr>
<td></td>
<td>➢ Openness</td>
</tr>
<tr>
<td></td>
<td>➢ Accessibility</td>
</tr>
<tr>
<td><strong>Data analysis and mining</strong></td>
<td>➢ Text data analysis</td>
</tr>
<tr>
<td></td>
<td>➢ Image data analysis</td>
</tr>
<tr>
<td></td>
<td>➢ Video data analysis</td>
</tr>
</tbody>
</table>
3 The Reference Architecture of Big Data Platform
Achieve a manual ticketing to the computer ticketing transition
Achieve a regional center networking ticket
Achieve a national network ticketing
Shift from enterprise systems to e-commerce system
Adaptation needs to achieve a revenue of liquidation;
Version 4.18 implements the 180 days in prior booking.
Meet the needs of centralized management and organization of new ticketing policy

4 Typical Application of Big Data in CR
Challenges

- 7,000,000 orders and its stub data in each day in average
- PB level structured data
- High concurrence
- Log data mining
- Operating data mining

Requirements

- Massive data transmission
- Massive data storage
- Massive data query
- Booking ticket analysis by big data
- User behavior analysis by big data

4 Typical Application and Development
### 4 Typical Application and Development

<table>
<thead>
<tr>
<th></th>
<th>Hadoop cluster</th>
<th>Traditional DB</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1 day</strong></td>
<td>36.162 seconds</td>
<td>25 minutes (1500 seconds)</td>
</tr>
<tr>
<td><strong>1 month</strong></td>
<td>36.409 seconds</td>
<td>12 hours (43200 seconds)</td>
</tr>
<tr>
<td><strong>1 year</strong></td>
<td>49.384 seconds</td>
<td>&gt;1 day</td>
</tr>
</tbody>
</table>

Passenger travel times by railway during the named period (54,670,000 ticket sales records)
Consider the long-term trend, cycle regularity, expertise and random factors, the prediction model can be mathematically described by

\[ X_t = W_y \sum_{i=1}^{p} \Phi_i X_{t-i} + W_{y_1} X_{t-\tau_1} + W_{y_2} X_{t-\tau_2} + W_{\alpha} \alpha_t \]
4 Typical Application and Development

3 Big Data visualization for real time status for ticket sold
Big Data visualization for the 2016 Dragon Boat Festival