Enterprise Railway Data Service Platform Based China Railway Safety Application

Zhongtieke (Beijing) Software Technology Co. Ltd

Agenda

1 Introduction

The development of China Railway, the information systems , PB data of CR safety , the demand of application scene and the solution of “platform + application”

2 Railway Data Service Platform

The overall architecture, technical architecture, functional architecture and design features of RDSP

3 Application In China Railway Safety

The typical application scenario and the result obtained

4 Conclusion

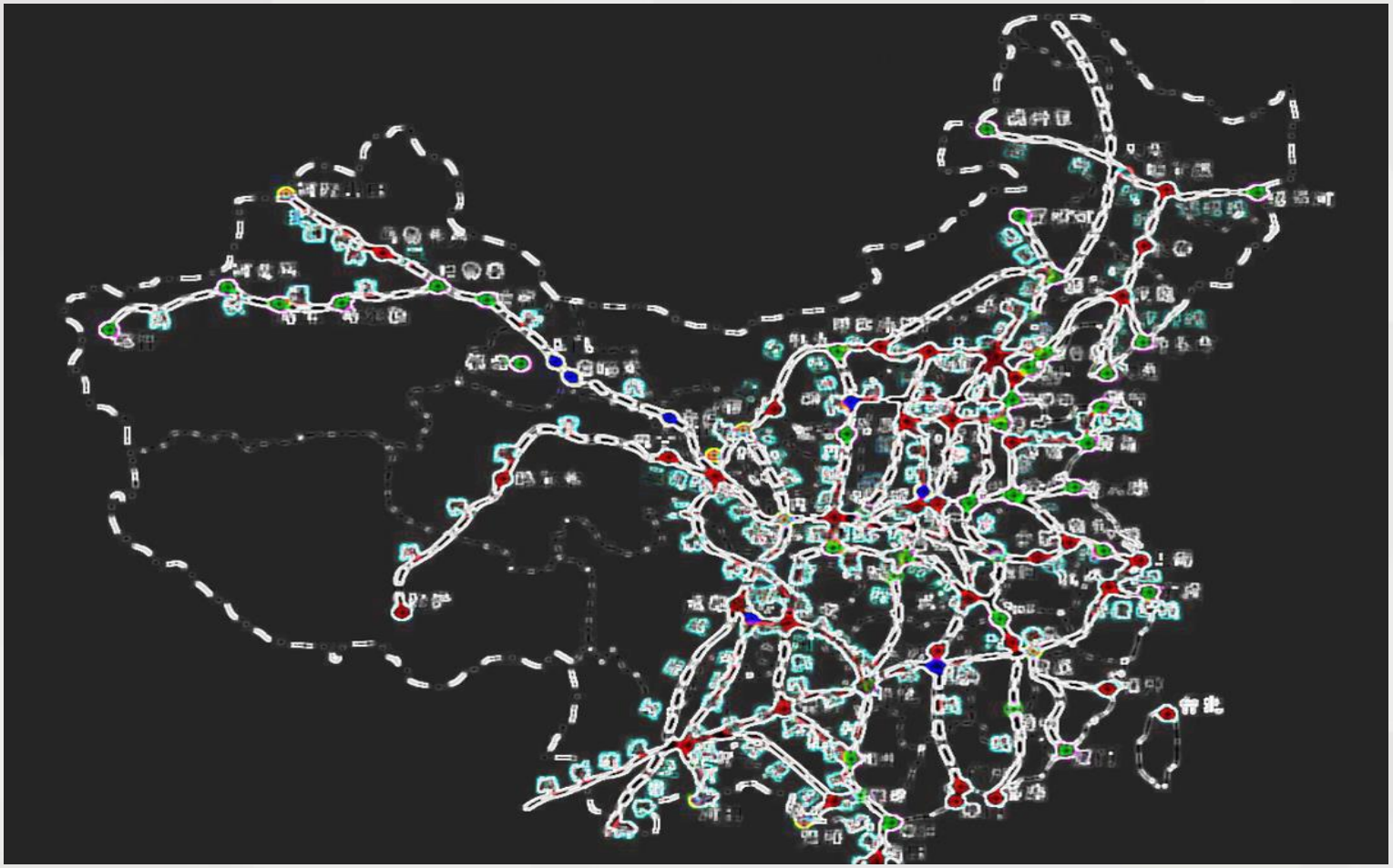
Summarize the results of the application in CR safety and look forward to the future work from the platform and application

CHAPTER ONE

Introduction

The Rapid Development of China Railway

China has the world's largest high-speed rail networks with nearly **20,000 km** of track in service.



The Safety Information Systems of China Railway

More Than 60 Railway Safety Systems，eg. 5T、6A、6C

☐Rolling stock safety monitoring(5T)

Trace Hotbox Detection System

Trackside Acoustic Detection System

Truck Performance Detection System

Trouble of Moving Freight Car Detection System

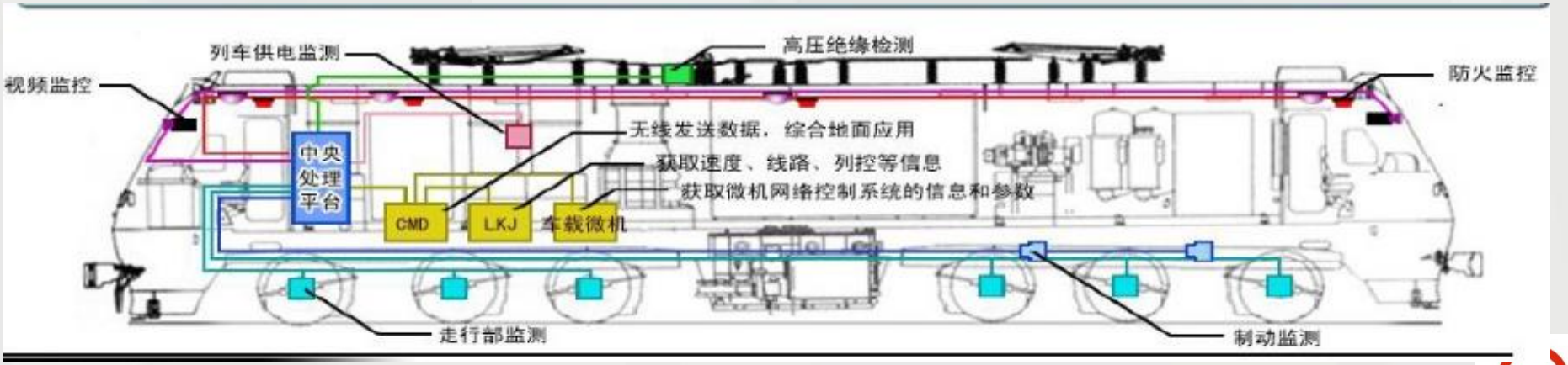
Train Coach Running Safety Diagnosis System

☐ Locomotive vehicle safety protection (6A)

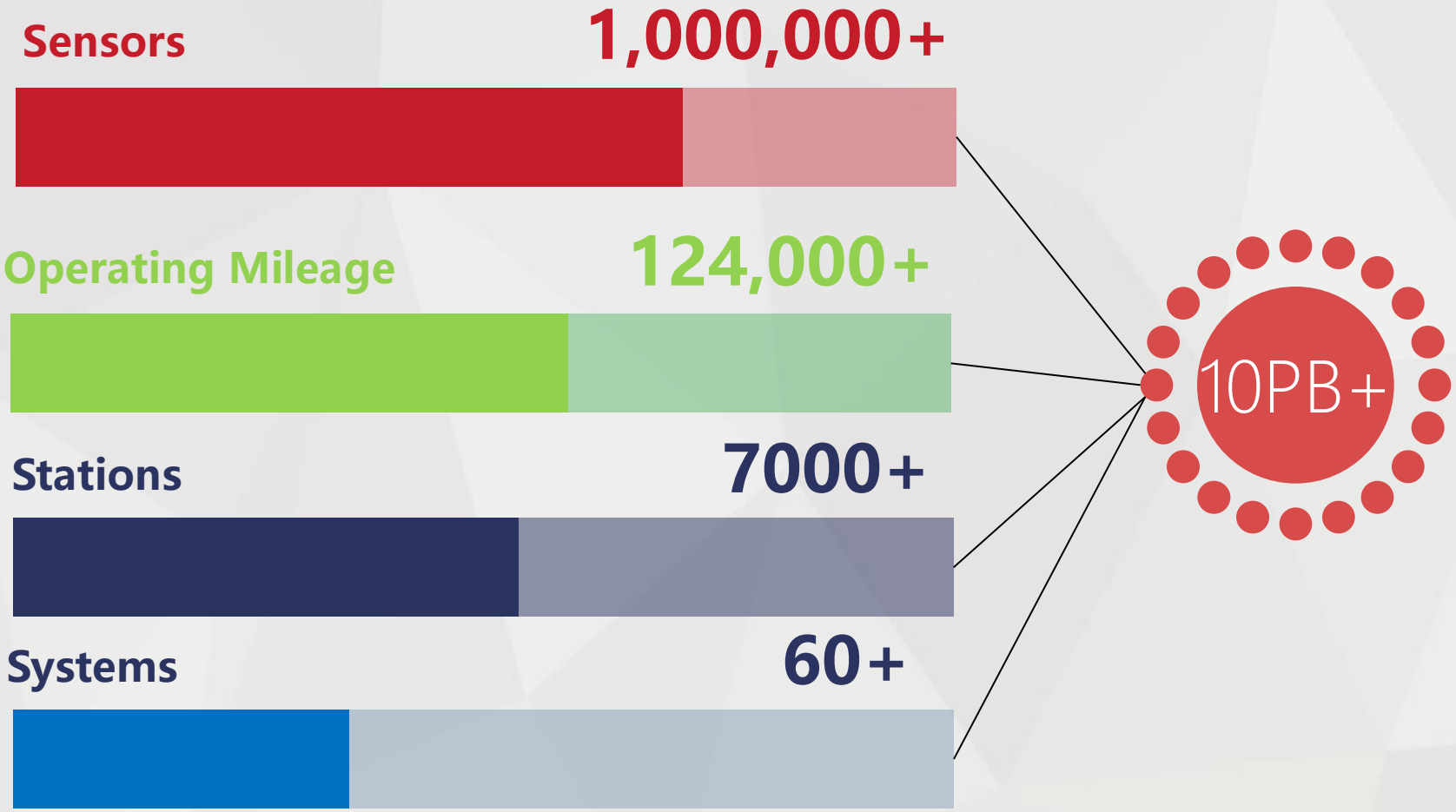
Air brake/ fire-proofing/ high-voltage insulation / power supply/ running gear monitoring/ video monitoring record

☐ Catenary and power supply detection (6C)

High speed pantograph , pantograph pan, ground power supply detection



The Data Volume of China Railway Safety



The Requirement of China Railway Safety

The Solution-->“Platform+Application“

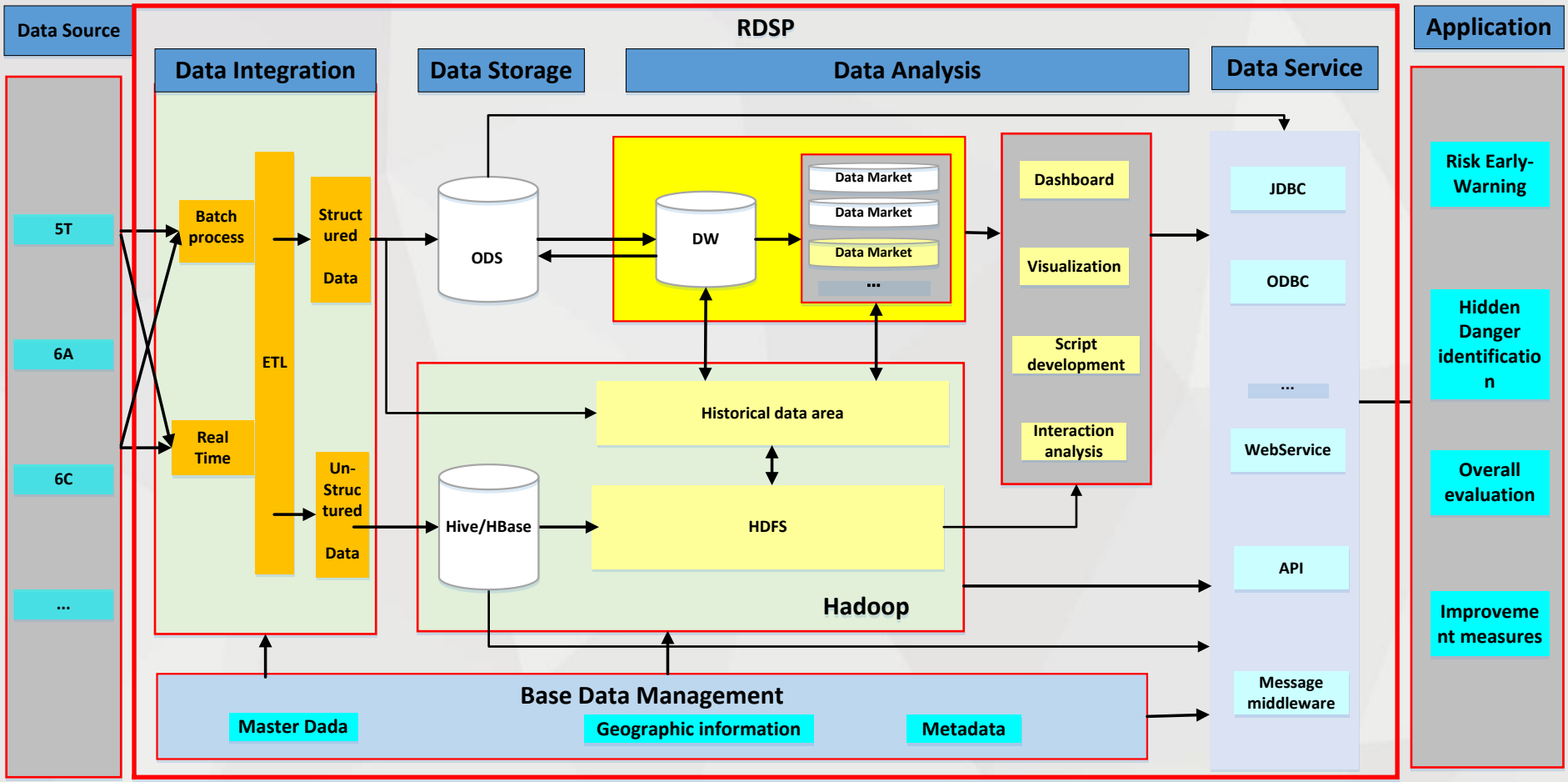
- Risk early-warning and hidden danger identification
- Qualitative analysis of accident and fault
- Improvement and solution recommendation
- Overall evaluation of safety

Railway Data Service Platform

CHAPTER TWO

Railway Data Service Platform

The Overall Architecture of RDSP



The Technological Architecture of RDSP

Data Integration

- Unstructured Data Streamsets/ftp/sftp
- Structured Data Streamsets/sqoop

Data Storage

- HDFS/HBase
- Ceph/HBase
- SparkStreaming /Flink
- Spark/Mapreduce
- Unstructured data annotation
- Data cleaning and quality management

Data Analysis

- Spark MLib /Mahout
- Spark Graphx
- Tensorflow
- Echarts/D3
- Greenplum
- Map service component
- kylin
- Railway professional data analysis model

Data Service

DAAS

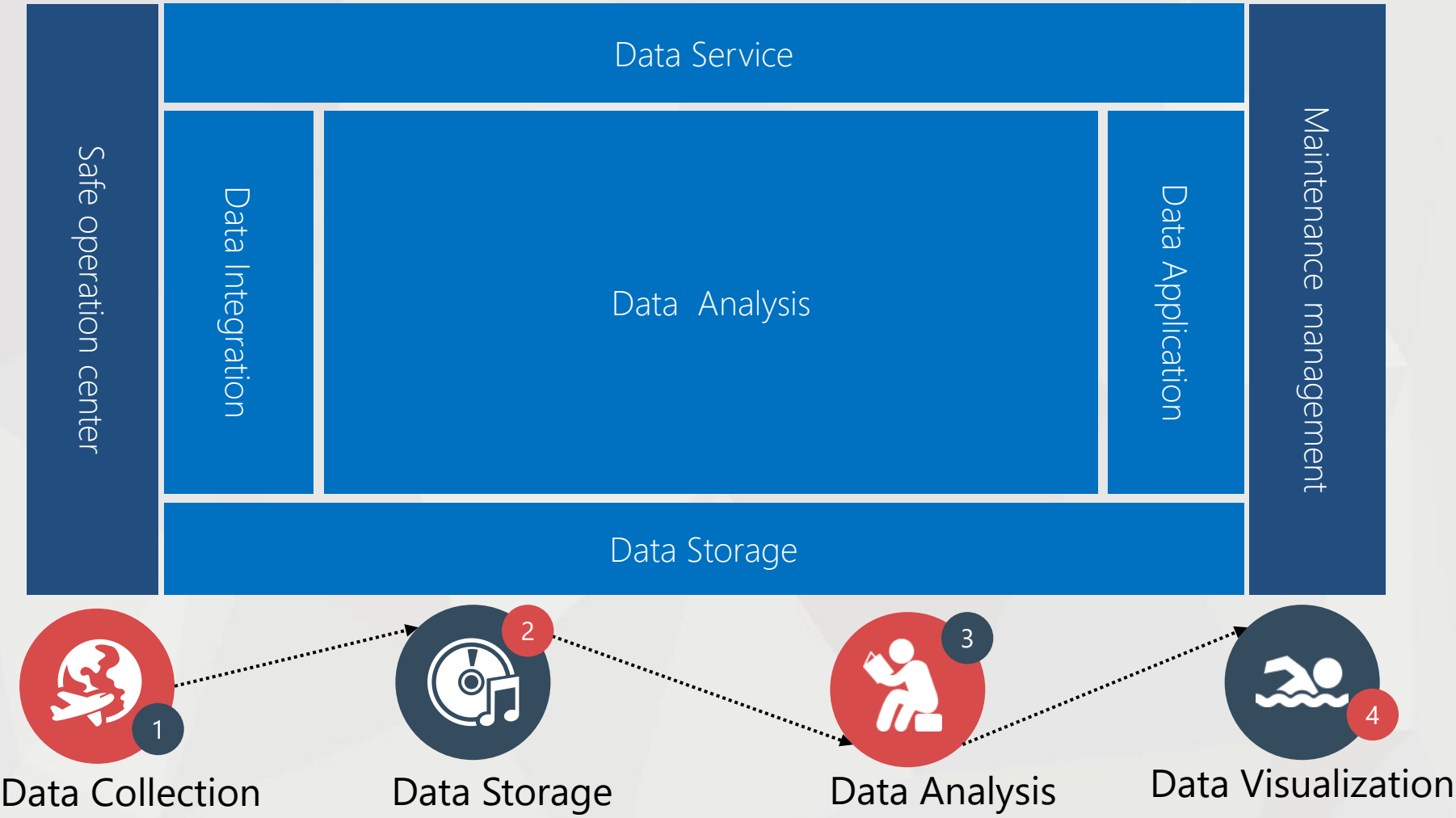
- Greenplum
- Postgres
- Redis
- ElasticSearch
- NOSQL(Hbase Phoenix)

- JDBC/ ODBC
- Restful API

Basic Data

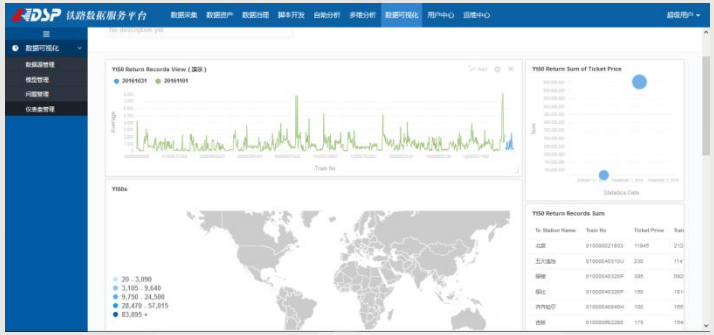
- Master Data
- Geography Information
- Metadata

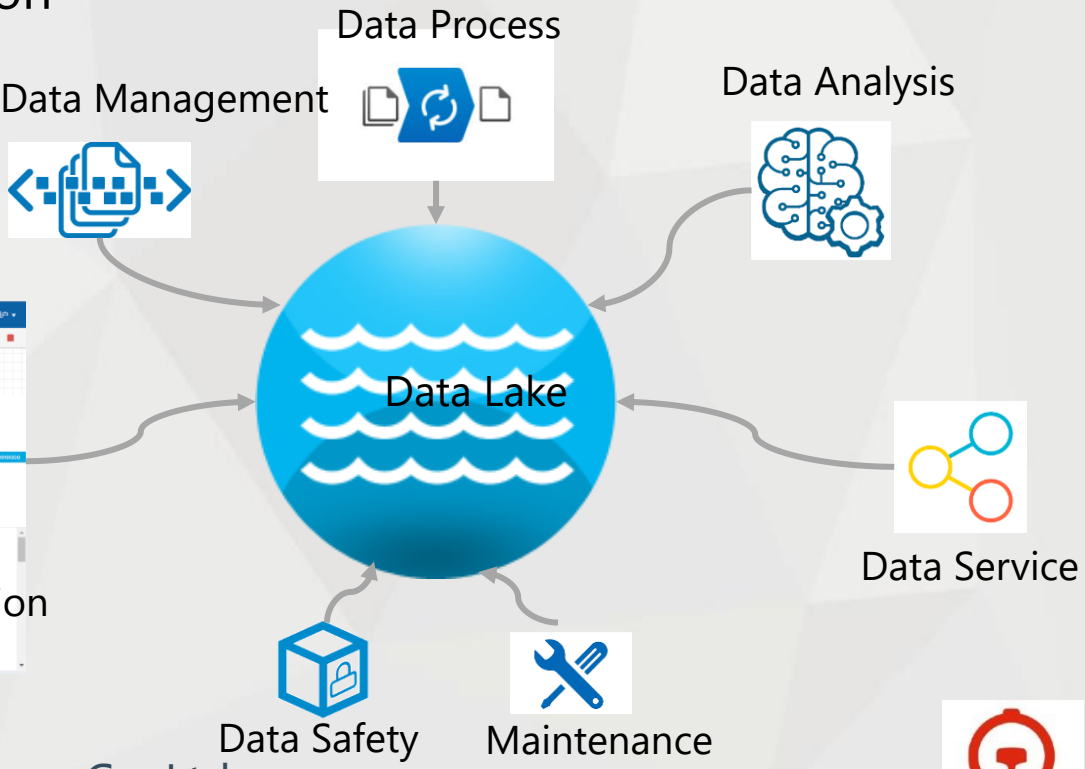
The Functional Architecture of RDSP



The Feature & Innovation of RDSP

- ✓ Storage and Computation are decoupled
- ✓ Distributed Computing Architecture
- ✓ Elastic Stretching Capacity
- ✓ Drag-and-drop Data Integration
- ✓ One-stop Self-help Analysis
- ✓ Container based Microservice





CHAPTER THREE

Application In China Railway Safety

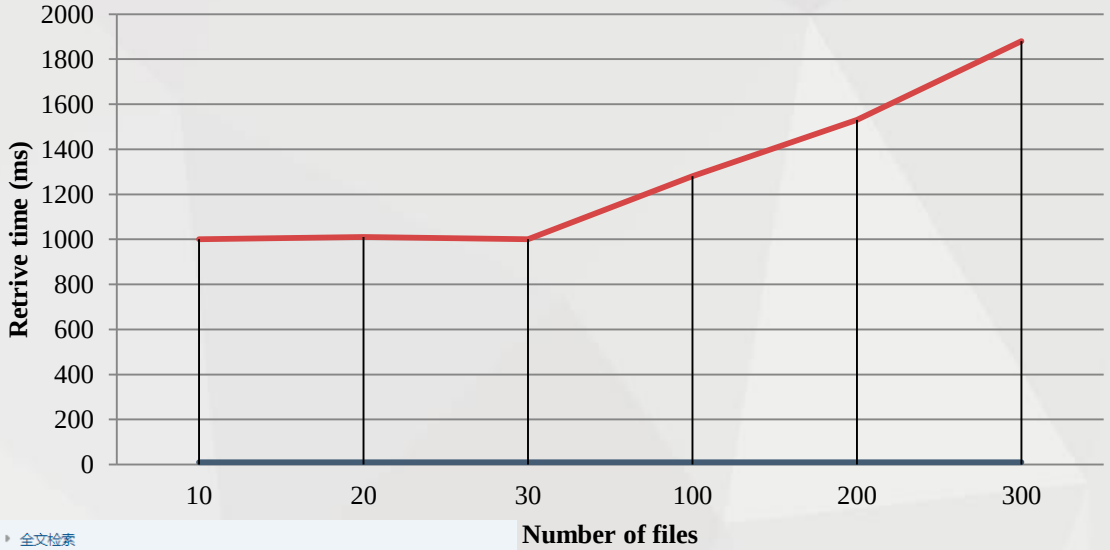
1. Data Storage and Retrieval

- Electronic distributed storage and full-text retrieval, solving the problem of too many paper files when targeting one
- Efficient retrieval, almost real-time, and distributed storage can be extended

1. Data Storage and Retrieval

- almost real-time retrieval

- extended storage



2. Risk Early-warning and Hidden Danger Identification

risk early-warning

hidden danger Identification

Risk alert and hidden danger elimination are intelligent recognition of harmful factors and hazards, which may lead to hidden dangers, accidents or faults.

2. Risk Early-warning and Hidden Danger Identification

current measurements:

According to safety instructions, exanimate and eliminate one by one, taking historical record and expert experience into account.

flaw 1

Cannot target hidden dangers precisely, and the examination has a wide scope, but no focus.

flaw 2

Due to different skill levels, people cannot find the real problem.

flaw 3

Only eliminate risks and hidden danger which can be seen, cannot find potential risks effectively.

2. Risk Early-warning and Hidden Danger Identification

Reports of risk, hidden danger, accident and failure

Equipment list

Data of examination and maintenance

Data beyond railway, such as weather

Integrated multi-source data, big data platform analyzes and mines those data deeply, through statistical model, machine learning and deep learning algorithm, to target risks and hidden dangers more efficiently, and furthermore to predict accidents and faults happening in the future.

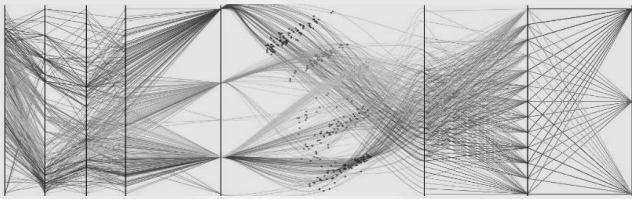
What Our Platform Do

1. Key points stand out
2. Target precisely
3. Identify potential risks and hidden dangers

2. Risk Early-warning and Hidden Danger Identification

Dimension reduction

Project risk and hidden danger data with high dimension into low dimension
-linear method：PCA，LDA
-nonlinear method：Isomap，GDA



Correlation analysis

Correlation among risks, hidden dangers and accidents
-Correlation coefficient：Pearson
-covariance matrix
-information entropy

Time series

Based on historical data, recognize seasonal pattern of hidden dangers, accidents and faults appearance
-AR model
-MA model
-ARMA model

MA model
mt=et - q1et-1 - q2et-2 - ¼ - qqet-q

AR model
Xt=j1Xt-1+ j2Xt-2 + … + jpXt-p + mt (*)
Xt=j1Xt-1+ j2Xt-2 + … + jpXt-p +et

Regression model

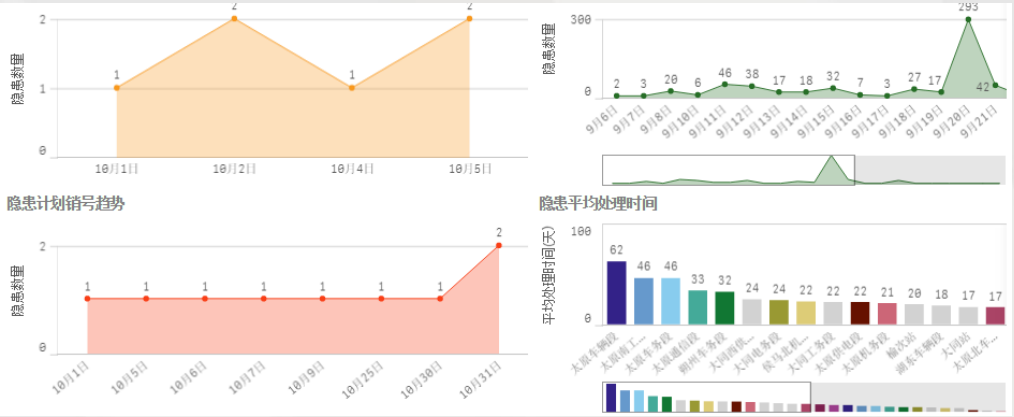
Build a regression model, using risks and hidden dangers to predict accidents in the future
- linear & nonlinear regression

2. Risk Early-warning and Hidden Danger Identification

application within single major-Track

Through analysis of multi-source track data and GIS display, know comprehensively key risk points and hidden dangers along the railway.

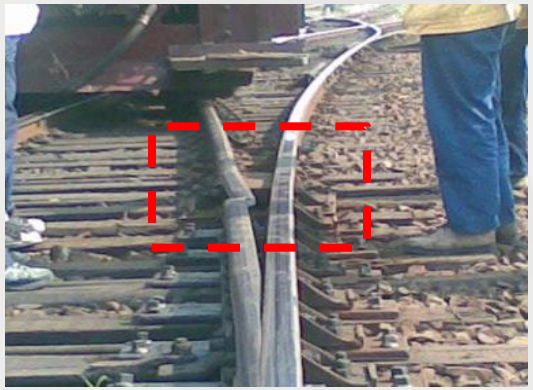
铁路安全业务系统名称	数据内容
工务安全生产管理信息系统	便携超限、车载晃车、人工检查、人工维修、轨检小车、轨检超限、等数据，动态病害库、静态病害库。
防灾系统	风设备报警、风速监测、雨量监测、雨设备报警、地震报警等数据
工务安全监测管理系统	线路监测的病害详情、轨温监测数据、雨量监测数据
PWMIS系统	钢轨、道岔等基础设备台账数据
车载系统	高低、垂加、水加等超限数据
检测车数据综合应用系统	高低、轨向、轨距、水平、三角坑、垂加、水加

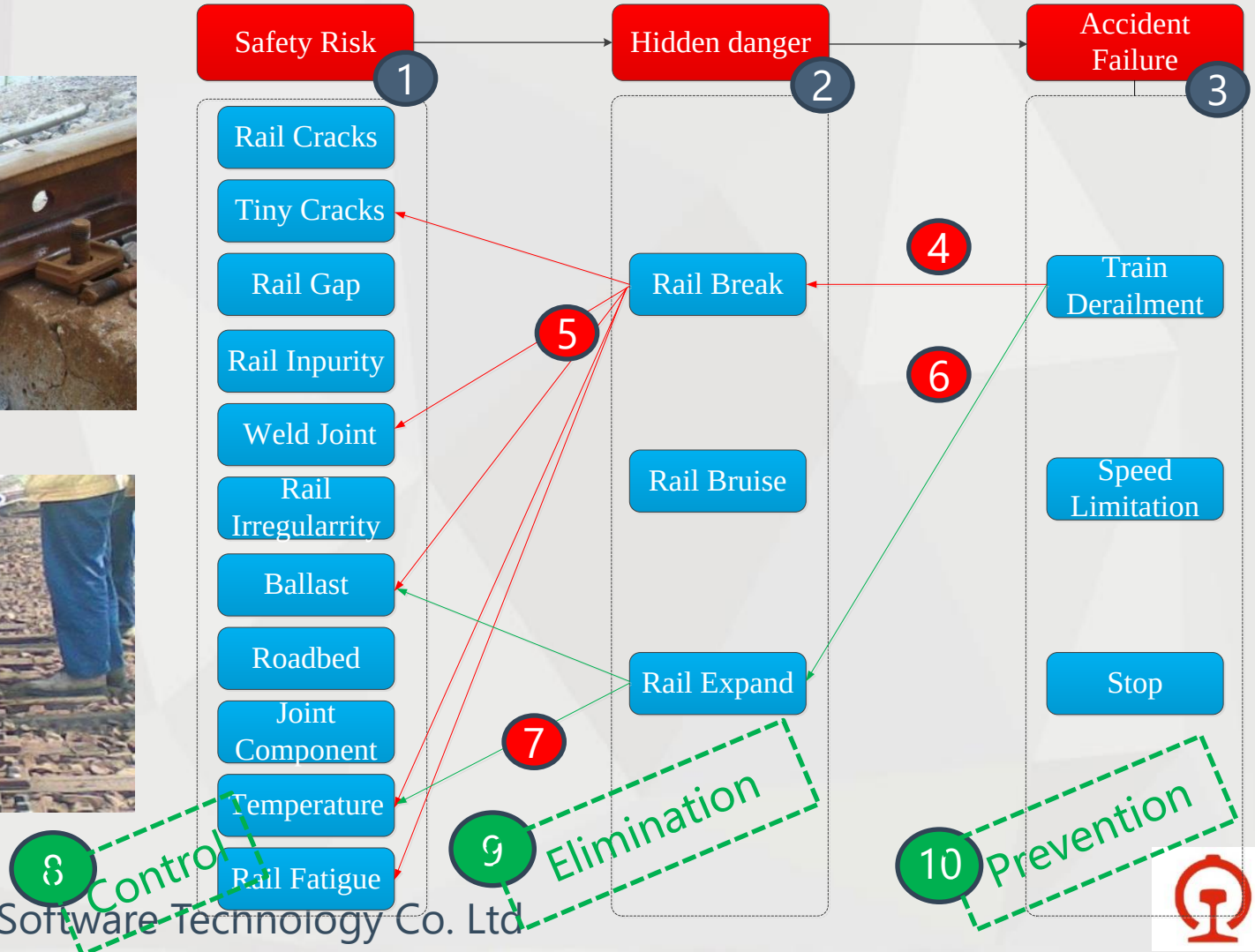


Meanwhile find potential risks and hidden dangers, to instruct risk control and hidden danger elimination.

2. Risk Early-warning and Hidden Danger Identification

application within single major-Track



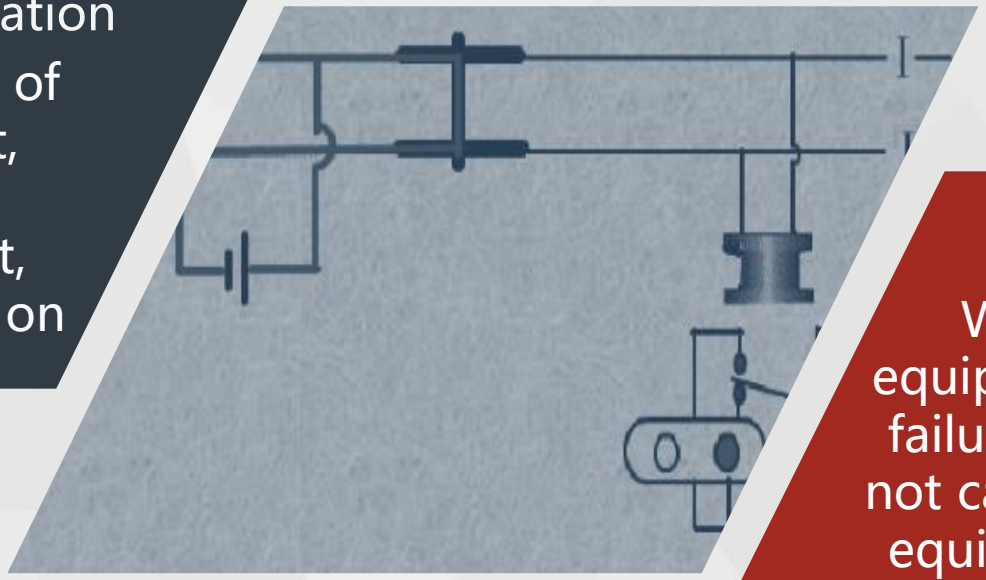


2. Risk Early-warning and Hidden Danger Identification

application across majors-Signal & Communication

Signal and communication

Various type of equipment, such as track circuit, track and so on



Possible causes

When one equipment have a failure, it may be not caused by that equipment itself, but by problems from other majors such as one failure in track.

2. Risk Early-warning and Hidden Danger Identification

application across majors-Signal & Communication

Officer on duty in signal and communication finds the track red light appearance

Signal and communication

Filed check: no vehicle or foreign body

track circuit: normal

Track

This morning just maintain fasteners.

Signal and communication

Finally it is because a short circuit between fasteners and the basic track.

Record ASAP

2. Risk Early-warning and Hidden Danger Identification

application across majors-Signal & Communication

1

the track red light appearance

Phenomenon of accidents or failures

2

Analysis on big data platform

-monitoring data of track circuit
-maintaining data of track circuit

3

-80% caused by wet
-15% caused by insulation damaged
-5% caused by such as rain and thunder

Trace responsibility accidents or failures

4

Improvement and maintenance

Find the corresponding responsible unit to repair and improve it

3. Qualitative Analysis of Accident and Failure

What Our Platform Do

- Deeply investigate accidents and failures, focus on weak aspects such as employees, management, equipment and so on.
- Help supervisors to trace responsibilities.
- Establish knowledge base of accident & failure causes and corresponding measurements, to achieve a better before-event proof and after-event maintenance.

3. Qualitative Analysis of Accident and Failure

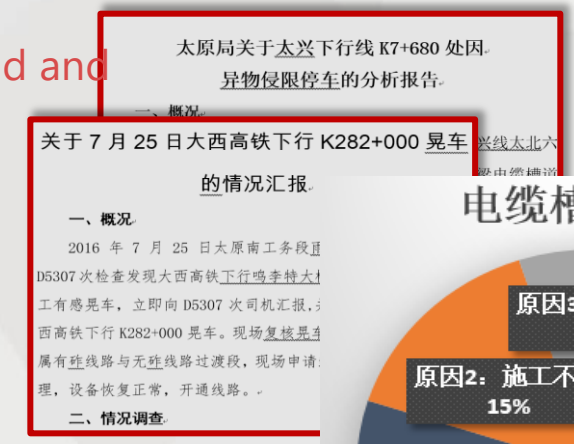
Parse report file , locate and extract the name of accidents or failures, the location, causes and so on
-name entity recognition (NER)

$$P(Y|X) = \frac{1}{Z}\prod_{i=0}^{n-1}\Psi_i(Y_i, Y_{i+1}|X)$$

经调查分析，本次电缆槽盖板刮起侵限的问题，是由于近期太原地区连续暴雨大风的恶劣天气，加之当日凌晨雷雨大风气候，致使桥上 K7+720 处下行线路左侧栏杆外固定电缆槽盖板的抱箍螺栓在大风持续作用下拽脱，抱箍脱落，造成该处一段 6 米长的电缆槽盖板（一端仍由抱箍固定在电缆槽上，另一端翻过 0.9 米高的桥梁栏杆），搭在线路轨枕头部侵限。

Accident Name：
Cable trench cover invasion
Cause:
Strong wind and heavy rain

Intelligent classification of causes
-semantic similarity analysis
-semantic classifiers


太原局关于太兴下行线 K7+680 处因异物侵限停车的分析报告

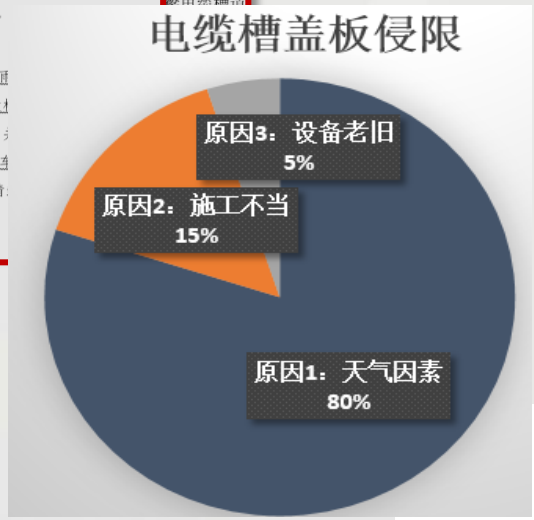
关于 7 月 25 日大西高铁下行 K282+000 晃车的情况汇报

一、概况

2016 年 7 月 25 日太原南工务段

二、情况调查

Calculate possibilities of causes



Investigate and trace responsibility of accidents or failures, based on possibilities of causes

$$p(\mathbf{y}, \mathbf{x}) = \prod_{t=1}^{T} p(y_t|y_{t-1})p(x_t|y_t)$$

4. Improvement and Solution Recommendation

When finding risks and hidden dangers, or after accidents or failure appearance, improvement and maintenance are needed. Traditional methods usually come from paper documents, manuals, or staffs' experience, which leads to difficulties and inefficiency.

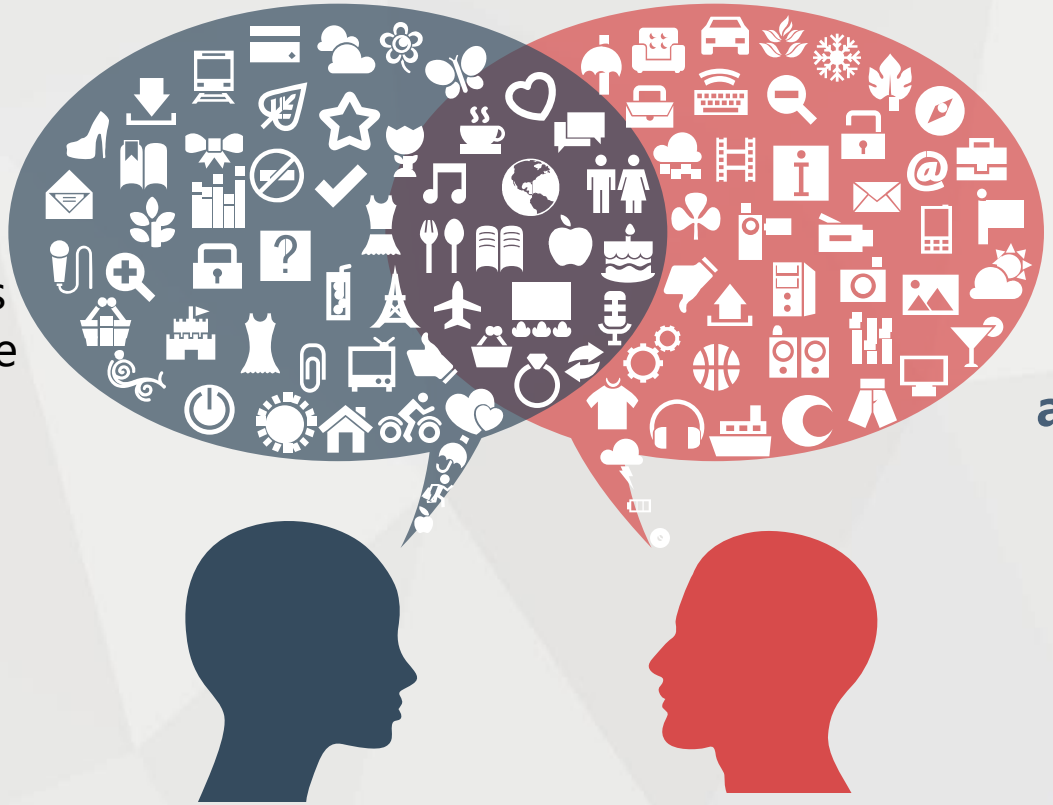
Current Problems

4. Improvement and Solution Recommendation

Intelligent QA

Through organizing non-order materials and information, intelligent question and answer system is building a knowledge classification model, to offer information of railway improvement, maintenance and method.

-anywhere, anytime
-always updated



functions

-related QA
-top questions
-hot words
-question intelligent completion

approaches

-build knowledge base and knowledge map through natural language processing (NLP)
-full-text retrieval of QA

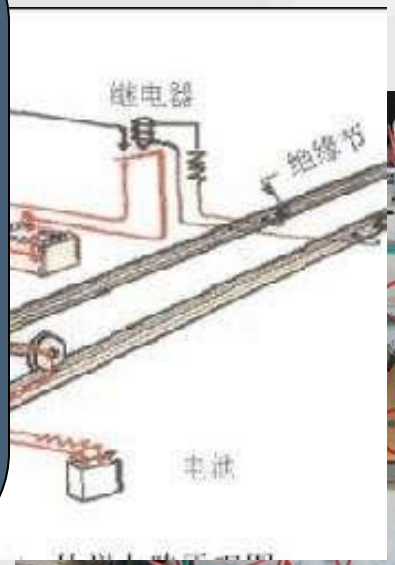
4. Improvement and Solution Recommendation

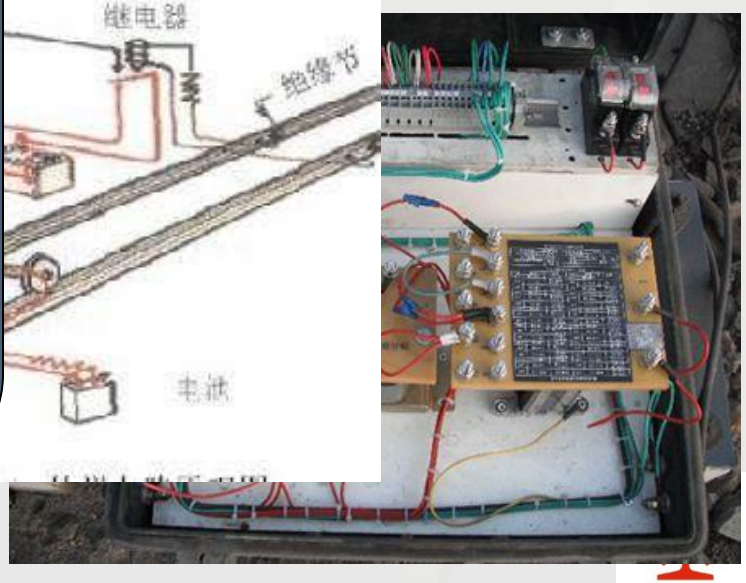
How can I help you ?

the track red light appearance

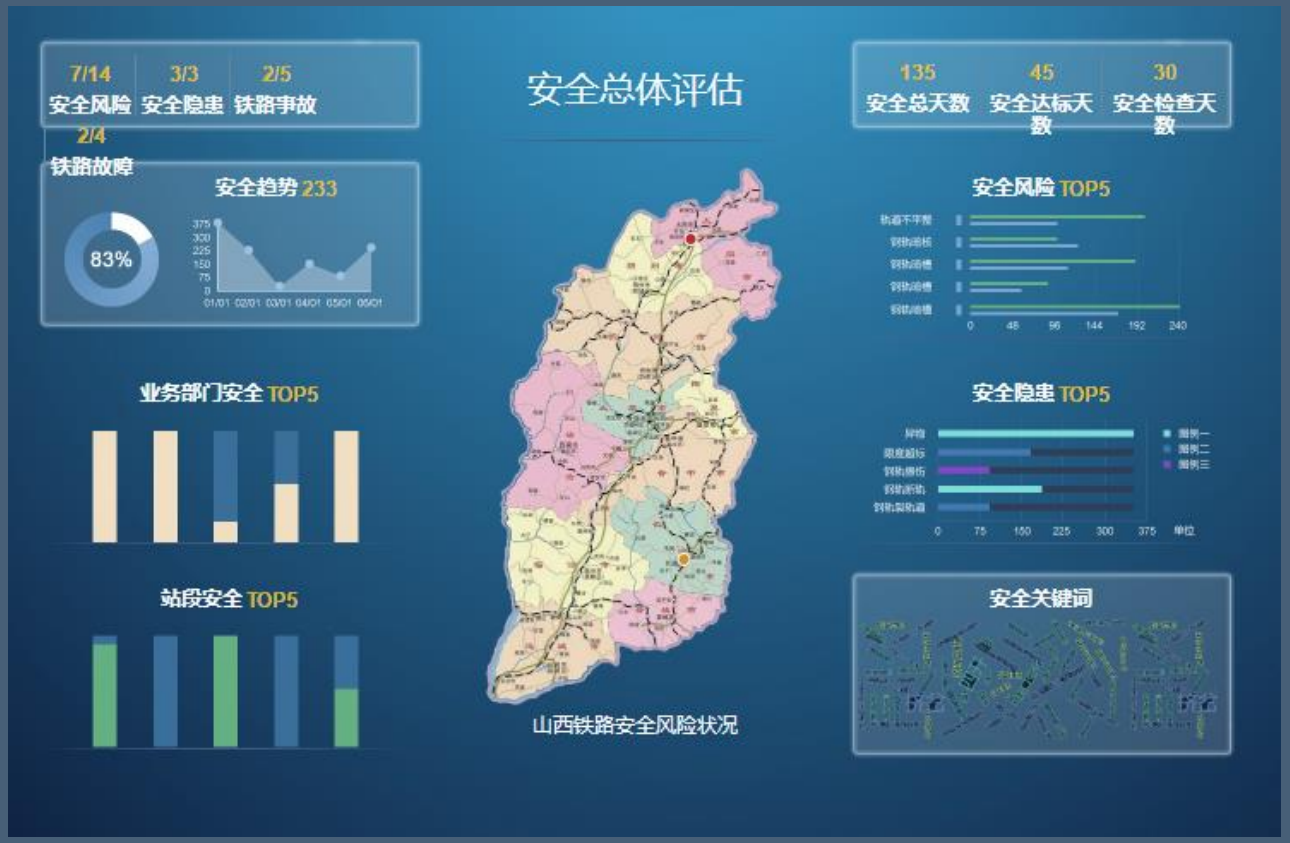
The track red light appearance is a usual accident. The main causes are the following: 1) wet-80%; 2) insulation damaged by human-15%; 3) weather such as rain and thunder-5%.

Measurements: 1) dry; 2) change the damaged part; 3) test and make sure track circuit normal.





5. Overall Evaluation of Safety



Use a dashboard to evaluate the overall safety condition and trend.

Bring advantages for supervisors and corresponding units to know the overall situation about safety.

5. overall evaluation of railway safety

Accidents and failures

Hidden dangers

Evaluating content

Risks
-level of risks, risk number, no. of risks under monitoring

Hidden dangers
-level of hidden dangers, number, no. of eliminated dangers

accidents and failures
-level, number

Safety evaluation standards

Build standards for railway bureau, majors and sections

Effects of maintenance

Risks

CHAPTER 4

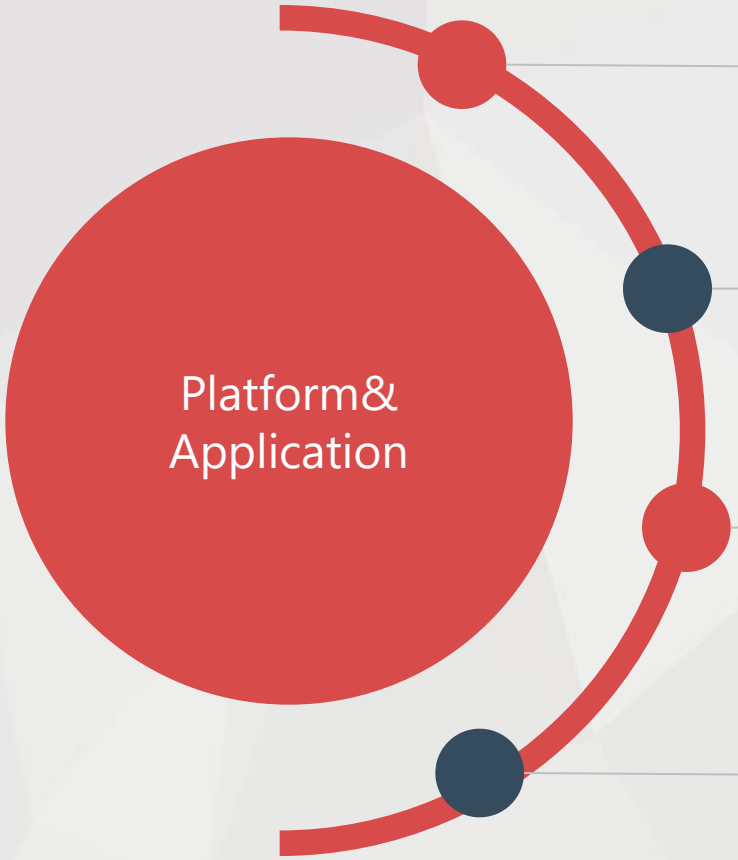
Conclusion

Conclusion

By the RDSP, accidents and failures are reduced by nearly 20%.

Before (without the RDSP)	After (with the RDSP)
Safety data is printed, unstructured	Safety data is digitized, structured
Query and statistics are difficult	Full text retrieval and big data analysis
Identification of risk and hidden danger by human being	Hidden danger and risk identification by Big data Analysis
Passive safety, Rectification after the accident	Active safety, Risk early-warning Overall evaluation

The Future Work



Enrich platform functions and modules

Improve the usability and performance of the platform

Expand the application to passenger and freight service

Encapsulate railway professional data analysis model

The RDSP is Moving Forward

Thank You !