

insight to !nspiration

Samsung SDS

Brightics™ v2.0 Suite

Product Overview



Copyright © 2016 Samsung SDS Co., Ltd. All rights reserved

삼성SDS

SAMSUNG

AGENDA

I. Brightics Overview

II. Brightics Architecture

III. Brightics in Detail

IV. Brightics Use Cases

V. Technical Support



I

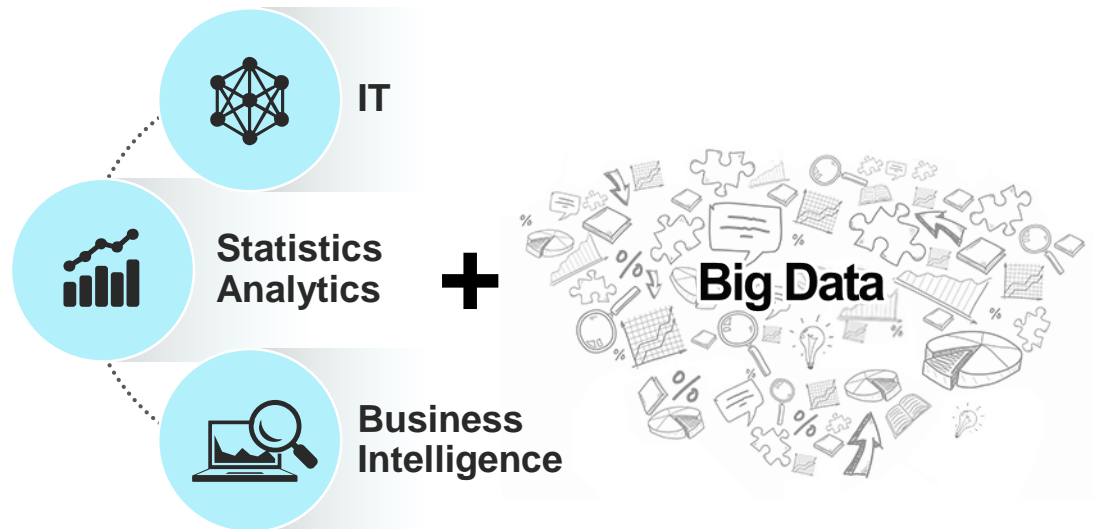
Brightics Overview

Why Big Data Analytics Platform?

Why Brightics?

Why Big Data Analytics Platform?

Analytics platform is required to offer a variety of fundamental big data technologies and integrated user environment.



✓ Requirements

- **Rapid processing and analysis of massive data**, which is impossible with traditional analysis approaches
- **Analysis tool and automation functionality** to support different levels of users and meet different goals
- **Technology suitable for big data visualization** to analyze/verify large-scale data
- **Decision support functionality** optimized for complex customer environment

💡 Required Functionality for Analysis Platform

- ✓ Advanced statistical/analysis functions to analyze massive and diverse data
- ✓ Production environment for easy modeling and rapid verification
- ✓ Interactive visualization of big data
- ✓ Real-time analysis of unstructured data

Why Brightics™ ?

Brightics™ is an integrated platform for optimal big data analysis at complicated enterprise environments.



- ✓ **High performance big data analytics platform to enable rapid analysis of super large volume of data at a low cost** which is impossible with traditional analysis tools
- ✓ **Intuitive and interactive web-based visual analytics environment** to help find potential insights from data
- ✓ **Advanced analysis functionality supporting AI** including prescriptive analytics and deep learning



II

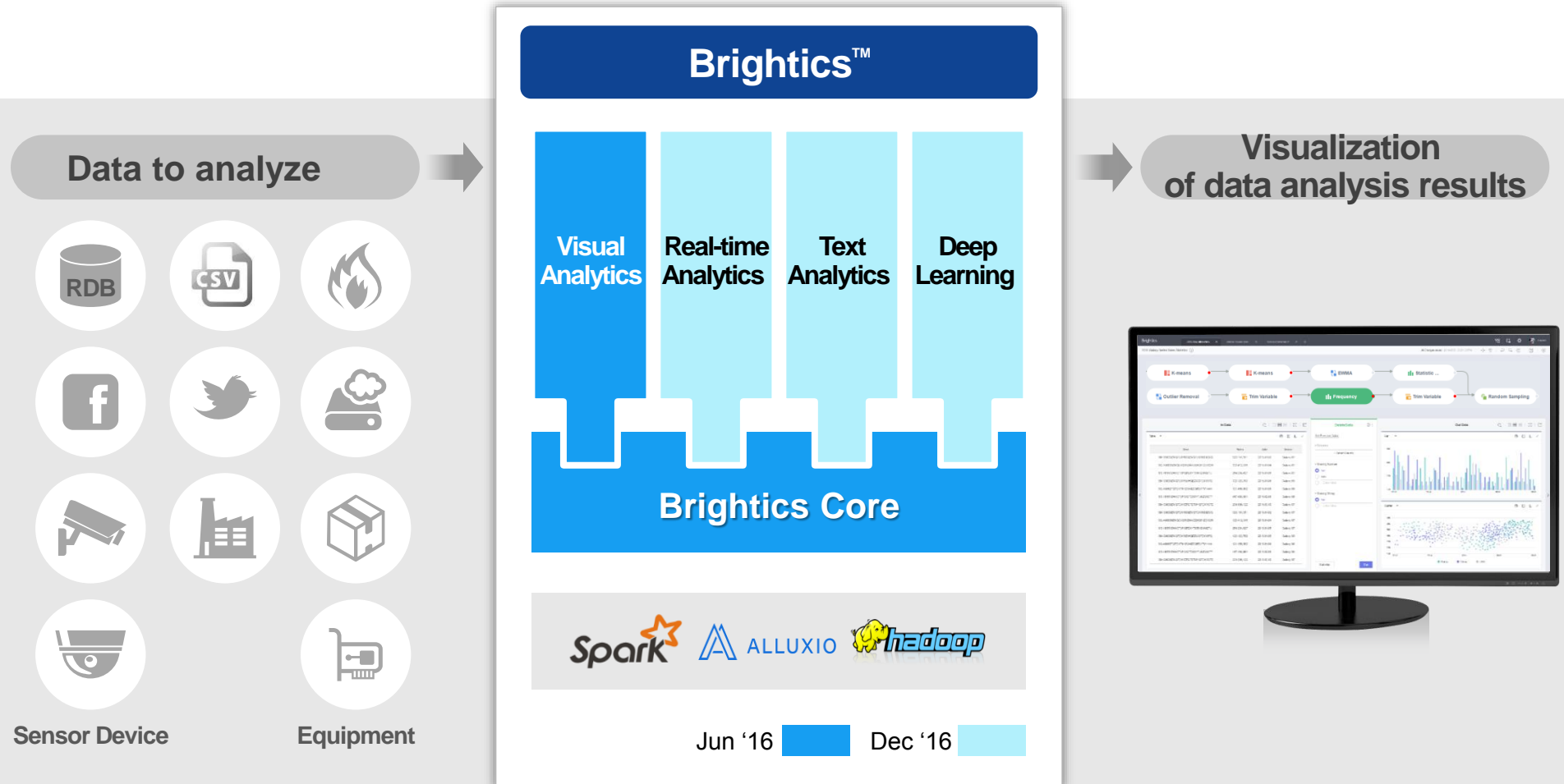
Brightics Architecture

Brightics Architecture

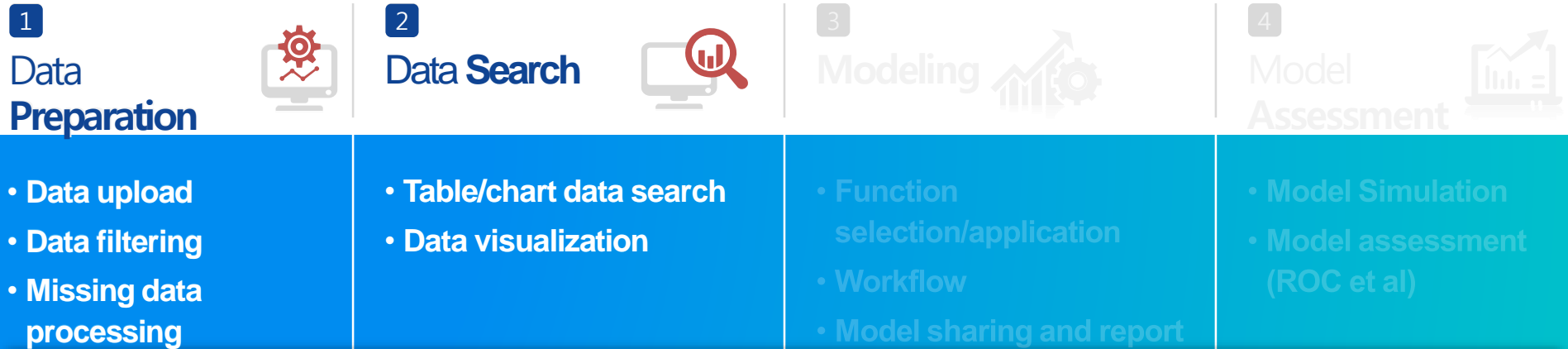
Key Components of Brightics

Brightics™ Architecture

Offers Hadoop ecosystem-based Brightics™ server and user analysis/visualization tool.



Key Components of Brightics



1 Data Preparation

Interface with various data sources

- Support for RDB, HDFS, CSV files
- Real-time data, unstructured text data (2nd half of the year)

2 Data Search

Visual & interactive data search and visualization

- Seamless data interface to analysis data and visualization charts enables intuitive data analysis

Fast distributed data processing

- Diverse high-performance statistics and advanced analysis functions rapidly analyze large-scale data

Key Components of Brightics

1 Data Preparation



- Data upload
- Data filtering
- Missing data processing

2 Data Search



- Table/chart data search
- Data visualization

3 Modeling



- Function selection/application
- Workflow
- Model sharing and report

4 Model Assessment



- Model Simulation
- Model assessment (ROC et al)

3 Modeling

Interface with various data sources

- Integrated analysis environment for visual analytics to see data and workflow at a glance
- Real-time analysis, text analytics, prescriptive analytics and deep learning functionalities to be offered (2nd half of the year)

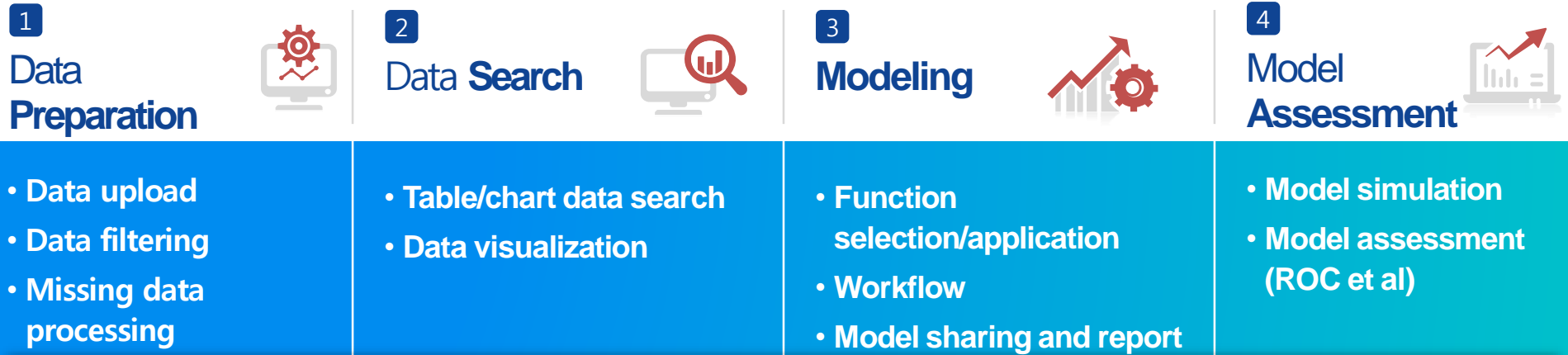
Easy-to-use analysis modeling

- Quick menu (Delete, connect, copy) to minimize the number of mouse clicks in case of creating workflow
- Function search with key word selection and entry and multiple pre-processing functions available

Model sharing and report

- Import and export functionality to share a model and reporting functionality for desired chart and tables

Key Components of Brightics



4 Model assessment

Model simulation

- Simulation results of analysis model according to the selection of data and parameters (To be offered, 2nd half of the year)

Analysis model assessment

- Based on evaluation function, visualize the accuracy of analysis functions with figures and graphs to offer a guideline for suitable model selection

5 Installation/management

Automated installation/configuration

- Service installation for big data analytics in one click available for non-big data experts

Ease of service management

- Real-time monitoring and management over installed services



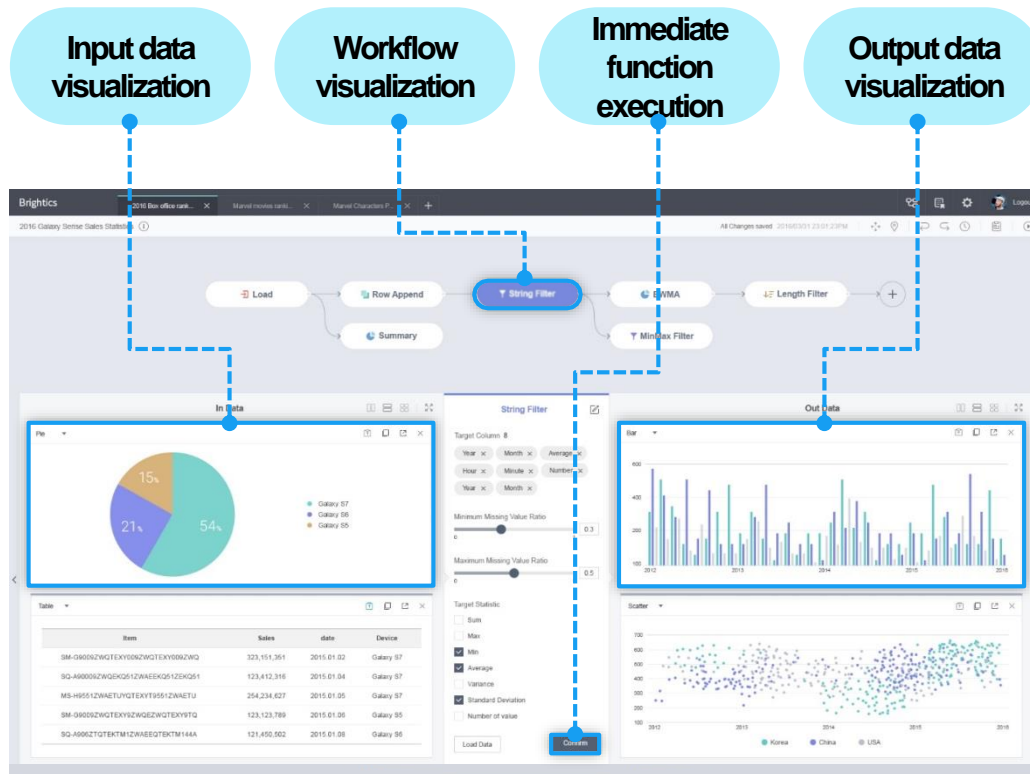
III

Brightics in Detail

User-oriented Web-based Integrated Analysis Environment
Visual & Interactive Data Search and Visualization
Easy-to-Use Analysis Modeling Environment
Automated Installation/Configuration
Ease of Service Management

1) User-oriented Web-based Integrated Analysis Environment

Data and workflow handling in a single window

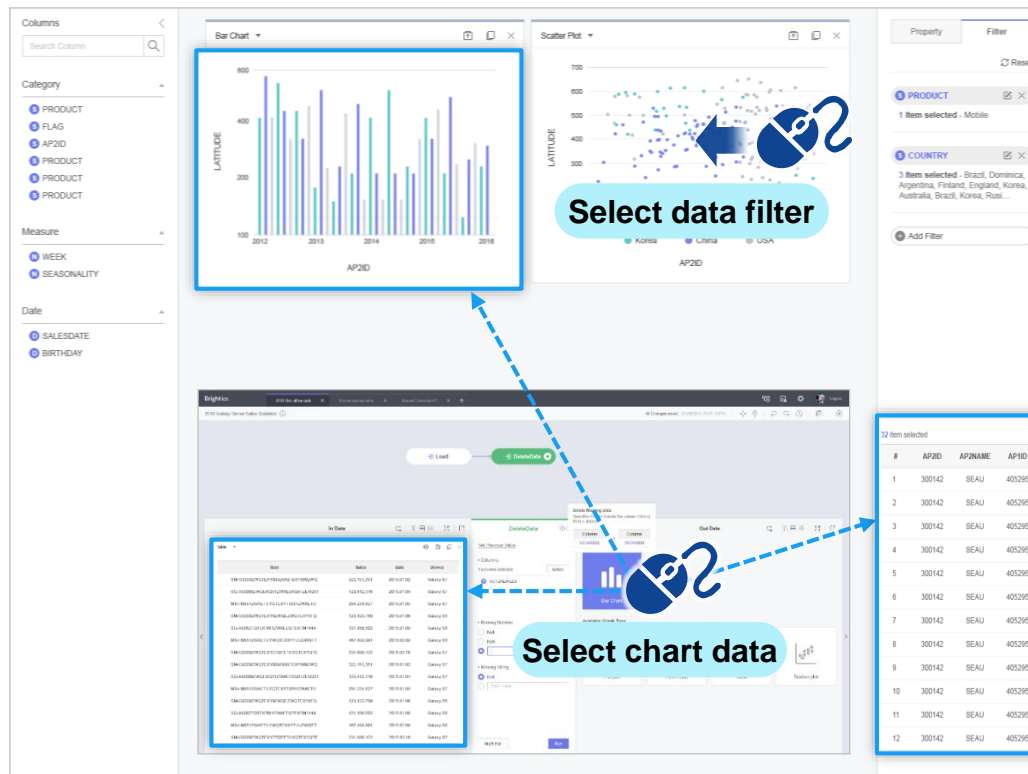


Key Characteristics

- **Workflow visualization**
Visualization of functions by category and execution status to see analysis flow at a glance
- **Input data visualization**
Charts and table-based visualization to select necessary variables for processing/analysis functions
- **Immediate function execution**
Run function immediately after changing variables to enhance the ease of testing
- **Automatic display of outputs**
Automated chart recommendations to best represent outputs of the analysis function

2) Visual & Interactive Data Search and Visualization

Seamless Data/Chart Interface and Report



Key Characteristics

- A variety of charts and reports**
 Analyze entire or partial data with 14 charts from various perspectives and offer reports on outputs
- Visual & interactive data search**
 Seamless interface to analysis data and visualized charts enables to intuitively select and analyze data
- Easy to select and analyze data**
 Data filter enables users to automatically obtain and immediately analyze detailed and additional data

3) Easy-to-use Modeling Environment

Keyword-based function search & assessment of analysis models

You can search other functions with this keywords.

Preprocessing Transform Timeseries **Trend** Extraction Filter

Manipulation **Missing Value** Regression Remove Table schema transform

Any other keywords? **Keyword search**

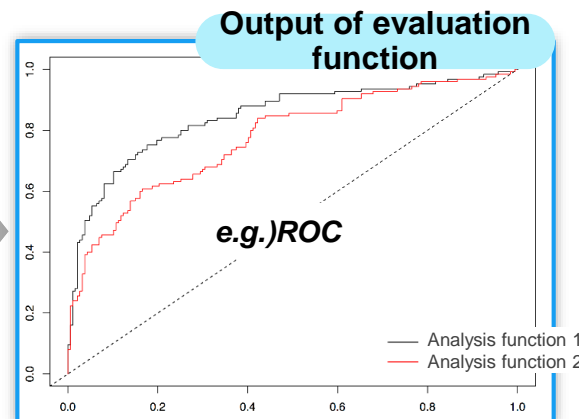
e.g. Logistic Regression

Analysis function 1

Analysis function 2

e.g. Decision Tree

Evaluation function



Key Characteristics

- **Ease of function selection**
Offer function search with keyword entry and selection and detailed description and examples of functions
- **Myriad of large-scale analysis functions**
Up to 120 data pre-processing, cluster analysis/identification, regression analysis functions (to be offered 2nd half of the year)
- **Comparison and assessment of analysis functions**
Based on evaluation function, visualize the accuracy of analysis functions with figures and graphs to offer a guideline on suitable models (2nd half of the year)

4) Automated Installation

Web UI-based easy installation and duplicate analysis server

Installation status monitoring

Advanced installation

Key Characteristics

- **Installation in one click**
General users can easily go through entire installation and configuration process on web UI (Enter only server information and ID/password)
- **Advanced installation**
Advanced users can select and configure desired services only
- **Duplicate analysis server**
Duplicate server for front office provides fault-tolerant services

5) Ease of Service Management

Real-time monitoring and management over installed services

Service	Status	Action	Remove
Hadoop	Running	<button>Start</button> <button>Stop</button>	<button>Remove</button>
Spark	Warning	<button>Start</button> <button>Stop</button>	<button>Remove</button>
Flink	삭제 중	<button>Start</button> <button>Stop</button>	<button>Remove</button>
Tachyon	Stop	<button>Start</button> <button>Stop</button>	<button>Remove</button>
Zookeeper	Running	<button>Start</button> <button>Stop</button>	<button>Remove</button>

Add new service

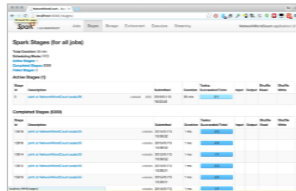
Service management

Job and resource monitoring

Hadoop Job



Spark Job



Resource for analysis



Key Characteristics

- **Service management**
Check whether installed service is properly working and promptly start/stop service
- **Job and resource monitoring**
Monitor Hadoop clusters and diverse job resources
- **Configuration change and prompt application**
Right after modifying memory and clusters, restart and promptly apply changes

....

IV

Brightics Use Cases

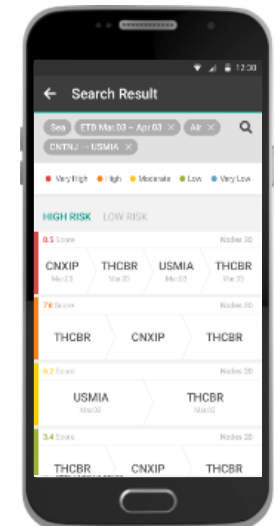
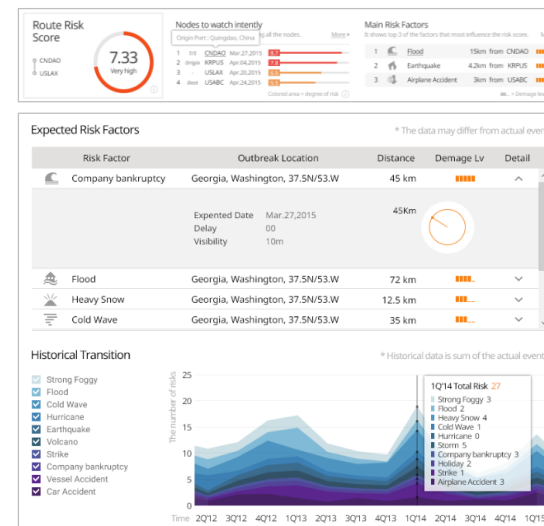
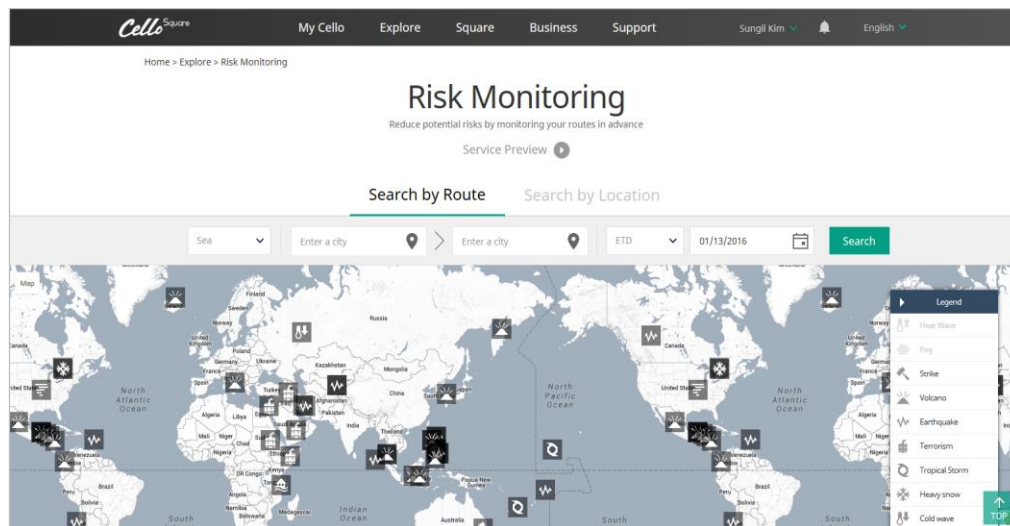
Cello Square Risk Monitoring

Business Values

- ✓ Enhanced credibility of Cello Square's external logistics service by collecting/analyzing logistics risks such as natural disaster, weather, and social issues and assessing shipment risks
- ✓ Assessment of 21 risks in total including marine/air/in-land transportation and structured +unstructured data

Secured technology

- ✓ Text mining: extracts logistics risks-related data from news articles
- ✓ Ordinal classification by the intensity of risks
- ✓ Real-time data collection/analysis: RSOE, OpenWeatherMap, Factiva data
- ✓ Developed logistics risk assessment methodology by leveraging failure mode effect analysis



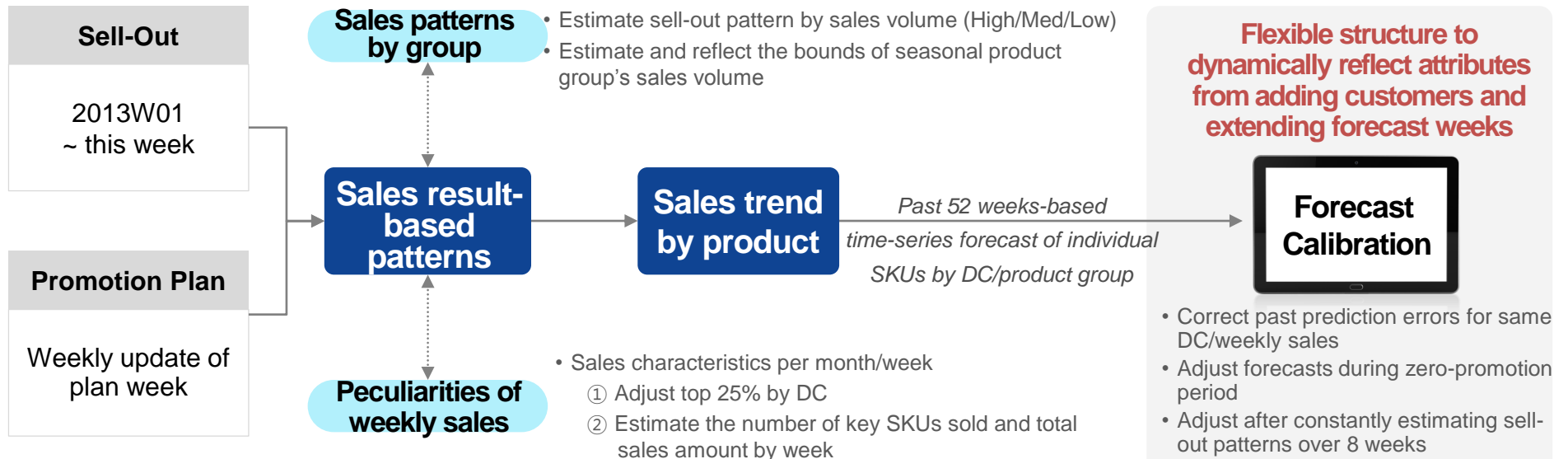
Product Demand/Sales Forecast

Business Values

- ✓ **Faster-decision making with dramatically shorter prediction time driven by automated and advanced analysis**
- ✓ **Suggestion on right promotion timing through demand/sales forecast**
- ✓ **More efficient business management by forecasting new product lifecycle**

Secured technology

- ✓ **Sell-out forecast**
 - Identifies factors affecting sales amount and estimates their impacts to predict sales amount
- ✓ **Forecast calibration**
 - Enhances the forecast accuracy by reflecting promotion effects and time-series forecast of individual models



Enhanced Anomaly Detection Algorithm

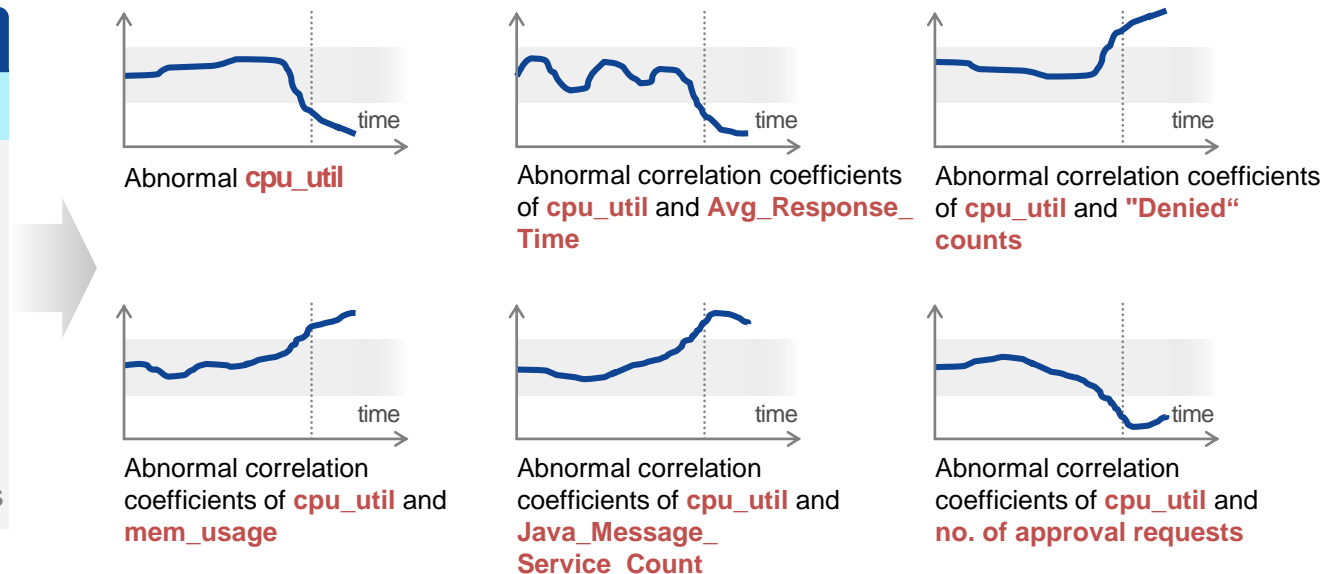
Business Values

- ✓ **Big data analytics-based innovative failure prevention framework responding to the complexity and rapid changes to data centers to create new businesses**
 - Environment for prompt failure analysis (data storage by management entity → data integration by service)
 - Multi-faceted anomaly detection (Threshold → performance and log data analysis-based)

Secured technology

- ✓ **Log analysis-based anomaly detection: computes the times words/patterns appeared and normal range**
- ✓ **Correlation analysis-based anomaly detection: analyzes correlations of time-series data such as performance, logs and application**

Type	metric
DB	cpu_util
DB	mem_usage
WEB-WAS	Avg_Response_Time
WEB-WAS	Java_Message_Service_Count
Logs	"Denied" counts
Application-specific	No. of approval requests



Building Energy Analysis

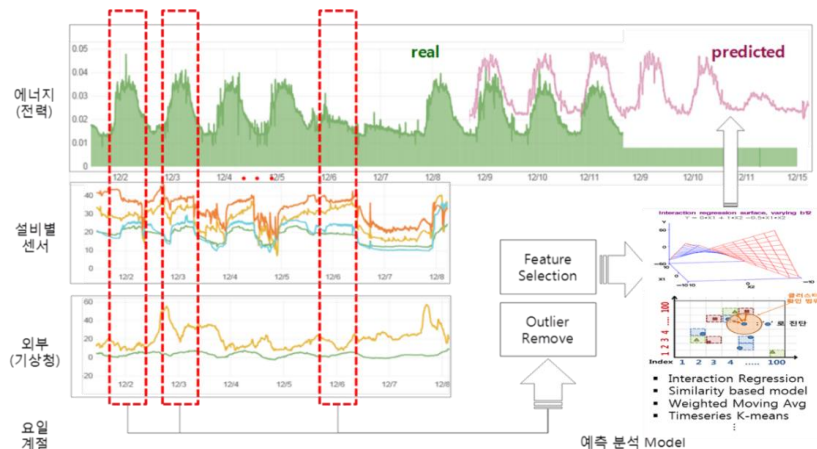
Business Values

- ✓ Energy demand forecast-based peak power decrease to save energy cost and offer operation guidelines
- ✓ Energy leak reduction and suggestion on equipment change timing by leveraging sensor anomaly detection technology

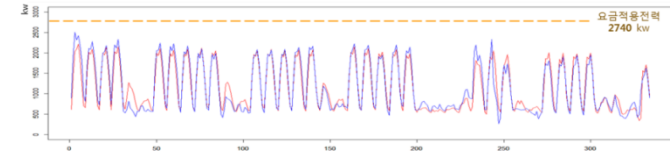
Secured technology

- ✓ Building energy demand forecast
 - Baseline verification (Interaction Regression, Similarity Based Model)
 - Time-series prediction (Time-series decomposition, ARIMA, Weighted Moving Average)
- ✓ Visualization to verify suspect factors
 - Time-series analysis (Time series K-means, Dynamic Time Warping)
 - Anomaly detection (Dynamic Bound, Asymmetric Outlier Remove)

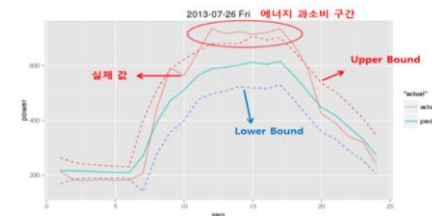
Energy baseline model



Peak power demand estimation



Sensor anomaly detection



Resident's Living Patten Profile Analysis

Business Values

- ✓ Created generation-specific profiles of living pattern by comprehensively analyzing diverse data collected from home
- ✓ Analytics platform for tailored data push service to generation according to profiles
- ✓ Commercialized analytics functionality for B2B business such as linking to B2B advertising business tailored to lifestyle
- ✓ Expanded analytics-based data push service

Secured technology

- ✓ Machine learning and statistical analysis to assume the profiles of each generation
- ✓ Time-series data analysis for event prediction

Families that leave home for school or office between 6 and 7 o'clock in the morning on weekdays



It is likely to rain this afternoon. Please close the window and take an umbrella before you go out.
Don't forget to bring your smartphone.

Senior couples who usually stay at home, not going outside



You have barely gone outside for the past 3 days. At 4 pm, the weather is clear with low UV index. How about taking a walk?

Families with children that go out frequently on weekends



We will have strong winds with high density of fine dust this afternoon. Children are advised to refrain from outdoor activities.

Security Analytics Model Development

Business Values

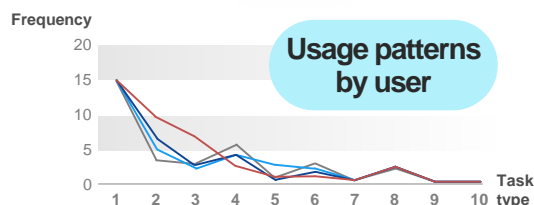
- ✓ Developed malicious code analysis model based on analytics rather than traditional rule-based signature approach to respond to APT (Advanced Persistent Threat) attacks
- ✓ Developed a model to detect signs of (personal) information leak by analyzing employees' document usage patterns

Secured technology

- ✓ Developed a model to detect potentially compromised PCs by analyzing security log data
- ✓ Developed a model to detect a suspicious user who leaked personal information

Security Log

CS호스트명	클라이언트 송신 Byte	서버 송신 Byte	접속시간
www.werpingad.com	251	392	[19/Sep/2014:09:06:45 +0900]
www.werpingad.com	309		[19/Sep/2014:09:07:29 +0900]
www.werpingad.com	309		[19/Sep/2014:09:07:33 +0900]
www.werpingad.com	265		[19/Sep/2014:09:07:43 +0900]
www.werpingad.com	265		[19/Sep/2014:09:08:13 +0900]
www.werpingad.com	220	392	[19/Sep/2014:09:09:32 +0900]
www.werpingad.com	220	392	[19/Sep/2014:09:09:34 +0900]



Attributes factors for extraction

Factors to identify potentially compromised PCs

Data traffic

Access interval/cycle

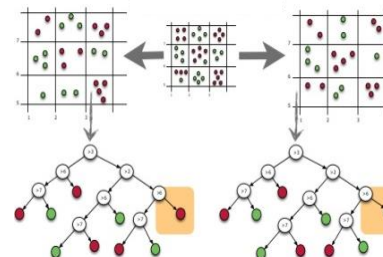
User-agent peculiarities

Document attachment and modification

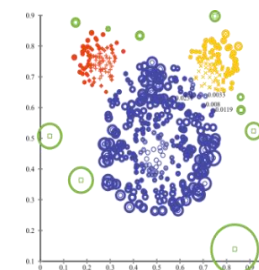
:

Models to detect hacking suspect PCs

Random Forest



Clustering





V

Technical Support

Requirements

Contact

Requirements

- ✓ OS : Linux Multi Node or Single Node
- ✓ CPU : 4 Cores per node
- ✓ RAM : 8G+ per node
- ✓ HDD : 100G+ per node
- ✓ JVM 1.8 or higher
- ✓ Hadoop 2.6
- ✓ Spark 1.6
- ✓ Alluxio 1.0
- ✓ PostgreSQL

- ✓ Web Browser
: Internet Explorer 11 or higher or Chrome



※ 분석 목적 및 데이터 크기에 따라 달라질 수 있습니다

Inquiries

“ Samsung is
the best partner
for your Company ”

☎ Contact

✔ Email

• brightics@samsung.com

✔ Product support

• Jaeyoung LEE (R&D Center) +82-2-6155-3664

✔ Policy support

• EunJoo LEE (R&D Center) +82-2-6155-3680

insight to !nspiration



삼성SDS

www.samsungsds.com

Copyright © 2016 Samsung SDS Co., Ltd. All rights reserved.