Big Data Analytics in Telecommunication

Nokia NET Technologies & Innovation / Norbert Kraft

code::dive conference / Wroclaw / 05-Nov-2015

Agenda

- Intro
- Telecommunication Data & Use Cases — *What we do …*
- Anomaly & Root Cause Detection — *Why we do …*
- Time Series Analysis & Prediction — *How we do …*
- Software & Methods

NOKIA

Short Introduction

- Software researcher & data analyst
- Nokia Technology & Innovation
- Long history in SW development
- Project Leader NDI research project:

‘Network Data Intelligence’

NOKIA

What is a data analyst?

A Person ...

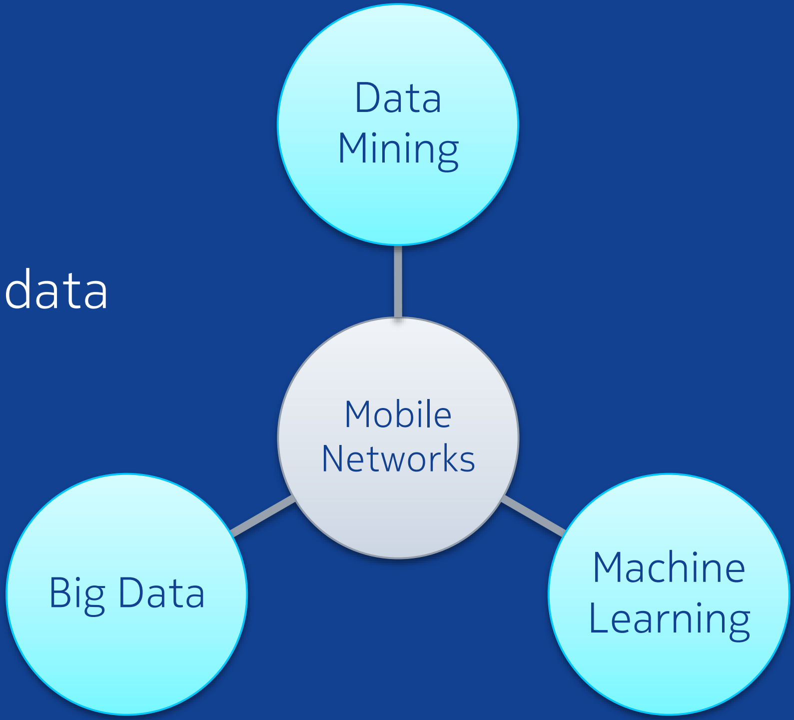
who knows more about programming than a mathematician ...

... and who knows more about statistics than a programmer.

Research Project

Network Data Intelligence

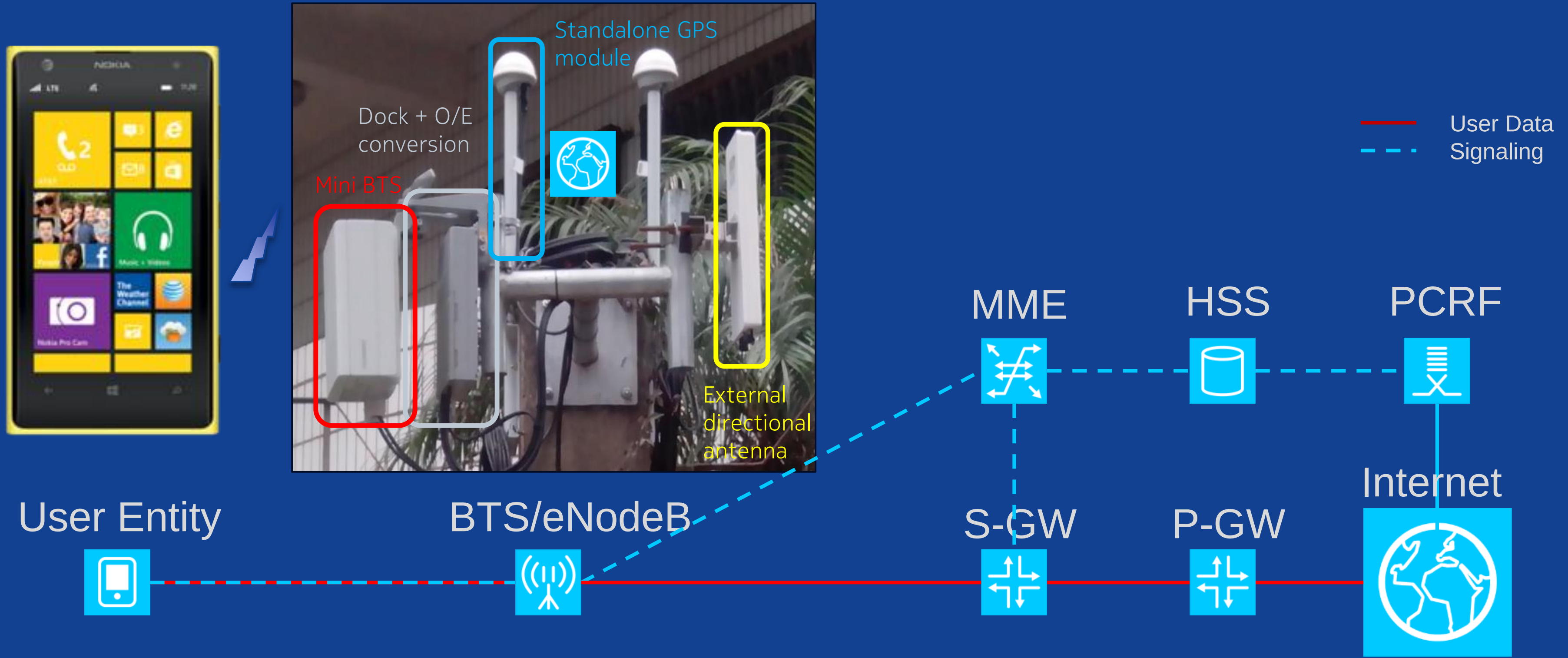
- Nokia research project
- Technology exploration
- Generate new insights in telecom data
- Raise new business opportunities



NOKIA

End to End Mobile Broadband

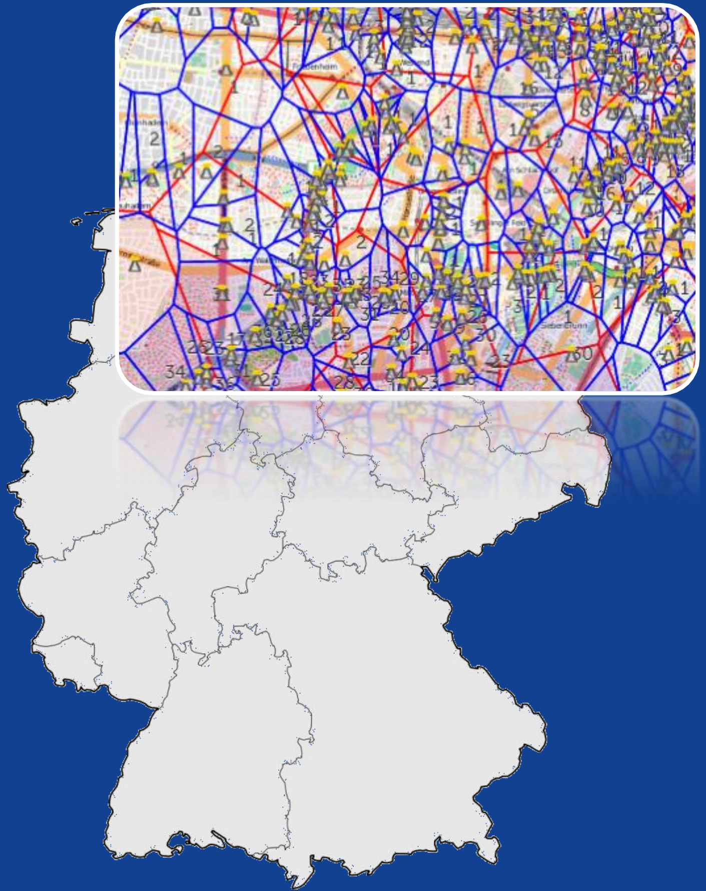
More Than an End Device



Some (estimated) Numbers …
German Telekom (2012)

- 36.6 Million Subscribers for German Telekom
- Total of 113 Million Subscribers in Germany
- ~70 000 Radio Cells in Germany
- ~100 Million GBytes traffic volume (*2011)
- xxx.xxx.xxx.xxx Number of Calls & SMS per Day
- xxx.xxx.xxx.xxx Number of Internet connections
- SmartPhone is always 'ON'

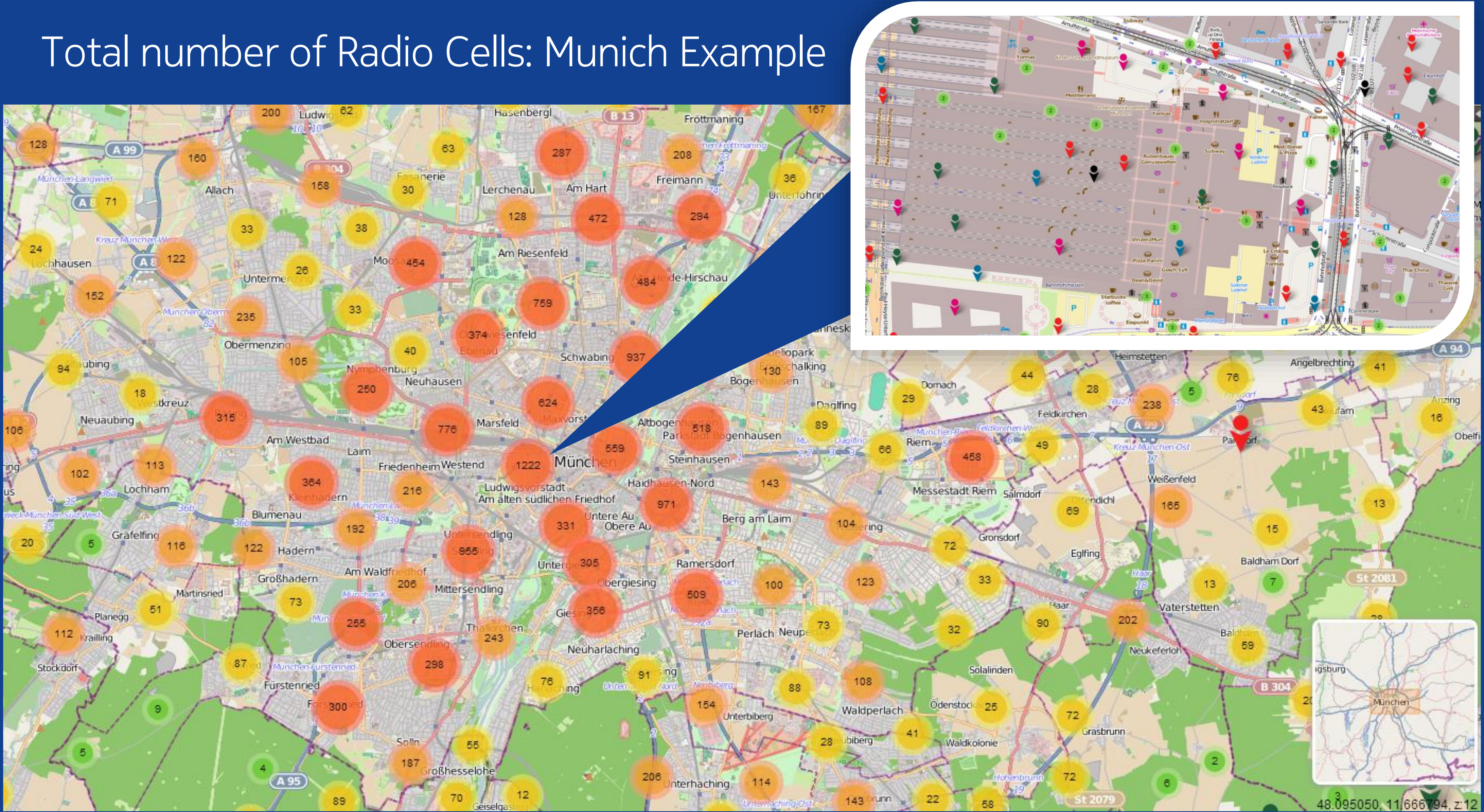
Radio Cell Layout of Munich



Source: Bundesnetzagentur from 2012

NOKIA

Total number of Radio Cells: Munich Example



Some Secret about Big Data (… you might have never heard)
Big data is useless … only Information counts

Id	IMSI	IMEI	Radio cell id	RNC	Error code	IP address	ports	…	…
xxx	2345		45	1	0x0af	10.1.	80		

Big Data

Information

Customer A has some trouble, specifically if he enters radio cell B during business hours, because he is using an old fashioned smart phone with a wrong setup and additionally his office location has a bad radio coverage.

NOKIA

Agenda

- Intro
- Telecommunication Data & Use Cases
- Anomaly & Root Cause Detection
- Time Series Analysis & Prediction
- Software & Methods

NOKIA

structured

unstructured

Highly structured

What the Operator (needs to …) know about …

Mobile Network Data on 'Signaling' not on Content!

User

- **User Identity**: IMSI, A/B No, Device ID & Type IMEI, MSIDN
- **Location**: Cell location, Higher precision w. triangulation on signal strength
- **Service Usage**: URL, IP ports, User Agent, Device OS
- **Data Volumes**: Bytes up/down, Call/SMS Length
- **Personal Data**: Tarif, Revenue, Sex, Address

Network

- Network Element Status
- Configuration Data
- Performance values
- Alarms
- SW Logs & Traces
- CDRs

Attention

Network Data is **Personal Data**

!!!!

But

- Strictly limited by (inter)national laws
- Very complex field under continuous change
- Different views in different countries
- Restrictions on use beyond network management scope
- Usage requires customer permission

Meta data

- Network operators have the right to use this data for management purposes
  - Billing
  - Fault diagnosis
  - Network improvement
  - Support activities

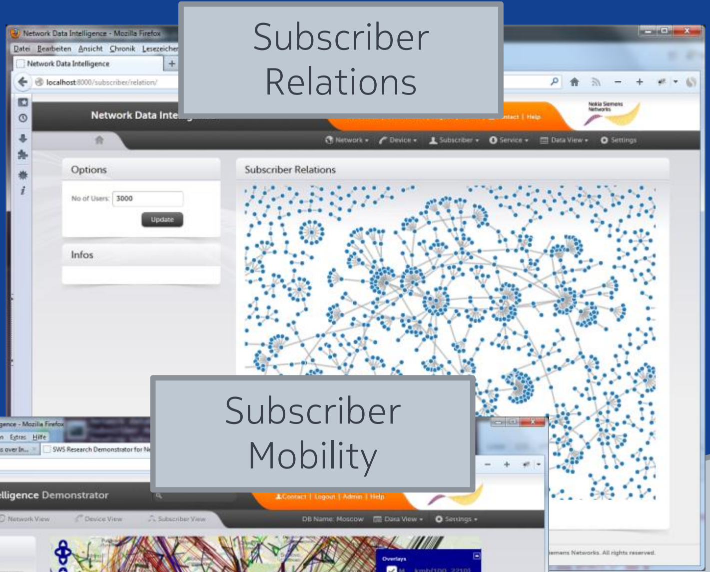
User data

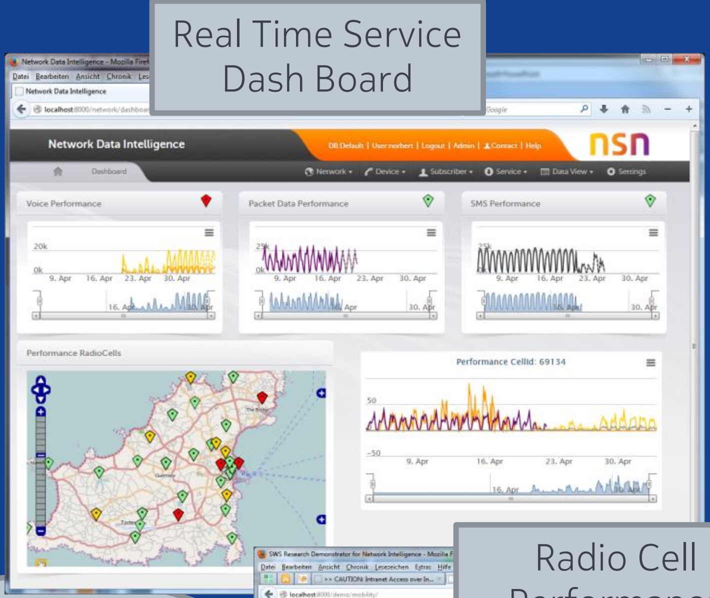
- Strictly limited access allowed with judicial order acc. local laws

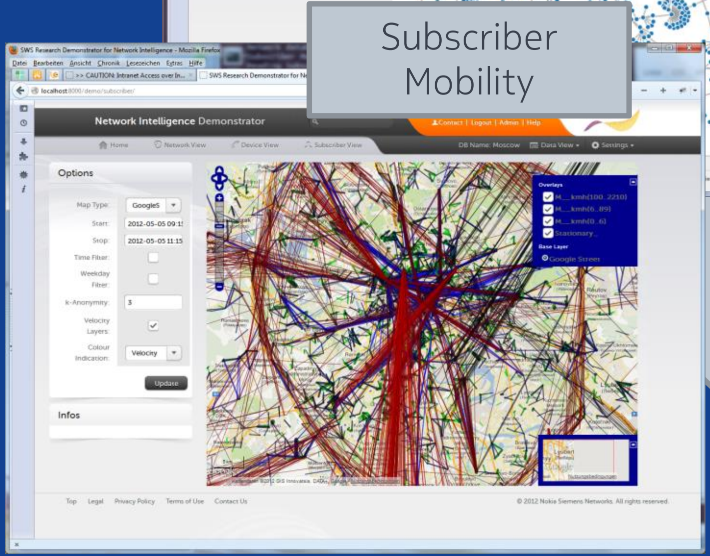
Map of Big Data Analytics Use Cases

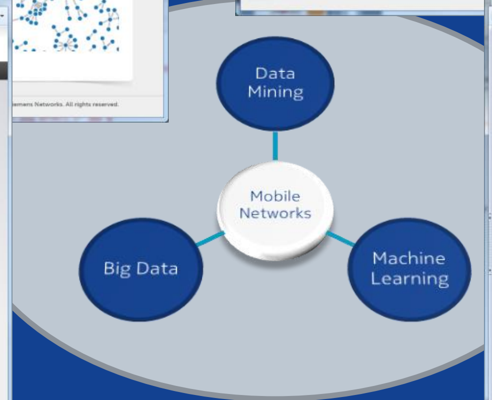
- **Network**
  - Planning
    - Radio cell Performance
    - User Mobility WiFi offload
    - Drop Call Probability
    - High-volume applications
    - High-volume websites
    - Peak data information
    - Roaming analysis
  - Operation
    - Failure analysis
    - Predict network outages
    - Video download experience
    - Service failures
    - Predictive HW maintenance
    - Chronic circuit problems
  - Security
    - Botnet detection
    - Intrusion detection
    - DOS attacks
- **Customer**
  - Product management
    - OTT tracking
    - Tariff simulation
    - Verifying new services, products & devices
    - Viral marketing
  - CRM
    - Fraud detection
    - Churn probability
    - Customer Segmentation
    - Loyalty offers
    - Service Up selling
    - Tracking specific customers (VIPs, dissatisfied)
  - Service
    - First best offer
    - Feedback analysis
    - Discourage SIM swapping
    - Pre-pay recharge message
    - Personalized portal
    - Troubleshooting support
    - Bill shock messages
- **External**
  - Public interest
    - Disaster detection
    - Crowd estimation
    - Traffic supervision
    - Big event management
  - Commercial interest
    - Advertisement
    - User mobility transportation, traffic planning
    - Site protection & supervision
    - Large site management
- **Off topic**
  - Social media
    - Sentiment analysis
    - Peer group analysis
    - Picture recognition information extraction

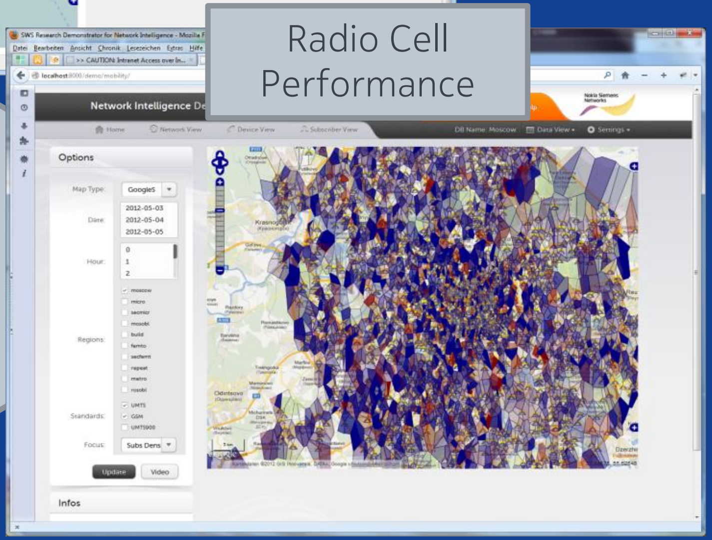
Use Cases from the NDI Project







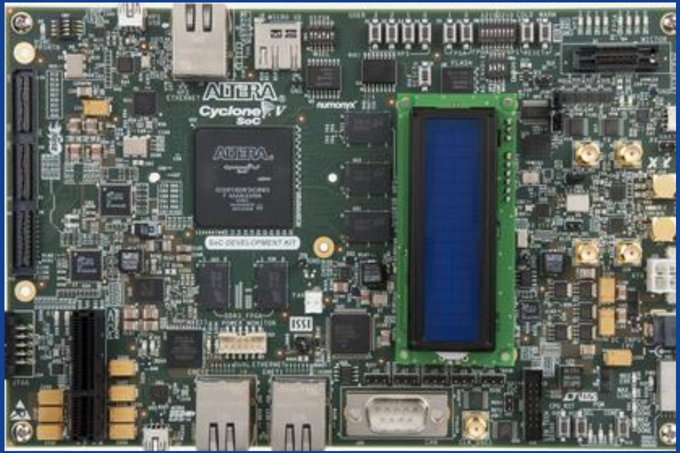


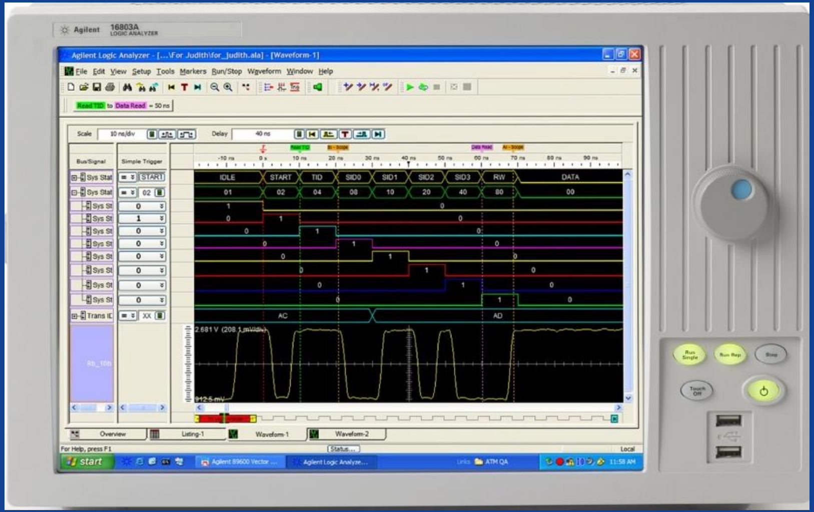


Basic Thoughts

Get to Know what’s going on - Example

Lets assume you are a hardware developer and your board is coming back from prototype manufacturing and it does not work …



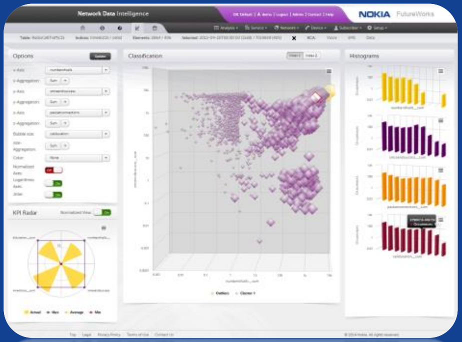


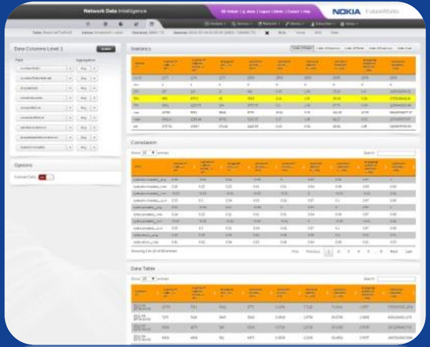
NOKIA

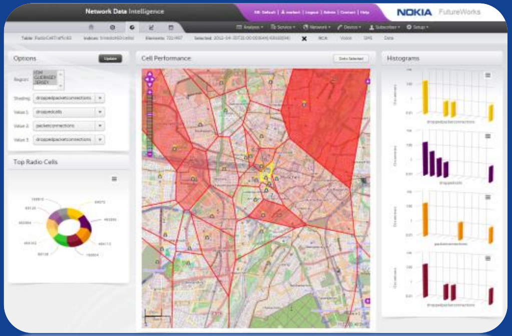
NDI Components

Idea: Analytics Engine

- Universal analysis tool
- Works on every mobile network data
  - Traffica (Nokia Network Data Collector)
  - Service KPIs
  - Log/Trace files
- View concept for specific aspects:
  - Time
  - Radio Cell
  - Parameter correlation
  - Statistics
  - Entity relations
- Applies selectable algorithms: prediction, clustering, regression, (un-)supervised learning, training mode
- Highly interactive

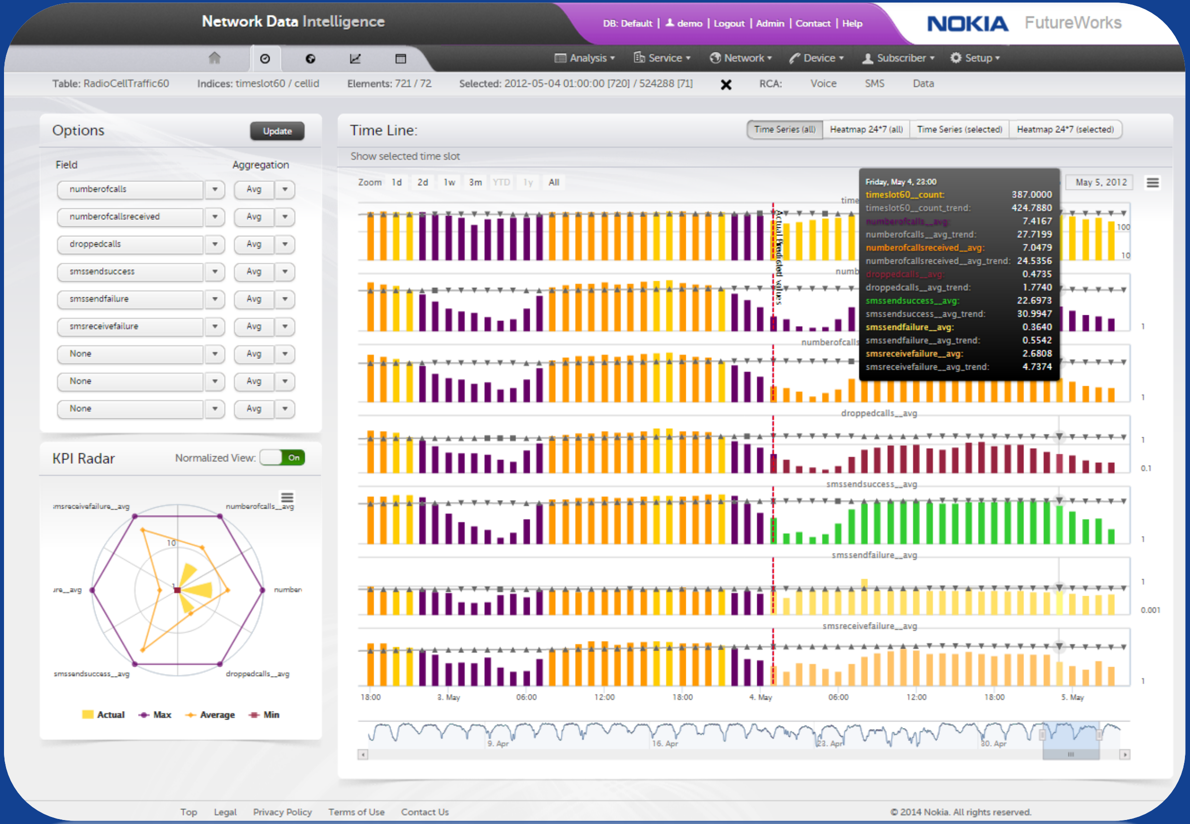






KPI Prediction & Time Slot Classification

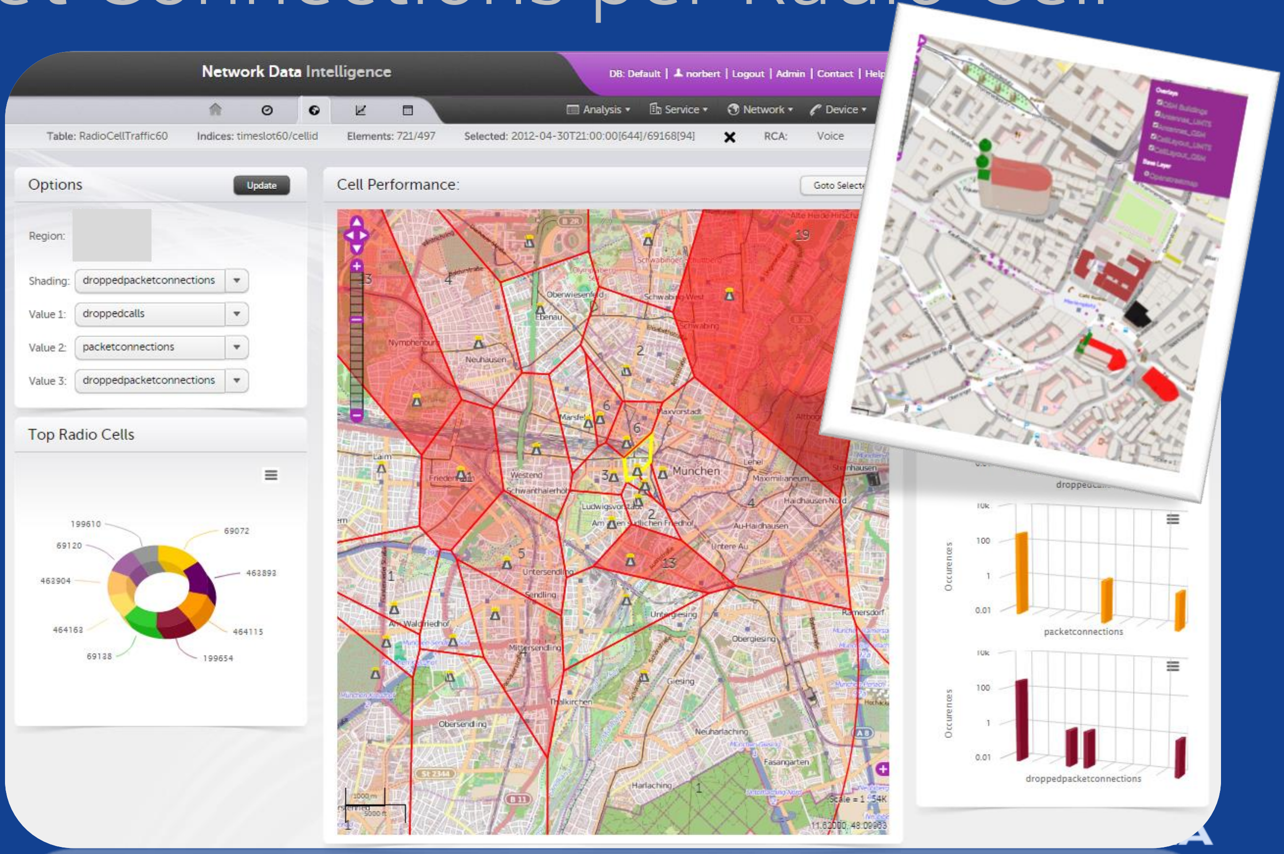
- Time slot classification on history
  - Normal behavior
  - Outliers
- Long term trend analysis
- KPI radar & prediction



Network Use Cases

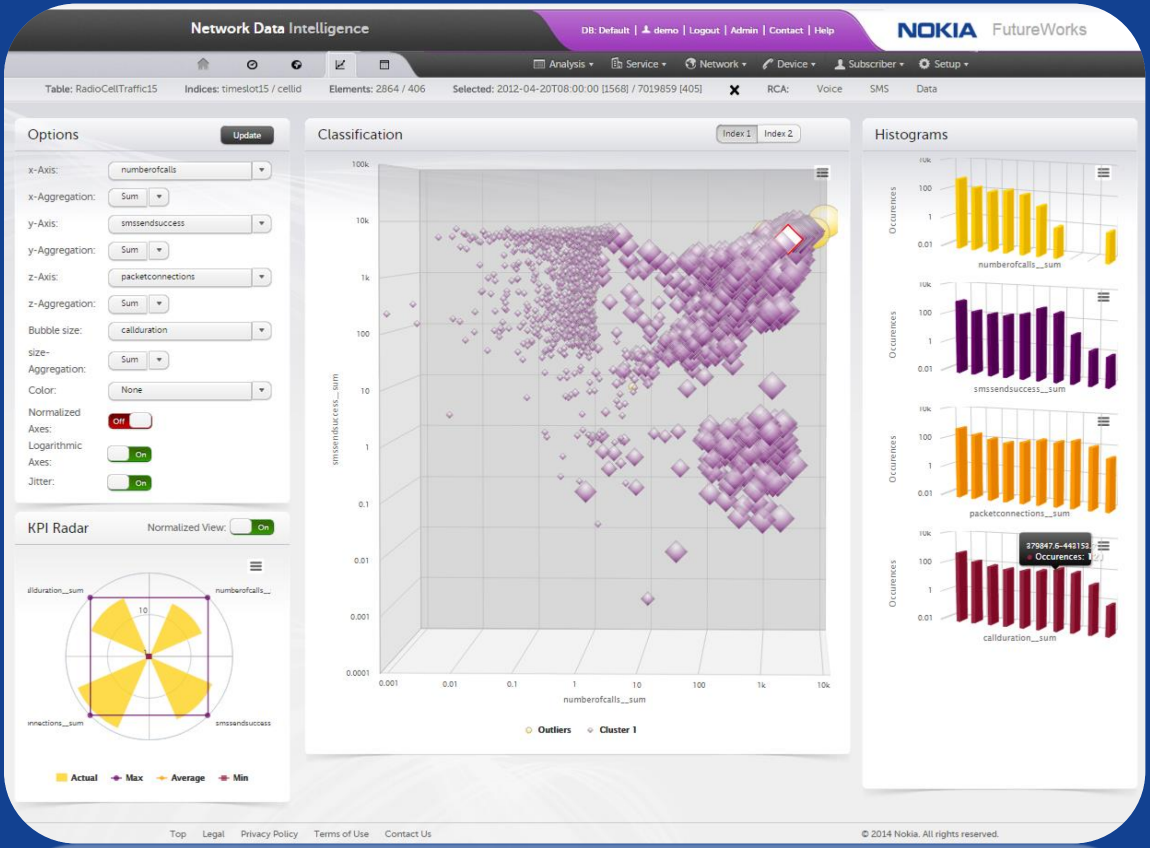
Dropped Packet Connections per Radio Cell

- Important SLA criteria
- History Needs to be continuously monitored
- Prediction turns monitoring to preventive activity
- Correlate network problems with 3D buildings



Network Use Cases
Parameter Correlation

- Show as many dimensions as possible
- Show relations between data



Agenda

- Intro
- Telecommunication Data & Use Cases
- Anomaly & Root Cause Detection
- Time Series Analysis & Prediction
- Software & Methods

NOKIA

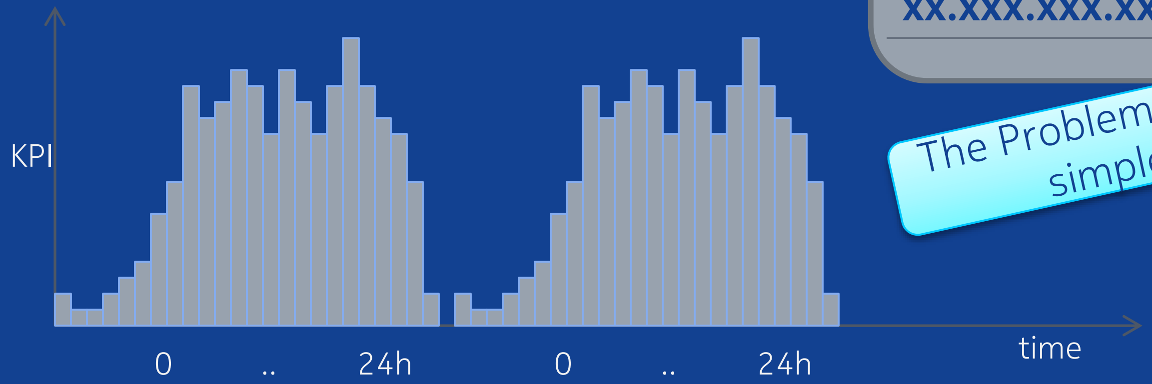
KPI Time Series

What does it tell you ?

xxx KPIs
x.xxx Hours, 15m intervals
xx.xxx Radio Cells, nodes

xx.xxx.xxx.xxx Time Series

The Problem: A huge amount of simple information



KPI
- Drop rates
- Alarm counters
- Performance values

Time
- Resolution: 15min, 1hour, 1 Day
- N months history

NOKIA

The Problem

… more information does not always help

There must be something more intelligent

NOKIA

KPI Time Series

Operation Challenges

What KPI is operating in (un-)normal state?

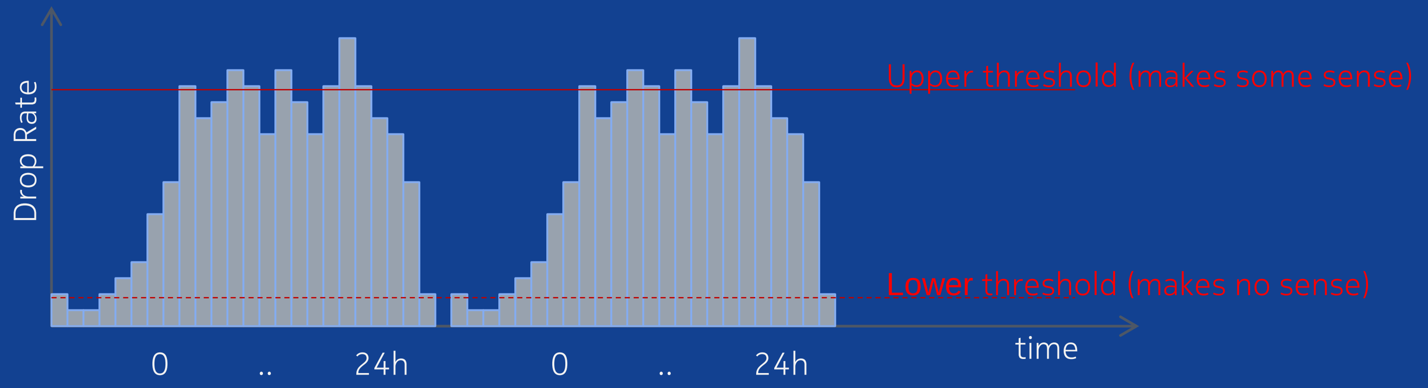
Critical situations: When, Where?

What is the reason for that?

What will happen tomorrow?

KPI Time Series

Operation Challenges - Answers Today: Thresholds



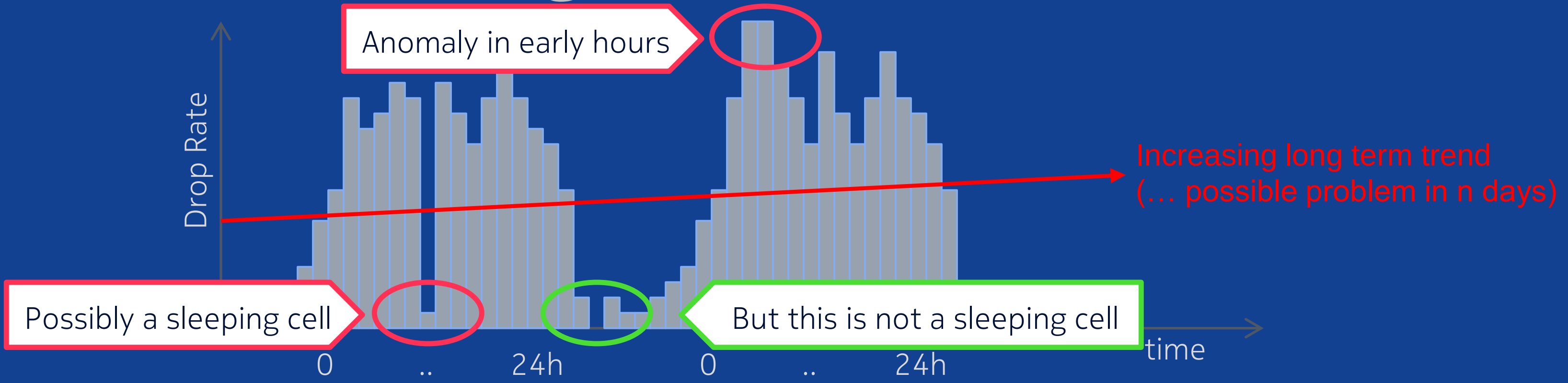
Weak Concept ...

- Thresholds depend on time, radio cell ...
- Hard to find the right boundary between good/bad
- Does not reflect seasonality
- Generate a lot of false positives

NOKIA

KPI Time Series

Operation Challenges - Hidden information



Solutions today:

- Experienced 'eyes'
- But
  - Too much KPIs … hundreds
  - Too much cells … thousands
  - Not enough experienced 'eyes' …
  - Error prone

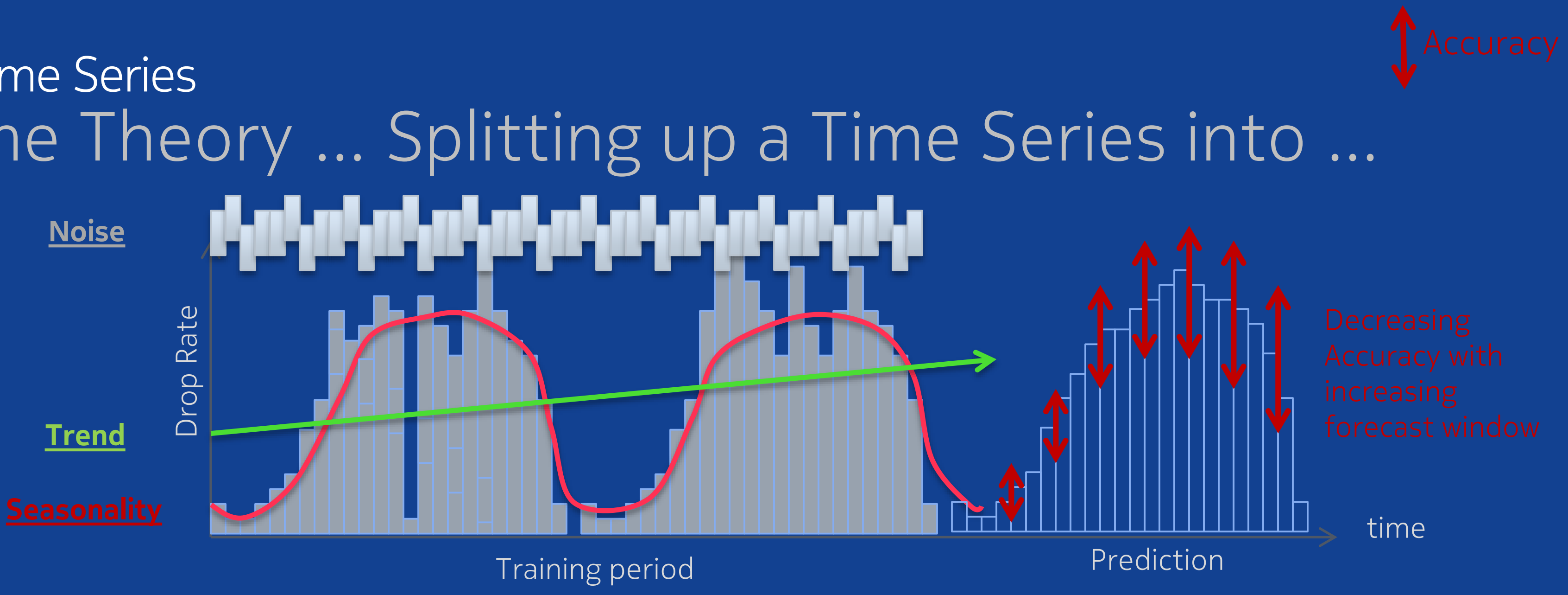
An intelligent system should tell me:

- Where to look at …
- Where are reasonable anomalies …
- Where are dependencies to other KPIs, entities
- What are possible reasons for this behavior

NOKIA

KPI Time Series

Some Theory … Splitting up a Time Series into …



Split system signal into
- Season component (day, week, month)
- Trend
- Noise

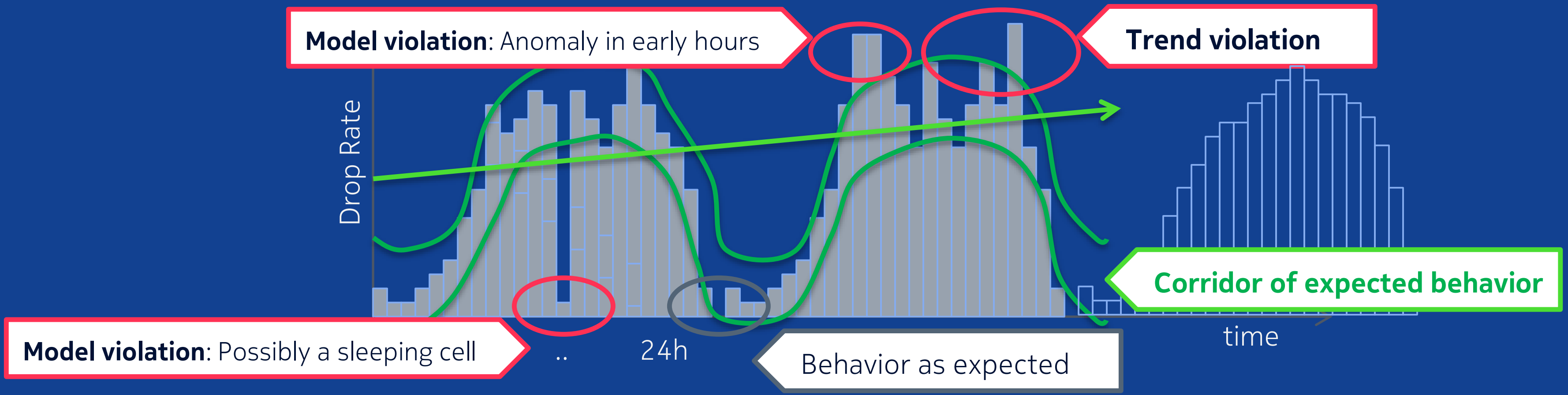
Set upp./low. threshold
- Season component
- Trend

Corridor of expectations

NOKIA

KPI Time Series

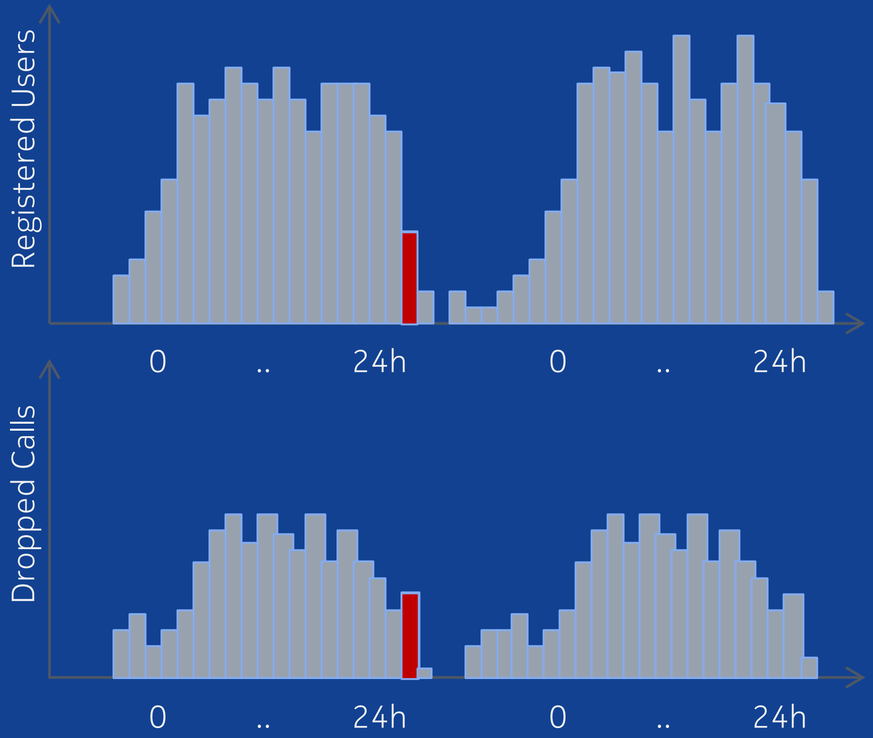
Corridor of ‘expected behavior’



NOKIA

Multi KPI Problems

Finding Anomalies between KPIs



- In most cases the ratio **dropped calls / registered users** is in a similar range
- By using clustering **outliers** can be detected
- Nevertheless the single KPIs might be in the trusted/expected corridor

NOKIA

Why Prediction is so Important

Some Facts …

Business Facts

- Radio cell outage impacts a large amount of customers
- Even if you rapidly realize a problem …. it is always too late …
- High acquisition costs for new customers

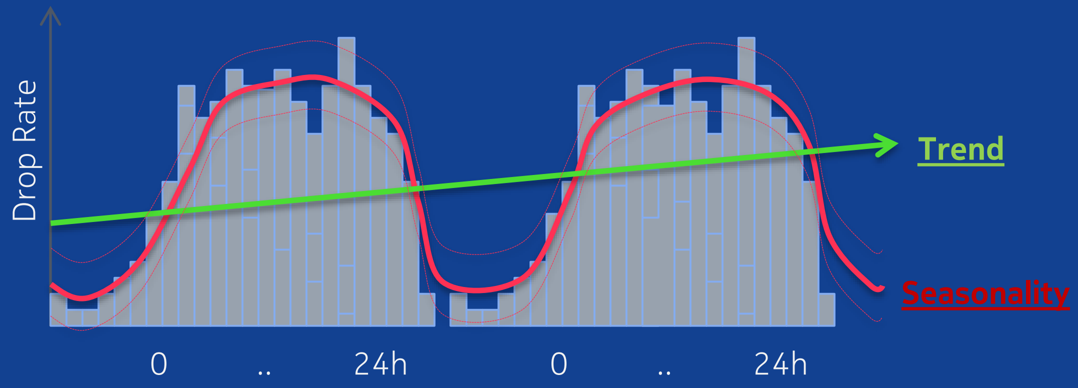
Idea: Change from ‚reactive‘ mode to ‚preventive‘ mode

- Make an estimation which cells are in trouble next time
- Trigger preventive maintenance actions

NOKIA

KPI Modelling

Now machine learning comes into play



Sizing & computation efforts increase ...

- 1 model per cell & KPI
- x.xxx.xxx number of models

Algorithms

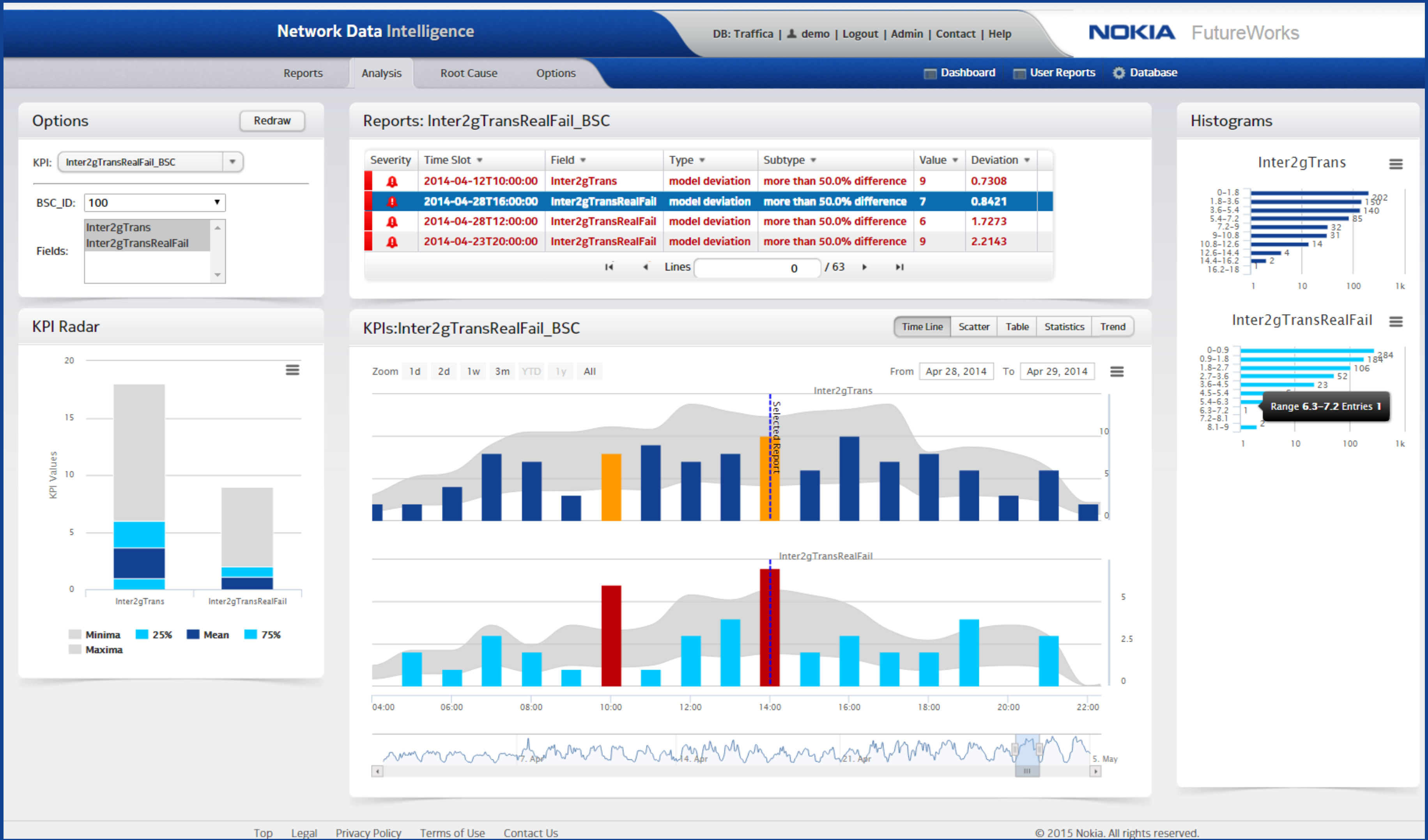
Trend

Linear Regression

Seasonality

k Nearest Neighbor or
Decision Tree or
(non)linear Regression
Logistic Regression

NOKIA


Network Data Intelligence
DB: Traffica | demo | Logout | Admin | Contact | Help
NOKIA FutureWorks
Reports
Analysis
Root Cause
Options
Dashboard
User Reports
Database
Options
Redraw
KPI: Inter2gTransRealFail_BSC
BSC_ID: 100
Fields:
Inter2gTrans
Inter2gTransRealFail
Reports: Inter2gTransRealFail_BSC
Severity
Time Slot
Field
Type
Subtype
Value
Deviation
2014-04-12T10:00:00 Inter2gTrans model deviation more than 50.0% difference 9 0.7308
2014-04-28T16:00:00 Inter2gTransRealFail model deviation more than 50.0% difference 7 0.8421
2014-04-28T12:00:00 Inter2gTransRealFail model deviation more than 50.0% difference 6 1.7273
2014-04-23T20:00:00 Inter2gTransRealFail model deviation more than 50.0% difference 9 2.2143
Lines 0 / 63
Histograms
Inter2gTrans
Inter2gTransRealFail
Range 6.3–7.2 Entries 1
KPI Radar
KPI Values
Inter2gTrans
Inter2gTransRealFail
Minima
25%
Mean
75%
Maxima
KPIs:Inter2gTransRealFail_BSC
Time Line
Scatter
Table
Statistics
Trend
Zoom 1d 2d 1w 3m YTD 1y All
From Apr 28, 2014 To Apr 29, 2014
Selected Report
04:00 06:00 08:00 10:00 12:00 14:00 16:00 18:00 20:00 22:00
Top
Legal
Privacy Policy
Terms of Use
Contact Us
© 2015 Nokia. All rights reserved.

Problem solving

A Simple Thing

- Find problem
  - Normal operation?
  - Hundred of KPIs
  - Thousands of cells
- Show deviations

Detect √

Explain ?

- Problem location
- Involved systems
- References to
  - Customer
  - Device type

- Change parameters
- Fix problem
- Invest

Solve

NOKIA

Why something happens …

Root Cause Analysis

Having a KPI is not sufficient …

- Drop call rate in cell A is too high
- Huge amount of handover failure during busy hours

Reasoning is important …

- Why does this happen?
- What is driving this problem?
- Is it a periodic or a one time problem?
- What do I need to do, how can I change it?

NOKIA

Root Cause Analysis

Information Complexity – Just a normal case: Handover failures

Questions:
- Errors specifically related to time, BSC, cell, end device
- Any typical errors (DX_CAUSES)

Timeslot60	Inter2gTransRealFail	BSC_ID	CELL	DX_CAUSE	LAC	PRB	MM_CAUSE	TAC	HOUR	M15	M05	M
2014-04-16T14:00:00	1	343	45831	2835	34301	306		01371900	14	0	2	14
2014-04-16T15:00:00	0	341	43993	778	34100	306	17	35477004	15	3	11	59
2014-04-16T15:00:00	1	351	56983	2835	35100	306		01264600	15	3	11	59
2014-04-16T15:00:00	0	342	48171	778	34202	306	17	35750805	15	3	11	59
2014-04-16T15:00:00	0	343	47493	3328	34300	306		35461002	15	3	11	59
2014-04-16T15:00:00	0	343	47593	778	34300	306	17	01362900	15	3	11	58
2014-04-16T15:00:00	0	342	40702	778	34201	306	17	35491105	15	3	11	58
2014-04-16T15:00:00	1	342	42221	2842	34200	306		35907605	15	3	11	58
2014-04-16T15:00:00	0	343	42593	3336	34302	306		35635505	15	3	11	58
2014-04-16T15:00:00	0	342	40611	778	34201	306	17	35487901	15	3	11	58
2014-04-16T15:00:00	0	343	45571	778	34300	306	17	35957605	15	3	11	58
2014-04-16T15:00:00	0	342	40051	778	34202	306	17	35907605	15	3	11	58

KPI

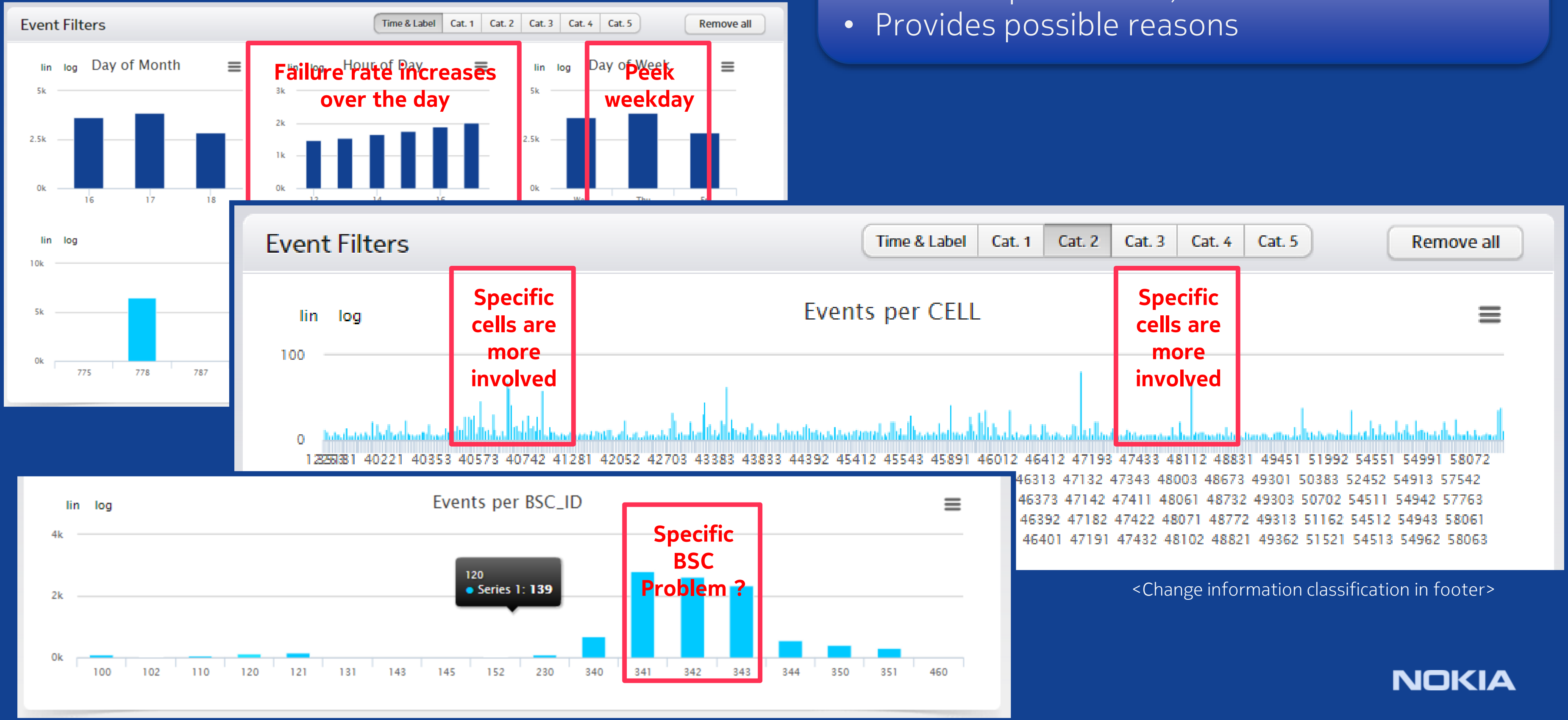
Lines 0 / 3624

3.624 messages

NOKIA

Root Cause Analysis
Let’s solve it graphically

- Must be an interactive drill down
- Reduces xx.xxx messages to a single problem
- Shows dependencies, anomalies
- Provides possible reasons



Root Cause Analysis

Solutions: Let’s use an algorithm for that

Timeslot60	Inter2gTransRealFail	BSC_ID	CELL	DX_CAUSE	LAC	PRB	MM_CAUSE	TAC	HOUR	M15	M05	M
2014-04-16T14:00:00	1	343	45831	2835	34301	306		01371900	14	0	2	14
2014-04-16T15:00:00	0	341	43993	778	34100	306	17	35477004	15	3	11	59
2014-04-16T15:00:00	1	351	56983	2835	35100	306		01264600	15	3	11	59
2014-04-16T15:00:00	0	342	48171	778	34202	306	17	35750805	15	3	11	59
2014-04-16T15:00:00	0	343	47493	3328	34300	306		35461002	15	3	11	59
2014-04-16T15:00:00	0	343	47593	778	34300	306	17	01362900	15	3	11	58
2014-04-16T15:00:00	0	342	40702	778	34201	306	17	35491105	15	3	11	58
2014-04-16T15:00:00	1	342	42221	2842	34200	306		35907605	15	3	11	58
2014-04-16T15:00:00	0	343	42593	3336	34302	306		35635505	15	3	11	58
2014-04-16T15:00:00	0	342	40611	778	34201	306	17	35487901	15	3	11	58
2014-04-16T15:00:00	0	343	45571	778	34300	306	17	35957605	15	3	11	58
2014-04-16T15:00:00	0	342	40051	778	34202	306	17	35907605	15	3	11	58

Lines 0 / 3624

Root Cause Analysis

Show dependencies for error type

778

CLEAR/A ONHOOK DURING SET-UP PHASE

Feature	Probability
M	30.094
HOUR	12.699
M05	5.611
M15	1.537
PRB=306	1
MM_CAUSE=17	0.999
BSC_ID=341	0.283
BSC_ID=342	0.247
BSC_ID=343	0.228
LAC=34100	0.114

Lines: 41

Strong time dependency

Involvement of BSCs

NOKIA

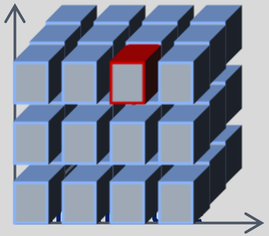
Algorithmic View

Report types	Sub types	Algorithm	Use case
Model deviation (single field)	Difference between model an real value above limit (x%, configurable) Difference between model & real value but also value near zero	Nearest Neighbor (KNN)	Unexpected value ‘sleeping cell’
Trend violation (single field)	Raised if trend slope above/below limit	Linear Regression	Long term trend analysis
Classification error (multiple field)		DBSCAN	Detecting irregular pattern KPI (combinations not seen before)
Root cause analysis (multiple field)	Gaussian Naive Bayes	Gaussian Naive Bayes	Find driving factor for specific dx_causes Show relations to other attributes (cell, customer, …)

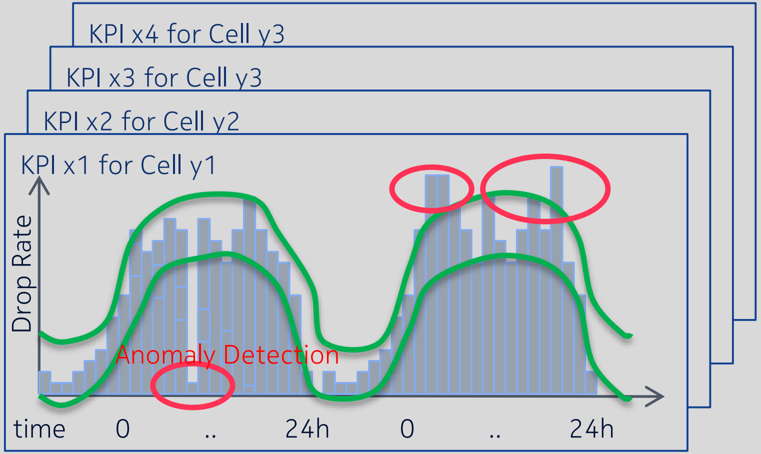
Summary
From Simple KPIs to a guided problem solving workflow …

Trained Model For expected KPI behavior

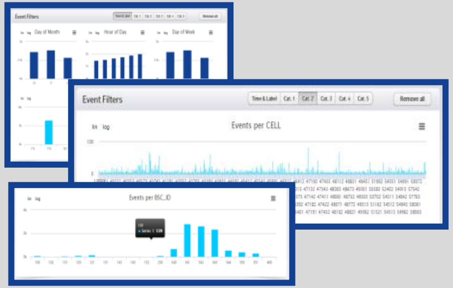
Anomaly Detection

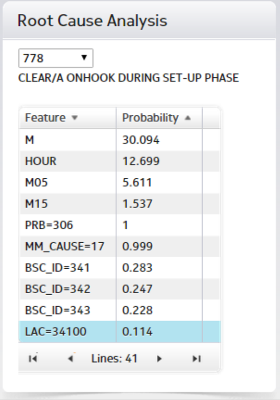


Problem categorization



Problem evaluation





Root Cause Analysis

778

CLEAR/A ONHOOK DURING SET-UP PHASE

Feature	Probability
M	30.094
HOUR	12.699
M05	5.611
M15	1.537
PRB=306	1
MM_CAUSE=17	0.999
BSC_ID=341	0.283
BSC_ID=342	0.247
BSC_ID=343	0.228
LAC=34100	0.114

Lines: 41

Root Cause & correlation

Agenda

- Intro
- Telecommunication Data & Use Cases
- Anomaly & Root Cause Detection
- Time Series Analysis & Prediction
- Software & Methods

NOKIA

Prediction of Time Series (ARMA) – Step 1

Raw Data

Time	KPI
Time t	KPI value
Time t-1	KPI value
Time t-2	KPI value
Time t-3	KPI value
Time t-4	KPI value

Step 1: Computing Moving Averages

Time	KPI	Mov AVG
Time t	KPI value	MA KPI
Time t-1	KPI value	MA KPI
Time t-2	KPI value	MA KPI
Time t-3	KPI value	MA KPI
Time t-4	KPI value	MA KPI

Moving average window

- Auto
- Regression
- Moving
- Average

NOKIA

How to do Time Series Prediction

Prediction of Time Series (ARMA) – Step 2

Step 1: Computing Moving Averages

Time	KPI	Mov AVG
Time t	KPI value	MA KPI
Time t-1	KPI value	MA KPI
Time t-2	KPI value	MA KPI
Time t-3	KPI value	MA KPI
Time t-4	KPI value	MA KPI

Lag window

Step 2: Generating lag window

Time	KPI	KPI t-1	KPI t-2	KPI t-3	Mov AVG	M AVG t-1	M AVG t-2	M AVG t-3
Time t	KPI value	KPI value	KPI value	KPI value	MA KPI	MA KPI	MA KPI	MA KPI
Time t-1	KPI value	KPI value	KPI value	KPI value	MA KPI	MA KPI	MA KPI	MA KPI
Time t-2	KPI value	KPI value	KPI value	KPI value	MA KPI	MA KPI	MA KPI	MA KPI
Time t-3	KPI value	KPI value	KPI value	KPI value	MA KPI	MA KPI	MA KPI	MA KPI
Time t-4	KPI value	KPI value	KPI value	KPI value	MA KPI	MA KPI	MA KPI	MA KPI

NOKIA

Results

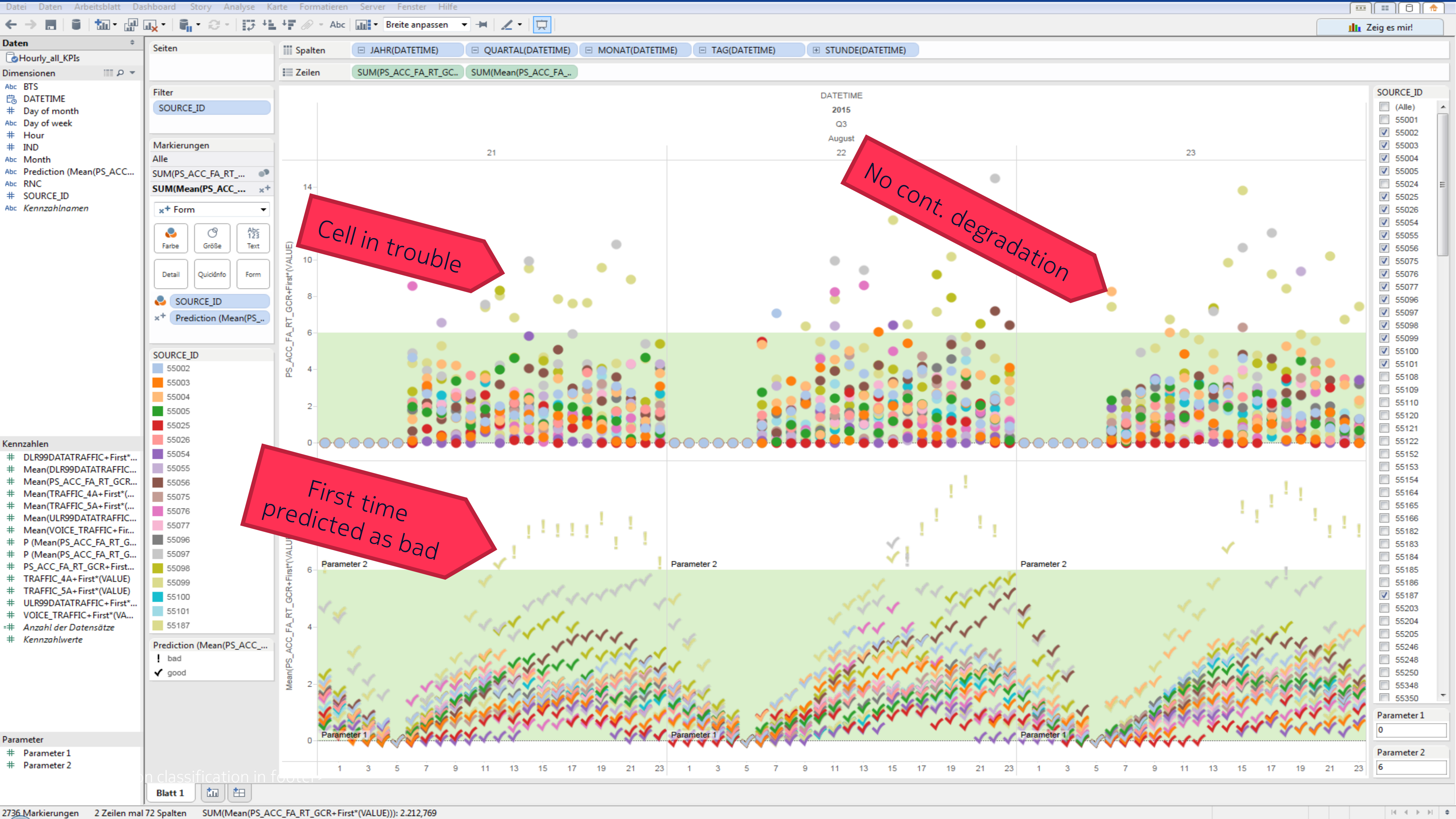
Confusion Matrix

KPI:	PS_ACC_FA_RT_GCR
Pos./Neg. boundary	6%
Number of cells	233 (out of top BTS with highest voice traffic)
MA Window	6 hours
Lag time window	48 hours
Prediction window	1 hour
Ind. test data	12177 Positives / 785 Negatives(=KPI violation)

Method	Accuracy		Predicted as Positive	Predicted as Negative	Precision	Recall
Neural Networks	93,944	Real Positive	True Positives:	False Negatives:	0.939	1
		Real Negative	False Positives:	True Negatives:	?	0
Non Linear Regression	**98,681**	Real Positive	**True Positives: 12099**	**False Negatives: 78**	**0.992**	**0.994**
		Real Negative	**False Positives: 93**	**True Negatives: 692**	**0.899**	**0.882**
Logistic Regression	**98.388**	Real Positive	**True Positives: 12155**	**False Negatives: 62**	**0.988**	**0.995**
		Real Negative	**False Positives: 147**	**True Negatives: 638**	**0.911**	**0.813**
Decision Tree	96.607%	Real Positive	True Positives: 11836	False Negatives: 143	0.976	0.988
		Real Negative	False Positives: 287	True Negatives: 409	0.741	0.588
Ensemble Trees	97.338	Real Positive	True Positives: 12087	False Negatives: 90	0.979	0.993
		Real Negative	False Positives: 255	True Negatives: 530	0.855	0.675

TRUE NEGATIVES:	Correctly predicted KPI violations
FALSE NEGATIVES	Dummy KPI violations incorrectly show as violation
FALSE POSITIVES	KPI violations not found

NOKIA


Cell in trouble
No cont. degradation
First time predicted as bad
DATETIME
2015
Q3
August
21
22
23
Parameter 2
Parameter 1
SOURCE_ID
Prediction (Mean(PS_ACC_...
bad
good
2736 Markierungen
2 Zeilen mal 72 Spalten
SUM(Mean(PS_ACC_FA_RT_GCR+First*(VALUE)): 2.212,769

Agenda

- Intro
- Telecommunication Data & Use Cases
- Anomaly & Root Cause Detection
- Time Series Analysis & Prediction
- Software & Methods

Network Data Intelligence
Research Areas

Applications Use Case
- Customer experience management
- Churn prediction
- Customer segmentation
- Geo Analytics
- Social Analytics

Methods
- Data mining
- Machine Learning
- Statistical Analysis
- Predictive Analysis
- Tools

Storage
- Data warehouses
- NoSQL Databases
- Column stores
- In memory DBs

Processing
- Hadoop engines
- Real time analytics

Representation
- Dashboards
- Data visualization
- Rich clients
- Collaboration platforms

Use Cases

Technology Enabler / Platform component

NOKIA

Generalized Big Data Analytics Stack

What do you need?

Big Data Analytics Stack

- Use Cases
- Visualization, Charting, Drill Down Views
- Analytics Algorithms (K-Means, KNN...)
- Data Storage (Relational, NoSQL)
- Aggregation, Filtering, Distributed Computing
- Import, Formatting, Type Conversion
- Data Sources: net elements, protocols

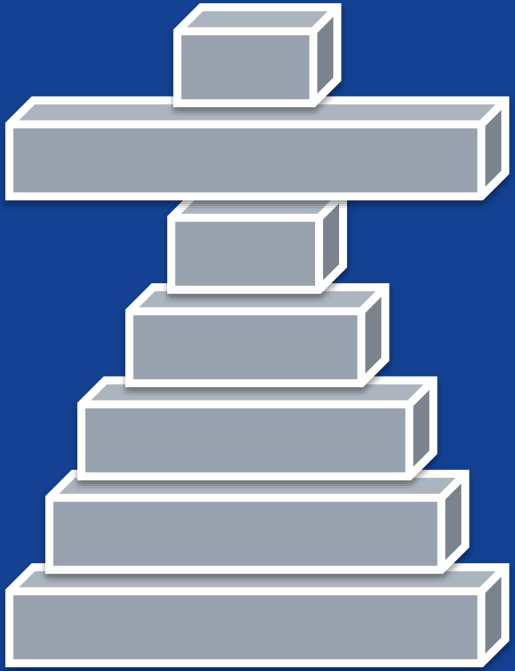
NOKIA

Generalized Big Data Analytics Stack

How does it correspond to the efforts?

Big Data Analytics Stack

- Use Cases
- Visualization, Charting, Drill Down Views
- Analytics Algorithms (K-Means, KNN...)
- Data Storage (Relational, NoSQL)
- Aggregation, Filtering, Distributed Computing
- Import, Formatting, Type Conversion
- Data Sources: net elements, protocols



Development Effort

NOKIA

Generalized Big Data Analytics Stack
Software Choices

Big Data Analytics Stack

- Use Cases
- Visualization, Charting, Drill Down Views
- Analytics Algorithms (K-Means, KNN…)
- Data Storage (Relational, NoSQL)
- Aggregation, Filtering, Distributed Computing
- Import, Formatting, Type Conversion
- Data Sources: net elements, protocols

Requires good analyst background

Needs to be complemented by powerful DB

Problems with Big Data

Not very popular

High entry barrier

Crowded place

No DB replacement

Analytic Tools

RapidMiner
Knime
SPSS

Tableau, QlikView

Mahout

Hive

Parallel processing
Hadoop, Spark, Storm

Pig

Language Stacks

SCI-Kit (Python)

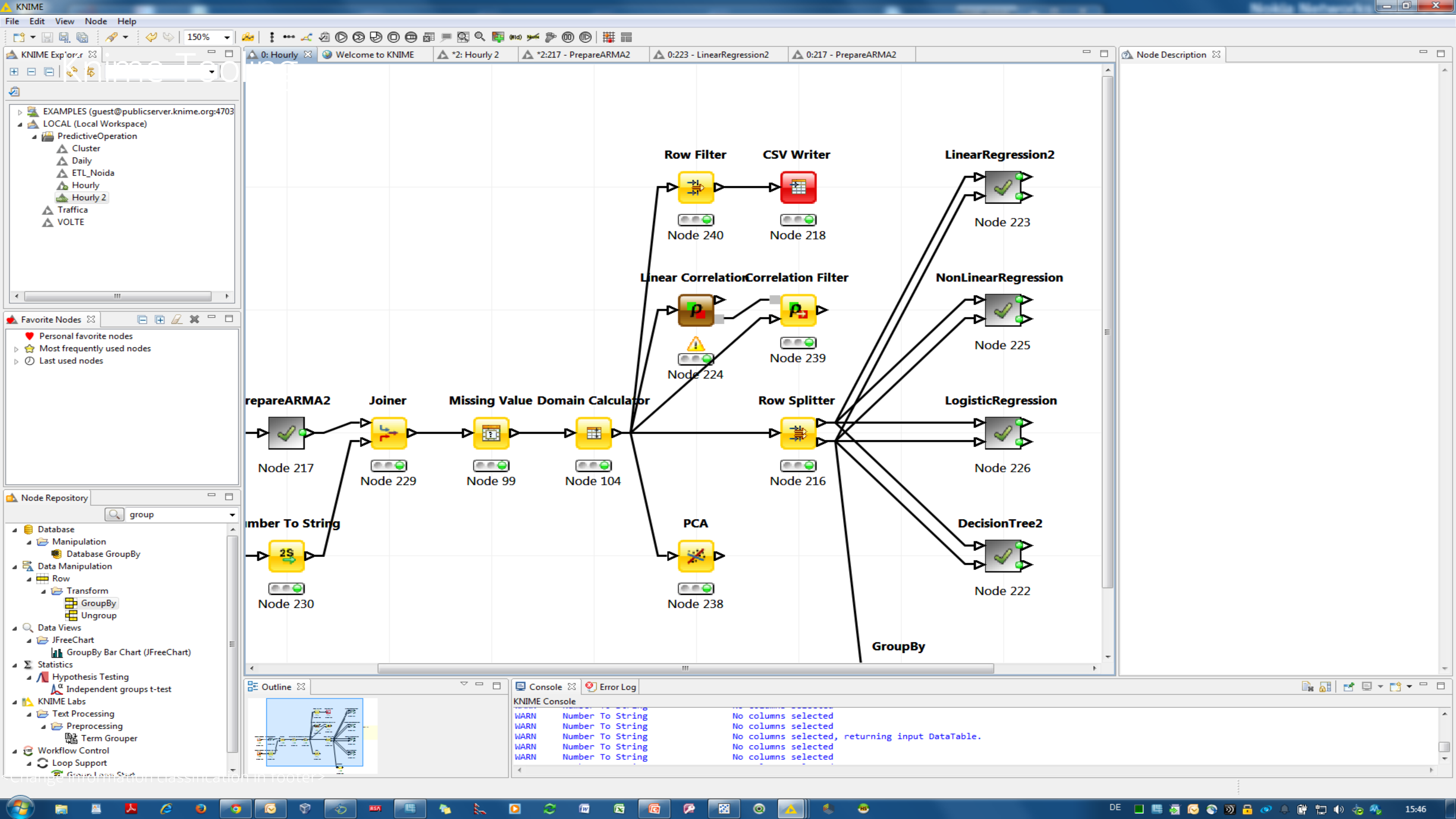
R
Python

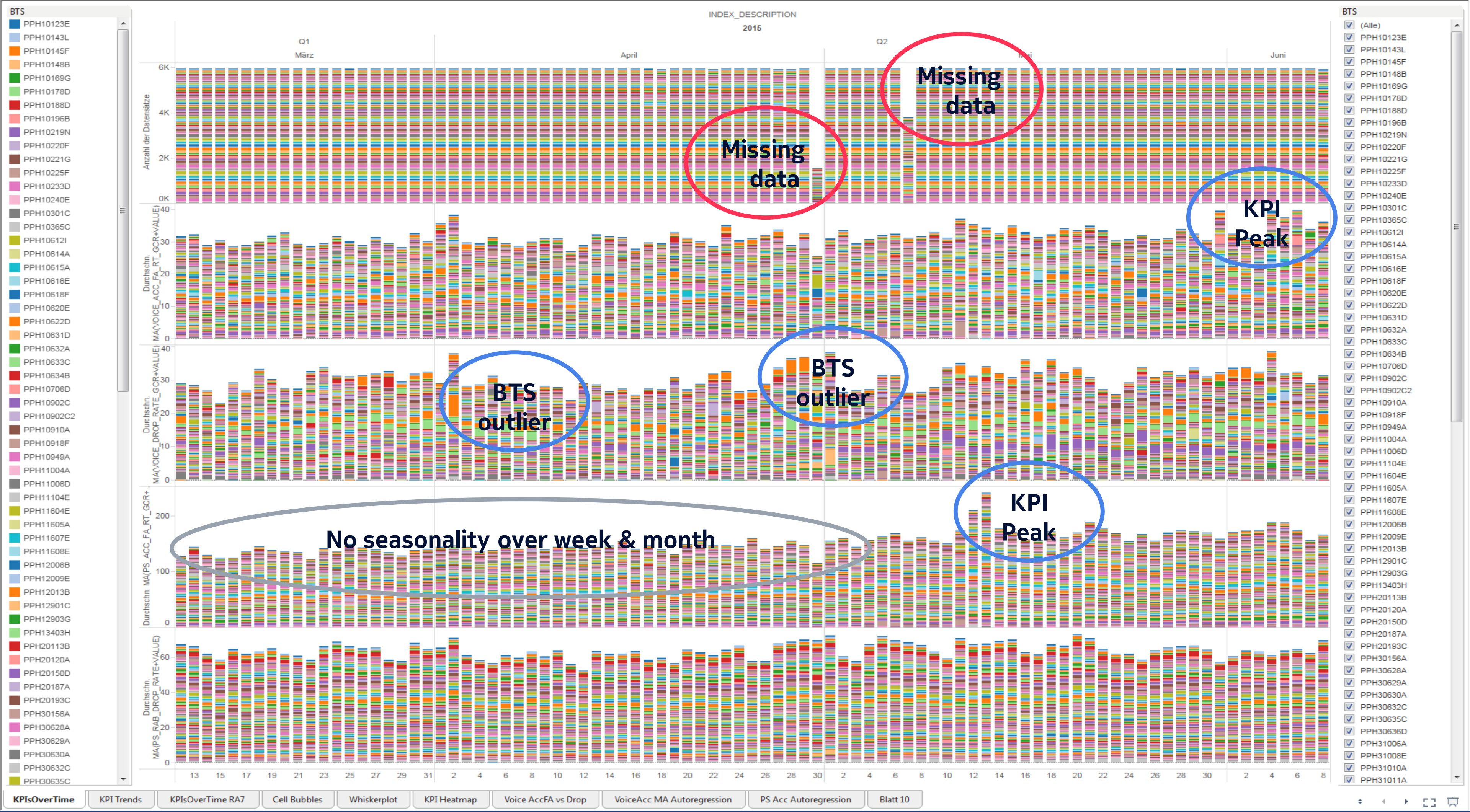
Pandas (Python)

Big Data DBs

Oracle
Teradata
NoSQLs
Vertica

NOKIA


KNIME
File Edit View Node Help
150%
KNIME Explorer
0: Hourly
Welcome to KNIME
*2: Hourly 2
*2:217 - PrepareARMA2
0:223 - LinearRegression2
0:217 - PrepareARMA2
Node Description
EXAMPLES (guest@publicserver.knime.org:4703
LOCAL (Local Workspace)
PredictiveOperation
Cluster
Daily
ETL_Noida
Hourly
Hourly 2
Traffica
VOLTE
Favorite Nodes
Personal favorite nodes
Most frequently used nodes
Last used nodes
Node Repository
group
Database
Manipulation
Database GroupBy
Data Manipulation
Row
Transform
GroupBy
Ungroup
Data Views
JFreeChart
GroupBy Bar Chart (JFreeChart)
Statistics
Hypothesis Testing
Independent groups t-test
KNIME Labs
Text Processing
Preprocessing
Term Grouper
Workflow Control
Loop Support
Row Filter
Node 240
CSV Writer
Node 218
LinearRegression2
Node 223
Linear Correlation
Node 224
Correlation Filter
Node 239
NonLinearRegression
Node 225
repareARMA2
Node 217
Joiner
Node 229
Missing Value
Node 99
Domain Calculator
Node 104
Row Splitter
Node 216
LogisticRegression
Node 226
mber To String
Node 230
PCA
Node 238
DecisionTree2
Node 222
GroupBy
Outline
Console
Error Log
KNIME Console
WARN Number To String No columns selected
WARN Number To String No columns selected
WARN Number To String No columns selected, returning input DataTable.
WARN Number To String No columns selected
WARN Number To String No columns selected
<Change information classification in footer>
DE
15:46



INDEX_DESCRIPTION
2015
Q1
Q2
März
April
Mai
Juni
Anzahl der Datensätze
Durchschn. MA(VOICE_ACC_FA_RT_GCR+VALUE)
Durchschn. MA(VOICE_DROP_RATE_GCR+VALUE)
Durchschn. MA(PS_ACC_FA_RT_GCR+...
Durchschn. MA(PS_RAB_DROP_RATE+VALUE)
Missing data
Missing data
KPI Peak
BTS outlier
BTS outlier
No seasonality over week & month
KPI Peak
BTS
PPH10123E
PPH10143L
PPH10145F
PPH10148B
PPH10169G
PPH10178D
PPH10188D
PPH10196B
PPH10219N
PPH10220F
PPH10221G
PPH10225F
PPH10233D
PPH10240E
PPH10301C
PPH10365C
PPH10612I
PPH10614A
PPH10615A
PPH10616E
PPH10618F
PPH10620E
PPH10622D
PPH10631D
PPH10632A
PPH10633C
PPH10634B
PPH10706D
PPH10902C
PPH10902C2
PPH10910A
PPH10918F
PPH10949A
PPH11004A
PPH11006D
PPH11104E
PPH11604E
PPH11605A
PPH11607E
PPH11608E
PPH12006B
PPH12009E
PPH12013B
PPH12901C
PPH12903G
PPH13403H
PPH20113B
PPH20120A
PPH20150D
PPH20187A
PPH20193C
PPH30156A
PPH30628A
PPH30629A
PPH30630A
PPH30632C
PPH30635C
(Alle)
PPH30636D
PPH31006A
PPH31008E
PPH31010A
PPH31011A
KPIsOverTime
KPI Trends
KPIsOverTime RA7
Cell Bubbles
Whiskerplot
KPI Heatmap
Voice AccFA vs Drop
VoiceAcc MA Autoregression
PS Acc Autoregression
Blatt 10


QlikView x64 Personal Edition - [D:\userdata\zzzzzgqo\Desktop\HourlyNoPivot.qvw*]
Datei Bearbeiten Ansicht Auswahl Layout Einstellungen Lesezeichen Reports Extras Objekt Fenster Hilfe
Ausgangsstatus Zurück Vorwärts Sperren Freigeben
KPIs per BTS
KPIs per Cell
Kopie von KPIs per Cell
RNC
PXRNC001
PXRNC002
PXRNC003
PXRNC004
PXRNC005
PXRNC008
PXRNC009
KPI_DESCRIP...
PS_ACC_FA_RT_GCR
PS_RAB_DROP_RATE
VOICE_ACC_FA_RT_GCR
VOICE_DROP_RATE_GCR
BTS
CELL
Auswahlstatus
KPI_DESCRIPTION
NOT DLR99DATATRAFFIC, TRAFFIC_4A, TRAFFIC_5A, ULR99DATATRAFFIC, VOICE_TRAFFIC
39 von 301
INDEX_DESCRIPTION_string
96 von 2687
Avg (VALUE)
Very weak alarm status
KPI_DESCRIPTION
61613
61006
7786
7778
7771
7750
7653
7652
7650
VOICE_DROP_RATE_GCR
VOICE_ACC_FA_RT_GCR
PS_RAB_DROP_RATE
PS_ACC_FA_RT_GCR
Max (VALUE)
Only a few corresponding alarms with KPIs above limit
Drücken Sie F1 für Hilfe
2015-08-28 15:55:28*
10:57

Generalized Data Analytics Stack

Final Choices

Big Data Analytics Stack

- Use Cases
- Visualization, Charting, Drill Down Views
- Analytics Algorithms (K-Means, KNN…)
- Data Storage (Relational, NoSQL)
- Aggregation, Filtering, Distributed Computing
- Import, Formatting, Type Conversion
- Data Sources: net elements, protocols

Analytic Tools

RapidMiner
Knime
SPSS

Tableau, QlikView

Mahout

Hive

Parallel processing
Hadoop, Spark, Storm

Pig

Language Stacks

SCI-Kit (Python)

R
Python

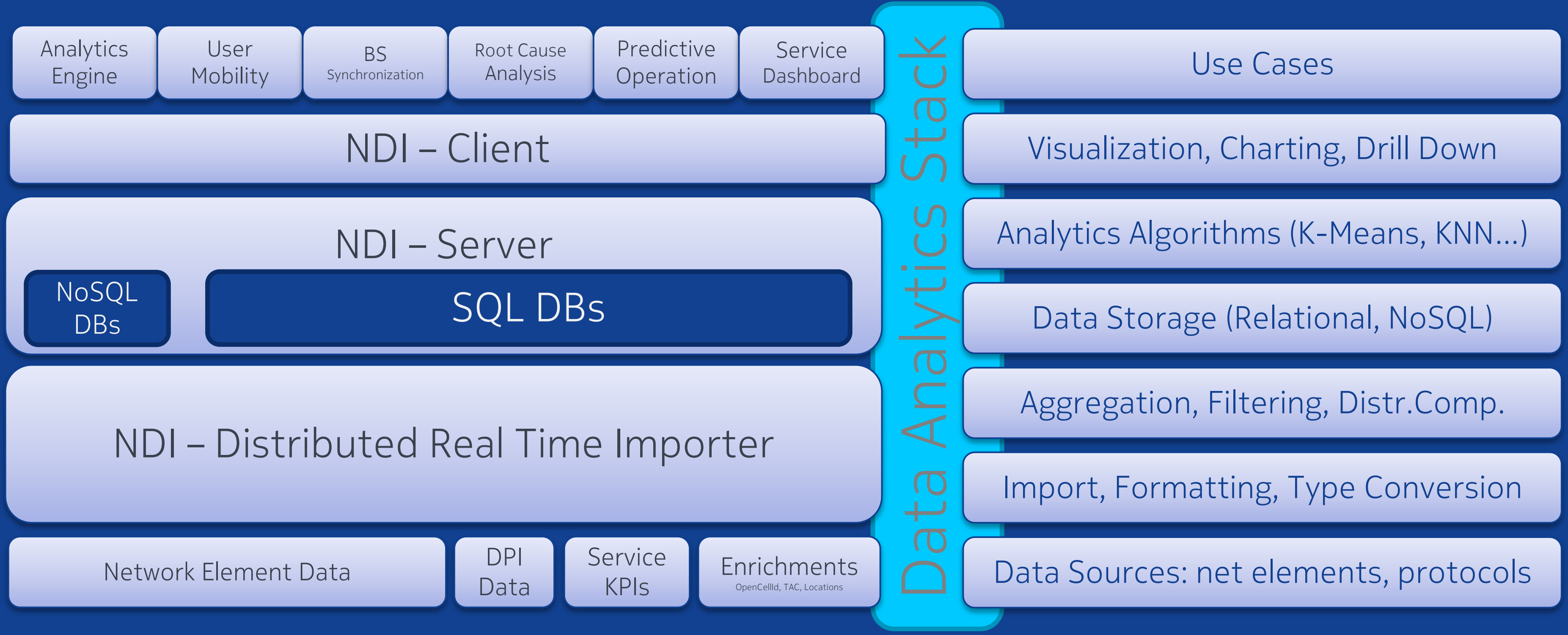
Pandas (Python)

Big Data DBs

Oracle
Teradata
NoSQLs
Vertica

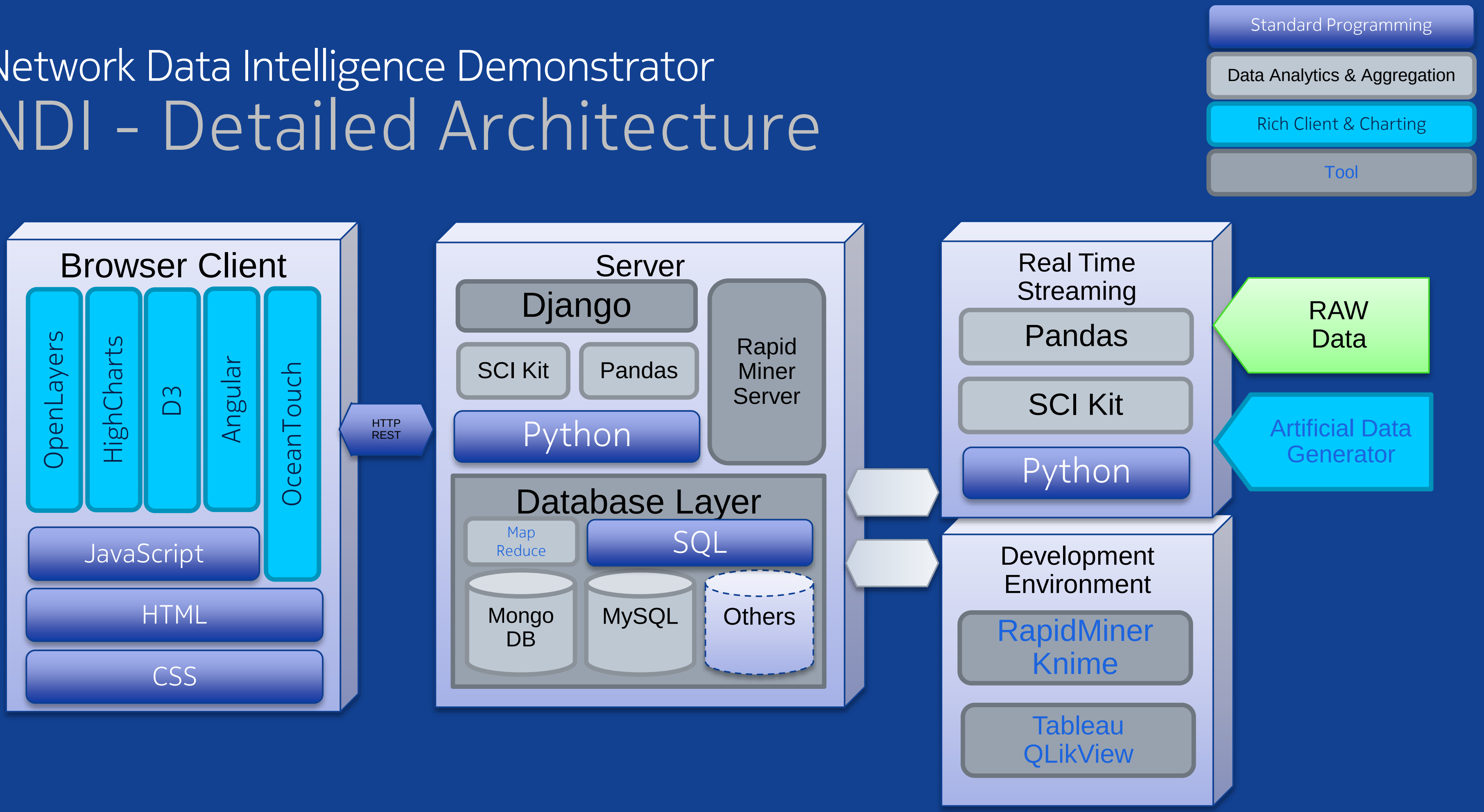
Generalized Data Analytics Stack

NDI Software Architecture



Network Data Intelligence Demonstrator

NDI - Detailed Architecture



NOKIA

Network Data Intelligence
Reasons to use Python

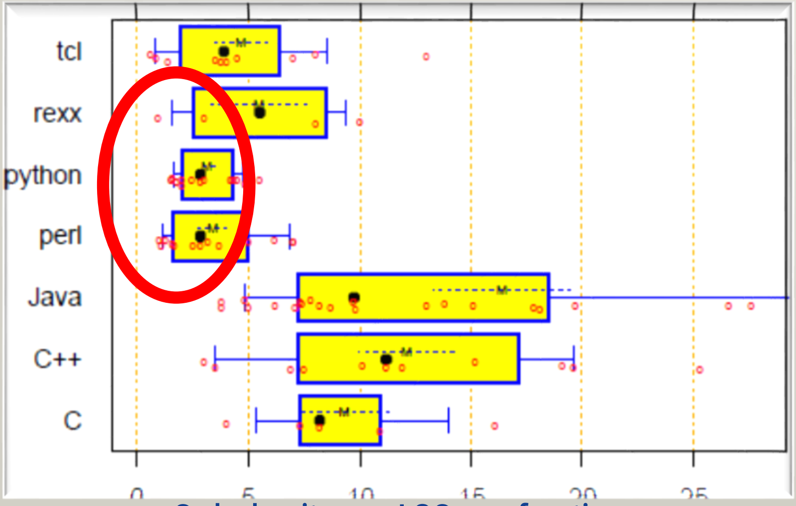
Rapid Prototyping
- Interpreted, very dense coding, eclipse supp.(PyDev)
- Zero turnaround, no SW production

High Functionality
- Data analytics, Statistical computing, Text mining
- Modern 'R'

Next generation programming
- Object oriented, Functional programming
- Closures, Duck typing, memory management, Lambdas

Huge library & community
- Uses most C-libs on Linux - therefore very fast
- Communication stacks, encryption, …

Support for server programming
- Django (HTTP based)
- Twisted (event based)



Code density – no LOCs per function

Source: Lutz Prechelt: An empirical comparison of C, C++, Java, Perl, Python, Rexx, and Tcl for a search/string-processing program

Google: "… we use Python, where we can, C/C++, where we must …"

NOKIA

Experiences / Details
Python - Possible Counter Arguments

Speed

- Think big and optimize your algorithms
- Use full potential of your database

No support for parallel programming (GIL)

- Use secure message passing

Python 2/3 incompatibility

- That's more or less done
- Most packages have Python 3 support today

Unstable – nothing for product development

- That's just wrong
- Our code just worked from the beginning, no GC pain, no bugs, …

Our programmers love it

NOKIA

Important NDI Components

Pandas

- R-extension for Python
- ‘in-memory’ SQL
- Fast – native C-arrays
- Data types:
  - Series, DataFrame, Panel, 4D
- Vector operations
- IO operations (CSV, DBs, …)
- Descriptive statistics
- Group by, sort, indexing
- Merge, join, concatenate
- Reshape, pivoting
- Time series analysis

SCI-Kit

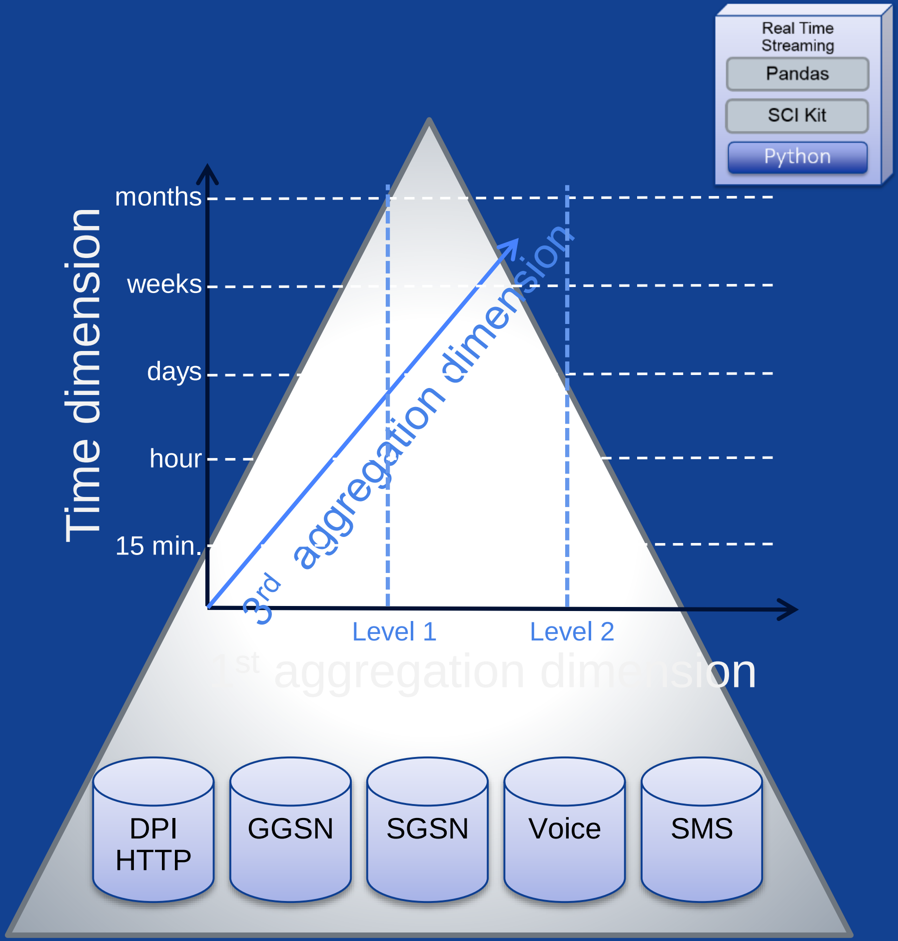
- (Un)-Supervised learning
  - Decision trees, …
- Classification
  - SVM, nearest neighbors, random forest, …
- Clustering
  - k-Means, spectral clustering, mean-shift, …
- Regression
  - SVR, ridge regression, Lasso, …
- Data pre-processing
  - Normalization

NOKIA

Network Data Intelligence

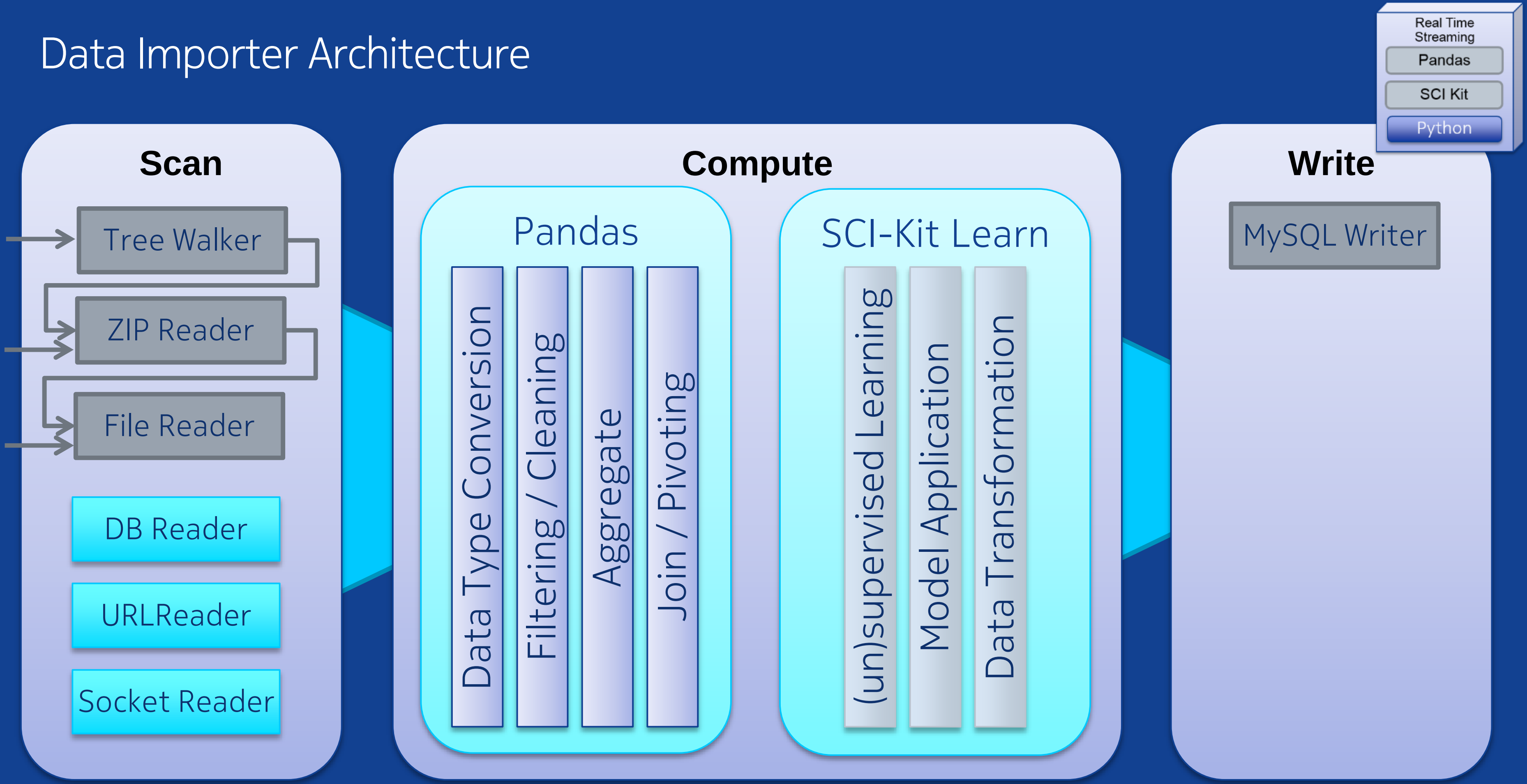
Importer Data Model

- Describes transformation of raw data to a n-dimensional OLAP cube
- Dimensions
  - 1st aggregation dimension
    - CGI, IMSI, URL, IMEI, Service …
  - 2nd aggregation dimension
    - Time
  - 3rd aggregation dimension on demand
    - IMSI / CGI (e.g. for mobility)
- Aggregation levels
  - Time: 15min, hour, day, week, month
  - CGI: MCC, MNC, LAC, Class ID
  - IMEI: TAC, SNR
  - IMSI: MCC, MNC, MSIN

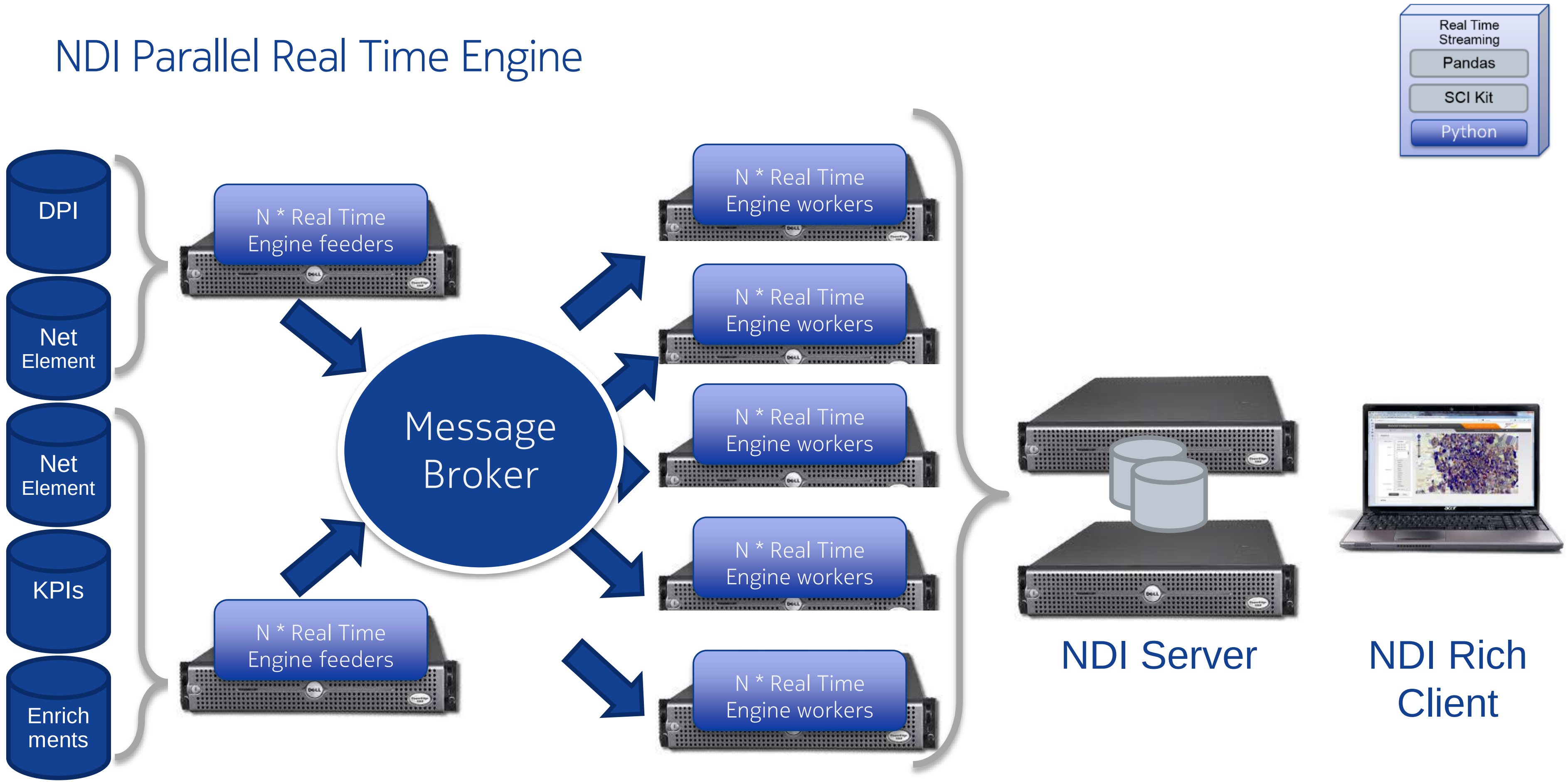


NOKIA

Data Importer Architecture



NDI Parallel Real Time Engine



Speed: Up to 150.000 msg/sec on 4 server system

NOKIA

Important Importer Components

Celery

- Distributed task messaging
- Tight python integration
  - @task
  - Django integration
  - Django management support module
- Based on several message brokers
  - RabbitMQ
  - Redis
  - ...others
- Inter / intra node operation

RabbitMQ

- Message Broker
- Communication patterns
  - 1:1, 1:n, n:1
  - Publish / subscribe
- Multiple queue support
- Complex message routing
- AQMP standard
- Multi language support
- Written in Erlang

Rich Client Development

Basic Thoughts about Browsers …

Virtual machines

- Container for complex applications
- Standardized environment (better than any alternative)

… with multiple DSL support

- HTML5: DOM structure
- CSS3: look & feel
- JavaScript: behavior

… comes with development environment

- Extremely fast and well tested
- Available everywhere: PCs, Tabs, Mobiles
- Powerful choices: Chrome, Firefox (not IE)

And they are for free

Web application types

Client rendering

- JavaScript application loaded via HTTP
- Requires client framework
  - Angular, Ember
- Complex JavaScript application
- Triggers locally handled
- Fast local interaction
- Data access via JSON
- Lots of JavaScript
- Behaves like an application

Server rendering

- Fully rendered web page downloaded
- Requires server framework
  - Java: GWT, JSF, Portlets …
- Pure server programming
- All interactive events across server
- No local interaction
- Data embedded in DOM
- Mostly server programming
- Less interactive

NOKIA

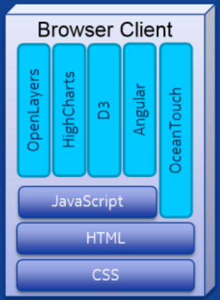
Rich Client Development

Programming with JavaScript

- A reasonable language for application development
- No. 8 ⇑ in Tiobe index
- Standardized (ECMA 5..6)
- Functional / object oriented, Closures
- Dynamic Typing, Interpreted & broadly available
- Huge amount of upcoming frameworks: Charting, Angular …

JavaScript is hot

Rich Client Development
Client Components/Libraries

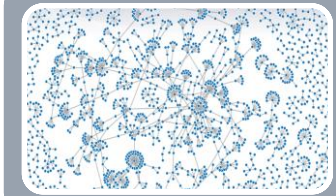


Openlayers

- All kind of maps(OSM, Google …)
- Graphics layers

HighCharts

- Complex charts
- Highly interactive(zoomin/shifting, event handling, …)



D3

- Complex graphics (Force directed graphs, trees ---
- Low level graphics and event handling

Complexity

NOKIA

Rich Client Development

The Need for Angular

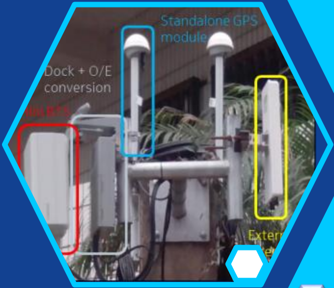
- Angular capabilities
  - MVC architecture
  - 2 way binding
    - JavaScript <-> HTML/DOM
  - Expressions
  - Directives
    - ng-show, ng-repeat
  - Routing
    - Introduces sub commands to URL

- ‘Tomcat / JBOSS’ for rich client development
- Great simplification for development
- Reduces amount of necessary JavaScript code

NOKIA

Big Data Analytics in Telecommunication

Wrap Up



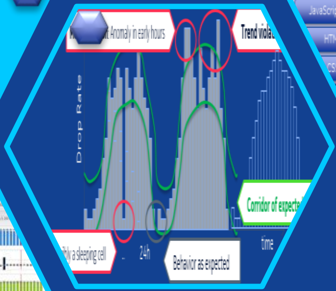
Big Data

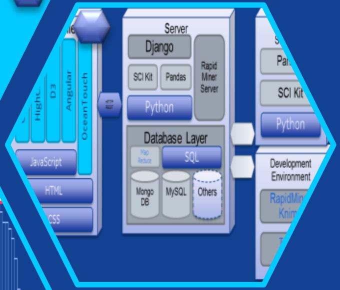
Use Cases

Telecom

Analytics

Software





NOKIA

Thank You!

norbert.kraft@nokia.com