Skyminer Big Data Solution
Features Highlights
Purpose

1. Add value to the data by providing a unified archive and efficient analysis tools

2. Help using monitoring data to improve operations, detect and anticipate failures, and optimize systems
Big Data problem?

- Store all data for any duration?
- Correlation between data sources?
- Further analysis to detect unknown information?
- Learning model to anticipate failures?

How efficiently is data stored and used?
Skyminer Solution

- Systems generate... Large data amount
- Usual Real-Time Processing
- Big Data Storage
- Virtual keep all data forever... Scale to any size!
- Correlation between data sources!
- Automated analysis for new information?
- Learning model to anticipate failures?

AND THEN...

Company Proprietary – Sensitive Information
Primary Goals

1. Time series monitoring & analytics
2. Low cost system
3. Fast
4. Flexible & Scalable
5. Fault tolerant
6. Incorporates useful analysis features
7. Open to other systems
How does it work?

ONE single database with a Time Series Web Service frontend

- A Time Series Database frontend (based on KairosDB)
- A NoSQL Database as storage backend (Apache Cassandra) – We never query from Cassandra directly.
Large data sets: volume variety and throughput

Monics
- Number of metrics: Low
- Throughput: High
- Storage needs: high

Compass
- Number of metrics: High
- Throughput: Low
- Storage needs: Medium

Epoch IPS
- Number of metrics: High
- Throughput: High
- Storage needs: High

Neuralstar
- Number of metrics: High
- Throughput: High
- Storage needs: High

... And other data sources
A Typical Skyminer System(s)

Fault-Tolerant Small Cluster
Using Apache Cassandra DB

Start Small – Scale out at lower cost when needed

- Low cost
- Quick start
- Easy administration
- Fault management
- Scale to any size
- Data replication
- Best performances
Interoperability Features

1. All features are provided as web services (HTTP / REST + JSON)
2. Open APIs
3. Interoperable data format based on JSON
4. Intuitive Web UI for starting using the system
5. APIs include:
   - Data acquisition
   - Data querying
   - Analysis features (prediction, correlations)
Existing acquisition connectors

1. Spectrum and Signal Monitoring - Monics data
2. Equipment M&C - Compass data
3. Satellite Control - EPOCH telemetry
4. Enterprise Network Management - Neuralstar
5. Other connectors exist (or can be built)
Toolbox for leveraging Data value
Skyminer Exclusive Features

1. Analytics API via ad-hoc queries
   • Time Series Query engine
   • Horizontal aggregations (Down Sampling) – statistical features calculated over time
   • Series combining (vertical aggregations)
   • Predictors
   • Correlations analysis

2. User Interface
   • Web UI with Correlations User Interface
   • Integrated Dashboard

Company Proprietary – Sensitive Information
Skyminer Exclusive Features

3. Numerical Analysis
   • Fully integrated with R
     ○ The most popular data analysis environment

4. Reporting
   • Plugin for BIRT (or its commercial alternative OpenText Analytics)
Query Engine & aggregations

- Ad-hoc queries and statistics calculation
- Business Intelligence features already implemented (aggregate, drill & pivot)
  - Data aggregates: Min, Max, Sum, Average, Count, Rate, Std Deviation...etc
  - Multi-level Group-by feature using tags, value, or time
- Filter by tags values
# Aggregations: Available aggregators

- Min
- Max
- Avg
- Sum
- Std Dev
- Scale
- Rate (Derivative)
- Least Square
- Count
- Percentile

- MinMax
- Filter (with predicate)
- First
- Last
- Interpolation
- Gaps marker
- Alias
- Untag/Retag
- Time Shift
- Formula
- Scripted aggregator
- Recorder

KairosDB

---

Skyminer
1. Generic predictive analysis (in the query engine) is being implemented

Below this threshold there is a noticeable impact (failure, degradation)

Predictive analysis indicates when and how-long the system or service will be affected by the degradation

2. Several predictors: linear (exponential Smoothing, holt, least squares), or dynamic with Dynamic Linear Model (DLM)
Correlations Analysis

• Explore and Find correlations within large data sets
  • Correlations may indicate relationship in behaviours

• Several analysis methods
  • Linear
  • DTW (Dynamic Time Warping)
  • Other methods can be added as new modules

• Several correlation API and user interfaces
  • SEARCH – Find data correlated to a reference
  • MATRIX - Explore one-to-one correlations between large number of series
1. Integrated Dashboard

2. Most Skyminer analytic features available: aggregators (downsampling, vertical), group-by, filters, and predictors

3. Build a dashboard from various queries in a few mouse clicks from a rich Web user interface
Using BIRT reporting tool plugin
Using R Numerical Analysis Environment

```r
# Load the library
library("skyminer")

# Create a metric queries
metric1 = SkyminerMetric('metric1', aggregator = aggregator.avg(1, TimeUnit.HOURS, alignSampling)
metric2 = SkyminerMetric('metric2', aggregator = aggregator.avg(1, TimeUnit.HOURS, alignSampling)

# Query & prepare results
query = SkyminerMetricQuery(list(metric1, metric2)
response = executeQuery(query, 'http://192.168')
series = getSeriesByTag(response, 'host')
timestampsAsDate = convertTimestampsToDate(series)

# plot results
plot(timestampsAsDate, series[, 'value'])
```
Data Storage back-end can be changed in one line of configuration file

=> Pick-up the best of the moment for the use case
Key elements
Features

• **System exists and is operational in production since 2014**

• **System is:**
  • Fast
  • Scalable (1 to N nodes)
  • Fault tolerant (1 to N replicas)
  • Easy to backup (e.g. Cassandra snapshots files)
  • Modular and evolutive
  • Open (to change and to other systems)
Conclusion

Skyminer is a simple system, but featuring a rich data processing toolbox for time series

• Unlimited aggregation capabilities

• Analytic features
  • Predictions
  • Correlations
  • Time shifted queries

• Integration with R