

# Projekt MONSOON – návrh platformy pre analýzu veľkých dát v priemysle

Martin Sarnovský<sup>1,2</sup>, Peter Bednár<sup>1,2</sup>

<sup>1</sup>Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická Univerzita v Košiciach, Letná 9, 042 00 Košice

<sup>2</sup>Ekonomická fakulta, Technická Univerzita v Košiciach, Letná 9, 042 00 Košice

martin.sarnovsky@tuke.sk, peter.bednar@tuke.sk

**Abstrakt.** Predkladaný článok predstavuje základnú myšlienku a ciele projektu MONSOON, ktorý je zameraný na oblasť využitia analýzy veľkých dát v priemysle. Hlavným cieľom je predstaviť projekt vrátane jeho hlavných oblastí nasadenia a popísať konceptuálnu architektúru softvérovej platformy používanej v projekte. Charakterizujeme jednotlivé prípady použitia, ktorými sú prediktívne úlohy v dvoch hlavných doménach, produkcii hliníka a výroby plastových výrobkov. Jadrom projektu bude softvérová platforma postavená na technológiách pre spracovanie veľkých dát, ktorá umožňuje ukladanie a spracovanie dát z prevádzok z rôznych odvetví priemyslu. Platforma bude integrovaná so systémami prevádzok a umožní dátovým analytikom vyvíjať, testovať a nasadzovať prediktívne metódy pre optimalizáciu výrobných procesov prevádzok.

**Kľúčová slova:** procesný priemysel, big data technológie, analýza dát.

## 1 Úvod

MONSOON (MOdel based coNtrol framework for Site-wide OptimizatiON of data-intensive processes)<sup>1</sup> je 3-ročný výskumný SPIRE projekt ktorý je zameraný na návrh a implementáciu infraštruktúry pre využitie analýzy veľkých dát v oblasti procesného priemyslu. Hlavnou úlohou projektu je aplikovanie dátovo-orientovanej metodológie pre optimalizáciu produkčných procesov, ktorá je založená na budovaní prediktívnych modelov. Projekt je založený na koncepte zdieľanej medzi-sektorovej dátovo analytickej platformy, ktorá zhromažďuje a spracováva veľké dáta z viacerých prevádzok. Zdieľanie analytickej platformy medzi jednotlivými odvetviami priemyslu potom umožní, okrem optimalizácie procesov a šetrenia výrobných nákladov, aj transfer najlepších praktík a znalostí medzi jednotlivými doménami. V rámci projektu sú zahrnuté dve prostredia pre validáciu a demonštrovanie možností navrhovanej platformy: spoločnosť pre produkciu hliníka vo Francúzsku a spoločnosť pre výrobu plastových dielov v Portugalsku. V nasledujúcich kapitolách bližšie predstavíme

---

<sup>1</sup> <https://www.spire2030.eu/monsoon>

hlavné prípady použitia v oboch doménach a popíšeme konceptuálnu architektúru navrhovanej analytickej platformy.

## **2 Prípady použitia**

### **2.1 Produkcia hliníka**

V súčasnosti je proces výroby hliníka na veľmi vysokej technologickej úrovni, ktorej optimalizácia má vyše 130 ročnú históriu. Aj napriek tomu však existujú oblasti, v ktorých je možné využiť moderné metódy dátovej analýzy na optimalizovanie výrobného procesu, ako napr. prevencia procesných anomálií v elektrolytických peciach, prediktívna detekcia defektov na anódach a zníženie emisií v ovzduší. Práve produkcia anód odstraňovanie defektov v anódach predstavuje v súčasnosti najväčšiu variabilnú zložku produkčných nákladov pri výrobe hliníka. Ciele pri optimalizovaní môžu byť zamerané na zníženie nákladov pri produkcii kvalitných anód resp. pri optimalizácii prevádzky elektrolytických pecí s cieľom čo najlepšieho vyťaženia kvality anód pri zredukovaní energetických nákladov. V rámci projektu boli stanované dva ciele:

- Prediktívna údržba zariadení – hlavným cieľom je odhalenie procesných odchýlok a zlyhaní zariadení pri výrobe anód, ktoré majú vplyv na ich výslednú kvalitu (napr. na hustotu a homogenitu)
- Predikcia kvality anód – hlavným cieľom je detegovanie chybných anód a ich vyradenie z výrobného procesu.

Prediktívna údržba zariadení je jedným z najbežnejších prípadov využitia analýzy veľkých dát pri optimalizácii výrobných procesov. Hlavným cieľom je skrátenie časov pri výpadkoch. Ako vstupné dáta sa používajú senzorické dáta zozbierané z jednotlivých zariadení, ktoré tvoria produkčnú linku pri výrobe anód. Vstupné dáta sú korelované s historickými záznamami o poruchách zariadení. Cieľom je nielen detegovať známe poruchy, ale aj detegovať neočakávané správanie mimo bežnej prevádzky. V prípade prediktívnej kvality anód je cieľom zahrnúť parametre celého výrobného procesu anód spolu a odvodiť z nich indikátor kvality anódy.

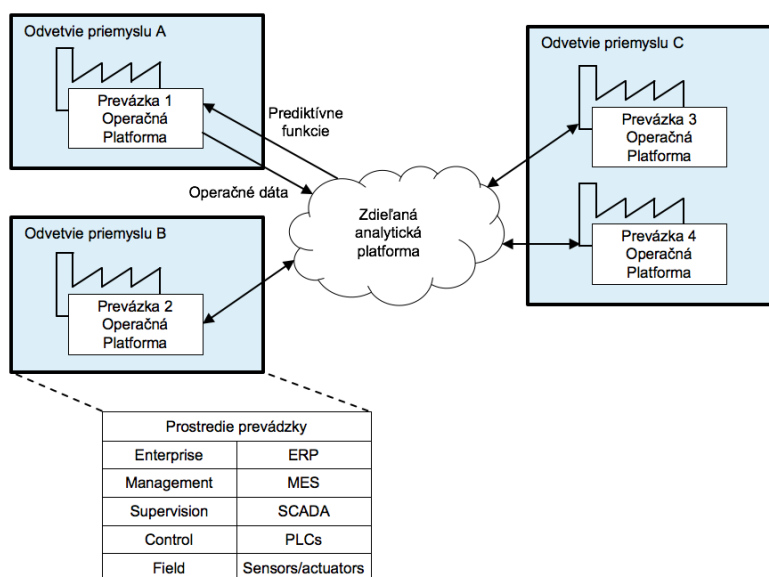
### **2.2 Výroba plastových dielov**

V prípade výroby plastových dielov je projekt zameraný na výrobu dielov technológiou injekčného vstrekovania. Pomocou tejto technológie je možné vyrábať veľké množstvo rôznych typov dielov určených napr. pre medicínsku oblasť, automobilový priemysel, športové potreby, hračky a pod. Produkcia plastových dielov pokrýva masovú výrobu dielov pri ktorých sa nepožadujú vysoké kvalitatívne a estetické normy výrobkov až po kusovú výrobu vysokokvalitných dielov. Charakteristika dielov závisí jednak na použitej materiáli a jednak na parametroch výrobného transformačného procesu pri vstrekovaní. V rámci projektu je výskum zameraný na dva prípady. V prvom ide o výrobu kapsúl pre kávy, ktoré sú vyrábané vo veľkých počtoch

s nízkou variabilitou a relatívne nízkymi požiadavkami na kvalitu. Do výrobného procesu je zaradený systém kontroly kvality, ktorý pokrýva 100% výrobkov. Druhý prípad v doméne výroby plastových výrobkov reprezentuje výrobu technických dielov pre automobilový priemysel, ktoré zahŕňajú inštaláciu kovových dielov vo výrobku. Okrem procesu vstrekovania, výrobný proces zahŕňa aj proces osádzania kovových dielov. V oboch prípadoch je hlavným cieľom predikovanie porúch zariadení a notifikácia riadiacich pracovníkov, čo umožňuje efektívnejšie operatívne odstavenie výroby s cieľom zamedziť produkcii chybných výrobkov. Ďalším cieľom je optimalizácia procesných parametrov počas vstrekovania s cieľom skrátiť výrobný cyklus jedného výrobku pri zachovaní jeho kvality.

### 3 Architektúra riešenia

MONSOON platforma pozostáva z dvoch hlavných častí, ktoré sú zobrazené na Obr. 1. *Operačná platforma* je množina komponentov, ktoré sú nasadené priamo v prostredí konkrétnej prevádzky. Hlavnou úlohou operačnej platformy je spracúvať operačné dáta z prevádzky v reálnom čase. Tento komponent je integrovaný a prepojený s rôznymi, existujúcimi a už nasadenými systémami v produkčnom prostredí ako napr. ERP, manažérske systémy, dátami zo zariadení a senzorov, atď.. Relevantné dáta sú následne z operačnej platformy prenášané do zdieľanej analytickej platformy.



Obr. 1. Architektúra MONSOON platformy

*Zdieľaná, medzi-sektorová dátová analytická platforma* je jadrom MONSOON riešenia. Predstavuje škálovateľné distribuované prostredie, ktoré sa používa pre zber a ukladanie dát z rôznych prevádzok z rôznych odvetví priemyslu. Táto platforma slúži

predovšetkým pre dátových analytikov alebo procesných manažérov. Je postavená na technológiách pre veľké dáta, vrátane distribuovaného úložiska a technológií pre spracovanie veľkých dát. Okrem toho poskytuje viaceré nástroje pre podporu pokročilých analytických procesov. Jej súčasťou je množina vývojárskych nástrojov, ktoré umožňujú navrhovať a implementovať prediktívne metódy pre potreby optimalizácie procesov jednotlivých prevádzok. Vývojárske nástroje sú podporované množinou simulačných nástrojov, ktorých primárnou úlohou je vyhodnotenie implementovaných metód v testovacích prostrediach a ich nasadenie do reálneho prevádzkového prostredia. Pri implementácii riešenia sme použili viacero existujúcich technológií pre spracovanie veľkých dát. Ako jadro sme použili rámec Hadoop v distribúcii Hortonworks HDP, ktorý sme konfigurovali Apache Ambari (používaný aj na manažment). Používame Apache YARN [1] pre manažment zdrojov, (CPU, pamäť, disk, atď.), Hadoop Distributed File System (HDFS) ako distribuované a škálovateľné úložisko [2]. Pre spracovanie dát v reálnom čase používame Apache Spark [3], vývoj prediktívnych modelov bude podporovaný knižnicami strojového učenia ako napr. MLlib [4] alebo Mahout. Pre streamovanie dát používame Apache Kafka a Apache ZooKeeper slúži pre konfiguračný manažment.

## **4 Záver**

Článok mal za úlohu predstaviť projekt MONSOON, jeho hlavné oblasti nasadenia, popísať konkrétne jednotlivé prípady nasadenia a architektúru riešenia. Architektúra riešenia obsahuje z funkčného pohľadu dva hlavné komponenty celej platformy - operačnú platformu zbierajúcu dáta z prevádzky reálnom čase, ktoré sú potom zozbierané, ukladané a analyzované v zdieľanej analytickej platforme. Tá navyše umožňuje jej využitie naprieč rôznymi sektormi, čo môže viesť k transferu znalostí alebo šíreniu skúseností medzi jednotlivými odvetvami. Prvotná verzia analytickej platformy bola implementovaná a nasadená použitím technológií pre spracovanie veľkých dát v doménach produkcie hliníku a výroby plastových komponentov.

## **Literatura**

1. Kumar Vavilapalli, V., Murthy, A.C., Douglas, C., Agarwal, S., Konar, M., Evans, R., Graves, T., Lowe, J., Shah, H., Seth, S., Saha, B., Curino, C., O'Malley, O., Radia, S., Reed, B., Baldeschwieler, E.: Apache Hadoop YARN: Yet Another Resource Negotiator. SOCC '13 Proc. 4th Annu. Symp. Cloud Comput. . 13, 1–3 (2013).
2. White, T.: Hadoop: The definitive guide. (2012).
3. Han, Z., Zhang, Y.: Spark: A Big Data Processing Platform Based on Memory Computing. In: Proceedings - International Symposium on Parallel Architectures, Algorithms and Programming, PAAP. pp. 172–176 (2016).
4. Meng, X., Bradley, J., Yavuz, B., Sparks, E., Venkataraman, S., Liu, D., Freeman, J., Tsai, D., Amde, M., Owen, S., Xin, D., Xin, R., Franklin, M.J., Zadeh, R., Zaharia, M., Talwalkar, A.: [seminal] MLlib: Machine Learning in Apache Spark. J. Mach. Learn. Res. 17, 1–7 (2016).

**PodĎkování:** Tento článok bol podporený projektom H2020-SPIRE-2016 Project MONSOON “MOdel based control framework for Site-wide OptimizatiON of data-intensive processes”

**Annotation:**

*Project MONSOON – design of the big data analysis platform for process industries*

The main objective of the paper is to describe the main objectives of the MONSOON project. MONSOON (MOdel-based coNtrol framework for Site-wide OptimizatiON of data-intensive processes) is a 3-year integrated SPIRE project, that aims to establish data-driven methodology to support identification and exploitation of optimization potentials by applying model-based predictive controls so as to perform plant and site-wide optimization of production process. MONSOON is based on the concept of a cross-sectorial big data platform, a scalable analytical platform that will support collection, storage and processing of data from multiple industry domains. The analytical platform will contain development environment to build these functions and simulation environment to evaluate the models. The platform will be shared among multiple sites from different industry sectors. Cross-sectorial sharing will enable transfer of knowledge across different domains. The project consider two main processs industry domains, the aluminium production factory from France and the plastic molding industry from Portugal.