

## Základní informace

Jméno projektu	Airport Data Hub
Zkratka	ADH
Vedoucí	Mgr. Martin Hlavatý <martin.hlavaty@profinit.eu>
Konzultanti	RNDr. Filip Zavoral, Ph.D. <Filip.Zavoral@mff.cuni.cz>
Anotace	<p>Cílem projektu je navrhnout a částečně implementovat nový způsob integrace provozních dat mezinárodního letiště (náhrada monolitické integrační databáze, tzv. Airport Operational Database). Řešení bude inspirováno architekturou microservices, nicméně realizace bude ve formě CEP (Complex Event Processing) systému s řadou oddělených modulů komunikujících prostřednictvím zpráv. Klíčovým požadavkem je zajištění vysoké dostupnosti, spolehlivosti (zprávy se nesmí ztratit a i duplikace zpráv je problém) a v neposlední řadě škálovatelnosti.</p> <p>Projekt bude pracovat s reálným řešením používaným Letištěm Praha, a.s. (CAODB). Z důvodu velkého rozsahu nebude v rámci projektu implementována kompletní náhrada stávajícího řešení, nýbrž budou nahrazeny jen vybrané moduly.</p>

## Motivace

Současné řešení (CAODB) používané Letištěm Praha, a.s. je postavené na dávkovém zpracování centrálním databázovým serverem. Od vstupu na výstup trvá cesta řádově jednotky minut, což je pro řadu aplikací příliš pomalé. Navíc je řešení obtížně škálovatelné, proto letiště začalo uvažovat o jeho modernizaci. Vzhledem k tomu, že většina operací je asynchronních, je namísto použití klasického Enterprise Service Busu vhodnějším řešením messaging.

## Popis projektu

Systém ADH bude sloužit k odesílání odchozích a zpracování příchozích provozních dat od různých externích i interních subjektů letiště, jako jsou například informace o časech příletů letadel různých společností. Rozsah projektu pokryje implementaci tří vstupních a tří výstupních rozhraní. Mezi systémy (moduly) letiště například patří CDM (collaborative decision making), kontrola letů a externě musí se systémem komunikovat i resource management system (parkovacích míst, tzv. stojánek, či nástupních mostů) a další. Rozsah tohoto projektu pokryje i přepis vstupního modulu a jednoho z funkčních modulů systému. Nově navržený systém musí být v rámci rozsahu napojen na stávající řešení letiště, dokud se nepřepíše i zbytek stávajícího systému (není v rozsahu práce).

Systém musí být schopen přijímat a odesílat zprávy v různých formátech a různými kanály (volání webových služeb, JMS, OAQ ...) a musí poskytovat rozhraní pro doplnění dalších formátů a kanálů.

Systém musí být dekomponovaný do modulů, přičemž jednotlivé moduly budou mít jasně definované rozhraní a budou vzájemně komunikovat prostřednictvím zpráv nebo voláním služeb. Jedním z modulů bude i stávající systém CAODB, díky čemuž bude možné, i přes omezený rozsah projektu, implementovat řešení poskytující všechny funkce současného systému.

Systém bude navržen tak, aby jednotlivé moduly mohly být realizovány v různých technologiích a běžet na samostatném HW. Systém (Kafka cluster) musí být částečně spuštěn na Azure cloudu, kde může být funkcionality samostatná i redundantní.

Systém musí fungovat v **near real-time**, tedy průchod dat systémem (od přijetí zprávy po vygenerování související odchozí zprávy pro externí systém) musí za standardního provozu v průměru trvat znatelně méně než stávající systém, a to v **řádu sekund** místo minut (stávající systém). (bez započtení času průchodu dat skrz modul obsahující dávkově pracující CAODB).

V systému nesmí existovat single point of failure. Nasazování nových verzí systému musí být možné s odstavkou v rámci sekund. Způsob tohoto nasazení bude upřesněn v analýze.

Jedním z nejdůležitějších požadavků je zajištění **kauzality dat** (na úrovni jednoho letu, tj. přistání nebo odletu) – data musí být zpracována v tom pořadí, v jakém byla obdržena. Zároveň musí mít systém schopnost zpracování zpráv paralelizovat pro nesouvisející zprávy.

Navrhnutý systém musí splňovat **vysokou dostupnost** a umožnit **disaster recovery**. Jedním z řešených problémů je i potřeba distribuovaných transakcí, respektive dvoufázového commitu, které uvažovaná technologie implicitně nenabízí. Kromě klasických funkčních testů bude provedena i sada HA a DR testů (seznam scénářů bude navržen v průběhu projektu a bude zahrnovat například přerušení komunikace mezi jednotlivými uzly systému či "zabití" procesů na úrovni OS). Cílem těchto testů bude otestovat, že nedochází ke ztrátě dat a že jedna zpráva není ani za extrémních okolností zpracována vícekrát. Dalším cílem je samozřejmě ověřit robustnost systému a schopnost provozu v režimu 24/7.

V rozsahu projektu tedy bude:

- Implementace alespoň tří vstupních rozhraní pro příjem dat
- Implementace alespoň tří výstupních rozhraní pro odeslání dat
- Napojení nového řešení na současný systém (CAODB)
- Přepis vstupního modulu CAODB do nového řešení. Tento modul mimo jiné zajišťuje identifikaci letu, ke kterému se přijatá data vztahují (externí systémy nepracují s identifikátory letu v CAODB).
- Přepis dalšího vybraného modulu CAODB (například kontroly letu, CDM, meteorologické informace, predikce počtu pasažérů,...) do nového řešení. Tento modul bude identifikován v analytické fázi projektu.

## Platforma, technologie

Očekává se implementace s využitím Apache Kafka a SAP ASE (stávající platforma). Předpokládá se využití event-driven architektury. Konkrétní použité technologie budou ujasněny v rámci analýzy provedené na začátku projektu.

## Odhad náročnosti

Počet řešitelů: 5

Termín dokončení: 9 měsíců od zahájení

Plán prací: 1-2. měsíc – detailní analýza požadavků a upřesnění požívaných technologií  
3.-8. měsíc – implementace  
6. měsíc – kostra aplikace (integrace ADH se stávajícím řešením, schopnost přijímat a odesílat zprávy)  
8. měsíc – dokončen přepis vybraných modulů  
9. měsíc – dokončování, dokumentace, provedení HA a DR testů

## Vymezení projektu

*Projekt je zaměřen na následující oblasti (zaškrtněte vyhovující):*

Diskrétní modely a algoritmy	
<input type="checkbox"/>	diskrétní matematika a algoritmy
<input type="checkbox"/>	geometrie a matematické struktury v informatice
<input type="checkbox"/>	optimalizace
Teoretická informatika	

	Teoretická informatika
Softwarové a datové inženýrství	
x	softwarové inženýrství
x	vývoj software
	webové inženýrství
x	databázové systémy
	analýza a zpracování rozsáhlých dat
Softwarové systémy	
	systémové programování
x	spolehlivé systémy
x	výkonné systémy
Matematická lingvistika	
	počítačová a formální lingvistika
	statistické metody a strojové učení v počítačové lingvistice
Umělá inteligence	
	inteligentní agenti
	strojové učení
	robotika
Počítačová grafika a vývoj počítačových her	
	počítačová grafika
	vývoj počítačových her

## Poznámky

*Vzhledem k tomu, že řešení bude realizováno pro reálné letiště a v rámci projektu se bude pracovat s produkčními daty, nebude možné zveřejnit všechny části výsledného projektu.*