



## POSUDEK OPONENTA BAKALÁŘSKÉ PRÁCE

**Jméno studenta:** Jiří Šimůnek

**Název práce:** Procesy ETL

**Autor posudku:** Hana Skalská

**Cíl práce:** Seznámení s procesy ETL pro přenášení, transformaci a čištění dat z datových zdrojů do datových skladů a datových tržišť. Analýza specializovaných nástrojů ETL, aplikace pro reálná data pomocí Integration Services a MS SQL Server.

Povinná kritéria hodnocení práce	Stupeň hodnocení (známka)					
	A	B	C	D	E	F
Práce svým zaměřením odpovídá studovanému oboru	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vymezení cíle a jeho naplnění	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zpracování teoretických aspektů tématu	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zpracování praktických aspektů tématu	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Adekvátnost použitých metod, způsob jejich použití	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Hloubka a správnost provedené analýzy	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Práce s literaturou	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Logická stavba a členění práce	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jazyková a terminologická úroveň	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Formální úprava a náležitosti práce	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vlastní přínos studenta	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Využitelnost výsledků práce v teorii (v praxi)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

### Vyjádření k výsledku anti-plagiátorské kontroly

K dokumentu nebyly nalezeny podobnosti na „Odevzdej.cz“, shoda s jinými texty je 0%.

### Dílicí připomínky a náměty:

- Teoretická část práce by zasluhovala pečlivější vymezení a definování složek procesu a větší péči (obsahovou, stylistickou i gramatickou) věnovat celkovému zpracování.
- Použití zkratk zpravidla předchází v celé práci jejich vysvětlení, které bývá uvedeno až následně s odstupem v navazující části.
- V popisu Extrakce by bylo vhodné důsledněji rozlišovat datové zdroje od softwarových komponent a nástrojů, nebo případně vysvětlit (definovat) použité pojmy.
- V praktické části by bylo užitečné popsat záměr aplikace. Například věta „Důvodem k migraci...“ mohl předcházet odstavec s popisem původní databáze a popisem požadované cílové databáze, nebo měl být cíl upřesněný na začátku kapitoly 6.

- Charakteristika aplikace je sice zmíněná v úvodní části v metodice, avšak vysvětlení procesu ETL, který se řeší v aplikaci, by mělo být součástí aplikace. Zejména chybí popis původní struktury dat, jejich vlastnosti a očekávaný cíl, aby bylo možné porozumět některým následným krokům.

#### **Celkové posouzení práce a zdůvodnění výsledné známky:**

Práce se zabývá základními teoretickými a praktickými otázkami procesů extrakce, transformace a přenosu dat do datového skladu nebo tržiště (ETL). Stručně charakterizuje jednotlivé fáze procesu, zmiňuje a charakterizuje vybrané softwarové nástroje a platformy, které jsou k dispozici pro volbu realizace jednotlivých etap procesu.

Vlastní aplikace používá SQL Server Integration Services (SSIS). Popisuje datový tok, jeho vizualizaci kontrolním grafem, použité transformační funkce, postupy ošetření duplicit, mapování transformovaných dat a sjednocení datových formátů.

Pozornost věnuje diagnostice chyb, implementaci a využití tabulky ImportLog. Testuje záznam chyb a možnost zpětného dohledání pomocí historie proběhlých procesů.

V závěru se aplikace zabývá optimalizací procesu, popisuje zkušenosti z praktického nasazení a vytvoření balíčku cyklického spuštění procesu podle časového plánu.

Aplikace je poměrně komplexní. Součástí práce je SSIS balíček procesu ETL a návod spuštění této aplikace.

#### **Otázky k obhajobě:**

Charakterizujte hlavní rizika procesu ETL a možnosti jejich kontroly v reálném prostředí.

**Práci doporučuji k obhajobě.**

**Navržená výsledná známka: B**

**V Hradci Králové, dne 7. prosince 2019**

---

**podpis**