

Přílohy

Příloha 1 Kód parsovacího programu

```
<?php
include_once("simple_html_dom.php");

//vstup: array
//zobrazí pole přehledně
function p($array){
    echo "<pre>";
    print_r($array);
    echo "</pre>";
}

//vstup: string
//výstup: string
//vrací text bez přebytečných mezer (trim) i tvrdých
function my_trim($str)
{
    $new_str = trim(str_replace("&nbsp;"," ", $str));
    return $new_str;
}

$mapa_tymu = file_get_contents("./tymy-mapa-upravena.csv");

$mapa_lig = array("NBA","Extraliga", "KHL", "NHL","LaLiga2", "2.
Bundesliga", "Championship", "Ligue 2", "Serie B","Synot liga", "1.
Liga", "Bundesliga", "Champions League", "Eredivisie", "Europa
League", "Jupiler League", "LaLiga", "Ligue 1", "Premier League",
"Primeira Liga", "Primera Division", "Serie A", "Super Lig", "Brazil
A", "Division 2");
$mapa_lig_nahradit = array("Basketbal-NBA","Hokej-Extraliga", "Hokej-
KHL", "Hokej-NHL","Fotbal-Druha spanelska liga", "Fotbal-Druha nemecka
liga", "Fotbal-Druha anglicka liga", "Fotbal-Druha francouszka liga",
"Fotbal-Druha italska liga","Fotbal-Ceska liga", "Fotbal-Ceska liga",
"Fotbal-Nemecka liga", "Fotbal-Liga mistru", "Fotbal-Nizozemska liga",
"Fotbal-Evropska liga", "Fotbal-Belgicka liga", "Fotbal-Spanelska
liga", "Fotbal-Francouszka liga", "Fotbal-Anglicka liga", "Fotbal-
Portugalska liga", "Fotbal-Spanelska liga", "Fotbal-Italska liga",
"Fotbal-Turecka liga", "Fotbal-Brazilska liga", "Fotbal-Druha ceska
liga");

$mapa_mesicu =
array("01"=>"Jan", "02"=>"Feb", "03"=>"Mar", "04"=>"Apr", "05"=>"May", "06"
=>"Jun", "07"=>"Jul", "08"=>"Aug", "09"=>"Sep", "10"=>"Oct", "11"=>"Nov", "1
2"=>"Dec");

//rozdělení řádků
$rad = explode("\r\n",$mapa_tymu);

//inicializace pole
$zapas = array();

//převedení do pole
foreach($rad as $r){
    $text = explode(";", $r);
```

```

    $zapas[$text[0]] = $text[1];
}

//otevření adresáře
$adresar = opendir("./odds/");

//otevření souborů pro zápis
$soubor_vysledek = fopen("./vysledek-html.txt", "a");
$soubor_neexistujici = fopen("./neexistujici.txt", "a");

while($name = readdir($adresar))
{
    $i++;

    if($name!="." && $name!="..")
    {
        //otevření souboru se surovými daty
        $adr = "./odds/".$name;
        $html = file_get_html($adr);

        //vyhledání tabulky
        foreach($html->find('table[id=tournamentTable]') as $bet)
        {
            $datum = "0000-00-00";
            $j = 0;

            //vyhledání řádků tabulky
            foreach($bet->find('tr') as $tr)
            {
                $vystup = "";

                //načtení ligy
                if($tr->find('th[colspan=7]'))
                {
                    $liga = my_trim($tr->find('a', 2)->plaintext);

                    foreach($mapa_lig as $k => $v)
                    {
                        if(strpos($liga, $v) !== FALSE)
                        {
                            $liga = $mapa_lig_nahradit[$k];
                            BREAK;
                        }
                    }
                }

                //načtení data zápasu
                if($tr->find('th'))
                {
                    if($tr->find('span[class=datet]'))
                    {
                        $pd = explode(" ", $tr->find('span[class=datet]', 0)-
>plaintext);
                        $d = $pd[0];
                        $m = array_search($pd[1], $mapa_mesicu);
                        $r = $pd[2];
                        $datum = $r."-".$m."-".$d;
                    }
                }
            }
        }
    }
}

```

```

//načtení času zápasu
if($tr->find('td[class=table-time]'))
{
    $pt = explode(":",$tr->find('td[class=table-time]', 0)-
>plaintext);

    if(!isset($pt[1]))
    {
        echo $tr->find('td[class=table-time]', 0)-
>plaintext."<br />";
        echo $name;
    }

    $cas = $pt[0].":".$pt[1].":00";
    $datetime = $datum." ".$cas;
    $vystup .= "\"".$liga."\"";
    $vystup .= ";";
    $vystup .= "\"".$datetime."\"";
}

//načtení týmů domácích a hostů a jejich správné přemapování
if($tr->find('td[class=name]'))
{
    $z = explode(" - ",$tr->find('td[class=name]', 0)-
>plaintext);

    $cisty_tym1 = my_trim($z[0]);
    $cisty_tym2 = my_trim($z[1]);
    if(strpos($cisty_tym1,"(") != FALSE)
    {
        $cisty_tym1 =
substr($cisty_tym1,0,strlen($cisty_tym1)-6);
    }
    if(strpos($cisty_tym2,"(") != FALSE)
    {
        $cisty_tym2 =
substr($cisty_tym2,0,strlen($cisty_tym2)-6);
    }

    if(isset($zapas[$cisty_tym1])){
        $t1 = $zapas[$cisty_tym1];
    }
    else
    {
        $t1 = $cisty_tym1;
        fwrite($mapa_tymu1,
$liga.";".$z[0].";".$z[1].";".$cisty_tym1.";".$cisty_tym2.";".$datetim
e."\r\n");
    }
    if(isset($zapas[$cisty_tym2])){
        $t2 = $zapas[$cisty_tym2];
    }
    else
    {
        $t2 = $cisty_tym2;
        //jestliže neexistuje tým, uloží se do tabulky
        pro pozdější ruční domapování
    }
}

```

```

        fwrite($soubor_neexistujici,
$liga.";".$z[0].";".$z[1].";".$cisty_tym1.";".$cisty_tym2.";".$datetim
e."\r\n");
    }

    $z = $t1."-".$t2;
    $vystup .= ";";
    $vystup .= "\"".$z."\"";
}

//načtení a vyhodnocení výsledů zápasu
if($str->find('td[class=table-score]'))
{
    $sc = explode("&nbsp;", $str->find('td[class=table-score]',
0)->plaintext);
    $g = explode(":", $sc[0]);

    if(isset($sc[1]) || !isset($g[1]))
    {
        if(!isset($g[1]))
        {
            $vysl = "wtf";
        }
        else
        {
            if($sc[1] != "award." || $sc[1] != "canc." || $sc[1]
!= "postp." || $sc[1] != "abn.")
            {
                $vysl = 0;
            }
            else
            {
                $vysl = NULL;
            }
        }
    }
    else
    {
        if($g[0]*1>$g[1]*1)
        {
            $vysl = 1;
        }
        elseif($g[0]*1<$g[1]*1)
        {
            $vysl = 2;
        }
        elseif($g[0]*1===$g[1]*1)
        {
            $vysl = 0;
        }
    }

    $vystup .= ";";
    $vystup .= "\"".$str->find('td[class=table-score]', 0)-
>plaintext."\"";
    $vystup .= ";";
    $vystup .= $vysl;
    $vystup .= "\r\n";
}

```

```
        //zápis do souboru
        fwrite($soubor_vysledek, $vystup);
    }
}
}
}

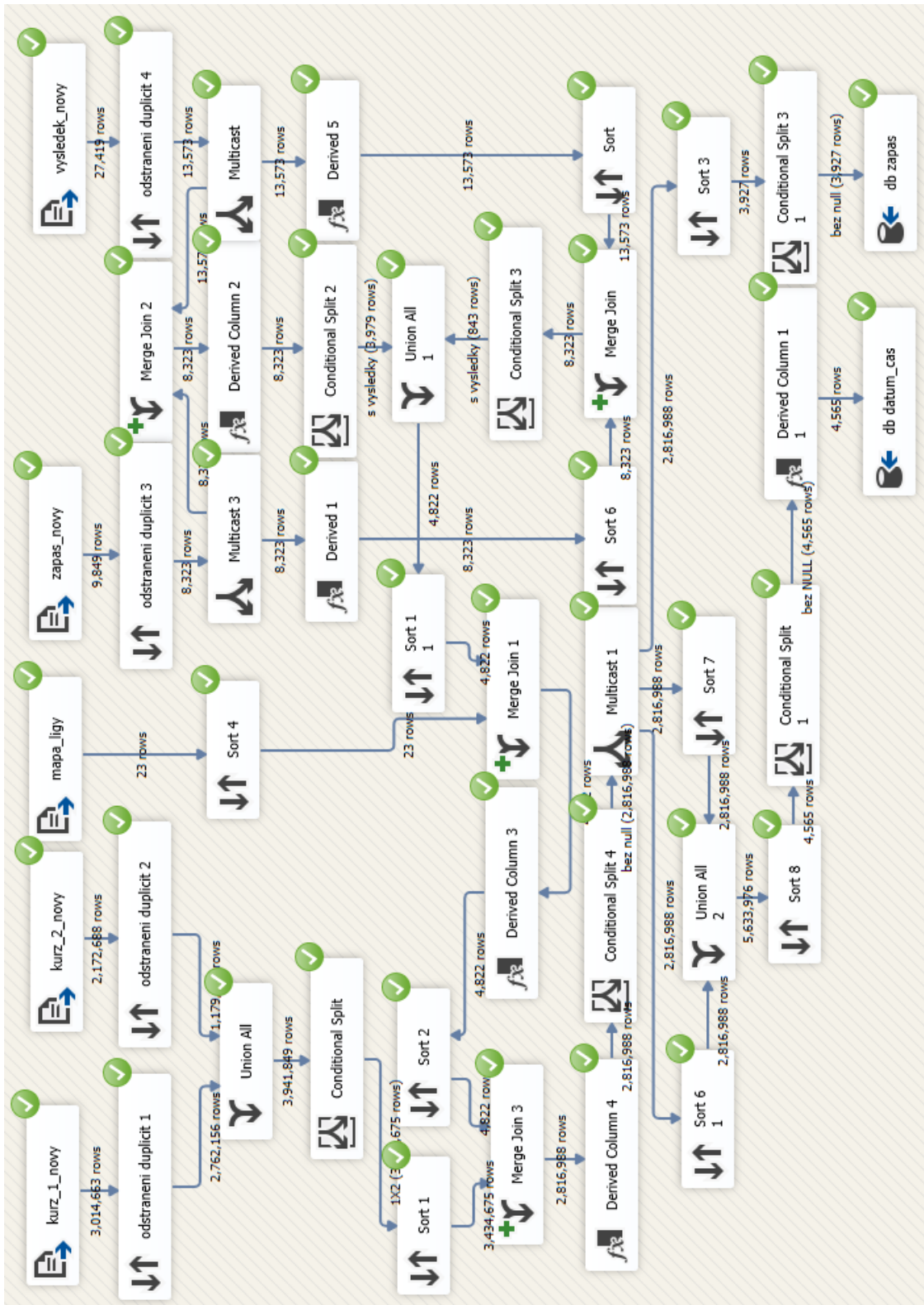
closedir($adresar);

fclose($soubor_vysledek);
fclose($soubor_neexistujici);

$cas = microtime(true)-$start;
echo "ok - ".round($cas,2);

?>
```

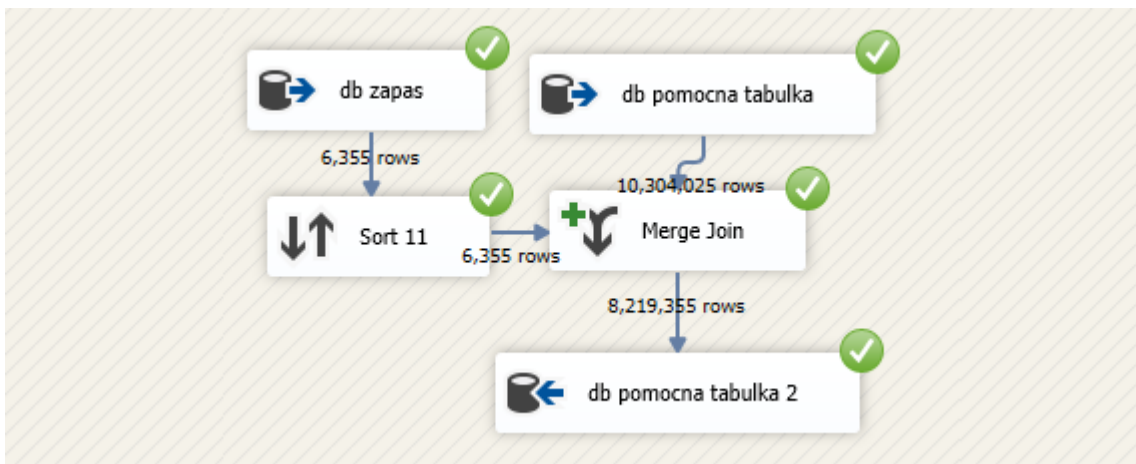
Příloha 2 ETL diagram pro dimenzi ZÁPAS a DATUM_ČAS nový



Příloha 3 ETL diagram pro tabulku faktu KURZ nový část 1



Příloha 4 ETL diagram pro tabulku faktu KURZ nový část 2



Příloha 5 ETL diagram pro tabulku faktu KURZ nový část 3

