

KLIMATOLOGICKÁ DATABÁZE CLIDATA – DATOVÝ MODEL A JEHO APLIKACE

Climatological data base CLIDATA. Data model and its application. Second part of CLIDATA® information about database data model used in CLIDATA®. The paper describes the main data tables in CLIDATA® climatological application and some possibilities how to use them. That application is in routine use since 1st January 2000 with the Czech Hydrometeorological Institute in Ostrava. The actual numbers of records in main data tables (RDATA, MDATA, NDATA, EDATA, MET_PHENOMENA and INTENSITY_RAINFALL) together with the number of stations inside is given in the table. At the last part of the paper holds information about data mining from CLIDATA® application.

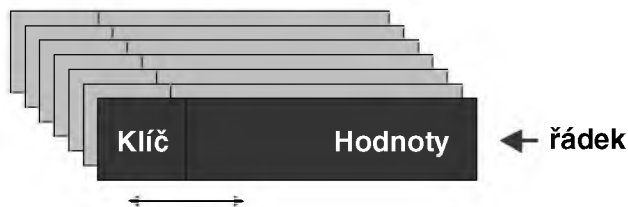
KLÍČOVÁ SLOVA: CLIDATA – databáze klimatologická – Oracle

ÚVOD

V předcházející části volného seriálu informací o aplikaci CLIDATA® [2, 3] jsme se zabývali popisnými údaji o stanicích, měřených prvcích a o typech pozorování a pozorovacích rozvrhů. Protože aplikace prošla od počátku roku 2000 zatěžkavací zkouškou rutinního provozu bez zásadních problémů, lze konstatovat, že navržené schéma je plně životaschopné.

DATOVÉ MODELY – PŘEHLED

Způsob fyzického uložení dat v databázi je nazýván datovým modelem databáze. Pro snadné pochopení datového modelu databáze je třeba vědět, že data jsou v databázi uložena v jednotlivých datových tabulkách, které obsahují předem definované sloupce a většinou předem neurčený počet řádků (obr. 1). Každý řádek v datové tabulce by měl obsahovat minimálně dvě části – identifikaci (klíč) databázového řádku a hodnoty.



Obr. 1 Obecná struktura databázové tabulky.
Fig. 1. General structure of the data base table.

Pro klimatologickou databázi jsou k dispozici tři základní druhy datového modelu – prvkový, časový a plně normalizovaný. Je zřejmé, že jednotlivé existující klimatologické databázové aplikace vykazují od základních typů drobné odchylky, v zásadě však pouze kombinují výhody jednotlivých typů. Důležitou vlastností struktury záznamu je poměr objemu klíče k objemu hodnot. Před zahájením vývoje aplikace CLIDATA® jsme analyzovali výhody i nevýhody jednotlivých možností; náš výsledný model je v jednotlivých datových tabulkách kombinací všech tří základních typů. Pro běžného uživatele není typ datového modelu důležitý, protože má většinou k dispozici aplikační grafické rozhraní. Uživatel vyšší úrovně však přistupuje k datům i mimo aplikaci (např. přes SQL Net, nebo přes ODBC) a znalost datového modelu je pro něj důležitá.

Prvkový datový model

Datová tabulka navržená podle tohoto datového modelu obsahuje na každém řádku jednoznačnou definici stanice a termínu měření nebo pozorování, např. sloupce:

Stanice ID, Datum nebo
Stanice ID, Rok, Měsíc, Den, Hodina, Minuta
a hodnoty k definovaným prvkům, např. sloupce:
Teplota, Úhrn srážek, Oblačnost nebo
Sluneční svit, Úhrn srážek, Vlhkost, Směr větru apod.

S takto definovanou tabulkou se programátorům pracuje snáze, protože výpočty jsou obvykle prováděny nad jedním sloupcem tabulky a pro velkou část všech výpočtů lze použít skupinové funkce¹. Pro uživatele však skrývá tato tabulka velkou nevýhodu. Pokud je nutné rozšířit pozorování o nový prvek nebo sdružit prvky jinak, než bylo předem definováno, je nutný zásah do struktury databáze i do struktury aplikace.

Časový datový model

Datová tabulka navržená podle tohoto datového modelu obsahuje na každém řádku jednoznačnou definici stanice, prvku a neúplně časové určení měření nebo pozorování, popřípadě neobsahuje termín pozorování vůbec (v aplikaci CLICOM je termín pozorování určen definicí prvku). Takto navržený model může obsahovat tento klíč:

Stanice ID, Prvek, Rok, Měsíc, Hodina, Minuta nebo
Stanice ID, Prvek, Rok, Měsíc, Den
a hodnoty vztažené k nedefinovanému časovému údaji na pevných pozicích, např. sloupce:
Hodnota_01, Hodnota_02, Hodnota_03, ... Hodnota_31
pro dny v měsíci pro první typ identifikace nebo
Hodnota_01, Hodnota_02, Hodnota_03, ... Hodnota_24
pro hodiny v průběhu dne pro druhý typ identifikace apod.

U takto definované tabulky je programování komplikovanější, protože výpočty jsou většinou prováděny nad všemi sloupci v záznamu; nelze jednoduše použít zmíněné skupinové funkce. Tato nevýhoda je vyvážena mimořádnou variabilitou

¹ Skupinovou funkci rozumíme jednoduchý způsob výpočtu charakteristik v prostředí SQL nad celým sloupcem v tabulce nebo nad jeho definovanou částí. V časových datových modelech je nutno definovat podobné výpočty nad více sloupci a nelze použít jednoduché a rychlé skupinové funkce.

tou a uživatelskou volností v uložení hodnot různých časových řad měření a pozorování bez nutnosti zásahu programátora.

Plně normalizovaný datový model

Datová tabulka navržená podle tohoto datového modelu obsahuje na každém řádku jednoznačnou definici stanice, prvku a časového určení měření nebo pozorování, např. pole: **Stanice ID, Prvek, Rok, Měsíc, Den, Hodina, Minuta** nebo **Stanice ID, Prvek, Datum, Čas** a jednu hodnotu vztaženou k definované identifikaci.

Takto definovaná tabulka je přehledná a snadno se do ní přistupuje všem uživatelům. Identifikátor na každém řádku (klíč) je však několikanásobně delší než vlastní hodnota a takto navržená databáze má ve svém důsledku větší část diskového prostoru věnovanou indexovým souborům a její celková velikost postupně přesahuje možnosti postupného hardwarového rozšiřování tohoto prostoru.

DATOVÝ MODEL CLIDATA

Optimální kombinace základních typů datových modelů je použita ve všech čtyřech hlavních částech aplikace CLIDATA® (Metadata, Data, Produkty, Systémová správa). Seznam základních tabulek v datovém modelu části Data aplikace CLIDATA® je uveden v tab. 1.

V aplikaci jsou i další datové tabulky, rozšiřující její možnosti (např. MDATA_COUNT pro počty dní vyhovující definovaným limitům, MDATA_COUNT_PHENOMENA pro počty dní s výskytem meteorologického jevu, RDATA_M pro stanice měřící na lodích a bójích, popřípadě pro výsledky expedičních měření aj.) a systém samozřejmě umožňuje definovat různé pohledy (databázová view) na jednotlivé tabulky a vytvářet kombinace jednotlivých datových tabulek.

Všechny datové tabulky jsou svázány s informacemi o metadatech [2], které jsou uloženy ve zvláštních tabulkách a nejsou přímou součástí datového modelu. Pravidlo absolutní konzistence databáze znemožňuje ukládat data dosud nedefinovaných stanic, nedefinovaných prvků, dat mimo určený rozsah pozorování (věcný i časový) nebo mimo definované termíny pozorování.

Tabulka RDATA_N

Záznam v této tabulce je definován klíčem ve tvaru **Stanice - Prvek - Čas pozorování - Rok a Měsíc** a hodnotami pro celý kalendářní měsíc. Do této základní tabulky byla uložena všechna již dříve digitalizovaná klimatologická data za období od roku 1961. Pro vybrané stanice jsou zde k dispozici data za celé období pozorování (např. Praha - Klementinum) a postupně jsou digitalizovány infor-

mace z vybraných kvalitně pozorujících stanic za období před rokem 1961. Od roku 2000 jsou do tabulky RDATA_N průběžně importována data z databáze CLICOM, klimatická datová věta z automatických meteorologických stanic a na pobočce ČHMÚ v Ostravě jsou manuálně pořizována zbývající data přímo v aplikaci CLIDATA® (pobočka ČHMÚ v Ostravě již pro základní práci s daty nepoužívá databázový systém CLICOM). Struktura tabulky RDATA_N odpovídá časovému datovému modelu (jeden záznam obsahuje informace o jednom meteorologickém prvku v jednom pozorovacím termínu za jeden měsíc pro jednu stanici). Každý datový údaj je doprovázen dvěma příznaky, první blíže specifikuje data (např. výskyt ledu na punčošce vlhkého teploměru nebo neměřitelný úhrn srážek, druhý jejich kvalitu (údaj pochybný, odvozený, odhadnutý, vypočtený). Pokud je ve sloupci **Čas pozorování** zapsáno 'AVG', pak jde o průměrné denní hodnoty, 'SUM' takto indikuje denní součty, 'MAX' a 'MIN' denní extrémy. Tato tabulka může obsahovat jak data získaná měřením či pozorováním, tak hodnoty vypočtené (např. relativní psychrometrickou vlhkost nebo evapotranspiraci), vždy se však vztahují k definovanému termínu pozorování nebo dni.

Tabulka RDATA_R

Struktura tabulky je shodná s tabulkou RDATA_N, do této tabulky však ukládáme data pravidelně vznikající v průběhu dne (15minutová data z automatických stanic, hodinová data ze zpráv SYNOP, popřípadě obdobná data z jiných zdrojů). Pro základní práci s tabulkou platí totéž co v předchozím případě.

Tabulka MDATA (MDATA_COUNT, MDATA_COUNT_PHENOMENA)

Tato tabulka je první ze série tabulek, obsahujících odvozená data, průměry nebo součty za definované období (měsíce, dekády, pentády). Její indexová část sestává ze sloupců **Stanice - Prvek - Čas pozorování - Typ - Funkce - Rok - Režim pozorování**, pak následují data včetně roční hodnoty. Tyto hodnoty mají pouze jeden příznak určující jejich kvalitu. Data pro jednotlivá období jsou rozlišitelná pomocí sloupce **Typ** (0 – měsíční hodnoty, 1 – 3 dekádové hodnoty, 4 – 9 pentády), povaha dat sloupcem **Funkce** ('AVG' znamená průměr, 'SUM' součet, 'MAX' nejvyšší hodnotu a 'MIN' nejnižší hodnotu). Ve sloupci **Režim pozorování** je rozlišeno, zda zdrojová data jsou uložena v tabulce RDATA_R nebo RDATA_N. To např. umožňuje vypočíst a uložit průměrné hodnoty, získané ze synoptických i klimatologických pozorování.

Výpočet těchto hodnot probíhá v databázi plně automaticky a není tudíž třeba jej speciálně ošetřit. V případě potřeby by

Tab. 1 Základní datové tabulky aplikace CLIDATA.

Table 1. Basic data tables of the CLIDATA application.

Tabulka	Popis	Model
RDATA_N	Základní tabulka obsahující hodnoty měřené nebo pozorované nepravidelně v průběhu dne (např. v klimatologických termínech).	časový
RDATA_R	Základní tabulka obsahující hodnoty měřené nebo pozorované pravidelně v průběhu dne (např. v synoptických termínech nebo každých 15 minut).	časový
MDATA	Tabulka odvozených dat vztažených ke kalendářním pentádám, dekádam a měsícům.	časový
EDATA	Tabulka měsíčních a ročních dlouhodobých průměrů a extrémů vztažených k zadanému období. Pro standardní uložení je minimální délka období 10 let.	prvkový
NDATA	Tabulka normálových měsíčních a ročních hodnot za doporučené normálové období.	prvkový
MET_PHENOMENA	Tabulka meteorologických jevů.	normalizovaný
INTENSITY_RAINFALL	Tabulka minutových úhrnů srážek.	normalizovaný

bylo možné vždy odvozená data vypočítat. Počet přístupů především k měsíčním datům je v klimatologické praxi však tak častý, že využití popsané praxe by bylo z hlediska zatížení systému neúměrně velké. Na druhou stranu je však třeba ošetřit ty případy, kdy došlo ke změně vstupních dat (např. během jejich validace). O to pečují speciální programové prostředky, které se nazývají trigger².

Ne všechna data v této tabulce jsou však vypočtena. Platí zde výjimka, která byla vyvolána možností zpracovat údaje z ročenek či jiných publikací, obsahujících pouze měsíční hodnoty. V tomto případě mají uložené hodnoty speciální příznak, poukazující na povahu jejich získání.

Tabulka MDATA_COUNT má podobnou, avšak jednodušší strukturu. Pokud jde o klíč, je tvořen sloupci **Stanice - Prvek - Čas pozorování - Typ - Rok - Režim pozorování - Způsob porovnání - Porovnávaná hodnota**. U způsobu porovnání existují pouze dvě možnosti, větší (G) a menší (L), porovnávaná hodnota je libovolné číslo. Měsíční údaje jsou doplněny pouze jedním příznakem.

Tabulka MDATA_PHENOMENA má kratší klíč, složený pouze ze sloupců **Stanice - Rok - Jev**. Výpočet probíhá z tabulky MET_PHENOMENA (viz níže), jevy lze při výpočtu různě sdružovat (např. den se srážkou je definován výskytem alespoň jednoho z jevů dešť, dešťová přeháňka, sněžení atd.).

Tabulka EDATA

Tato tabulka je nezbytná pro validaci údajů, i když její použití v klimatologické praxi je bezesporu širší. Její indexová část sestává ze sloupců **Stanice - Prvek - Měsíc - Čas pozorování - Režim pozorování - Uživatel**. Z uvedeného vyplývá, že je možné stejně jako v předchozím případě rozlišovat mezi pravidelnými a nepravidelnými režimy pozorování. Sloupec **Uživatel** umožní oprávněnému uživateli, aby v případě potřeby zadal databázi výpočet pro jiné než definované období. Tabulka obsahuje především nejvyšší a nejnižší hodnotu (včetně data výskytu), nejvyšší zjištěné rozdíly oproti předchozímu termínu pozorování a oproti stejnému pozorovacímu termínu předchozího dne a konečně empirické pravděpodobnosti 1, 2, 5, 10, 20 ... 80, 90, 95, 98 a 99 %; poslední informace je délka zpracovaného období.

Tabulka NDATA

Jako podklad pro výpočet této tabulky slouží jedna z tabulek RDATA a tabulka MDATA. Princip výpočtu je definován v [1], oproti doporučené verzi je poněkud rozšířen. Klíč tabul-

² Trigger je zvláštní programová jednotka, která je automaticky spuštěna při definovaném zásahu do databáze (např. RDATA_R_TRIGGER_ON_UPDATE je spuštěn při změně hodnoty v tabulce RDATA_R). Dobře navržená sada triggerů spolehlivě zajišťuje konzistenci databáze.

Obr. 2 Formulář tabulky MDATA.

Fig. 2. MDATA table form.

Obr. 3 Formulář tabulky MDATA_COUNT.

Fig. 3. MDATA_COUNT table form.

ky je tvořen sloupci **Stanice - Prvek - Měsíc - Čas pozorování - Režim pozorování - Období**. Jejich smysl je stejný jako u předchozí tabulky s výjimkou posledního sloupce, který umožňuje výpočet standardních normálů pro různá období (např. 1961–90, které je defaultní, ale i 1931–60, 1901–50 atp.). Kromě normálové hodnoty jsou vypočteny empirické pravděpodobnosti z denních hodnot stejně jako v tabulce EDATA, a z měsíčních hodnot pro pravděpodobnosti výskytu 10, 20 ... 80, 90 %.

Pokud ovšem jsou počítány normály počtu dní s ..., které vycházejí z tabulky MDATA_COUNT, resp. MDATA_PHE-

NOMENA, pak empirické pravděpodobnosti pro denní hodnoty pochopitelně neexistují.

Tabulka MET_PHENOMENA

Pro tuto tabulku byly vyvinuty speciální symboly jako písmo TrueType, usnadňující především pořizování dat (např. ●, ☼, ~, ☽). Protože jde o upravenou normalizovanou tabulku, obsahuje klíč sloupce **Stanice - Jev - Rok - Měsíc - Den - Čas začátku jevu - Čas konce jevu**. Jev je blíže specifikován svou intenzitou a proměnlivostí; pokud je jeho začátek či konec stanoven pouze odhadem (zkratkou), pak je i tato informace k dispozici.

Tabulka INTENSITY_RAINFALL

Tato tabulka má jen tři sloupce, **Stanice - DatumČas - Úhrn**, přičemž první dva tvoří klíč a ve třetím je minutový úhrn srážek. Tabulka je aktuálně naplňována importem z automatických stanic, historická data jsou rovněž postupně importována. V kombinaci s tabulkou MET_PHENOMENA je uživatel schopen analyzovat jednotlivé srážkové situace a pracovat i s dlouhodobými intenzitami srážek. V tabulce nejsou uloženy nulové úhrny k jednotlivým minutám v období srážkové situace, které vznikají při malé srážkové intenzitě.

DATA APLIKACE CLIDATA

Aplikace CLIDATA[®] je od ledna roku 2000 ve zcela rutinním provozu na pobočce ČHMÚ v Ostravě, jsou zde však uložena všechna klimatologická a vybraná meteorologická data z celé České republiky. Přímý přístup do datových tabulek je možný ze všech pracovišť ČHMÚ napojených na ústavní WAN se standardním zabezpečením proti neoprávněnému přístupu – znalost jména a hesla je nutná pro připojení a jméno klienta počítače, popřípadě jeho IP adresa je evidována v systémové části databáze.

Datový obsah databáze

Jednotlivé datové tabulky jsou naplňovány importem dat z jiných prostředí (např. ze zpravodajství GTS, z automatických meteorologických stanic, z předcházejících databází apod.), pořízením dat prostřednictvím pořizovacích formulářů CLIDATA[®] (formuláře lze snadno přizpůsobit papírovým dokumentům) a v neposlední řadě výpočtem (v aplikaci jsou ukládána i vypočtená a odvozená data, která jsou používána několikanásobně častěji než data originální). V tabulce 2 jsou orientační počty záznamů a stanic v jednotlivých tabulkách (stav k 20. prosinci 2000).

Záznamy v jednotlivých tabulkách jsou spolu logicky svázány. Změna v základních tabulkách RDATA_N a RDATA_R (oprava nebo vymazání hodnoty) je možná pouze pro nezkontrolovaná data. Tato změna a vložení nového záznamu vyvolá automaticky následující akce:

- porovnání se seznamem výpočtových rovnic vytvoří požadavek na přepočtení příslušných záznamů (např. změna tla-

Obr. 4 Formulář tabulky EDATA

Fig. 4. EDATA table form.

Obr. 5 Formulář tabulky NDATA.

Fig. 5. NDATA table form.

Tab. 2 Základní statistika záznamů v aplikaci CLIDATA.

Table 2. Basic statistics of records in CLIDATA application.

Tabulka	Počet záznamů	Počet stanic
RDATA_A	4 538 983	954
RDATA_R	1 573 733	200
MDATA	1 103 977	953
EDATA	96 124	737
NDATA	62 854	696
MET_PHENOMENA	5 336 258	907
INTENSITY_RAINFALL	254 193	60

- ku vzduchu v jednom termínu na tzv. referenční stanici znamená přepočít průměrného denního tlaku na této stanici a všech tlaků vodní páry a relativní psychrometrické vlhkosti v daném termínu na všech stanicích používajících pro psychrometrický výpočet hodnoty z této referenční stanice),
- porovnání s definicí příslušného prvku (teplota, vlhkost, ...) vytvoří požadavek na přepočít příslušných záznamů v tabulce MDATA,
 - porovnání s tabulkou systémových parametrů (definice normálových a extrémních období) a s definicí prvků vytvoří požadavek na přepočít příslušných záznamů v tabulkách EDATA a NDATA,
 - změna již dříve vložené hodnoty vyvolá odložení původní hodnoty do tabulky HISTORIC_VALUE_R pro pravidelná a do tabulky HISTORIC_VALUE_N pro klimatická data, včetně jména uživatele, který změnu provedl a času provedené změny,
 - porovnání se seznamem měřených prvků na stanici vytvoří požadavek na případný zápis do tabulky INVENTORY (speciální tabulka obsahující statistiku chybějících a automatickým limitům nevyhovujících hodnot).

Všechny výše uvedené akce ukládají záznamy do pomocných systémových tabulek, které jsou postupně zpracovávány databázovými automaticky spouštěnými úlohami. Uživatel nemá možnost do tohoto postupu nijak zasahovat, databázový administrátor pouze sleduje průběh všech operací v připravených formulářích a LOG souborech.

Vypočtené termínové a denní hodnoty

Výpočet hodnot uložených v tabulkách RDATA probíhá plně automaticky podle připravených rovnic. Uživatel s příslušným oprávněním může definovat nové rovnice a přidělovat

je příslušným stanicím, již přidělené rovnice nelze měnit bez ztráty dříve vypočtených hodnot. Definice nových rovnic nevyžaduje žádné programové zásahy do aplikace. Například pro výpočet průměrné denní teploty je v systému uložena rovnice č. 5, která je přidělena všem měřeným termínovým teplotám. Pro výpočet denního úhrnu srážek z hodinových úhrnů srážek (od 00:00 do 23:00 hodin SEČ) je v systému uložena rovnice č. 9, která je přidělena příslušným prvkům SRA24, pro výpočet hodinových úhrnů srážek z 15minutových úhrnů na automatických stanicích je v systému rovnice č. 12, která je přidělena prvku SRA1H na příslušných stanicích. Komplexnost výpočtových možností ukazuje rovnice č. 13 pro výpočet tlaku vodní páry ve všech termínech měření (příklady v tabulce 3).

Identifikace pro každou rovnici definuje hodnotu uloženou po provedeném výpočtu v části klíče **Čas pozorování** (XX:XX je kód pro skutečný čas). Pokud tedy uživatel požádá databázi o sdělení časů pozorování například prvku teplota (T) v tabulce RDATA_N, dostane časy 07:00, 14:00, 21:00 a AVG nebo pro prvek denní úhrn srážek (SRA) dostane časy 07:00 a SUM.

Zálohování datových tabulek v databázi

Všechna základní data jsou včetně svých popisných metadat pravidelně zálohována. Vzniklá datová záloha je uložena nezávisle na databázovém serveru. Interval pravidelné zálohy je v současné době nedostatečný (jednou za měsíc), všechna data jsou tedy stále k dispozici i v jiných aplikacích. Od roku 2001 bude celý systém převeden ze systému Client-Server do replikačního systému Server-Server a celá databázová struktura bude včetně obsahu na dvou navzájem propojených serverech v Praze a v Ostravě. Navržené replikační procesy budou udržovat jednotnost datového obsahu na obou serverech

Tab. 3 Příklady symbolického zápisu výpočtových rovnic.

Table 3. Examples of a symbolic notation of computational equations.

Číslo rovnice	Identifikace	Rovnice – symbolický zápis
5	AVG	{(T.07:00)+{T.14:00}+2*{T.21:00}}/4
9	SUM	{SRA1H.SUM}
12	XX:XX	{SRA15M.SUM-3}
13	XX:XX	round(decode({TV.XX:XX.FLAG1},L', power(10, -9.09685*(273.16/({TV.XX:XX}+273.16)-1) -3.56654*log(10,273.16/({TV.XX:XX}+273.16)) +0.87682*(1-({TV.XX:XX}+273.16)/273.16) +0.78614), power(10, 10.79574*(1-273.16/({TV.XX:XX}+273.16)) -5.028*log(10,({TV.XX:XX}+273.16)/273.16) +1.50475e-4*(1-power(10,-8.2969*({TV.XX:XX}+273.16)/273.16-1))) +0.42873e-3*(power(10,4.76955*(1-273.16/({TV.XX:XX}+273.16))-1) +0.78614))- 1e-6*decode({F.XX:XX},0,1110,1,985,2,834,3,783,4,750,5,730,6,715,7,700,8,691,9,684,10,679,11, 675,12,672,13,669,14,667,665) *({T.XX:XX}- {TV.XX:XX}) *power(10, log(10,{P.XX:XX.REFER}) +({SPEC.ELEV}+nvl({SPEC.P.HEIG},0)- {SPEC.ELEV.REFER}- {SPEC.P.HEIG.REFER}) /(18400+67.53*({T.XX:XX}- {T.XX:XX.REFER})/2+0.003*({SPEC.ELEV}+nvl({SPEC.P.HEIG},0)- {SPEC.ELEV.REFER}- {SPEC.P.HEIG.REFER})))*) decode({TV.XX:XX.FLAG1},L', (1+1.1e-6*{TV.XX:XX}/0.0008)*0.8821, (1+1.4e-6*({TV.XX:XX}-10)/0.0008),1)

a zároveň nedovolí jejich oboustrannou destrukci. V závislosti na investičních možnostech ČHMÚ bude replikační struktura postupně rozšířena na všechny pobočky ČHMÚ. Na jednotlivých regionálních serverech by měla být fyzicky uložena data z příslušného regionu a pro specifické operace (např. prostorová kontrola) nebo se specifickým oprávněním budou dostupná i data z jiných oblastí.

ZACHÁZENÍ S DATY

Pokud jsme se v předchozí části věnovali datům z hlediska jejich uložení do databáze, pak jsme si vědomi toho, že čtenářům zůstalo utajeno, jakým způsobem (a jakými metodami) jsou data do databáze ukládána, jak jsou kontrolována a jakými postupy je lze vybrat. V následujícím se pokusíme na tyto otázky odpovědět.

VSTUP DAT

Požizování dat

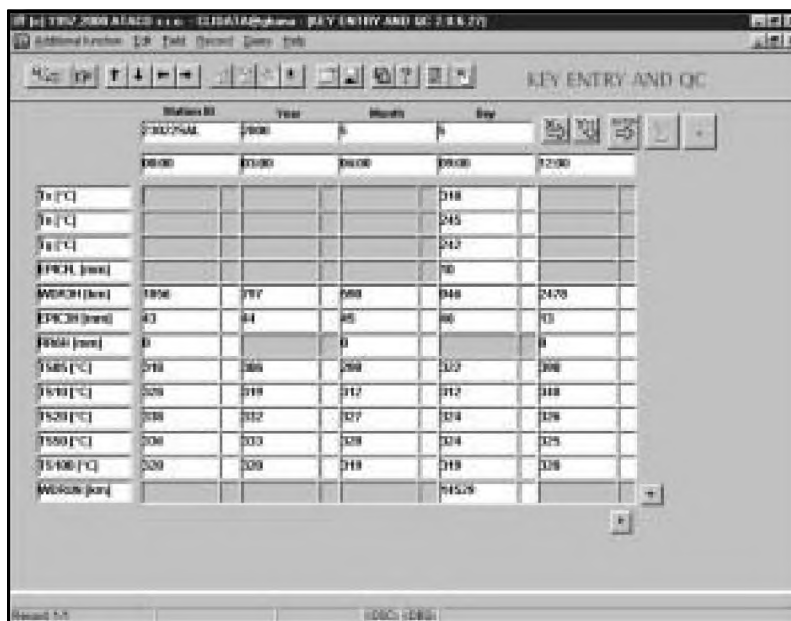
Pro data na tiskopisech zůstává jejich manuální digitalizace jediným způsobem, jak je převést do elektronického formátu. Postupy cestou rozpoznávání písma bohužel ve zdrcující většině případů selhávají a u starších dokladů o nich nelze ani uvažovat pro jejich spornou čitelnost.

Bylo již řečeno, že aplikace CLIDATA® nepoužívá speciální pořizovací formuláře; data jsou zapisována přímo do databáze, čímž je umožněno jejich okamžité použití, pochopitelně s vědomím toho, že dosud neprošla kontrolou. Protože tabulky RDATA_R a RDATA_N nejsou pro přímé pořizování vhodné, byl vytvořen formulář, který dovoluje, aby bylo pořizování pro uživatele pohodlné, především tím, že sloupce a řádky jsou uspořádány ve stejném pořadí jako na papírové předloze.

Takových formulářů lze vytvořit libovolné množství. Pořizovací tiskopis vytváří systémový administrátor ve dvou krocích; v prvním vybere potřebné prvky a ve druhém je uspořádá do odpovídajícího pořadí, obvykle podle termínů pozorování.

Papírový tiskopis obsahuje obvykle větší počet sloupců a řádků, než který lze zobrazit na obrazovce. Proto tiskopis po obrazovce roluje způsobem, který je předem volitelný (pořizování po sloupcích nebo po řádcích). Pořizování je možné kdykoliv přerušit a opět po libovolné době se k rozpracovanému formuláři vrátit. Pokud mezi tím byla již některá data opravena, jsou na obrazovce odlišena šedou barvou a do takto vyznačených sloupců již nelze vstoupit. Šedá jsou rovněž pole prvků, pro které nemá stanice definovaná pozorování, tiskopis lze proto použít pro celou skupinu stanic.

Autoři si byli při analýze této části vědomi toho, že při pořizování dochází z různých příčin k chybám, kterým prakticky nelze zamezit. Jako opatření vedoucí k minimalizaci jejich počtu jsme použili dvě nezávislé metody. První z nich



Obr. 6 Pořizovací formulář.

Fig. 6. Key entry form.



Obr. 7 Obrazovka definic pořizovacích formulářů.

Fig. 7. Key entry form definitions.

jsou součty, které mohou být v tiskopise vytvořeny buď po sloupcích, nebo po řádcích. Druhá umožňuje opětovné pořízení (přezkoušení) již vytvořených dat, pochopitelně před jejich validací.

Pro usnadnění práce jsou v pořizovacím tiskopisu tlačítka, umožňující předvyplnění sloupce definovanou hodnotou (často nulou, např. u charakteristik sněhové pokrývky mimo zimní období). K dispozici jsou i funkční klávesy a posouvací lišty.

Pokud při uložení zůstane v tiskopise některé políčko prázdné, chápe se, že tento údaj chybí a je mu automaticky

přidělen příslušný příznak. Příznaky se pořizují i tehdy, je-li třeba numerickou hodnotu prvku ještě třeba nějakým způsobem upřesnit (např. u neměřitelné výšky sněhové pokrývky).

Zvláštním případem je pořizování měsíčních dat, obvykle z ročenek. Tento postup se uplatňuje především tehdy, kdy je třeba získat taková data a pro jejich značný objem nelze postupovat standardně, tj. od termínových pozorování. V takovém případě mají všechna data příznak, definující tento stav.

Vstup dat z automatických stanic

Automatické stanice předávají data buď neprodleně v průběhu dne (dobrovolnické), nebo je předávají po ukončení měsíce na disketě (profesionální, kde je aktuální potřeba informací řešena tvorbou zprávy SYNOP). Formálně jde v obou případech o elektronický dokument, který je odpovídajícími programy zpracován tak, aby data byla odpovídajícím způsobem uložena. U dat na disketě jde o standardní postup, systém čte uložený soubor a konvertuje jej do patřičného formátu.

Data z dobrovolnických automatických stanic jsou uložena na odpovídajícím serveru jako textový soubor. Aplikace CLIDATA® detekuje jeho přítomnost, spustí příslušnou databázovou úlohu, která přenesla data a uloží je v odpovídajícím formátu. Totéž platí pro meteorologickou část dat, získávaných na stanicích AIM, případně pro srážkové úhrny z hydrologických stanic povrchových vod.

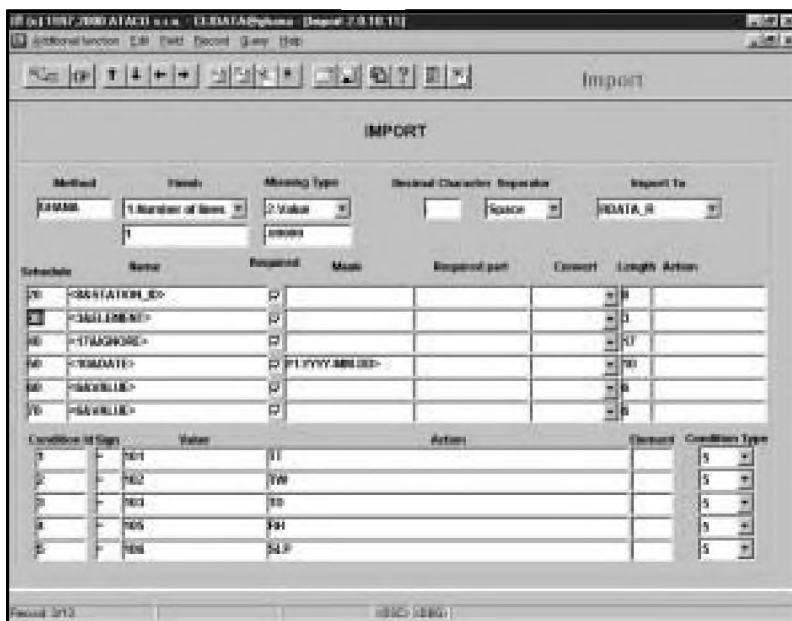
Pokud lze získávat data i z automatických stanic jiných organizací, je postup obdobný s tím, že předem musí být známa struktura a formát předávaných dat.

Vstup dat z GTS

Na telekomunikačních okruzích jsou k dispozici především zprávy SYNOP, METAR, TEMP a případně i další, pokud by vznikla potřeba je archivovat. Struktura těchto zpráv je známa, v praxi se uplatňuje analogický postup jako u informací z automatických stanic. Navíc mezinárodní výměna dat umožňuje, aby byla získána i zahraniční data, což je výhodné především při kontrole dat nebo při zpracování informací z oblastí, ležících poblíž hranic.

Vstup dříve zpracovaných dat

Již zpracovaná data jinými databázovými systémy, případně data uložená v tabulkových kalkulátorech (např. Quattro), lze do systému vkládat rozličnými způsoby. Pro soubory DataEase (soubory *.DBM) jsou připraveny postupy převádějící jeden typ uložených dat na druhý. Výsledek importu (zpráva o jeho průběhu) je uložen v textových souborech *.LOG, kde je dokumentován případný důvod, proč data přenesena nebyla (obvykle díky tomu, že původní data buď obsahují neodstraněné chyby, protože při importu jsou údaje zběžně kontrolovány, nebo jsou nedostatečně definována metadata v části rozsahu pozorování).



Obr. 8 Definice importní metody – textový soubor.

Fig. 8. Import method definition – text file.



Obr. 9 Definice kontrolních rovnic.

Fig. 9. QC formula definition.

Druhou standardní možností je import textových souborů. Protože ty mohou mít různý tvar (data na pevných pozicích nebo oddělena čárkou – soubory *.CSV) a odlišnou vnitřní strukturu (uspořádání) dat, je použita parametrizace importní metody tak, aby nebylo zapotřebí pro každý případ modifikovat importní program.

Protože prakticky všechny aplikace umožňují export textových souborů, zdá se být zbytečné rozšiřovat importní metody o další možnosti. Při importu z textových souborů je příslušný LOG soubor vytvořen jen v případě detekovaných problémů během importu.

Import informací o intenzitě srážek

Aplikace nemá přímý nástroj na vyhodnocování ombrografických záznamů. Proto jsou všechny údaje získané klasickými metodami do databáze importovány po jejich zpracování jinými metodami; jde tedy o analogii importu dříve zpracovaných dat. Data jsou ukládána do plně normalizované tabulky spolu s informacemi z automatických stanic, které takové údaje měří přímo.

KONTROLA DAT

Kontrola pomocí limitů

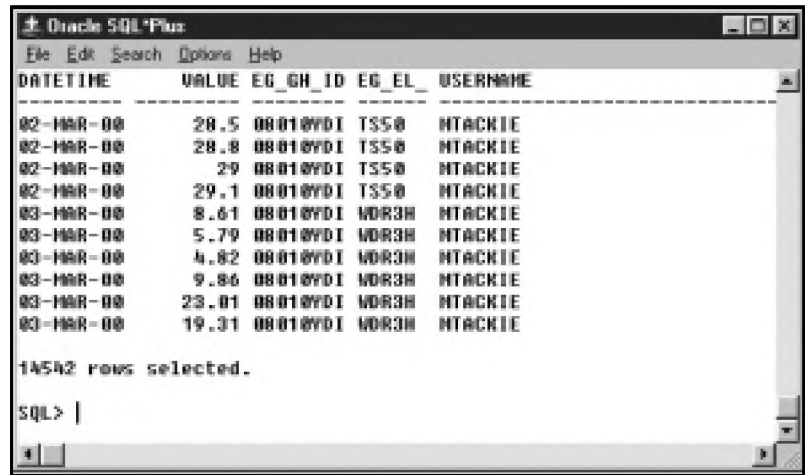
Pokud není rozloha území, pro něž jsou archivována data v aplikaci CLIDATA[®] příliš velká, pak lze použít obecně platné limity (horní a dolní) pro většinu meteorologických prvků, přičemž některý z nich vyplývá přímo z jejich definice (např. horní limit u relativní vlhkosti vzduchu je 100). U kódovaných veličin je situace jednodušší, protože rozsah kódu je předem znám. Toto omezení se používá pouze v pořizovacích tiskopisech a při importu dat, detekované hodnoty jsou evidentní chyby a není třeba dále o nich uvažovat.

Kontrola v rámci jedné stanice

Podstatná část chyb je zjištělná pomocí analýzy dat v rámci jedné stanice. V tabulce EDATA (obr. 4) jsou uloženy pro každou stanici, prvek a termín pozorování extrémní hodnoty, nejvyšší rozdíl hodnot mezi kontrolovaným a předchozím termínem pozorování a rovněž nejvyšší rozdíl hodnot oproti stejnému termínu předcházejícího dne (kontrola logického sledu hodnot v časové řadě), a to vždy v rámci měsíců. Tyto hodnoty vymezují rámec, v němž se hodnoty mohou pohybovat, pochopitelně s ohledem na okolnost, že čas od času jsou dosud nalezené extrémní hodnoty překročeny nebo podkročeny. Jde tedy v tomto případě o vážné varování, že hodnota je nejvýše nepravděpodobná a bude ji třeba zřejmě změnit.

Podkladem pro kontrolu je rovněž soustava rovnic, kterou administrátor systému zapisuje jako formalizovaný text do vytvořené tabulky. Z toho plyne, že zásadně není třeba žádné dodatečné programování a rovněž platí, že soustavu rovnic lze omezit, rozšířit nebo upravit podle skutečné potřeby. Rovnicemi se nekontroluje pouze jeden prvek, ale vzájemné vazby mezi nimi a dodržování předepsaných pravidel. Jde např. o případ, kdy teplota v termínu pozorování je vyšší než maximální denní teplota nebo kdy ve zprávě SYNOP při N = 9 je ww <41, 43, 45, 47, 49>. Do rovnic lze zařadit i používání konstant a logické analýzy případů, kdy při celodenních teplotách nad 0 °C se výška sněhové pokrývky nezměnila, aniž by bylo pozorováno sněžení.

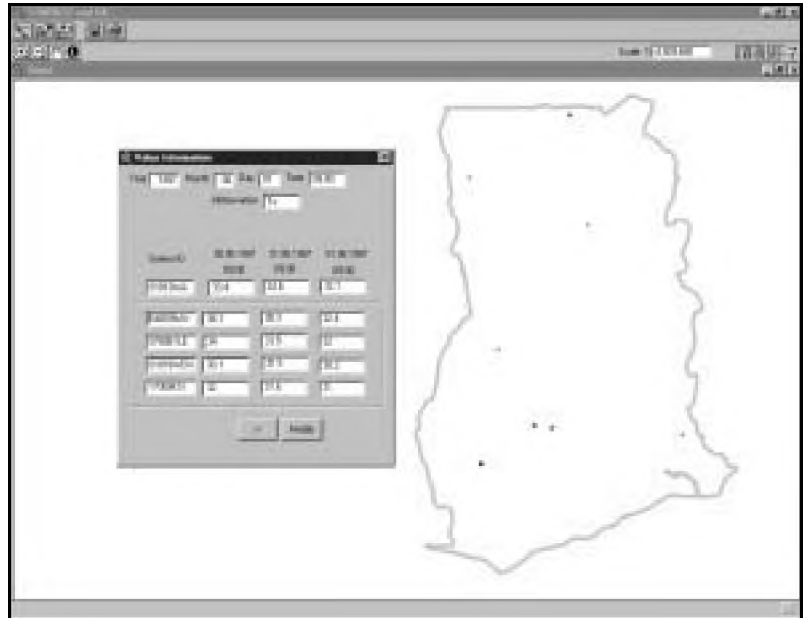
Ve všech uvedených příkladech je zobrazeno chybové hlášení, zvýrazněné červenou barvou. Toto hlášení poukazuje buď na evidentní chybu (zpráva typu error), nebo na podezřelá data (zpráva typu warning), kde je na uvážení revizora, jak se při takové situaci zachová. Při změně dat je třeba v některých případech definovat jejich příznak, tedy způsob, jakým



DATETIME	VALUE	EG_GH_ID	EG_EL	USERNAME
02-HAR-00	28.5	00010YD1	T550	NTACKIE
02-HAR-00	28.8	00010YD1	T550	NTACKIE
02-HAR-00	29	00010YD1	T550	NTACKIE
02-HAR-00	29.1	00010YD1	T550	NTACKIE
03-HAR-00	8.61	00010YD1	W0R3H	NTACKIE
03-HAR-00	5.79	00010YD1	W0R3H	NTACKIE
03-HAR-00	4.82	00010YD1	W0R3H	NTACKIE
03-HAR-00	9.86	00010YD1	W0R3H	NTACKIE
03-HAR-00	23.01	00010YD1	W0R3H	NTACKIE
03-HAR-00	19.31	00010YD1	W0R3H	NTACKIE

Obr. 10 Obsah tabulky HISTORIC_VALUE_R.

Fig. 10. HISTORIC_VALUE_R table content.



Obr. 11 Příklad oblastní kontroly pro maximální teplotu v Ghaně (červen 1997).

Fig. 11. An example of regional check for maximum air temperature in Ghana (June 1997).

byla nová hodnota získána. Možností je zapsání správné hodnoty, získané správným přečtením zápisu, který se během pořizování nepodařilo dobře vyhodnotit; v tomto případě se příznak netvoří. Ponechání původní hodnoty vyvolá automatickou tvorbu příznaku '+'. Příznakem se odlišuje zapsání odhadnuté hodnoty ('E') nebo vypočtené (objektivně stanovené) hodnoty ('C').

Původní hodnoty jsou uloženy do tabulek HISTORIC_VALUE_R, resp. HISTORIC_VALUE_N (obr. 10), čímž je umožněno pro případné budoucí analýzy se k originálním hodnotám vrátit a znovu je podrobit revizi. Zároveň je uveden autor změny. Pokud při této kontrole zůstane po otevření pořizovacího tiskopisu kurzor v prvním (nepodbarveném) poli, je to indikace toho, že v celém formuláři nebyla chyba indikována.

Oblastní (územní) kontrola

Problém oblastní kontroly spočívá v objektivním porovnání hodnot blízko ležících stanic. Takové porovnání však musí brát v úvahu územní proměnlivost meteorologických prvků, především však jejich závislost na nadmořské výšce. Proto jako základ porovnávání bylo vybráno empirické rozdělení kontrolované veličiny [4]. Tím je zohledněn fakt, že zdaleka ne všechny prvky mají rozdělení blízké normálnímu, ale především poskytnuta možnost porovnat vzájemně stanice, ležící v rozličných klimatických oblastech.

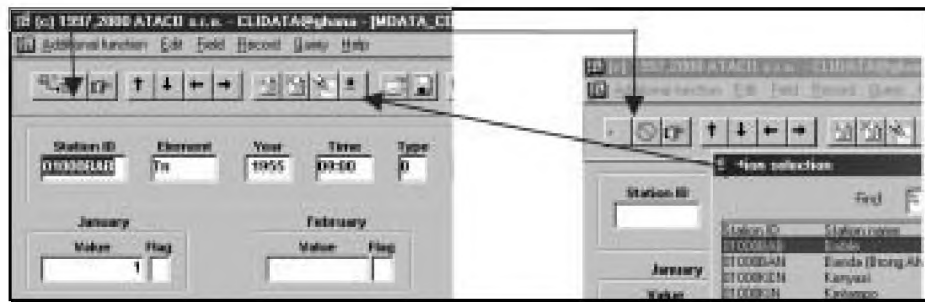
Podkladem pro kontrolu je tabulka EDATA, proto ji můžeme provádět pro všechny prvky, pro které má výpočet těchto hodnot smysl. Pokud ovšem jde o stanici nově založenou nebo stanici s krátkým obdobím pozorování, je poskytnuta možnost si hodnoty pro tabulku EDATA odvodit; prozatímní hodnoty jsou používány po jistou dobu, pak jsou nahrazeny hodnotami vypočtenými standardním postupem.

Pro mapový podklad je pravděpodobnost výskytu kontrolované veličiny převedena do barevné stupnice. Pokud hodnoty leží ve stejných nebo sousedících intervalech, mapa se nezobrazí, protože je splněn předpoklad běžné územní variability meteorologického prvku. V opačném případě se na mapě objeví barevné obdélníky (obr. 11) a je na schopnosti revizora si vytvořit náležitou představu o dané situaci.

Vyskytne-li se osamocená jedna hodnota, vymykající se ostatním, půjde pravděpodobně o chybu. Ale i zde je třeba uvážit, jaké jsou příčiny takové odchylky, např. u letních bouřkových srážek. Mimořádné jsou i případy teplotních inverzí, při nichž stanice na horách jsou vyjádřeny zcela jinak než za běžných situací. Obecně to znamená, že kontrola se musí provádět se znalostí synoptické situace.

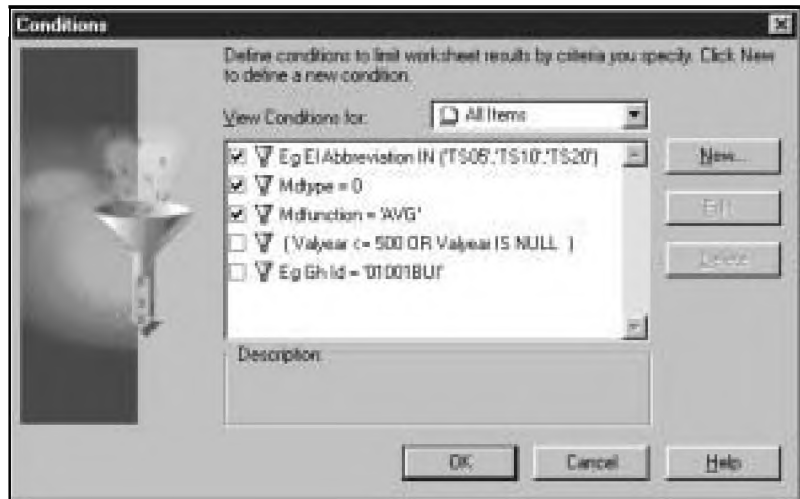
Pracovat ovšem pouze s grafickou informací by nebylo dostatečné. Proto se po přesunutí kurzoru na stanici po kliknutí myši objeví tabulka s numerickými údaji jako podklad pro následující rozhodování – uživatel si může zobrazit hodnoty pro jiný výběr stanic nebo modifikovat hodnotu pro aktuální stanici. Teprve pak zapíše správnou (opravenou) hodnotu nebo hodnoty, pokud by bylo třeba opravit údaje u více stanic. Provedené změny jsou okamžitě podrobeny předcházejícím stupňům kontroly.

Po kontrole všech prvků se provedené změny zapíše do tabulky RDATA_R a RDATA_N a zároveň se změni validační příznak



Obr. 12 Příklad dotazu do formuláře.

Fig. 12. An example of a query to the form.



Obr. 13 Zadání pro výběr dat.

Fig. 13. Conditions definition for data selection.

Year	01	02	03	04	05	06	07	08	09	10	11	12	Value
1958	34.4	35.0	34.3	34.6	33.6	31.2	29.8	29.3	31.6	32.6	34.1	34.3	32.8
1967	34.9	36.7	36.7	34.6	33.8	31.9	30.2	30.0	30.7	31.4	32.9	33.4	33.8
1958	34.1	35.6	35.9	34.2	33.1	31.1	29.8	30.4	30.7	32.3	32.8	33.1	33.8
1959	36.3	36.3	35.4	35.4	33.2	32.6	30.4	30.2	29.8	32.8	33.7	33.2	33.2
1960	34.3	37.1	36.2	34.6	34.8	32.1	30.1	29.9	30.7	31.4	34.1	33.7	33.2
1961	36.7	37.0	37.6	34.4	33.3	31.1	29.6	29.0	30.8	33.0	34.2	35.6	33.6
1962	36.6	37.2	36.6	34.4	32.6	30.8	30.6	29.7	30.6	31.7	33.2	34.0	33.1
1967	34.9	36.7	34.4	34.1	33.1	31.8	31.1	30.6		31.8	34.8	34.2	
1964	34.6	37.1	36.6	34.7	33.1	31.6	29.8	28.9	29.3	30.7	32.1	32.2	32.6
1965	34.1	35.2	35.8	33.7	31.8	30.4	29.4	28.6	30.0	31.9	33.9	33.7	32.3
1966	36.6	36.7	36.8	33.7	32.9	30.4	30.2	29.7	30.6	31.7	33.4	33.6	33.8
1967	34.8	35.9	36.8	34.4	33.2	30.6	29.8	28.6	29.7	31.3	32.7	33.0	32.4
1968	34.1	35.2	35.8	32.9	32.3	30.5	29.7	30.0	30.3	32.1	32.6	34.0	32.4
1969	36.4	36.3	36.1	34.0	32.8	31.4	29.8	29.2	30.6	31.2	33.0	33.7	32.7
1970	36.2	36.3	36.7	34.9	32.1	31.5	29.2	29.9	30.8	32.8	34.3	34.6	33.8
1971	36.4	36.6	34.2	33.6	33.4	31.0	29.7	29.6	29.8	32.3	34.1	33.4	32.7
1972	36.6	36.2	34.8	32.8	32.8	31.0	30.3	29.8	30.8	32.0	34.3	34.8	32.9
1973	34.7	37.9	37.2	35.0	34.1	31.5	30.9	30.3	30.6	32.6	34.6	34.8	33.8

Obr. 14 Vybraná data – průměrné maximální teploty.

Fig. 14. Select data – mean maximum air temperatures.

tak, aby bylo zřejmé, že data jsou validována. Tato indikace znemožňuje případné neoprávněné změny údajů. Pokud by se dodatečně zjistilo, že data byla opravena nedostatečně nebo chybně, pouze administrátor systému může provést patřičné změny a chybný údaj opět opravit. Administrátor může dodatečnou opravu nebo opravy umožnit i jiným oprávněným uživatelům. Nejde však o zcela standardní proces – pro změnu již validovaných dat musí být dostatečný důvod a vysoké oprávnění uživatele.

Kontrola meteorologických jevů

Tento problém je dosud ve stádiu úvodních analýz. Zdá se, že jedinou vhodnou možností je opět mapová analýza. Ta má však své úskalí ve skutečnosti, že kvalita pozorování meteorologických jevů ve staniční síti je vysoce variabilní a totéž v nemalé míře platí i o doprovodných časových údajích. Mnoho jevů není na dobrovolnických stanicích zaznamenáno vůbec a z pochopitelných důvodů jsou v noci zaznamenány pouze markantní jevy. Proto hledání optimálního postupu dosud stále naráží na celou řadu objektivních problémů.

VYBAVOVÁNÍ DAT

Dotaz do formuláře

Předem je třeba varovat před prohlížením v tabulkách. Jsou rozsáhlé, tudíž nepřehledné a většina z nich ani není pro uživatele přístupná. Proto všechny příklady z předcházející části nejsou vyobrazeními tabulek, nýbrž formulářů, které bylo třeba zvlášť připravit. Pokud potřebujeme zjistit obvykle pouze jednu konkrétní informaci, přepneme formulář do dotazovacího módu a zapíšeme zadání dotazu (obr. 12). Pokud databáze obsahuje vyhledávané údaje, zobrazí je.

Tvorba pohledů (view)

Prohlížení dat v databázových tabulkách je uživatelsky komplikované a nepřínáší očekávané výsledky. Administrátor může pro jednotlivé uživatele definovat zvláštní databázové pohledy – view, které jsou zpřístupněny buď v aplikačních formulářích, nebo v jiných grafických rozhraních, popřípadě pro oprávněné uživatele přímo v prostředí SQL. Vždy je nutné na cestě mezi uživatelem a daty zadat omezení formou vhodné podmínky při tvorbě daného pohledu. Počet záznamů v základních tabulkách dosahuje několikamilionových hodnot a v každém záznamu je až 93 položek pro danou kombinaci klíče. Pohledy mohou kombinovat záznamy z několika tabulek (nejčastěji kombinace datové a popisné tabulky se jménem stanice, pozorovaného prvku nebo jednotkou měření).

Grafické rozhraní (Discoverer 3.1)

Uživatelsky nejjednodušší metodou, jak vybrat z databáze potřebná data, je použít vhodné grafické rozhraní. V aplikaci CLIDATA® používáme Oracle Discoverer, který zobrazuje hodnoty v tabulkách, velmi podobných tabulkovému kalkulaátoru Excel. Pro běžnou práci uživatel nemusí znát zásady pro-

```
SQL> select m.eg_gh_id Stanice, m.year Rok, '01' Měsíc,
        m.valmon01 SRAM, n.normal Normál,
        m.valmon01-n.normal Abs.Odchylka,
        m.valmon01/n.normal*100 Rel.Odchylka
from mdata m, ndata n
where m.eg_gh_id=n.eg_gh_id
and m.eg_el_abbreviation='SRA'
and m.eg_el_abbreviation=n.eg_el_abbreviation
and m.mdtype=0
and m.mdfunction='SUM'
and m.time=n.time
and m.regular='N'
and n.month='01'
```

Obr. 15 Výpis dotazu z prostředí SQL.

Fig. 15. Query output from the SQL.

	AVG				MAX				MIN				SUM					
	TT	Ts	Ta	RR	SUNHR	TT	Ts	Ta	SUNHR	TT	Ts	Ta	RR	SUNHR	TT	Ts	Ta	RR
January	27.5	21.9	33.2	9.4		28.8	23.8	34.5		26.3	20.3	33.3	15.8					
February	28.9	21.9	35.9	18.0		31.0	24.8	37.8		25.7	17.7	33.0	34.5					
March	28.9	22.0	34.8	27.2	8.9	31.3	26.3	38.0	0.0	34.5	19.5	29.8	140.1	153.9				
April	28.9	22.5	33.8	85.6	8.9	29.9	24.5	38.6	4.8	24.3	20.0	26.3	236.6	200.0				
May	28.3	23.0	33.5	56.1		29.9	25.3	34.9		26.1	20.6	30.8	188.1					
June	27.0	22.8	31.4	48.0		28.8	24.2	33.4		25.4	20.8	29.0	285.6					
July	25.9	21.5	30.2	17.3		27.5	23.8	32.3		23.5	19.5	25.9	49.1					
August	25.5	21.3	29.7	10.9		27.7	22.3	31.5		23.0	19.5	26.2	39.5					
September	27.8	22.0	33.0	44.6	7.2	28.3	23.3	33.9	0.1	24.5	20.5	28.1	82.0					
October	27.5	22.1	32.9	31.5	9.8	29.2	23.7	34.7	1.9	25.4	21.2	30.0	169.1	188.0				
November	27.4	21.6	33.1	25.4	5.8	28.8	23.5	34.4	2.4	25.2	21.2	30.0	74.3	186.7				
December	27.1	21.4	32.9	29.2		29.1	24.0	34.6		22.2	18.2	29.0	47.5					
Year	27.4	22.1	32.7	85.6		27.3	25.3	38.0		22.2	18.2	28.8	1362.2					

Obr. 16 Příklad výstupu z Oracle Discovereru ve formátu HTML.

Fig. 16. An example of the Oracle Discoverer output in HTML format.

gramování, stačí, když vybere odpovídající stanici, prvek a požadované období v předpřipraveném databázovém dotazu.

Pochopitelně se výběr může provést pro více stanic, prvků i období. Tyto parametry se zapisují do tabulek, které systém automaticky nabízí. Podmínky nemusí být zcela elementární, je možné je vytvořit i výpočtem. Nutná je však určitá znalost vnitřní struktury databáze získaná v průběhu základního školení aplikace CLIDATA.

Výstup (obr. 14) je možné třídit, měnit uspořádání sloupců, druh a velikost fontu i jeho barvu, doplnit jednoduchými výpočty, zvýraznit v něm výjimečné hodnoty (vyšší nebo naopak nižší než zadaná hodnota zvlášť pro každý sloupec), použít grafické zobrazení nebo zvolit další postupy, jejichž popis však přesahuje rámec této informace. Pokud za základ vybraných dat zvolíme optimálně navržený pohled, můžeme velmi snadno připravit tabelaci výstupu obdobnou ročenkovým výstupům nebo podklady podle přání jednotlivých uživatelů.

Pro konečný výstup si může uživatel zvolit z mnoha možností (XLS, TXT, PRN, CSV, HTM a další). Výstupy lze předpřipravit tak, že uživatel pouze změní připravené parametry v BAT souboru a celý dotaz spustí, aniž by musel jakkoliv do procesu tvorby výstupu zasahovat.

Programovatelný výstup (SQL)

I když pomocí grafického rozhraní je možné získat prakticky všechny potřebné informace, přesto se může vyskytnout případ, kdy nabízené možnosti nestačí. Pro tento případ je k dispozici programovací jazyk SQL (PL/SQL). Relativně jednoduchými příkazy lze zadat různě komplikované dotazy, umožňující optimální výběr dat (obr. 15). Pro celkovou složitost problematiky odkazujeme zájemce na odbornou příručku nebo specializovaný kurz.

ODBC driver

Operační systémy a některé programové produkty obsahují přímé propojení do databáze Oracle, umožňující přímý přenos dat mezi nimi a aplikací CLIDATA[®]. Tak je možné předávat data pro zpracování programovacími jazyky (Visual Basic, Delphi) nebo specializovanými programovými balíky (Mathematica, Systat 9) k dalšímu zpracování. Pro pokročilého uživatele je tato možnost mimořádně výhodná. V naší aplikaci je tato možnost použita pro oboustranné propojení mezi ní a grafickým informačním systémem ArcView (viz oblastní kontrola dat).

Internet

Pomocí Oracle Reports nebo Oracle Discoverer je možné vytvářet off-line www stránky. Jejich tvorba se tím automatizuje do té míry, že je možné je nejen plánovat (vytvářet podle časového schématu), ale připravovat je nejen jako obecnou informaci, ale i jako specializovaný výstup pro vybraného uživatele, který si ji vyzvedne po zapsání odpovídajícího hes-

la z www nebo ftp serveru. Kombinace naplánovaných úloh a předem definovaných parametrizovaných dotazů umožňuje vysoce bezpečné a přítom aktuální off-line pohledy na data v databázi.

ZÁVĚR

Specifikace datového modelu je jednou z nejdůležitějších částí při vývoji databázové aplikace. Použitý datový model určuje způsob práce programátora. Běžný uživatel, který má k dispozici dobré aplikační rozhraní, nepotřebuje mít o datovém modelu žádné znalosti. Podrobný popis možností vstupu dat do aplikace CLIDATA[®] a možností jejich výběru je vždy součástí instalace aplikace na daném pracovišti. Nedílnou součástí aplikace jsou i tzv. povinné produkty, jejichž popis bude obsahovat některé z příštích pokračování informací o aplikaci CLIDATA[®].

Literatura

- [1] Calculation on Monthly and Annual 30-year Standard Normals. WCDP-10, WMO/TD-No.341. Geneva, WMO 1989.
- [2] *Coufal, L. – Tolasz, R.*: Klimatologická databáze CLIDATA (část metadata – informace o měřicích stanicích). Meteorol. Zpr., **52**, 1999, č. 2, s. 55–58.
- [3] *Tolasz, R.*: Nové směry ve vývoji klimatologických databází. Meteorol. Zpr. **51**, 1998, č. 3, s. 85–87.
- [4] *Tolasz, R. – Coufal, L.*: The Static Data Quality Process as Part of CLIDATA Application. In: Österreichische Beiträge zu Meteorologie und Geophysik, Heft 19 Proceedings of the 2nd European Conference on Applied Climatology, s. 155. Wien, Zentralanstalt für Meteorologie und Geodynamik 1998.

Lektor RNDr. M. Hradil, rukopis odevzdán v únoru 2001.