

Zpracování dat a databáze, Big Data

Daniel Žížala



Data v zemědělství

- Data - údaje získané měřením, pozorováním
- v zemědělství velice variabilní

- **Agronomická data**

- Výnosová data
- Obsah živin
- Půdní analýzy

- **Data ze strojů**

- Spotřeba
- Diagnostika
- Dráha pojezdů

- **Data o počasí**

- Srážky, teplota

Data z různých zdrojů

- **Senzorová data**
- Transakce
- **Podniková data**
- Údaje o přihlášení uživatelů
- **Obrazová data**
- Vztahy
- Sociální média
- **Mobilní aplikace**
- Email
- **Prostorová data**
- **Signál GNSS**
- Internet věcí



Zpracování dat

- evidování a zpracování velkého množství údajů o velkém množství objektů.
 - základem jsou informace, které jsou potřeba pro rozhodování
- Tradiční postup pro větší objem dat:
 - extrakce relevantních informací z různých zdrojů
 - transformace – zpracování dat do podoby korektních a konsolidovaných dat
 - load - naplnění zpracovaných dat do cílového systému datového skladu – databáze
 - analytické nástroje pro provádění složitých výpočtů nad daty ze skladu tzv. Data Mining (dolování dat) založené na pokročilých analýzách dat za pomoci matematických a statistických metod
- Problém:
 - pro ukládání dat v tradičních databázích je potřeba mít jasnou strukturu dat, nastavené vztahy, jsou obtížně škálovatelné
 - datové sklady nejsou připraveny na zpracování a analýzy velkých objemů komplexních dat ve velmi krátkém čase (někdy real-time) nebo nákladově efektivním způsobem.
 - využití databází (geodatabází) – SQL řešení – pomalé nebo s velkou redundancí dat -> nároky na finance



BIG DATA

- V současné době výrazný nárůst dat
- Základní problém
 - Uložení, management a zpracování datových sad na normálních desktopových počítačích je nemožné
 - Je proto potřeba využít nové nástroje pro správu dat
- Možnosti:
 - Nové typy databází – key-value databáze, NoSQL databáze, grafově orientované databáze a další modifikace klasického přístupu
 - Big data
 - Zpracování petabytů dat
 - Zpracování dat volně strukturovaných (pokud mají vůbec nějakou strukturu), vysoce distribuovaných se vzrůstajícím objem.



BIG DATA

„Buzzword“ ~~Velká data~~

“ Oblast ve které se točí hodně peněz. Business model jak vydělat na datech

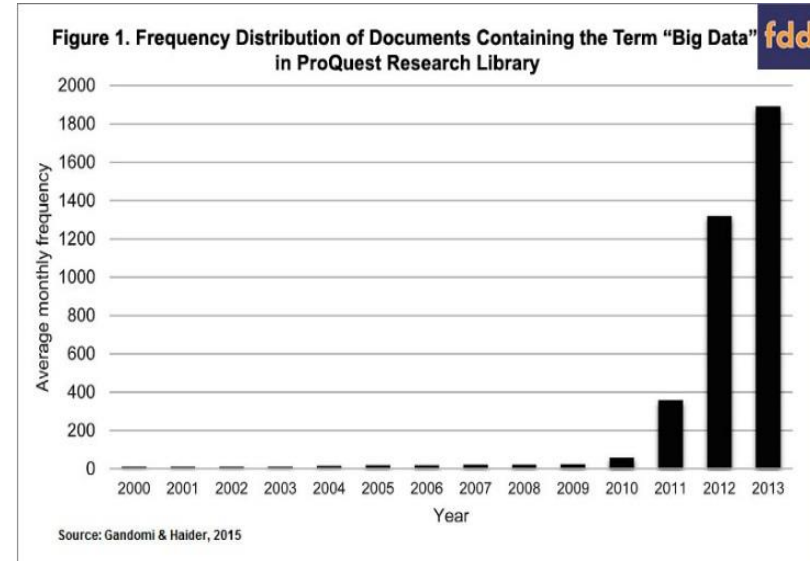
“ všichni o nich mluví, píší, ale jen velice málo se skutečně zpracovávají a používají (Gartner, 2012).

“ Big Data jsou velkoobjemové, rychlé a velice variabilní informační praktiky které vyžadují nákladově efektivní, inovativní formy zpracování informací pro lepší pochopení a rozhodování (Gartner IT Glossary, 2012).

“ schopnost agregovat informace pro nalezení trendů, vzorců a schémat (Todd Janzen)

“ termín aplikovaný na soubory dat, jejichž velikost je mimo schopnosti zachycovat, spravovat a zpracovávat data běžně používanými softwarovými nástroji v rozumném čase. (Gartner IT Glossary, 2012).

“ schopnost extrahovat informace a dovednost nahlédnout tam, kde to nebylo dříve možné. (Steve Sonka, 2015)



BIG DATA

Hovoříme o 3V (4V) dimenzionalitě

- » **Volume (objem)**
 - › Objem vznikajících dat narůstá exponenciálně
- » **Velocity (rychlost)**
 - › Stoupá rychlost vzniku dat a potřeba rychlého zpracování
 - › Úlohy vyžadující okamžité zpracování velkého objemu dat – např. zpracování obrazu v reálném čase
- » **Variety (variabilita, různorodost)**
 - › Stoupá variabilita dat
 - › Úlohy zpracování nestrukturovaných dat (prostorová data, data z logů) nebo např. multimediálních dat
- » **Veracity (věrohodnost)**
 - › Nejistota v důsledku inkonzistence, neúplnosti, nejasnosti apod.



BIG DATA

- nové způsoby zpracování a analýzy velkých objemů dat, např. nástroj Apache Hadoop
- frameworky pro zpracování, ukládání a analýzu velkého množství distribuovaných, nestrukturovaných dat
- stavěny pro zvládání petabytů a exabytů dat distribuovaných přes více uzlů současně
- Systémy zajišťují přístup k datům, jejich rozdělení do více replikovaných dílů, paralelní zpracování na jednotlivých uzlech, uložení dat a jejich distribuce do analytického prostředí
- Důležitou komponentou je vizualizace dat – zobrazení hledaného problému do „lidské řeči“



BIG DATA

Nové sofistikované technologie vyžadující odborné znalosti a různé stupně propojení

Potřeba zapojit

- Primární sektor – farmáře
- Vědce a vývojáře
- Softwarové a hardwarové specialisty
- ...

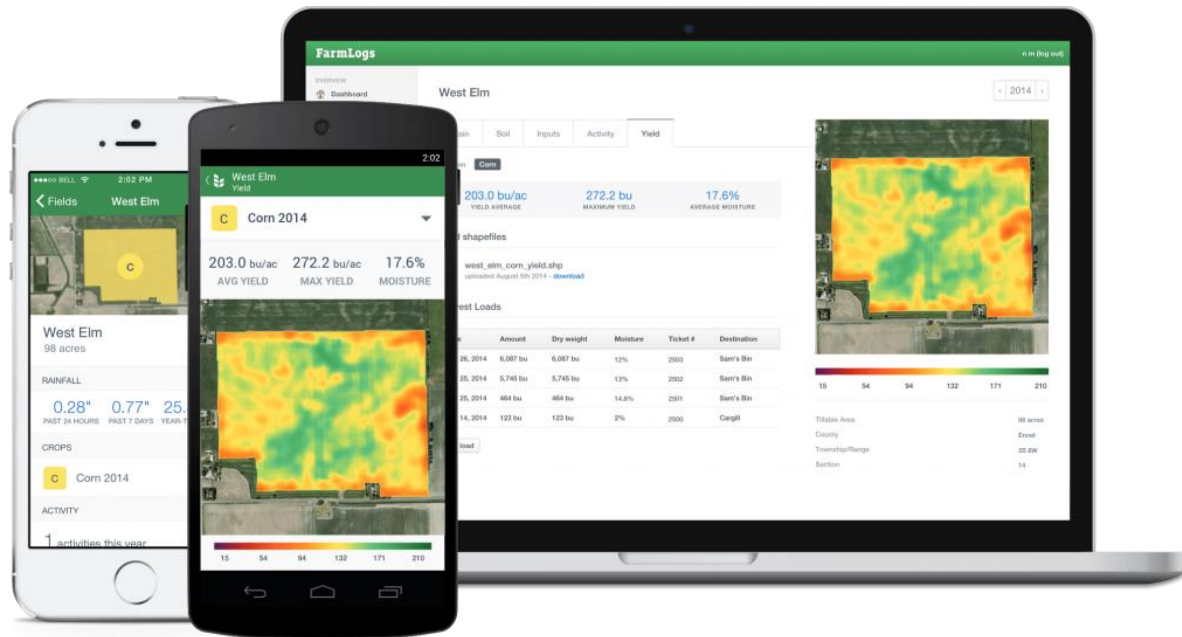
Spolupráce a výměna dat (srážky, půda, data od farmářů apod.) je nezbytná

Platforma Centra precizního zemědělství může v tomto ohledu posunout věci výrazně dopředu



Příklady z praxe

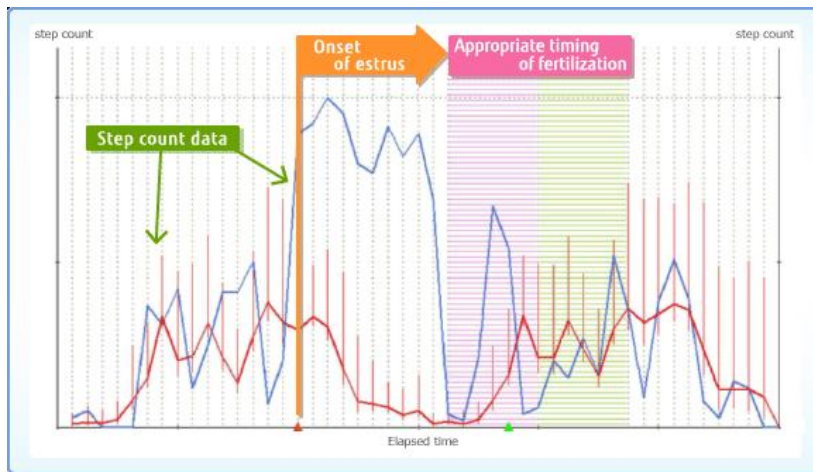
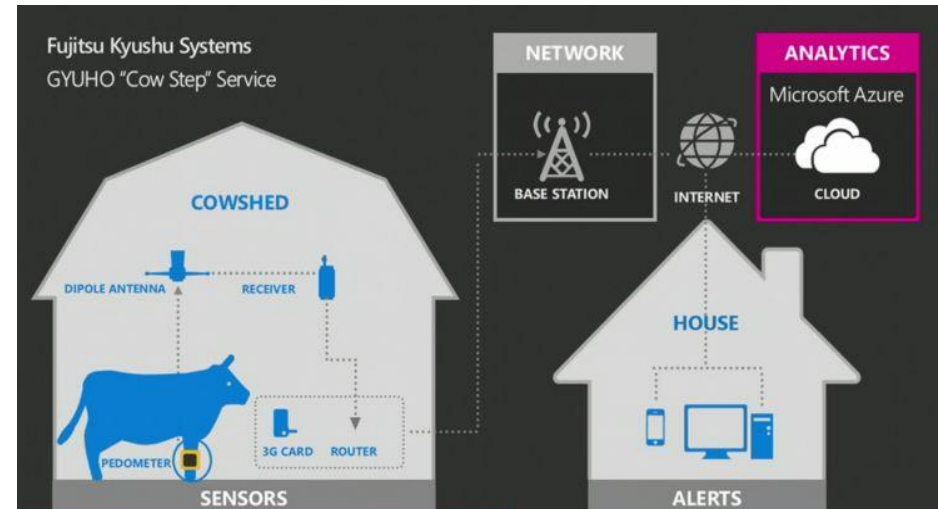
- » **Aplikace FarmLogs**
- » Údaje o stavu půdy či srážkových úhrnech a další data přímo z terénu jsou doplněny o analýzy. Farmář si je může zobrazit na počítači, tabletu nebo smartphonu a svou práci a zdroje pak může přizpůsobit aktuálním podmínkám.



Příklady z praxe

» Služba GYUHO SaaS

Japonský projekt, který zemědělcům ukazuje, kdy je nejvhodnější doba k inseminaci dobytka.



Daniel Žížala

zizala.daniel@vumop.cz



DĚKUJI ZA POZORNOST

