

Big data v praxi ...

Ing. Milan Paštrnák, PhD
Atos IT Solutions and Services s.r.o.

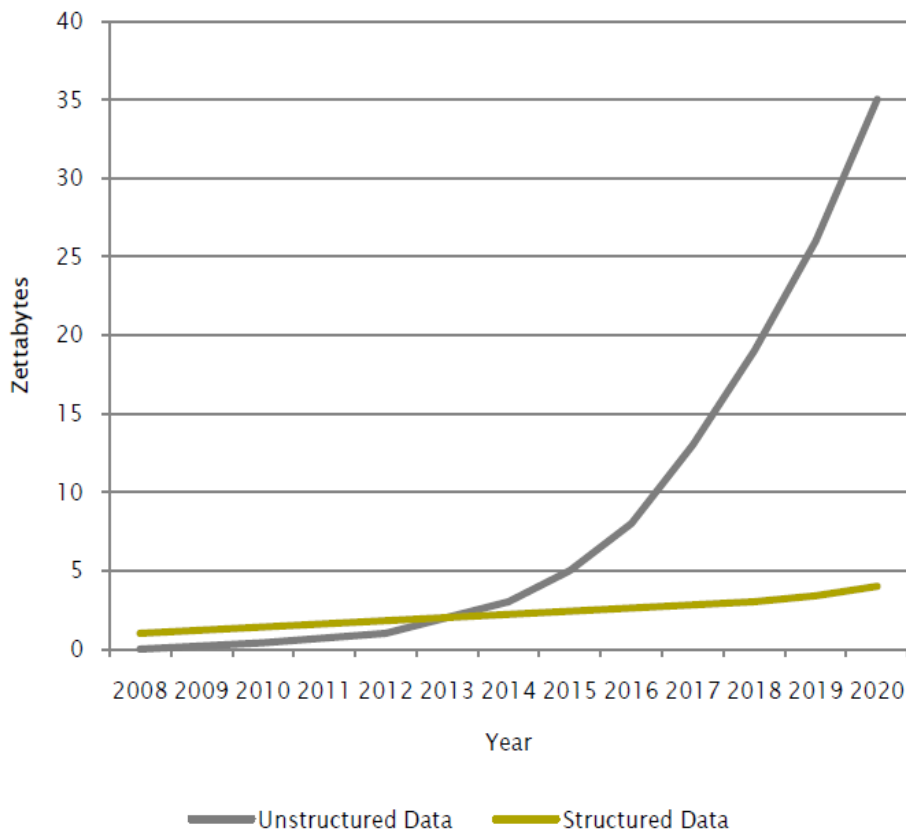
eFOCUS

Atos

Big data - intro

- ▶ Každý deň vytvoríme **2.5 quintillion bytes** dát – na porovnanie: 90% súčasťných dát bolo vytvorených v posledných 2 rokoch ...
- ▶ Dáta prichádzajú **zovšadiaľ**: senzory poveternostných podmienok, príspevky na stránkach sociálnych sietí, digitálne obrázky a video uploadované online, transakčné záznamy z online platieb alebo GPS signál z mobilných zariadení ...
- ▶ Data sety, ktorých rozsah je mimo možností štandardne používaných software-ových nástrojov takéto dáta zozbierať, usporiadať a spracovať v prijateľnom čase na spracovanie.

Prečo Big data



Typy dát

- ▶ **Štruktúrované dáta:** postavené na relačných databázach (t.j. plná atomickosť, konzistencia, izolovanosť a tvárnosť), referenčná integrita, silná podpora typov a schém
- ▶ **Semi-štruktúrované dáta:** Metadata, XLS tabuľky, XML dáta definované XSD schémami
- ▶ **„Kvázi“ štruktúrované:** excelovské webové click-stream dáta (obsahujúce nekonzistenciu v dátových hodnotách a typoch)
- ▶ **Neštruktúrované:** najmä dokumenty obsahujúce text, obrázky, videá a ďalšie objekty, ktoré je pre ďalšie využitie potrebné analyzovať

Big data – vertikálny pohľad

Atos

eFOCUS

Buisness aplikácie Big data

Všeobecné kategórie non-CRUD použitia

- ▶ Recommendations & Matching
- ▶ Graphs
- ▶ Search
- ▶ Data Normalization
- ▶ News feed
- ▶ Analysis & Monitoring

Odporúčania

Frequently Bought Together

Customers buy this item with [Case Logic Com](#)

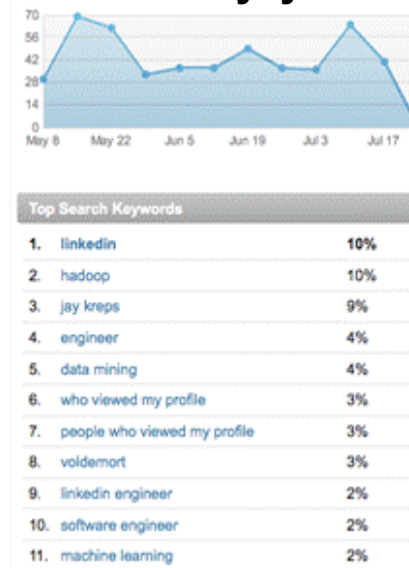


Price For Both: \$

[Add both to Cart](#)

These items are ship

Analýzy



Vyhľadávače

logistic regression

Keep filter selections

[Search](#) [More](#)

Refine By

Current Company

All Companies

Bank of America (38)

Accenture (28)

SAS (28)

IBM (21)

HSBC (20)

Show more...

Enter company name

Relationship

All LinkedIn Members

1st Connections (2)

2nd Connections (124)

RSS kanály

Stephen Holsapple

LinkedIn Debuts Embeddable Job Application Tool For Companies
techcrunch.com

In April, LinkedIn launched new customizable plug-ins, which are features from the professional social network that can be customized and embedded on websites with minimal effort. LinkedIn plug-ins launched with the ability to...

Like • Comment • Send a message • Share • 47 seconds ago

Xuan Nguyen LinkedIn is making it easier for our users to apply for jobs.

LinkedIn Launches Button That Lets You Apply for Jobs
mashable.com

LinkedIn just made it easier to apply for a job, thanks to the launch of its new "Apply With LinkedIn" button.

Like • Comment • Send a message • Share • 2 minutes ago

Vertikálny pohľad na Big data

| Big data charakteristika (príklady) | Verejný sektor | Telekomunikácie a média | Financie | Výroba, predaj a služby | Energetika a utility |
|---------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------|
| Asociatívne korelačné vzory | <ul style="list-style-type: none"> ▶ Transparentné služby ▶ Popora eGovernance | <ul style="list-style-type: none"> ▶ Analýza správania ▶ Cross-selling (previazané produkty) ▶ Ponuky zamerané na lokality | <ul style="list-style-type: none"> ▶ Cross-selling založený na podobnej spotrebe | <ul style="list-style-type: none"> ▶ Zvyšovanie produktivity a spájania výrobných kapacít | <ul style="list-style-type: none"> ▶ Efektivita prevádzky ▶ Optimalizácia spotreby |
| Vytváranie nových produktov a služieb | <ul style="list-style-type: none"> ▶ Stimulácia rezortov k poskytovaniu služieb verejnosti efektívnejším spôsobom | <ul style="list-style-type: none"> ▶ Web-based zákazníci ▶ Služby porovnávania cien ▶ Zľavové kampane zamerané na špecifické segmenty | <ul style="list-style-type: none"> ▶ Poistenia orientované na bydlisko \ zamestnanie ▶ Produkty založené na údajov zo záväzkov | <ul style="list-style-type: none"> ▶ Budovanie a rozširovanie výrobných liniek založené na logoch senzorov | <ul style="list-style-type: none"> ▶ Budovanie a rozširovanie výrobných liniek založené na logoch senzorov |
| Podpora rozhodovania | <ul style="list-style-type: none"> ▶ Detekcia nezrovnalostí napr. dane, penzia ▶ Automatizované vyplňanie formulárov k zníženiu prevádzkových nákladov | <ul style="list-style-type: none"> ▶ Optimalizácia práce a prac. zariadení ▶ Cenová optimalizácia ▶ Optimalizácia umiestňovania zdrojov | <ul style="list-style-type: none"> ▶ Real-time detekcia podvodov | <ul style="list-style-type: none"> ▶ Optimalizácia inventúr ▶ Optimalizácia logistiky ▶ Cenová Optimalizácia | <ul style="list-style-type: none"> ▶ Inteligentné dodávky ▶ Automatizované diaľkové odpočty |

Use-case – Finančný sektor

Vymedzenie

- ▶ Spoločnosti poskytujúce finančné služby celosvetovo vyhodnocujú a analyzujú „big data“ aby boli o krok vpredu pred ich konkuretnmi. They need to
 - ▶ Detekcia a prevencia podvodov
 - ▶ Maximize prevádzkovej efektívnosti

Buisness obmedzenia

- ▶ Eliminácia spojených rizík s podvodmi
- ▶ Časovo náročné a komplexné úlohy

Technické obmedzenia

- ▶ Objem a rozličnosť dát určených na analýzu
- ▶ Tvorenie modelov z transakčných dát pre real-time nástroje na vyhodnocovanie rizík a detekciu podvodov

Aplikovanie Big Data

Použitie Big data riešení

- ▶ Spúšťanie komplexných query korelujúcich množstvá aktivít v rôznych data setoch na predchádzanie podvodov s ukradnutými kreditnými kartami.
- ▶ Nové metódy prevencie zakomponované vo vyhľadávacích algoritmoch a prevádzke systémov
- ▶ **Big data akcelerátory:** Real-time detekcia podvodov, efektívita prevádzky

Očakávané prínosy

Business Case

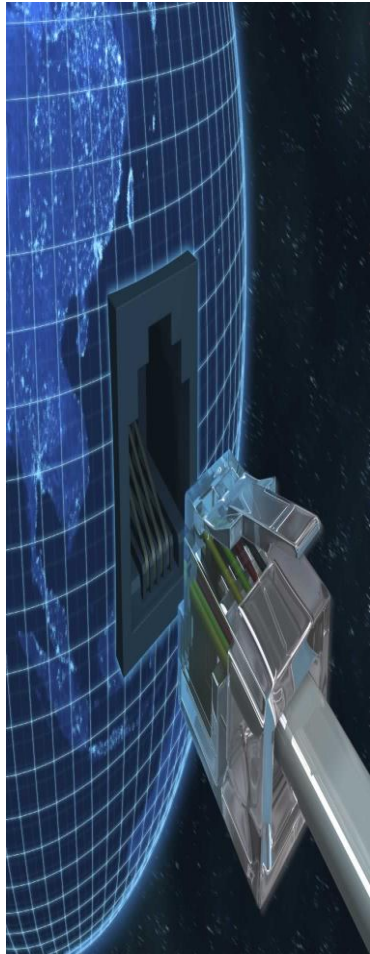
- ▶ Straty z titulu podvodov
- ▶ Maximalizácia prevádzkovej efektívnosti

Vplyv

- ▶ Redukcia strát z dôvodu podvodov s kreditnými kartami cez big data analytickú real-time infraštruktúru
- ▶ Spúšťanie náhodných & interaktívnych query maximalizujúce prevádzkovú efektívnosť



Use-case – Telekomunikácie



Vymedzenie

- ▶ S ustálením počtu mobilných zákazníkov dosiahla súťaž medzi mobilnými operátormi konečný bod a ich hlavný cieľ sa mení zo získavania nových zákazníkov na:
 - ▶ Udržanie zákazníkov po skončení viazanosti
 - ▶ Efektivita kampane

Aplikovanie Big Data

Použitie Big data riešení

- ▶ Big data otvára BI služby pre operátorov rýchlo identifikovať problémy zákazníkov na zvýšenie spokojnosti
- ▶ Big data vedia spracovať rôzne typy dát o rôznych službách a na základe získaných dát aplikovať komplexnú analýzu vedúcu k vytvoreniu detailného profilu zákazníka
- ▶ **Big data akcelerátory:** Zníženie nekvality prenosov, cielený marketing, cross selling produktov

Business obmedzenia

- ▶ Sprostredkovanie CDR's na billing systém
- ▶ Zlepšiť rýchlosť a kvalitu billing procesov a realácie kampaní

Technické obmedzenia

- ▶ Real time zbieranie informácií
- ▶ Možosti spracovávať bilióny záznamov o telefónnych hovoroch
- ▶ Integrovaná celopodniková infraštruktúra

Očakávané prínosy

Business Case

- ▶ Zvýšenie revenue
- ▶ Zákaznícka spokojnosť a udržateľnosť Vplyv
 - ▶ Vypočítané analýzy poskytované prepojavacím uzlom znižujú dĺžku prepojov a čakaní čím zvyšujú revenu a spokojnosť zákazníkov
 - ▶ BI služby zabezpečujú neprerušované služby pre klientov pri rotácii zamestnancov

Use-case – Zdravotníctvo

Vymedzenie

- ▶ Rozsiahle a komplexné dátové sety ako napríklad MRI záznamy, snímky, genetický výskum sa stávajú základnou normou pre
 - ▶ Určenie príčin ochorení
 - ▶ Vývoj nerušivých testov a predikovanie chorôb

Business obmedzenia

- ▶ Včasná detekcia možných život ohrozujúcich podmienok na JIS
- ▶ Umožňuje lekárom verifikovať nové klinické hypotézy

Technické obmedzenia

- ▶ Real time analyza & korelácia dátových streamov fyziologických dát
- ▶ Začlenenie a integrácia zdrojov audio, video, elektronických diagnostických prostr.

Aplikovanie Big Data

Použitie Big data riešení

- ▶ Výskum má k dispozícii nástroje na data mining pre porovnanie diagnóz a následných liečebných postupov
- ▶ Otvára sa možnosť rýchleho prístupu k elektronickým záznamom vrátane digitálneho zobrazovania
- ▶ Vytváranie multi-dimenz. informácií o pacientoch, liekoch a financiách
- ▶ **Big data akcelerátory:** Zvyšovanie kvality starostlivosti o pacienta ale aj efektívnosti poskytnut. starostlivosti

Value Proposition

Business Case

- ▶ Zvýšenie kvality starostlivosti
- ▶ Redukcia nákladov

Vplyv

- ▶ Vytvorenie podporných systémov rozhodovania založených na výhodách big data analýze pre uľahčenie výberu najlepšieho spôsobu liečby
- ▶ Štandardizácia elektronického zdravotného záznamu medzi zariadeniami znižujúce celkové prevádzkové náklady

Use-case – Energetika a utility



Vymedzenia

- ▶ V ropnom priemysle sú úniky ropy jedným z najkritickejšie sledovaných tém. Presná lokalizácia je kľúčovou pri odhadovaní závažnosti problému v prípade havárií.
 - ▶ Problémová analýza

Business obmedzenia

- ▶ Nejednoznačnosť pri presnom určovaní možných únikov
- ▶ Zhromaždenie dát vyžaduje čas a nákladnosť pri zhromažďovaní

Technické obmedzenia

- ▶ Ťažkopádne a často manuálne analýzy dát
- ▶ Neintegrovane a premiešané zdroje štruktúrovaných a neštruktúrovaných dát

Aplikovanie Big Data

Použitie Big data riešení

- ▶ Meniace sa odhady na základe použitia rôznych zdrojov informácií môžu byť analyzované použitím techník analýz neurčitostí pre spresnenie a následnú akcieschopnosť
- ▶ Analýza štruktúrovaných & neštruktúrovaných dát na spresnenie odhadov
- ▶ **Big data akcelerátory:** Zníženie strát ziskov, reportovanie dodržiavanie regulácii

Value Proposition

Business Case

- ▶ Zníženie strát ziskov
- ▶ Eliminácia dopadov havárií

Vplyv

- ▶ Efektívna a rýchla detekcia únikov ropy pomáha koordinovať zásahy pri havarijných stavoch a tak znižovať straty ziskov

Big data - referencie

Atos poskytuje rozsiahly rad služieb ktoré umožnia zákazníkom:

- ▶ Preskúmať reálnu hodnotu, ktorú môžu získať z dynamicky meniaceho sa trhu
- ▶ Pochopiť pripravenosť na využitie týchto možností
 - ▶ Integrácia Big data s ich business and IT stratégiou
 - ▶ Získať vnútorný pohľad na buisness použitím pokročilých analyt. techník
 - ▶ Využiť špecifickú množinu BDaaS platforiem na pilot a „proof of concept“ implementáciu
 - ▶ Implementácia Big data riešení
 - ▶ Prevádzka kompletnej Atos BDaaS platformy ako zero-capex riešenia

Strategický partneri



Hlavné referencie

The logo for e-plus+ is displayed in a green, sans-serif font.



The logo for ATOC is displayed in a blue, stylized, sans-serif font.

The logo for KARSTADT is displayed in white, uppercase, sans-serif font on a dark blue rectangular background.

Big data – horizontálny pohľad

Atos

eFOCUS

Technologické stavebné bloky BD - I

File Storage

Potrebné pre ukladanie rozsiahlych neštruktúrovaných dát. File system orientované riešenia poskytuje dobrá priamočiara možnosti škálovateľnosti.

Hadoop Distributions

Hadoop – free software framework podporujúci „data-intensive computing“. Na dosiahnutie potrebnej stability firmy ako napr. Cloudera rozvíjajú použitie a akceptáciu Hadoop možností.

Connectors/ Integration

Prevažne proprietárne integračné nástroje na prepojenie big data s dátami organizácie.

Analytické platformy

Vysoko výkonné analytické nástroje priamo zabudované vo veľkých paralelných databázach. Analytické platformy riešiace business výzvy vyžadujúce analýzu nad terabytmi až petabytmi dát.

Technologické stavebné bloky BD - II

Tools & Analytics

Software-ové produkty pre dátových analytikov a profesionálov na prechod k Big Data, zlepšujúce schopnosti využívať nové paternity, vzťahy a určujúce faktory obsiahnuté vo všetkých typoch dát neštruktúrovaných, štruktúrovaných, semi-štruktúrovaných – čo bolo predtým nemožné.

Spreadsheets/ Visualizácia

Kvôli veľkým objemom spracovávaných dát je pomerne zložité identifikovať skutočný význam analýzy, kde vizualizácia môže výrazne urýchliť prírmanie rozhodnutí.

Analytical application

Špecifické software-ové aplikácie používané na vyhľadávanie, monitorovanie, reportovanie a analýzu „machine-generated“ datasets.

BD Ecosystem

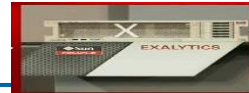
Analytical Applications



Spreadsheets/ Visualization



Tools & Analytics



Analytical Platform



Hadoop Distributions



Connectors/ Integration



File Storage



Atos Research & Innovations a Big data

Atos

eFOCUS

Big Data Public Private Forum

Vytvoriť udržateľnú technologickú komunitu Big Data v Europe !

Ciele projektu sú:

- Vytvorenie prostredia poskytujúcu korektné informácie a technologický pohľad na vývoj v technológiách big data
- Zameranie sa na kľúčových „decision takers“ v EU; ako vo vedeckých komunitách tak aj v IT technologických firmách
- Propagácia prvých reálnych použití „big data“ technológií v praxi
- Identifikácia existujúcich bariér ako sú rôzne regulácie a normy spomaľujúce nasadenia big data technológií a jednotný postup pri ich riešení v EU



Helix Nebula je tiež BD

- ▶ Najväčšie výskumné Európske centrá (CERN, EMBL, ESA) vytvorili s IT dodávateľmi konzorcium s cieľom vytvoriť platformu pre uspokojenie IT požiadaviek Európskych vedcov
- ▶ 2 ročný pilotný projekt - Helix Nebula bol testovaný na 3 projektoch:
 - ▶ CERN projekt ATLAS – large hadron collider
 - ▶ EMBL – whole genome assembly
 - ▶ ESA – pozorovanie sopiek a zemetrasení

▶ 12.3.2013 ESA prezentovala Helix Nebula výsledky pilotného projektu na konferencii V Holandsku

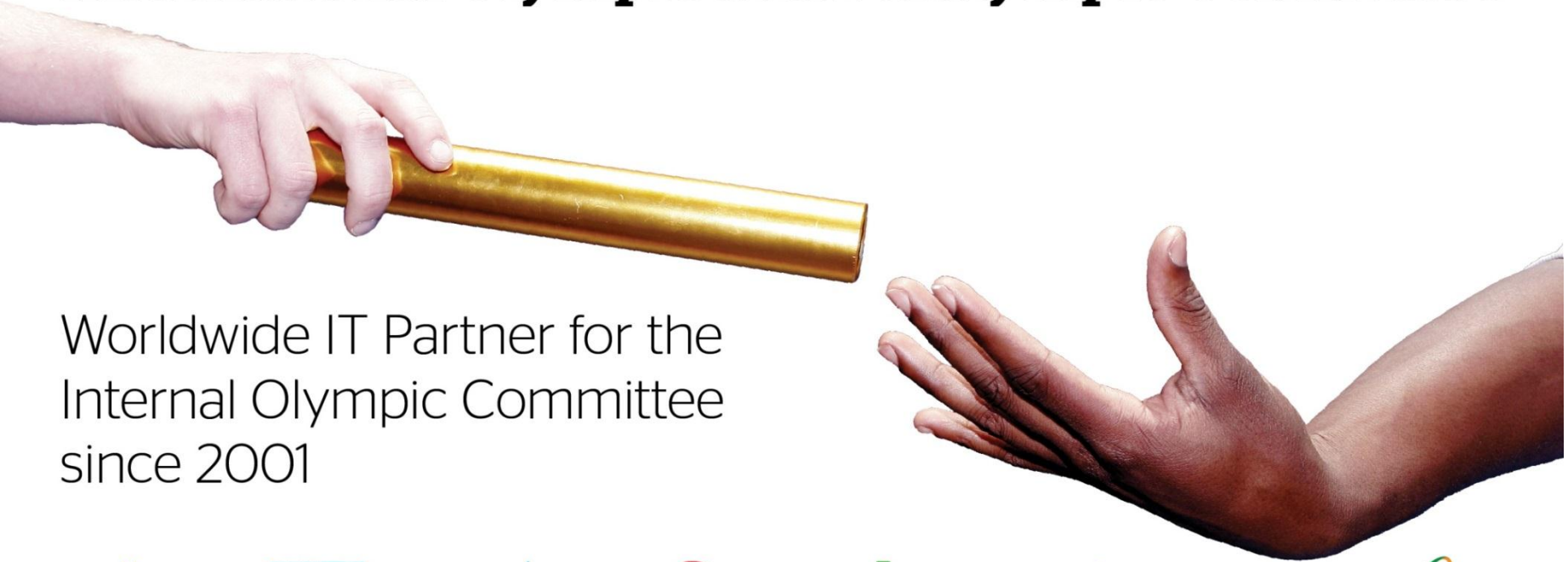


Olympijské hry a Big data ?

Atos

eFOCUS

Atos and the International Olympic and Paralympic Committee



Worldwide IT Partner for the International Olympic Committee since 2001



Proud to be a Worldwide IT Partner for the International Paralympic Committee

our personal bests to everyone's progress



The world's largest sports event



our personal bests to everyone's progress

Atos | 
Worldwide IT Partner

eFOCUS

Prime technology integrator

Stronger integration for team victory

International Olympic Committee



our personal bests **to** everyone's progress



Prepáčte!

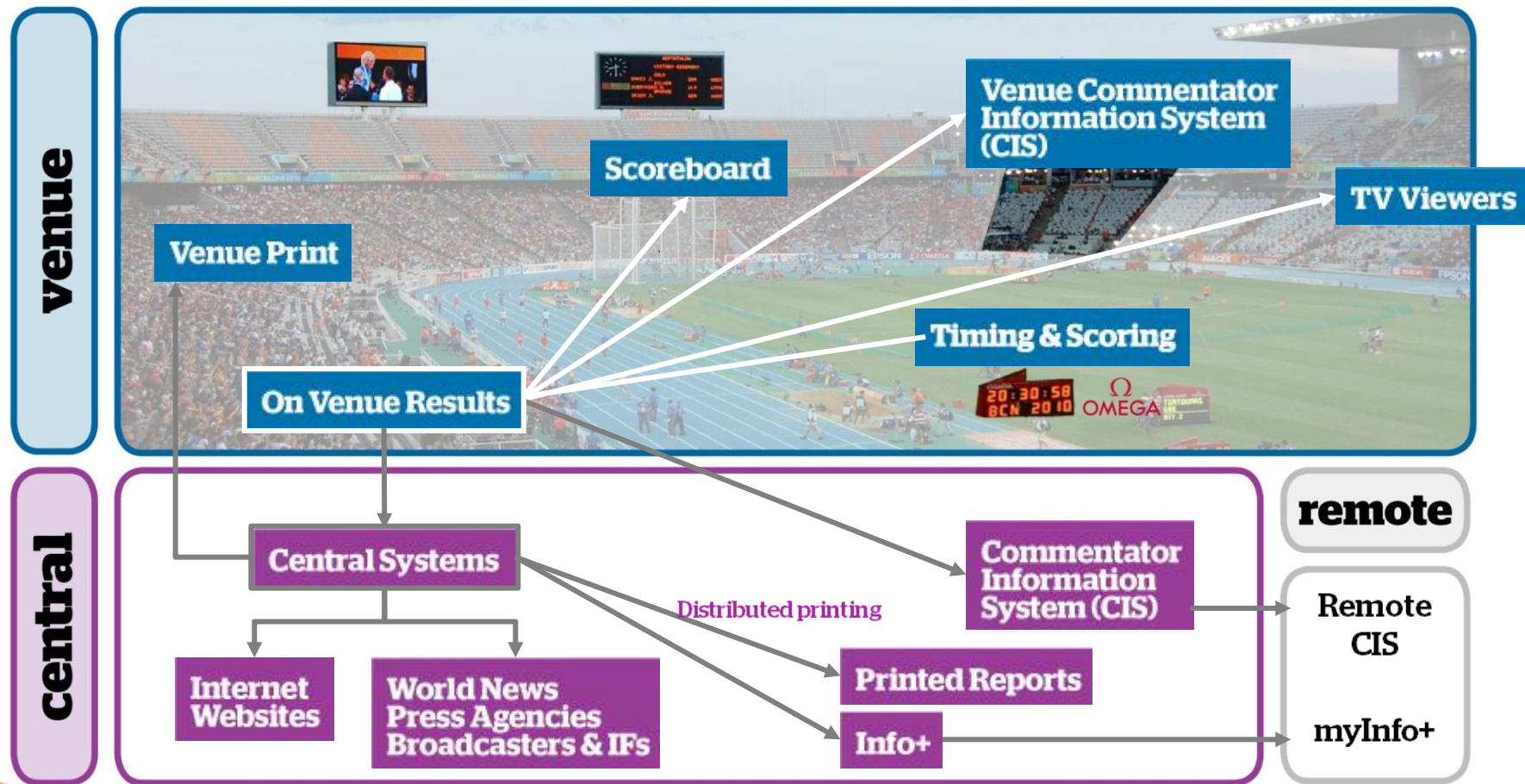
Môžete zopakovať tento
beh ?

Práve sme reboot-ovali

... ☹️

Behind the Scenes

Information Diffusion System (IDS)



Big Data examples in the Olympics

- Supporting 36 Olympic competitions and 21 Paralympic competitions
- 30% more competition data than ever before for media worldwide.
- Combining all results from all 36 Competition Venues in a single Data Feed.
 - 1.8M incoming messages, 53M distributed to customers worldwide.
- Providing results, schedule, weather, transport , biographies and other essential info from all 15,000 athletes to the Olympic Family via the Atos Info+ and myInfo+ systems.
- More than 20,000 pictures and biographies for Athletes and Officials.
- More than 200,000 pictures and records for all Accredited persons.





2020

A vision of social networks **now** and **then**

London 2012 will be the favorite, most liked and trended sporting event in history. But this is just the start. By the time of the 2020 Games, social networks will be integral to every aspect of our lives and an enormous generator of sporting data

2012

4 billion

estimated global audience for London 2012

57%

of the world's population

2020

5 billion

expected to rise to at least 5bn by 2020, as two thirds of the world's population gain access to mobile broadband

66%

of the world's population

2012

1 billion

people expected to share London 2012 content via social networks

2020

3000%

forecast increase in data traffic by 2020

2012



25% proportion of viewers who subscribe to a mobile broadband network

2020

100% estimated by 2020

hub.olympic.org

2012

1,000

athletes joining the Olympic Athletes Hub at hub.olympic.org

2020

10,500

all athletes will be on the hub by 2020

2012

1,300,000 likes

of London 2012 on Facebook

2020

5.2 million likes

estimated for the 2020 Games

2012

20,000 tweets/second

estimated peak during London 2012

2020

660,000 messages/second

via single social interface



2012

1.3 billion

total active users of Facebook, YouTube and Twitter

2020

4.3 billion

estimated total subscribers to a social interface by 2020

2012

700,000 followers

of London 2012 on Twitter

2020

2.8 million followers

estimate for the 2020 Games

2012



16 million Most followed Olympic athlete is Lebron James (Basketball)

2020

64 MILLION Who will be the most followed athlete?

2012

Official London 2012 pages exist on

Facebook
Twitter
YouTube
& Flickr

2020

There will be

One Games channel

that you will access through your social network

Growth in data traffic from 2012 to 2020

2012

30x

2020

Ďakujem za pozornosť

m.pastrnak@atos.net

Atos

eFOCUS