# Fog, Cloud Computing, Big Data

# Cloud Computing, Cloud Service

**Cloud computing** je služba, která nabízí na vyžádání přístup ke sdílenému fondu konfigurovatelných výpočetních zdrojů. Tyto služby mohou být zpřístupněny rychle s malým úsilím. Poskytovatelé cloudových služeb využívají datová centra pro jejich cloudové služby a cloudové zdroje. Aby byla zajištěna dostupnost datových služeb a zdrojů, poskytovatelé často udržují prostor v několika vzdálených datových centrech.

Cloud computing umožňuje:

* Umožňuje přístup k organizačním datům kdekoli a kdykoli.
* Zjednodušuje IT organizační operace tím, že se přihlásí pouze k potřebným službám.
* Eliminuje nebo snižuje potřebu místního IT vybavení, údržby a správy.
* Snižuje náklady na vybavení, energii, fyzické požadavky na provoz a potřeby školení personálu.
* Umožňuje rychlé reakce na rostoucí požadavky na objem dat.

Cloud computing umožňuje svým modelem „**pay-as-you-go**“ organizacím přistupovat k nákladům na výpočetní techniku a úložiště jako k nástroji než k infrastruktuře. To znamená, že nejsou potřené peníze pro vybudování infrastruktury. Tyto peníze mohou být rovnou využity ke zpracování dat.

**Amazon AWS** poskytuje cloudcomputingové služby. Protože je služba navržena tak, aby byla na vyžádání, uživatelé mohou začít s nízkými hardwarovými prostředky (paměť a CPU) a podle potřeby je rozšiřovat.

Cloud computing nabízí služby v následujících oblastech:

* **Infrastructure as a Service (IaaS)** – Hardware včetně serverů a dalších komponent infrastruktury je dodáván poskytovatelem a upravován na vyžádání. Poskytovatel zajišťuje údržbu systému, zálohování a plánování kontinuity.
* **Platform as a Service (PaaS)** – Poskytovatel poskytuje platformu, servery, úložiště a operační systémy uživatelům k vývoji a spouštění aplikací.
* **Mobile PaaS (mPaaS)** – Poskytovatelé dodávají vývojové funkce pro návrháře a vývojáře mobilních aplikací.
* **Software jako služba (SaaS)** – Software, jako je zasílání zpráv, zpracování dat IoT, zpracování mezd, hraní her a příprava daní, je licencován na základě předplatného a je hostován na cloudových serverech.

**Cloud Service**: Cloudové služby jsou služby nabízené poskytovateli cloudu, které jsou hostovány mimo prostory a jsou dostupné na vyžádání. Cloud zákazníci mají přístup ke sdílenému fondu konfigurovatelných výpočetních zdrojů, které lze rychle zřídit a uvolnit s minimálním úsilím správy.

Cloudové služby jsou skvělým nástrojem pro rozšíření funkčnosti IoT systému, díky vysoké dostupnosti. Zpracování a ukládání dat lze přesunout z IoT zařízení na cloud. Takto jsou data a zdroje jsou vždy dostupné pro jakékoli zařízení v systému, pokud má zařízení připojení k internetu. Poskytovatelé cloudových služeb d**bají na zabezpečení dat** zákazníků. IoT systémy, které vyžadují serverovou komunikaci, jako je cloudová analýza dat, by mohli těžit ze služeb cloud computingu.

Příklady cloudových služeb:

* **Amazon AWS** (cloud computing)
* **IFTTT** ('If This Then That') - umožňuje vytvářet speciální adresy URL zdrojů a mapovat je na konkrétní akce IFTTT.
* **Zapier** – podobně jako IFTTT umožňuj cloudovou automatizaci. Je zaměřen na podnikání a podporuje více aplikací a akcí.
* **Built.io** – pomáhá vývojářům vytvářet aplikace. Podobné jako IFTTT a Zapier, ale nabití pokročilé funkce.
* **Webex** Teams – zaměření na aplikace, poskytuje kompletní sadu pro spolupráci týmů (velice podobné jako Microsoft Teams)

# Fog Computing Model

**Fog computing** model identifikuje distribuovanou výpočetní infrastrukturu blíže k okraji sítě. Umožňuje okrajovým zařízením spouštět aplikace lokálně a přijímat okamžitá rozhodnutí. To snižuje datovou zátěž sítí, protože nezpracovaná data není nutné posílat přes síťová připojení. Zvyšuje odolnost tím, že umožňuje zařízením IoT fungovat, když se ztratí připojení k internetu. Zvyšuje také zabezpečení tím, že zabraňuje přenášení citlivých dat za hranice, kde jsou potřeba.

Fog computing rozšiřuje cloudové připojení blíže k okraji. Umožňuje koncovým zařízením, jako jsou inteligentní měřiče, průmyslové senzory, robotické stroje a další, připojit se k místnímu integrovanému výpočetnímu, síťovému a úložnému systému.

Fog computing zahrnuje kombinaci hardwarových a softwarových řešení. Některé platformy Fog computing podporují speciální operační systém nazvaný Cisco IOx. Tento operační systém v podstatě kombinuje Cisco IOS a open source Linux. To umožňuje IoT switchi provozovat IOS a aplikaci Fog na bázi Linuxu bez nutnosti interakce s cloudem.

Všechny fog aplikace monitorují nebo analyzují data v reálném čase ze zařízení připojených k síti. Na základě těchto dat provádějí akce jako je zamknutí dveří, přiblížení pomocí videokamery, nebo odeslání upozornění. Tato akce může zahrnovat M2M (machine-to-machine) komunikaci a M2P (machine-to-people) interakci.

Jako příklad fog computingu lze uvést inteligentní semafor. Tento semafor reaguje na data nasbírána senzory. Tato data jsou v reálném čase zpracována a semafor podle nich upraví provoz. Semafor se také řídí podle ostatních semaforů v okolí a poskytne tak plynulou dopravu. Data z klastrů systémů inteligentních semaforů jsou odesílána do cloudu k analýze dlouhodobých vzorců provozu.



# Big data

## Definice Big dat

**Big data** jsou větší a komplexnější datové sady především z nových zdrojů. Tyto datové sady jsou tak objemné, že tradiční software pro zpracování dat s nimi jednoduše nedokáže pracovat. Tyto obrovské objemy dat lze ale využít k řešení obchodních problémů, které byste předtím nebyli schopni vyřešit.

Do problémů velkých dat nepatří pouze **objemná data**. Do Big dat patří i data, která jsou potřeba zpracovat **v reálném čase** (data in motion).

Pro lepší rozpoznání Big dat se udávají čtyři parametry:

* **Objem (volume)** – Popisuje množství přenášených a ukládaných dat. Současným úkolem je objevit způsoby, jak co nejefektivněji zpracovat rostoucí množství dat.
* **Rychlost (velocity)** – popisuje rychlost, jakou jsou tato data generována. Například data vygenerovaná miliardou akcií prodaných na newyorské burze nelze jen tak uložit pro pozdější analýzu. Datová infrastruktura musí být schopna okamžitě reagovat na požadavky aplikací přistupujících a streamujících data.
* **Rozdílnost (variety)** – popisuje typ dat, která jsou jen zřídka ve stavu, kdy jsou dokonale připravena pro zpracování a analýzu. Velkým přispěvatelem do Big Data jsou nestrukturovaná data, která podle odhadů představují 70 až 90 %
* **Pravdivost (veracity)** – Jedná se o proces, který zabraňuje tomu, aby nepřesná data kazila vaše datové sady. Když si například lidé zaregistrují online účet, často použijí nepravdivé kontaktní údaje. Zvýšená pravdivost při shromažďování dat snižuje množství potřebného čištění dat.

## Zdroje Big dat

Zdrojem big dat může být inteligentní měření, správa zásob a sledování majetku, logistika. Dalším zdrojem big dat mohou být sociální sítě, které ukládají informace o uživatelích. Zdroje dat a i data lze rozdělit do tří druhů: sociální, mechanická a transakční.

Sociální big data jsou získána od uživatelů sociálních sítí. Tato data mapují chování a preference uživatelů. Tato data jsou nadále využívána k marketingovým účelům.

Mechanická data jsou generovaná zejména v průmyslu různými senzory. Tato data jsou například generovaná při monitorování dopravy. Do této kategorie patří také IoT sítě.

Transakční data jsou generována online i offline. Tato data jsou generována díky finančním transakcím, záznamech o zboží a objednávkami.

## Otevřená a privátní data

**Otevřená data** se dají popsat jako jakýkoli obsah, informace nebo data, které mohou lidé volně používat, znovu používat a distribuovat bez jakýchkoli právních, technologických nebo sociálních omezení. Tato data jsou úplná a snadno dostupná. Jedná se například o jízdní řády, příjmy států, rozpočty, databáze, seznam poskytovatelů sociálních služeb, kalendář ministra nebo měření čistoty ovzduší. Pocházejí z univerzit, nevládních organizací, soukromých firem nebo veřejné správy.

To co jsou **privátní data** nebo také soukromá data se pořád vyvíjí. Od uživatelů je stále požadováno více dat pro společnosti. Regulace těchto dat se po světě liší. Mezi tyto údaje patří jméno, pohlaví, věk a další. Pro využití těchto údajů se využívá openPDS. Namísto vyplnění osobních údajů se využívají odpovědi na určité otázky.

## Strukturovaná a nestrukturovaná data

Jedná se o další způsob rozdělení dat.

**Strukturovaná data** jsou zadávána a udržována v pevných polích v souboru nebo záznamu. Jsou snadno zadávána, klasifikována a dotazována počítačem. Struktura dat usnadňuje počítači zpracování dat a minimalizuje chyby. Příkladem strukturovaných dat jsou vyplňované adresační údaje při objednávkách.

Pokud je datová sada dostatečně malá jsou strukturovaná data zpravována pomocí SQL (Structured Query Language). SQL je programovací jazyk vytvoření pro procházení a vyhledávání dat v relačních databázích. SQL funguje pouze u strukturovaných datech.

**Nestrukturovaná data** postrádají organizaci. Nejsou tedy zpracovaná. Nemají pevné schéma, které identifikuje typ dat. Také nemají nastavená způsob zadávání nebo seskupování dat. Příklady nestrukturovaných dat zahrnují obsah fotografií, zvuku, videa, webových stránek, blogů, knih, časopisů, bílých knih, prezentací v PowerPointu, článků, e-mailů, wiki, dokumentů pro zpracování textu a textu obecně. Strukturovaná i nestrukturovaná data jsou cenná pro jednotlivce, organizace, průmyslová odvětví a vlády. Je důležité, aby organizace vzaly všechny formy dat a určily způsoby, jak tato data formátovat, aby je bylo možné spravovat a analyzovat.

## Data at Rest a Data in Motion

**Data at Rest** neboli data v klidu jsou klasická data uložena na fyzické lokaci například pevný disk nebo datové centrum. Tato data následují tradiční postup analýzy: Uložení > Analyzování > Upozornění > Akce. Data jsou uložena v databázi, až poté jsou analyzována. Následná rozhodnutí jsou provedena podle potřeby.

**Data v pohybu** vyžadují zpracování v reálném čase před tím, než jsou irelevantní nebo zastaralá. Analýza je provedena co nejdříve. Zařízení na okraji sítě okamžitě rozhodují na základě dynamicky analyzovaných dat. Proces analýzy je: Analýza > Akce > Upozornění > Uložení. Pořadí prvních třech akcí se může lišit. Hlavním rozdílem je uložení dat po analýze. Například při automatizovaném pěstování plodin je půda zavlažena při nedostatku vody. Až po zavlažení je odesláno upozornění o zavlažování a nakonec jsou data uložena.

Vzhledem k vlastnostem Big Data již není možné duplikovat a ukládat všechna tato data v centralizovaném datovém skladu. Nově vznikající implementace zařízení zahrnují velké množství senzorů zachycujících a zpracovávajících data. Rozhodnutí a akce se musí odehrávat na okraji, kde a kdy jsou data vytvářena. Díky tomu, že senzory získávají větší výpočetní výkon a stávají se více kontextovými, je nyní možné přiblížit inteligenci a analytické algoritmy ke zdroji dat. V tomto případě zůstávají data v pohybu tam, kde byla vytvořena, a prezentují náhledy v reálném čase, což vede k lepším a rychlejším rozhodnutím.

## Úložiště dat

Denně je vytvořeno 2.5 kvintilionů bajtu dat. Takové množství dat nemohou tradiční technologie datové sklady zachovat. I přes cloudová úložiště, která jsou k dispozici od společností jako Amazon, Google, Microsoft a mnoha dalších, se zabezpečení dat stává velkým problémem. Řešení pro big data musí být bezpečná, musí mít vysokou odolnost proti chybám, musí být schopna horizontálního škálování a musí používat replikaci, aby se data neztratila. Ukládání big dat není jen o jejich ukládání, ale také o jejich správě.

Při ukládání big dat narazíme na hlavně těchto 5 problémů:

* **Správa** – Existuje jen málo standardů pro sdílení dat a k dispozici jsou tisíce nástrojů pro správu dat.
* **Zabezpečení** – Ověření, přístup a účtování je obtížné zajistit.
* **Nestrukturovaná data** – nestrukturovaná data se obtížně analyzují a také se obtížně hledají.
* **Vstup a výstup** – Velké projekty, které produkují extrémně velké soubory datových řešení pro ukládání daní s množstvím I/O požadavků, které vydávají.
* **WAN** – Čím více se úložiště přesouvá do cloudu, tím větší tlak bude kladen na připojení WAN.

Data mohou být uložena na pevném disku, servetu nebo v datovém centru.

## Exponenciální růst dat a přesnost dat

V dnešní době exponenciálně dat přibývá. Je to následkem rozšíření technologii a internetu. Také aplikace sbírají čím dál více informací. V IoT systémech je také generované velké množství dat.

Větší přesnost dat lze ovlivnit mnoha způsoby. U IoT systémů to mohou být kvalitnější senzory. Z databází lze také odstranit nereálné hodnoty. U sociálních dat lze přesnost ovlivnit větším množstvím dat. Pro přesnost sociálních dat je nutné mít data od různých věkových i zájmových skupin. Z nasbíraných dat je také nutné odstranit nepotřebné údaje.

## IoT systémy v reálném světě

Příkladem využití iot v praxy může být například inteligentní semafor, který podle nasbíraných dat senzory upravuje provoz. Může tak předejít zácpám a zrychlit tak provoz.

Dalším příkladem mohou být chytré domácnosti. Tyto domácnosti mohou pomocí senzorů udržovat teplotu a také dům zabezpečit.

Mezi další příklady patří automatizované zemědělství, využití v logistice, monitorování sladů.

# Adafruit IO

ADAFRUIT\_IO\_KEY = 'aio\_XXXXX0123456789'

ADAFRUIT\_IO\_USERNAME = 'pavelnovak'

import time

import Adafruit\_DHT

from Adafruit\_IO import Client, Feed

DHT\_READ\_TIMEOUT = 60

DHT\_DATA\_PIN = 22

aio = Client(ADAFRUIT\_IO\_USERNAME, ADAFRUIT\_IO\_KEY)

temperature\_feed = aio.feeds('temperature')

humidity\_feed = aio.feeds('humidity')

dht11\_sensor = Adafruit\_DHT.DHT11

while True:

humidity, temperature = Adafruit\_DHT.read\_retry(dht11\_sensor, DHT\_DATA\_PIN)

if humidity is not None and temperature is not None:

print('Temp={0:0.1f}\*C Humidity={1:0.1f}%'.format(temperature, humidity))

temperature = '%.2f'%(temperature)

humidity = '%.2f'%(humidity)

aio.send(temperature\_feed.key, str(temperature))

aio.send(humidity\_feed.key, str(humidity))

else:

print('Failed to get DHT22 Reading, trying again in ', DHT\_READ\_TIMEOUT, 'seconds')

time.sleep(DHT\_READ\_TIMEOUT)