# I benefici dell'Edge Computing

## White paper n. 226

Revisione 0

di Steven Carlini


## Sintesi

Grazie all'IoT sta diventando sempre più diffuso l'uso di contenuti che richiedono un consumo intensivo della larghezza di banda, mentre cresce costantemente il numero di “oggetti” collegati. Allo stesso tempo si assiste alla convergenza delle reti di telecomunicazione mobile e delle reti dati in architetture di cloud computing. Per far fronte alle esigenze attuali e future, la potenza di calcolo e lo spazio di archiviazione vengono trasferiti alla periferia della rete per ridurre i tempi di trasmissione dei dati e incrementare la disponibilità. Grazie all'edge computing, i contenuti ad elevato consumo di larghezza di banda e le applicazioni sensibili sono più prossimi all'utente o all'origine dei dati.


APC Questi e altri white paper fanno parte della biblioteca Schneider Electric creata dal **Data Center Science Center di Schneider Electric**
DCSC@Schneider-Electric.com

Schneider Electric

# Definizione di edge computing

L'edge computing avvicina le funzioni di controllo e acquisizione dei dati, l'archiviazione di contenuti ad elevata larghezza di banda all'utente finale. Questa tecnologia è inserita in un endpoint logico di una rete (Internet o rete privata) nell'ambito di una più ampia architettura di cloud computing.

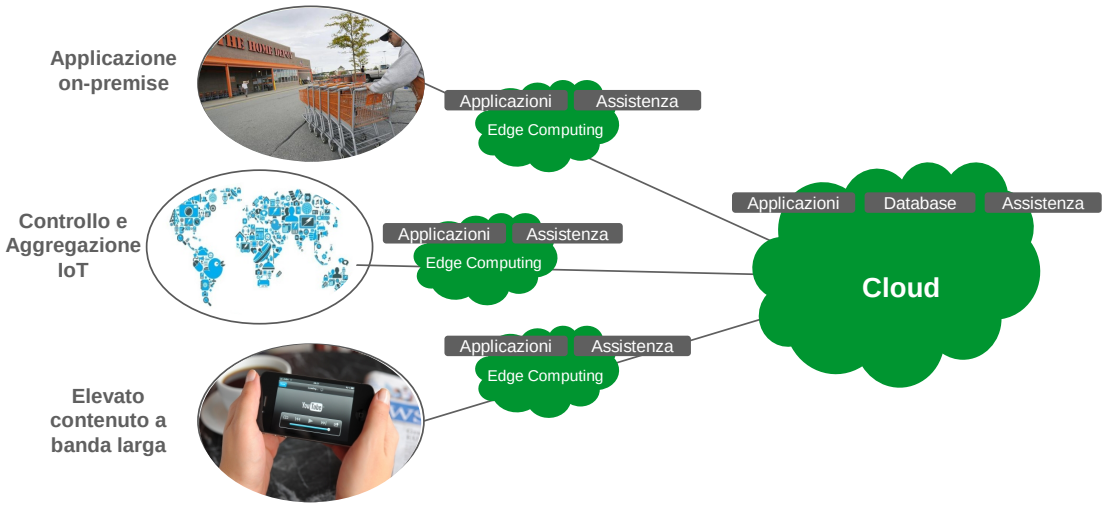



**Figura 1**

*Schema basilare del cloud computing con dispositivi periferici*

In questo white paper vengono discusse le tre applicazioni principali dell'edge computing.

1. Uno strumento per la raccolta di innumerevoli informazioni dagli "oggetti" locali come aggregazione e punto di controllo.
2. Un fornitore di archiviazione e fornitura locale di contenuti ad elevata larghezza di banda nell'ambito di una rete di distribuzione di contenuti.
3. Un'applicazione on-premise e uno strumento di elaborazione per la replica di servizi cloud e l'isolamento del Data Center dal cloud pubblico.

Prima di illustrare le applicazioni e le soluzioni, è opportuno tenere presente il funzionamento delle reti e di Internet.

# Come funziona Internet

## Trasmissione dei dati "da est a ovest"

Le origini dati vengono convertite in pacchetti che vengono trasmessi in rete tramite il protocollo di rete denominato IP (Internet Protocol). L'instradamento di Internet viene gestito da un altro protocollo denominato BGP (Border Gateway Protocol). Internet è stato progettato per rimanere attivo anche in caso di tempi di fermo prolungati e per aggirare i problemi. Il protocollo BGP non tiene conto della temporizzazione per l'instradamento dei dati, ma considera semplicemente il numero di hop tra due reti che tentano di comunicare. Gli hop possono essere realmente congestionati oppure il percorso fisico può essere lungo ma con meno hop anziché molto breve con più hop. La **Figura 2** contiene una mappa di molti hop a lunga distanza negli Stati Uniti.[1] Il protocollo BGP funziona efficacemente in termini di affidabilità ed è una tecnologia fondamentale per Internet, però, si dimostra meno valido dal punto di vista delle prestazioni in termini di latenza (ritardi, instabilità e congelamento delle immagini).

[1] http://conferences.sigcomm.org/sigcomm/2015/pdf/papers/p565.pdf

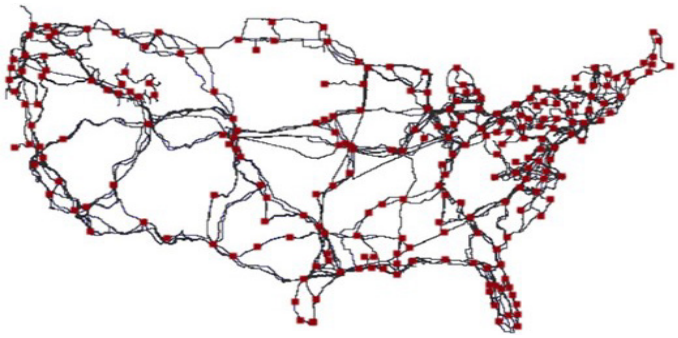


**Figura 2**

*Mappa dei vari hop di rete negli Stati Uniti*

### Trasmissione dei dati "da nord a sud"

Come illustrato nella **Figura 3**, dall'interno all'esterno della tipica rete di Data Center su cloud, il flusso di dati passa da un'interfaccia server fisica attraverso switch top-of-rack (ToR) o end-of-row (EoR). Da ogni switch ToR, i dati passano attraverso uno switch di aggregazione e gli switch di aggregazione provvedono all'instradamento dei dati tramite uno switch principale che rappresenta l'ingresso e l'uscita principale del Data Center. Ognuno di tali switch trasferisce i dati e viene considerato un hop della rete con il relativo rallentamento dati associato e la possibile congestione della rete. In caso di eccesso di sottoscrizioni in un livello di rete (in altre parole, se la larghezza di banda non è dimensionata per l'uscita di picco), possono verificarsi ulteriori rallentamenti durante i periodi di utilizzo intensivo.

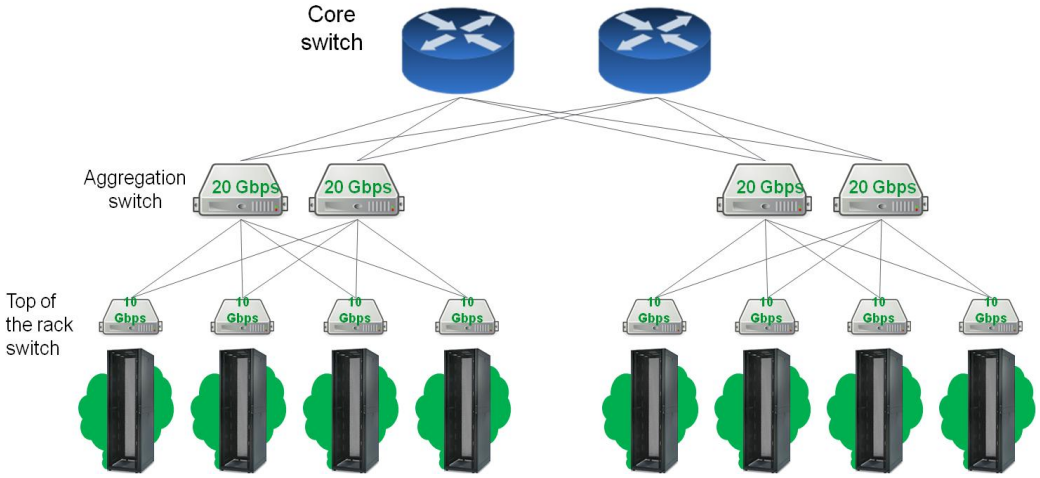



**Figura 3**

*Rete del Data Center*

## Applicazione n. 1: distribuzione di contenuti ad elevato consumo di larghezza di banda

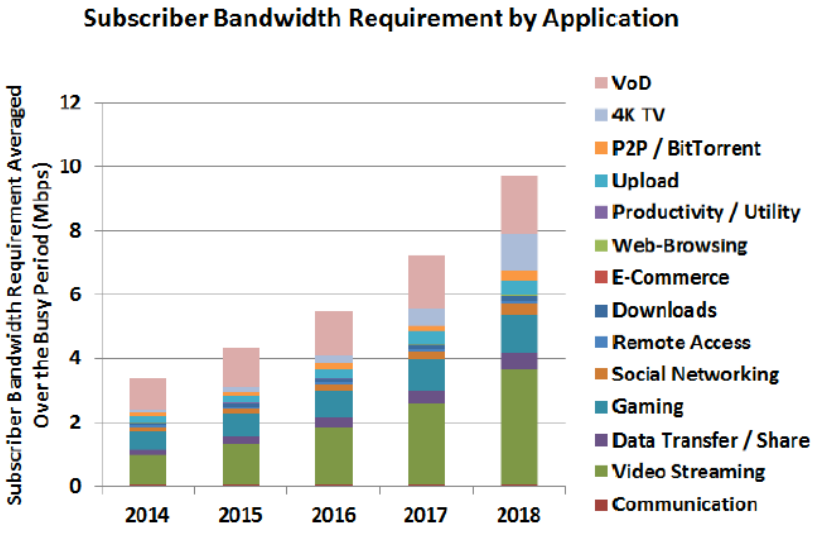
La latenza è il tempo che intercorre tra il momento in cui un pacchetto dati viene trasmesso e il momento in cui raggiunge la sua destinazione (solo andata) e torna indietro (andata e ritorno). Anche se gran parte dei dati viaggia in una sola direzione, questo tempo è quasi impossibile da misurare. Per questo motivo, il tempo di andata e ritorno da un singolo punto è il metodo più comune di misurazione della latenza. La tipica latenza di andata e ritorno è inferiore a 100 millisecondi (ms), mentre quella desiderata è inferiore a 25 ms.

La larghezza di banda si riferisce alla velocità di trasmissione dei dati nella rete. Le velocità massime delle apparecchiature di rete sono rese note dai rispettivi produttori, ma la velocità effettiva ottenuta in una determinata rete è quasi sempre inferiore al valore nominale di picco. Una latenza eccessiva crea intasamenti del traffico che impediscono ai dati di sfruttare la massima capacità della rete. L'impatto della latenza sulla larghezza di banda della rete può essere temporaneo (alcuni secondi), come un semaforo, o costante, come un ponte con una singola corsia. La massima probabilità di congestione della rete è correlata ai contenuti

video che richiedono un elevato consumo della larghezza di banda. Come mostrato nella **Figura 4**, VoD, TV a 4K e streaming video sono le applicazioni ad elevato consumo di larghezza di banda che stanno registrando la crescita più rapida.[2].

**Figura 4**

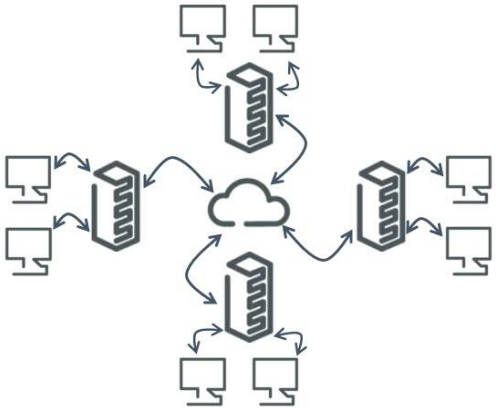
*Crescita delle applicazioni ad elevato consumo di larghezza di banda*




Per ridurre la congestione della rete e migliorare lo streaming dei contenuti ad elevato consumo della larghezza di banda oggi e in futuro, i fornitori di servizi stanno provvedendo all'interconnessione di un sistema di computer su Internet che funga da cache di contenuti più prossima all'utente. In tal modo, il contenuto può essere distribuito rapidamente a numerosi utenti duplicandolo su più server e indirizzandolo agli utenti in base alla prossimità. I computer che fungono da cache di contenuti rappresentano un esempio di edge computing (**Figura 5**).

**Figura 5**

*Schema di una semplice rete di distribuzione di contenuti (CDN)*



## Applicazione n. 2: Edge computing come aggrega-zione IoT e punto di controllo

Le tecnologie che in futuro consentiranno di rendere "intelligente" tutto ciò che ci circonda (città, agricoltura, automobili, sanità e cosi via) richiedono la distribuzione massiccia di sensori IoT (Internet of Things). Un sensore IoT è un nodo non costituito da un computer o un oggetto dotato di indirizzo IP per collegarsi a Internet.

Poiché il prezzo dei sensori diminuisce progressivamente, il numero di oggetti IoT connessi aumenta in maniera vertiginosa. Cisco stima che entro il 2020 l'IoT raggiungerà i 50 miliardi di dispositivi connessi a Internet[3]. L'IoT può automatizzare le operazioni in vari modi:

[2] ACG Research, The value of content at the edge, 2015, p. 4

[3] Dave Evans, The Internet of Things: How the Next Evolution of the Internet Is Changing Everything, Cisco Internet Business Solutions Group, p. 3

- Acquisizione automatica di informazioni sulle risorse fisiche (macchinari, apparecchiature, dispositivi, strutture, veicoli) per monitorarne lo stato o il funzionamento.
- Utilizzo di tali informazioni per fornire visibilità e controllo in modo da ottimizzare processi e risorse.

La tecnologia M2M (Machine to Machine) si riferisce alle tecnologie che consentono ai sistemi wireless o cablati di comunicare con altri dispositivi dello stesso tipo. La tecnologia M2M è considerata parte integrante dell'IoT e apporta numerosi vantaggi all'industria e alle aziende in generale, grazie a una vasta gamma di applicazioni nella Smart City.

L'IIoT (Industrial Internet of Things), che include l'utilizzo dei dati dei sensori, il controllo delle comunicazioni tra macchinari e le tecnologie di automazione, genera grandi quantità di dati e traffico di rete. I sistemi IT industriali proprietari e le tecnologie di rete stanno migrando ai sistemi IT commerciali principali che utilizzano le reti IP (Internet Protocol) per le comunicazioni.

L'esplorazione di giacimenti di petrolio e gas è un esempio di applicazione dell'IIoT. I vari droni volanti detti “bot di raccolta dati aerei” che esaminano i cantieri durante l'esplorazione alla ricerca del petrolio stanno generando grandi quantità di dati in forma di video ad alta definizione. Tali cantieri sono difficili da coordinare, vista la presenza di flotte di enormi carrelli, gru e scavatori rotanti. I metodi precedenti per la gestione del traffico utilizzavano elicotteri pilotati dall'uomo per la videosorveglianza. I droni autopilotati possono fotografare i cantieri 24 ore su 24, fornendo ai responsabili una visione aggiornata della distribuzione delle risorse. La tecnologia dell'edge computing consente ai droni di trasmettere i dati in tempo reale e ricevere istruzioni in maniera tempestiva.

**Figura 6**

*Esplorazione alla ricerca di gas e petrolio I droni raccolgono massicce quantità di dati sui giacimenti petroliferi e utilizzano l'edge computing per trasferire in tempo reale dati e istruzioni per gli spostamenti*

## Applicazione n. 3: Applicazioni on-premise

La necessità di preservare o migliorare la disponibilità delle risorse informatiche e delle relative reti è quasi sempre prioritaria. L'architettura del cloud computing è sempre stata centralizzata, ma con l'avvento dell'edge computing può diventare più distribuita. Il vantaggio principale è che qualunque tipo di interruzione è limitato a un solo punto della rete anziché all'intera rete. Un attacco DoS (Denial of Service) distribuito o un'interruzione dell'alimentazione di lunga durata, ad esempio, sarebbero limitati al dispositivo di edge computing e alle applicazioni locali installate su tale dispositivo e non a tutte le applicazioni in esecuzione in un Data Center centralizzato.

Le aziende che hanno provveduto alla migrazione al cloud computing off-premise possono sfruttare la maggiore ridondanza e disponibilità proprie dell'edge computing. Le applicazioni aziendali critiche o le applicazioni necessarie per lo svolgimento delle funzioni aziendali basilari possono essere duplicate in sede. Come analogia si può immaginare una piccola città che utilizza un'enorme fornitura idrica condivisa come fonte principale, come illustrato

nella **Figura 7**. Se questa fornitura idrica si interrompe a causa di un problema dell'impianto principale o della rete di distribuzione, è previsto un serbatoio di emergenza ubicato in città.

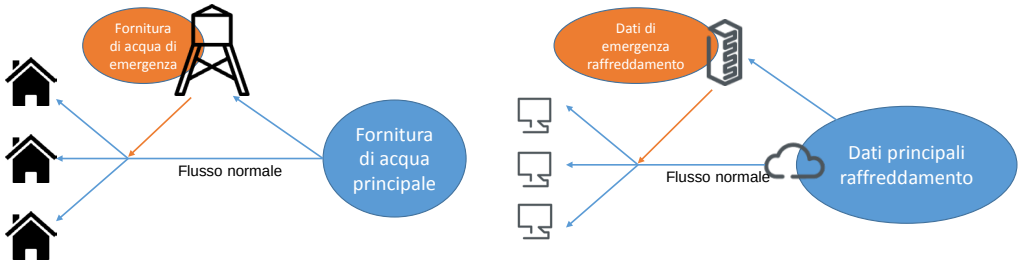



**Figura 7**

*Una rete idrica cittadina come metafora dell'edge computing.*

## Tipi di edge computing

In generale, esistono tre tipi di edge computing, come illustrato nella **Figura 8**.

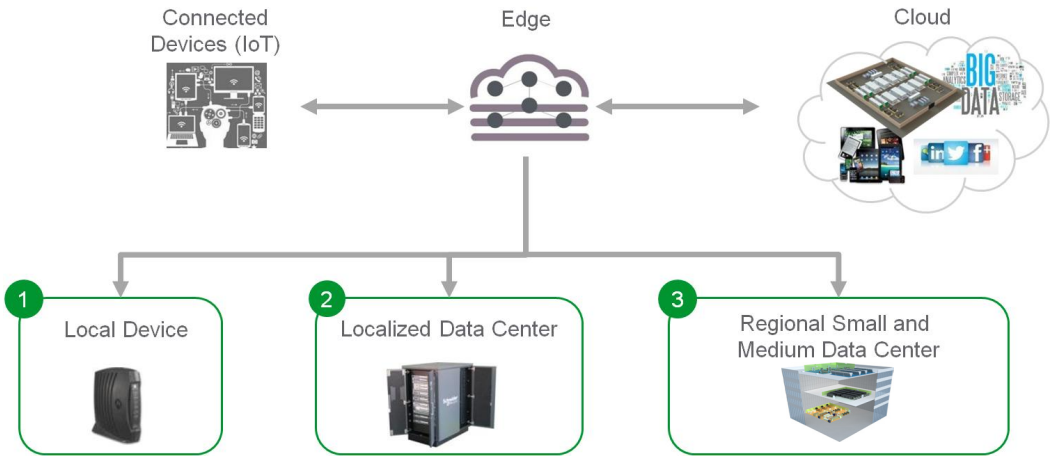



**Figura 8**

*Tipi di edge computing*

**Dispositivi locali:**
Dispositivi dimensionati per assolvere a uno scopo definito e specificato. La distribuzione è "immediata" e i dispositivi sono adatti ad abitazioni e piccoli uffici. Alcuni esempi sono la gestione di un sistema di sicurezza per l'edificio (applicazione Intel SOC) o l'archiviazione di contenuti video locali su un DVR. Un altro esempio è un gateway di archiviazione su cloud, ossia un dispositivo locale e, in genere, un'applicazione di rete o server che converte API di archiviazione su cloud, ad esempio SOAP o REST. I gateway di archiviazione su cloud consentono agli utenti di integrare l'archiviazione su cloud nelle applicazioni senza spostare le applicazioni sul cloud stesso.

**Data Center localizzati (1-10 rack):**
Questi Data Center forniscono funzionalità di elaborazione e archiviazione significative e possono essere distribuiti rapidamente in ambienti esistenti. Questi Data Center spesso sono disponibili come sistemi configurabili su ordinazione, pre-progettati e quindi assemblati in sede, come mostrato nella **Figura 9** (a sinistra). Un'altra forma di Data Center localizzato è costituita dai micro Data Center prefabbricati, che vengono assemblati in fabbrica e consegnati in sede, come mostrato nella **Figura 9** (a destra). Questi sistemi costituiti da un unico modulo possono essere installati in armadi particolarmente resistenti (impermeabili, anticorrosione, ignifughi, ecc.) oppure in normali armadi informatici per uffici. Le versioni a singolo rack possono utilizzare l'edificio, il raffreddamento e l'alimentazione esistente, consentendo di risparmiare sui costi di capitale necessari rispetto alla realizzazione di un sito dedicato. L'installazione richiede la scelta di una posizione vicina alla fonte di alimentazione dell'edificio e della fibra. Le versioni con più rack sono più funzionali e flessibili grazie alle maggiori dimensioni, ma richiedono tempi di pianificazione e installazione più lunghi e un

sistema di raffreddamento dedicato. Tali sistemi da 1-10 rack sono adatti a una vasta gamma di applicazioni che richiedono bassa latenza e/o elevata larghezza di banda e/o sicurezza o disponibilità supplementari.

**Figura 9**

*Un micro Data Center configurabile su ordinazione (a sinistra) e un micro Data Center prefabbricato (a destra)*


**Data Center regionali:**
I Data Center con oltre 10 rack che si trovano più vicini all'utente e all'origine dati rispetto ai Data Center su cloud centralizzati vengono definiti Data Center regionali. Grazie alle loro dimensioni, offrono maggiori funzionalità di elaborazione e memorizzazione rispetto ai Data Center localizzati da 1-10 rack. Anche se sono prefabbricati, richiedono tempi di realizzazione superiori rispetto ai Data Center localizzati, a causa di possibili problemi correlati alla realizzazione, alle autorizzazioni e alla conformità alle normative locali. Devono essere dotati, inoltre, di impianti di alimentazione e raffreddamento dedicati. La latenza dipende dalla prossimità fisica agli utenti e ai dati, nonché dal numero di hop tra origine e destinazione.

## Conclusioni

L'edge computing consente di superare i problemi correlati alla latenza, per cui le aziende possono sfruttare appieno le opportunità offerte da un'architettura di cloud computing. I carichi di lavoro generati dallo streaming video ad elevato consumo di larghezza di banda stanno causando latenze e congestioni della rete. I Data Center edge avvicinano il contenuto ad elevato utilizzo di larghezza di banda all'utente finale e le applicazioni sensibili alla latenza ai dati. La potenza di calcolo e le funzionalità di archiviazione vengono inserite direttamente nella periferia della rete in modo da ridurre i tempi di trasporto e incrementare la disponibilità. Tipi di edge computing includono dispositivi locali, Data Center localizzati e Data Center regionali. Quelli che forniscono velocità di distribuzione e capacità allineata alla domanda futura di applicazioni IoT sono i Data Center localizzati da 1-10 rack. Questi Data Center possono essere progettati e distribuiti in maniera semplice e rapida con varianti prefabbricate o configurabili su ordinazione.

## Note sull'autore

**Steven Carlini** è responsabile marketing per le soluzioni per Data Center di Schneider Electric. Nella sua carriera ha contribuito ad alcune delle soluzioni più innovative che hanno trasformato il panorama e l'architettura dei Data Center. Ha conseguito un diploma di laurea BSEE dell'Università dell'Oklahoma e un MBA in International Business dell'Università di Houston. La sua esperienza sul campo è ampiamente riconosciuta. Partecipa come relatore a numerosi eventi riguardanti il settore dei Data Center.

# Risorse

**Vantaggi economici dei micro Data Center a singolo rack**
White paper n. 223

**Scelte pratiche per la realizzazione di piccole sale server e micro Data Center**
White paper n. 174

**Sfoglia tutti i white paper**
whitepapers.apc.com

**Sfoglia tutti i TradeOff Tools™**
tools.apc.com

## Contatti

Per feedback e commenti relativi a questo white paper:

Data Center Science Center
dcsc@schneider-electric.com

Per formulare richieste specifiche sulla progettazione del Data Center:

Contattare Schneider Electric all'indirizzo
www.apc.com/support/contact/index.cfm