

Le criptovalute sono davvero l'unica causa dell'aumento di prezzo delle GPU?

Principalmente, la colpa della temporanea sparizione delle schede grafiche è stata data alle **criptovalute**. Un altro bersaglio sono stati i produttori di **GPU**, **Nvidia** e **AMD**, accusati di non produrne abbastanza e i partner **Asus**, **EVGA**, **Gigabyte**, **MSI**, per averne aumentato i prezzi.

In realtà i maggiori responsabili di tutto questo sono in particolare i produttori di **DRAM**. Nel 2015 e nel 2016, le DRAM erano facilmente reperibili e costavano pochissimo, quindi investire in fonderie aggiuntive per produrre ancora più DRAM per soddisfare le esigenze di un mercato già saturo non sembrava una buona idea.

In sostanza, i prezzi e la richiesta di DRAM stavano precipitando. Nel frattempo, stava aumentando la domanda delle **NAND**, le quali sono spesso fabbricate nelle stesse fonderie delle DRAM. La produzione richiede tempo e può costare molto, il che significa che i piani vengono messi in atto mesi o addirittura anni prima, prevedendo di vendere più NAND e meno DRAM (e di passare alla NAND 3D, ma questo è un'altra storia).

Gli smartphone hanno iniziato a utilizzare più DRAM e NAND: AMD ha lanciato **Ryzen**, e le battaglie CPU del 2017 hanno spinto gli utenti ad aggiornare sempre più spesso le componenti del PC. Le automobili sono diventate un mercato sempre più proficuo per i circuiti integrati DRAM: la maggior parte delle auto moderne ha da **4 a 8 GB** di DRAM e nei modelli con funzionalità avanzate come la tecnologia **Lane Assist** e **Self-Driving** la quantità di DRAM usata può quadruplicare. Inoltre sono stati prodotti milioni di piccoli dispositivi **IoT**, ciascuno con una piccola parte di DRAM.



Ma il punto focale di tutto ciò, è che le **schede grafiche** hanno bisogno di molta DRAM. Con più domanda che offerta, i prezzi potevano solo salire; anche senza i minatori, i prezzi delle schede grafiche sarebbero aumentati comunque verso la fine del 2017 e l'inizio del 2018: i kit da **16GB** di memoria **DDR4-2400** e **DDR4-2666** costavano tra i **40** e i **50€** nella metà del 2016; quegli stessi kit oggi vengono venduti per **140€** o più. Gli stick DDR4, vengono fabbricati nella stessa struttura dei GDDR5, GDDR5X e HBM2, e devono lottare contro il tempo sulla linea di produzione.

Che cosa significa questo per le schede grafiche? Il prezzo da contratto per il set GDDR5 è stato tra i **30** e i **40€** per **8GB** quando sono state lanciate la **GeForce 1080** e **1070** di Nvidia, insieme alla **Radeon RX 480/470** di AMD. Oggi, i prezzi del contratto per lo stesso set GDDR5 da 8GB sono intorno agli **80€**. Ma ogni livello della catena di approvvigionamento vuole la sua parte, quindi se il costo base aumenta di 40€ su una scheda grafica, complessivamente il prezzo di vendita complessivo aumenterà di 80€.

Se già per i set GDDR5 è un momento difficile, è ancora peggiore per gli HBM2; 8GB di HBM2 possono arrivare a costare 140€ e usare questo sistema, è già comunque più costoso a causa della

necessità di un **interposer al silicio**. Sulla base di queste informazioni, si può intuire che **Vega 56** e **Vega 64** non torneranno mai agli obiettivi iniziali di vendita del **MSRP** ammontanti a 330€ e 400€, il che fa sì che nonostante possano competere con la **GTX 1070/1070 Ti/1080** per quanto riguarda le prestazioni, diventano una scelta sconsigliata con un prezzo del **25** o **50%** in più.



Nvidia e AMD non hanno ufficialmente alzato i prezzi sulle loro schede grafiche pre-costruite, ma con i modelli **Founders Edition** per lo più di serie (tranne il 1080 Ti), vale la pena notare che le carte **FE** in genere costano 40€ in più rispetto al MSRP di base. I modelli di Founders Edition aiutano a eliminare almeno un livello nella catena di approvvigionamento. In via ufficiale, sembra che i prezzi dei contratti per i produttori di schede grafiche siano aumentati, il che in parte giustifica il costo più elevato della DRAM. Anche ora, le GPU “budget friendly” che originariamente costavano tra i **90** e i **110€** sono vendute tra i **130** e i **160€**

I produttori di DRAM (Samsung, SK-Hynix, Micron) stanno aumentando la loro produzione di DRAM e costruendo nuove strutture a causa della grande richiesta. Ma anche con l’aumento della produzione dovrebbe passare un po’ di tempo prima che i prezzi delle DRAM si avvicinino a quelli del 2016.

Cosa succederà nel 2018? AMD sembra essere in silenzio stampa e, a parte una riduzione di 7nm della Vega destinata alle applicazioni di *machine learning*, non ci si aspetta di vedere alcuna nuova scheda grafica principale lanciata quest’anno.

Per Nvidia la storia è diversa, con la **GTX 2080/2070** o **GTX 1180/1170** molto diffusa che verrà lanciata in agosto o settembre e, se si considerano le architetture di **Turing**, **Ampere** e **Volta**, sembra che le nuove GPU potrebbero risultare simili al Volta **GV100**, tranne che per il **Tensor Cores** e il supporto **FP64**.

La ragione per l’aumento dei prezzi relativo alle parti della serie 10 è la stessa, i costi della DRAM più elevati, combinati con l’aumento della domanda da parte di più settori (giochi, automotive, IA e crittografia).

NVIDIA e i rumors sulla nuova generazione di GPU “Ampere”

Sul sito tedesco Heise.de è apparso quello che sembra essere il nome della prossima generazione di **GPU** di **Nvidia**. Pare infatti che l'architettura **GPU** che prenderà il posto della serie **Pascal** e che Nvidia intende presentare alla GTC 2018 per il settore gaming si chiamerà “**Ampere**”. Attualmente non sono disponibili dettagli, ma si vocifera che Nvidia salterà direttamente da Pascal ad Ampere, almeno per quanto riguarda la linea **GeForce**. Il sito afferma che la famiglia di schede grafiche **Ampere** succederà quella delle GeForce serie 10 basate sull'architettura Pascal.

Nvidia che nel **Q3 FY18**, ha registrato un fatturato record di **2,64 miliardi di dollari**, ha visto aumentare entrambi i mercati, Datacenter e Gaming, in maniera straordinaria, con il mercato del gaming con una crescita pari al **59%** e **19%** in più per quello dei datacenter. Questi risultati ci fanno capire che le GPU basate su Pascal stanno ancora andando forte mentre le GPU **Volta** hanno a loro volta dato una spinta al mercato Datacenter con un incremento dei guadagni di quasi 100 miliardi di dollari rispetto al primo trimestre dell'anno fiscale.

Quest'ultime stanno mostrando il loro potere facendosi strada in ogni azienda operante nel mercato del cloud computing come ad esempio **Amazon** e **Baidu**. Nvidia ha anche altri prodotti di punta oltre alle V100 basate su Tesla sviluppate attorno alle GPU Volta, come le schede **Xavier SOC** e **Drive Pegasus**, che ricoprono soluzioni a scheda video dedicata o in versione SOC.

Anche se questo rumor sembra dirci il contrario, la roadmap Nvidia ci dà le GPU Volta in versione GeForce e Quadro in uscita per l'inizio dell'anno prossimo.