

# Power Week 2025

18-19-20 novembre 2025

IBM Innovation Studio Paris

#pw2025

## S45 - IBM Power11, entrez dans l'ère de l'IA souveraine

L'Architecture Unifiée pour l'Ère de l'IA Hybride et de la Donnée Souveraine

19 novembre 13h30-14h30

Salle Hedy Lamarr



François Guérin

**NOVAHE**  
lab.  
by constellation

Transformation Manager

[francois.guerin@constellation.fr](mailto:francois.guerin@constellation.fr)



**IBM**  
**common**  
FRANCE

# Retour vers le futur, nextGen



ina.fr

# La Vague de Complexité IT

Les défis stratégiques des entreprises:

- 目 Prolifération des données
- ⌚ Consommation énergétique
- 🔒 Souveraineté des données

Les 3 piliers SI pour y répondre

- 🚀 La Puissance de Calcul (Compute)
- 🔧 Les Archi. logicielles (Software & Tools)
- 🗃️ La Gestion des Données (Flux & Data)



L'**Open Source** exige performance et sécurité de niveau entreprise.

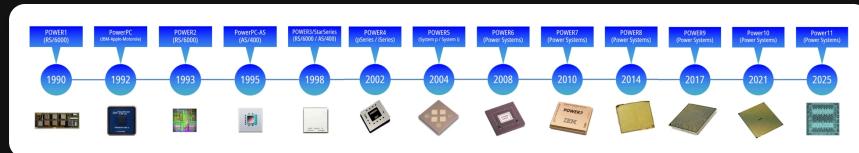


Le "**Cœur Métier**" doit s'interfacer avec l'IA générative.



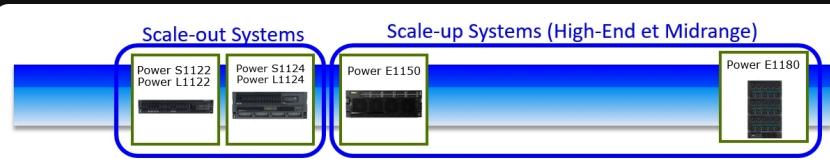
L'**IA** doit tourner en privé, sur ses propres données.

# IBM POWER



• 35 ans d'histoire.

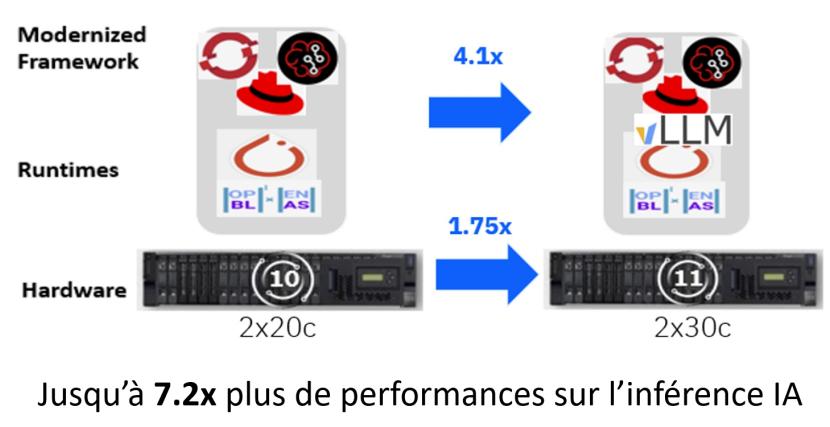
Depuis 1990, 11ième génération, toujours au cœur de l'innovation



## Performance (P10 vs P11)

+20% Augmentation du nombre de coeurs par chip

+15% Augmentation performance par coeur



## i comme intégration

- Beaucoup de coeurs disponibles intégrés aux serveurs, et souvent inutilisés.
- Linux disponible,
- Carte accélératrice AI disponible.

# Le Fondement : Vision Architecturale de Power11



2x

performance par watt que les systèmes x86

60%

d'énergie et d'émissions de CO<sub>2</sub> en moins que les POWER9

28%

meilleure efficacité du serveur avec le nouveau mode Energy Efficient par rapport au mode Maximum Performance

1. Performance is based on Quantitative Performance Index (QPI) data as of May 15, 2025 from IDC available at <https://www.idc.com/about/qpi>. IBM Power E1150 (4x30c Power11) QPI of 241,000E versus HPE Compute Scale-up Server 3200 (4x60-core Intel cores at 1.9GHz) QPI of 208,898 and utilizations of 75% for E1050 based on IBM Power Performance Utilization Guarantee and 40% for x86 based on IT Economics assessments of customer environments running a total of 13,800 x86 cores where average measured IBM peak utilization is 16%. Peak utilization of 16% is derived from a weighted average of x86 workloads in four large enterprise client IT environments using a total of the compared 13,861 x86 cores in production and test environments. For additional information on x86 workload analysis contact the IBM IT Economics team, IT.Economics@us.ibm.com

2. Energy consumption is based on maximum input power: IBM Power E1150 with maximum power of 5,200 W <https://www.redbooks.ibm.com/redpapers/pdfs/redp05624.pdf>; Superdome Flex 280 with maximum power of 4,740 W <https://www.hpe.com/en/renew/datasheets/280.pdf> ([www.hpe.com/en/renew/datasheets/280.pdf](https://www.hpe.com/en/renew/datasheets/280.pdf))

## Philosophie de Conception



### \*Performance Durable

Ratio performance/watt.



### \*Sécurité et Souveraineté

Contrôle total des données.



### \*Flexibilité Hybride

Intégration Red Hat OpenShift.



### \*La clé de voûte : PowerVM

Allocation dynamique.

# Pilier 1 (La puissance): IBM i sur Power11

## La Plateforme Critique, Prête pour l'IA

La robustesse et l'intégration légendaires d'IBM i.

Ce que Power11 apporte :



Piste d'envol  
Une fondation matérielle surpuissante.



Savoir faire métier  
Données et objets métiers fiables



Intégration transparente  
Appels IA à latence quasi nulle.

*"Votre forteresse de données s'ouvre sur le monde de l'IA."*



# Préambule : Lexique IA



## Modèle génératif (LLM/VLM/GAN/...)

Un système d'IA conçu pour créer du contenu nouveau (texte, images, sons, etc.) en apprenant les motifs et structures à partir de données existantes.

## Fine-tuning

Optimisation d'un modèle d'IA pré-entraîné pour qu'il intègre en profondeur les spécificités d'un domaine (ex: terminologie)

## Agent

Système autonome qui analyse l'environnement, décline une stratégie, et exécute des actions pour accomplir une tâche ou un objectif défini.

## Prompt

Consigne donnée à une IA générative pour guider sa réponse. On parle de prompt système (instructions) ou utilisateur (demande).

## RAG

Une technique qui combine une IA générative avec une recherche d'informations dans une base de données externe (ex : documents privés) pour contextualiser la réponse.

## Outils (pour les agents)

Modules connectés à un agent pour élargir ses capacités et lui permettre d'interagir avec différents environnements (ex : recherche d'informations, calculs avancés, workflows déterministes).

## RedHat OpenShift

Ce n'est pas "juste un Linux" ! Pensez d'abord "Conteneur" : c'est une "boîte magique" (un Tupperware) qui contient une app et tout ce dont elle a besoin pour fonctionner.

OpenShift, c'est le chef d'orchestre qui peut gérer des milliers de ces "boîtes" pour qu'elles travaillent ensemble sans se marcher dessus.

Power11 intègre une évolution majeure des accélérateurs IA dans chaque cœur.



\*Au-delà de l'inférence  
"Fine-tuning" de modèles sur la puce.

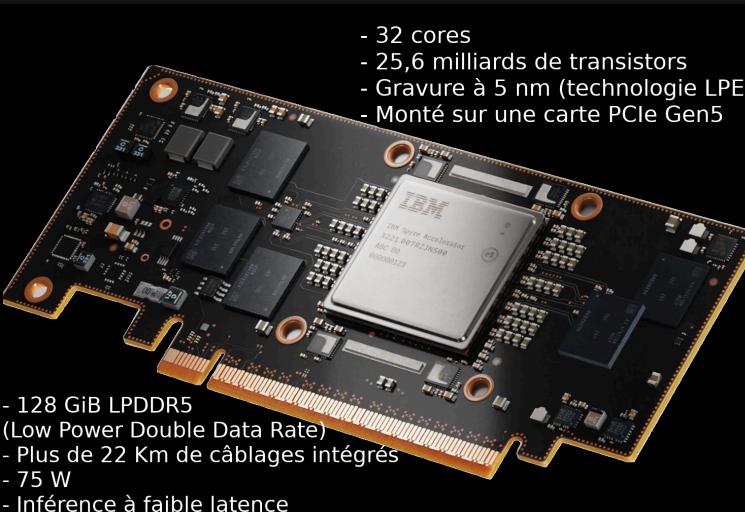


\*Optimisé pour les LLM  
Conçu pour l'IA générative.

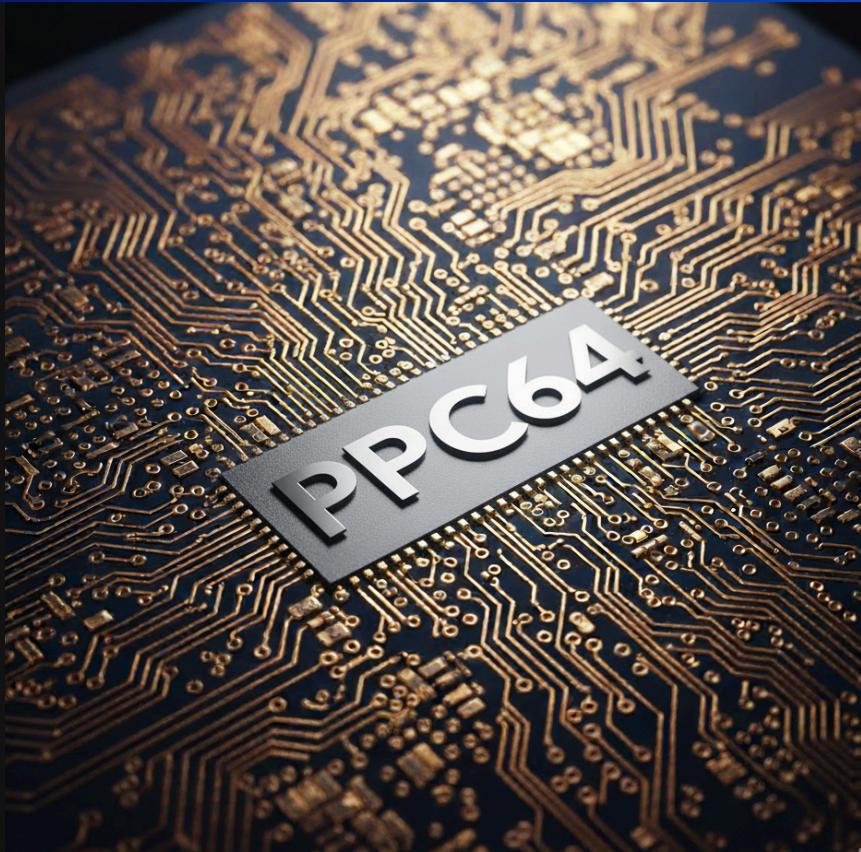


\*Le message clé : L'IA Souveraine.  
Latence minimale, sécurité maximale.

## Le "cerveau" de la plateforme



# Pilier 3 (gestion des flux data): Linux on Power11



## La Plateforme pour l'IA d'Entreprise

Linux sur Power11 est la plateforme de choix pour les workloads exigeants.



**\*\*Idéal pour l'IA Générative :\*\***

Bandes passantes mémoire et caches supérieurs pour les LLM.



**\*\*Consolidation Extrême :\*\***

Un TCO imbattable pour vos clusters OpenShift.

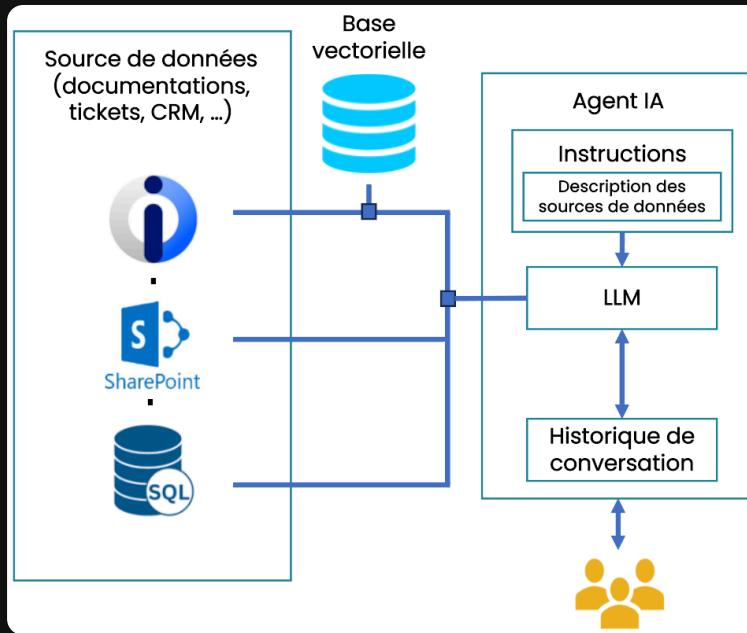


**\*\*Le meilleur des deux mondes :\*\***

L'écosystème que vos développeurs aiment, sur le matériel que vous méritez.

# La Synergie : L'IA au Service du Métier

## Scénario : Conseiller en assurance



*"Utilisation typique, Agent + RAG"*

1. Contexte (IBM i) : Dossier client (30 ans d'historique) sur LPAR IBM i.

2. Requête (API) : "Résume, identifie les risques et propose une offre."

3. Traitement (Linux) : Un LLM privé sur LPAR Red Hat OpenShift.

4. Accélération (Power11) : Calcul par les accélérateurs IA intégrés.

5. Résultat (IBM i) : Réponse complète argumentée en quelques secondes, y compris avec les nouveaux produits.

# Conclusion : Power11, une Plateforme Stratégique



## Souveraineté des Données et de l'IA

Reprenez le contrôle en exécutant l'IA la plus avancée au cœur de votre infrastructure.



## Efficacité Énergétique

Consolidez plus que jamais et réduisez votre empreinte carbone tout en augmentant votre capacité de calcul.



## Innovation sans Risque

Modernisez en connectant vos systèmes critiques à l'IA, sans migration complexe ni risque pour la sécurité.

## Remerciements



Fabrice Joubaire



Antoine Guillaume

# Questions & Réponses



François Guérin

Architecte logiciel -- Novahé



[francois.guerin@constellation.fr](mailto:francois.guerin@constellation.fr)

<https://www.novahe.fr>

Merci de votre attention.