

Ключевые технологии программного обеспечения для следующего поколения PRIMEPC - Post-FX10

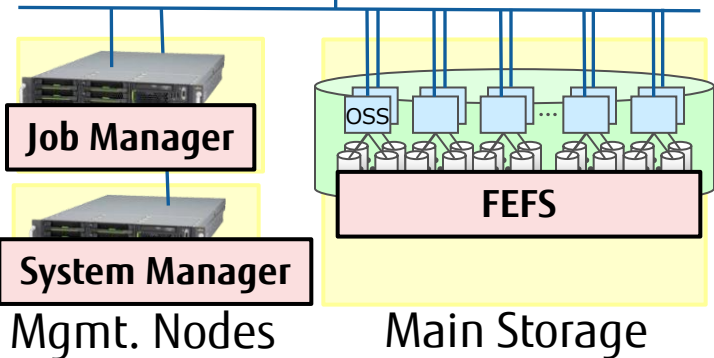
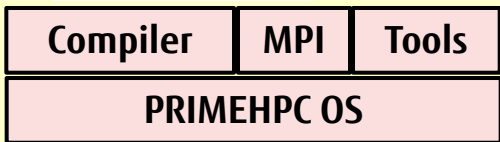
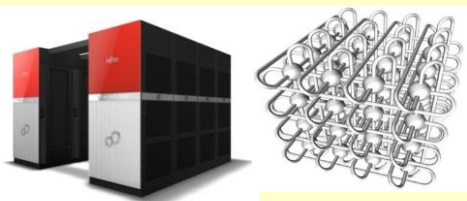
Ключевые вопросы программного обеспечения для последующих НРС



- Масштабируемость
- Управляемость
- Энергоэффективность
- Производительность

Технологии для масштабируемости

PRIMEHPC Series



Подоплека:

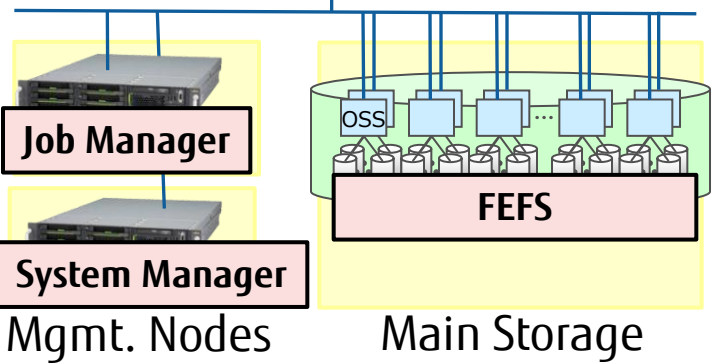
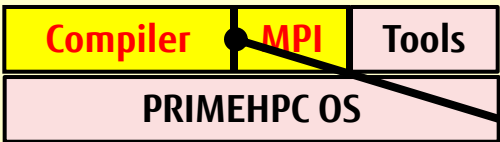
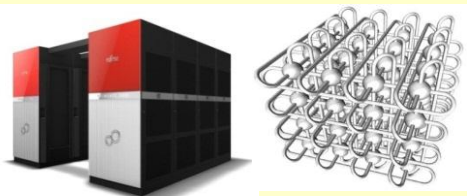
Производительность всегда является самым высшим приоритетом в HPC приложениях. Масштабируемость является наиболее важной для производительности при массивно параллельных вычислениях.

Вопросы в этой области:

- Улучшение распараллеливания приложений
- Снижение влияний неустойчивостей ОС
- Сохранение масштабируемости производительности ввода/вывода

Технологии для масштабируемости

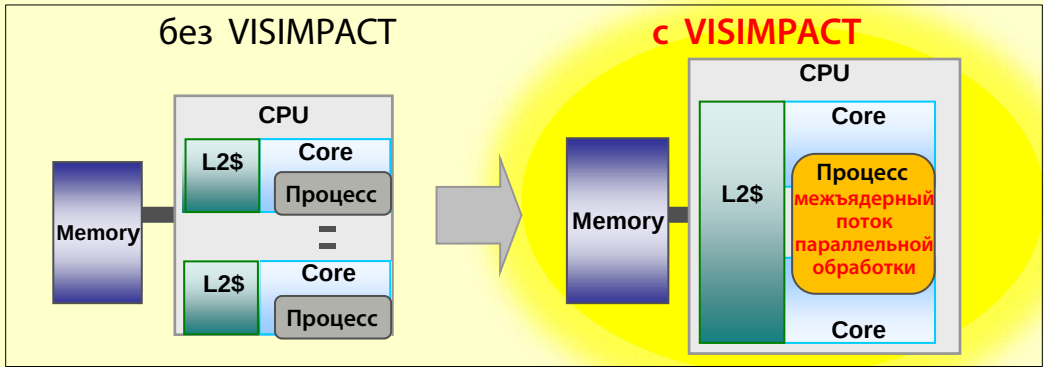
PRIMEHPC Series



■ Улучшение распараллеливания приложений

Технология: **VISIMPACT + Tofu Optimized MPI**

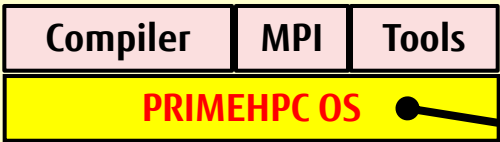
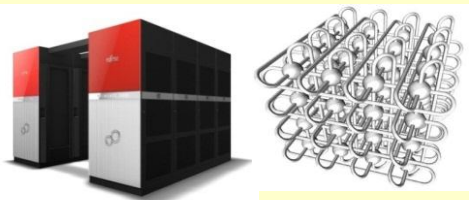
- Простота и эффективность межядерного распараллеливания, автоматизированная с помощью VISIMPACT



- Гибридное распараллеливание потоков-процессов с оптимизированными для Tofu MPI библиотеками

Технологии для масштабируемости

PRIMEHPC Series

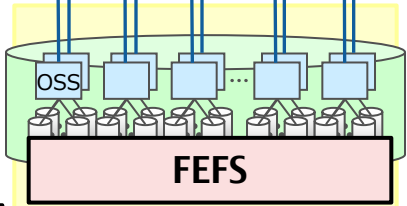


Job Manager



System Manager

Mgmt. Nodes



Main Storage

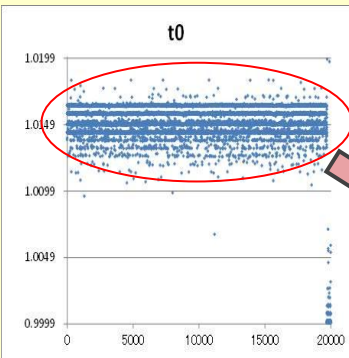
■ Снижение влияний неустойчивостей ОС

Технология: Настроенная ОС Linux для PRIMEHPC

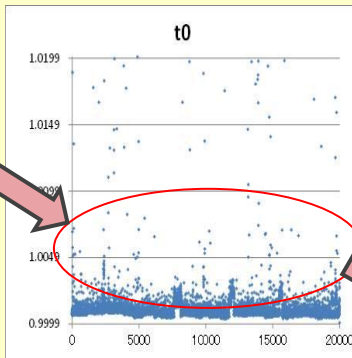
минимизированный негативный эффект от неустойчивости ОС в конечном счете достигается:

- Привязанные к ядрам ЦПУ технологии
- Сознательно выбираемые и настраиваемые системные службы

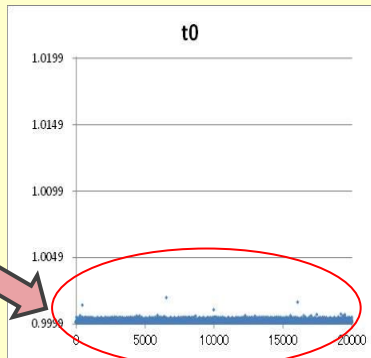
х86 кластер без TCS



х86 кластер с TCS



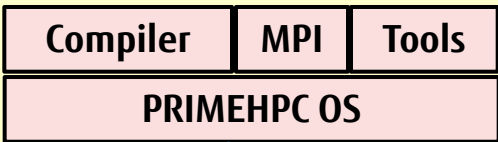
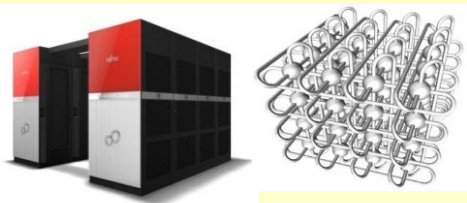
FX10 OS с TCS



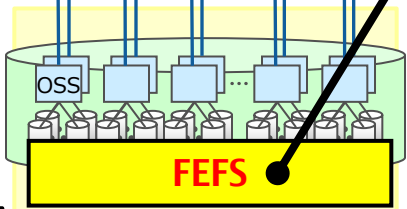
TCS: Technical Computing Suite (Пакет системного программного обеспечения Fujitsu)

Технологии для масштабируемости

PRIMEHPC Series



Mgmt. Nodes



Main Storage

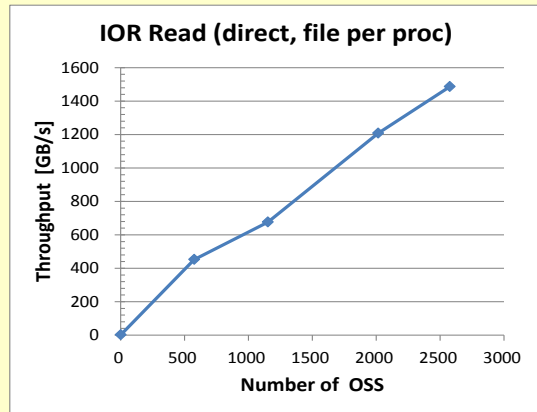
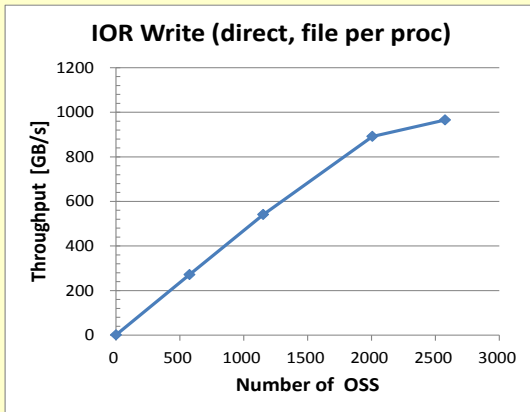
■ Сохранение масштабируемости производительности ввода/вывода

Технология: FEFS (Экзбайт Файловая система Fujitsu)

Масштабируемая файловая система на основе Lustre

- Поддерживает емкость до 8 Экзбайт
- Достигнута исключительная производительность на К компьютере

Запись : **965 GB/s** Чтение : **1,486 GB/s**



Collaborative work with RIKEN

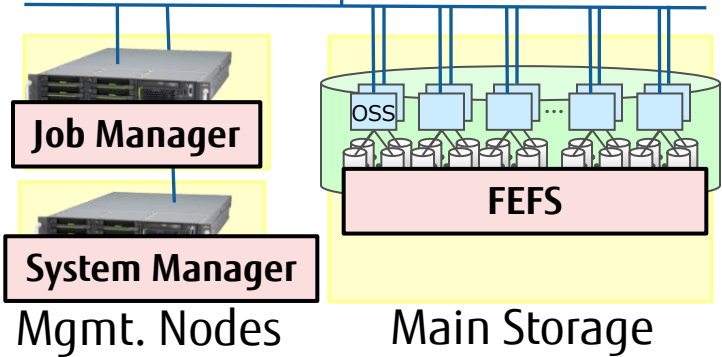
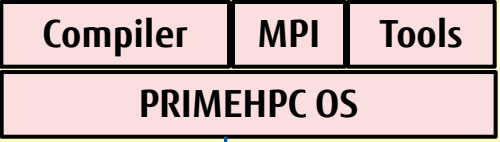
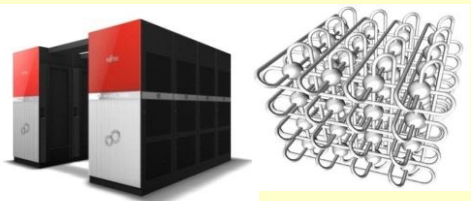
Ключевые вопросы программного обеспечения для последующих НРС



- Масштабируемость
- **Управляемость**
- Энергоэффективность
- Производительность

Технологии для управляемости

PRIMEHPC Series



Подоплека:

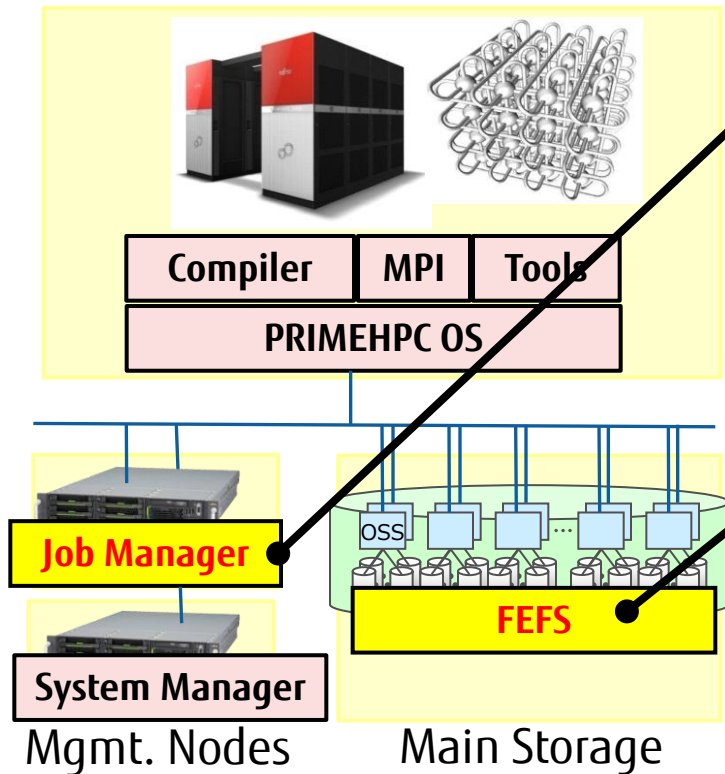
Для системного администратора ключевым вопросом является управление системой. Мы уже достигли системы с почти 100,000 узлов на К компьютере.

Вопросы в данной области:

- **Доступность**
- **Работоспособность**

Технологии для управляемости

PRIMEHPC Series



■ Доступность

Технология: Автоматическое управление восстановлением узлов после сбоев

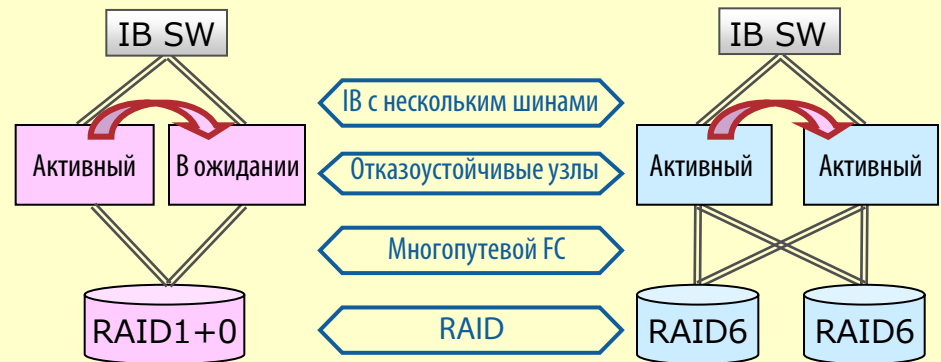
- Немедленное восстановление за счет оборудования в горячем ожидании

Технология: Автоматическое восстановление MDS/OSS после сбоев

- Мониторинг системой управления TCS

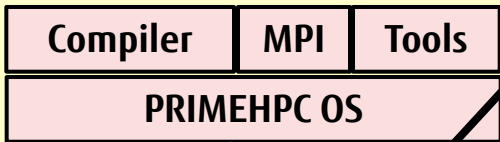
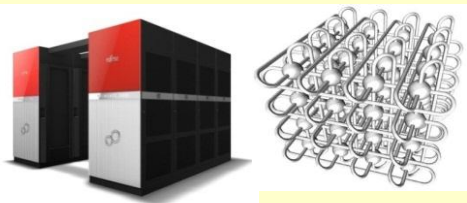
MDS: Активный/ В ожидании

OSS: Активный/ Активный

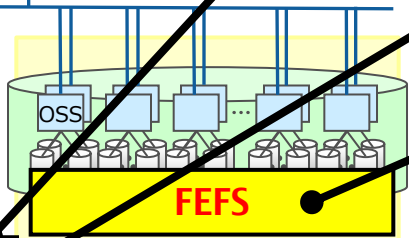


Технологии для управляемости

PRIMEHPC Series



System Manager



Main Storage

Mgmt. Nodes

■ Работоспособность

Технология: Центрическое управление производит единый системный образ для:

- Установки системы / обновления
- Мониторинга состояния узла (аппаратное/программное обеспечение)
- Контроль питания / мониторинг
- Поддержка PRIMEHPC / x86 гибридных кластерных систем

Технология: Гибкое управление кластером

обеспечивает различное разделение на физическом/логическом уровнях.

Технология: QoS/квотирование директорий в FEFS

упрощает совместное использование глобальной системы хранения поперек мультикластерной системы.

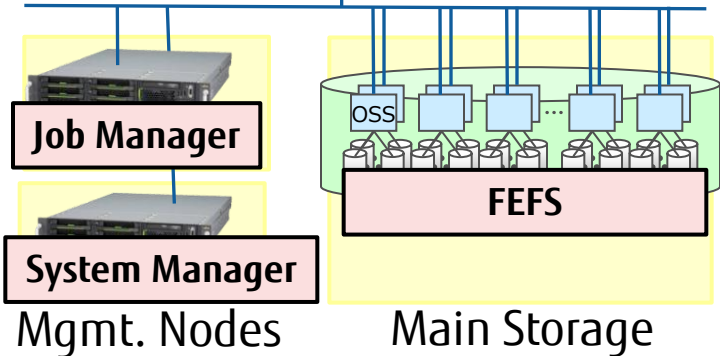
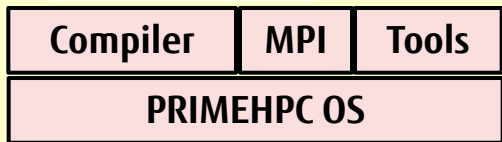
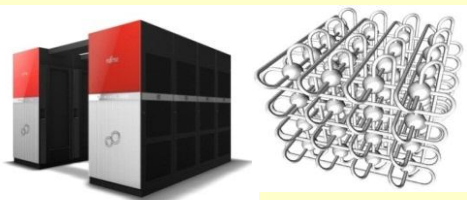
Ключевые вопросы программного обеспечения для последующих НРС



- Масштабируемость
- Управляемость
- Энергоэффективность
- Производительность

Технологии для энергоэффективности

PRIMEHPC Series



Подоплека:

$$\text{Энергоэффективность} = \frac{\text{“Раальная” производительность}}{\text{Общее энергопотребление}}$$

Требования пользователей:

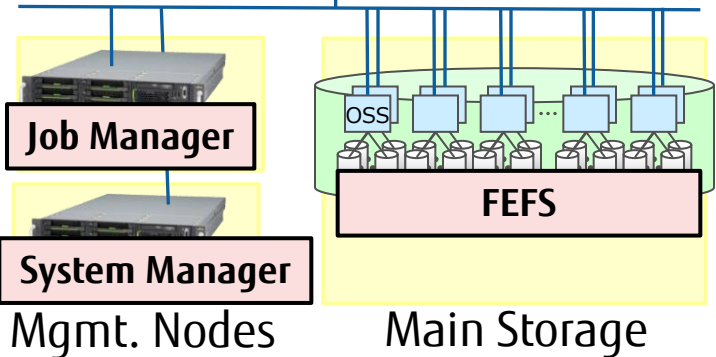
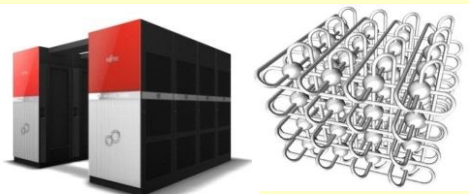
- Увеличение “фактической” производительности системы
- Сохранение общей мощности на целевых значениях

Вопросы в данной области:

- **Максимальная эффективность приложений**
- **Максимальное использование ресурсов**
- **Управление питанием всей системы**

Технологии для энергоэффективности

PRIMEHPC Series



■ Максимальная эффективность приложений

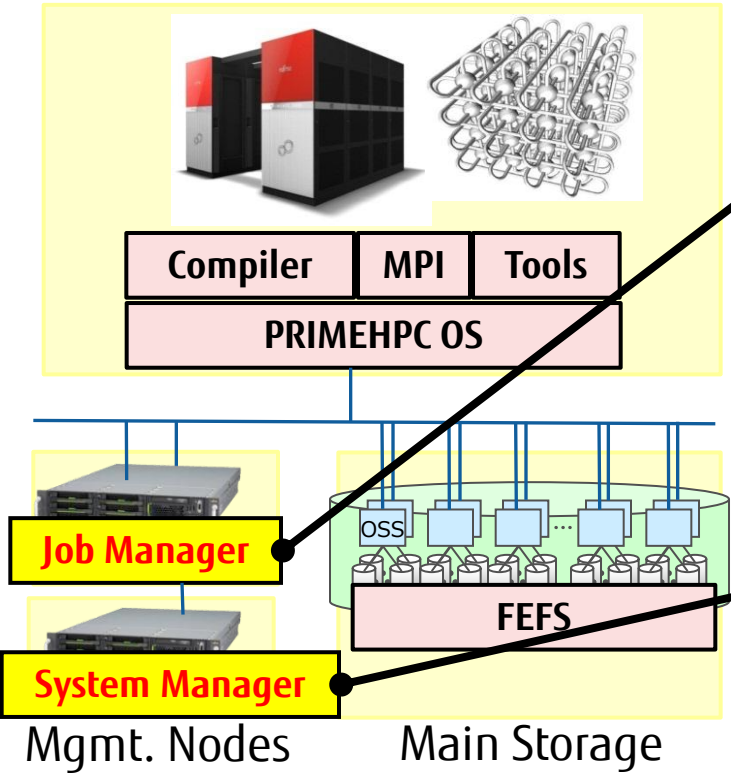
Технология: Оптимизированные ОС и языки программирования достигают хорошей эффективности для многих приложений

Application	Nodes	Efficiency
LINPACK	88,128	93%
NICAM	81,920	8%
Seism3D	82,944	18%
PHASE	82,944	20%
RSDFI	82,944	52%
FrontFlow/blue	80,000	3%
Lattice QCD	82,944	16%
ZZ-EFSI	82,944	46%

Таблица составлена др. Minami, RIKEN

Технологии для энергоэффективности

PRIMEHPC Series



■ Максимальное использование ресурсов

Технология: Различные методы распределения заданий для увеличения использования узлов/ядер даже на тороидальной системе

- Распределение режим тора/сетки
- Распределение узлов однонаправленное/разделяемое
- Гетерогенное гибридное распределение параллельных заданий

■ System-wide Power Management

Технология: Централизованное управление питанием помогает интегрировать вокруг центра покрытие питанием с:

- Интерфейс для управления питанием узлов или систем хранения
- Мониторинг потребления электроэнергии
- Управление режимом энергосбережения совместное с планировщиком заданий

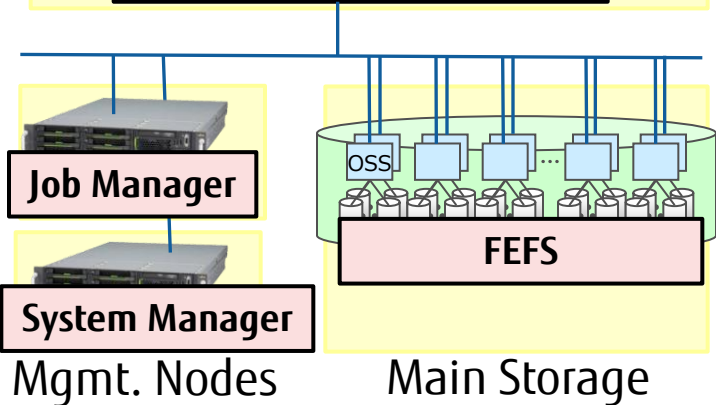
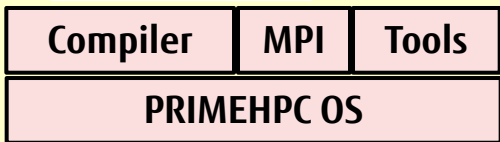
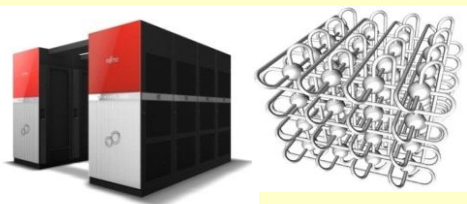
Ключевые вопросы программного обеспечения для последующих НРС



- Масштабируемость
- Управляемость
- Энергоэффективность
- **Производительность**

Технологии для производительности

PRIMEHPC Series



Технология:

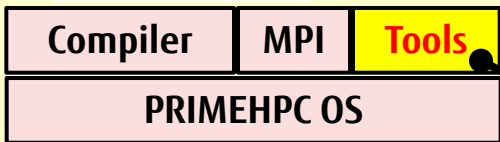
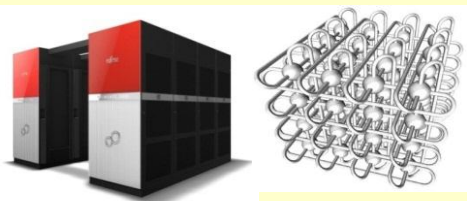
Из-за возрастающей сложности узла и сетевой архитектуры разработка приложений становится все более сложной.

Вопросы в данной области:

- **Настройка и отладка**
- **Переносимость**

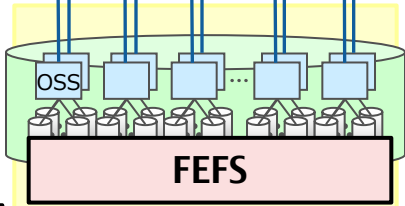
Технологии для производительности

PRIMEHPC Series



System Manager

Mgmt. Nodes

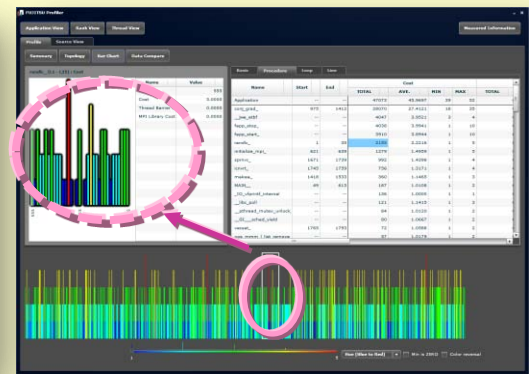
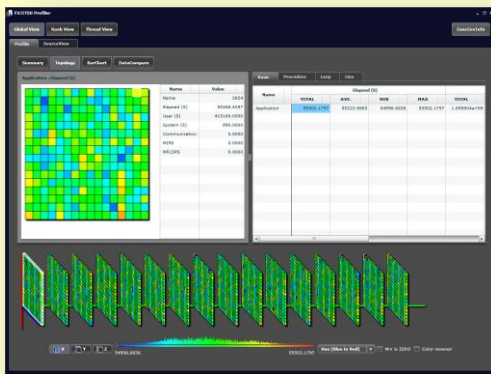


Main Storage

■ Настройка и отладка

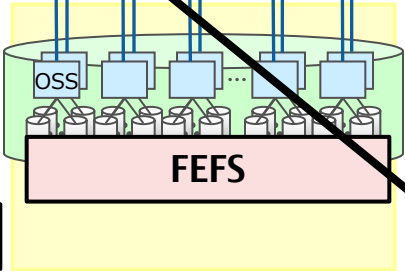
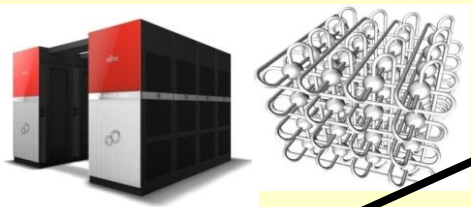
Технология:

- Поддержка стандартных отладчиков (DDT)
- Профилировщик
 - GUI детализованная информация PA
 - Оптимизация коммуникаций в Tofu
- Оптимизация ранжированных карт (RMATT, Rank Mapping Optimization)



Технологии для производительности

PRIMEHPC Series



Mgmt. Nodes

Main Storage

■ Переносимость

Технология: Совместимость К компьютером

- Двоичная совместимость
- Поддержка аналогичной архитектуры 1сру/узел, VISIMPACT, интерконнект Tofu

Технология: Продолжение поддержки:

- Последние международные стандарты Fortran 2008, C 11, C++ 11
- Стандарты де-факто расширения GNU C/C++, OpenMP 4.0, MPI 3.0

Технология: Поддержка типичных Linux OS

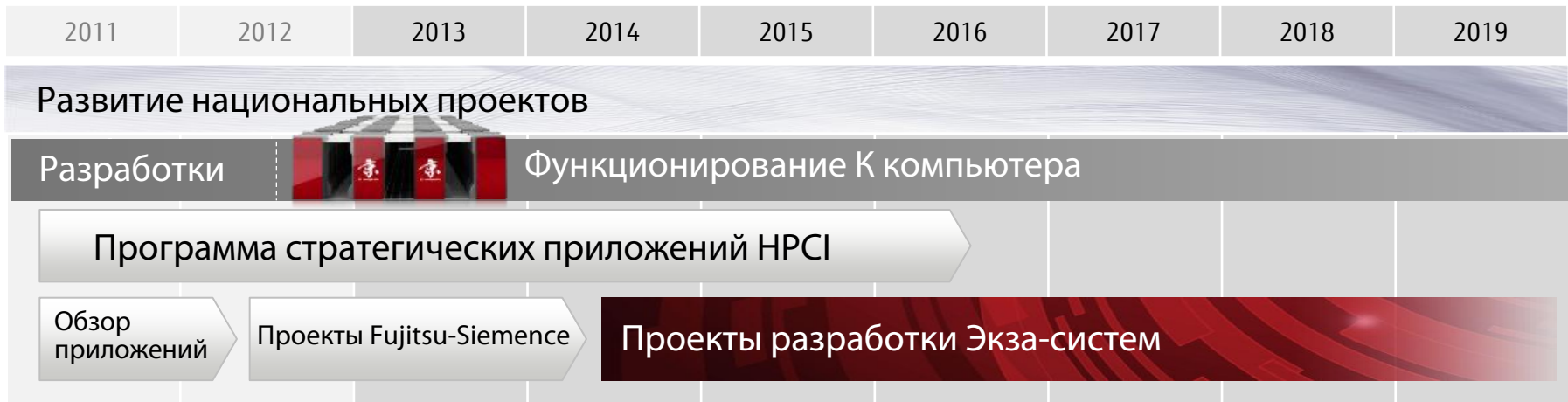
- POSIX -совместимые Linux и общие библиотеки

Сейчас мы готовы к 100PFlops!

Что дальше?

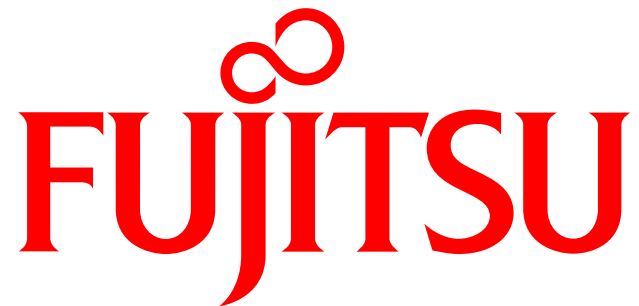
Активности для вычислений Экза-масштаба

■ Дорожная карта Японского национального проекта_а



■ Участие в совместном проектировании программного обеспечения для система Экза-масштаба

- Облегченное Микро-ядро близкое к Linux
- Повышение пропускной способности файлового ввода/вывода



формируем завтра вместе с вами