

# PureData for Analytics

*The Simple Appliance for Serious Analytics*

2012-12-05

한국IBM 소프트웨어그룹 정보관리사업부 이덕수 실장



# Agenda

- 01 IBM PureSystems Family 소개**
- 02 PureData for Analytics Overview**
- 03 PureData for Analytics Architecture**
- 04 PureData for Analytics What's New?**

# 01

# PureSystems

소개

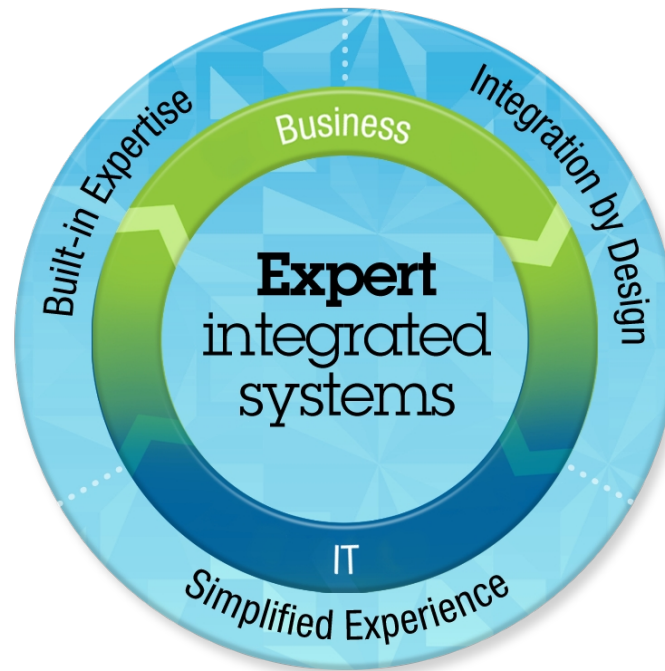
Built-in expertise ■ Integration by design ■ Simplified experience

# PureSystems

새로운 전문가통합시스템

## 내재화된 전문성

- 전문가가 업무를 캡처하고, 자동화하는 수준 보장 (인프라부터 애플리케이션까지)



## 통합에 최적화된 설계

- 하드웨어와 소프트웨어를 충분히 사전에 튜닝, 통합 (준비된 단일 시스템)

## 간편한 사용자 경험

- IT 라이프사이클의 모든 부분을 간편하게 개선
- 전체 시스템 통합 관리
- 최적화된 솔루션의 광범위한 개방형 에코시스템 제공

# IBM PureSystems Family

IBM의 시스템 관리 및 통합의 모든 노하우를 탑재한 전문가통합시스템

## Infrastructure



### PureFlex

인프라 서비스

Delivering Infrastructure Services

## Application Platform



### PureApplication

플랫폼 서비스

Delivering Platform Services

## Data Platform



### PureData

데이터 서비스

Delivering Data Services

# IBM PureData System

업무부하의 유형이 다르다면 시스템 역시 다르게 최적화되어야 합니다



**PureData**  
System for Transactions

전자상거래

고가용성, 확장성, 통합성을 보유,  
대용량 트랜잭션을 처리하는 데이터베이스 서비스

**PureData**  
System for Analytics  
*Netezza 기술 기반*

고객 분석

최소한의 관리로 petabytes 데이터를  
손쉽게 로드, 복잡한 분석 및 보고서 실행

**PureData**  
System for  
Operational Analytics

실시간 사기 감지

다양한 형태의 분석업무와 트랜잭션 처리가 혼재  
된 업무를 위한 다수의 사용자 환경 운영을 위한  
데이터웨어하우스 서비스

# 02

## PureData System for Analytics Overview

Built-in expertise ■ Integration by design ■ Simplified experience

# IBM PureData System for Analytics

분석 업무에 최적화 된 전용 DW 어플라이언스

## 내재화된 전문성

- 인덱스나 성능 튜닝 필요를 최소화
- In-Database 분석을 위한 병렬화/최적화
- 비대칭 초병렬 처리(AMPP)

## 디자인 단계부터 고려된 통합

- 서버, 스토리지, 데이터베이스를 하나로 통합
- 자동 병렬화 및 자원 최적화
- 엔터프라이즈급 보안성과 플랫폼 관리

## 간편한 사용자 경험

- 설치부터 데이터 로딩까지 수 시간 내 가능
- 유지관리 비용 최소화
- 표준 인터페이스 지원
- 다른 빅데이터 플랫폼과의 쉬운 연동 (Hadoop, BigInsights, Streams)

Powered by  
Netezza Technology

New





# Appliance Simplicity

## 인덱스 생성 및 튜닝 불필요

### ORACLE

```
CREATE TABLE "MRDWDDM"."RDWF_DDM_ROOMS_SOLD" ("ID_PROPERTY" NUMBER(5, 0) NOT NULL ENABLE, "ID_DATE_STAY" NUMBER(5, 0) NOT NULL ENABLE, "CD_ROOM_POOL" CHAR(4) NOT NULL ENABLE, "CD_RATE_PGM" CHAR(4) NOT NULL ENABLE, "CD_RATE_TYPE" CHAR(1) NOT NULL ENABLE, "CD_MARKET_SEGMENT" CHAR(2) NOT NULL ENABLE, "ID_CONFO_NUM_ORIG" NUMBER(9, 0) NOT NULL ENABLE, "ID_CONFO_NUM_CUR" NUMBER(9, 0) NOT NULL ENABLE, "ID_DATE_ARRIVAL" NUMBER(5, 0) NOT NULL ENABLE, "CU_REV" "CU_REV_PROJ_NET_QY_DAYS_STAY_CU" CHAR(1) NOT NULL STORAGE(FREELISTS PARTITION BY RANGE THAN (600) PCTFR STORAGE(INITIAL "DDM_ROOMS_SOLD LESS THAN (1200) STORAGE(INITIAL "DDM_ROOMS_SOLD LESS THAN (1800) STORAGE(INITIAL "DDM_ROOMS_SOLD LESS THAN (2400) STORAGE(INITIAL "DDM_ROOMS_SOLD LESS THAN (3000) STORAGE(INITIAL "DDM_ROOMS_SOLD LESS THAN (MAXVA STORAGE(INITIAL "DDM_ROOMS_SOLD
```

### ORACLE Indexes

```
CREATE INDEX "MRDWDDM"."RDWF_DDM_ROOMS_SOLD_IDX1" ON "RDWF_DDM_ROOMS_SOLD" ("ID_PROPERTY" , "ID_DATE_STAY" , "CD_ROOM_POOL" , "CD_RATE_PGM" , "CD_RATE_TYPE" , "CD_MARKET_SEGMENT" ) PCTFREE 10 INITRANS 6 MAXTRANS 255 STORAGE ( FREELISTS 10) TABLESPACE "DDM DATAMART INDEX L" NOLOGGING
```

### ORACLE Bitmap index

```
CREATE BITMAP INDEX "CRDDB"."SNAPSHOT_MONTH_IDX13" ON "SNAPSHOT_OPPTY_MONTH_HIST" ("SNAPSHOT_YEAR" ) PCTFREE 10 INITRANS 2 PCTFREE MAXTRANS 255 STORAGE(INITIAL 4194304 NEXT 4194304 MINEXTENTS 2 MAXEXTENTS 2147483645 PCTINCREASE 0 FREELISTS 1 FREELIST GROUPS 1 BUFFER_POOL 1 BUFFER PARTITI 4194304
```

### ORACLE Table Clusters

```
CREATE CLUSTER "MRDW"."CT_INTRMDRY_CAL" ("ID_YEAR_CAL" NUMBER(4, 0), "ID_MONTH_CAL" NUMBER(2, 0), "ID_PROPERTY" NUMBER(5, 0)) SIZE 16384 PCTFREE 10 PCTUSED 90 INITRANS 3 MAXTRANS 255 STORAGE(INITIAL 83886080 NEXT 41943040 MINEXTENTS 1 MAXEXTENTS 1017 PCTINCREASE 0 INITRANS: FREELISTS 4 FREELIST GROUPS 1 BUFFER_POOL RECYCLE) TABLESPACE "TSS_FACT" ;
```

### PureData for Analytics

```
CREATE TABLE MRDWDDM.RDWF_DDM_ROOMS_SOLD (
  ID_PROPERTY numeric(5, 0) NOT NULL ,
  ID_DATE_STAY integer NOT NULL ,
  CD_ROOM_POOL CHAR(4) NOT NULL ,
  CD_RATE_PGM CHAR(4) NOT NULL ,
  CD_RATE_TYPE CHAR(1) NOT NULL ,
  CD_MARKET_SEGMENT CHAR(2) NOT NULL ,
  ID_CONFO_NUM_ORIG integer NOT NULL ,
  ID_CONFO_NUM_CUR integer NOT NULL ,
  ID_DATE_CREATE integer NOT NULL ,
  ID_DATE_ARRIVAL integer NOT NULL ,
  ID_DATE_DEPART integer NOT NULL ,
  QY_ROOMS integer NOT NULL ,
  CU_REV_PROJ_NET_LOCAL numeric(21, 3) NOT NULL ,
  CU_REV_PROJ_NET_USD numeric(21, 3) NOT NULL ,
  QY_DAYS_STAY_CUR smallint NOT NULL ,
  CD_BOOK_SOURCE CHAR(1) NOT NULL)
```

istribute on random;

- No indexes
- No Physical Tuning/Admin
- Stripe data randomly, or by Columns

# 03

## PureData System for Analytics Architecture

Built-in expertise ■ Integration by design ■ Simplified experience

# Inside the IBM PureData System for Analytics

## *Optimized Hardware + Software*

- Hardware accelerated AMPP
- Purpose-built for high performance analytics

▪ Requires no tuning

## *SMP Hosts*

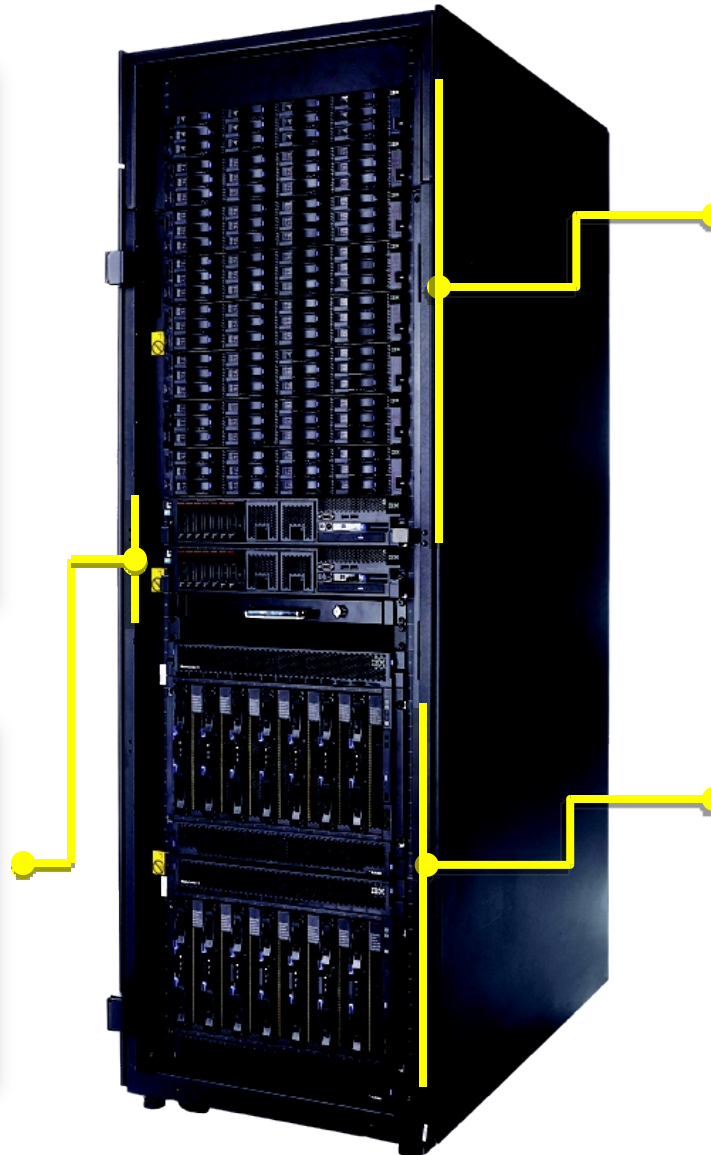
- SQL Compiler
- Query Plan
- Optimize
- Admin

## *Disk Enclosures*

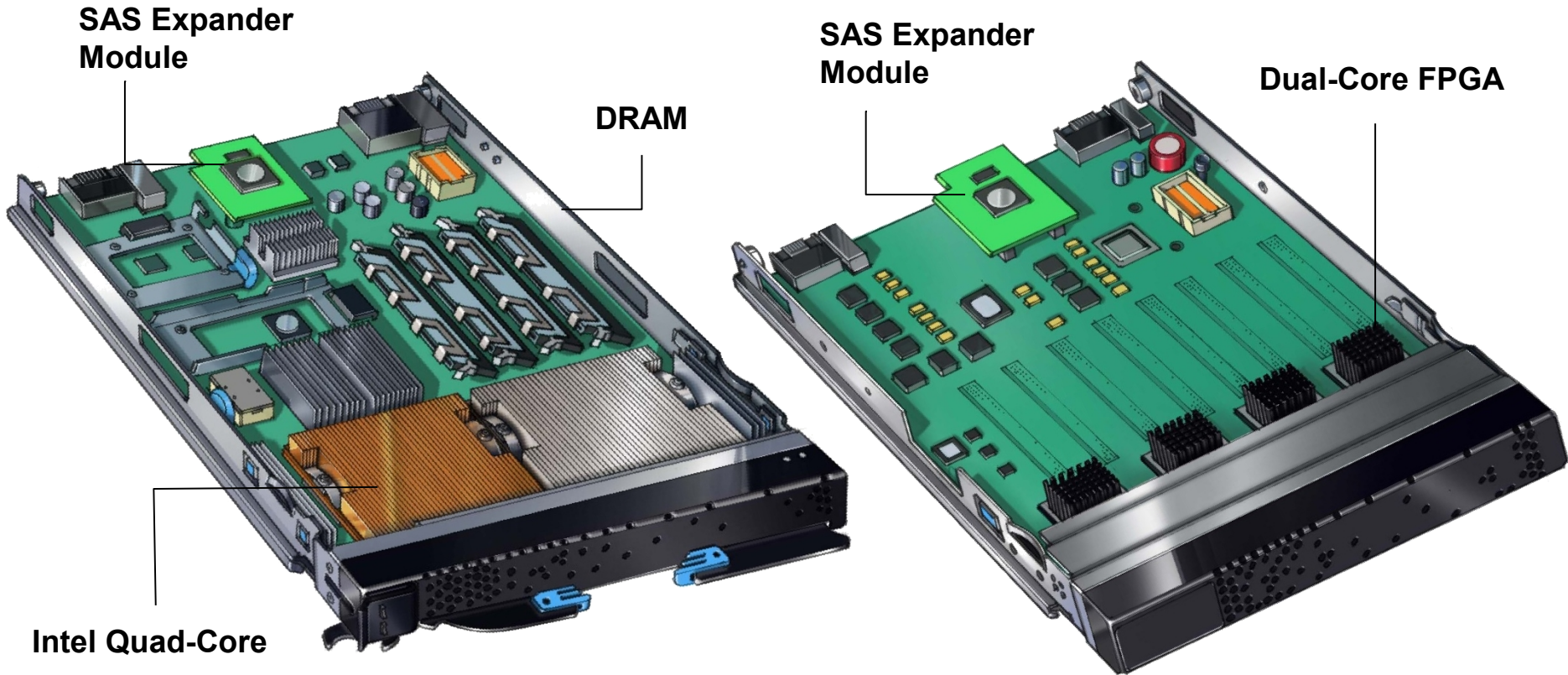
- User data, mirror, swap partitions
- High speed data streaming

## *Snippet Blades™*

- Hardware-based query acceleration with FPGAs
- Blistering fast results
- Complex analytics executed as the data streams from disk



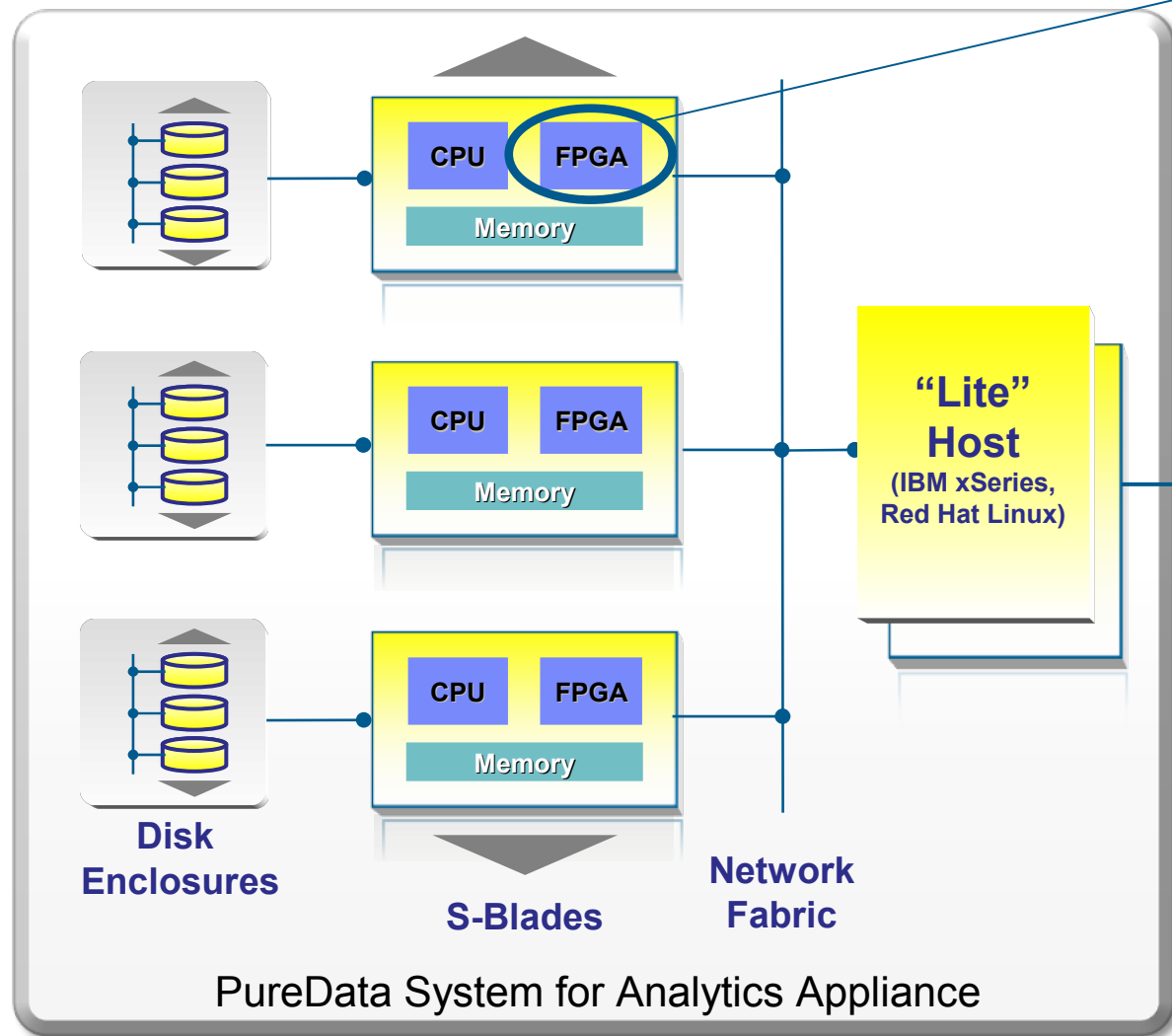
# Snippet-Blade™ (S-Blade) Components



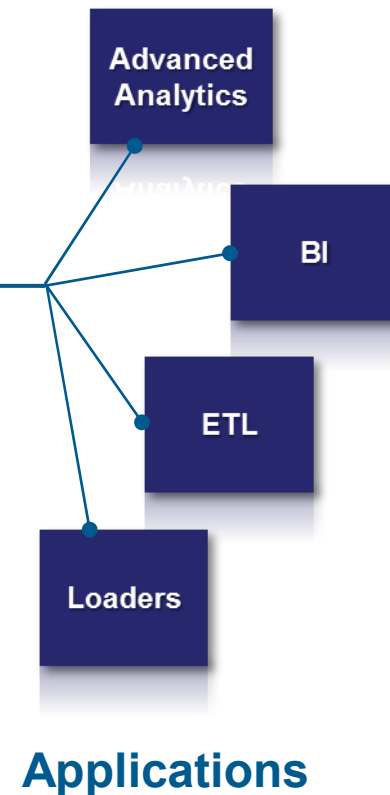
**IBM BladeCenter Server**

**Netezza DB Accelerator**

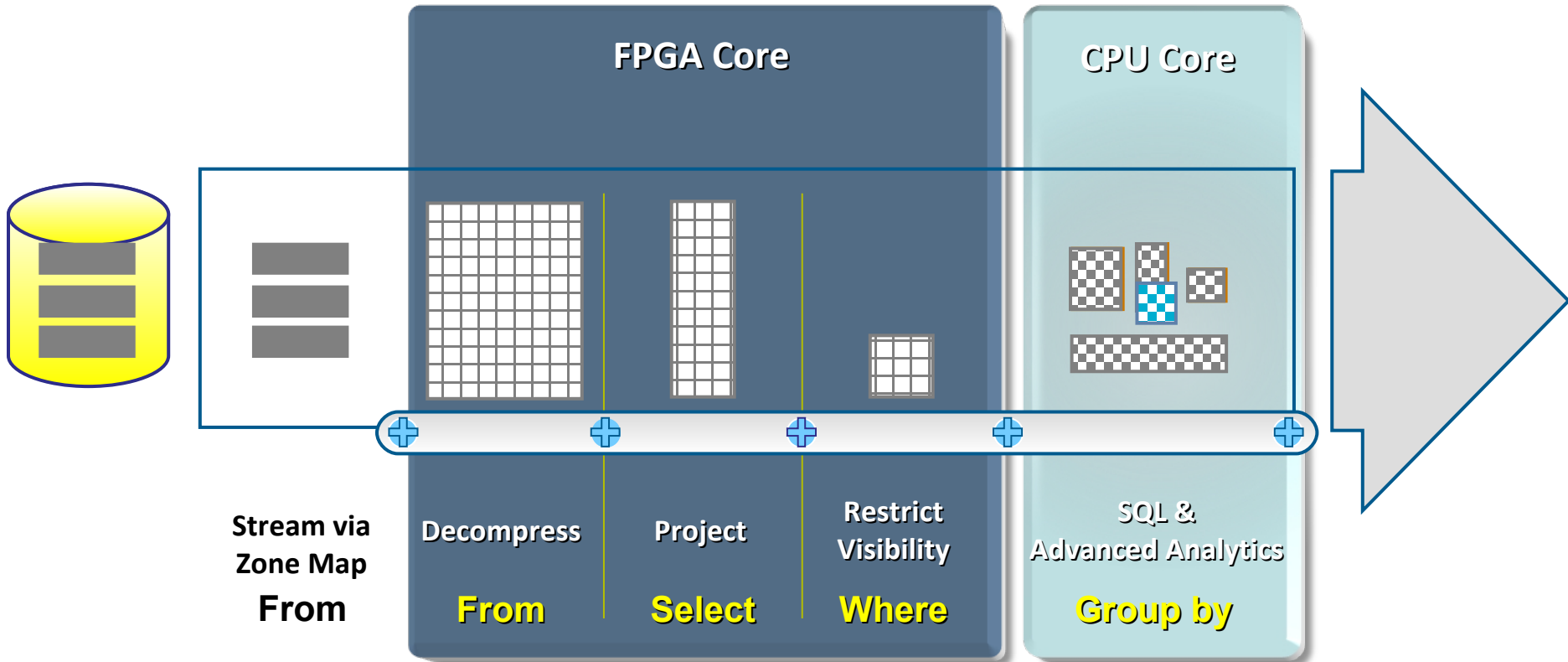
# AMPP Architecture



*Field Programmable Gate Array* =  
a blank canvas until it's configured

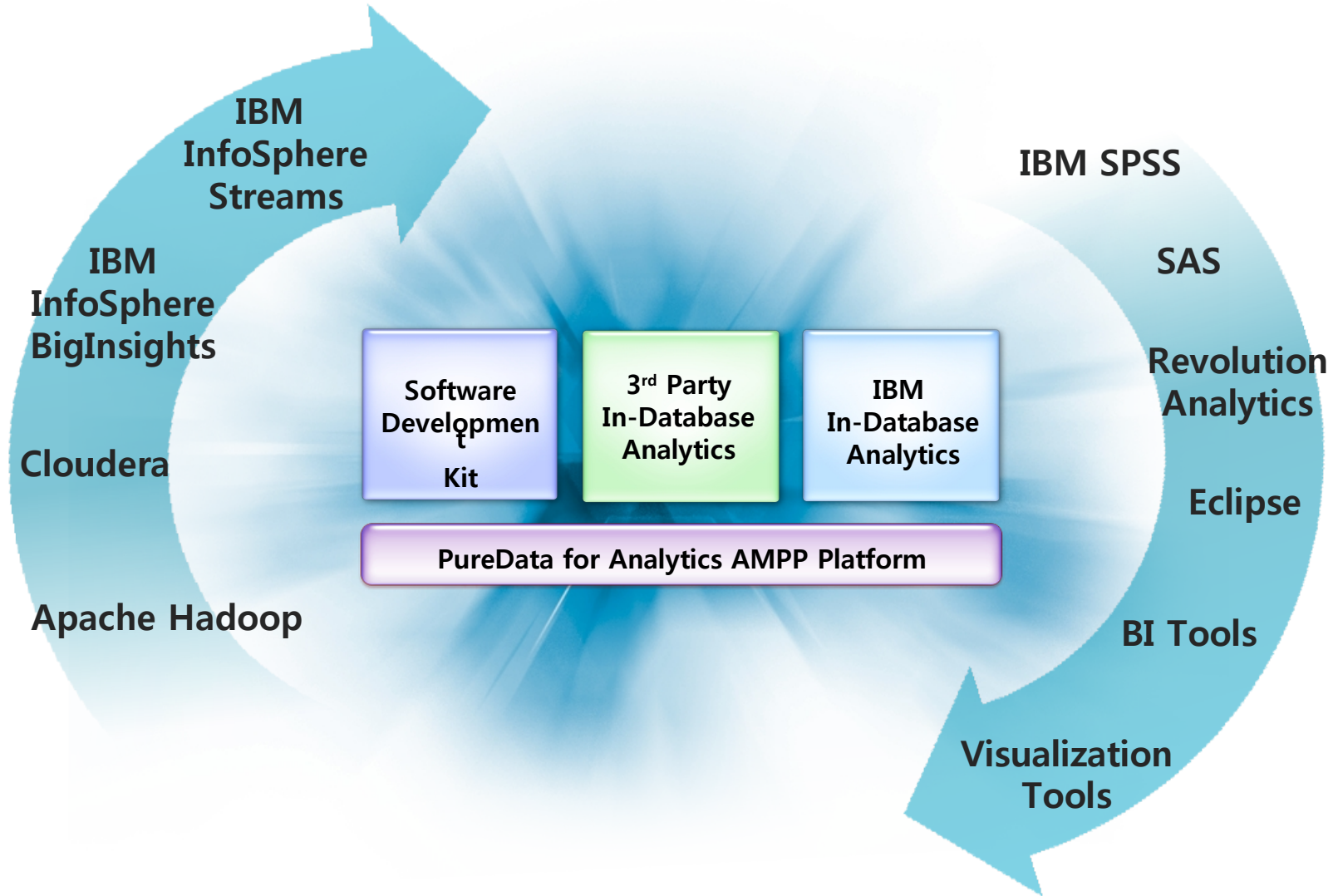


# 고유의 S-Blade Data Stream Processing



```
Select State, Age, Gender, count(*) From MultiBillionRowCustomerTable Where BirthDate < '01/01/1960'
And State in ('FL', 'GA', 'SC', 'NC') Group by State, Age, Gender Order by State, Age, Gender
And State in ('FL', 'GA', 'SC', 'NC') Group by State, Age, Gender Order by State, Age, Gender
```

# IBM Netezza Analytics 2.0: 임베드된 분석 엔진



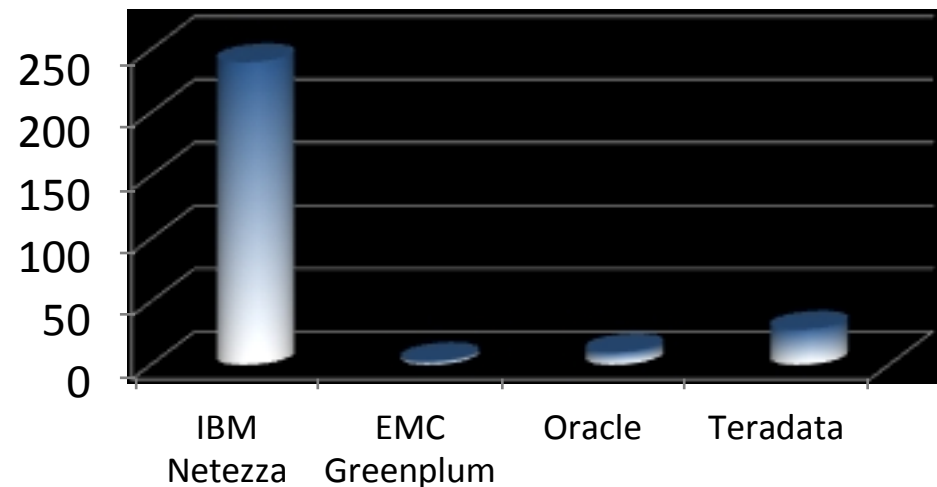
# Netezza In-DB Analytics Version 2.0

업계에서 가장 다양한 분석 함수 라이브러리 제공

**Netezza  
In-Database  
Analytics**

- Transformations
- Mathematical
- Geospatial
- Predictive
- Statistics
- Time Series
- Data Mining

The MOST In-Database  
Analytic Functions



- ✓ No data movement
- ✓ Analyze deep and wide data
- ✓ High performance, parallel computation



# 04

## PureData System for Analytics What's New?

Built-in expertise ■ Integration by design ■ Simplified experience

# Netezza에서 PureData로: What's New?

*Improved Concurrency, Performance, I/O efficiency and Manageability*

## Better Performance



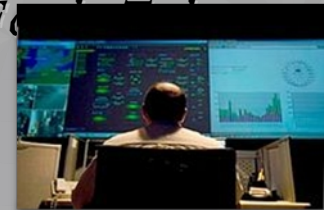
- Tactical Query\*에 한해 기존 네티자 어플라이언스 대비 데이터 처리량 최고 20X 상승
- Directed Data Processing으로 쿼리 동시 처리 능력 강화
- 더 세밀해진 Zone Map으로 디스크 I/O 감소

## Improved Management & Efficiency



- 업무부하 밸런싱 자동화
- 기타 세세한 개선 사항:
  - Optimizer efficiency
  - Memory management
  - Communications protocols
  - Workload management

## Improved Resiliency and Fault Tolerance



- 기존 예비 S-Blade, Passive에서 Active 구성 변화로 Failover시 성능 저하 최소화

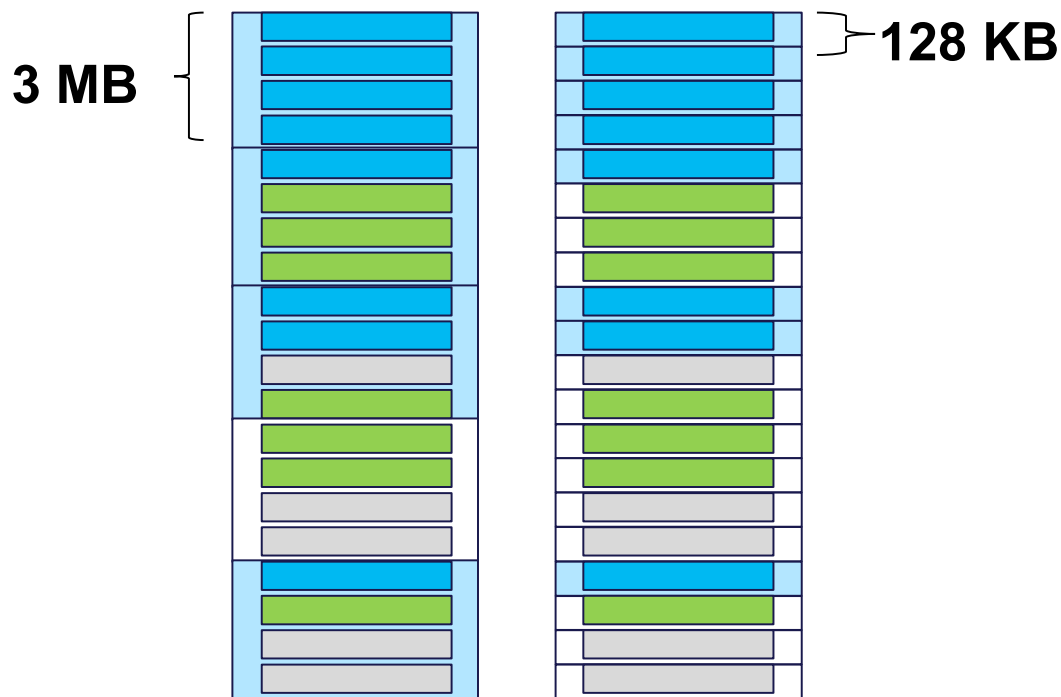
**Tactical Query:** 하나의 Entity에 대한 계산이나 집계를 위해 진행되는 쿼리 (예: 스타벅스 도곡역 지점의 매출)

# What's New: Page Granular Zone Maps

데이터 처리량 향상의 비결

*24X finer granularity*

where col = October



**Total 12 MB**  
(4 x 3 MB)

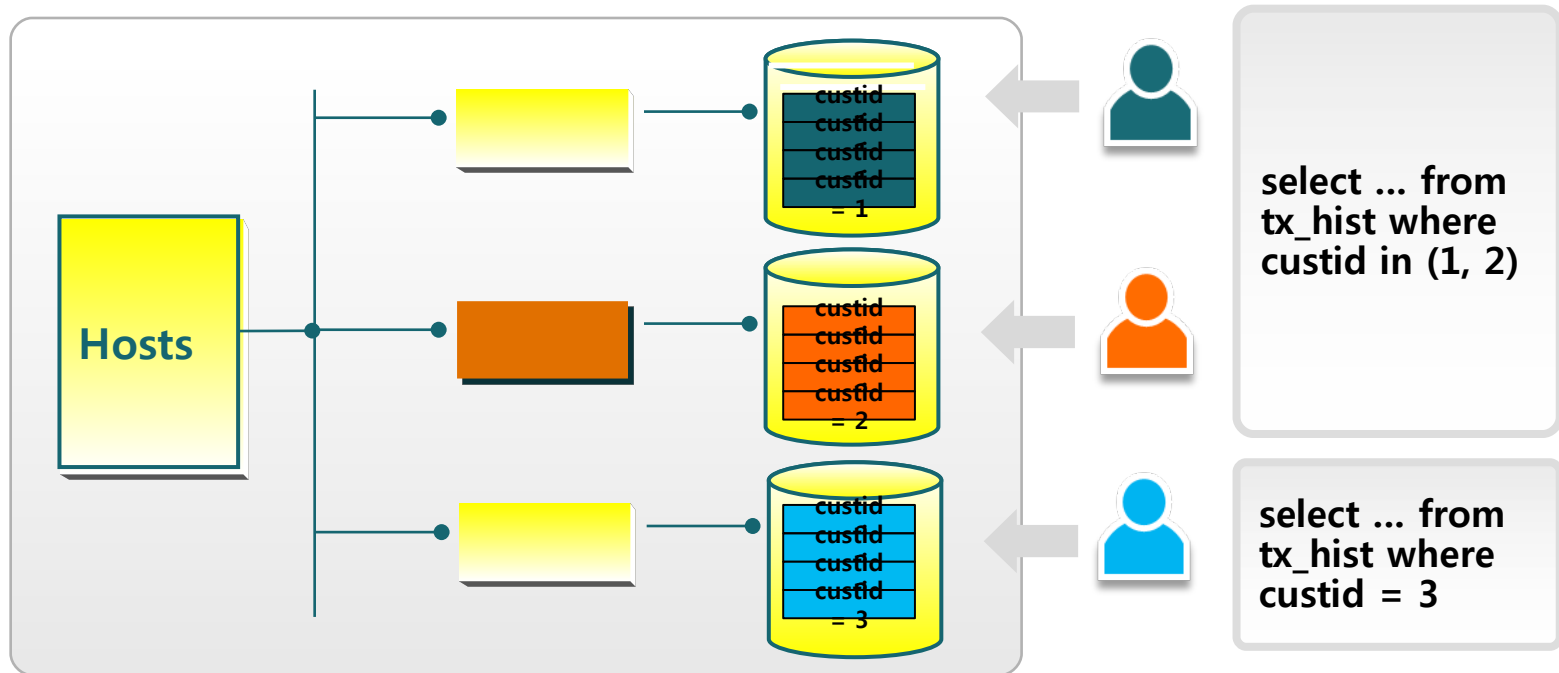
**Total 1 MB**  
(8 x 128KB)

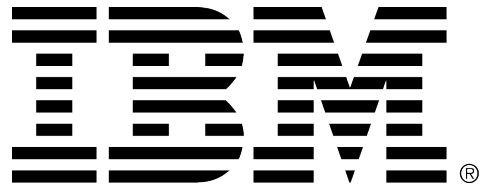
# What's New: Directed Data Processing

## Concurrency 향상

- ❖ 디스크상 경합이 발생하지 않는 쿼리는 동시 실행 (기존에는 순차적으로 처리)

예: custid (1,2)와 custid (3)에 대한 쿼리 동시 실행





© International Business Machines Corporation 2012

International Business Machines Corporation New Orchard Road Armonk, NY 10504

IBM, the IBM logo, PureSystems, PureFlex, PureApplication, PureData and ibm.com are trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide.

A current list of IBM trademarks is available on the Web at [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml)

All rights reserved.

WAP12402-USEN-01