



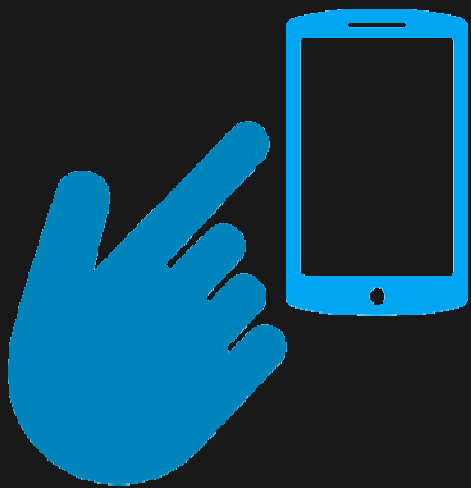
빅데이터를 위한 최적의 플랫폼

IBM 소프트웨어 정보관리 사업부 최석재 부장



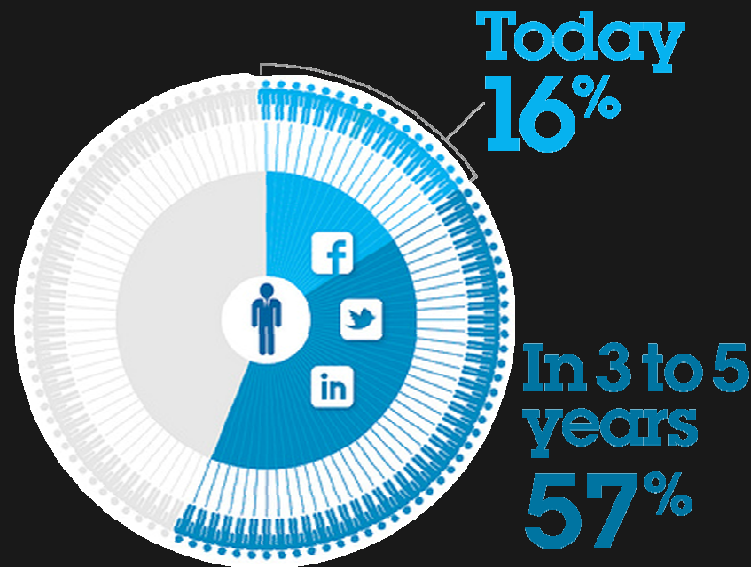
IT환경의 큰 변화 도래

모바일



모바일 사용자의 90%는 손이 닿는 곳에 항상 모바일 기기를 가지고 있음¹

소셜



고객과 소통하기 위해 소셜 미디어를 사용하는 CEO의 비율²

빅데이터 & 분석



디지털 콘텐츠가 2015년에 약 8 제타바이트까지 증가될 것으로 예상³

¹ IBM Forum 2012, Smarter Commerce Prague

² IBM GBS 2011 IBV Study, "The power of cloud: driving business model innovation"

³ IDC, IDC Predictions 2012: Competing for 2020, Doc #231720, December 2011

변화의 대응하기 위한 기업의 변화

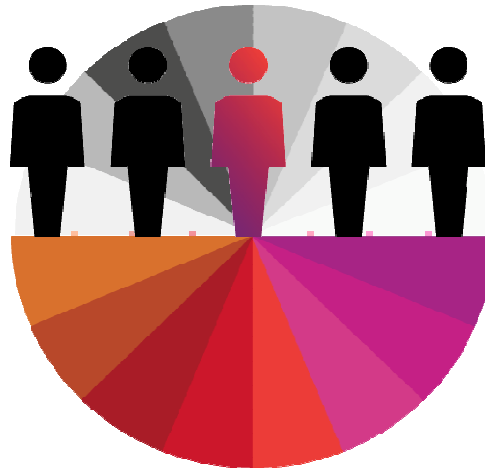
신규 어플리케이션, 빅데이터
, 분석 작업 가속화



**34%의 신규
IT프로젝트는
계획보다 길어짐**

From a commissioned study conducted by Forrester Consulting on behalf of IBM

전문 통합 시스템으로
IT효율성 향상



**단지 5개중에 1개의
신규프로젝트에 IT 예산의
50%이상을 투자함¹**

IBM, *Data center operational efficiency best practices*, April 2012.

클라우드 인프라 및
애플리케이션 플랫폼을
단순화



**기업의 90%가
클라우드 구축 계획
2015년 까지**

* IBM GBS 2011 IBV Study, "The power of cloud: driving business model innovation"

빅데이터 환경으로의 요구사항 증가

Increasing
데이터의 크기
용량확장 요구

50x
2010 2020
35 ZB
by 2020

Increasing
데이터의 속도
보다 높은 성능 요구

초당 수백만
트랜잭션

Increasing
데이터의 다양성
새로운 기술 요구

수십억의
디바이스 &
센서들

BigData IT환경을 통해 고객이 원하는 것은

복잡도 감소
(Reduce complexity)

민첩성 증대
(Accelerate time to value)

비용 절감
(Improve IT economics)

PureData



Data Platform

Delivering Big Data Platform Services

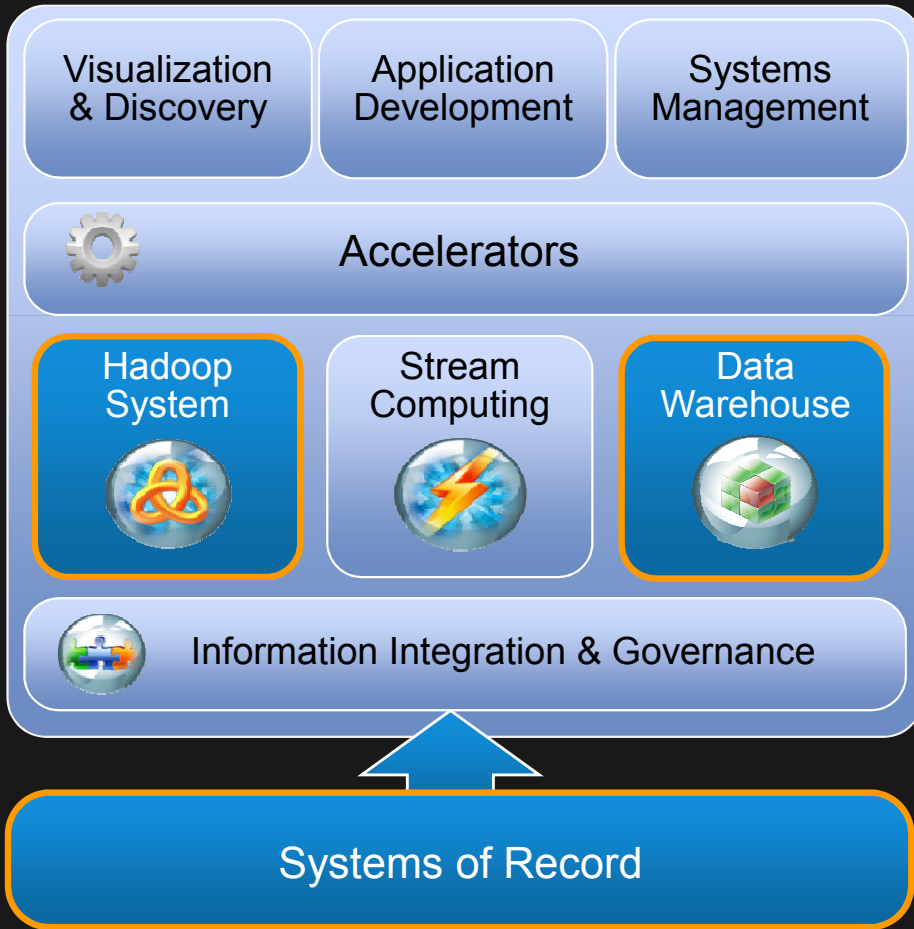
- 최적화된 성능 보장
- 데이터 적재 후 사용가능 하기 까지 최단 시간 소요
- 통합 관리
- 자동화된 유지 관리
- 일원화된 지원

PureSystems

PureData System

빅데이터를 위한 전문가 통합 시스템

IBM Big Data Platform



PureData

For Hadoop

Optimized system to accelerate big data analytics and online archive with appliance simplicity



For Analytics

Optimized system delivering data services for analytics & reporting

For Operational Analytics

Optimized system delivering data services for operational analytics

For Transactions

Optimized system delivering data services for transactions

워크로드에 따라 최적화된 시스템이 필요함

Order Management

Sales Analysis

Real Time Fraud Detection

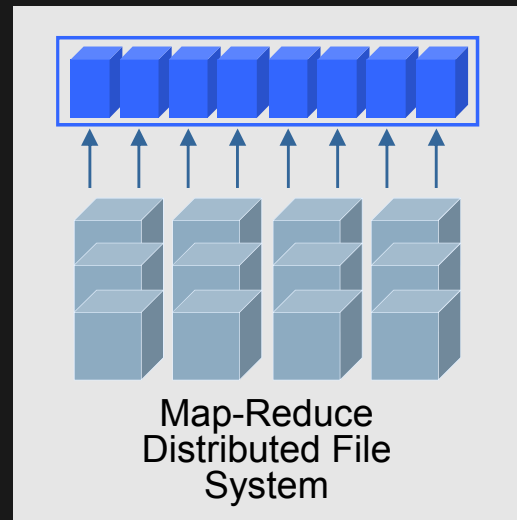
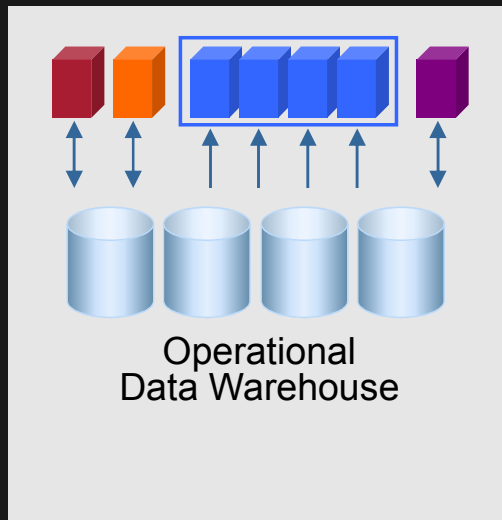
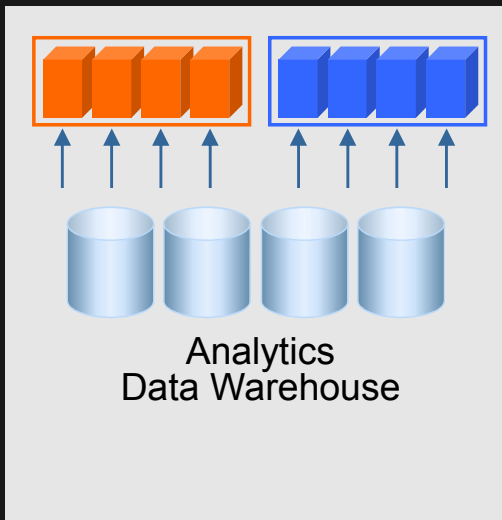
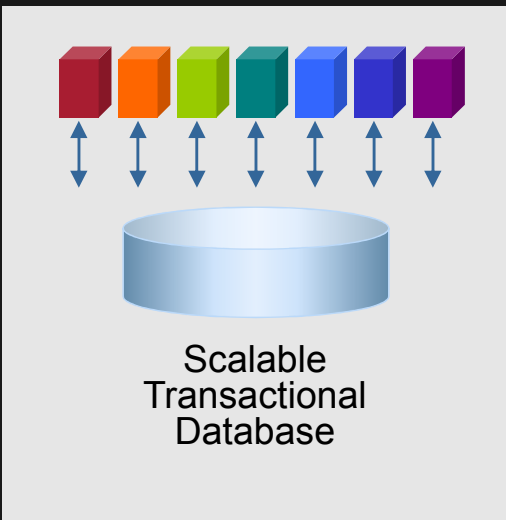
Customer Analysis

Transaction Processing

Reporting and Analytics

Operational Analytics

Any Structure Analytics



IBM PureData System

빠르고 쉽게 빅데이터 환경의 여러 요구사항을 충족해주는 시스템



PureData
System for Transactions

For apps like order management ...

트랜잭션 성능과 확장성을 극대화한 데이터 베이스 클러스터 서비스 시스템

PureData
System for Analytics

For apps like Sales Analysis...

고성능 대용량 (peta-scale) 분석과 단순함을 극대화한 DW 서비스 시스템

PureData
System for Operational Analytics

For apps like Real-time Fraud Detection...

고성능과 실시간 트랜잭션 처리를 균형 있게 갖춘 Operational DW 서비스 시스템

PureData
System for Hadoop

New

For Exploratory Analysis & Queryable Archive

빅데이터 분석 과 어플라이언스의 장점을 하나로 모은 Hadoop 데이터 서비스 시스템

PureData

System for Analytics



분석 데이터 워크로드에 가장 최적화된 시스템

- 기존 custom 시스템에 비해 **10-100배** 빠른 성능
- 쿼리 동시성과 성능부분에 있어 **20배** 성능 향상
- 특히 받은 **MPP 하드웨어 가속 기능**
- No 인덱스, No 튜닝, No 스토리지 관리
- 고성능 데이터 적재
- in-database 분석을 위한 다양한 기능

Powered by
Netezza Technology

PureData System for Analytics 하드웨어 구성

N2001 Model



12 Disk Enclosures

- 288 600 GB SAS2 Drives
 - 240 for User Data
 - 14 for S-Blades
 - 34 Spare
- RAID 1 Mirroring

2 Hosts (Active-Passive)

- 2 6-Core Intel 3.46 GHz CPUs
- 7x300 GB SAS Drives
- Red Hat Linux 6 64-bit

7 PureData for Analytics S-Blades™

- 2 Intel 8 Core 2+ GHz CPUs
- 2 8-Engine Xilinx Virtex-6 FPGAs
- 128 GB RAM + 8 GB slice buffer
- Linux 64-bit Kernel

½ Rack 에서 4 Rack
까지 확장 가능

- User 데이터 용량: 192 TB*
- 데이터 스캔 속도: 450 TB/hr*
- Load 속도 (per system): 5+ TB/hr

- 전력량: 7.5 kW
- 발열량: 27,000 BTU/hr

* 4배 압축 가정

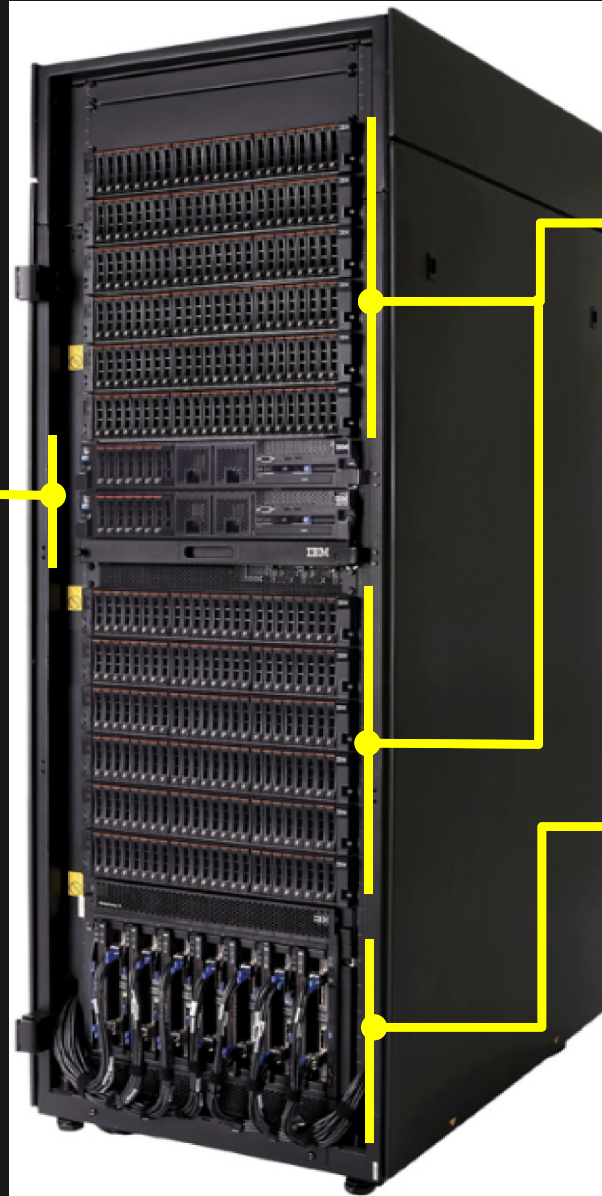
PureData System for Analytics 구성

최적화된 하드웨어 + 소프트웨어

- 최적화된 AMPP 하드웨어
- 고성능 분석을 위한 설계
- 튜닝이 필요 없음

SMP Hosts

- 쿼리 컴파일
- 쿼리 플랜
- 옵티마이저
- 관리



디스크 인클로저

- 유저 데이터, 미러, 스왑 파티션
- 고성능 데이터 스트리밍

Snippet Blades™

- FPGA를 이용한 하드웨어 기반의 쿼리 가속기
- 빠른 결과 추출
- 디스크에서 바로 읽어 스트림으로 복잡한 분석이 가능



최대 **2000배** 까지 속도 향상
매달 30% 용량 증가

"Netezza has allowed us to reduce the complexity of regulatory reporting and processing of exchange data from days down to minutes."

- Edwin Marcial, Chief Technology Officer



기본 교육만으로 **6개월간**
운영
기존 Oracle 시스템 대비 **200배**
빠름

"Allowing the business users access to the Netezza box was what sold it."

- Steve Taff, Executive Dir. of IT Services



1 PB on Netezza
7년의 이력 데이터 보관
매년 **100-200%** 데이터 증가

"NYSE ... has replaced an Oracle IO relational database with a data warehousing appliance from Netezza, allowing it to conduct rapid searches of 650 terabytes of data."

- ComputerWeekly.com



버팔로 대학 연구원들은 10경
(10^{18}) 연상의 수행 시간을 기존
27시간 에서 12분으로 단축
시켰음

"Once we had the data on Netezza we were able to do the same analysis and much more complex analysis in minutes."

- Dr. Murali Ramanathan, SUNY Buffalo



PureData System for Analytics 주요 장점

분석쿼리에 대한
성능 가속

- **3배 빠른 성능¹**
빅데이터 분석을 위해
- **rack² 당 128 GB/sec 디스크 스캔 속도**
빅데이터를 빠르게 처리하기 위해

데이터 센터의 운용 효율성
증대

- **rack³ 당 50%의 데이터 저장 용량 증가**
데이터 센터를 최적화 하기 위해
- **더 많은 용량, 더 적은 전력 소모**
Oracle 과 Teradata와 비교해서

단순화 및 쉬운 관리

- **향상된 시스템 관리 및 탄력성**
관리를 위해 적은 시간을 사용하지만 더 나은 value 전달
- **70% 적은 지원 서비스 호출**
더 많은 스페어 드라이브와 빠른 디스크 regeneration⁴

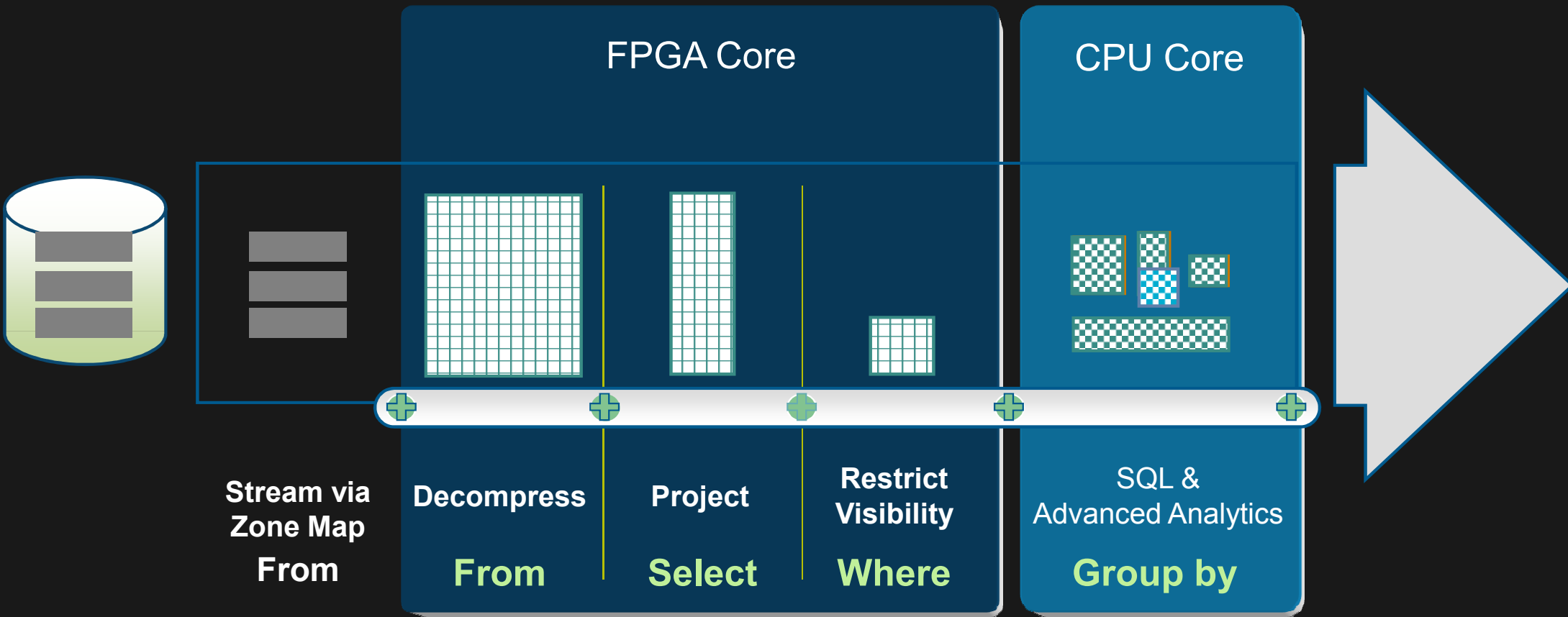
¹ Based on a comparison of the IBM PureData System for Analytics N2001 to the IBM PureData System for Analytics N1001. The performance speed refers to the query times on both macro-analytic and mixed workload tests as conducted in IBM engineering lab benchmarks. The N2001 query times were an average of 3x faster than those of the N1001. Individual results may vary.

² 128 GB/sec scan rate assuming an average of 4x compression across the system. Individual results may vary.

³ Capacity of IBM PureData System for Analytics N2001 compared to previous generation IBM PureData System for Analytics N1001.

⁴ Each N2001 rack contains 34 hot spare drives and 240 active drives for a ratio of 1 spare per 7 drives. Each N1001 rack contains 4 hot spare drives and 92 active drives for a ratio of 1 spare per 23 drives. The N2001 has 3.3x more spares per active drive. Frequency of disk related service calls expected to decrease by 70% assuming the same drive failure rates.

PureData System for Analytics 수행 원리



```

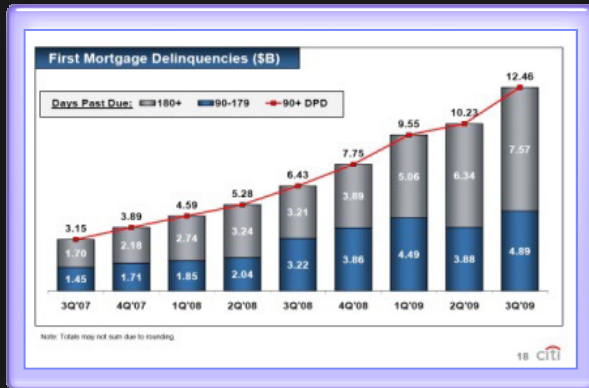
Select State, Age, Gender, count(*)
From MultiBillionRowCustomerTable
Where BirthDate < '01/01/1960' And State in ('FL', 'GA', 'SC', 'NC')
Group by State, Age, Gender Order by State, Age, Gender
    
```

PureData System for Analytics 분석 기능

단순 리포트를 넘어 분석 및 최적화 까지..

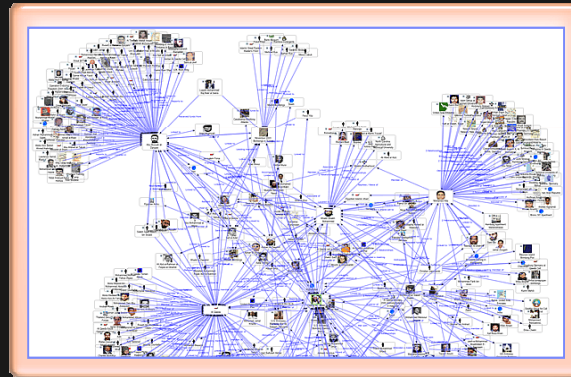
Optimization

BI Reporting and Ad-Hoc Analysis



- What happened?
- When and where?
- How much?

Predictive Analytics

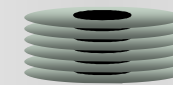
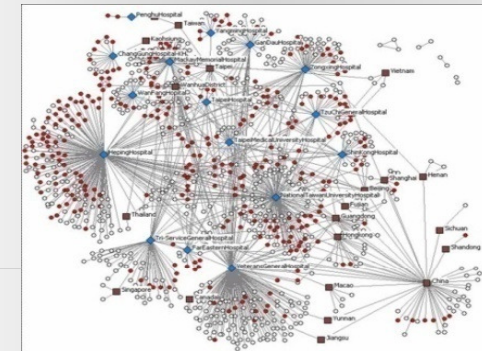
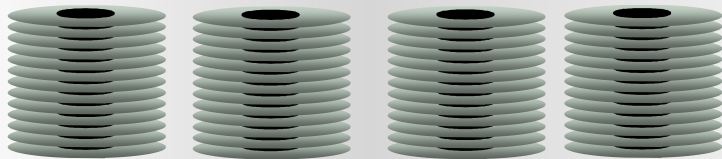
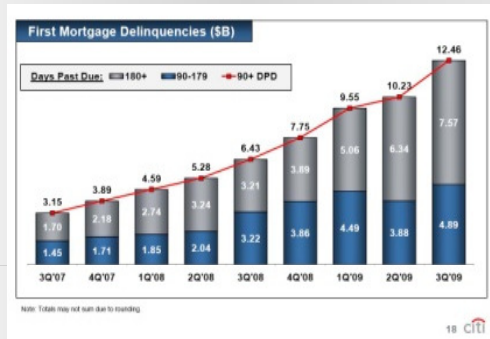


- What will happen?
- What will the impact be?



- What is the best choice?

Big Data Meets Deep Analytics



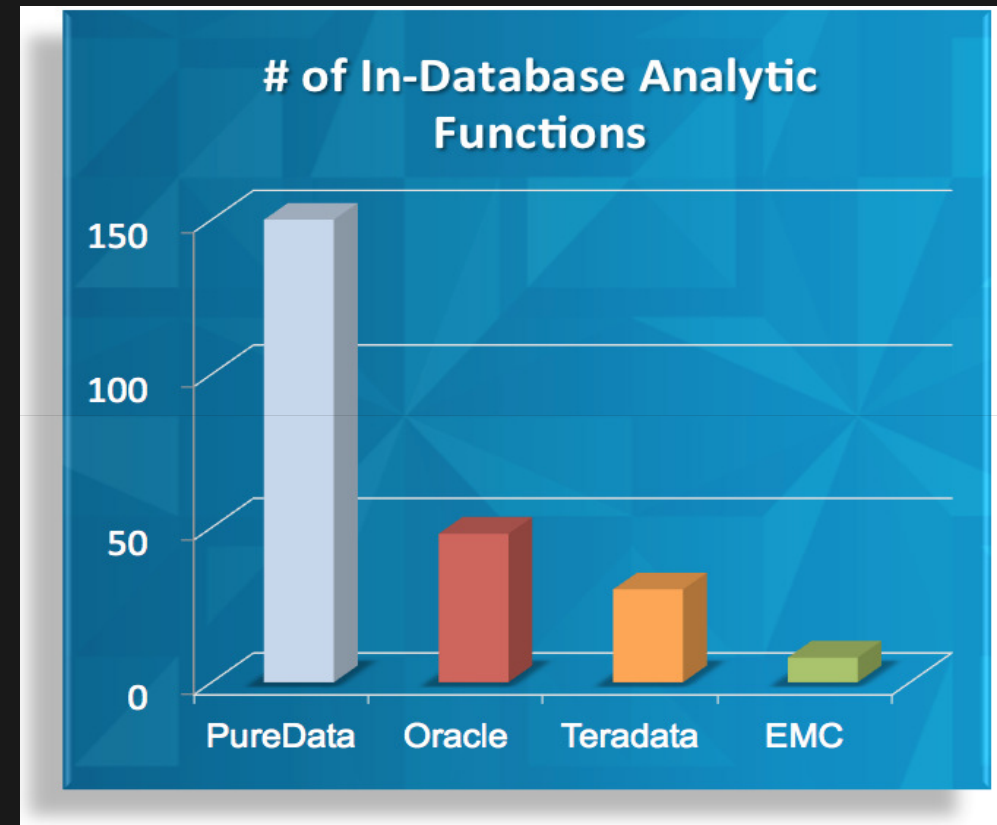
무제한 분석 환경

Integrated by Design

IBM Netezza Analytics Version 2.0

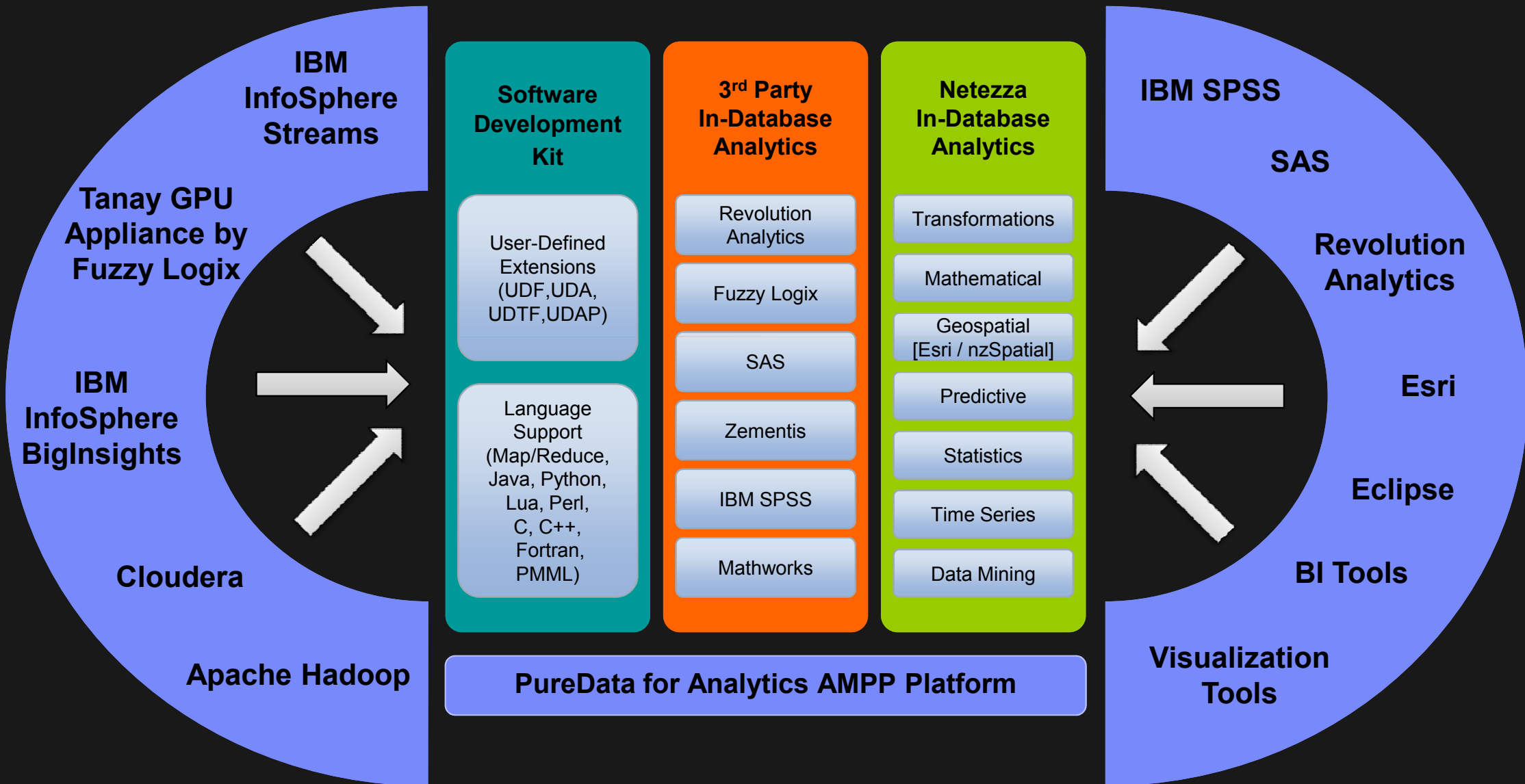
Netezza In-Database Analytics 2.0

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

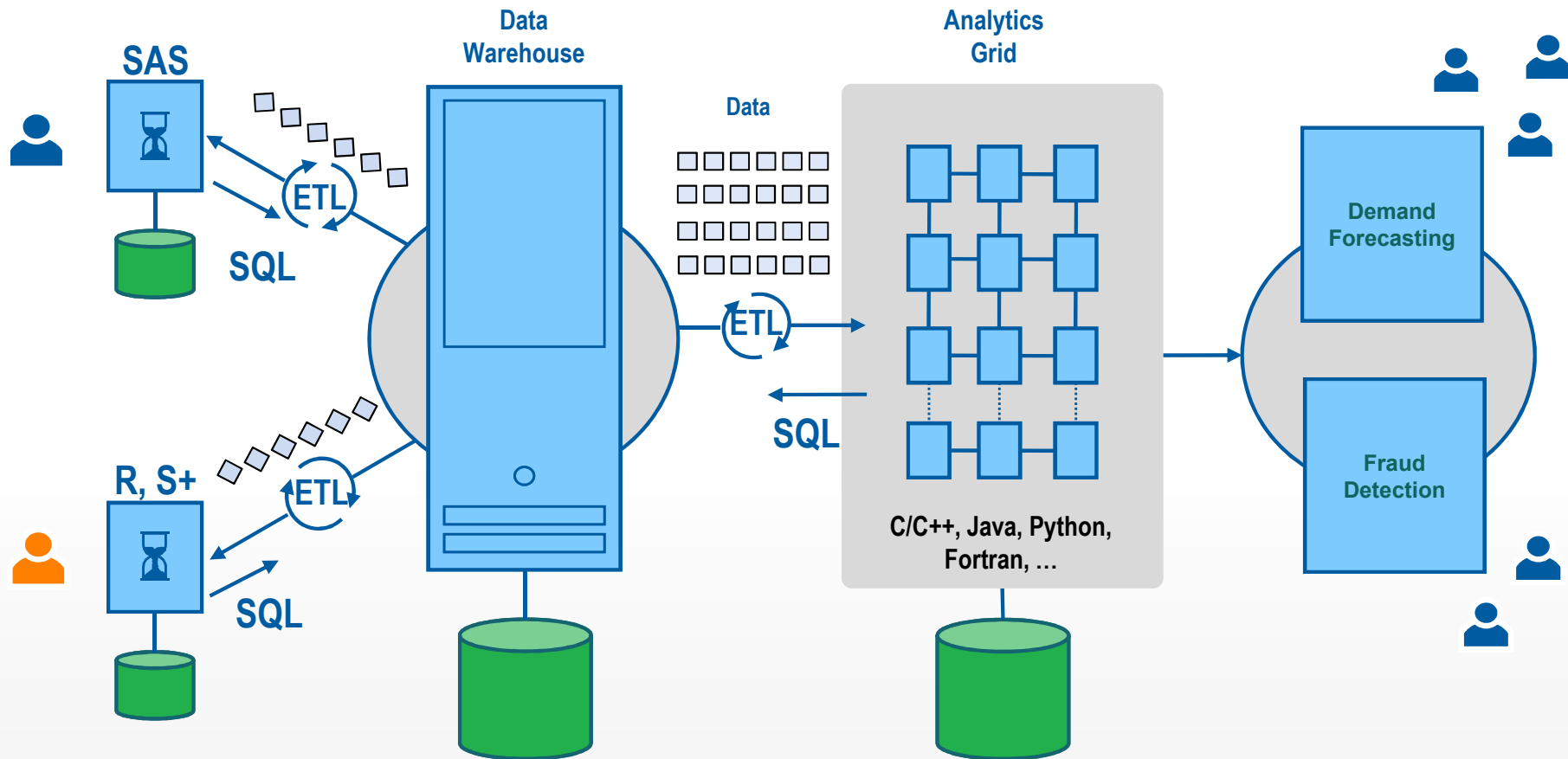


- ✓ 데이터 이동 없음
- ✓ Analyze deep and wide data
- ✓ 고성능, 병렬 컴퓨팅

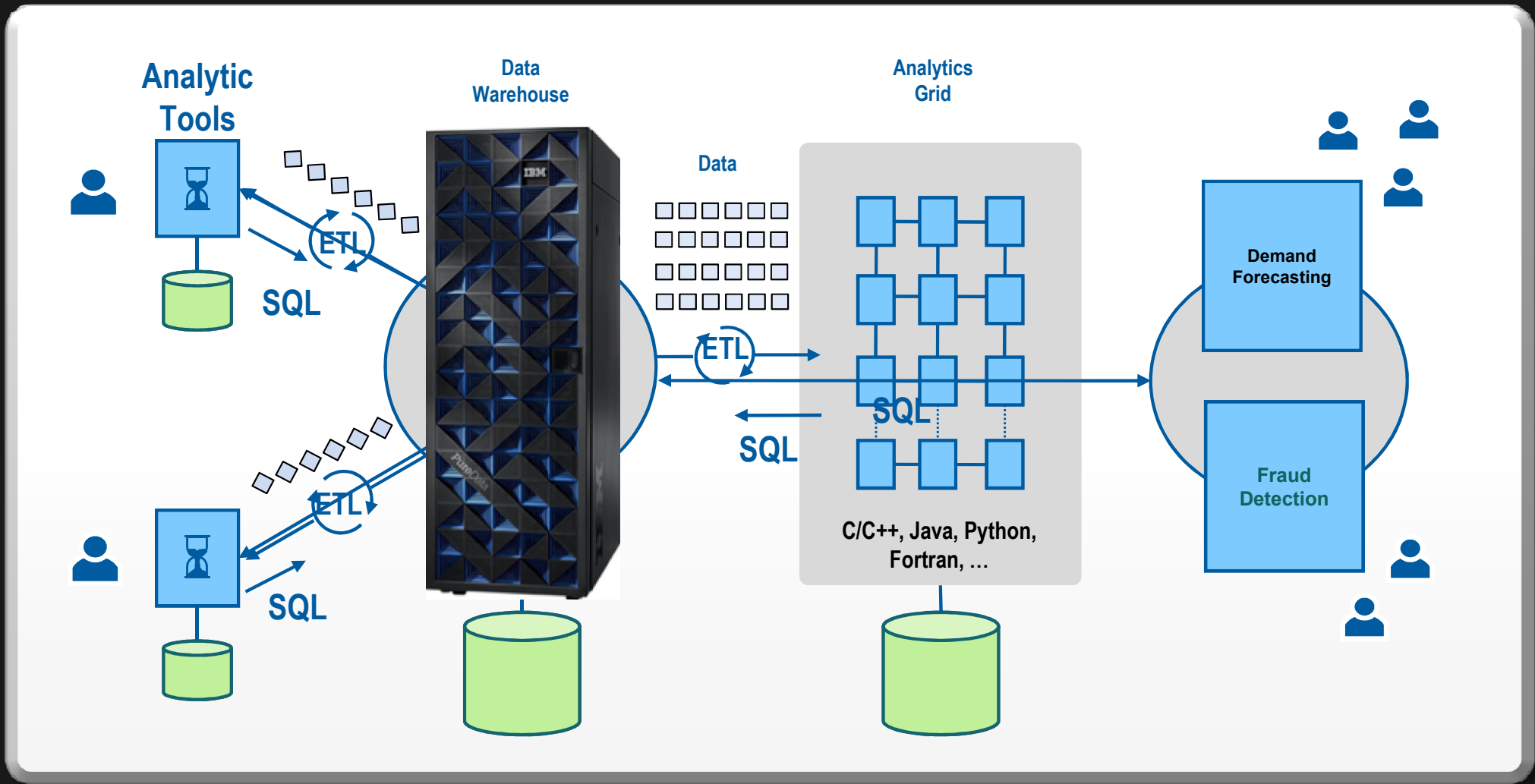
IBM Netezza Analytics Ecosystem



Advanced Analytics – 기존 방식

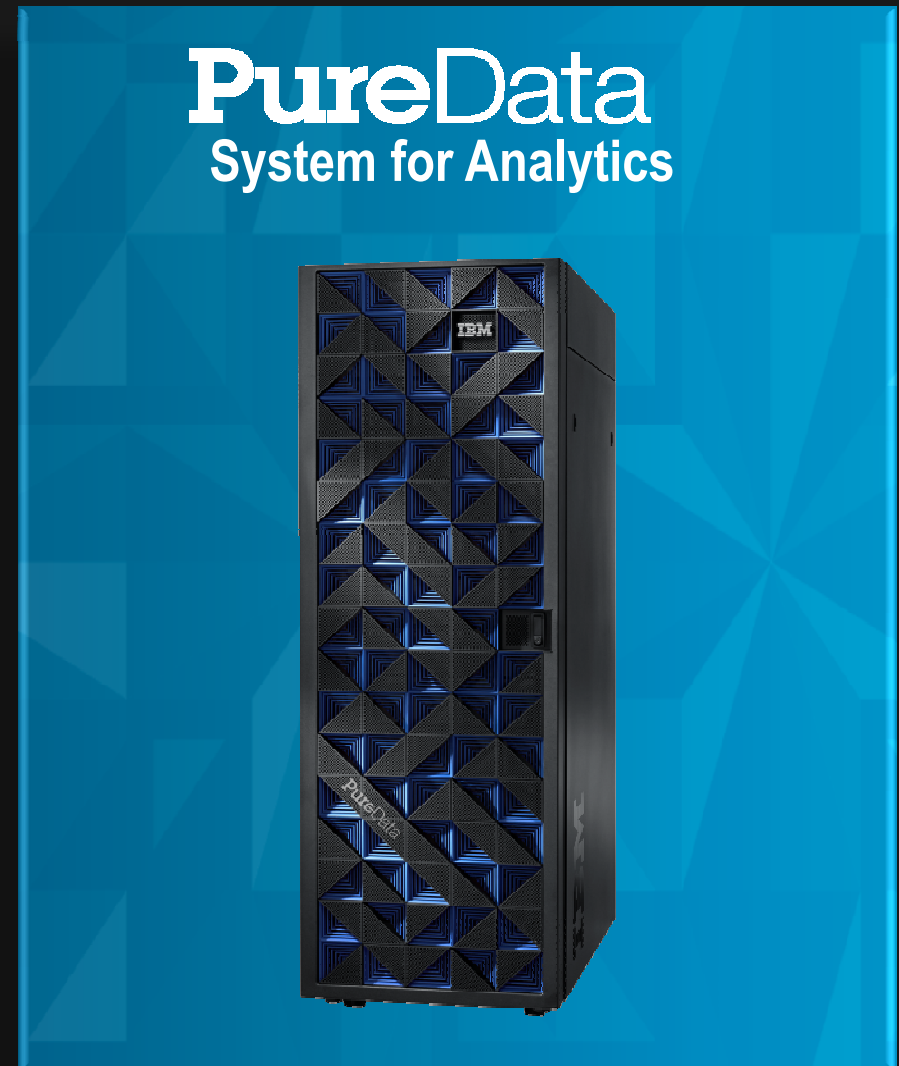


IBM Netezza Analytics – 복잡한 분석의 단순화



IBM PureData System for Analytics 의 가치

- 현재 마켓에서 가장 빠른 가치를 보여주는 (time to value) 제품
- 빅데이터에 최적화된 분석 성능 제공
- 빠르고 민첩한 적용을 위한 단순한 관리
- 분석 성능을 높여줄 방대한 양의 분석 함수들



IBM®

