



# Introduction to SAS

Education Team, SAS Korea

**THE  
POWER  
TO KNOW.**



## Contents...

**Chapter 1** SAS Software 소개

**Chapter 2** SAS System 기본적인 환경

**Chapter 3** 데이터 핸들링 및 레포트 출력

**Chapter 4** SAS 9

**THE  
POWER  
TO KNOW.**

# Chapter 1: SAS Software 소개

## 1.1 SAS Institute Inc. , SAS Korea

### 1.2 SAS System 소개

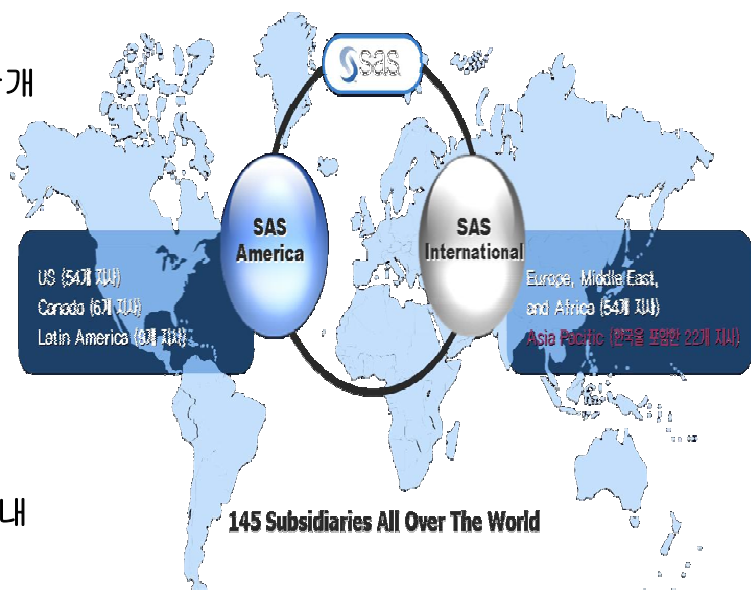
### 1.3 SAS Solution 소개

3



## 1.1 SAS Institute Inc.

- 설립연도: 1976
- 전세계 110여개 국에 걸친 4만개 이상 고객 사,
- 2006 포춘지 선정 100대 기업 중 96개사가 고객
- 세계 최대의 개인소유 소프트웨어 업체
- 1990년 한국지사 설립으로 국내 데이터분석 시장에 선두진입



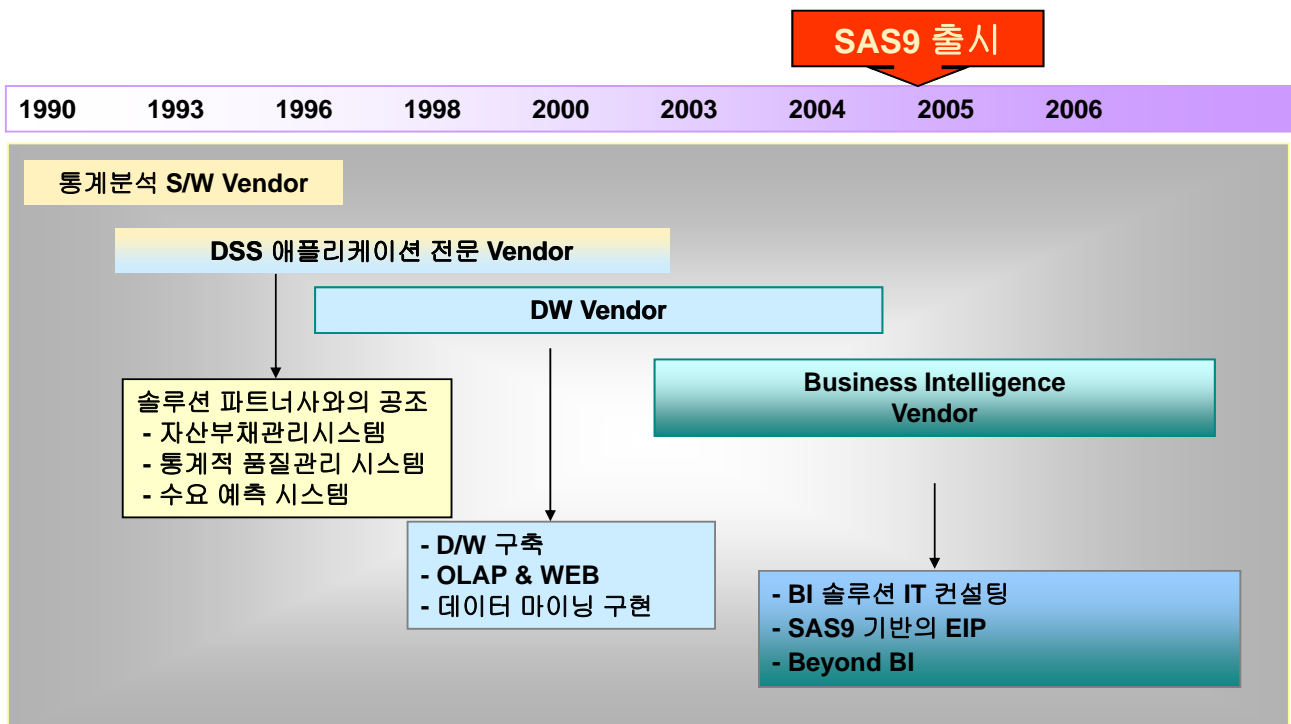
4



# SAS Korea (연혁)

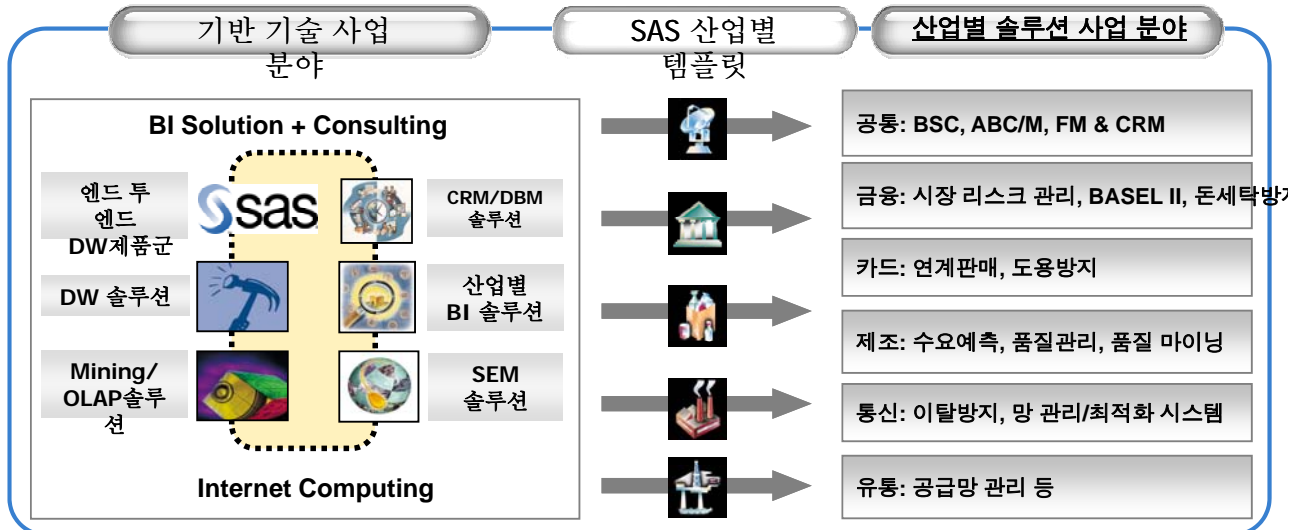
설립 연도	1990년
자 본 금	17억 원
매 출 액	2005년 - 300억 달성 (전년 대비 28% 성장)
고 객 사	금융, 제조 유통, 정부, 공공, 교육기관 등 600여 site
직 원 수	125명 ('06 기준)
서 비 스	컨설팅, 기술지원, 교육
국내 주요 솔루션	품질 마이닝(6 시그마), 바젤 II (은행권 리스크 관리, 용/운영/시장), 사기방지, 둔세탁방지, 비즈니스예측(수요예측, 매출 추정, 조기 경보) 재무관련 솔루션, 고객 인텔리전스 등 200여개 기업 솔루션

# SAS Korea (Biz. 전략)



# SAS Korea (주요 사업 내용)

SAS는 각 산업별 Business Intelligence Solution인 CRM, SPM, SEM, DW, 분석, 데이터마이닝, OLAP, 바젤II 솔루션, 공급망 인텔리전스 등의 강력한 기술력을 기반으로 산업별로 다양한 Needs를 충족하는 최고의 Business Intelligence Solution을 지원

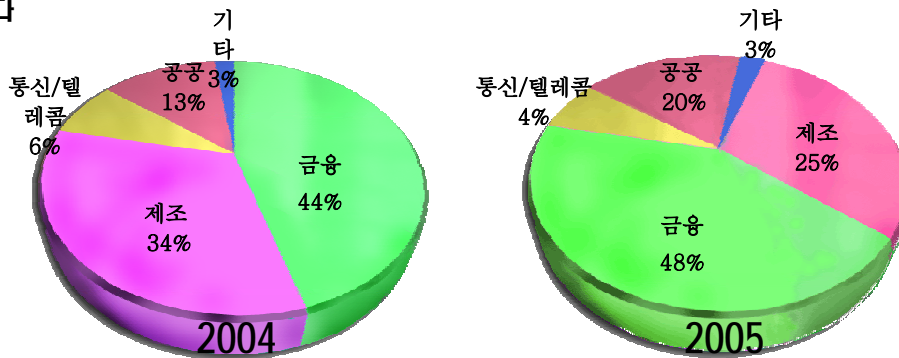


7



# SAS Korea (산업별 고객 현황)

국내에서는 금융산업 솔루션이 40% 이상을 차지하고 있으며, 나머지 제조, 유통/서비스, 통신, 공공, 병원/제약 산업에서도 지속적인 성장으로 국내 성공사례가 타 지역으로 전파되어 성공적인 실적을 보이고 있다



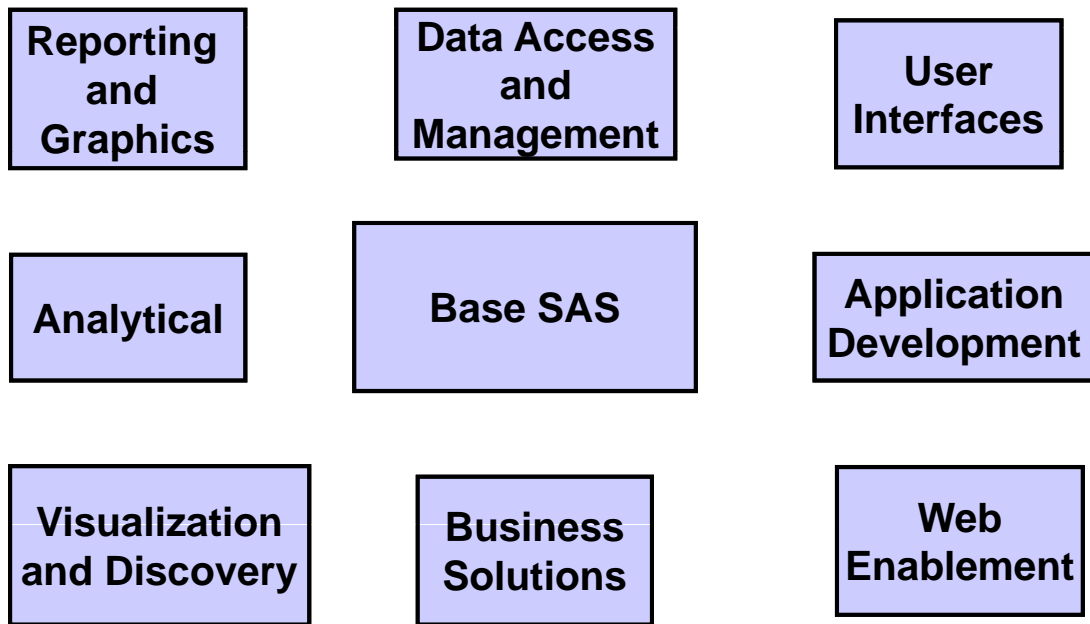
## 주요 국내 고객사



8

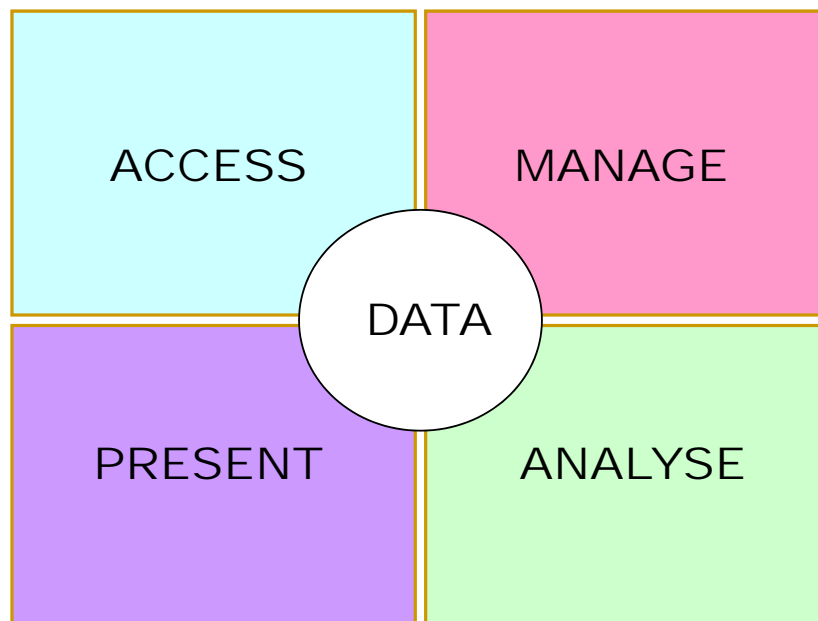


## 1.2 SAS System 소개



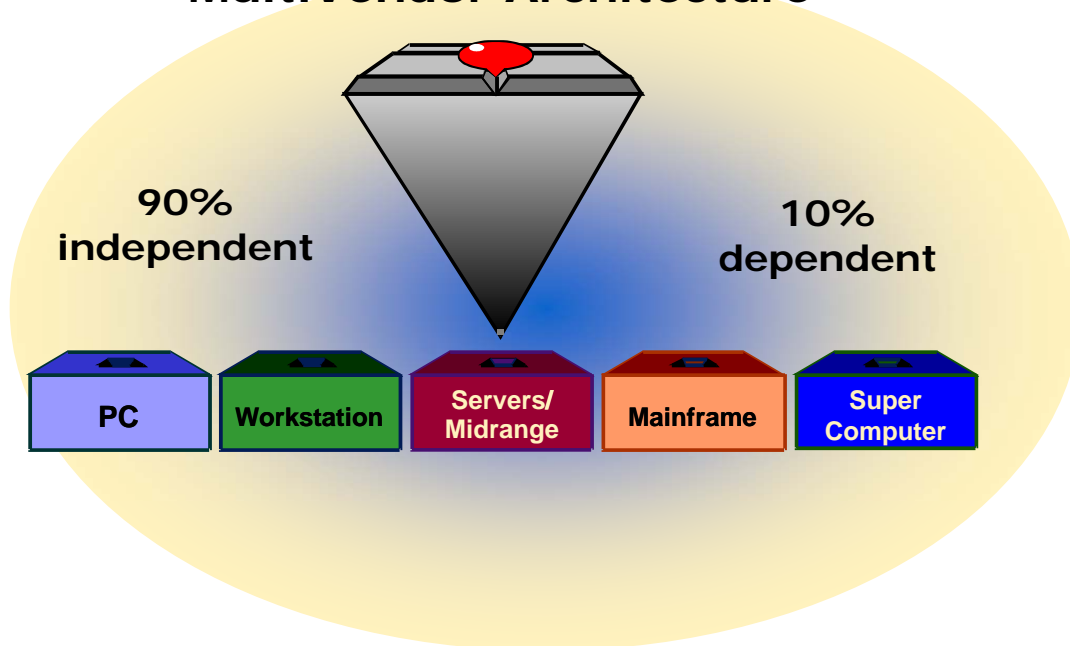
9

## Data Driven Tasks

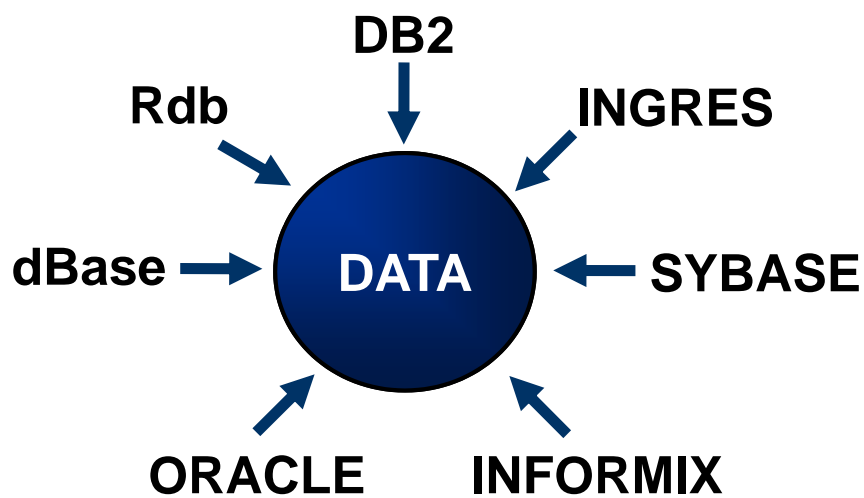


10

## MultiVendor Architecture



## MultiEngine Architecture™



# 1.3 SAS Solution 소개

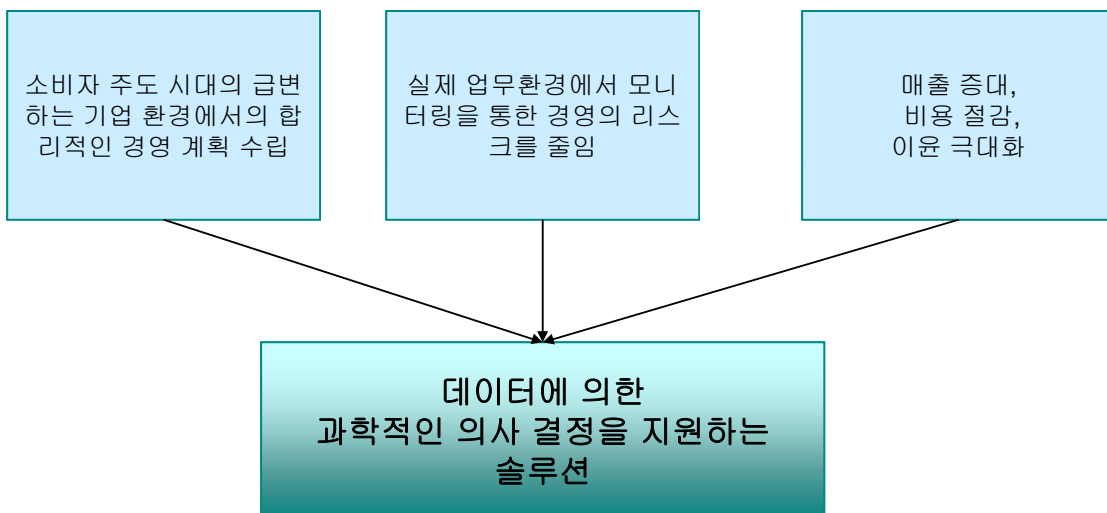
## Enterprise Trends Yesterday's Manager



## Enterprise Trends Tomorrow's Manager

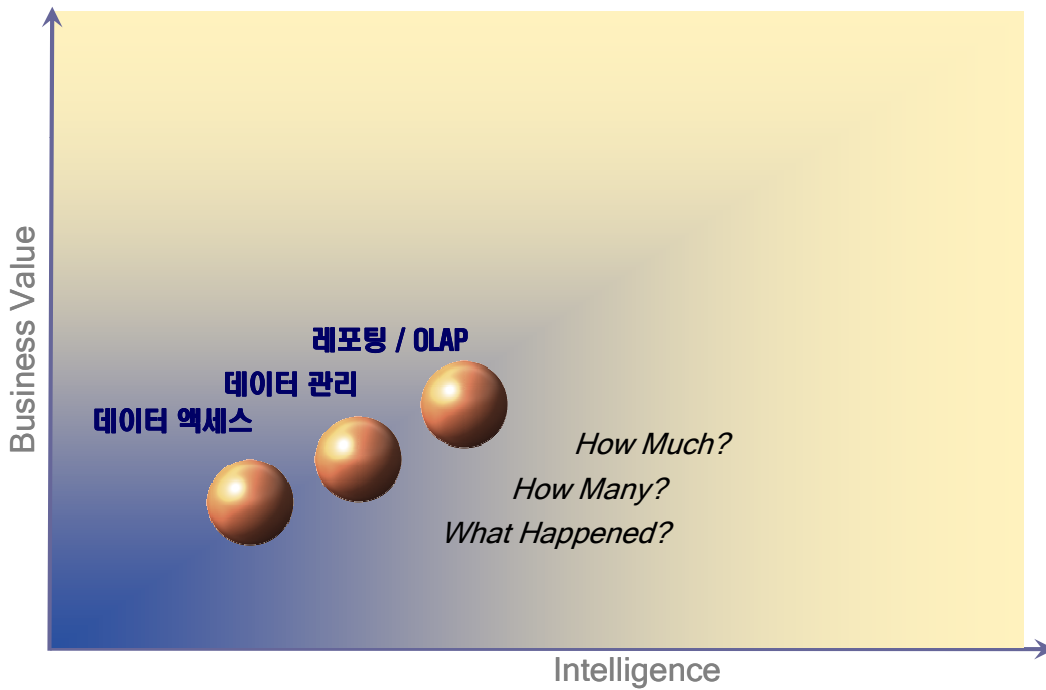


## BI(Business Intelligence)의 목적



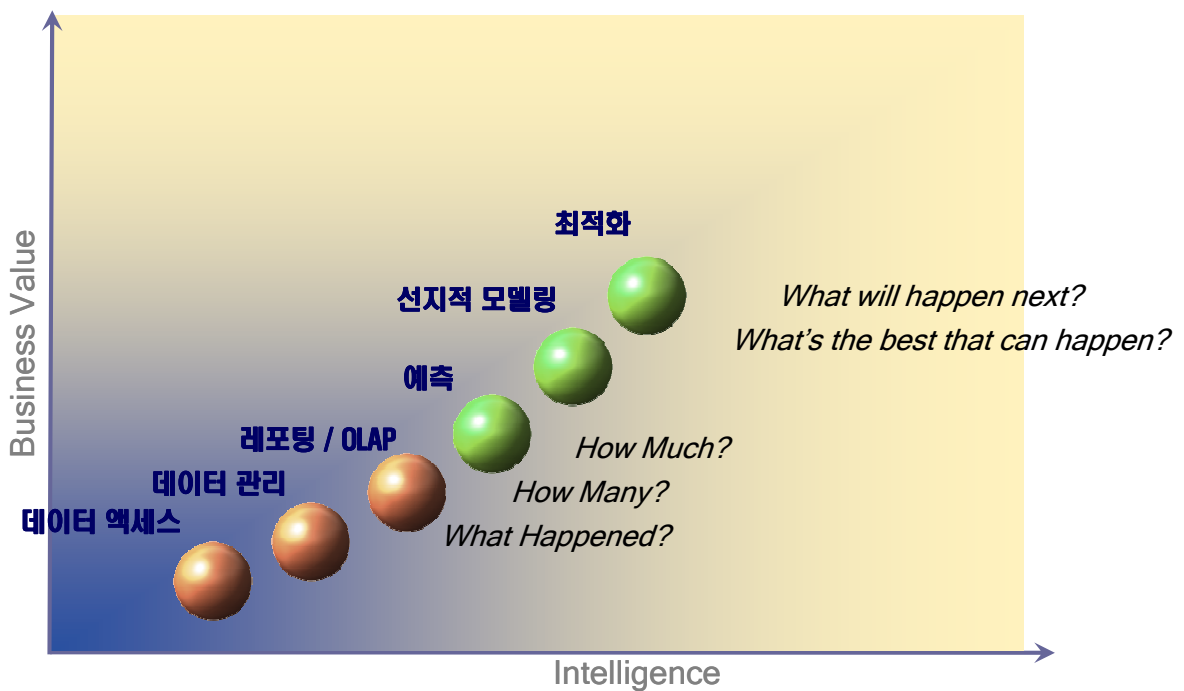
“향후 12개월간 기업이 IT 지출에 있어 우선 투자 분야는 비즈니스 인텔리전스, 웹기반 애플리케이션 및 산업에 특화된 솔루션- IDC 2006년 초 보고서”

## 과거의 BI (Business Intelligence)



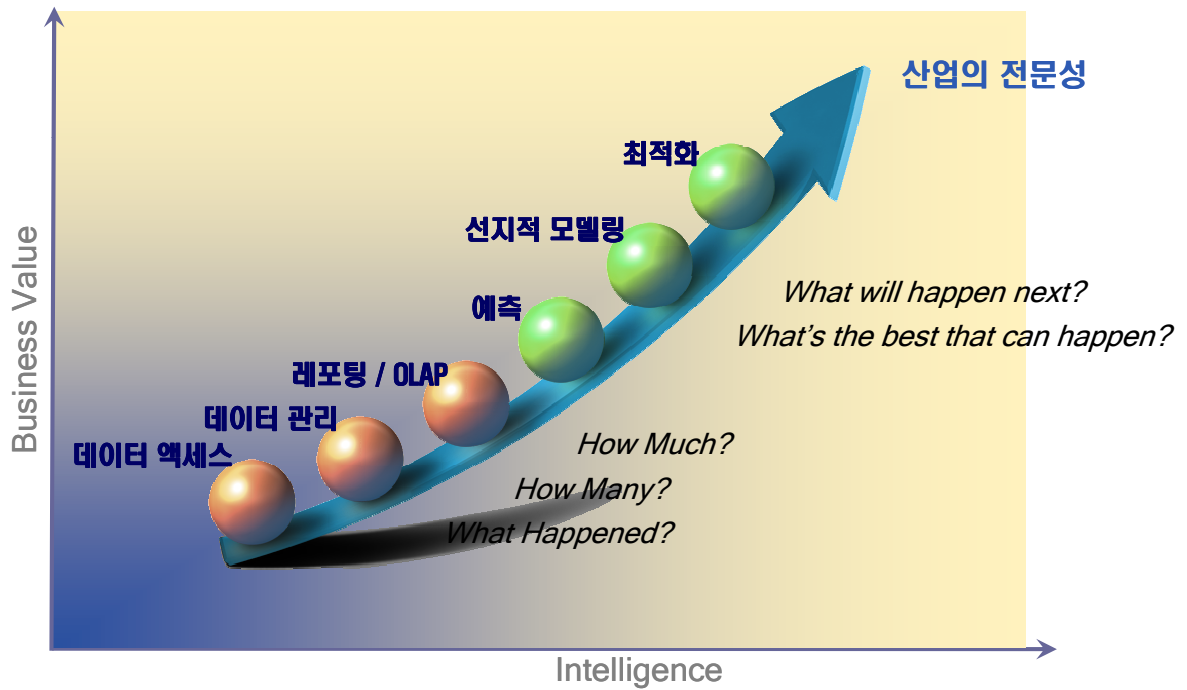
15

## 현재의 BI (Business Intelligence)

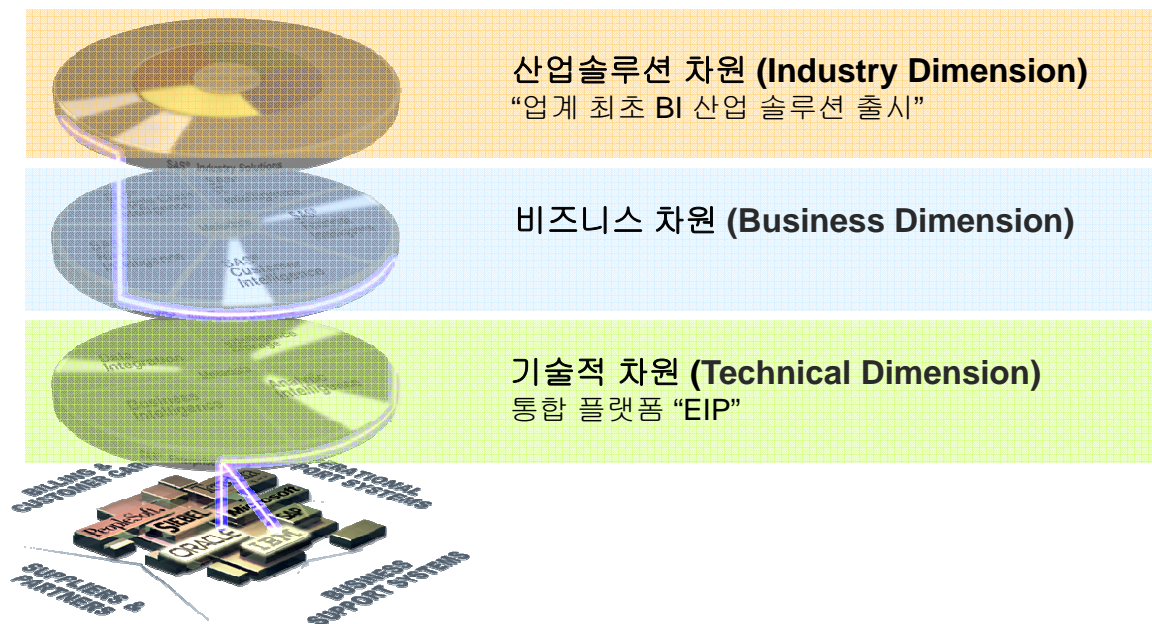


16

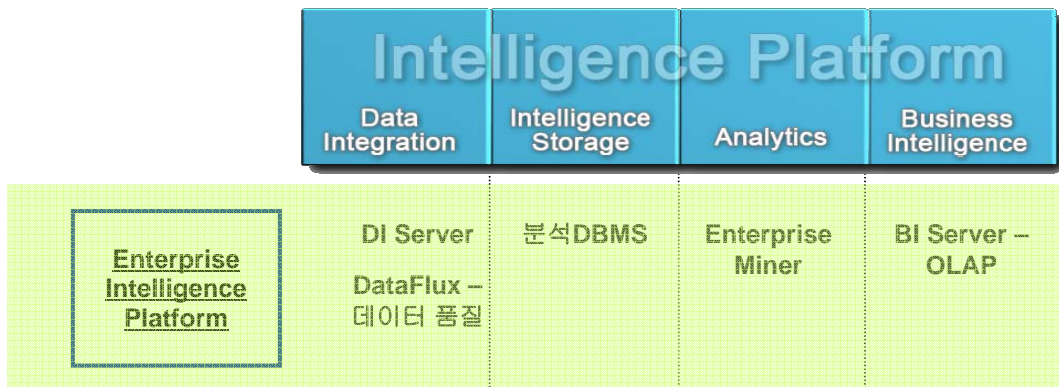
# SAS "Beyond BI"™



## SAS "end-to-end BI" 의 3차원적 접근



## 기술적 차원 (Technical Dimension)

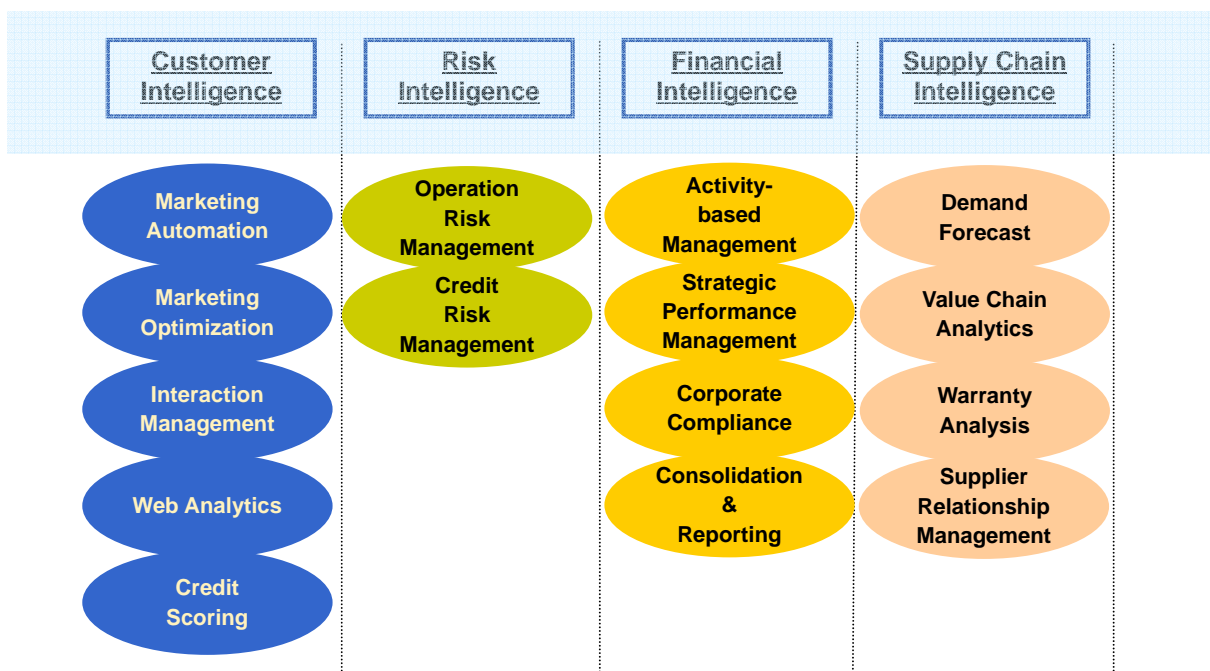


### 통합 플랫폼 EIP

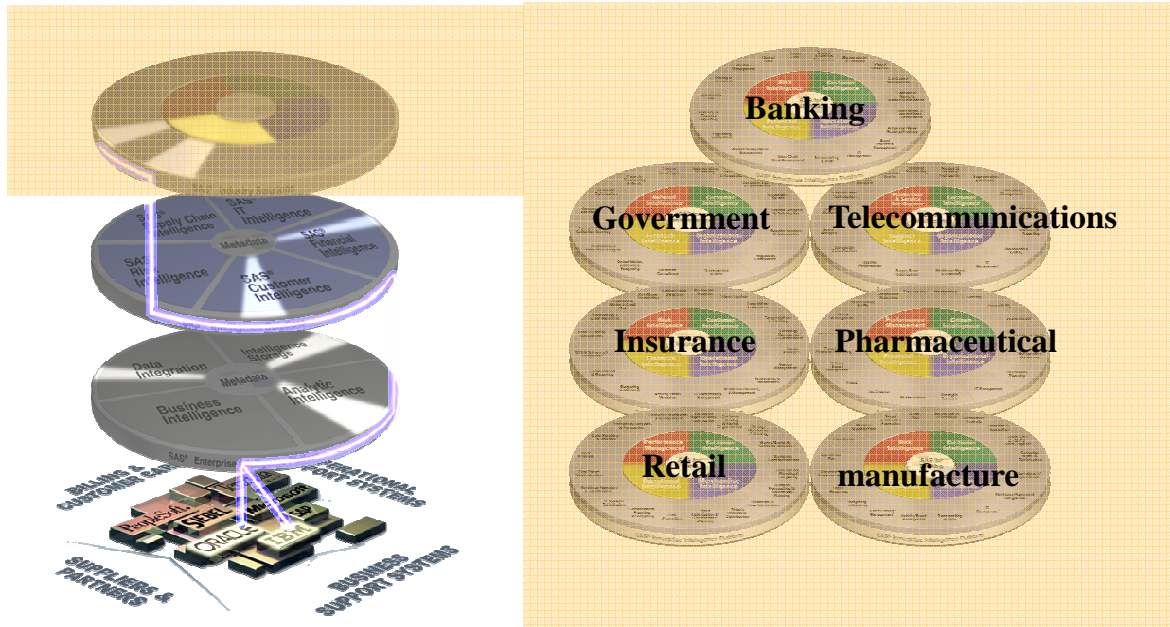
SAS는 Enterprise Datawarehouse를 기반으로 각종 비즈니스 솔루션의 Plug-in이 가능한 Intelligence Solution의 Platform을 제공 -> SAS EIP

이러한 통합 Platform을 바탕으로, 이 기존 시스템간의 data의 통합 및 관리 및 정합성을 보장

## 비즈니스 차원 (Business Dimension)



## 산업솔루션 차원 (Industry Dimension)



21

## Industry-Specific Application Industry Solutions



22

# Chapter 2: SAS Environments

## 2.1 SAS System 구성 및 Workspace

## 2.2 SAS Library 연결 및 사용방법

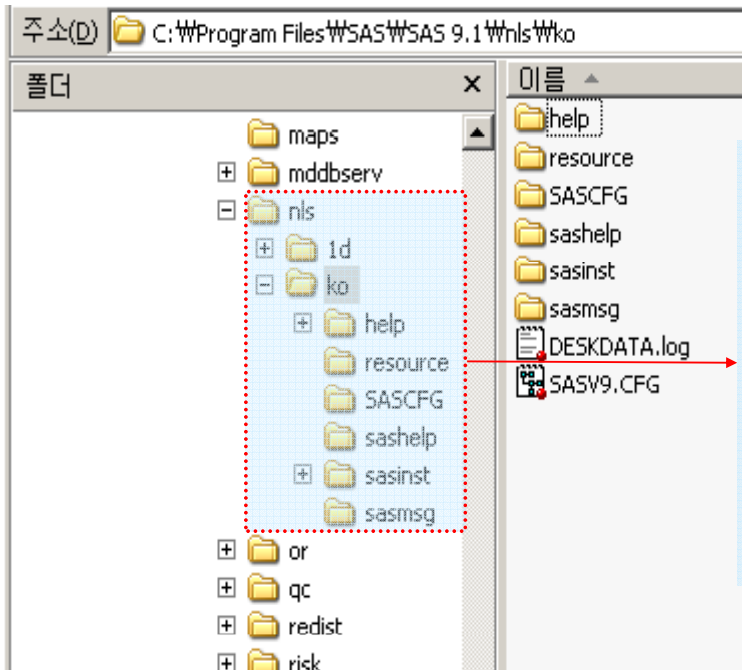
## 2.3 SAS Options 사용 방법

# 2.1 SAS System의 구성 및 Workspace

### Windows 설치 폴더

이름	종류	수정된 날짜
access	파일 폴더	2005-04-26 오후 1:58
af	파일 폴더	2005-04-26 오후 1:58
assist	파일 폴더	2005-04-26 오후 1:58
common	파일 폴더	2005-04-26 오후 1:56
connect	파일 폴더	2005-04-26 오후 1:58
core	파일 폴더	2005-04-26 오후 1:56
creditscrnod	파일 폴더	2005-04-26 오후 2:04
dbcs	파일 폴더	2005-04-26 오후 2:02
dmine	파일 폴더	2005-04-26 오후 2:02
doc	파일 폴더	2005-04-26 오후 1:58
dquality	파일 폴더	2005-04-26 오후 2:01
eis	파일 폴더	2005-04-26 오후 1:58
ets	파일 폴더	2005-04-26 오후 1:58
fsp	파일 폴더	2005-04-26 오후 1:59
gis	파일 폴더	2005-04-26 오후 2:01
graph	파일 폴더	2005-04-26 오후 1:59
hpf	파일 폴더	2005-04-26 오후 2:04
iml	파일 폴더	2005-04-26 오후 1:59
insight	파일 폴더	2005-04-26 오후 1:59
intrnet	파일 폴더	2005-04-26 오후 3:55
inttech	파일 폴더	2005-04-26 오후 1:58
lab	파일 폴더	2005-04-26 오후 1:59
maps	파일 폴더	2005-04-26 오후 2:02
mddbserve	파일 폴더	2005-04-26 오후 1:58
nls	파일 폴더	2005-04-26 오후 2:06
or	파일 폴더	2005-04-26 오후 1:59
qc	파일 폴더	2005-04-26 오후 2:00
redist	파일 폴더	2005-04-26 오후 1:58
risk	파일 폴더	2005-04-26 오후 2:02
sasstest	파일 폴더	2005-04-26 오후 2:08
share	파일 폴더	2005-04-26 오후 2:00
stat	파일 폴더	2005-04-26 오후 2:00
sview	파일 폴더	2005-04-26 오후 2:00

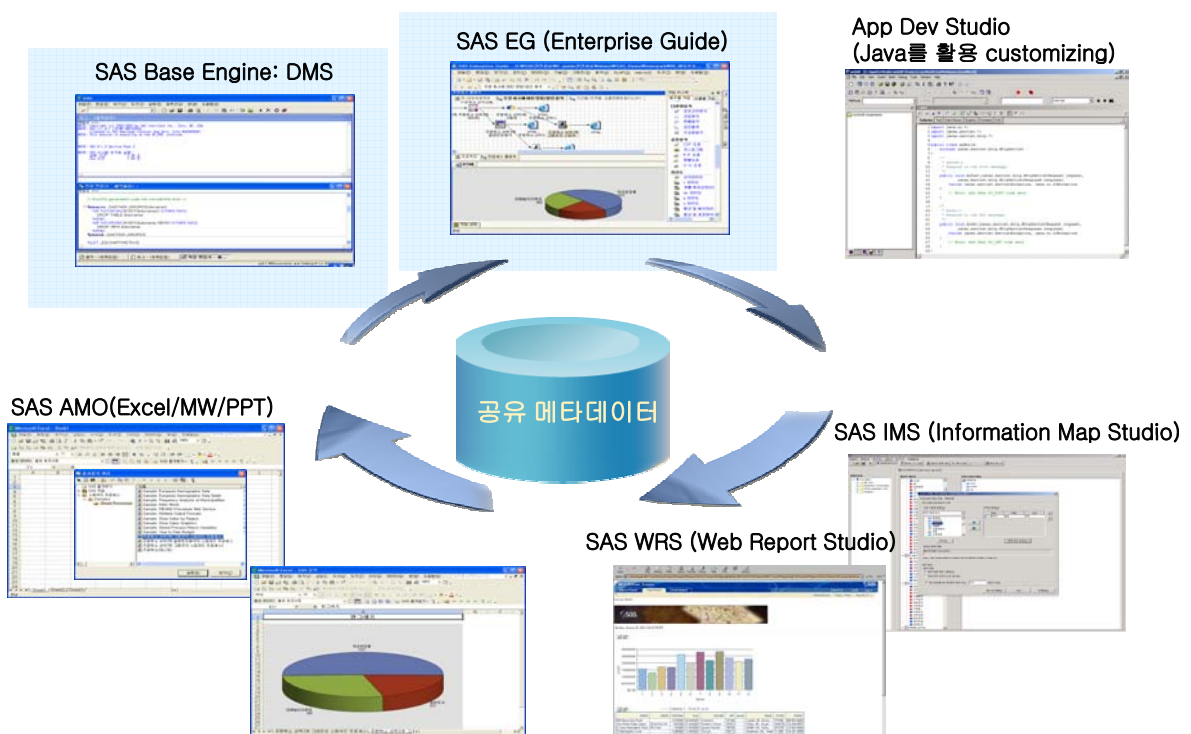
# nls (National Language Support)



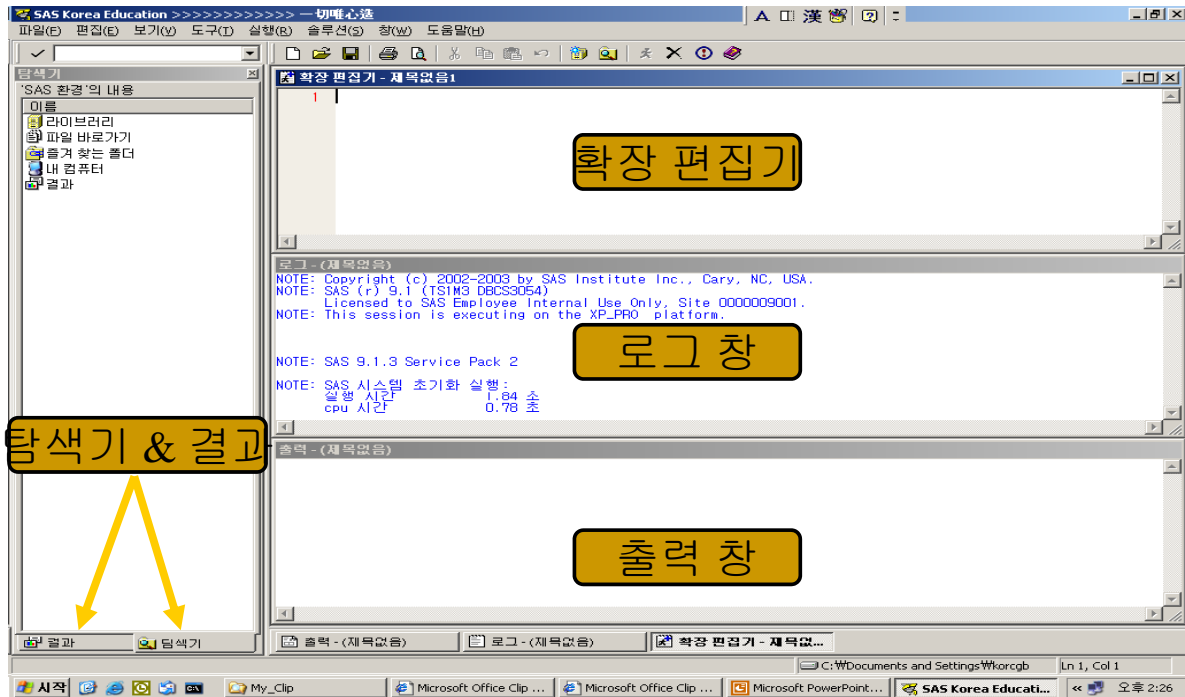
**nls 폴더**  
 SAS System을 여러 가지 환경 (DBCS, Korean, Japanese 등)에서 사용하기 위한 시스템 폴더들을 포함하고 있음.

**nls/ko 폴더**  
 2-Byte 데이터 처리 지원  
 SAS System을 한글 환경에서 사용할 수 있도록 지원하기 위한 파일들을 가지고 있음.

# 다양한 SAS의 Workspace



# SAS Base Engine: DMS (Display Management System) 환경



SAS System은 하나의 메인 창이 여러 개의 창을 포함하고 있는 Multi-Window Application.

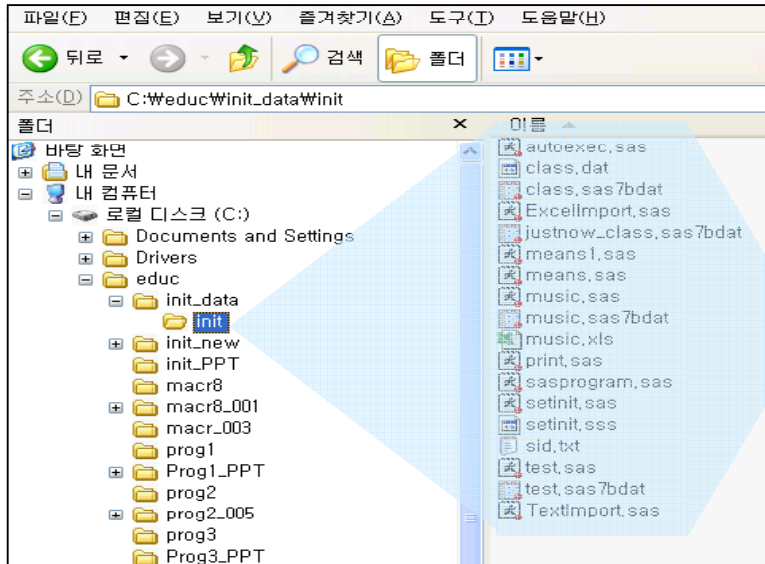
## SAS System의 5개 기본 창.

- 확장 편집기
  - SAS 프로그램을 편집하고 실행할 수 있는 창
- 로그 창
  - SAS 프로그램의 실행과 관련된 여러 가지 정보를 보여주는 창.
  - Note, Warning, Error 등의 정보가 표시됨.
- 출력 창
  - SAS 프로그램 실행 후 생성되는 결과를 보여주는 창
- 탐색기
  - SAS 파일(데이터셋, 카탈로그 등)을 탐색하고 관리하는 작업 수행
- 결과 창
  - SAS 프로그램 실행에 의해 생성된 결과물을 컴포넌트 단위로 탐색하고 관리할 수 있는 기능을 제공함.

## 2.2 SAS Library 연결 및 사용방법

### SAS 라이브러리(Library)란 무엇인가?

- SAS 파일(SAS 데이터 테이블, 카탈로그 등)들의 저장소
- OS의 파일 시스템에 존재하는 폴더 또는 디렉토리를 SAS 시스템에 연결하는 것



#### SAS 파일의 확장자

- SAS Dataset (\*.sas7bdat)
- SAS Catalog (\*.sas7bcatalog)
- SAS Program (\*.sas)

29

### LIBNAME 문장으로 연결(할당)

```
LIBNAME libref 'SAS-data-library' <options>;
```

예) libname **init** v9 "c:\educ\init\_data\init";

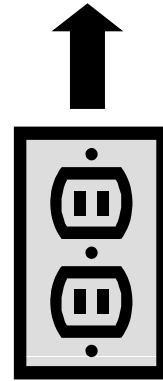
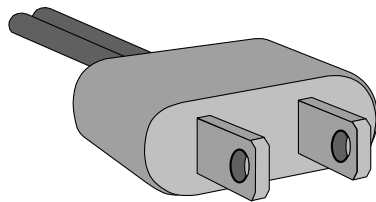
규칙:

- Libref의 명은 반드시 8자리 이내
- 첫 번째 자리는 영문 또는 \_
- 두 번째 자리부터는 영문, 숫자 및 \_

30

LIBNAME 문장을 실행하였을 때, 현재 사용중인 OS 환경의 실제 위치가 SAS 시스템 내에서 지정한 **Libref** 라는 명으로 연결

<b>Windows</b>	'c:\workshop\winsas\prog1'
<b>UNIX</b>	'/users/userid'
<b>z/OS (OS/390)</b>	'userid.prog1.sasdata'



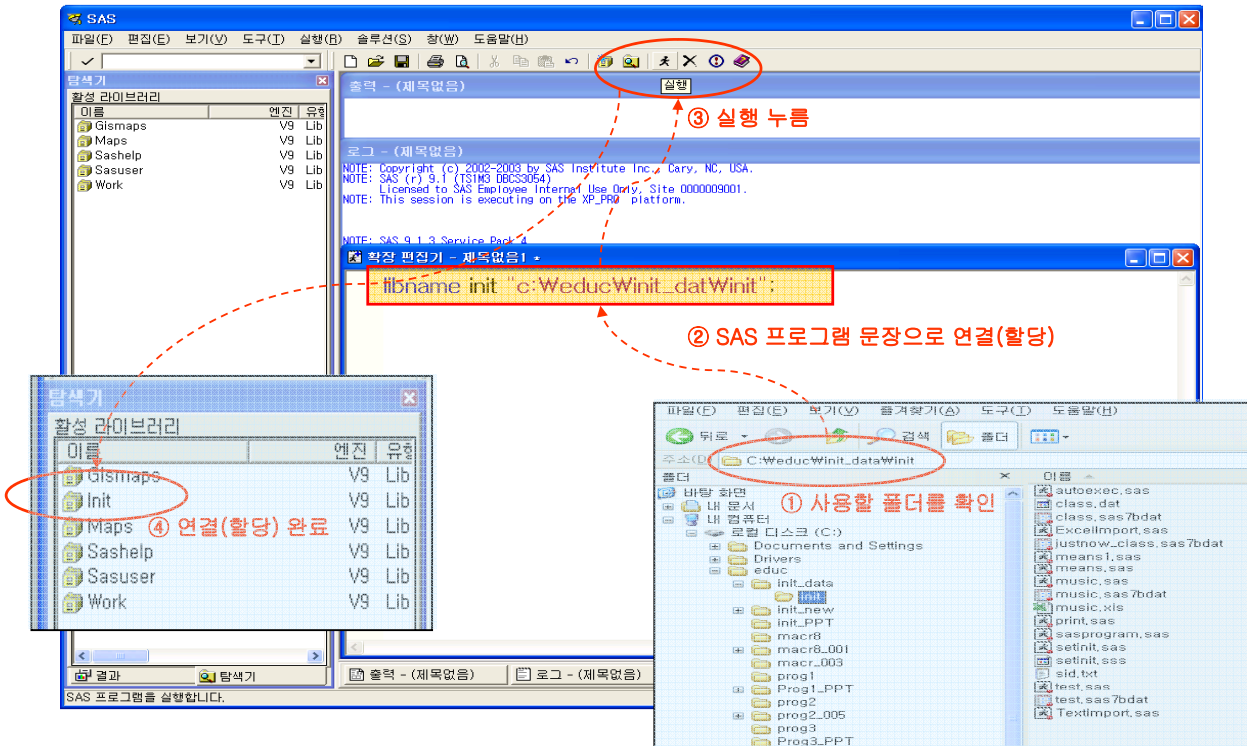
## SAS Library의 사용방법

### 라이브러리(Library)의 종류

- **영구(Permanent) 라이브러리 :**  
SAS System을 종료해도 라이브러리 내의 파일들은 보관됨.  
*예) SASHELP, SASUSER, 모든 사용자 정의 라이브러리*
- **임시(Temporary) 라이브러리 :**  
SAS System을 종료하면 임시 라이브러리 내의 생성된 모든 파일은 자동으로 삭제되어 없어짐.

*예) WORK*

연결된 사용자 라이브러리는 현재 사용중인 SAS System이 종료되면 연결이 자동으로 해제됨. 따라서 SAS System을 재 시작하면 연결(할당)을 다시 해 주어야 함.

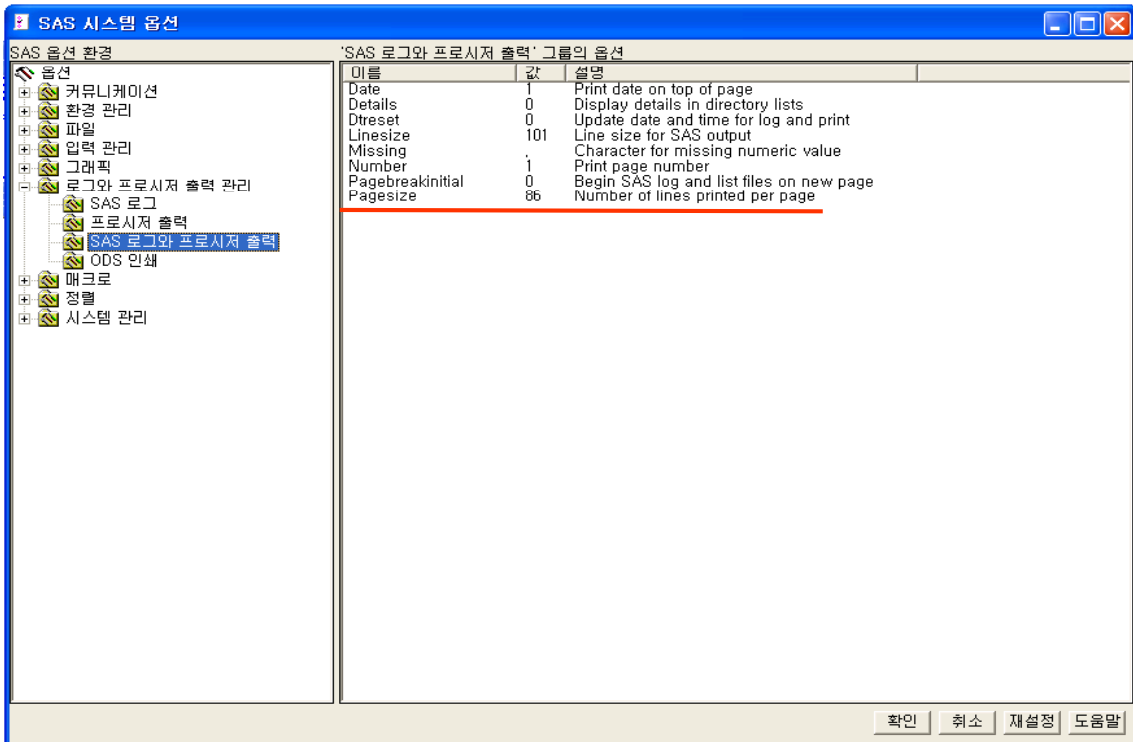


## 2.3 SAS Options 사용방법

SAS System 옵션은 SAS System의 시작에 필요한 초기 환경 설정, 입력 및 출력의 처리등과 관련된 여러 가지 환경을 구성하기 위해 사용되는 값들을 설정하는 역할을 함.

옵션의 종류에 따라서 사용자가 지정하거나 변경할 수 있는 것이 있고, 시스템이 자동적으로 결정하여서 사용자가 지정 또는 변경할 수 없는 것도 있음.

SAS System 옵션 창을 이용하는 방법 : [도구]->[옵션]->[시스템...]을 선택



35



## Exercises

- 사용자 라이브러리의 연결
- OPTIONS 수정(Linesize 및 Pagesize 수정)

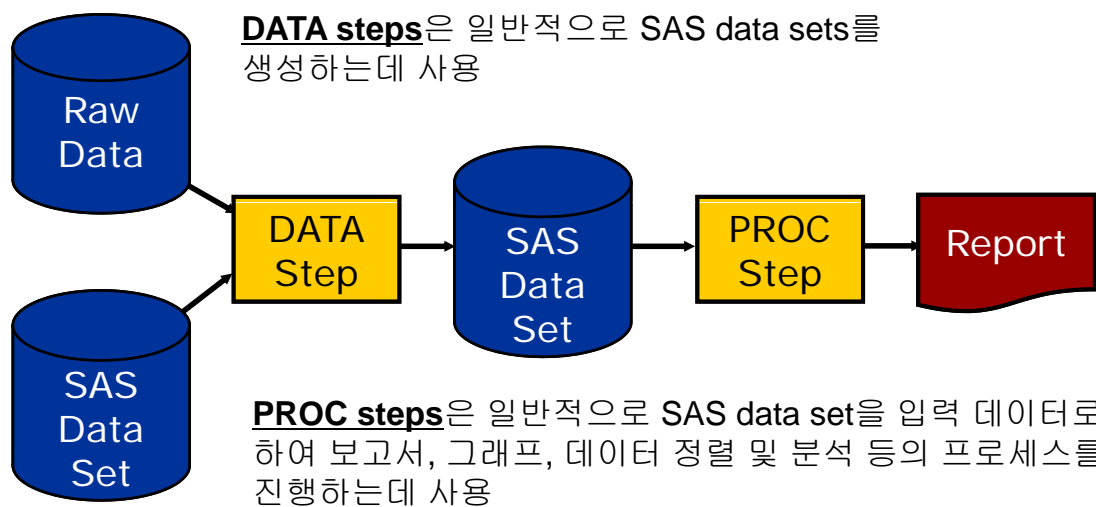
36

# Chapter 3: 데이터 핸들링 및 레포트 출력

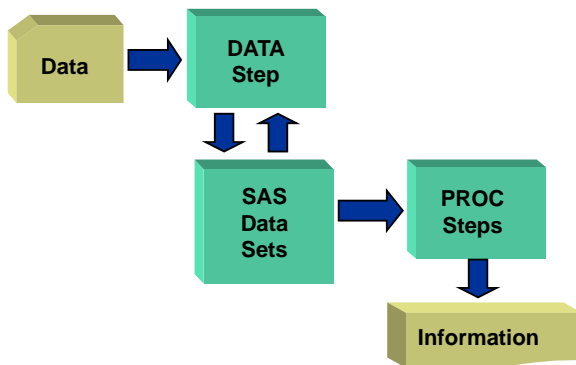
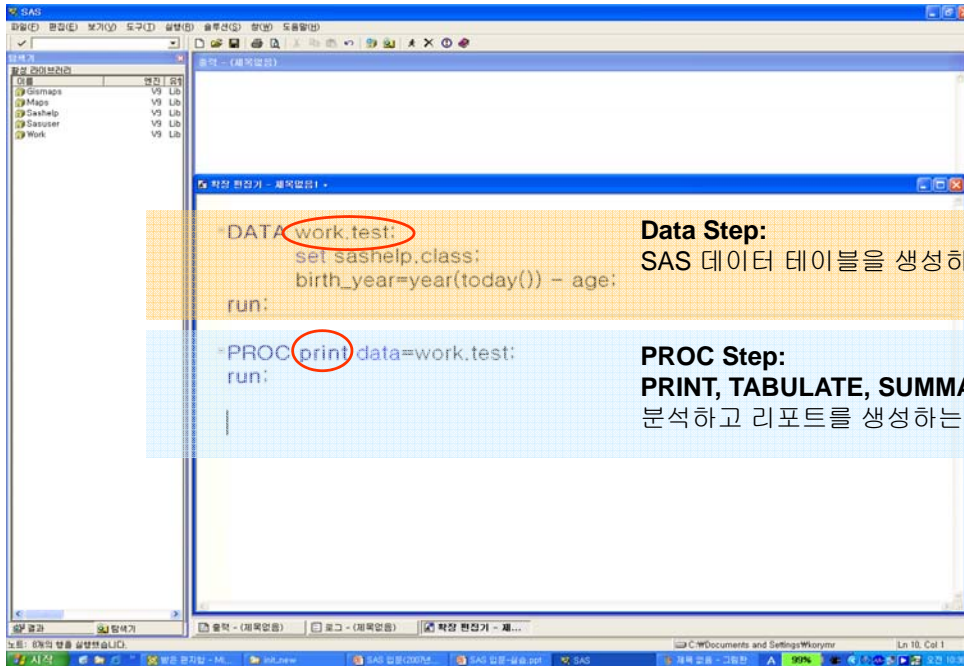
3.1 Introduction to SAS Programs
3.2 SAS data set 구성
3.3 SAS Data set 생성
3.4 SAS Insight를 이용한 데이터 탐색
3.5 Tabulate 프로시저를 이용한 데이터 출력

## 3.1 Introduction to SAS Programs

SAS Program은 순차적인 단계로 진행됨.



- SAS Program의 기본 구성 단위: Data step, Proc step
- 각 Step의 구성 단위: 문장(Statements)



### Step Boundaries

SAS Step은 다음 둘 중 하나로 시작

- **DATA step**
- **PROC step**

```

➔ data work.staff;
   infile 'raw-data-file';
   input LastName $ 1-20 FirstName $ 21-30
         JobTitle $ 36-43 Salary 54-59;
➔ run;
➔ proc print data=work.staff;
➔ proc means data=work.staff;
   class JobTitle;
   var Salary;
➔ run;

```

STEP의 끝은 다음의 문장까지

- **RUN**
- **QUIT**
- 다른 STEP의 시작

## 3.2 SAS data set 구성

Name	Sex	Age	Height	Weight
Alfred	M	14	69	112.5
Alice	F	13	56.5	84
Barbara	F	13	65.3	98
Carol	F	14	62.8	102.5
Henry	M	14	63.5	102.5
James	M	12	57.3	83
Jane	F	12	59.8	84.5
Janet	F	15	62.5	112.5
Jeffrey	M	13	62.5	84
John	M	13	59	99.5
Joyce	F	14	51.3	50.5
Judy	F	14	64.3	90
Louise	F	12	56.3	77
Mary	F	15	66.5	112
Philip	M	16	72	150
Robert	M	12	64.8	128
Ronald	M	15	67	133
Thomas	M	11	57.5	85
William	M	15	66.5	112

**Data Set Table**

Column Variable Field

Row Observation Record

41

SAS data sets은 descriptor portion과 data portion으로 구성

**Descriptor  
Portion**

**General data set information**

\* data set name            \* data set label  
 \* date/time created \* storage information  
 \* number of observations

**Information for each variable**


\* Name    \* Type            \* Length    \* Position  
 \* Format \* Informat \* Label

**Data  
Portion**

42


# 3.3 SAS data set 생성

## 생성방법 1 : 텍스트(Text) 데이터 가져오기



class.dat

➔



class.sas7b.dat

**Data Step**

class.dat - 메모장

```
Alfred,M,14,69,112.5
Alice,F,13,56.5,84
Barbara,F,13,65.3,98
Carol,F,14,62.8,102.5
Henry,M,14,63.5,102.5
James,M,12,57.3,83
Jane,F,12,59.8,84.5
Janet,F,15,62.5,112.5
Jeffrey,M,13,62.5,84
John,M,12,59,99.5
Joyce,F,11,51.3,50.5
Judy,F,14,64.3,90
Louise,F,12,56.3,77
Mary,F,15,66.5,112
Philip,M,16,72,150
Robert,M,12,64.8,128
Ronald,M,15,67,133
Thomas,M,11,57.5,85
William,M,15,66.5,112
```

```
DATA work.class;
length name $ 12 sex $ 1;
infile 'c:\Weduc\Winit\Wclass.dat' dlm='.';
input name sex age height weight;
run;


proc print data=work.class;
run;

proc contents data=work.class;
run;
```

VIEWTABLE: Init.Class


	Name	Sex	Age	Height	Weight
1	Alfred	M	14	69	112.5
2	Alice	F	13	56.5	84
3	Barbara	F	13	65.3	98
4	Carol	F	14	62.8	102.5
5	Henry	M	14	63.5	102.5
6	James	M	12	57.3	83
7	Jane	F	12	59.8	84.5
8	Janet	F	15	62.5	112.5
9	Jeffrey	M	13	62.5	84
10	John	M	12	59	99.5
11	Joyce	F	11	51.3	50.5
12	Judy	F	14	64.3	90
13	Louise	F	12	56.3	77
14	Mary	F	15	66.5	112
15	Philip	M	16	72	150
16	Robert	M	12	64.8	128
17	Ronald	M	15	67	133
18	Thomas	M	11	57.5	85
19	William	M	15	66.5	112

## 생성방법 2: 엑셀(Excel) 데이터 가져오기



music.xls

➔



class.sas7b.dat

**Proc Step**

	A	B	C	D	PROJ
1	YEAR	QTR	MONTH	DATE	
2	1993	1	1	93-01-01	클래
3	1993	1	1	93-01-01	가요
4	1993	1	1	93-01-01	가요
5	1993	1	1	93-01-01	가요
6	1993	1	1	93-01-01	가요
7	1993	1	1	93-01-01	JAZZ
8	1993	1	1	93-01-01	가요
9	1993	1	1	93-01-01	JAZZ
10	1993	1	1	93-01-01	가요
11	1993	1	1	93-01-01	JAZZ
12	1993	1	1	93-01-01	클래
13	1993	1	1	93-01-01	기타
14	1993	1	1	93-01-01	가요

```
PROC import datafile= 'music.xls'
out=init.music
sheet='sheet1$';
run;

PROC print data=init.music;
PROC contents data=init.music;
run;
```

VIEWTABLE: Init.Music

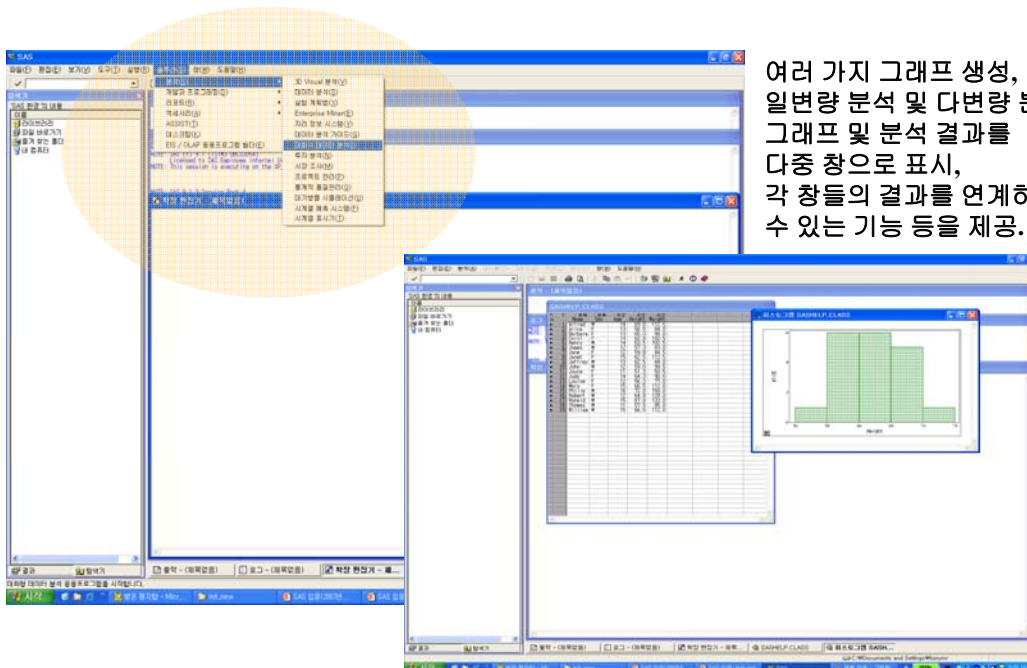
	YEAR	QTR	MONTH	DATE	PRODU
1	1993	1	1	01JAN1993	클래식 Tape
2	1993	1	1	01JAN1993	가요 CD
3	1993	1	1	01JAN1993	가요 CD
4	1993	1	1	01JAN1993	가요 CD
5	1993	1	1	01JAN1993	가요 CD
6	1993	1	1	01JAN1993	JAZZ Tape
7	1993	1	1	01JAN1993	가요 Tape
8	1993	1	1	01JAN1993	JAZZ CD
9	1993	1	1	01JAN1993	가요 Tape
10	1993	1	1	01JAN1993	JAZZ CD
11	1993	1	1	01JAN1993	클래식 Tape
12	1993	1	1	01JAN1993	기타
13	1993	1	1	01JAN1993	가요 CD
14	1993	1	1	01JAN1993	JAZZ CD
15	1993	1	1	01JAN1993	가요 Tape
16	1993	1	1	01JAN1993	JAZZ Tape

### 생성방법 3: Import Wizard

- 여러 가지 데이터 소스로부터 SAS 데이터셋을 생성하는 변환 과정을 자동화된 마법사 형식으로 제공하는 어플리케이션
- [파일]->[데이터 가져오기...]를 선택하면 마법사가 실행됨.
- 가져올 수 있는 데이터 유형
  - Microsoft Excel
  - Microsoft Access
  - Lotus File
  - dBase File
  - Text File(Delimited, Tab delimited, CSV)
  - Text File(User-Defined Format)

## 3.4 SAS Insight를 이용한 데이터 탐색

데이터 탐색과 분석을 Point & Click 방식으로 수행할 수 있는 GUI 환경을 제공하는 제품.



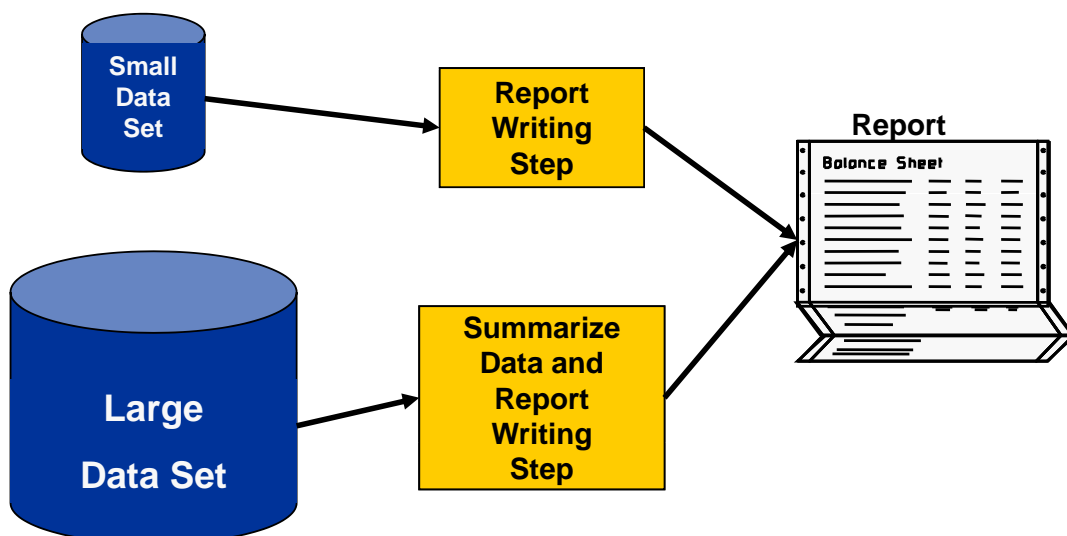
여러 가지 그래프 생성, 일변량 분석 및 다변량 분석 수행, 그래프 및 분석 결과를 다중 창으로 표시, 각 창들의 결과를 연계하여 분석할 수 있는 기능 등을 제공.

## 제공되는 기능들

- 데이터 탐색
  - ✓ 산점도, 히스토그램, 상자도표, 등고선 도표, 3차원 도표 생성
  - ✓ 데이터 검색, 정렬, 편집
  - ✓ 특정 관측치를 그래프 또는 분석에 제외할 수 있음.
- 일변량 숫자형 데이터의 분포 분석
  - ✓ 기술 통계량 계산, Q-Q 도표 생성, 모수적 분포 적합, 분포에 대한 적합도 검정
- 반응 변수와 설명 변수 간의 관계에 대한 모형 적합
  - ✓ 일반 선형 모형(분산분석, 공분산 분석) 적합
  - ✓ 일반화 선형 모형(로지스틱 회귀, 포아송 회귀) 적합
- 변수들간의 관계 분석
  - ✓ 상관 분석, 주성분 분석, 정준 상관 분석

## 3.5 프로시저를 이용한 데이터 출력

### Summary Reports



## PROC TABULATE Output

Average Salary for Cary and Frankfurt

	Location		All
	CARY	FRANKFURT	
	Salary	Salary	Salary
	Mean	Mean	Mean
JobCode			
FLTAT1	\$26,200	\$25,000	\$25,667
FLTAT2	\$35,000	\$36,200	\$35,500
FLTAT3	\$43,400	\$44,667	\$43,875
All	\$34,882	\$34,583	\$34,759

49

## PROC TABULATE

TABULATE step에 사용되는 문장

```
PROC TABULATE DATA=SAS-data-set <options>;  
  CLASS class-variables;  
  VAR analysis-variables;  
  TABLE page-expression,  
         row-expression,  
         column-expression </ option(s)>;  
RUN;
```

50

```

PROC TABULATE data=work.test;
  class age sex;
  var height;
  table age, sex * (height * mean)*f=10.2;
run;

```

Age	Sex	
	F	M
	Height	Height
	Mean	Mean
11	51.30	57.50
12	58.05	60.37
13	60.90	62.50
14	63.55	66.25
15	64.50	66.75
16	.	72.00

51



### 3. Exercises

- 1) SAS Insight를 이용한 탐색적 데이터 분석
- 2) TABULATE Procedure 실행
  - 연령별 성별 몸무게의 합계, 인원수, 평균을 산출

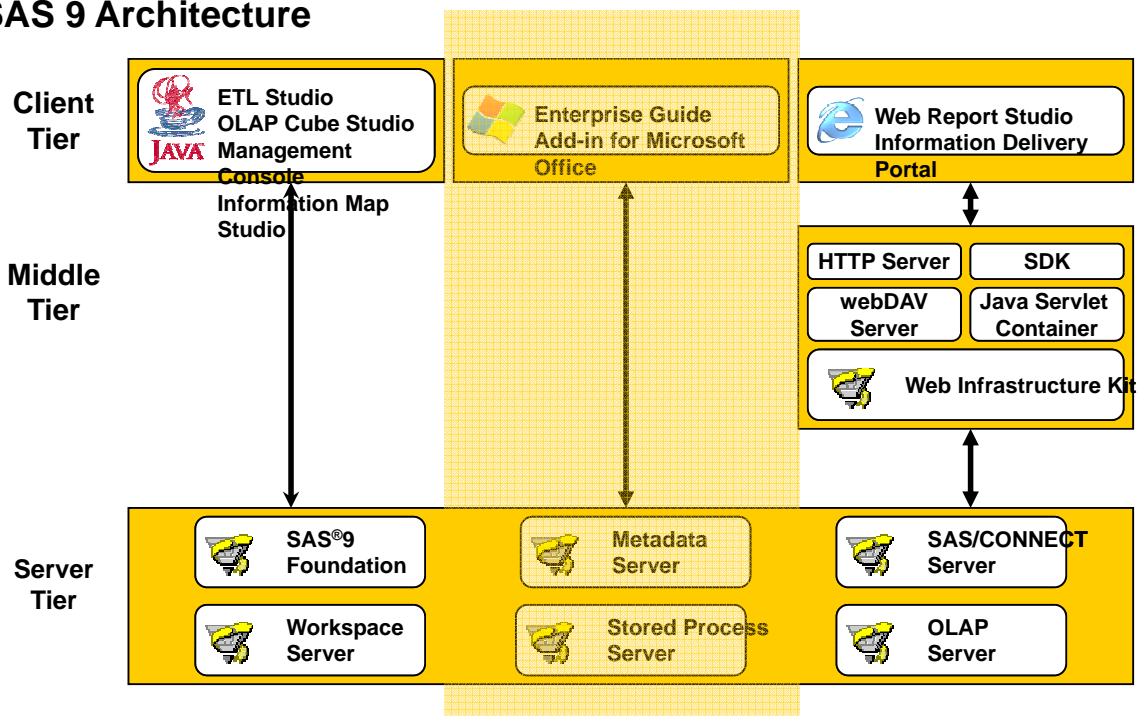
52

# Chapter 4: SAS 9

4.1 Introduction to SAS 9
4.2 SAS EG (Enterprise Guide)
4.3 SAS AMO (Add-In for Microsoft Office)

## 4.1 Introduction to SAS 9

### SAS 9 Architecture



## Client Tier

사용자가 직접 보면서 작업을 하는 컴퓨터는 Client Tier에 속하게 된다.

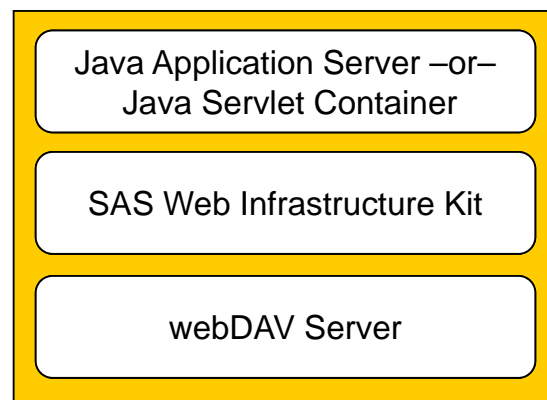
SAS 클라이언트 애플리케이션의 3가지 유형

자바 클라이언트	Java run-time environment (JRE) 환경 상에서 실행되는 애플리케이션.
윈도우 클라이언트	Microsoft Windows 환경 상에서 실행되는 애플리케이션.
브라우저 클라이언트	Middle Tier의 Java Application Server 또는 Servlet Container와 연결 되어서 웹 브라우저 상에서 실행되는 애플리케이션. 웹 브라우저만 클라이언트 머신에 설치됨.

## Middle Tier

웹 애플리케이션이 위치하고 실행되는 컴퓨터는 Middle Tier를 구성함. 또한 이들 애플리케이션들이 실행되는데 필요한 다음의 인프라스트럭처를 가짐.

- Java Application Server 또는 the Java Servlet Container
- SAS Web Infrastructure kit (WIK)
- webDAV server.



## Server Tier

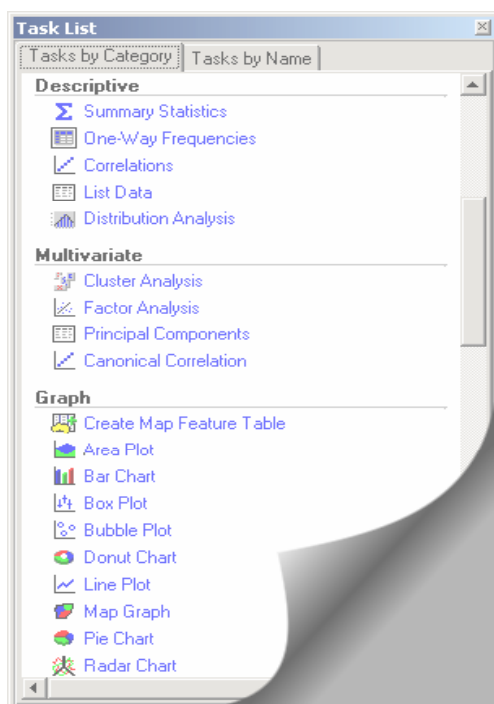
BI 툴들이 액세스하는 하나 이상의 SAS 서버가 설치되어 있는 컴퓨터들.

SAS 서버의 여러 가지 유형:

<b>Metadata Server</b>	전사 환경에 있는 SAS 애플리케이션들에 의해 사용될 메타데이터를 관리하고 클라이언트의 요청에 필요한 메타데이터를 전송함.
<b>Workspace Server</b>	클라이언트 애플리케이션을 대신하여 SAS 코드를 실행하고 결과를 전송함.
<b>Stored Process Server</b>	SAS Stored Processes를 실행하고, 결과를 전송하는 작업을 수행함.
<b>OLAP Server</b>	OLAP 클라이언트들에게 데이터로부터 생성된 큐브에 대한 액세스 제공

....

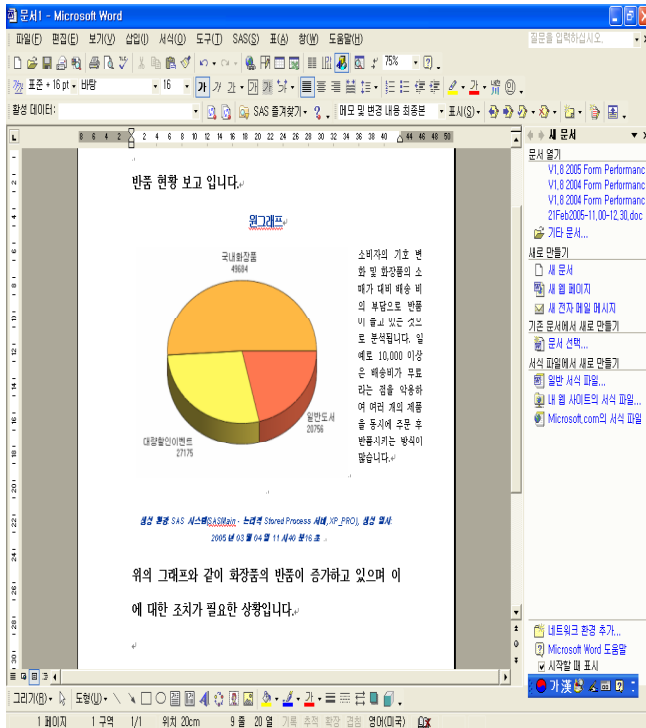
## 4.2 SAS® Enterprise Guide



60개 이상의 레포팅, 그래픽, 분석 TASK

SAS® Enterprise Guide는 SAS 코드를 직접 사용하여 실행하는 것 또는 도움말을 찾는 것 보다는 **GUI(Graphic User Interface)** 를 통하여 많은 **tasks** 를 **speeds** 하게 처리

## 4.3 SAS® Add-In For Microsoft Office



- SAS 서버가 액세스할 수 있는 모든 데이터 소스를 포인트-클릭 방법을 사용하여 Excel 내에서 직접 볼 수 있다.
- Excel이 탐색할 수 있는 한계를 넘어서는 데이터를 탐색할 수 있다- 64K Cell 이상.
- SAS의 powerful한 모든 분석 기능을 excel에서 사용할 수 있음 (forecasting, simulation 가능)
- 다차원 분석 가능(Pivot table로 자동 conversion 해 줌)
- Enterprise Guide에서 만든 프로그램인 stored process를 excel에서 수행 할 수 있음

59

## Exercises

- SAS EG를 이용하여 통계량을 산출 및 그래프 작성
- SAS AMO (Add-in for Microsoft Office) 실행

60